



THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

Applications of higher-order quantum maps

Kaumudibikash Goswami
Master of Technology



[ORCID: 0000-0002-6603-8082](https://orcid.org/0000-0002-6603-8082)

*A thesis submitted for the degree of Doctor of Philosophy at
The University of Queensland in 2021
School of Mathematics and Physics*

Abstract

Higher-order maps account for the most general transformations in quantum mechanics—they transform input quantum maps to output quantum maps. In this thesis, I focus on the application of such higher-order maps representing two important processes: (1) a particular type of non-Markovian quantum dynamics, and (2) indefinite causal order.

First, I will introduce our work on the non-Markovian evolution of quantum states. Such a situation occurs when the past system-environment interaction leads the environment to memorize the information about the system. The measure of the non-Markovianity is given by the temporal correlation among the operations on the system. One way to estimate such correlation is to employ a full tomography of the process matrix, which is computationally intractable. Instead, we utilize machine learning algorithms without using tomographically complete measurement, to estimate the measure of non-Markovianity. We apply our model to a dataset obtained from an optical experiment and show that we are able to predict the measure of non-Markovianity with 90% accuracy.

I then introduce various aspects of indefinite causal order. Ordinary experience points only to a definite causal structure, but quantum mechanics admits a more general causal structure that allows for the superposition of different causal orders. That means that the statements, "event A is in the causal past of event B", and "event B is in the causal past of event A" can both be true. The quantum switch can physically realise such an indefinite causal structure where a control quantum system controls the order of operations acting on a target quantum system. Superposition in the control leads to the superposition of the order. We experimentally realise a quantum switch with the control qubit being the polarisation of light and the target qubit being the transverse spatial mode of light. We first verify our quantum switch by measuring a 'causal witness'. We then demonstrate an advantage of the quantum switch for classical communication.

Finally, I present our work on classical communication through the most general quantum causal structures, represented by the process matrix. We formulate different classical capacities of a bipartite process matrix. We also study one-way and bidirectional communication protocols through such a process and establish the relevant communication bounds. These enable further analyses of higher-order maps that take advantage of the full suite of processes that quantum mechanics allows.

Declaration by author

This thesis is composed of my original work, and contains no material previously published or written by another person except where due reference has been made in the text. I have clearly stated the contribution by others to jointly-authored works that I have included in my thesis.

I have clearly stated the contribution of others to my thesis as a whole, including statistical assistance, survey design, data analysis, significant technical procedures, professional editorial advice, financial support and any other original research work used or reported in my thesis. The content of my thesis is the result of work I have carried out since the commencement of my higher degree by research candidature and does not include a substantial part of work that has been submitted to qualify for the award of any other degree or diploma in any university or other tertiary institution. I have clearly stated which parts of my thesis, if any, have been submitted to qualify for another award.

I acknowledge that an electronic copy of my thesis must be lodged with the University Library and, subject to the policy and procedures of The University of Queensland, the thesis be made available for research and study in accordance with the Copyright Act 1968 unless a period of embargo has been approved by the Dean of the Graduate School.

I acknowledge that copyright of all material contained in my thesis resides with the copyright holder(s) of that material. Where appropriate I have obtained copyright permission from the copyright holder to reproduce material in this thesis and have sought permission from co-authors for any jointly authored works included in the thesis.

Publications included in this thesis

1. **K. Goswami**, C. Giarmatzi, C. Monterola, S. Shrapnel, J. Romero, F. Costa, [Experimental characterisation of a non-Markovian quantum process](#), *Physical Review A* **104**, 022432 (2021). Open access via [arXiv:2102.01327](#).
2. **K. Goswami**, and F. Costa, [Classical communication through quantum causal structures](#), *Physical Review A* **103**, 042606 (2021). Open access via [arXiv:2007.05051](#).
3. **K. Goswami**, Y. Cao, G. A. Paz-Silva, J. Romero, and A. G. White, [Increasing communication capacity via superposition of order](#), *Physical Review Research* **2**, 033292 (2020). Open access via [arXiv:1807.07383](#).
4. **K. Goswami**, C. Giarmatzi, M. Kewming, F. Costa, C. Branciard, J. Romero, and A. G. White, [Indefinite Causal Order in a Quantum Switch](#), *Physical Review Letters* **121**, 090503 (2018). Open access via [arXiv:1803.04302](#).

Submitted manuscripts included in this thesis

No manuscripts submitted for publication

Other publications during candidature

Review article

1. **K. Goswami**, J. Romero, [Experiments on quantum causality](#), *AVS Quantum Science* **2**, 037101 (2020). Open access via [arXiv:2009.00515](#).
Review article commissioned by the Editor.

Non-refereed publications

1. J. Romero, **K. Goswami**, C. Giarmatzi, F. Costa, C. Branciard, A. G. White, [Spatial modes for testing indefinite causal order](#), *Complex Light and Optical Forces XII*; 10549, 1054908 (2018).

Conferences and presentations

1. **K. Goswami**, [Indefinite causal order](#), presented a talk at *Quantum Software and Information (QSI) online seminar* (November, 2020).
2. **K. Goswami**, [Exploring indefinite causal order](#), presented a talk at *Quantum Foundations, Technology and Applications (QFTA)* (October, 2019).

3. **K. Goswami**, J. Romero, A. White, [Communicating via ignorance](#), *APS March Meeting Abstracts*, R28.002 (2019).

Contributions by others to the thesis

No contributions by others.

Statement of parts of the thesis submitted to qualify for the award of another degree

No works submitted towards another degree have been included in this thesis.

Research involving human or animal subjects

No animal or human subjects were involved in this research.

Acknowledgments

I express my heartfelt gratitude to my advisory team for providing me invaluable support throughout my Ph.D. I am indebted to Andrew White, with whom any kind of discussion is a pleasure to take part. His inspirational guidance and invaluable professional advice kindled the right spirit in me to continue my research. My sincere regard to Jacqui Romero who took me under her wing and taught me the nuances of optical experiments. It has been a great pleasure working with you on the beautiful causality experiments. I am grateful to Fabio Costa for introducing the process formalism to me, I learnt a lot from your intuitive approach. It is always a pleasure to discuss physics with you. I thank Till Weinhold and Marcelo Almeida for the warm welcome upon my arrival in Australia. Not only you guys helped me in the lab whenever I needed, but you also gave me several cherishable memories—from playing cards at the Queenstown airport to discussing martial arts over beer.

My sincere gratitude to Murray Kane, his prompt actions helped me navigate through the unfortunately long waiting period for my visa. If it were not for him, probably I would miss the excellent research opportunity. I am extremely grateful to Shayan Garani Srinivasa. He gave me a research position at IISC, Bangalore during the time I almost lost hope for an Australian visa. The position was not only a big emotional and financial support, it allowed me to learn basic Quantum Shannon theory which I went on using in my Ph.D.

I would like to thank fellow Ph.D. students: Azzwa, Jihun, Rob, Michael, Ming, Leo, Mahdi, Markus, Raphael, Lewis; all of you created a great research atmosphere. Especially thanks to Michael for helping me with the real estate agency, to Azzwa for all the food you cooked, and to Jihun for all the stay-over and movie-nights.

A special acknowledgement to the UQ EQUUS team: Angie, Lisa, Joyce, and Tara. I am grateful for all the administrative works that helped me to attend international conferences without any hassle. It is your constant efforts that help all the students to have an excellent academic atmosphere.

I am grateful to my friends Niraj, Uttam, Priya, Atul, and Ali for all fun and light-hearted conversations. All of you helped me to live in the moment and fight stress. I thank Kaushani and my parents for the unwavering support. Nothing would have been possible without my mother's unending encouragement and my father's pleasant smile; they provided an incredible source of inspiration in my life.

Financial support

This research was supported by an Australian Government Research Training Program Scholarship.

Keywords

quantum information, process matrix, transverse spatial mode of light, indefinite causal order, quantum switch, non-Markovianity.

Australian and New Zealand Standard Research Classifications (ANZSRC)

ANZSRC code: 020603, Quantum Information, Computation and Communication, 50%

ANZSRC code: 020604, Quantum Optics, 50%

Fields of Research (FoR) Classification

FoR code: 0205, Optical Physics, 50%

FoR code: 0206, Quantum Physics, 50%

This thesis is dedicated to my late father.

Contents

Abstract	ii
Contents	x
List of Figures	xiii
List of Tables	xv
1 Introduction	1
2 Conceptual Background	5
2.1 Elementary objects in Quantum Mechanics	5
2.1.1 Quantum state	5
2.1.2 Joint quantum state	7
2.1.3 Quantum Channel	9
2.1.4 Quantum measurement	11
2.1.5 Choi-Jamiołkowski (CJ) isomorphism	14
2.2 Process Framework	16
2.2.1 Examples	19
2.3 Background on information theory	20
2.3.1 Classical Shannon theory	21
2.3.2 Quantum Shannon theory	25
2.4 Experimental Elements	30
2.4.1 Polarisation of light	30
2.4.2 Transverse spatial mode of light	33
3 Characterising a non-Markovian quantum process	39
3.1 Introduction	39
3.1.1 Initially separable system-environment	40
3.1.2 Initially correlated system-environment	41
3.1.3 Operational difficulty with dynamical maps	42
3.1.4 Resolving the initial correlation problem	42

3.2	Our work	43
3.3	Theory	44
3.3.1	Formulation of quantum processes	44
3.3.2	Our non-Markovian process	45
3.3.3	Generating data and labels	47
3.4	Experiment	49
3.5	Polynomial Regression	50
3.6	Conclusion	53
4	Indefinite causal order in a quantum switch	55
4.1	Introduction	55
4.2	Quantum switch	56
4.3	Causal witness	58
4.3.1	Taking experimental imperfections into account	59
4.4	Experiment	60
4.4.1	Implementation of unitaries	63
4.4.2	Polarisation measurement at the output	64
4.4.3	Result	65
4.5	Conclusion	66
4.6	Author contribution	67
5	Communicating via ignorance	69
5.1	Introduction	69
5.2	Background	70
5.2.1	Estimating the Stokes parameter of the quantum switch	72
5.3	Experiment	73
5.3.1	Effects of experimental imperfections	76
5.4	Results	76
5.4.1	Two depolarizing channels	76
5.4.2	Depolarizing and dephasing channel	78
5.4.3	More general combinations of depolarizing and dephasing channels	79
5.4.4	Bit flip and bit-phase-flip channels	80
5.4.5	Holevo capacity from path superposition	81
5.5	Conclusion	84
5.6	Author contribution	85
6	Classical communication through quantum causal structures	87
6.1	Introduction	87
6.2	Classical communication through a quantum channel	88
6.3	The Process framework	90

6.4	Motivation	92
6.4.1	Causal inequality	93
6.5	One-directional communication through an indefinitely ordered process	94
6.5.1	Asymptotic setting	95
6.5.2	Classical capacities of a quantum process	98
6.5.3	Bounds on the classical capacities of a quantum process	100
6.6	Broadcast communication	103
6.7	Conclusion	106
6.8	Author contribution	107
7	Conclusion	109
	Bibliography	111

List of Figures

2.1	Bloch sphere representation of a qubit.	7
2.2	Three types of dilations.	13
2.3	Two versions of CJ representation.	15
2.4	Example of two quantum processes.	17
2.5	Linking local operations to processes.	18
2.6	Quantum state encoded in polarisation.	32
2.7	Modal decomposition.	34
2.8	Quantum information using spatial mode	36
3.1	Quantum processes representing open quantum dynamics.	43
3.2	A process based on a specific instance of unitary operations U_i and U_j	46
3.3	Experimental setup.	48
3.4	Scatter plots for the second degree polynomial regression on the experimental dataset.	50
4.1	Space-time diagram of the <i>quantum switch</i>	56
4.2	Causal witness S	58
4.3	Experimental schematic.	62
4.4	Preparation method for our input state.	62
4.5	Top view of the setup for realising the unitary operations \hat{A} and \hat{B} using a set of special inverting prisms R, and pairs of cylindrical lenses C.	63
4.6	Spatial transformations.	63
4.7	Stokes parameters $\langle \hat{X} \rangle_{\hat{A}, \hat{B}}$ obtained by measuring the polarisation of the output control qubit in the diagonal basis.	65
5.1	Classical communication through a quantum switch.	70
5.2	Predicted spatial mode of target qubit outputs.	75
5.3	Schematic of quantum switch.	75
5.4	Logarithm of Holevo capacity χ , of two identical depolarizing channels, vs depolarizing channel strength, q	77
5.5	Logarithm of Holevo capacity χ , where one channel is a depolarizing channel of varying strength q , and the other is a σ_3 channel.	78

5.6	Predicted logarithm of Holevo capacity χ for the channels in Eq. 5.17 vs. depolarization noise strength, q	80
5.7	Logarithm of Holevo capacity χ , where one channel is a depolarizing channel of varying strength q , and the other one is a dephasing channel of strength p	81
5.8	Predicted Holevo capacity χ for the channels in Eq. 5.19 vs noise parameter, p	82
5.9	Experimental Holevo capacity χ for $p \geq 0.5$ region.	82
6.1	One-way communication protocol	92
6.2	Different communication setups	93
6.3	Pictorial depiction of Lemma 1	97
6.4	Pictorial depiction of Theorem 1.	98
6.5	Encoding and decoding schemes.	101

List of Tables

2.1	Angles for wave plates to measure polarisation components.	32
3.1	Pairs of q and R to model our joint pmf.	48
3.2	Angles for motorised wave plates to implement U_i and U_j	49
3.3	polynomial regression on the experimental data of size 1000.	51
3.4	k-fold cross validation on our experimental dataset.	51
3.5	Second degree polynomial trained only on the experimental data of size 1000.	52
3.6	Different machine learning algorithms trained on the experimental data of size 1000.	52
4.1	Data for the different combinations of unitary operations \hat{A} and \hat{B}	61
4.2	Phases and angles for the unitary operations realised in our experiment.	64
5.1	Data used to calculate the Holevo capacity.	74
5.2	Phases and angles for the unitary operations realized in our experiment.	76

If all this damned quantum jumping were really here to stay, I should be sorry, I should be sorry I ever got involved with quantum theory.

Erwin Schrödinger,
Physics and Beyond: Encounters and Conversations

Chapter 1

Introduction

The most general quantum transformation maps an input quantum operation to an output quantum operation. My research is focused on two specific applications of such higher-order maps: a non-Markovian quantum evolution and indefinite causal order.

We observe a non-Markovian dynamics when interaction between the system and the environment results in an environment retaining the past information about the system, i.e., a quantum evolution with memory. Conventionally, a non-Markovian evolution is characterized by dynamical maps. Such characterisation poses several challenges. For example, an evolution with an initial system-environment correlation results in dynamical maps which lack operational interpretations; we are forced to relax positivity, linearity, or consistency of the dynamical maps. In contrast, since the higher-order map transforms any input completely positive (CP) map to a valid output CP map, it gives a sound operational representation of non-Markovian processes. In a non-Markovian dynamics, we encode all information about the system-environment correlations in the higher-order process. Any local operation performed on the system is an input of the process.

The higher-order map also helps to represent a process exhibiting indefinite causal order. The notion of a fixed causal order is predominant in our daily experience; an event A either occurs before or after an event B . Surprisingly, quantum mechanics allows superposition of different causal orders—the event A can both be in the causal past and causal future of event B . Such exotic causal structures have relevance in quantum gravity as well as applications in several computation and information-processing tasks. A quantum switch is a physically realisable example of a device that demonstrates indefinite causal order. In a quantum switch, the order of operations acting on a target quantum system is coherently controlled by a control quantum system. Superposition in the control leads to superposition in the causal order.

Although the quantum switch presents interesting applications in various information-processing tasks, it cannot violate the causal inequality—an inequality for causal structures analogue to Bell's inequality. Interestingly, certain indefinitely ordered processes violate the causal inequality. Although the physical implementation of such processes is not known yet, it is interesting to explore different properties of such general indefinitely ordered processes.

Chapter summary

My thesis is arranged as follows: in Chapter 2, I first introduce elementary quantum objects (states, transformation, and measurement). Then I introduce properties of the higher-order processes. Next I introduce both classical and quantum information theoretic concepts. Finally, I present two experimental platforms for quantum information processing: polarisation and transverse spatial mode of light.

Having developed the preliminary concepts in the Chapter 2, Chapter 3 discusses my experiment on a non-Markovian process. Although a higher-order map successfully represents a non-Markovian process, a full characterisation of such dynamics requires tomography of the higher-order map, which is both computationally and experimentally difficult. We propose a more efficient solution. We employ machine learning models to estimate the amount of non-Markovianity—as quantified by an information-theoretic measure—with tomographically incomplete measurement. We test our model on a quantum optical experiment, and we are able to predict the non-Markovianity measure with 90% accuracy. Our experiment paves the way for efficient detection of non-Markovian noise appearing in large-scale quantum computers.

In Chapter 4, I present our work on the experimental implementation of the quantum switch. We realise a photonic quantum switch, where polarisation coherently controls the order of two operations on the transverse spatial mode of the photons. Our setup avoids the limitations of earlier implementations: the operations cannot be distinguished by spatial or temporal position. We verify the ‘indefiniteness’ in the causal order in our quantum switch by measuring a causal witness. We show that our quantum switch has no definite causal order, by constructing a causal witness and measuring its value to be 18 standard deviations beyond the definite-order bound.

In Chapter 5, I present an information-theoretic advantage of the quantum switch. Classically, no information can be transmitted through a depolarising—that is a completely noisy—channel. We show that by combining a depolarising channel with another channel in an indefinite causal order—that is, when there is superposition of the order that these two channels were applied—it becomes possible to transmit significant information. We consider two limiting cases. When both channels are fully-depolarising, the ideal limit is communication of 0.049 bits; experimentally we achieve $(3.4 \pm 0.2) \times 10^{-2}$ bits. When one channel is fully-depolarising, and the other is a known unitary, the ideal limit is communication of 1 bit. We experimentally achieve 0.64 ± 0.02 bits. Our results offer intriguing possibilities for future communication strategies beyond conventional quantum Shannon theory.

In Chapter 6, I present our theoretical work on the most general causal structures. We formulate different classical capacities for a bi-partite quantum process. We find that a one-way communication protocol through an arbitrary process cannot outperform a causally separable process, i.e., we can send at most one bit per qubit. Next, we study bi-directional communication through a causally separable process. Our result shows, a bi-directional protocol cannot exceed the limit of one way communication protocol. Finally, we generalise this result to multi-party broadcast communication protocol through a definite ordered process.

Finally, in Chapter 7, I conclude with a summary and outlook of the results presented in this thesis.

Chapter 2

Conceptual Background

Any information processing task revolves around a suitable encoding-decoding model [1]. In quantum physics, information is encoded in a *quantum state*, and extracted through a *measurement operation*. The transformation between input and output quantum states is governed by a *quantum operation* [2]. The central aspect of my research is focused beyond this traditional description. Specifically, my work involves the most general types of transformations between the elementary quantum operations formalised by *quantum supermaps* [3–6] or *process-matrix* [7]. My work is based on two specific applications of such higher order quantum processes— firstly, representation of a *non-Markovian quantum process*, i.e. a quantum process with memory [5, 8–12], and a more striking scenario of a quantum process without a definite causal structure— *indefinite causal order* [4, 5, 7, 13–15].

Before delving deeper into application of the process-matrix, it is important to develop the conceptual tools. Hence I dedicate this chapter to introduce the basic building blocks of quantum information. I first explain the conventional framework of the quantum mechanics, then I introduce the process formalism, next, I will introduce some important information-theoretic measures, and finally, I will shed light—pardon the pun—on optical implementation of the elementary quantum operations. I am going to leave the concept of exotic quantum processes to subsequent chapters.

2.1 Elementary objects in Quantum Mechanics

2.1.1 Quantum state

The notion of state is perhaps the most fundamental concept in quantum physics. The state incorporates the ‘complete’ description of a quantum physical system, containing all retrievable information about the system. The physical embodiment of states depend on the specific quantum systems in consideration. For instance, in a Harmonic oscillator, the quantum state is given by the different energy levels [16]; on the other hand in quantum optics, the state can be described by the polarisation or the spatial mode or temporal mode [17]. A quantum state can be either a *pure state*, or it could be a statistical ensemble of different pure states, referred to as a *mixed state*. Mathematically, a pure state is

described by a vector $|\psi\rangle$ in a Hilbert space \mathcal{H} with a unit norm,

$$\| |\psi\rangle \| = \sqrt{\langle \psi | \psi \rangle} = 1 \quad (2.1)$$

Here, $\langle \cdot | \cdot \rangle$ represents the inner-product. Each vector can be expanded in terms of an orthonormal basis (ONB) $\{|i\rangle\}$ as

$$|\psi\rangle = \sum_{i=0}^{d-1} \alpha_i |i\rangle \quad (2.2)$$

Here $\{|i\rangle\}$ being an ONB, $\langle i | j \rangle = \delta_{ij}$, d is the dimension of the Hilbert space \mathcal{H} , the coefficient of the basis vectors, in general, belong to the complex field, $\alpha_i \in \mathbb{C}$. The unit-norm condition in Eq. (2.1) implies $\sum_i |\alpha_i|^2 = 1$. In matrix representation, we express the ket-vectors as column matrices. As shown in Eq. (2.2), a striking property of a quantum state is that it can be in a linear combination of several pure states, also known as a *superposition* state. The superposition of states, which brings indefiniteness from a very fundamental level and is without a classical analogue, has acquired a paramount attraction—both theoretically and experimentally—since the early inception of the quantum mechanics [18].

In practice, no state is completely pure, so we need a description for mixed states. A quantum system is in a mixed state when the state fluctuates according to an underlying probability distribution. Mathematically, it is represented by a *density operator* $\rho \in \mathcal{L}(\mathcal{H}_A)$, here $\mathcal{L}(\mathcal{H})$ is the set of all linear bounded operators on the Hilbert space \mathcal{H} . As an example, if a quantum system fluctuates among the states $\{|\psi_i\rangle \in \mathcal{H}\}$, with associated probabilities p_i ($\sum_i p_i = 1$), then the resulting mixed state $\rho \in \mathcal{L}(\mathcal{H})$ is given by the convex combination $\rho = \sum_i p_i |\psi_i\rangle \langle \psi_i|$. Note that, density operator representation of a pure state $|\psi_1\rangle$ is simply the outer product, $\rho = |\psi_1\rangle \langle \psi_1|$, this can be seen as a special case when the underlying distribution is concentrated to only one pure state, i.e., it has only one nonzero coefficient p . The density operator ρ is Hermitian ($\rho^\dagger = \rho$) and positive ($\rho \geq 0$). The unit norm of the ensemble pure states ensures $\text{Tr}(\rho) = 1$. The above conditions make the eigenvalues of ρ real, positive and they sum up to one, thus density operators generalise classical probability distributions. A density operator admits an operator basis decomposition. For a density operator that describes a *qubit*—two-dimensional quantum systems—the decomposition is as follows:

$$\rho = \frac{1}{2} \left(\mathbb{1} + \sum_{i=1}^3 n_i \sigma_i \right). \quad (2.3)$$

With $\{n_i \in \mathbb{R}\}$ being real numbers and $\{\sigma_i\}$ being the Pauli operators¹ described as

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (2.4)$$

Geometrically, one can represent the qubits using a three-dimensional unit sphere, also known as the ‘Bloch sphere’. As shown in Fig. 2.1, on a Bloch sphere, the north and south pole of the sphere

¹Note that, different conventional notations exist for Pauli operators: $\sigma_1 \equiv \sigma_x \equiv X$, $\sigma_2 \equiv \sigma_y \equiv Y$, and $\sigma_3 \equiv \sigma_z \equiv Z$. In the subsequent chapters, I will explicitly mention which notation I am going to use.

represents the eigenvectors of σ_3 — $|0\rangle$ and $|1\rangle$, also known as the *computational basis*. The equatorial planes contains the eigenvectors of σ_1 and σ_2 . The pure states are represented by points that lie on the surface of the sphere whereas, the mixed states are represented by points that lie inside, with the centre representing the maximally mixed state $1/2$. Thus we can infer ‘purity’ of a state by its radius $|\vec{n}| = \sqrt{\sum_i n_i^2}$, or by evaluating $\text{Tr}[\rho^2] = (1 + |\vec{n}|^2)/2$. Note that, for pure states, $|\vec{n}|=1$, hence $\text{Tr}[\rho^2] = \text{Tr}[\rho] = 1$, whereas for the maximally mixed state, $|\vec{n}|=0$, and $\text{Tr}[\rho^2] = 1/2$.

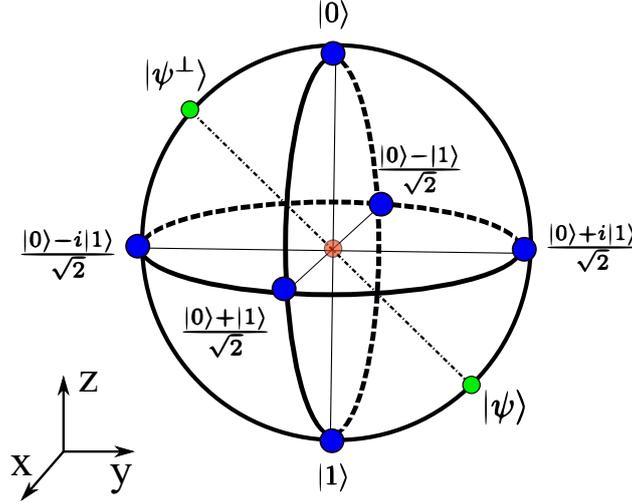


Figure 2.1: Bloch sphere representation of a qubit. The axes represent eigen vectors of different Pauli matrices: z for σ_1 , y for σ_2 , and z for σ_3 . The pure states lie on the surface of the sphere, with polar opposite points are orthogonal states. For example, the green circles represent the states $|\psi\rangle$ and $|\psi^\perp\rangle$ where $\langle\psi|\psi^\perp\rangle=0$. The mixed states lie inside the sphere, with the maximally mixed state (the red circle) is at the centre.

2.1.2 Joint quantum state

So far I have discussed the single quantum system. However, it is possible to have a large number of quantum systems having a collective state representation referred to as a *joint state*. For the sake of simplicity, I am considering only two quantum systems belonging to Hilbert spaces \mathcal{H}_A and \mathcal{H}_B with dimensions d_A and d_B respectively. This joint state, also called a *bipartite state*, in this case defined over the product Hilbert space $\mathcal{H} \equiv \mathcal{H}_A \otimes \mathcal{H}_B$, of dimension $d_A \cdot d_B$. To represent a pure joint state, first I introduce the ONB of the composite Hilbert space. If $\{|i\rangle\}$ and $\{|j\rangle\}$ are the ONBs of the Hilbert spaces \mathcal{H}_A and \mathcal{H}_B respectively, the product Hilbert space \mathcal{H} admits the ONB $\{|i\rangle \otimes |j\rangle\}$. Thus a pure state $|\psi\rangle_{AB}$ can be written as

$$|\psi\rangle_{AB} = \sum_{i=1}^{d_A-1} \sum_{j=1}^{d_B-1} \alpha_{ij} |i\rangle_A \otimes |j\rangle_B. \quad (2.5)$$

Here $\alpha_{ij} \in \mathbb{C}$, and $\sum_{i,j} |\alpha_{ij}|^2 = 1$.

Schmidt decomposition and state purification

The bipartite state mentioned in Eq. (2.5) has a useful decomposition. First we define a matrix X with elements $X_{ij}=\alpha_{ij}$, and rank r . Performing a singular value decomposition on the matrix X reveals that there exist strictly positive *Schmidt coefficients* $\{\lambda_k > 0\}_{k=1}^r$, ONBs $\{|e_k\rangle \in \mathcal{H}_A\}$ and $\{|f_k\rangle \in \mathcal{H}_B\}$ such that

$$|\psi\rangle_{AB} = \sum_{k=1}^r \sqrt{\lambda_k} |e_k\rangle_A \otimes |f_k\rangle. \quad (2.6)$$

The rank r is also known as *Schmidt rank*. One important aspect of Schmidt decomposition is that it gives an alternative physical meaning of a single system mixed state. A mixed state can always be interpreted as a joint pure state with an auxiliary quantum system which is later discarded, Fig. 2.2(a). Mathematically,

$$\begin{aligned} \text{for every mixed state } \rho_A &= \sum_{k=1}^r p_k |e_k\rangle\langle e_k| \\ \text{there exists a pure state } |\psi\rangle_{AR} &= \sum_{k=1}^r \sqrt{p_k} |e_k\rangle_A \otimes |\bar{e}_k\rangle_R. \end{aligned} \quad (2.7)$$

Here, the bar, $|\bar{e}_k\rangle$, represents the complex conjugate. This can be easily verified by evaluating the density operator $|\psi\rangle\langle\psi|_{AR}$ and then performing a partial trace on R . Physically, a partial trace is equivalent to discarding the auxiliary system.

Separable and entangled state

A bipartite state is a *product state* when it is simply a tensor product of two pure states. For instance, two uncorrelated states $|\tau\rangle \in \mathcal{H}_A$ and $|\phi\rangle \in \mathcal{H}_B$ results in the bipartite product state in \mathcal{H} is $|\psi\rangle_{AB} = |\tau\rangle_A \otimes |\phi\rangle_B$. Physically, a product state implies a complete lack of correlation between two quantum systems.

The two quantum systems can also be classically correlated. This happens when the systems' state are governed by a joint probability distribution. If the bipartite system acquires a state $|\tau_i\rangle \otimes |\phi_j\rangle$ according to a joint probability distribution $p(i, j)$, then the resulting correlated mixed state is

$$\rho_{AB} = \sum_{i,j} p(i, j) |\tau_i\rangle\langle\tau_i|_A \otimes |\phi_j\rangle\langle\phi_j|_B. \quad (2.8)$$

Clearly, the maximum classical correlation occurs when $p(i, j) = p(i)\delta_{ij}$, i.e. at a given instance the state of the system is confined to $|\tau_i\rangle \otimes |\phi_i\rangle$ with probability $p(i)$ — $\rho = \sum_i p(i) |\tau_i\rangle\langle\tau_i| \otimes |\phi_i\rangle\langle\phi_i|$. The states admitting the form in Eq. (2.8) are called *separable states*.

Any state that cannot be written in the above mentioned decomposition, is an *entangled state*. Entanglement has been a pivotal resource in many computation and information processing tasks [19–22]. The phenomena of entanglement is unique to the quantum physics: the measurement outcomes of two space-like separated entangled systems show correlations which cannot be explained

by classical physics. This non-local behaviour, as famously described by Einstein's 'spooky action at a distance' was challenged by Einstein and coauthors, in the EPR paper [23]. The authors proposed such a state, in the end, can always be described by underlying hidden variables, indicating that the formulation of quantum mechanics was incomplete. In reply, based on measured statistics, Bell proposed an inequality that cannot be violated under the assumption of local hidden variable theory [24, 25]. Bell's inequality was first experimentally violated by Aspect et. al. [26]. The experiment ensured that entanglement is a genuine non-classical property. Several follow-up experiments have been conducted since then and finally Hensen et al. [27] achieved a loop-hole free violation of Bell's inequality. A pure state is entangled if and only if its Schmidt rank is strictly greater than one. An important class of entangled states are the maximally entangled states. One such maximally entangled state is $|\Phi\rangle_{AB}$ with both Hilbert spaces having an equal dimension d and sharing an ONB $\{|i\rangle\}$.

$$|\Phi\rangle_{AB} = \frac{1}{\sqrt{d}} \sum_{i=0}^{d-1} |i\rangle_A \otimes |i\rangle_B. \quad (2.9)$$

Note that, the above state is already in the Schmidt form with the Schmidt rank being the dimension of the Hilbert space and all the Schmidt coefficients being $1/\sqrt{d}$. This is indeed a prerequisite for a state to be a maximally entangled state.

2.1.3 Quantum Channel

The state of a dynamic quantum system undergoes evolution through time. In case of a closed system—free from environmental interaction—the evolution is noiseless, and is governed by the famous Schrödinger equation [28]

$$i\hbar \frac{\partial}{\partial t} |\psi(t)\rangle = H |\psi(t)\rangle, \quad (2.10)$$

with H being the evolution Hamiltonian, $\hbar = h/2\pi$ where h is the Planck's constant. Solution of this partial differential equation states that a noiseless quantum state evolution is given by a *unitary transformation*:

$$|\psi(t)\rangle = U(t) |\psi(0)\rangle. \quad (2.11)$$

Here $U(t) = e^{-iHt/\hbar}$ is an exponential of a Hamiltonian operator H , hence a unitary transformation. In general, we drop the time parameter t . In density operator representation, for an input density operator ρ , the output density operator ρ_{out} is

$$\rho_{\text{out}} = U \rho U^\dagger. \quad (2.12)$$

A unitary transformation is perfectly reversible, hence no information about the system is lost. However, in practice no system is isolated, resulting in leakage of information about the system to the environment. Thus while the collective system-environment joint state still undergoes a noiseless evolution, our lack of knowledge about the environment renders the net evolution of the system noisy and irreversible. Such general class of system evolution can be modeled by a *quantum channel*.

A quantum channel is a linear, completely positive, and trace-preserving (CPTP) map transforming an input quantum state to an output quantum state. The map is *positive* because it maps any input quantum state to a valid output quantum state. A quantum channel is *completely* positive (CP) because any arbitrary extension of the input quantum state to an auxiliary quantum system still produces a valid output quantum state. The trace-preserving condition indicates the trace of the output state is equal to the trace of the input; physically it means that the output quantum state is not conditioned on a measurement outcome. Furthermore, when acted on a convex mixture of quantum states, the evolution ensures that the output state is the same convex mixture of the individual inputs, i.e. convex linearity is respected. We represent a quantum channel $\mathcal{N}:X_I \rightarrow X_O$ by a map that transforms any density operators in $X_I \equiv \mathcal{L}(\mathcal{H}_{X_I})$ to a density operator in $X_O \equiv \mathcal{L}(\mathcal{H}_{X_O})$.

Kraus decomposition: The action of a CPTP map on an input state can be understood through *Kraus decomposition* [2]. The Kraus representation of a map involves a set of linear bounded operators $\{K_i \in A_I \otimes A_O\}_{i=1}^r$, also known as *Kraus operators* such that for an input state $\rho \in A_I$, the output state $\rho_{\text{out}} \in A_O$ is

$$\rho_{\text{out}} = \mathcal{N}(\rho) = \sum_{i=1}^r K_i \rho K_i^\dagger. \quad (2.13)$$

The trace-preserving condition of the channel implies $\sum_i K_i^\dagger K_i = \mathbb{1}_{A_I}$. The Kraus operators could be interpreted as various errors affecting the quantum state. However, the Kraus decomposition is not unique, and different decomposition are related by an isometry. Specifically, given $\{K_i\}_{i=1}^r$ is a set of Kraus operators for the channel \mathcal{N} , then an $r' \times r$ isometry V , with elements $\{V_{i,j} : 1 \leq i \leq r', 1 \leq j \leq r\}$ satisfying $\sum_{k=1}^{r'} \bar{V}_{k,i} V_{k,j} = \delta_{ij}$, defines a new set of Kraus operators $\{K'_i\}_{i=1}^{r'}$ given by

$$K'_i = \sum_{j=1}^r \sum_{k=1}^{r'} V_{i,j} K_j \quad (2.14)$$

Stinespring dilation

All quantum evolution are unitary. However, as I have mentioned before, it is our lack of knowledge of the surrounding environment that requires us to formulate general quantum channels focusing solely on the evolving quantum system. *Stinespring dilation* theorem [29]—often referred to as the Church of the larger Hilbert space—establishes a relation between a quantum channel and its unitary extension with the help of an auxiliary environmental state. Stinespring dilation, Fig. 2.2(b), states that for any quantum channel $\mathcal{N}:A_I \rightarrow A_O$, there exists an isometry $V : A_I \rightarrow A_O E_O$, with E_O being the output of the environmental state, such that for all input state $\rho \in \mathcal{H}_{A_I}$,

$$\mathcal{N}(\rho) = \text{Tr}_{E_O}(V \rho V^\dagger). \quad (2.15)$$

One can relate the Kraus operators $\{K_i\}$ of a channel and the corresponding stinespring isometry V as

$$V = \sum_i K_i \otimes |i-1\rangle. \quad (2.16)$$

Thus Stinespring's theorem pictures the evolution of every quantum system as an interaction with its environment, followed by discarding the environment. We can always extend an isometry $V : A_I \rightarrow A_O E_O$, to a unitary $U : A_I E_I \rightarrow A_O E_O$ acting on the inputs of the system A_I and a dummy environmental state $|0\rangle\langle 0| \in E_I$. Thus Eq. (2.15) can be rewritten as

$$\mathcal{N}(\rho) = \text{Tr}_{E_O}(V\rho V^\dagger) = \text{Tr}_{E_O}[U(\rho_{A_I} \otimes |0\rangle\langle 0|_{E_I})U^\dagger]. \quad (2.17)$$

2.1.4 Quantum measurement

In the previous section I discussed how information about a quantum object is encoded in its state. To complete the information processing task we need to extract the encoded information. Thus realising a suitable measurement device is crucial. The measurement outcome of the quantum mechanics is inherently probabilistic. Any sequence of experimental outcome is governed by a specific probability depending on the input quantum state and the measurement device. This makes the quantum theory a framework for estimating probabilities of the measurement outcomes. To capture the statistical essence of the measurement procedure, first we need to develop the concept of *observables*, also known as *effects* [2]. An observable is a physical quantity such as position, momentum, polarisation of light, or orbital angular momentum. Mathematically they can be modelled by a self-adjoint operator. For a finite-dimensional case, we represent an observable with a Hermitian matrix. Given an input state ρ , the *expectation value* of observing A is given by

$$\langle A \rangle_\rho = \text{Tr}(A\rho). \quad (2.18)$$

The probabilistic nature of quantum mechanics results in a non-deterministic and irreversible evolution of states. Precisely, given an input state ρ , the measurement of an observable A collapses the state into one of its eigenstates $|\psi_a\rangle$ with probability $|\langle \psi_a | \rho | \psi_a \rangle|^2$ and the measurement outcome is the corresponding eigenvalue a .

Complementarity and uncertainty principle

Unlike in classical measurement, not all quantum mechanical observables can be measured simultaneously. A pair of observables that cannot be measured simultaneously is called *complementary observables*, for example position and momentum, polarisation along different axes and so on. Two observables, A and B , are simultaneously measurable only when they commute, $[A, B] = (AB - BA) = 0$. A general non-commuting pair of observables respects the following operator inequality:

$$\langle \Delta A \rangle \langle \Delta B \rangle \geq \frac{1}{2} |\langle [A, B] \rangle|. \quad (2.19)$$

Here, $\langle \Delta A \rangle_\rho = \sqrt{\langle A^2 \rangle_\rho - \langle A \rangle_\rho^2}$ represents the variance of the observable A given an input state ρ . As the inequality in Eq. (2.19) holds for any input state, we have dropped the subscripts ρ . An important

example involves the position (x) and momentum (p_x) operators. They satisfy the commutation relation $[x, p_x] = i\hbar$. Writing Eq. (2.19) for x and p , we get the famous Heisenberg uncertainty relation [30],

$$\langle \Delta x \rangle \langle \Delta p_x \rangle \geq \frac{\hbar}{2}. \quad (2.20)$$

Positive Operator Valued Measure (POVM)

To infer complete statistical knowledge of an input state we need to make a complete set of measurements—as in a Positive Operator Valued Measure (POVM). A POVM represents a collection of operators $\{A_m\}$, we call each element A_m in the set a POVM-element. The index m denotes the specific measurement outcome. For an input state, the probability of obtaining m is given by the expectation value of the corresponding POVM element A_m ,

$$p(m) = \langle A_m \rangle_\rho = \text{Tr}(A_m \rho). \quad (2.21)$$

Eq. (2.21) is also known as *Born's rule*. As $\langle A_m \rangle_\rho$ represents a probability — $0 \leq \text{Tr}(A_m \rho) \leq 1$, any POVM-element A_m must satisfy the operator inequality $0 \leq A_m \leq \mathbb{1}$, with 0 , and $\mathbb{1}$ representing the null and identity operators respectively. The sum of all observables denotes a deterministic outcome, i.e., an outcome with probability one, and hence we have a ‘completeness’ relation,

$$\sum_m A_m = \mathbb{1}. \quad (2.22)$$

The completeness relation is owing to the fact the probabilities sum up to one. Each POVM-element A_m is associated with a *measurement operator* M_m such that $A_m = M_m^\dagger M_m$. This factorisation is not unique and the relationship between the measurement operator and the observable can be parameterized by a unitary U_m as follows, $M_m = U_m \sqrt{A_m}$. Since, a unitary is a deterministic transformation, it does not have any influence on the $p(m)$. Hence, without loss of generality, one can assume a minimal measurement $M_m = \sqrt{A_m}$. We can perform either a *selective measurement* (where a specific outcome is read out) or a *non-selective measurement* (measurement is performed without reading the outcome) to get post measurement states ρ_s and ρ_{ns} respectively. For a given input ρ and measurement operator M_m , the post-measurement states are

$$\rho_s = \frac{M_m \rho M_m^\dagger}{p(m)} \quad \text{and} \quad \rho_{ns} = \sum_m M_m \rho M_m^\dagger. \quad (2.23)$$

A special class of measurement operator is called a *Projection valued measure (PVM)* which is represented by a set of orthogonal *projectors* $\{P_m\}$, the orthogonality condition implies,

$$P_m P_{m'} = \delta_{mm'} P_m. \quad (2.24)$$

The completeness relation ensures $\sum_m P_m = \mathbb{1}$. A simple example of PVM is measurement in a specific ONB $\{|m\rangle\}$. The measurement operators in such scenario are simply the outer-product of the basis states: $P_m = |m\rangle\langle m|$.

Naimark's dilation

Naimark's dilation theorem gives a physical interpretation of POVM and establishes a relationship with PVM. The key ingredients are an isometry that takes a quantum system as an input and a joint state (the evolved quantum system and an auxiliary state) as output. Naimark's theorem states that every POVM can be effectively realised by the process of a projective measurement on the auxiliary state following the isometric evolution, Fig. 2.2(c). Specifically, for every POVM-element A_m , there exists an isometry V such that

$$A_m = V(\mathbb{1} \otimes |m\rangle\langle m|)V^\dagger, \text{ for all } m. \quad (2.25)$$

The isometry can be expressed in terms of the minimal measurement operator $\sqrt{A_m}$ as

$$V = \sum_m \sqrt{A_m} \otimes |m\rangle. \quad (2.26)$$

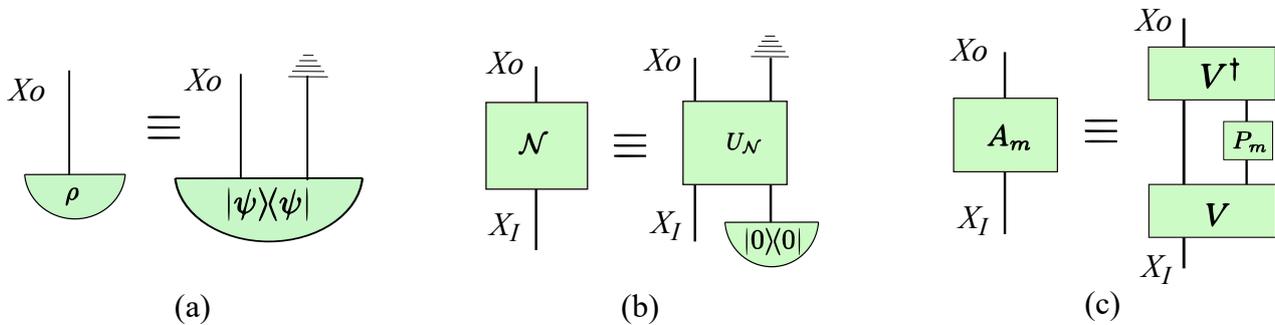


Figure 2.2: Three types of dilations. (a) Purification of state: any mixed state ρ can be purified to a pure joint state $|\psi\rangle$. We get back the mixed state by tracing out (shown by the inverted ground) the auxiliary system, see Eq. (2.7). (b) Stinespring dilation: any noisy channel \mathcal{N} can be extended to a joint unitary $U_{\mathcal{N}}$ followed by tracing out the auxiliary system, see Eq. (2.17). (c) Naimark's dilation: any POVM element A_m can be dilated to an isometric evolution V and a projector $P_m = |m\rangle\langle m|$ on the auxiliary system, see Eqs. (2.25), and (2.26).

Measurement as complete positive (CP) maps

If we compare the Kraus decomposition of a channel in Eq. (2.13) with the post-measurement state ρ_{ns} in Eq. (2.23) and observe that the completeness relationship in Eq. (2.22) suggests $\sum_m M_m^\dagger M_m = \mathbb{1}$, we come to the conclusion that a non-selective measurement scheme is in fact a quantum channel (CPTP) with the measurement operators $\{M_m\}$ being Kraus operators. This observation enables us to associate a map \mathcal{M}_m with each POVM-element A_m that gives the probability of observing the outcome m given an input state ρ ,

$$\mathcal{M}_m(\rho) = \text{Tr}(M_m \rho M_m^\dagger) = \text{Tr}(A_m \rho). \quad (2.27)$$

The map \mathcal{M}_m is a *completely positive* map (CP). However it is not *trace preserving*, as $M_m^\dagger M_m \neq \mathbb{1}$. The completeness relationship ensures that the sum of all such maps is a CPTP map, with with the

combination of all outputs representing an event of unit probability i.e.

$$\sum_m \mathcal{M}_m(\rho) = \sum_m \text{Tr}(A_m \rho) = \text{Tr} \left[\left(\sum_m A_m \right) \rho \right] = \text{Tr}(\rho) = 1. \quad (2.28)$$

As this CP map gives out a probability, we say that they have a trivial output, with the dimension of the output Hilbert space $d_O=1$.

It is worth mentioning, that the notion of POVMs could be further generalised to *quantum instruments*. The subtle difference is that a quantum instrument gives more freedom in choosing the post-measurement state. For example, we can simply discard the post-measurement state and replace with an arbitrary state, the corresponding CP map is given by $\mathcal{M}_m(\rho) = \sigma_m \text{Tr}(A_m \rho)$. Note that sum of all such CP maps again produces a CPTP map, $\mathcal{M}(\rho) = \sum_m \mathcal{M}_m(\rho) = \sum_m \sigma_m \text{Tr}(A_m \rho)$. The key difference between a POVM and a quantum instrument is that, the output dimension of a quantum instrument can be greater than one $d_O \geq 1$. This makes a POVM a special case of a quantum instrument.

2.1.5 Choi-Jamiołkowski (CJ) isomorphism

In the previous sections, I have introduced elementary quantum objects: quantum state, evolution, and measurement. I have shown all quantum evolutions are CPTP maps and measurements are associated with CP maps. It is often useful to represent a linear quantum operator and a quantum map in terms of a positive semidefinite operator, we can do this via the *Choi-Jamiołkowski (CJ) isomorphism* [31, 32]. Physically, CJ isomorphism of a CP map (a linear operator) represents the output when the map (operator) acts on an unnormalized maximally entangled state $|\mathbb{1}\rangle\rangle = \sum_i |i\rangle \otimes |i\rangle$, with $\{|i\rangle\}$ being an ONB of the input Hilbert space \mathcal{H}_{X_I} . In case of an operator $A \in X_I \otimes X_O$, we consider a ‘*pure CJ isomorphism*’ defined as ²

$$|A\rangle\rangle = (\mathbb{1} \otimes A) |\mathbb{1}\rangle\rangle. \quad (2.29)$$

We represent the inverse CJ map as

$$A |\psi\rangle = \left[\langle \bar{\psi} |^{X_I} \otimes \mathbb{1}^{X_O} \cdot |A\rangle\rangle^{X_I X_O} \right] \quad (2.30)$$

Here, the bar, $\langle \bar{\psi} |$, represents the complex conjugate. Two important examples of pure CJ isomorphism are unitary operations and measurement-preparation operations.

a. Unitary operation: Consider a unitary $U \in X_I \rightarrow X_O$ with a matrix representation $U = \sum_{jk} u_{jk} |j\rangle\langle k|$. Here $|k\rangle$ is an ONB of the input Hilbert space and $|j\rangle$ is an ONB of the output Hilbert space. The unitary condition imposes $\sum_l \bar{u}_{lk} u_{lj} = \delta_{jk}$. The corresponding CJ vector $|U\rangle\rangle$ then becomes

$$|U\rangle\rangle^{X_I X_O} = \sum_{jk} u_{jk} |k\rangle^{X_I} \otimes |j\rangle^{X_O}. \quad (2.31)$$

²As shown in Fig. 2.3, there are two definitions of CJ representation available in the literature. In this chapter, I am describing the one in Refs. [5, 6] (see Fig. 2.3(a)–(c)). The other definition is given by an overall transpose [7, 33] (see Fig. 2.3(d)–(f)). Two versions of CJ representations result in two different rules of linking different quantum operations and processes (see Fig. 2.5). In the subsequent chapters, I will explicitly mention which definition I am using. In Figs. 2.3, 2.4, and 2.5, I aim to capture the mathematical formalism through pictures. For a detailed description, the readers are encouraged to see Coecke’s work on “quantum picturalism” [34].

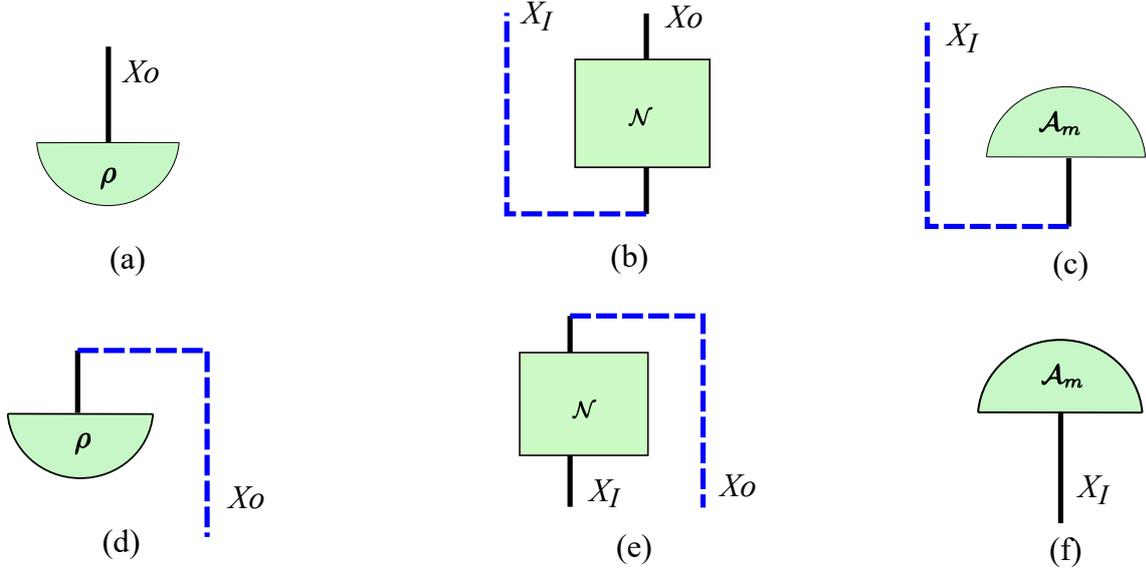


Figure 2.3: Two versions of CJ representation for quantum states (a,d), channels (b,e), and measurements (c,f). The blue dashed lines represent the necessary transformation. In (a)-(c), I show CJ representation as shown in Eq. (2.34) [5, 6]. Note the blue dashed lines go upward representing unnormalised maximally entangled states isomorphic to the relevant systems (also called cup in Ref. [34]). On the other hand, (d)-(f) shows a CJ representation with an overall transpose [7, 33]. For example, CJ of the channel \mathcal{N} is given by $N = [\sum_{i,j} |i\rangle\langle j| \otimes \mathcal{N}(|i\rangle\langle j|)]^T$. The blue dashed lines in this case go downward representing transpose of unnormalised maximally entangled states isomorphic to the relevant systems (also called cap in Ref. [34]).

b. Measurement-preparation: Next consider a measurement preparation operator $A = |\psi\rangle\langle\phi|$. Physically it means that we perform a projective measurement $|\phi\rangle\langle\phi|$ at the input Hilbert space and prepare a post-measurement state $|\psi\rangle$ at the output Hilbert space. The corresponding CJ representation then becomes

$$|A\rangle\rangle^{X_I X_o} = |\bar{\phi}\rangle^{X_I} \otimes (|\psi\rangle)^{X_o}. \quad (2.32)$$

Next I consider the ‘mixed CJ isomorphism’ which is particularly relevant for representing CP maps. For a linear map $\mathcal{M} \in X_I \rightarrow X_o$ we have the following CJ representation

$$M^{X_I X_o} = [\mathbb{1} \otimes \mathcal{M}(|\mathbb{1}\rangle\rangle\langle\langle\mathbb{1}|)] \in X_I \otimes X_o. \quad (2.33)$$

$$= \sum_{i,j=0}^{d_{X_I}-1} [|i\rangle\langle j|^{X_I} \otimes \mathcal{M}(|i\rangle\langle j|)^{X_o}]. \quad (2.34)$$

Here, d_{X_I} is the dimension of the system X_I . In the case of a linear operator A acting on a density operator $\rho = \rho A^\dagger$, we can relate pure and mixed CJ representation as

$$M = [[A]] = |A\rangle\rangle\langle\langle A|. \quad (2.35)$$

Note the use of the notation $[[.]]$ in this case, which I will often use in the subsequent chapters. The inverse map of a mixed CJ representation is

$$\mathcal{M}(\rho) = \text{Tr}_{X_I} [(\rho^{X_I} \otimes \mathbb{1}^{X_o})^T \cdot M^{X_I X_o}] \quad (2.36)$$

Here, the superscript T represents the transpose operation. CJ isomorphism provides an important definition for complete positivity of maps, a map is CP if and only if its CJ representation is a positive semidefinite operator, i.e. $M^{X_I X_O} \geq 0$. If the map \mathcal{M} is CPTP, then the Choi representation gives an additional constraint

$$\text{Tr}_{X_O} M^{X_I X_O} = \mathbb{1}^{X_I}. \quad (2.37)$$

Two important examples of quantum maps are *trace and replace maps* and POVMs.

Trace and replace map: A trace and replace map $\mathcal{M}^X(\rho) = \sigma \text{Tr}(\rho)$ represents preparation of a fixed state σ regardless of the input state, i.e. the input state is discarded. The corresponding CJ representation is

$$M^{X_I X_O} = \mathbb{1}^{X_I} \otimes (\sigma)^{X_O}. \quad (2.38)$$

POVMs: I have defined a POVM in Eq. (2.27). The corresponding CJ representation of such maps is transpose of the POVM-element, i.e.

$$M^{X_I} = A_m^{T X_I}. \quad (2.39)$$

With these two maps, we can easily infer the CJ representation of an element of a quantum instrument. For instance, a CP map of the form $\mathcal{M}_X(\rho) = \sigma_m \text{Tr}(A_m \rho)$ has a CJ representation

$$M^{X_I X_O} = A_m^{T X_I} \otimes (\sigma_m)^{X_O}. \quad (2.40)$$

2.2 Process Framework

Elementary quantum maps, as introduced in the previous section, transforms quantum states to quantum states either deterministically (via CPTP maps) or probabilistically (via CP maps). However, it is possible to allow for more general transformations. For instance, optimal cloning of a unitary operation is an example of a transformation from a CPTP map to another CPTP map [35]. A transformation from a quantum state to a CP map is also possible, an important example is a programmable quantum processor [36], where a control quantum state—or *program register*, as in Ref. [36]—implements a CP map on a target quantum state referred to as a *data register*. The framework that encompasses such higher order transformation is formalised by a *quantum supermap* [3] or process matrix [7]. The input CP maps $\{\mathcal{M}_i\}_{i=1}^n$ can be considered as operations performed by a group of n experimentalists in their local laboratories. The higher order map or process matrix W represents the background. The local parties can receive a quantum or classical input from the background, perform quantum operation on it and then send the output back to their surrounding. Depending on the background process, the local parties can either communicate with each other or not (i.e. it could be a no-signalling process). The process W and all the local operations collectively create a new CP map \mathcal{N} with CJ representation,

$$N = M_1 * M_2 * \dots * M_n * W. \quad (2.41)$$

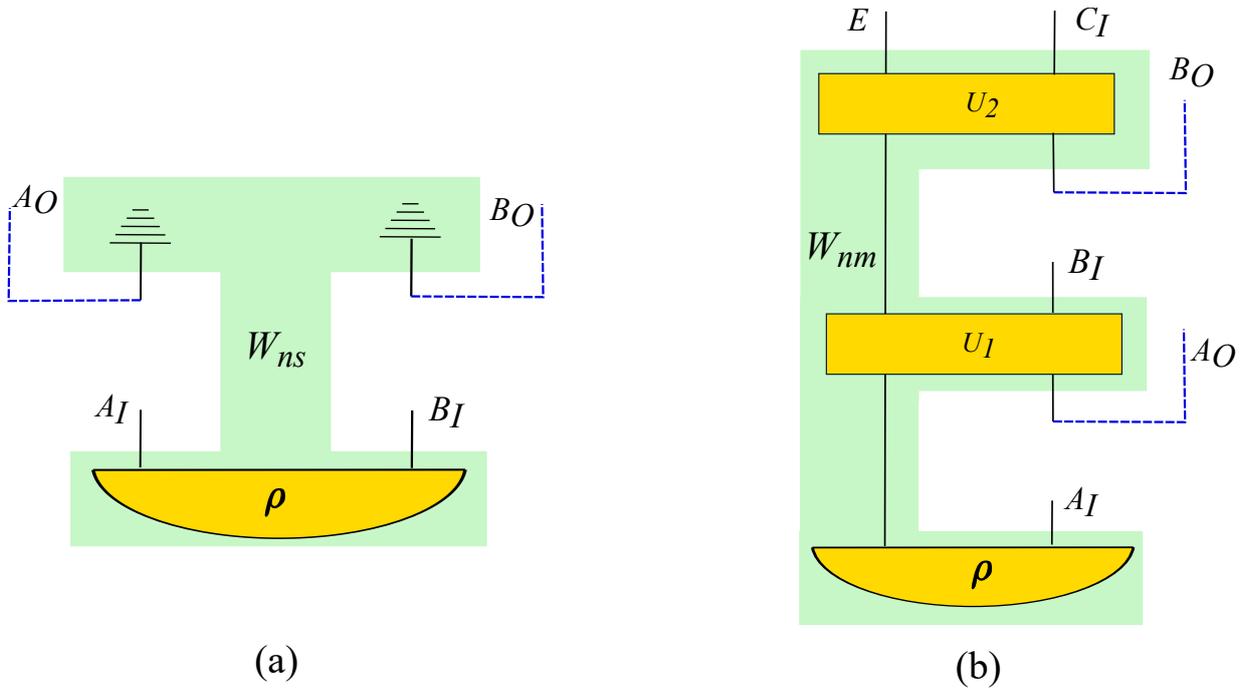


Figure 2.4: Example of two quantum processes. The green shade represent the process, with the construction shown in yellow. The empty slots are available to the local parties applying their quantum operations; e.g., $\mathcal{M} : A_I \rightarrow A_O$ can be applied to the slot $A_I A_O$. The dashed blue wires denote the CJ isomorphism of the output systems, which are necessary to represent a process by a positive semidefinite matrix. In (a), I show a no-signalling process W_{ns} . This is a typical Bell-type scenario where an entangled state ρ is shared between two parties. The inverted ground denotes tracing out the corresponding quantum system since the output dimension of a quantum measurement is one. In (b), I show a process that allows signalling between the local operations. The above example shows a non-Markovian process W_{nm} (see Chapter 3) composed of an initially correlated state ρ , and the subsequent entangling operations U_1 and U_2 . Interestingly, unlike (a) and (b) the most general processes do not show a fixed causal order between the local parties. We refer to them as indefinite causal ordered processes (see Chapters 4, 5, and 6).

Here, M_i is the CJ representation of the map \mathcal{M}_i . The operator ‘*’ is called the *link product* [6] representing the concatenation among the local operations. For two operators $A \in \mathcal{X}$ and $B \in \mathcal{Y}$, the link product $A * B$ is defined as follows

$$A * B := \text{Tr}_{\mathcal{X} \setminus \mathcal{Y}} [(\mathbb{1}^{\mathcal{X} \setminus \mathcal{Y}} \otimes A)^{T_{\mathcal{X} \setminus \mathcal{Y}}} \cdot (B \otimes \mathbb{1}^{\mathcal{Y} \setminus \mathcal{X}})]. \quad (2.42)$$

Here the superscript $T_{\mathcal{X} \setminus \mathcal{Y}}$ denotes partial transpose over the shared systems. Note that, when the Hilbert spaces associated with all local operators completely overlap with that of the process matrix, i.e. $(\mathcal{X} \setminus \mathcal{Y}) \cup (\mathcal{Y} \setminus \mathcal{X}) = \emptyset$, the link product reduces to a scalar quantity. This has an important implication in generalising Born’s rule. To elucidate the motivation behind such generalisation, let me first recapitulate Born’s rule, introduced in Eq.(2.21). Born’s rule gives the probability of a particular measurement outcome given an input quantum state. The interpretation can be straightforwardly extended for multiple POVM elements acting on different systems. For example, for a joint state $\rho^{X_I Y_I}$

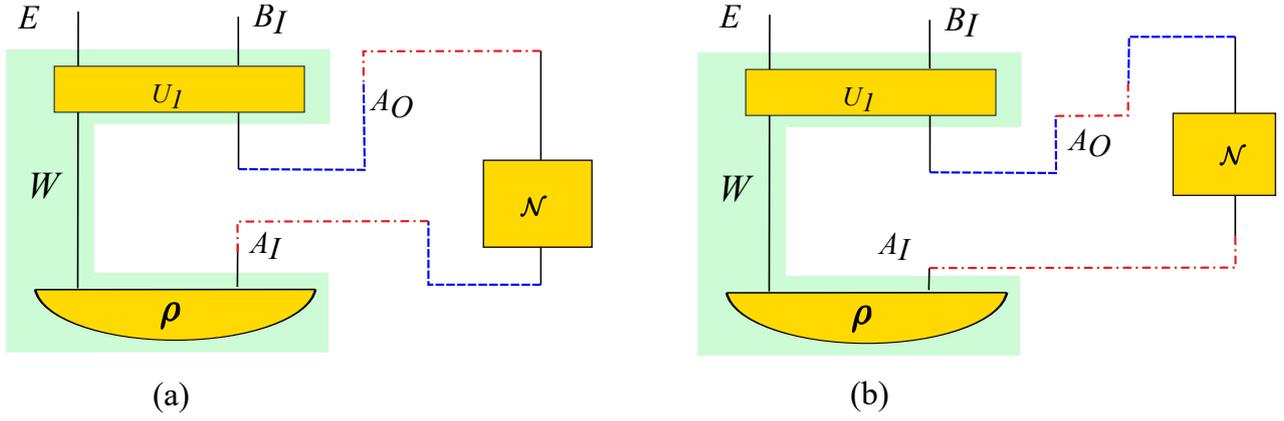


Figure 2.5: Linking local operations to processes. Red dashed lines show the necessary bending of wires to join the process and the local operation. Depending on the definition of CJ representation, we have two ways of linking. For Eq. (2.34), we have (a) admitting the link product (2.42). Note here red dashes bend downward, i.e., partial transpose on the overlapping system is necessary. In (b), we have CJ representation of Ref. [7, 33]. The red dashes bend upward. Hence, no transpose is required while linking processes to the local operations.

on which we are applying two POVMs $A_m^{X_I X_O}$ and $B_n^{Y_I Y_O}$, the probability of occurrence of the outcomes m and n given the initial state ρ is $P(A_m, B_n | \rho) = \text{Tr}[(A_m \otimes B_n) \cdot \rho]$. Interestingly, such interpretation is problematic when multiple CP maps act on the same system sequentially. The challenge arises due to the collapse of the quantum state. Each measurement induces a post-measurement state as shown in Eq. (2.23). To mitigate this, in Refs. [37, 38] the authors proposed an interesting solution. As shown in Fig. 2.4(b), the background process $W \in \otimes_{i=1}^n A_I^{(i)} \otimes A_O^{(i)}$ depicts a comb-like structure, which is indeed referred to as ‘*quantum comb* [5, 6]’, composed of the initial state ρ and open slots for the subsequent CP maps $\{\mathcal{M}_i : A_I^{(i)} \rightarrow A_O^{(i)}\}_{i=1}^n$. The overall concatenation represents a probability. Thus with the help of Eqs. (2.41), and (2.42) we can find the *Generalised Born’s rule*, i.e. probability of observing the intermediate CP maps, given the initial state is

$$\begin{aligned}
 & P(\mathcal{M}_1, \mathcal{M}_2, \dots, \mathcal{M}_n | \rho) \\
 &= \text{Tr} \left[\left(M_1^{A_I^{(1)} A_O^{(1)}} \otimes M_2^{A_I^{(2)} A_O^{(2)}} \otimes \dots \otimes M_n^{A_I^{(n)} A_O^{(n)}} \right) \cdot W^{A_I^{(1)} A_O^{(1)} A_I^{(2)} A_O^{(2)} \dots A_I^{(n)} A_O^{(n)}} \right]. \quad (2.43)
 \end{aligned}$$

In general, the set of valid process matrices should result in well-defined probabilities for all possible local operations. To obtain a valid probability the process matrix should satisfy the following two conditions:

Non-Negativity: The probability must be non-negative, for all local operations even if the local operations are entangled among themselves. To elaborate, consider two local operations with CJ representation $M_1^{A_I A_I A_O}$ and $M_2^{B_I B_I B_O}$. The local operations are connected by a joint state $\rho^{A_I B_I}$. The background is modelled by a bipartite process matrix $W^{A_I A_O B_I B_O}$. The joint state and the process matrix can be thought of as a new process $\rho^{A_I B_I} \otimes W^{A_I A_O B_I B_O}$. With this, the non-negativity of probability

requires

$$\text{Tr} \left[\left(M_1^{A_I A_I' A_O} \otimes M_2^{B_I B_I' B_O} \right) \cdot \left(\rho^{A_I B_I'} \otimes W^{A_I A_O B_I B_O} \right) \right] \geq 0 \quad (2.44)$$

$$\text{for all, } M_1^{A_I A_I' A_O} \geq 0, M_2^{B_I B_I' B_O} \geq 0, \rho^{A_I B_I'} \geq 0. \quad (2.45)$$

The above condition is satisfied only when the process matrix is a positive semidefinite,

$$W \geq 0. \quad (2.46)$$

Normalisation The second criteria is that for all CPTP local operations, the probability must be normalised. Hence, in case of two local CPTP operations with CJ representations $M_1^{A_I A_O}$ and $M_2^{B_I B_O}$, a valid process $W^{A_I A_O B_I B_O}$ should give

$$\text{Tr} \left[\left(M_1^{A_I A_O} \otimes M_2^{B_I B_O} \right) \cdot W^{A_I A_O B_I B_O} \right] = 1, \quad (2.47)$$

$$\text{for all, } M_1^{A_I A_O} \geq 0, M_2^{B_I B_O} \geq 0,$$

$$\text{such that, } \text{Tr}_{A_O} M_1^{A_I A_O} = \mathbb{1}^{A_I}, \text{Tr}_{B_O} M_2^{B_I B_O} = \mathbb{1}^{B_I}.$$

The relevant constraints associated with Eq. (2.47) are

$$\text{Tr}(W) = d_{A_O} d_{B_O}, \quad (2.48)$$

$${}_{A_I A_O} W = {}_{A_I A_O B_O} W, \quad (2.49)$$

$${}_{B_I B_O} W = {}_{A_O B_I B_O} W, \quad (2.50)$$

$$W = {}_{A_O} W + {}_{B_O} W - {}_{A_O B_O} W. \quad (2.51)$$

Here ${}_X W = \text{Tr}_X(W) \otimes \mathbb{1}^X / d_X$ is the ‘trace and replace operator’ with d_X representing the dimension of X . The above-mentioned constraints can be generalised to multi-partite process and are as follows [11]

$$W \geq 0 \quad (2.52)$$

$$\text{Tr} W = d_O \quad (2.53)$$

$$W = P(W) \quad (2.54)$$

Here, $d_O = d_{A_O^{(1)}} d_{A_O^{(2)}} \dots d_{A_O^{(n)}}$ is the dimension of all output Hilbert spaces. P is a projector onto the linear subspace $\mathcal{P} \subset A_I^{(1)} \otimes A_O^{(1)} \otimes \dots \otimes A_I^{(n)} \otimes A_O^{(n)}$.

2.2.1 Examples

I will end this section with two important examples of processes.

Quantum states: These are no-signalling processes, a typical example is a setup of Bell’s experiment. As shown in Fig. 2.4(a), a joint state is shared between the input systems A_I and B_I , and there are open slots for local operations. The corresponding process is

$$W^{A_I B_I A_O B_O} = \rho^{A_I B_I} \otimes \mathbb{1}^{A_O B_O}. \quad (2.55)$$

Let us consider the local parties are applying two POVM-elements $E_1^{A_I}$ (CJ representation $M_1^{A_I A_O}$) and $E_2^{B_I}$ (CJ representation $M_2^{B_I B_O}$) on the each system of the joint state. Note that we have $\text{Tr}_{A_O} M_1^{A_I A_O} = E_1^{T A_I}$ and $\text{Tr}_{B_O} M_2^{B_I B_O} = E_2^{T B_I}$. Then the probability of performing the respective POVM elements according to Eq. (2.43) reduces to Born's rule, i.e.

$$\begin{aligned} P(M_1^{A_I A_O}, M_2^{B_I B_O}) &= \text{Tr} \left[\left(M_1^{A_I A_O} \otimes M_2^{B_I B_O} \right) \cdot W^{A_I A_O B_I B_O} \right], \\ &= \text{Tr} \left[\left(E_1^{T A_I} \otimes E_2^{T B_I} \right) \cdot \rho^{A_I B_I} \right]. \end{aligned} \quad (2.56)$$

Quantum Channels: Suppose a party A prepares a quantum state ρ^{A_O} sends it to party B 's input system B_I through a background channel $\mathcal{N} : A_O \rightarrow B_I$. The corresponding process representation of the channel is $W^{A_O B_I}$. The party B performs a measurement to obtain a POVM element E^{B_I} . According to Born's rule, Eq. (2.21), the probability of observing the said POVM element is $P(E|\rho) = \text{Tr}[E \cdot \mathcal{N}(\rho)]$. On the other hand, using Eq. (2.43) we find

$$P(E|\rho) = \text{Tr} \left[\left(\rho^T \right)^{A_O} \otimes E^{B_I T} \cdot W^{A_O B_I} \right], \quad (2.57)$$

$$= \text{Tr} \left[E^{B_I T} \cdot \text{Tr}_{A_O} \left\{ \left(\rho^T \right)^{A_O} \otimes \mathbb{1}^{B_I} \cdot W^{A_O B_I} \right\} \right]. \quad (2.58)$$

We can absorb the transpose in the definition of the POVM-element E^{B_I} . The remaining part of the equation suggests, $\mathcal{N}(\rho) = \text{Tr}_{A_O} \left\{ \left(\rho^T \right)^{A_O} \otimes \mathbb{1}^{B_I} \cdot W^{A_O B_I} \right\}$. Comparing with the inverse CJ representation in Eq. (2.36), we deduce

$$W^{A_O B_I} = \mathbb{1}^{A_O} \otimes \mathcal{N}(|\mathbb{1}\rangle\rangle\langle\langle\mathbb{1}|)^{B_I}. \quad (2.59)$$

Thus process representation of channel is equivalent to its CJ representation as in Eq. (2.33).

2.3 Background on information theory

The goal of a scientific theory is twofold, firstly to explain prior experimental observations, and then to propose a falsifiable test. Extracting information through empirical evidence is the pillar of scientific enquiry. This makes it imperative to attribute a precise meaning to the term *information*. Intuitively, the information content, quantified in *bits*, denotes how much knowledge we can acquire from a random measurement—the higher the randomness the larger the information content.

A measure of information was first introduced by Hartley in the 1930s. Essentially, his measure was the logarithm of the dimension. In 1948, Claude Shannon conceptualized the modern information theory in his seminal work [39]. Shannon's contribution was so profound that the field is often referred to as *Shannon theory*. Since then the field has flourished with the contributions of Kullback and Leibler [40], Fano [41], Brillouin [42], Jaynes [43] and so on. The tools of information theory soon found applications in a wide range of subjects—physics [42], statistics [44], computer science [45], biology [46] and so on. The evolution of information theory is closely intertwined with physics. As we will see shortly, the fundamental information-theoretic measure, entropy, was actually motivated

by thermodynamics. In the same light, there has been an active discussion about the physical nature of information. Landauer famously phrased, ‘Information is physical’ and in his seminal work [47], he proposed a quantification of the energy cost of erasing one bit of information. The *Landauer bound* was subsequently confirmed in several experiments [48–51]. Quantum mechanics has a great impact on information theory. Quantum entanglement is an enabling resource that augments conventional information processing tasks, resulting an independent research area, quantum Shannon theory [52]. A significant aspect of my research involves this nascent and intriguing field.

In this section I will introduce basic information-theoretic measures, leaving specific use of those measures to subsequent chapters. I will first give a brief overview of classical information theory and then extend it to its quantum analogue.

2.3.1 Classical Shannon theory

The information signifies how much knowledge we can acquire from a system. The amount of knowledge depends on our anterior ignorance about state of the system. If we are certain about the state, we do not learn anything new by a further measurement — the net information content is zero. In contrast, a maximally chaotic system gives us more opportunity to gain knowledge. In this sense, one would argue the information is closely related to the thermodynamic entropy — a measure of randomness of the system. Acquisition of information decreases the thermodynamic entropy of the environment, as opined by Szilard [53]. On the other hand, erasing of information increases thermodynamic entropy of the environment, as postulated by Landauer [47] and supported by Bennett [54].

In classical physics, we model a state by a random variable. For the purpose of my thesis, I will only focus on the discrete random variable. We represent each event in a random variable X by a numerical label x with the corresponding probability $p_X(x)$. We assign the information content $I_X(x)$ to an event x , defined as

$$I_X(x) \equiv -\log_2(P_X(x)). \quad (2.60)$$

The base 2 in the logarithm indicates that the number of possible states is represented as a power of 2, and the unit of information is a *binary unit (bit)*.

Shannon entropy of a random variable

The information content in Eq. (2.60) is applicable to a particular event. However, we are more interested in the net information of the overall system. Hence, we look for an average information content, also known as *Shannon entropy*. The Shannon entropy $H(X)$ of a random variable X is defined by

$$H(X) = \sum_x p_X(x) I_X(x) = -\sum_x p_X(x) \log_2(P_X(x)). \quad (2.61)$$

Before introducing further properties of Shannon entropy, I will digress a bit (again, no pun intended). Although the term ‘entropy’ is reminiscent of thermodynamics and statistical mechanics, Shannon’s

original intention was to build a model for communication theory. In Shannon's own words, "My greatest concern was what to call it. I thought of calling it 'information,' but the word was overly used, so I decided to call it 'uncertainty.' When I discussed it with John von Neumann, he had a better idea. von Neumann told me, 'You should call it entropy, for two reasons. In the first place your uncertainty function has been used in statistical mechanics under that name, so it already has a name. In the second place, and more important, no one knows what entropy really is, so in a debate you will always have the advantage [55].'"

Even though Shannon's original goal was different, the thermodynamic interpretation of Shannon's entropy is a research field in its own right and a full discussion on the same is beyond the scope of this thesis. Nevertheless, I would like to emphasise a particular physical relevance: If we consider a macrostate, where each of its microstates is associated to a probability $p_X(x)$, then Gibb's entropy, S_{Gibbs} , of the macrostate [56] is

$$S_{\text{Gibbs}} = -k_B \sum_x p_X(x) \ln(p_X(x)) = (k_B \ln 2) \cdot H(X). \quad (2.62)$$

Eq. (2.62) gives Shannon entropy an operational interpretation: the Shannon entropy is the amount of bits required to fully describe the ensemble. It has the following properties:

Non-negativity: The Shannon entropy is non-negative for any discrete random variable X ,

$$H(X) \geq 0. \quad (2.63)$$

This is true because information content is always non-negative, and hence the average information content is also nonnegative.

Concavity: The Shannon entropy is concave in the probability mass function $p_X(x)$. If we consider a random variable X is composed of a probabilistic mixture of two random variables X_1 and X_2 , such that $p_X(x) = qp_{X_1}(x) + (1 - q)p_{X_2}(x)$, then

$$H(X) \geq qH(X_1) + (1 - q)H(X_2). \quad (2.64)$$

The intuition behind this is that a random mixture increases the uncertainty of the state, and hence increases the entropy.

Permutation invariance: The Shannon entropy remains unchanged under permutation, e.g., if we shuffle a deck of cards, the entropy remains the same. This is true because Shannon's entropy depends only on the probability distribution, which remains unchanged under permutation.

Minimum and maximum value: The minimum value of the Shannon entropy is zero, which is the case when only one event occurs with a certain probability. Similarly, Shannon entropy attains a maximum value when all the events occur with an equal probability. For a random variable X with n possible events, the maximum possible value of Shannon's entropy is $\log_2 n$. Hence the following inequality holds,

$$0 \leq H(X) \leq \log_2 n. \quad (2.65)$$

Conditional entropy

Next let us consider two correlated random variables X and Y described by the joint probability mass function $P_{XY}(x,y)$ with events of X being $\{x\}$ and that of Y being $\{y\}$. The conditional entropy $H(X|Y)$ signifies the Shannon entropy of X given knowledge of Y and is given by

$$H(X|Y) = - \sum_{x,y} p_{XY}(x,y) \log_2(p_{X|Y}(x|y)). \quad (2.66)$$

As expected, if the random variables are independent, $p_{XY}(x,y) = p_X(x)p_Y(y)$, the knowledge of Y does not decrease the uncertainty of X . In that case $H(X|Y) = H(X)$, i.e., information about Y does not increase uncertainty of X . Hence we have

$$H(X) \geq H(X|Y), \quad (2.67)$$

that is, conditioning does not increase Shannon entropy. Another important property of classical conditional entropy is that it is always non-negative,

$$H(X|Y) \geq 0. \quad (2.68)$$

Joint entropy

Joint entropy extends the notion of Shannon entropy to multiple random variables. If X and Y are two random variables with joint probability distribution $p_{XY}(x,y)$, then the joint entropy $H(X,Y)$ is defined as

$$H(X,Y) = - \sum_{x,y} p_{XY}(x,y) \log_2(p_{XY}(x,y)). \quad (2.69)$$

The joint entropy can be related to the conditional entropy:

$$H(X,Y) = H(X) + H(Y|X) = H(Y) + H(X|Y). \quad (2.70)$$

We can extend the definition for n random variables and get

$$H(X_1, X_2, \dots, X_n) = H(X_1) + H(X_2|X_1) + \dots + H(X_n|X_{n-1}, \dots, X_1). \quad (2.71)$$

Applying Eq. (2.67) to the above equation, we get the sub-additivity of Shannon entropy,

$$H(X_1, X_2, \dots, X_n) \leq \sum_{i=1}^n H(X_i), \quad (2.72)$$

with equality if and only if the random variables are independent.

Mutual information

Mutual information quantifies the correlation between two random variables. For two random variables X and Y , with joint distribution $p_{XY}(x,y)$, the mutual information is given by

$$I(X;Y) = H(X) - H(X|Y) = \sum_{x,y} p_{X,Y}(x,y) \log_2 \left(\frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)} \right). \quad (2.73)$$

When two random variables are uncorrelated, knowledge of Y does not change uncertainty of X , $H(X|Y)=H(X)$, leaving the mutual information $I(X;Y)=0$. On the other hand, when X and Y are maximally correlated, knowledge of Y , removes the randomness in X , i.e. $H(X|Y)=0$. This is possible when Y can be uniquely represented as a function of X ($Y=f(X)$). In that case the mutual information is simply the Shannon entropy of X , $I(X;Y)=H(X)$. Mutual information has two important properties:

Symmetry: $I(X;Y) = I(Y;X)$ and

Positivity: Simple rearrangement of Eq. (2.67) yields $I(X;Y) \geq 0$.

Relative entropy

Relative entropy gives a quantification of the separation between two probability distributions. To define relative entropy, we first define the support of a function. The support of a function $f(x)$ is the set of its arguments having non-zero images, that is $\text{supp}(f) = \{x: f(x) \neq 0\}$. For instance, a probability distribution containing one impossible event, $p = \{0.8, 0.1, 0.1, 0\}$, has support $\text{supp}(p) = \{0.8, 0.1, 0.1\}$.

The relative entropy $D(p||q)$ of two probability distributions $p(x)$ and $q(x)$ is defined as

$$D(p||q) = \begin{cases} \sum_x p(x) \log_2 \left(\frac{p(x)}{q(x)} \right), & \text{if } \text{supp}(p) \subseteq \text{supp}(q) \\ +\infty, & \text{otherwise} \end{cases} \quad (2.74)$$

Note that, the relative entropy is not symmetric, $D(p||q) \neq D(q||p)$, so it is not a proper distance measure. It is sometimes called a quasi-distance between two probabilities [1]. Its definition can also be problematic at times. For instance, if one distribution is defined as $p(0)=1$ and $p(1)=0$ and the other distribution is defined as $q(0)=1-\varepsilon$ and $q(1)=\varepsilon$, then the relative entropy is infinity even for a very small but non-zero ε , i.e. $D(p||q)=\infty$.

It is useful to relate the relative entropy to mutual information:

$$I(X;Y) = D(p_{XY}||p_X p_Y). \quad (2.75)$$

Here, $p_{XY}(x,y)$ is the joint distribution and $p_X(x)$ and $p_Y(y)$ are the marginal distributions. Thus operationally, mutual information represents the quasi-distance between the joint probability and the distribution of the corresponding random variables represented by the product of its marginals.

Conditional mutual information

Given random variables X , Y , and Z the conditional mutual information $I(X;Y|Z)$ is given by

$$I(X;Y|Z) = H(Y|Z) - H(Y|X,Z) \quad (2.76)$$

$$= H(X|Z) - H(X|Y,Z) \quad (2.77)$$

$$= H(X|Z) + H(Y|Z) - H(X,Y|Z) \quad (2.78)$$

Thus conditional mutual information formulates how the correlation between two random variables changes given knowledge of a third random variable. Depending on the specific scenario, knowledge of a particular random variable can either increase or decrease the mutual information. For example, if X , Z , and Y form a Markov chain ($X \rightarrow Z \rightarrow Y$), that is each random variable is dependent only on the previous random variable, $p_{XYZ}(x,y,z) = P_{Y|Z}(y|z)p_{Z|X}(z|x)p_X(x)$, then knowledge of Z uncorrelates X and Y , $I(X;Y|Z)=0$, although $I(X;Y) \neq 0$. On the other hand, if two independent random variables X and Y constitutes a new random variable $Z=f(X,Y)$, then knowledge of Z induces correlation between X and Y , $I(X;Y|Z) \neq 0$ although $I(X;Y)=0$. An example of this scenario could be when X and Y corresponds to two independent coin tosses with head being represented by 0 and tail being 1. If a third random variable Z , corresponds to the parity of the toss results, $Z = \{z : z = (x + y) \bmod 2\}$, then with the knowledge of Z , and the result of one coin toss, we immediately know the result of the other one.

2.3.2 Quantum Shannon theory

The initial motivation of quantum Shannon theory started with the advent of optical computing. In the 1960s, Sudarshan and Glauber did pioneering work on quantum optics [57, 58], which motivated scientists to utilise quantum systems to perform classical information processing tasks. An important question was how much classical information one can encode in a quantum state. Holevo, in 1973, answered this in his seminal paper [59]. The bound is now referred to as *Holevo bound*, which states one can encode at most one classical bit per qubit. Around the same time Helstrom developed the theory of quantum estimation and hypothesis testing [60]. The 1980s has seen some important developments in quantum information theory. In 1982 Richard Feynman described how a quantum computer can outperform classical computers when it comes to simulation of quantum systems [61]. This was the first proposal utilising quantum information and an important departure from the previous decade which focused only on exploiting classical information encoded in quantum systems. In the same year, Wootters and Zurek [62], and Dieks [63] independently proved the *no-cloning theorem*. The no-cloning theorem states that there is no universal unitary operation that can copy an arbitrary quantum state. Two years later, in 1984, Bennett and Brassard proposed a groundbreaking work on establishing a secret key using a quantum channel [64]— the famous BB84 protocol. In 1990s, the Ekert [65] and B92 [66] protocols were developed. Additionally, this decade has seen tremendous development of quantum information science, ranging from superdense coding [19], quantum teleportation [20], Shor's factorisation algorithm [67], classical capacity [68], entanglement assisted classical capacity [69], and

quantum capacity [70] of quantum channels. The 2000s, saw works on state-merging protocols [71], superactivation of quantum capacity [72], unification of different quantum protocols from a resource theoretic framework [73], and network quantum Shannon theory [74]. All these developments contributed to quantum Shannon theory being established as an interesting and actively researched field.

In this section, I will introduce some basic entropic quantities relevant to quantum Shannon theory.

von Neumann entropy

The von Neumann entropy or quantum entropy is the quantum analogue of Shannon entropy. For a density operator ρ^{A_I} (the superscript means $\rho \in A_I$), the corresponding von Neumann entropy is given by

$$S(A_I)_\rho \equiv -\text{Tr}(\rho \log_2 \rho). \quad (2.79)$$

Note that, here I am putting the system A_I as an argument with the density operator as a subscript. This representation is particularly helpful when we talk about joint quantum entropy or conditional quantum entropy. However, when there is no possibility of confusion, I will use a simpler notation $S(\rho)$. Some characteristics of the von Neumann entropy:

Relationship with eigenvalues: We observe an interesting property of von Neumann entropy when we perform a spectral decomposition on the density operator, $\rho = \sum_\lambda \lambda |\lambda\rangle\langle\lambda|$, where $\{\lambda\}$ are the eigenvalues and $\{|\lambda\rangle\}$ are the eigenvectors. It turns out that the von Neumann entropy of the density operator ρ is equal to the Shannon entropy of the distribution defined by the eigenvalues $\vec{\lambda} = \{\lambda\}$,

$$S(A_I)_\rho = H(\vec{\lambda}) = -\sum_\lambda \lambda \log_2 \lambda. \quad (2.80)$$

Non-negativity: The von Neumann entropy is non-negative for all density operators, i.e. $S(A_I)_\rho \geq 0$.

Minimum and maximum value: The minimum value of the von Neumann entropy is zero, as in the case of pure states:

$$S(A_I)_\rho = 0, \text{ if } \rho \text{ is a pure state.} \quad (2.81)$$

Intuitively, when there is no ambiguity in the state preparation such that only one state is definitely prepared, we say that the state is pure. In such cases, we can always measure the state with certainty.

On the other hand, the maximum value is equal to the logarithm of the dimension as in the case of a maximally mixed state, when we are maximally ignorant of the state preparation. For a maximally mixed state, $\rho = \mathbb{1}/d$, the von Neumann entropy is

$$S(A_I)_\rho = \log_2 d. \quad (2.82)$$

Concavity: If a density operator ρ is obtained by mixing density operators $\{\rho_x\}$ with corresponding probability distribution $p(x)$, i.e. $\rho = \sum_x p(x)\rho_x$, then concavity of von Neumann entropy ensures

$$S(A_I)_\rho \geq \sum_x p(x)S(A_I)_{\rho_x} \quad (2.83)$$

This is true because mixing always introduces more uncertainty and hence increases entropy.

Unitary invariance: Under unitary transformation, the von Neumann entropy remains unchanged,

$$S(A_I)_\rho = S(A_I)_{U\rho U^\dagger}. \quad (2.84)$$

This is true because unitary transformation only affects the eigenvectors.

Joint quantum entropy:

Joint quantum entropy extends the notion of von Neumann entropy to multi-partite systems. For a bipartite quantum system $\rho \in A_I \otimes B_I$, the joint quantum entropy is simply

$$S(A_I B_I)_\rho = -\text{Tr}(\rho^{A_I B_I} \log_2 \rho^{A_I B_I}). \quad (2.85)$$

We observe the first radical departure from classical Shannon theory when we compare joint quantum entropy of an entangled quantum system with the entropy of its subsystems. Let us consider a pure bipartite entangled state $|\psi\rangle \in \mathcal{H}_{A_I} \otimes \mathcal{H}_{B_I}$ with Schmidt decomposition $|\psi\rangle = \sum_i \sqrt{\alpha_i} |i\rangle \otimes |i\rangle$. As the density operator $\rho^{A_I B_I} = |\psi\rangle\langle\psi|$ represents a pure state, the von Neumann entropy is zero $S(A_I B_I)_\rho = 0$. However, when we consider the marginal states $\rho^{A_I} = \text{Tr}_{B_I} \rho^{A_I B_I}$ and $\rho^{B_I} = \text{Tr}_{A_I} \rho^{A_I B_I}$, we observe both of them have equal and non-zero von Neumann entropies dictated by the Schmidt numbers,

$$S(A_I)_\rho = S(B_I)_\rho = -\sum_i \alpha_i \log_2 \alpha_i \geq 0. \quad (2.86)$$

Here $S(A_I)_\rho = S(\rho')$ represents quantum entropy of the subsystem A_I when the joint state is described by $\rho^{A_I B_I}$. Similarly for $S(B_I)_\rho = S(\rho'')$.

A maximally entangled state, e.g. a Bell state, is an extreme case of the example above. The Marginals of a maximally entangled state are maximally mixed. For Bell states the von Neumann entropy of each subsystem is 1 bit—we are fully ignorant of the parts, whereas the joint state has zero entropy—we are certain of the whole system.

Arbitrary partition on a multi-partite pure state: Another interesting property of joint quantum entropy is that when we apply an arbitrary partition on a joint pure state, the resulting two subsystems have the same von Neumann entropy. For example, if we have a pure state $\rho^{A_I B_I C_I}$, then

$$S(A_I B_I C_I)_\rho = 0 \quad (2.87)$$

$$S(A_I)_\rho = S(B_I C_I)_\rho \quad (2.88)$$

$$S(B_I)_\rho = S(A_I C_I)_\rho \quad (2.89)$$

$$S(C_I)_\rho = S(A_I B_I)_\rho \quad (2.90)$$

Additivity: An interesting property of the von Neumann entropy is that it is additive for a product state.

If a joint state $\tau^{A_I B_I} = \rho^{A_I} \otimes \sigma^{B_I}$, then the von Neumann entropy follows

$$S(\tau) = S(\rho) + S(\sigma). \quad (2.91)$$

Conditional quantum entropy

The conditional quantum entropy represents how acquiring knowledge about one subsystem affects the entropy of the other. For a density operator $\rho^{A_I B_I}$, the conditional entropy $S(A_I|B_I)_\rho$ is

$$S(A_I|B_I)_\rho = S(A_I B_I)_\rho - S(B_I)_\rho. \quad (2.92)$$

Similar to the classical counterpart in Eq. (2.67), conditioning does not increase quantum entropy,

$$S(A_I)_\rho \geq S(A_I|B_I)_\rho. \quad (2.93)$$

However, unlike Eq. (2.68), conditional quantum entropy can be negative. A simple example is when ρ is a Bell state: The joint quantum entropy vanishes $S(A_I B_I)_\rho = 0$, but the subsystem has $S(B_I)_\rho = 1$, B_I being a maximally mixed state, hence the conditional quantum entropy $S(A_I|B_I)_\rho = -1$. Thus a negative conditional entropy signifies entanglement. The intuition behind the negative value can be found in *state merging protocols* [71]. The simplest version of state merging protocol considers an n -copy of a bipartite state. There are two parties each sharing specific subsystems of the bipartite state, and one party wants to send their share to the other party. The parties can use both quantum and classical channels to communicate. It turns out for a large number of copies, the sender needs to use a quantum channel $nS(A|B)$ times. Interestingly, for a state with negative conditional entropy, the sender does not need any quantum channel (not even a classical channel)—they can simply keep the entangled bipartite state and use it for future quantum communication using teleportation. This future possibility of quantum communication gives an operational interpretation of negative conditional entropy.

Conditional entropy is an interesting resource that its negative is a special quantity called *coherent information*,

$$I(A_I; B_I)_\rho \equiv -S(A_I|B_I)_\rho = S(B_I)_\rho - S(A_I B_I)_\rho. \quad (2.94)$$

Quantum mutual information

Quantum mutual information captures correlation (classical as well as quantum) between two subsystems. For a joint quantum state $\rho^{A_I B_I}$, the quantum mutual information $I(A_I; B_I)_\rho$ is given by

$$I(A_I; B_I)_\rho = S(A_I)_\rho + S(B_I)_\rho - S(A_I B_I)_\rho \quad (2.95)$$

$$= S(A_I)_\rho + I(A_I; B_I)_\rho \quad (2.96)$$

$$= S(B_I)_\rho + I(B_I; A_I)_\rho. \quad (2.97)$$

Quantum mutual information is always non-negative, $I(A_I; B_I)_\rho \geq 0$ for all ρ .

Holevo quantity

The Holevo quantity, originally defined by Holevo [59], determines how much classical information one can send using a quantum state. To give a definition of the Holevo quantity, I will first introduce the concept of a classical-quantum state. Suppose, a set of classical messages is defined by elements of

a random variable X associated with probability mass function $\{p_X(x)\}$. We can represent the entire ensemble of classical messages in density matrix formalism $\sum_x p_X(x) |x\rangle\langle x|^{X_I}$, where $|x\rangle$ represents an ONB. For each classical message $|x\rangle$ we can prepare a quantum state $\rho_x^{B_I}$ and send them to the receiver. In such a scenario, the overall state is given by a classical-quantum state $\sigma^{X_I B_I}$, where

$$\sigma^{X_I B_I} = \sum_x p_X(x) |x\rangle\langle x|^{X_I} \otimes \rho_x^{B_I}. \quad (2.98)$$

The receiver has access to only the quantum states, so he measures the density operator ρ^{B_I} where

$$\rho^{B_I} = \text{Tr}_{X_I} \sigma^{X_I B_I} = \sum_x p_X(x) \rho_x^{B_I}. \quad (2.99)$$

The Holevo quantity $\chi(\rho^{B_I})$ determines the classical information content of ρ^{B_I} and is given by

$$\chi(\rho^{B_I}) = I(X_I; B_I)_\sigma = S(B_I)_\rho - \sum_x p_X(x) S(B_I)_{\rho_x}. \quad (2.100)$$

An important property of the Holevo quantity is that it is bounded by the logarithm of dimension of B_I , $\chi(\rho^{B_I}) \leq \log_2 d_{B_I}$. For a qubit, the bound in 2.100 reduces to one bit. This means one can send at most one classical bit per qubit.

Conditional quantum mutual information (CQMI)

For a tri-partite state $\rho^{A_I B_I C_I}$, we have a conditional quantum mutual information $I(A_I; B_I | C_I)_\rho$ written as

$$I(A_I; B_I | C_I)_\rho = S(A_I | C_I)_\rho + S(B_I | C_I)_\rho - S(A_I B_I | C_I)_\rho. \quad (2.101)$$

The CQMI is always non-negative for any state ρ , $I(A_I; B_I | C_I)_\rho \geq 0$.

Quantum relative entropy

Quantum relative entropy gives a ‘quasi-distance’ measure between two density operators. To introduce quantum relative entropy, I first define the support of an operator. The support of an operator $A \in \mathcal{L}(\mathcal{H}_{A_I} \rightarrow \mathcal{H}_{A_O})$ is defined as

$$\text{supp}(A) \equiv \{|\psi\rangle \in \mathcal{H}_{A_I} : A|\psi\rangle \neq 0\}. \quad (2.102)$$

When A is Hermitian, it admits a spectral decomposition $A = \sum_i a_i |i\rangle\langle i|$. In that case $\text{supp}(A)$ is spanned by the eigenvectors,

$$\text{supp}(A) = \text{span}\{|i\rangle : a_i \neq 0\}. \quad (2.103)$$

The quantum relative entropy $D(\rho || \sigma)$ of two density operators ρ and σ is defined as

$$D(\rho || \sigma) = \begin{cases} -\text{Tr}(\rho[\log_2 \rho - \log_2 \sigma]), & \text{if } \text{supp}(\rho) \subseteq \text{supp}(\sigma) \\ +\infty, & \text{otherwise} \end{cases} \quad (2.104)$$

Note that similar to its classical analogue, $D(\rho||\sigma)\neq D(\sigma||\rho)$, hence the quantum relative entropy is not a true distance measure.

Quantum mutual information is a special case of quantum relative entropy, for a quantum state $\rho^{A_I B_I}$ with its marginals $\rho^{A_I} = \text{Tr}_{B_I}(\rho)$ and $\rho^{B_I} = \text{Tr}_{A_I}(\rho)$, the mutual information $I(A_I; B_I)_\rho$ is given by

$$I(A_I; B_I)_\rho = D(\rho||\rho' \otimes \rho''). \quad (2.105)$$

2.4 Experimental Elements

Asher Peres famously said, “Quantum phenomena do not occur in a Hilbert space. They occur in a laboratory” [75]. With that motivation, I am going to discuss how the mathematical concepts introduced in previous sections could be physically realised. There are several platforms on which quantum systems have been implemented: photons [76], cavity quantum electrodynamics [77], ion traps [78], superconducting qubits [79], nuclear magnetic resonance [80], to name a few. As my research interest lies in exploring quantum mechanics through optical phenomena, I will be focusing on photonic implementation. Specifically, I will discuss two convenient degrees of freedom of light: the polarisation and the transverse spatial modes.

2.4.1 Polarisation of light

Light being a transverse electromagnetic wave, the electric field associated with it undergoes oscillation perpendicular to the direction of propagation. The polarisation of light signifies the geometrical orientation of the vibration. The origin of polarisation is attributed to the *Spin* (also called *spin angular momentum* or SAM) of the quantized light field — ‘photons’ [81]. The word ‘spin’ was coined as it was physically interpreted as rotation of the elementary particle around some axis. However, this notion was later proved to be wrong due to the point-like nature of elementary particles. The classical analogue of spin does not exist for some elementary particles. For photons, the classical description of spin — polarisation — was well understood long before the birth of quantum mechanics. The generation and manipulation of polarisation was also well-known. Due to the availability of this technology, the polarisation is an important property used to realise a two-dimensional quantum system.

Representing quantum information with polarisation

Depending on the various directions of vibration, we ascribe different polarisation states to light. For example, we represent $|0\rangle \equiv |H\rangle$ with a horizontally polarised light, $|1\rangle \equiv |V\rangle$ with a vertically polarised light. To represent any other plane of vibration, we use superposition of these two states. Two other important orthogonal bases are diagonal $(|0\rangle + |1\rangle)/\sqrt{2} \equiv |D\rangle$ and anti-diagonal $(|0\rangle - |1\rangle)/\sqrt{2} \equiv |A\rangle$ basis and the right-circular $(|0\rangle - i|1\rangle)/\sqrt{2} \equiv |R\rangle$ and left-circular $(|0\rangle + i|1\rangle)/\sqrt{2} \equiv |L\rangle$ basis. Most natural sources of light are either unpolarised or partially polarised due to random change of the plane

of field oscillation. To represent such states, we need density matrix formalism. We represent a mixed polarised state with a density matrix ρ ,

$$\rho = \frac{1}{2} \begin{pmatrix} 1+S_1 & S_2-iS_3 \\ S_2+iS_3 & 1-S_1 \end{pmatrix}. \quad (2.106)$$

Here, S_1 , S_2 and S_3 are the Stokes parameters [82] given as

$$S_1 = \langle H|\rho|H\rangle - \langle V|\rho|V\rangle, \quad (2.107)$$

$$S_2 = \langle D|\rho|D\rangle - \langle A|\rho|A\rangle, \quad (2.108)$$

$$S_3 = \langle R|\rho|R\rangle - \langle L|\rho|L\rangle, \quad (2.109)$$

$$(2.110)$$

Comparing with Eq. (2.3), we can conclude that the Stokes parameters constitute the Bloch vector of the density matrix with the radius of the Bloch sphere being $|\vec{S}| = \sqrt{S_1^2 + S_2^2 + S_3^2}$ with $|\vec{S}|=1$ for a pure polarised light and $|\vec{S}|=0$ for an unpolarised light. Interestingly a Bloch-sphere equivalent for the polarisation, the Poincaré sphere, has been developed long before the advent of the quantum mechanics. We show it in Fig. 2.6(a).

Preparing an initial quantum state

Polarised light naturally occurs when light is reflected off the sea or snow, light scattered in the sky also has some degree of polarisation. However, generally most light sources are unpolarised. To obtain a light of desired polarisation, we need a polariser. Such polarisers can be obtained through calcite crystal, polarising beamsplitter (PBS), Glan-Taylor prism—all of which spatially separate the input light into the horizontal and vertical components, i.e. there are two projectors associated with those devices, $|H\rangle\langle H|$ and $|V\rangle\langle V|$. Alternatively, one can use a variable polarisation filter which filters out a linearly polarised component depending on the angle of the filter. For an angle θ , the relevant projector is $\Pi(\theta) := |\psi\rangle\langle\psi|$ with $|\psi\rangle = \cos\theta|H\rangle + \sin\theta|V\rangle$.

Performing a unitary transformation

To transform one particular polarisation state to a desired state, we perform a unitary transformation using a half waveplates (HWP) and a quarter waveplate. An HWP performs a rotation along the equator of the Bloch sphere and the corresponding unitary is represented as

$$U_{\text{HWP}}(\theta) := \begin{pmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{pmatrix}. \quad (2.111)$$

The QWP takes a linearly polarised light to an elliptically polarised light. The corresponding unitary is

$$U_{\text{QWP}}(\theta) := \frac{e^{i\pi/4}}{\sqrt{2}} \begin{pmatrix} 1+i\cos 2\theta & i\sin 2\theta \\ i\sin 2\theta & 1-i\cos 2\theta \end{pmatrix} \quad (2.112)$$

To produce a universal one qubit unitary, we use a combination of two QWPs and one HWP [83].

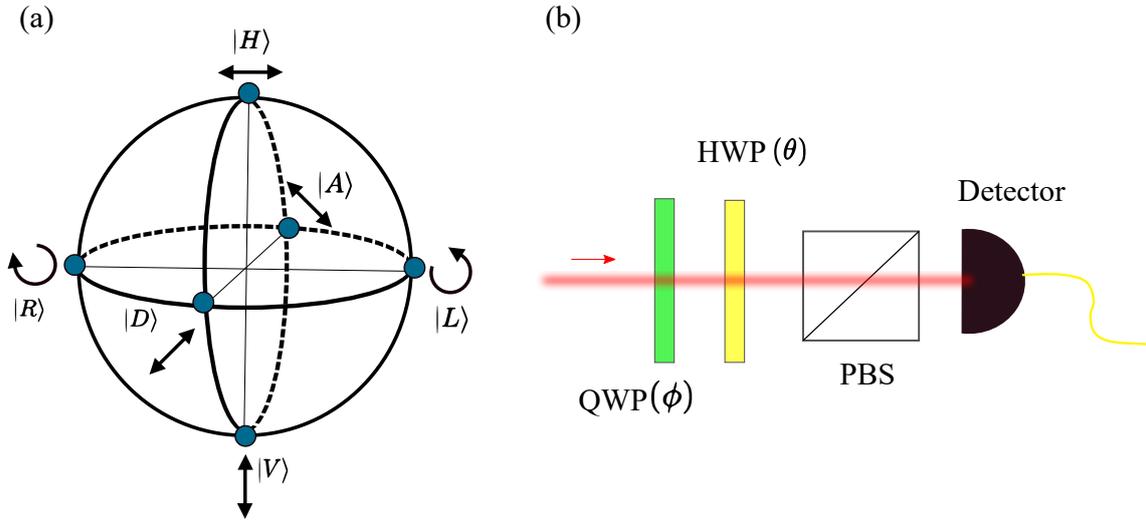


Figure 2.6: Quantum state encoded in polarisation. (a) Representing the quantum state in polarisation state of the light. The geometric representation was developed long before invention of the quantum mechanics. The representative sphere is called the Poincaré sphere. (b) Measuring the Stokes parameters. A combination of quarter waveplate (QWP), half waveplate (HWP), and polarising beamsplitter (PBS) can measure the Stokes parameters. The relevant angles are shown in Table 2.1.

Measuring the quantum state

As shown in Eq. (2.106), an arbitrary polarisation state depends on three Stokes parameters. Hence we measure those parameters to effectively reconstruct the density matrix. As shown in Fig.2.6(b), the necessary equipment are a HWP, QWP, a polarising beamsplitter, and a photon counter. Varying the angles of the waveplate, we measure different polarisation components. An overcomplete Stokes measurement is done by individually measuring all six polarisation components (horizontal, vertical, diagonal, anti-diagonal, right-circular and left- circular).

Table 2.1, shows angles of HWP and QWP to measure the corresponding polarisation components. From the photon count (n_X for the polarisation component $|X\rangle$) obtained in each setup, we construct the relevant Stokes parameter given by [82]

$$S_1 = \frac{n_H - n_V}{n_H + n_V}, \quad S_2 = \frac{n_D - n_A}{n_D + n_A}, \quad S_3 = \frac{n_L - n_R}{n_L + n_R}. \quad (2.113)$$

Component	HWP	QWP
$ H\rangle$	0	0
$ V\rangle$	$\pi/4$	0
$ D\rangle$	$\pi/8$	0
$ A\rangle$	$-\pi/8$	0
$ R\rangle$	0	$\pi/4$
$ L\rangle$	0	$-\pi/4$

Table 2.1: Angles for wave plates to measure polarisation components.

We can also perform Stokes measurement with minimal measurement, with the same setup as in Fig.2.6(b). To do an optimal measurement of Stokes parameters, we measure four variables $\{a_0, a_1, a_2, a_3\}$, with $S_1=a_1/a_0$, $S_2=a_2/a_0$, and $S_3=a_3/a_0$. Then we note that for a QWP angle ϕ and HWP angle θ , the intensity of the output light is [82]

$$I(\theta, \phi) = \frac{1}{2}(a_0 + a_1 \cos 2\theta + a_2 \cos 2\phi \sin 2\theta + a_3 \sin \phi \sin 2\theta). \quad (2.114)$$

Although the intensity of light is a classical concept, intensity is directly proportional to the photon count for monochromatic light. Thus we can replace $I(\theta, \phi)$, with photon count $n(\theta, \phi)$. Tweaking the angles of HWP and QWP, we find that

$$a_0 = n(0, 0) + n(\pi/4, 0) \quad (2.115)$$

$$a_1 = n(0, 0) - n(\pi/4, 0) \quad (2.116)$$

$$a_2 = 2n(\pi/8, 0) - a_0 \quad (2.117)$$

$$a_3 = 2n(\pi/8, \pi/2) - a_0. \quad (2.118)$$

Note that, although the above method reduces the number of measurements, generally in experiments overcomplete measurements are preferred. This is because the normalisation parameter changes depending on the imperfect experimental devices. For example, light might undergo losses non-uniformly along the waveplate angles. In the overcomplete measurement we naturally mitigate this by normalising the Stokes parameters using the angle-dependent factors.

2.4.2 Transverse spatial mode of light

Polarisation of light provides an excellent resource for quantum information processing, however, it is limited to two dimensions, that is one can pack only one bit of information per qubit. Utilising a ‘qudit’—a high-dimensional quantum state with d —levels—is advantageous for several reasons. Firstly, realising a high-dimensional quantum state paves the way to pack more information in a quantum system. A d -dimensional qudit allows a maximum information capacity of $I = \log_2 d$ bits. This means a smaller number of photons or particles can be used to achieve the same communication rate. Qudits are also advantageous in secure communication. The key advantage in quantum key distribution (QKD) is that it is impossible to perfectly copy an unknown quantum state—this provides security against an eavesdropping attack. Nevertheless, it is possible to clone imperfectly—as in optimal cloning—resulting in some leakage of information to the eavesdropper. Interestingly, the fidelity of optimal cloning, $F_{\text{clone}} = 1/2 + 1/(1 + d)$, declines with the dimension of the quantum system, from 83.3% for $d=2$ to approximately 50% for a very high-dimensional quantum system. This provides higher resistance against an eavesdropper. Consequently, the high-dimensional QKD protocol allows for higher error thresholds, from 13% for qubits to 50% for qudits [84]. One approach to access such higher-dimensional quantum systems, is to couple multiple qubits each encoded by a different degree of freedom of a quantum particle. Polarisation-path entangled photons [85, 86], time-frequency entangled photons [87, 88] are few such examples. Another approach is to use degree of freedom

that is naturally high-dimensional, such as the transverse spatial modes of light. Mair et al. [89] were the first to use the transverse spatial mode as a carrier of quantum information, they showed the entanglement of orbital angular momentum (OAM) of photons generated via spontaneous parametric downconversion (SPDC). Subsequently, there has been active research interest in this field [90–94].

$$\tilde{\psi}(x, y, z) = \alpha_{00} \Phi_{0,0}(x, y, z) + \dots + \alpha_{10} \Phi_{1,0}(x, y, z) + \dots + \alpha_{02} \Phi_{0,2}(x, y, z) + \dots + \alpha_{12} \Phi_{1,2}(x, y, z) + \dots$$

Figure 2.7: Modal decomposition of an arbitrary spatial mode of light. The figure is used from J. Pinnell *et al.*, “Modal analysis of structured light with spatial light modulators: a practical tutorial”, *JOSA A*, **37**, 2020 [95].

The transverse spatial mode of light is defined by the transverse profile of the monochromatic light field. A monochromatic light field $\Psi(x, y, z)$ propagating along z -axis, in paraxial approximation, is given by

$$\Psi(x, y, z, t) = \psi(x, y, z) \cdot e^{i(kz - \omega t)}, \quad (2.119)$$

where k is the wave-number and ω is the angular frequency of the light. The transverse profile $\psi(x, y, z)$ dictates the transverse spatial mode of the field. Two common families of spatial modes are the Hermite-Gaussian (HG) and Laguerre-Gaussian (LG) families. The HG modes are obtained by solving the wave-equation in the rectangular coordinate and given as

$$\text{HG}_{mn}(x, y, z) = H_m\left(\frac{\sqrt{2}x}{w(z)}\right) H_n\left(\frac{\sqrt{2}y}{w(z)}\right) \exp\left[-ik\frac{(x^2 + y^2)}{2q(z)} + i(m+n+1)\zeta(z)\right], \quad (2.120)$$

$$\text{where } \frac{1}{q(z)} = \frac{1}{R(z)} - i\frac{2}{kw^2(z)}, \quad w(z) = w_0 \sqrt{1 + \left(\frac{z}{z_0}\right)^2}, \quad R(z) = z \left[1 + \left(\frac{z}{z_0}\right)^2\right]. \quad (2.121)$$

Here w_0 is the beam width at $z=0$, $z_0 = kw_0^2/2$, $\zeta(z) = \tan^{-1}(z/z_0)$. $H_m(\cdot)$ and $H_n(\cdot)$ are the Hermite polynomials with mode numbers m and n respectively. The phase $(m+n+1)\zeta(z)$ is the Gouy phase [96]. The radius of curvature of the beam is denoted by $R(z)$. The LG modes are the paraxial solution of the wave equation in the cylindrical coordinates and given as

$$\text{LG}_p^l(r, \phi, z) = L_p^{|l|}\left(\frac{2r^2}{w^2(z)}\right) \exp\left[-ik\frac{r^2}{2q(z)} + i(2p + |l| + 1)\zeta(z)\right] \exp(-il\phi). \quad (2.122)$$

Here, l is the azimuthal mode index, p is the radial mode index, $L_p^{|l|}$ is the associated Laguerre polynomial, r and ϕ denotes the transverse coordinates. The Gouy phase of the LG mode is $(2p + |l| + 1)\zeta(z)$. The extra phase factor $\exp(-il\phi)$ in an LG beam gives rise to a helical phase front and gives rise to the OAM of the photon field. Note we omitted the intensity-normalisation constants in both Eqs. (2.120) and (2.122).

Representing quantum information with transverse spatial mode

We first note that transverse spatial modes form an orthonormal basis. Specifically, if a set of transverse spatial modes $\{\Phi_{a,b}(x,y,z)\}$, with variables a , and b being the mode numbers, represents solution of the paraxial wave equation in a specific coordinate, then it is an orthonormal set with the inner product defined by the overlap of the transverse field profiles,

$$\langle \Phi_{a',b'} | \Phi_{a,b} \rangle := \iint_{x,y=-\infty}^{\infty} \bar{\Phi}_{a',b'}(x,y,z) \Phi_{a,b}(x,y,z) dx dy = \delta_{aa'} \delta_{bb'}. \quad (2.123)$$

Here $\bar{\Phi}_{a',b'}(x,y,z)$ denotes the complex conjugate of the field pattern $\Phi_{a',b'}(x,y,z)$. As shown in Fig.2.7, given an ONB $\{\Phi_{a,b}(x,y,z)\}$, any arbitrary field pattern $\tilde{\psi}(x,y,z)$ can be decomposed in that basis,

$$\tilde{\psi}(x,y,z) = \sum_{a,b} \alpha_{ab} \Phi_{a,b}(x,y,z), \quad (2.124)$$

with all the coefficients being complex $\{\alpha_{ab} \in \mathbb{C}\}$ and $\sum_{ab} |\alpha_{ab}|^2 = 1$. In principle, the mode numbers can be any positive or negative real numbers ($-\infty \leq a, b \leq \infty$), allowing for arbitrarily high-dimensional quantum systems. In practice, the geometry of the experiment limits the state space to finite dimensions. The corresponding ONB can either have symmetric superposition, where the set of beams has the same Gouy number — $(m+n+1)$ for HG_{mn} beams and $(2p+|l|+1)$ for LG_p^l beams. For example, $\{HG_{11}, HG_{20}, HG_{02}\}$ is an orthonormal basis (ONB) of a three dimensional Hilbert space. In Fig.2.8(a), we show the Bloch sphere representation of a two dimensional Hilbert space with $HG_{10} \equiv |0\rangle$ and $HG_{01} \equiv |1\rangle$. It is also possible to allow asymmetric superposition, where beams of different Gouy families are taken as an ONB, for example we can consider a 3 dimensional ONB consisting of beams with OAM $\{-5, -2, 1\}$.

Preparation and measurement of the quantum state

Traditionally, the transverse spatial modes were generated using fixed diffraction gratings [98, 99]. Recent advancement in programmable spatial light modulators (SLMs) and deformable mirror displays (DMDs) has paved the way to achieve robust spatial modes with unparalleled control over its phase and amplitude profiles. To generate a desired spatial mode, we first display the computed hologram on the programmable device. When a fundamental Gaussian beam reflects off this hologram, the diffraction orders will contain the desired field (often optimised to be on the first order). The measurement procedure, also known as modal decomposition is the reverse of this process [95, 100]. To measure the overlap of an arbitrary field with a particular basis element, we shine the arbitrary field on the displayed transmission function of the basis element. The output beam turns into a fundamental Gaussian beam which is then collected to a single mode fiber and passed to a photon counter module. We show the generation and detection method in Figs. 2.8 (b) and (c).

Unitary transformation

Current technology allows for the realisation of some unitary operations in transverse spatial modes. Two important elements for this are rotating prisms and cylindrical lens pairs [101–103]. The light

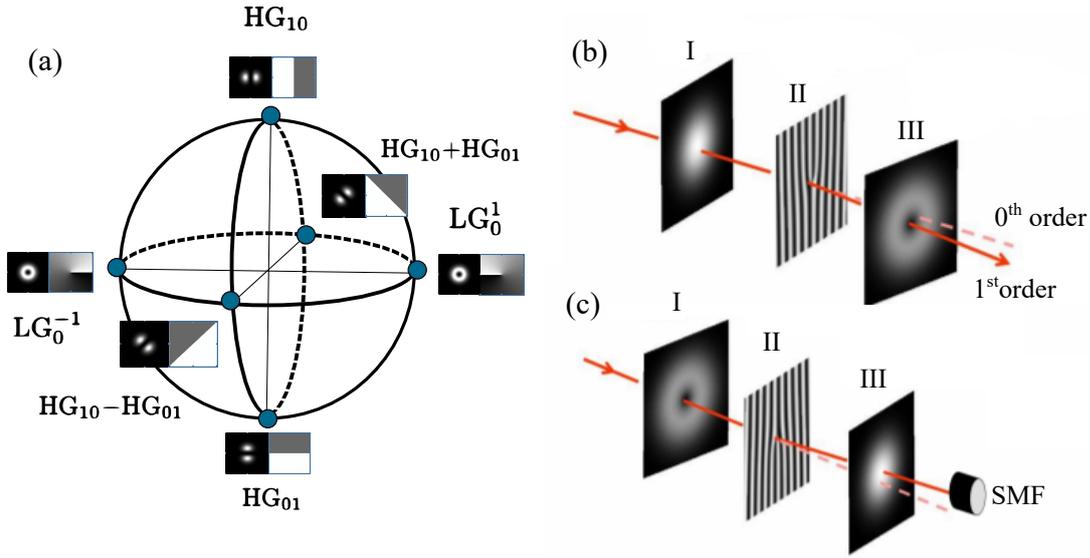


Figure 2.8: Quantum information using spatial mode of light. (a) Bloch sphere representation of spatial mode of light. (b) Generation of an orbital angular momentum (OAM) state of light. The incoming fundamental Gaussian mode of light (I) passes through a forked hologram (II). The resulting first order diffracted beam is the orbital angular momentum (OAM) state (III). (c) Measurement of an OAM state of light. To measure, the procedure is reversed. The incoming OAM state of light (I) when matches with the corresponding forked hologram (II), it turns back to the fundamental Gaussian beam in the first order diffraction (III). The resulting beam can be detected using a single mode fibre (SMF). Figs. (b) and (c) are taken from Ref. [97].

field undergoes three consecutive reflections when going through a rotating prism. Such a prism results in a rotation of the incoming spatial mode depending on the angle relative to optical axis. The rotation matrix U_1 , U_2 , for $d=2$, and 3 respectively, are given by [102]

$$U_1(\phi) = \begin{pmatrix} \cos \phi & \sin \phi \\ -\sin \phi & \cos \phi \end{pmatrix}, \quad U_2(\phi) = \begin{pmatrix} \cos^2 \phi & \frac{\sin 2\phi}{\sqrt{2}} & \sin^2 \phi \\ -\frac{\sin 2\phi}{\sqrt{2}} & \cos 2\phi & \frac{\sin 2\phi}{\sqrt{2}} \\ \sin^2 \phi & -\frac{\sin 2\phi}{\sqrt{2}} & \cos^2 \phi \end{pmatrix}. \quad (2.125)$$

Often while working with the transverse spatial mode, it is important to keep the polarisation unaffected. *Dove prisms* often also change the polarisation. Alternatively, one can use an M-shaped prism which changes the polarisation more predictably, as in a quarter waveplate [104].

One way to transform an HG beam to an LG beam is to use a pair of cylindrical lenses, which impart a selective phase shift depending on the particular basis element. The distance between the lens pairs decide how much phase shift is imparted. For example, we obtain a $\pi/2$ mode converter when a lens pair each of focal length of f , are placed $f\sqrt{2}$ distance apart. In such cases, an HG_{10} beam transmits without any change, but an HG_{01} experiences $\pi/2$ phase shift. As a result, an input beam $\psi_{in} = (HG_{10} + HG_{01})/\sqrt{2}$ transforms into $\psi_{out} = LG_0^1 = (HG_{10} + iHG_{01})/\sqrt{2}$, a Laguerre Gaussian beam. The two and three dimensional unitary matrices corresponding to a cylindrical lens pair, with a

variable phase ϕ is given by [105]

$$U_1^{\text{cyl}}(\phi) = \begin{pmatrix} e^{-i\phi/2} & 0 \\ 0 & e^{i\phi/2} \end{pmatrix}, U_2^{\text{cyl}}(\phi) = \begin{pmatrix} e^{-i\phi/2} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & e^{i\phi/2} \end{pmatrix}, \quad (2.126)$$

An arbitrary unitary, even in two dimension, requires a train of prisms and cylindrical lenses. Moreover, as rotating prisms and cylindrical lenses produces rotation around only two orthogonal axes, in higher dimension some unitary operations are inaccessible. Certain interferometric setup involving rotating prisms have been utilised to obtain high-dimensional unitaries. For example, in Ref. [106] Babazadeh et al., used rotating prisms, and spiral phase plate in an interferometric setup to implement four dimensional X , X^2 , and $X^3=X^\dagger$ gates for a set of beams with OAM $\{-2, -1, 0, 1\}$. In Ref. [107], Schlederer et al. showed four-fold cyclic transformation among the OAM $\{l\}_{l=-6}^5$. In Ref. [108] Romero and White proposed a volumetric holography technique to achieve arbitrary and controllable high-dimensional unitaries. Subsequently, in Ref. [109] Dahl et al. showed a 101 dimensional programmable X , Z and Fourier gate. Despite such unprecedented achievement, a recipe for arbitrary, programmable quantum gate with high fidelity is yet to materialise and is still an open area of research.

Chapter 3

Experimental characterisation of a non-Markovian quantum process

The following submitted manuscript is the basis of all of Chapter 3, except for the introductory material in Chapter 3.1.

[12] **K. Goswami**, C. Giarmatzi, C. Monterola, S. Shrapnel, J. Romero, F. Costa, [Experimental characterisation of a non-Markovian quantum process](#), *Phys. Rev. A* **104**, 022432 (2021).

See Section 3.6 for a breakdown of author contributions.

3.1 Introduction

In the previous chapter, I introduced the process formalism and promised some interesting applications involving the higher-order maps. In this chapter I will describe how a higher-order map can be used to represent a non-Markovian dynamic. In this work, I will use the process-matrix formalism to represent a non-Markovian higher-order quantum map. Non-Markovian dynamics emerges in open quantum system, i.e. when the system is open to interact with the environment. As shown in Steinspring's dilation theorem (Eq. (2.17)), the evolution of an open quantum system is defined by a joint unitary evolution acting on the system and the environment. To capture the evolution of a state for a time-duration t , we define a unitary U_t as

$$U_t = \exp(-iHt). \quad (3.1)$$

Here, H represents the time-independent Hamiltonian. To discuss dynamics of an open quantum system, let us consider two scenarios: a) initially separable system-environment, and b) initially correlated system-environment,

3.1.1 Initially separable system-environment

First we consider a separable initial system-environment state. For a system $\rho_s(0)$ and an environment σ_E at $t = 0$, the evolved system $\rho_s(t)$ at time t is

$$\rho_s(t) = \text{Tr}_E [U_t(\rho_s(0) \otimes \sigma_E) U_t^\dagger] = \mathcal{M}_t(\rho_s(0)). \quad (3.2)$$

Here \mathcal{M}_t is a dynamical map that evolves a system from time $T = 0$ to t . As shown in Steinspring's dilation in Eq. (2.17), \mathcal{M}_t is a CPTP map. A set of dynamical maps $\{\mathcal{M}_t : t \geq 0\}$ with $\mathcal{M}_0 = \mathbb{1}$ characterises the evolution of the system.

Although the joint system-environment state obeys the Schrödinger equation, an equivalent differential master equation for the system generally does not exist. Specifically, if the evolution of a quantum system admits a differential master equation, the system must be 'local in time'— $\rho_s(t+dt)$ depends only on $\rho_s(t)$ [110]. We refer to such time-local evolution as *Markovian* or *memoryless*. In contrast, in a *non-Markovian* evolution, the present state of the system depends on its past states.

To tackle such dynamics, we often approximate a non-Markovian evolution with a Markovian one for a large class of physical scenarios [111–114]. One such simple approximation is the Born-Markov approximation with following two assumptions.

1) *Born approximation*: The environment is reasonably large and the correlation between the system and the environment is sufficiently weak,

$$\rho_{SE}(t) \approx \rho_S(t) \otimes \sigma_E \text{ for all } t \geq 0. \quad (3.3)$$

2) *Markov approximation*: The environmental correlation produced by the system-environment interaction decays rapidly compared to the timescale of the evolution of the system. In other words, the environment is dissipative. The Markov assumption is applicable to an environment with a high temperature.

Under the Born-Markov approximation, the evolution of the quantum system is given by the Gorini, Kossakowski, Sudarshan, and Lindblad (GKSL) master equation [115, 116], also known as *Lindbladian*:

$$\frac{\partial \rho}{\partial t} = -\frac{i}{\hbar} [H, \rho] + \sum_{i=1}^{d_S^2-1} \gamma_i (L_i \rho L_i^\dagger + \{L_i^\dagger L_i, \rho\}). \quad (3.4)$$

Here, d_S is the dimension of the system, H is the Hamiltonian and the operators $\{L_i\}$ are called the *Lindblad operators*. The evolution respecting the GKSL master equation produces a set of dynamical maps which satisfy

$$\mathcal{M}_{t_1+t_2} = \mathcal{M}_{t_2} \circ \mathcal{M}_{t_1} \text{ for all } t_1, t_2 \geq 0. \quad (3.5)$$

The dynamical maps following Eq. (3.5) are called the *divisible* maps, also known as *CP-divisible*. These maps form a semigroup since they satisfy all but one group-theoretic postulate—the maps are

not invertible. This property allows us to express them in terms of the generator \mathcal{L} of the semigroup: $\mathcal{M}_t = \exp[\mathcal{L}t]$.

Interestingly, although all Markovian dynamics respect CP-divisibility, not all dynamics showing CP-divisibility is Markovian—CP-divisibility is not a sufficient condition to imply Markovianity [117].

3.1.2 Initially correlated system-environment

Let us consider scenarios where the initially prepared system is correlated with the environment. In practice this happens when preparation of the system disturbs the environment. In Ref. [118], Pechukas showed that such initial system-environment correlation fails to produce physical dynamical maps. To arrive at the conclusion, Pechukas introduced assignment maps $\zeta : S \rightarrow SE$, and defined the dynamical map \mathcal{M}_t as

$$\mathcal{M}_{t_2+t_1} = \text{Tr}_E [U_{t_2} \zeta(\rho_S(t_1)) U_{t_2}^\dagger], \quad (3.6)$$

where U_{t_2} is the joint unitary at time t_2 . The assignment maps are expected to obey following conditions:

1. *Linearity*: $\zeta(\sum_i p_i \rho_i) = \sum_i p_i \zeta(\rho_i)$,
2. *Consistency*: $\text{Tr}_E [\zeta(\rho)] = \rho$, and
3. *Positivity*: $\zeta(\rho) \geq 0$ for all ρ .

Pechukas subsequently showed that the above-mentioned conditions are met only when the initial system is uncorrelated with the environment, i.e., to allow for the initial correlation, one of the three conditions must be relaxed.

Relaxing the linearity constraint

In response to Ref. [118], Alicki [119] and Pechukas [120] decided to relax the linearity condition. However, there are several adverse consequences of giving up the linearity constraint. From a pragmatic point of view, quantum process tomography, and consequently the construction of dynamical maps is not feasible without linearity. Moreover, all experimental attempts to detect a non-linear Schrödinger evolution have failed [121]. A somewhat more severe consequence of doing away with linearity is that non-linearity allows for cloning of an arbitrary quantum state—the violation of the no-cloning theorem results in superluminal communication [22]. Finally, In Ref. [122], Jordan showed that linearity is valid with an assumption that the system can co-exist with another system without interaction.

Relaxing the consistency constraint

To preserve linearity and complete-positivity it is necessary to relax the consistency constraint. In Ref. [123] Rodríguez-Rosario et. al. showed that if the initial system-environment has ‘vanishing quantum discord’, i.e., they are only left with classical correlation, the linearity and the complete positivity condition can be ensured. Consequently, in Ref. [124], Shabani et. al. proposed that a

vanishing discord is a necessary and sufficient condition to achieve complete positivity. However, in Ref. [125] Brodutch et al. provided a counter-example where the complete positivity is not violated in the presence of a non-zero discord—the vanishing discord is not a necessary condition. This led to an erratum in [126].

Relaxing the positivity constraint:

The complete-positivity of the quantum maps results in meaningful experimental results by ensuring valid probabilities. The CP constraint is fundamental to many important results in quantum mechanics, e.g., Holevo bound, data processing inequality etc. Despite these, many researchers aimed to relax complete-positivity of the dynamical maps and favoured non-completely positive (NCP) maps [127, 128]. The assignment maps associated with such maps preserve positivity of only a subset of quantum states, $\rho_S : \zeta(\rho_S) \geq 0$, known as *compatible sets*.

3.1.3 Operational difficulty with dynamical maps

Previously I discussed how a subset of initial system-environment state allows for physical dynamical maps. This fails in the most general scenario. To perform a process tomography on the subsequent dynamical maps, we need to prepare a set of known initial states by applying projector on the initial system. The output states of the dynamical maps are subsequently measured (Fig. 3.1). This is not problematic for initially separable system-environment, as the dynamical maps are independent of the state-preparation, Fig. 3.1(a). However, in case of an initially joint state ρ_{SE} , a projection on the system affects the environmental system. Thus the reconstructed future dynamical maps are conditioned on the state-preparation, Fig. 3.1(b).

3.1.4 Resolving the initial correlation problem

From the discussion above, a major challenge in characterising a non-Markovian dynamics emerges: how to achieve an experimentally controllable input without disturbing the characteristics of the dynamics itself. To address the issue, the framework of quantum channels with memory by Kretschmann, and Werner [8], and its matrix representation, the framework of quantum combs [5, 6] play a pivotal role. Using this framework, in Ref. [10], the authors showed an operational approach: instead of describing the non-Markovian dynamics using dynamical maps, we can characterise the dynamics (initially correlated system-environment, and subsequent joint unitary evolution of the system-environment) using the higher-order maps (*process tensor formalism* as coined by Modi). Unlike the dynamical maps, the higher-order maps being positive semidefinite, Eq. (2.45), and respecting the normalisation, Eq. (2.47), it provides a physical description of the dynamics irrespective of any initial correlation. To characterise the evolution, we perform a tomography on the process matrix. As shown in Fig. 3.1(c), a tomography can be done by applying a tomographically complete set of orthogonal measure and prepare operations $\{\mathcal{A}_{kl}\}$ on the initial system, the orthogonality condition dictates that their CJ representations $\{A_{kl}\}$ obey $\text{Tr}[A_{kl}^\dagger A_{k'l'}^\dagger] = \delta_{kk'} \delta_{ll'}$. Subsequently after the joint unitary operation, we

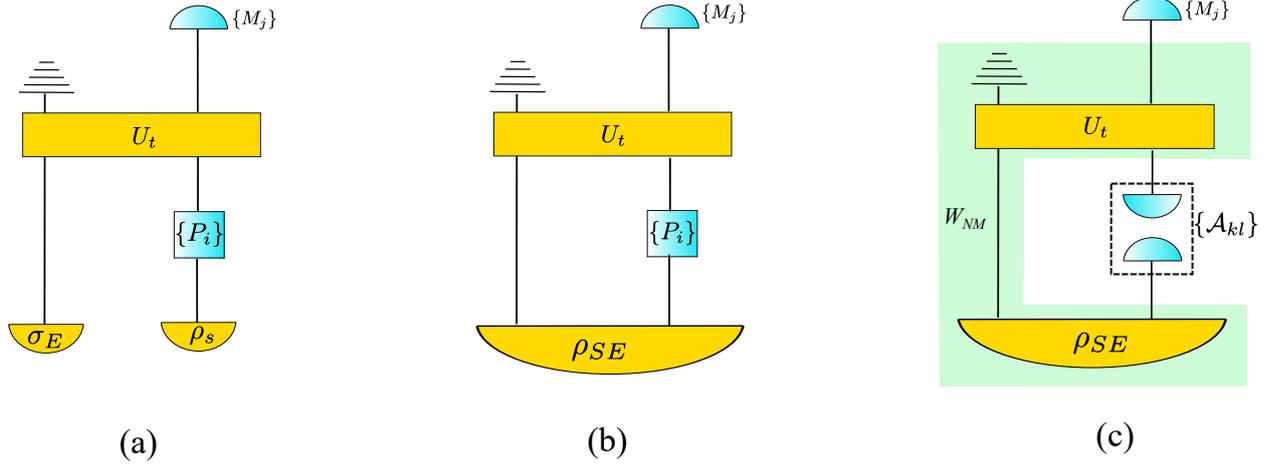


Figure 3.1: Quantum processes representing open quantum dynamics. (a) Initially separable system-environment state. Performing quantum process tomography on the subsequent dynamical map is not problematic as the set of projectors $\{P_i\}$ does not affect the environment. Hence the future dynamical map is independent of the state-preparation. (b) Initially correlated system-environment. The projectors $\{P_i\}$ affect the environment, i.e., the reconstructed dynamical maps are conditioned on the state-preparation. (c) The process formalism solves the problem. Instead of describing the evolution by dynamical maps, we can describe the dynamics by the process matrix W_{NM} . We can perform tomography to reconstruct the process matrix by applying tomographically complete measure and prepare operations $\{A_{kl}\}$ and performing state tomography at the output with $\{M_j\}$ measurement operators (see Ref. [129] for an experimental realisation).

perform the state tomography at the output with $\{M_j\}$ measurement operators. An experimental realisation of such scheme is shown by Ringbauer et.al. in Ref. [129]. Having introduced a background on non-Markovianity, I am now presenting my research.

3.2 Our work

One key feature of non-Markovian dynamics is that the environment retains some information about the past interaction with the system, which is manifested as temporal correlation among subsequent operations. A higher order process efficiently captures the noise arising from such kind of correlation. In a practical scenario where we want our quantum system to be isolated from the background, such non-Markovian noise has a deterrent effect [130, 131]. It is important to design an efficient method to characterise such non-Markovian noise [132].

Conventional noise-characterisation techniques, e.g. randomised benchmarking [133, 134] assumes Markovian evolution [135, 136]. In contrast, a complete characterisation of non-Markovian dynamics is possible via local operations on the system alone [11, 137]. A possible approach would be to perform a complete tomography on a multi-time process [138, 139]. However, such method requires measuring an exponential number of multi-time correlations. Here, we present a more resource-efficient solution. We employ machine learning models to estimate the amount of non-Markovianity—as quantified by

an information-theoretic measure— using tomographically incomplete measurement. In Ref. [140], it was shown that one can successfully train a machine learning model to estimate a measure of non-Markovianity, without full process tomography. This work, however, used only numerically simulated data, and was not tested in an experimental setting.

We use a quantum optics experiment to implement a non-Markovian process—specifically, a process with initial classical correlations between system and environment. We encode quantum states in the polarisation of photons and apply unitary transformations using waveplates. We introduce non-Markovian noise through correlated random unitaries, performed before and after a probe unitary. Our data comprises the Stokes parameters, obtained through a final measurement, conditional on choosing the probe unitary from a set of three unitaries. We train a suite of different supervised machine learning models to predict non-Markovianity—as quantified by an entropic measure introduced in Ref. [137]. Our method achieves a high accuracy in the estimation of non-Markovianity, even though the training data is far from being tomographically complete. The best results were achieved by a quadratic regression model (R^2 of 0.89 and Mean Absolute Error (MAE) of 0.045). Our work expands on the growing literature of machine-learning methods [141–145] and on the experimental characterisation of quantum non-Markovianity [129, 146–151].

In Section 3.3, we introduce the framework of the process matrix, the measure of non-Markovian noise, and procedure of our data acquisition. In Section 3.4, we describe our experiment. In Section 3.5, we analyse our experimental data using polynomial regression and present our results. In the Appendix, aside from polynomial regression on the experimental data, we present our results on the simulated data and our results obtained by other machine learning algorithms.

3.3 Theory

3.3.1 Formulation of quantum processes

Non-Markovian quantum processes are often described in terms of dynamical maps representing the evolution of the system’s reduced state [152]. However, such a description does not capture multi-time correlations mediated by the environment and can fail entirely in the presence of initial system-environment correlations [118, 153]. Here we use instead the process matrix formalism [7, 154], following a recent approach [137, 139] that has reformulated in operational terms the theory of quantum stochastic processes [155, 156]. We consider a scenario where a system of interest undergoes a sequence of arbitrary operations (such as unitaries or measurements) at well-defined instants of time. Let us label A, B, \dots the times at which the operations are performed (we can think of these labels as referring to “measurement stations”). The most general operation, say at A , is described by a Completely Positive (CP) map $\mathcal{M}^{A_I \rightarrow A_O}$ that maps the input system of the operation A_I to its output system A_O . The set of all measurement outcomes corresponds to a *quantum instrument* [157], namely a collection of CP maps $\mathcal{J}^A = \{\mathcal{M}^A\}$ that sum up to a CP and Trace Preserving (CPTP) map. Note that, as a particular case, the instrument can contain a single map, representing a deterministic operation

with no associated measurement (for example, a unitary transformation). Also, we typically take the last output system to be trivial (as the system is discarded afterwards), in which case the instrument reduces to a Positive Operator Valued Measure (POVM).

In a given quantum process, the joint probability for outcomes to occur at measurement stations A, B, \dots (corresponding to CP maps $\mathcal{M}^A, \mathcal{M}^B, \dots$) is given by

$$p(\mathcal{M}^A, \mathcal{M}^B, \dots | \mathcal{J}^A, \mathcal{J}^B, \dots) = \text{Tr}[W^{A_I A_O B_I B_O \dots} (M^{A_I A_O} \otimes M^{B_I B_O} \otimes \dots)], \quad (3.7)$$

where $M^{A_I A_O}, M^{B_I B_O}, \dots$ are the *Choi matrices* [32, 158] of the corresponding maps and $W^{A_I A_O B_I B_O \dots}$ is the process matrix that surrounds the measurement stations A, B, \dots and lives on the Hilbert space of their combined inputs and outputs. A Choi matrix, say $M^{A_I A_O} \in \mathcal{L}(\mathcal{H}^{A_I} \otimes \mathcal{H}^{A_O})$, that is isomorphic to a CP map $\mathcal{M}^A : \mathcal{L}(\mathcal{H}^{A_I}) \rightarrow \mathcal{L}(\mathcal{H}^{A_O})$, is defined as $M^{A_I A_O} := [\mathcal{J} \otimes \mathcal{M}(|\mathbb{1}\rangle\rangle\langle\langle\mathbb{1}|)]^T$. \mathcal{J} is the identity map, $|\mathbb{1}\rangle\rangle = \sum_{j=1}^{d_{A_I}} |jj\rangle \in \mathcal{H}^{A_I} \otimes \mathcal{H}^{A_I}$, $\{|j\rangle\rangle\}_{j=1}^{d_{A_I}}$ is an orthonormal basis on \mathcal{H}^{A_I} and T denotes matrix transposition in that basis and some basis of \mathcal{H}^{A_O} . The process matrix W is also known as process tensor [137], or comb [4], and it is equivalent to a quantum channel with memory [159].

In this formalism, it was found that the process matrix of a Markovian process should have the following form [11, 138, 160, 161]

$$W_M^{AB\dots} = \rho^{A_I} \otimes T^{A_O B_I} \dots, \quad (3.8)$$

where ρ is the density matrix of the initial state and by $T^{A_O B_I}$ we denote the Choi matrix of the channel $\mathcal{T}^{A \rightarrow B}$, defined as above but without the transposition—the same applies throughout the paper to all the Choi matrices of channels in a process matrix.

The form of a Markovian process matrix in Eq. (3.8) has a straightforward interpretation: just before the first operation (measurement station A), the system is in the initial state ρ . Between the first and second operation, the system evolves according to a CPTP map \mathcal{T} , which is uncorrelated with the initial state, and so on, with all evolutions independent of each other and of the initial state. Conversely, any process matrix that cannot be expressed in such a product form represents non-Markovian evolution, where the environment mediates correlations between the initial state and subsequent evolutions. To determine whether a process is Markovian, one needs first to reconstruct the process matrix from experimental data through process tomography—which generally involves non-destructive measurements at each station [138]—and then check if it W can be written in the product form [161]. In the following, we provide a method to detect non-Markovianity without having the full process matrix—instead, with incomplete data about the process, we can estimate with high accuracy a measure of non-Markovianity.

3.3.2 Our non-Markovian process

We experimentally implement a non-Markovian quantum process with memory. We implement a process with only two “stations”, A and B , and where the initial state is classically correlated with the evolution from A to B . This is a particular case of a non-Markovian process with classical memory [11].

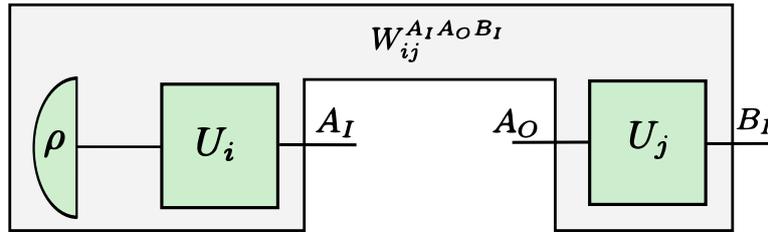


Figure 3.2: A process based on a specific instance of unitary operations U_i and U_j . The pair of unitaries occurs with probability $p(i, j)$. This makes our overall process W a convex combination of the constituent processes, W_{ij} , i.e., $W = \sum_{i,j} p(i, j) W_{ij}$. This process W , operationally, represents the environment. The experimentalist accesses the open slot $A_I A_O$ with a probe unitary U_K , as in Eq. (3.13) and B_I with a Pauli measurement.

We do this in two steps. We start with some initial state ρ followed by two operations U_i, U_j . The operations are unitaries from the Pauli group, $U_i \in \{\sigma_i, i = \{0, 1, 2, 3\}\}$ and $U_j \in \{\sigma_j, j = \{0, 1, 2, 3\}\}$, where $\sigma_0 = \mathbb{1}, \sigma_1 = X, \sigma_2 = Y, \sigma_3 = Z$. We insert A between U_i and U_j and B after U_j (Fig. 3.2). In this first step, for a given pair of unitaries (U_i, U_j) , we obtain the following Markovian process

$$W_{ij}^{A_I A_O B_I} = (\sigma_i \rho \sigma_i^\dagger)^{A_I} \otimes [[\sigma_j]]^{A_O B_I}. \quad (3.9)$$

In the second step, we simulate a non-Markovian environment by introducing correlations between the initial state and the unitary. This is done by sampling the processes W_{ij} according to some probability distribution $p(i, j)$. The resulting process matrix has the form

$$W^{A_I A_O B_I} = \sum_{i,j} p(i, j) W_{ij}^{A_I A_O B_I}. \quad (3.10)$$

To obtain processes with a varying degree of non-Markovianity, the distribution of the weights $p(i, j)$ is chosen according to the discrete random variables I and J governed by the joint probability mass function (pmf) $p(i, j) := p(I=i, J=j)$. From Eq. (3.10), it is clear that when the random variables I and J are independent, the overall process reduces to the product form of Eq. (3.9) and hence it becomes a Markovian process. To capture the non-Markovian effect, we model the joint probability $p(i, j)$ as

$$p(i, j) = p(i) [q \delta_{ij} + (1 - q) p(j)], \quad (3.11)$$

Here $q \in [0, 1]$ denotes the strength of correlation between the random variables I and J with $q=0$ being mutually independent events and $q=1$ being the maximum correlation, i.e. $\sigma_j = \sigma_i$. We assume that the marginal probabilities $p(i)$ and $p(j)$ to be the same probability mass functions. To define the probability mass function, for $p(i=0)$, we chose a random number uniformly distributed between 0 and $R \geq 1$. For the remaining $p(i \neq 0)$, we chose random numbers uniformly distributed between 0 and 1. We normalise the random numbers at the end to form a valid probability mass function. A high value of R signifies the evolution is less prone to error, i.e. the corresponding random operation is biased towards identity. Note that the process becomes Markovian with either $q=0$ or $R \rightarrow \infty$.

As a measure of non-Markovianity we use the quantum relative entropy [11, 52, 130, 137] between the process and the associated Markovian one:

$$S(\tilde{W}||\tilde{W}_{\text{Markov}}) := \text{Tr}[\tilde{W} \cdot (\log \tilde{W} - \log \tilde{W}_{\text{Markov}})], \quad (3.12)$$

where $\tilde{W}_{\text{Markov}} := \text{Tr}_{A_O B_I} \tilde{W} \otimes \text{Tr}_{A_I} \tilde{W}$ and $\tilde{W} := W/2$ is the process matrix normalised to have unit trace (obtained dividing the original process matrix by the dimension of the output system, A_O in this case).

In each realisation of $W_{ij}^{A_I A_O B_I}$ with a pair of unitaries U_i and U_j , we insert at A a unitary operation U_k and at B we perform state tomography. Each such process $W_{ij}^{A_I A_O B_I}$ has a circuit representation as shown in Figure 3.2 and an experimental realisation as shown in Fig 3.3. The unitary operations of A are a set of rotated Pauli operations

$$U_k = R_{\hat{n}}(\alpha) \sigma_k R_{\hat{n}}(\alpha)^\dagger, \quad (3.13)$$

where $k = \{0, 1, 2\}$ and $R_{\hat{n}}(\alpha)$ denotes a rotation by α , around an arbitrary axis \hat{n} in the Bloch sphere, given by

$$R_{\hat{n}}(\alpha) = \cos \frac{\alpha}{2} \mathbb{1} - i \sin \frac{\alpha}{2} (\hat{n} \cdot \vec{\sigma}), \quad (3.14)$$

$$\hat{n} \cdot \vec{\sigma} = \sin \beta \sin \gamma \sigma_1 + \cos \beta \sin \gamma \sigma_2 + \cos \gamma \sigma_3. \quad (3.15)$$

Briefly, the experimental procedure of realising a process with classical memory and taking data consists of the following steps: (1) Choosing a pair of variables (q, R) to obtain the weights $p(i, j)$, (2) Realising the processes W_{ij} , and for each one, taking data D_{ij} by running through the operations at A and B , and (3) Calculating the data $D(q, R) = \sum_{i,j} p(i, j) D_{ij}$. This final data is our input to a model that predicts the non-Markovianity of the process $W = \sum_{i,j} p(i, j) W_{ij}$.

To complete the set of training, validation, and test data for our model, we calculated the non-Markovianity for the realised processes—the label for each data $D(q, R)$. For that, we need the explicit description of the realised process matrix, which we can obtain from the above theoretical description.

We stress here that the input to the model that predicts the amount of non-Markovianity is data taken by inserting the operations A and B into the process. These provide incomplete information about the process. The full information would be provided by informationally complete operations, for example, a prepare-and-measure operation at A , and state tomography at B (with a minimum of 64 operations for a 3-qubit W , such as ours). In our case, while B performs state tomography, A performs 3 Pauli unitary operations. However, even with this incomplete information, the model is able to predict the chosen measure of non-Markovianity with $\approx 90\%$ accuracy.

3.3.3 Generating data and labels

One key point to consider in any predictive modelling is to avoid inherent bias in the training dataset. This bias can be manifested in terms of trivial transformation of the initial state. We account for this by choosing a suitable initial state, ρ , that leads to processes resulting non-trivial output data. Our choice

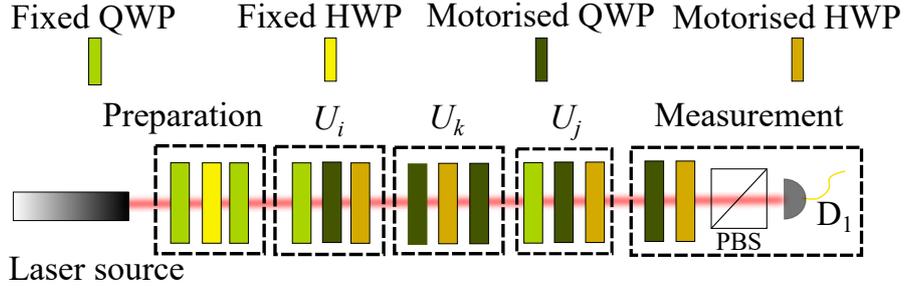


Figure 3.3: Experimental setup. We use polarisation of light to encode the quantum state. The experiment is divided into five stages — the first stage is state preparation, the second one implements the unitary U_i , the third stage represents unitary U_k , the fourth one denotes U_j and finally the last stage denotes polarisation measurements.

q	R
0.8	1
0.8	1.5
0.8	1.25
0.9	1
0.9	1.5
0.9	1.25
0.95	1
0.95	1.5
0.95	1.25
1	1

Table 3.1: Pairs of q and R to model our joint pmf as defined in Eq. (3.11). For each pair of q and R we generate 100 pmfs, thus for 10 pairs we have a total of 1000 datasets.

of state is $\rho = |\psi\rangle\langle\psi|$, with $|\psi\rangle = 0.16|0\rangle + 0.99e^{-i.0.16\pi}|1\rangle$. To model the probability mass function as in Eq. (3.11), we take the 10 pairs of q and R listed in Table 3.1.

For each pair, we generate 100 joint probability mass functions thus creating 1000 different processes as in Eq. (3.10) which are then divided into 100 groups classified by a given pair of q and R . Note that a specific instance of the experiment corresponds to a pair of unitaries sampled randomly from the underlying pmf. To experimentally realise the process in Eq. (3.10) described by a particular pmf, we need to perform repeated trials. In our experiment, we take 50 samples of each pmf. This finite sampling yields an experimentally realised process W_{exp} defined as

$$W_{\text{exp}} = \sum_{i,j} \tilde{p}(i,j) W_{ij}. \quad (3.16)$$

Here, $\tilde{p}(i,j)$ is the frequency of occurrence of the particular unitary pair (U_i, U_j) , and W_{ij} is the constituent process defined in Eq. (3.9). For each (U_i, U_j) , we apply unitary operation U_k at the second time-step as defined in Eq. (3.13) with $\alpha = \beta = \gamma = \pi/8$. As discussed earlier, we interpret U_k as an experimentally-controlled intervention, while W_{exp} simulates a noisy environment. Thus, in each instance, we have the state evolving through an overall unitary operation $U_j U_k U_i$. We measure the output state in the Pauli basis. Taking average over U_i and U_j , we get the mixed state

$\rho_k = \sum_{i,j} \tilde{p}(i,j) (U_j U_k U_i) \rho_{\text{in}} (U_i^\dagger U_k^\dagger U_i^\dagger)$. This state, when measured in σ_l basis, yields a Stokes parameter S_{lk} where

$$\begin{aligned} S_{lk} &= \text{Tr}(\sigma_l \rho_k) \\ &= \text{Tr} \left\{ \left(([U_k^*])^{A_l B_l} \otimes \sigma_l^{B_o} \right) \cdot W_{\text{exp}}^{A_l A_o B_l} \right\}. \end{aligned} \quad (3.17)$$

Note that both $k, l \in \{0, 1, 2\}$. For each process W_{exp} , we have total of 9 Stokes parameters—from now on we refer to them as *datapoints*. We evaluate the measure of the non-Markovianity associated with the process W_{exp} using Eq. (3.12) with $W = W_{\text{exp}}$ —from now on, we refer to these measures as *labels*. Thus, we have a total number of 1000 labeled data, each containing 9 datapoints and the corresponding label.

3.4 Experiment

We show the experimental schematic in Fig. 3.3. We start with a heavily attenuated laser with wavelength centred at 820 nm to create weak coherent states with 10000 counts per second. We encode the state in the photon's polarisation. Our experiment is divided into the following stages: state preparation, implementing the unitaries U_i , U_k and U_j , and state measurement. The polarisation state is prepared using a series of waveplates (Fig. 3.3). The arbitrary unitaries in polarisation were implemented using three waveplates, a half-waveplate (HWP) in between two quarter-waveplates (QWP) as in Fig. 3.3 [83]. To automate the transition between unitaries, we used motorised stages. Each U_i and U_j change within the Pauli group. For each of them we need only two motorised stages and a fixed QWP at 0° (the angles for the waveplates are given in Table 3.2). For the unitary U_k , we use three motorised stages. We control the motorised stages using a LabVIEW-controlled Newport XPS series motion controller (through a TCP/IP protocol) and a Newport SMC 100 motion controller (with serial communication to a computer). For preparing the state $|\psi\rangle$, we use another series of waveplates. Since the first QWP of U_i is set to a fixed angle at 0° , we can absorb that in the state preparation. After successful implementation of state preparation and the unitaries, we measure the Stokes parameter of the output light using a standard setup of QWP-HWP and polarising beamsplitter, as shown in Fig. 3.3.

Unitary	QWP	QWP	HWP
$\mathbb{1} \equiv \sigma_0$	0	0	0
$X \equiv \sigma_1$	0	$\frac{\pi}{2}$	$\frac{\pi}{4}$
$Y \equiv \sigma_2$	0	0	$\frac{\pi}{4}$
$Z \equiv \sigma_3$	0	$\frac{\pi}{2}$	0

Table 3.2: Angles for motorised wave plates to implement U_i and U_j .

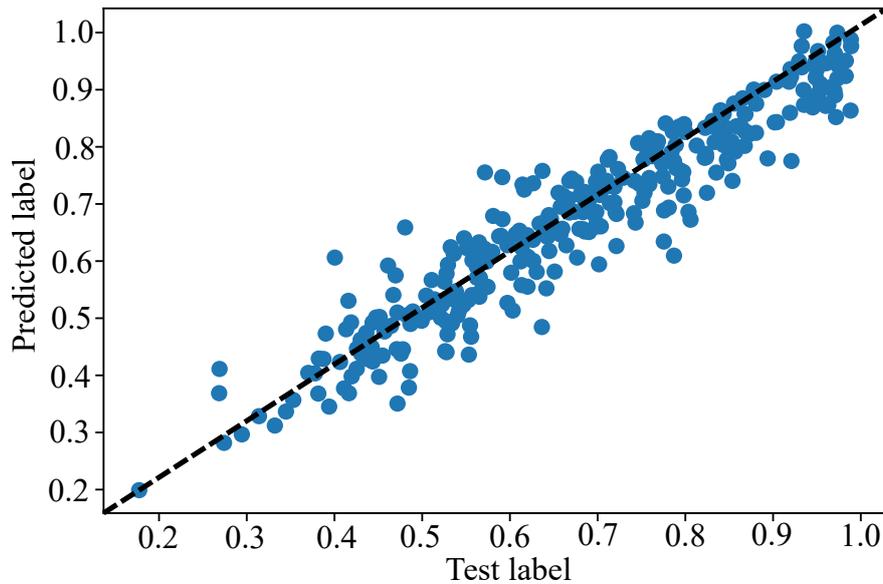


Figure 3.4: Scatter plots for the second degree polynomial regression on the experimental dataset. The y-axis represents the labels predicted by the regression model, and the x-axis represents the actual labels. The dashed black line is the best straight line that explains the data. The R^2 value associated with the plot is 0.89 and the MAE is 0.045.

3.5 Polynomial Regression

A regression model attempts to predict a relationship between a set of independent variables (datapoints) and an output variable (label) by utilising a polynomial function. Given a set of datapoints $\{x_i\}$, a polynomial regression model of degree n , finds the best prediction, \hat{y} , which is an n -degree polynomial with input arguments $\{x_i\}$. At first, to obtain a model, one uses a part of the labeled dataset, also known as *training dataset*. Once the model is obtained, to check its efficiency, one needs to employ a different group of data, known as *test dataset*. Hence, a common practice is to split the training and the test set in 7:3 ratio. To quantify the accuracy of the model of the dataset, we evaluate the R^2 value and the *Mean Absolute Error (MAE)* [162, 163]. To define these metrics, we first consider $\{y_i\}$ as our set of labels, with mean value of \bar{y} . We consider $\{\hat{y}_i\}$ as the predicted labels. With this, the metrics can be written as

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2},$$

$$MAE = \frac{\sum_i |y_i - \hat{y}_i|}{N}. \quad (3.18)$$

Here, $|\cdot|$ denotes the absolute value and N is the size of the dataset. An important aspect of a predictive algorithm is to minimise *overfitting*. The overfitting occurs when the model learns about the training set to the extent that it picks up random fluctuations in the training data to predict the labels. This results in failure to predict any additional data. To check for the overfitting, we observe R^2 and MAE score for both training data and test data. We show our results in Table 3.3. We conclude that a polynomial regression of degree 2 achieves the least overfitting with test R^2 value of 0.89 and MAE of

0.045. We show in Fig. 3.4 the scatter plot for the second degree polynomial regression. The figure demonstrates the scatter plot between the test label and the predicted label.

Deg	Train R^2	Train MAE	Test R^2	Test MAE
1	0.71	0.076	0.69	0.075
2	0.91	0.042	0.89	0.045
3	0.94	0.035	0.85	0.052

Table 3.3: polynomial regression on the experimental data of size 1000. We keep 30% of the experimental data as a test set and 70% of the same as a training set. We vary the degree of polynomial regression (Deg). To demonstrate overfitting, we show the R^2 and MAE for both training dataset and the test dataset. We observe that a polynomial regression of degree 2 achieves the least overfitting with test R^2 value of 0.89 and MAE of 0.045.

k-Fold Cross Validation: A potential issue is that a one round test-train split might result a selection bias because of the choice of test set. One way to account for it is to employ a k-fold cross validation technique [164]. In a k-fold cross-validation, the data-set is randomly divided into k equal sized groups. Out of the k groups, a single group is retained as the test set, and the remaining $k-1$ groups are the training set. Once done, in the next turn another group is selected without repetition and the entire process is iterated k-times. The results are then averaged to produce a single estimation. In our model, we use a commonly accepted value of $k=10$ [163]. We show our results in Table 3.4. This ensures an unbiased performance of our model.

Degree	R^2	MAE
1	0.69 ± 0.07	0.076 ± 0.007
2	0.89 ± 0.03	0.045 ± 0.004
3	0.87 ± 0.02	0.051 ± 0.004

Table 3.4: k-fold cross validation on our experimental dataset for polynomial regression with degree 1,2, and 3 with value of k being 10.

Varying the size of dataset: It is interesting to investigate whether the algorithm performs well while training on smaller datasets. To answer this, we fix the size of the test set to 300 and vary the length of the training set. We show our results for a second degree polynomial regression, in the Table 3.5. We observe that training set of size 210 achieves $R^2 = 0.87$ and $MAE = 0.051$. This suggests that even a small amount of experimental data is sufficient to achieve a reasonably good prediction.

Mixing with Simulated data: In practice, we may not have precise control over the environment. Hence, we ask whether assistance of simulated data augments the performance of the model. we investigate this by simulating a data set of length 14336. We proceed to vary the size of the simulated dataset and mix it with 70% of the experimental dataset to train and test on the remaining 30% of the experimental data. We observe that addition of simulated data deteriorates the performance of the model. To be precise, we see that the higher the number of simulated data, the worse the performance

LTD	Train R^2	Train MAE	Test R^2	Test MAE
70	0.86	0.057	0.15	0.123
140	0.96	0.031	0.84	0.053
210	0.94	0.036	0.87	0.051
280	0.92	0.040	0.87	0.049
350	0.91	0.042	0.88	0.047
420	0.91	0.043	0.88	0.047
490	0.91	0.043	0.89	0.046
560	0.91	0.043	0.89	0.046
630	0.91	0.043	0.89	0.046
700	0.91	0.042	0.89	0.045

Table 3.5: Second degree polynomial trained only on the experimental data of size 1000. We keep a fixed 30% of the experimental data as a test set and vary the length of training dataset (LTD). We show the R^2 and MAE for both training dataset and the test dataset. We observe even with 210 training dataset, we can achieve an R^2 value of 0.87 and MAE of 0.051.

of the model. This is due to the mismatch of the experimental and simulated training data. To circumvent this, we obtain simulated data with added white noise, potentially present in the setup. We also simulate the finite sampling that occurs in the experimental procedure (we draw 50 times from a probability distribution in Eq. 3.11). However, we do not observe an increase in performance.

Other machine learning algorithms: It is natural to expect other conventional machine learning algorithms might outperform the regression. We report this negatively. In this section, we demonstrate performance of several other standard machine learning algorithms, like K-Nearest Neighbour (KNN), Decision Tree, Random Forest, Support Vector Regression (SVR), and Gradient Boosting [163]. We split our experimental data into 70% training set and 30% test set. We show our results in Table 3.6. When we consider overfitting, Support Vector Regression (SVR) performs the best (test $R^2=0.79$, train $R^2=0.78$). Note that although Gradient boosting gives a better test R^2 , it overfits. This suggests that polynomial regression of degree 2 is still our best choice.

Algorithm	Train R^2	Test R^2	Test MAE
KNN	0.89	0.86	0.051
Decision Tree	1.0	0.64	0.081
Random Forest	0.98	0.88	0.049
SVR	0.78	0.79	0.069
Gradient Boosting	0.96	0.89	0.045

Table 3.6: Different machine learning algorithms trained on the experimental data of size 1000. We split the experimental data into 30% test set and 70% training set and report the test and train R^2 and test MAE.

3.6 Conclusion

Estimating non-Markovianity can be beneficial in practical scenarios, where the environment correlates the different time-steps of a quantum experiment. We show that with only partial information about an experimental setup, we obtain a measure of non-Markovianity with fairly high accuracy. We do this by employing different machine learning models that take as input experimental data obtained through a unitary operation and state tomography. We observe that a polynomial regression model of degree 2 achieves the best performance both in terms of overfitting and performance on the test set, which is sufficiently high ($R^2=0.87$) even with a small number of training data (500). A high score obtained by a regression model obviates the need to employ a more intensive learning algorithm, which reduces the time-complexity of the problem. This is especially beneficial to experiments where the opportunity to collect a large dataset is limited.

Our experiment is particularly interesting once we enter the large-scale quantum computation regime [165]. In this regime, correlated noise among the different gates is inevitable [166] and there is an growing interest in developing error-correcting codes for this kind of noise [167–170]. Hence, our approach provides a benchmark for further noise investigation on such multi-time-step processes.

Contributor	Statement of contribution	%
K. Goswami	Writing of text	60
	Simulated data generation	50
	Analyzing the experimental data	50
	Machine learning	50
	Performing the experiment	100
	Preparation of figures	100
	Experimental design	50
C. Giarmatzi	Writing of text	40
	Simulated data generation	50
	Analyzing the experimental data	50
	Machine learning	40
C. Monterola	Machine learning	5
S. Shrapnel	Machine learning	5
J. Romero	Proof-reading	50
	Supervision, guidance	50
F. Costa	Proof-reading	50
	Supervision, guidance	50
	Experimental design	50

Chapter 4

Indefinite causal order in a quantum switch

This Chapter is based on the publication:

[13] **K. Goswami**, C. Giarmatzi, M. Kewming, F. Costa, C. Branciard, J. Romero, and A.G. White. [Indefinite Causal Order in a Quantum Switch](#), *Phys. Rev. Lett.* **121**, 090503 (2018).

See Section 4.6 for a breakdown of author contributions.

4.1 Introduction

In the previous chapter, I showed how a process matrix can represent a non-Markovian dynamics. So far we have assumed that the local operations of the parties maintain a fixed causal order. Causal relations here are defined through the possibility of transmitting signals between events. A fixed causal order ensures if A sends some signal to B in a particular run of the experiment, B cannot send any signal to A in the same run. Operationally, an *event* associated with each party is an elementary quantum operation such as measurements, preparations, or transformations of a physical system. Note, causality in the relativistic sense naturally falls within this domain: if A is in the causal past of B it is possible to send a signal from A to B , while no signal exchange is allowed in reverse or for space-like separated events. Although, a fixed causal structure is central to our daily experience, quantum mechanics allows for a more generalised scenario with the possibility of superposition between two causal structures: “ A is in the causal past of B ” and “ B is in the causal past of A ”. We refer to such superposition of different causal orders as *indefinite causal order* [5, 7]. This counter-intuitive scenario is plausible in the light of general relativity where the causal order is dynamic rather than fixed [171]. Indefinite causal order is relevant in quantum foundations since it aims to combine general relativity and quantum mechanics [171, 172]. From a pragmatic perspective, an indefinite causal structure is advantageous to computation [173], communication complexity [174–177] and other information processing tasks [178–180]. To simulate an indefinite causal order experimentally, we implement a device called the *quantum switch* [5]. A quantum switch is composed of a target quantum system and a control quantum system. The control system controls the order of operations on the target. In this chapter, I will describe how we have implemented a quantum switch. We use polarisation of light as a control and transverse spatial mode of light as a target, the local operations are realised by a set of

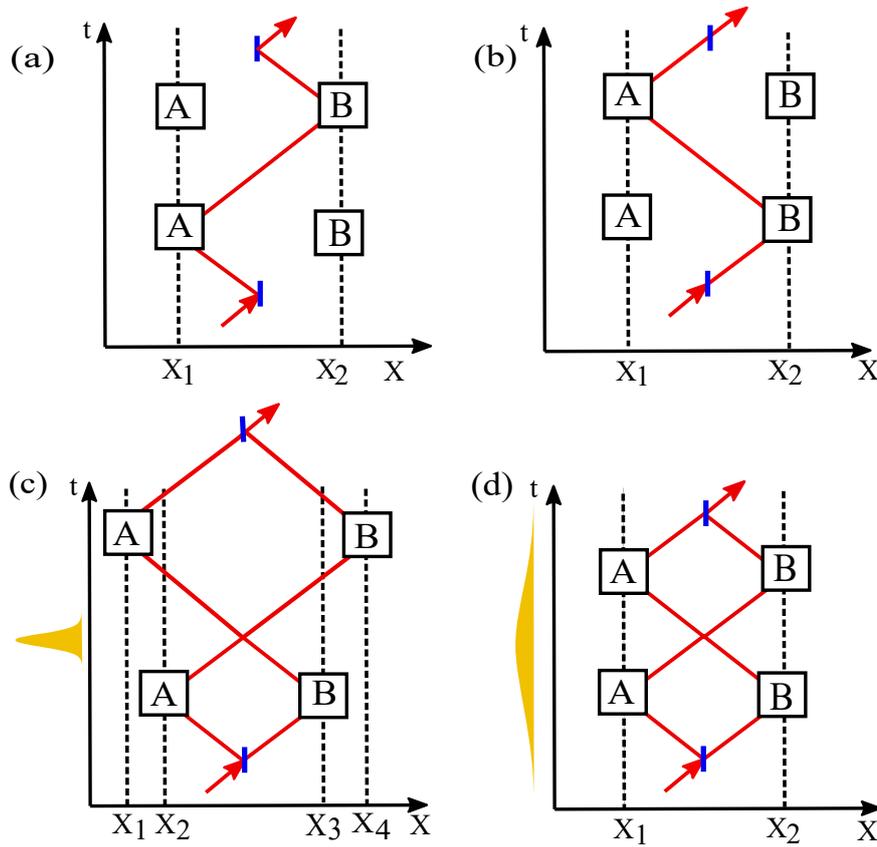


Figure 4.1: Space-time diagram of the *quantum switch*. A control qubit determines the order in which two quantum operations, \hat{A} and \hat{B} , are applied to a target qubit, ρ_t . In our experiment, the target qubit is a pure state, $\rho_t = |\psi\rangle\langle\psi|_t$. (a) When the control is $|0\rangle_c$, \hat{A} is applied before \hat{B} . (b) When the control is $|1\rangle_c$, \hat{B} is applied before \hat{A} . When the control is in the superposition $(|0\rangle + |1\rangle)_c/\sqrt{2}$, there is a superposition of the two orders, yielding the output state $|\phi\rangle = (\hat{B}\hat{A}|\psi\rangle_t \otimes |0\rangle_c + \hat{A}\hat{B}|\psi\rangle_t \otimes |1\rangle_c)/\sqrt{2}$. (c) In refs. [183, 184], the control is the transverse position at which a photon passes through a set of wave plates—consequently, the operations are performed in distinct spatial locations depending on the order. (d) In our experiment, the control is polarisation, hence each operation takes place in a fixed spatial location, independent of the order. The yellow pulses are graphical representations of the difference in temporal characteristics: in (c) the pulses are orders-of-magnitude shorter than the experiment and its internal components; in the (d) the pulses are orders-of-magnitude longer so that the operations are indistinguishable in time as well as space.

rotating prisms and cylindrical lens pairs. To verify the ‘indefiniteness’ of the causal order in our setup, we use a ‘causal witness’ [181] that distinguishes an indefinitely causal ordered process from causally separable ones. Since our experiment in 2018, there have been several realisations of quantum switch using different optical degrees of freedom. For an elaborate pedagogical review on these experiments, the readers are encouraged to read our review Ref. [182].

4.2 Quantum switch

A schematic of a quantum switch is shown in Fig. 4.1. In a quantum switch, two quantum operations \hat{A} and \hat{B} are applied to a target qubit ρ_t , their orders are controlled by a control qubit $|\psi\rangle_c$. When the

control qubit is in state $|0\rangle_c$, \hat{A} acts before \hat{B} . On the other hand, a control qubit in $|1\rangle_c$ ensures \hat{B} acts before \hat{A} . A superposition of the control qubit, for example $(|0\rangle + |1\rangle)/\sqrt{2}$, leads to a superposition of causal orders. Specifically, a switch represents a higher order map, \mathcal{S} , that transforms two input quantum channels \hat{A} and \hat{B} to a new quantum channel $\mathcal{S}(\hat{A}, \hat{B})$ with Kraus operators [2] $\mathcal{S}(\hat{A}, \hat{B})_{ij}$ given as

$$\mathcal{S}(\hat{A}, \hat{B})_{ij} = K_i^{(B)} K_j^{(A)} \otimes |0\rangle\langle 0|_c + K_j^{(A)} K_i^{(B)} \otimes |1\rangle\langle 1|_c. \quad (4.1)$$

Here, $\{K_i^{(A)}\}$ are the Kraus operators of \hat{A} — $\hat{A}(\rho_t) = \sum_i K_i^{(A)} \rho_t K_i^{(A)\dagger}$, with $\sum_i K_i^{(A)\dagger} K_i^{(A)} = \mathbb{1}$. Similarly, $\{K_i^{(B)}\}$ are the Kraus operators of \hat{B} . With this, the joint input state $\rho_t \otimes \rho_c$ transforms to the output state

$$\mathcal{S}(\hat{A}, \hat{B})(\rho_t \otimes \rho_c) = \sum_{i,j} \mathcal{S}(\hat{A}, \hat{B})_{ij} (\rho_t \otimes \rho_c) \mathcal{S}(\hat{A}, \hat{B})_{ij}^\dagger. \quad (4.2)$$

To represent the quantum switch map \mathcal{S} in process matrix formalism, we first note that the switch is composed of three parties— A corresponding to the operation \hat{A} , B corresponding to \hat{B} , and C is the party at the output of the quantum switch who generally performs a joint operation (operation on the joint target-control system) occurring after both A and B . There are two possible situations— A being in the causal past of B ($A \prec B \prec C$), and B is in the causal past of A ($B \prec A \prec C$). For a fixed input state $|\psi\rangle_t \otimes (|0\rangle_c + |1\rangle_c)/\sqrt{2}$, as in our experiment, the corresponding process matrix of the quantum switch is $W_{\text{switch}} = |w\rangle\langle w|$, where

$$|w\rangle = \frac{1}{\sqrt{2}} \left(|w^{A \prec B \prec C}\rangle \otimes |0\rangle_c^{C_c} + |w^{B \prec A \prec C}\rangle \otimes |1\rangle_c^{C_c} \right), \quad (4.3)$$

$$|w^{A \prec B \prec C}\rangle = |\psi\rangle_t^{A_I} \otimes |\mathbb{1}\rangle^{A_O B_I} \otimes |\mathbb{1}\rangle^{B_O C_I^t}, \quad (4.4)$$

$$|w^{B \prec A \prec C}\rangle = |\psi\rangle_t^{B_I} \otimes |\mathbb{1}\rangle^{B_O A_I} \otimes |\mathbb{1}\rangle^{A_O C_I^t}. \quad (4.5)$$

Here, $|\mathbb{1}\rangle^{A_O B_I} = |0\rangle^{A_O} \otimes |0\rangle^{B_I} + |1\rangle^{A_O} \otimes |1\rangle^{B_I}$ represents an identity channel from A 's outgoing space A_O to B 's incoming space B_I for the target qubit. Similarly for $|\mathbb{1}\rangle^{B_O C_I^t}$, $|\mathbb{1}\rangle^{A_O B_I}$ and $|\mathbb{1}\rangle^{A_O C_I^t}$. In the above equation, the superscripts (A_I, A_O , etc.) specify the Hilbert spaces in which each term is defined. In particular, C_I^c and C_I^t denote party C 's input Hilbert space of the control and target systems respectively. Here, dimensions of all the systems in Eqs. (4.4) and (4.5) are two.

Note that a quantum superposition of causal orders is distinct from a probabilistic mixture where the processes $W^{A \prec B \prec C} = |w^{A \prec B \prec C}\rangle\langle w^{A \prec B \prec C}|$ and $W^{B \prec A \prec C} = |w^{B \prec A \prec C}\rangle\langle w^{B \prec A \prec C}|$ occur with a probability q and $1-q$ respectively. We refer to such a probabilistic mixture of definite ordered processes as a *causally separable process* [7, 154, 181, 185] W_{sep} :

$$W_{\text{sep}} = q W^{A \prec B \prec C} + (1-q) W^{B \prec A \prec C}, \quad (4.6)$$

where $0 \leq q \leq 1$. To differentiate between an indefinite causal ordered process and a causally separable process, we need a causal witness. I discuss this in the next section.

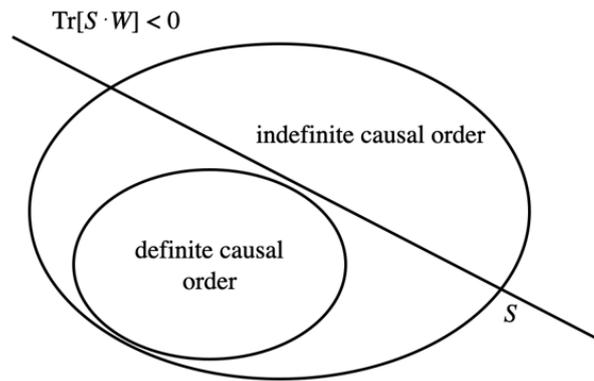


Figure 4.2: Causal witness S . A causal witness separates the processes that exhibit indefinite causal order from those that do not. For a process W , with indefinite causal order, there exists a causal witness, namely an operator S which results $\text{Tr}[S \cdot W] < 0$. For any non-separable process W_{sep} results $\text{Tr}[S \cdot W_{\text{sep}}] \geq 0$.

4.3 Causal witness

A causal witness is an observable which can certify a causally non-separable process [181, 186]. The notion of the causal witness is governed by the separating hyperplane theorem [187], which states that it is possible to find a hyperplane that separates two disjoint, closed-convex sets (see Fig. 4.2). As both causally-separable and non-separable processes form a closed convex set, they can be separated by a hyperplane. Specifically, for a particular causally non-separable process W (process that cannot be decomposed as in Eq. (4.6)) it is possible to construct a witness S such that

$$\langle S \rangle_W = \text{Tr}[SW] < 0. \quad (4.7)$$

Any causally separable process W_{sep} , on the other hand, follows $\langle S \rangle_{W_{\text{sep}}} = \text{Tr}[SW_{\text{sep}}] \geq 0$.

The search for a causal witness for a given causally nonseparable process matrix such as the quantum switch can be cast as a SemiDefinite Programming (SDP) problem that can be solved efficiently using convex optimisation techniques [181, 186].

To simplify the experiment and the construction of the causal witness, we only measure the output control system, i.e., we use a witness of the form $\mathbb{1}_{C_I^c} \otimes S_{C_I^c}$. Hence we use the reduced process matrix by tracing out the output target system C_I^c from the $W_{\text{switch}} = |w\rangle\langle w|$ [154, 181]. The reduced process matrix $\tilde{W}_{\text{switch}} = \text{Tr}_{C_I^c} |w\rangle\langle w|$ thus belong to the space $A_I \otimes A_O \otimes B_I \otimes B_O \otimes C_I^c$; it can be verified that its trace is 4, equal to the product of the dimensions of A and B 's outgoing spaces (note that in the quantum switch, C has no outgoing space—or equivalently, a trivial 1-dimensional one) [7, 154, 181].

The set \mathcal{S} of causal witnesses S , such that $\text{Tr}[SW] \geq 0$ for all causally separable process matrices $W \in \mathcal{W}_{\text{sep}}$ is simply the dual of the closed convex cone \mathcal{W}_{sep} (with respect to the Hilbert-Schmidt inner product defined by the trace). It is also a closed convex cone, and was characterized in Refs. [181, 186] in terms of linear and semidefinite constraints for the case of interest for the quantum switch, where

$\mathcal{W}_{\text{sep}}, \mathcal{S} \subset \mathcal{A}_I \otimes \mathcal{A}_O \otimes \mathcal{B}_I \otimes \mathcal{B}_O \otimes \mathcal{C}_I$. In order to prove the causal nonseparability of the quantum switch, it suffices to find a witness $S \in \mathcal{S}$ such that $\text{Tr}[S\tilde{W}_{\text{switch}}] < 0$. One may however want to optimise the choice of the witness in terms of its resistance to experimental noise. As shown in Refs. [181, 186], this can be done by minimising $\text{Tr}[S\tilde{W}_{\text{switch}}]$ over the cone \mathcal{S} . Fixing the value $\text{Tr}[S]$ to 8 (the product of the dimensions of A, B and C 's incoming spaces) ensures that the optimisation does not diverge to $-\infty$, and provides a practical interpretation for the optimal value of $-\text{Tr}[S\tilde{W}_{\text{switch}}]$ obtained as a result of the optimisation, as the random robustness of $\tilde{W}_{\text{switch}}$ with respect to white noise—i.e., $-\text{Tr}[S\tilde{W}_{\text{switch}}]$ is the minimal value of r that makes $\frac{1}{1+r}(\tilde{W}_{\text{switch}} + r\hat{I}/8)$ causally separable, where $\hat{I}/8$ represents here a maximally mixed state received by A, B and C .

The optimisation just described is a SDP problem. Further constraints can also be imposed on S . In our experiment, we want to be able to measure S by letting A and B only implement unitaries \hat{A}, \hat{B} taken from the set $\mathcal{U} = \{\hat{I}, \hat{X}, \hat{Y}, \hat{Z}, \hat{P} = \frac{\hat{Y} + \hat{Z}}{\sqrt{2}}, \hat{Q} = \frac{\hat{X} + \hat{Z}}{\sqrt{2}}\}$, and by letting C perform a measurement of \hat{X} . The statistics thus obtained allow us to calculate terms of the form $\text{Tr}[(\mathcal{A} \otimes \mathcal{B} \otimes \hat{X})\tilde{W}_{\text{switch}}]$, for $\hat{A}, \hat{B} \in \mathcal{U}$ (with \mathcal{A}, \mathcal{B} denoting their Choi representation). Note also that a trivial ‘‘measurement’’ of \hat{I} on the control qubit gives $\text{Tr}[(\mathcal{A}' \otimes \mathcal{B}' \otimes \hat{I})\tilde{W}_{\text{switch}}] = 1$ for any CP trace-preserving maps $\mathcal{A}', \mathcal{B}'$, so that such terms can also be trivially included in $\text{Tr}[S\tilde{W}_{\text{switch}}]$; in fact it suffices to consider the term with $\mathcal{A}' = \mathcal{B}' = \hat{I}/2$. Our setup thus allows us to calculate $\text{Tr}[S\tilde{W}_{\text{switch}}]$ for any witness of the form $S = \frac{1}{4}(\hat{I} + \sum_{\hat{A}, \hat{B} \in \mathcal{U}} \gamma_{\hat{A}, \hat{B}} \mathcal{A} \otimes \mathcal{B} \otimes \hat{X})$ with any real coefficients $\gamma_{\hat{A}, \hat{B}}$, which leads us to impose such a form in our optimisation problem.¹ Notice that we already included the constraint $\text{Tr}[S] = 8$ by introducing the factor $\frac{1}{4}$, as $\text{Tr}[\hat{I}] = 32$ for the identity operator \hat{I} acting here on the 2^5 -dimensional Hilbert space $\mathcal{H}^{\mathcal{A}_I} \otimes \mathcal{H}^{\mathcal{A}_O} \otimes \mathcal{H}^{\mathcal{B}_I} \otimes \mathcal{H}^{\mathcal{B}_O} \otimes \mathcal{H}^{\mathcal{C}_I}$.

The causal witness we measure in the experiment is the result of the following SDP problem:

$$\begin{aligned} & \text{minimize } \text{Tr}[S\tilde{W}_{\text{switch}}] \\ & \text{such that } S = \frac{1}{4}(\hat{I} + \sum_{\hat{A}, \hat{B} \in \mathcal{U}} \gamma_{\hat{A}, \hat{B}} \mathcal{A} \otimes \mathcal{B} \otimes \hat{X}) \in \mathcal{S}, \\ & \gamma_{\hat{A}, \hat{B}} \in \mathbb{R}. \end{aligned} \tag{4.8}$$

The nonzero coefficients $\gamma_{\hat{A}, \hat{B}}$ we thus obtained are listed in Table 4.1, together with the corresponding theoretically expected and experimentally measured values of the Stokes parameters $\langle \hat{X} \rangle_{\hat{A}, \hat{B}}$, that allow one to calculate the value of the causal witness, $\langle S \rangle = 1 + \frac{1}{4} \sum_{\hat{A}, \hat{B}} \gamma_{\hat{A}, \hat{B}} \langle \hat{X} \rangle_{\hat{A}, \hat{B}}$. The theoretically expected value is found to be $\langle S \rangle = \text{Tr}[S\tilde{W}_{\text{switch}}] \simeq -0.248$.

4.3.1 Taking experimental imperfections into account

Because of experimental imperfections in the operations that take place at events A, B and C , the witness measured in the experiment does not exactly match the ideal one. Therefore, it is in principle possible that the expectation value $\langle S_{\text{exp}} \rangle$ of the experimentally measured witness S_{exp} is negative for

¹We also checked that with our choice of unitary operations for A and B , allowing for measurements of \hat{Y} or \hat{Z} on the control qubit in C —i.e., also adding terms of the form $\gamma_{\hat{A}, \hat{B}}^{\hat{Y}} \mathcal{A} \otimes \mathcal{B} \otimes \hat{Y}$ and $\gamma_{\hat{A}, \hat{B}}^{\hat{Z}} \mathcal{A} \otimes \mathcal{B} \otimes \hat{Z}$ in S —does not help further decrease the value of $\text{Tr}[S\tilde{W}_{\text{switch}}]$ in our optimisation problem (4.8).

some causally separable process. This means that strictly speaking, an experimental negative value cannot be taken as a proof of causal nonseparability. To go around this problem, we derived a corrected causal bound $b < 0$ such that $\text{Tr}[S_{\text{exp}}W_{\text{sep}}] \geq b$ for all causally separable process matrices W_{sep} , where S_{exp} is estimated based on our understanding of the source of the experimental imperfections.

The main source of errors in the implementation of the unitaries lies in setting the angles of each optical element that realises the unitaries. We estimate the new causal bound with a simulation of our experimental causal witness S_{exp} , adding errors of 1° to the angles of each optical element in the unitaries and looking for the combination that minimises b (i.e., the worst-case scenario within the expected range of errors). We find that our witness is sufficiently robust to these errors: we varied randomly (up to $\pm 1^\circ$) the angles of each optical element in the unitaries and estimated the new causal bound with the following SDP:

$$\begin{aligned} & \text{minimize } \text{Tr}[S_{\text{exp}}W] \\ & \text{such that } W \in \mathcal{W}_{\text{sep}}. \end{aligned} \tag{4.9}$$

Using our simulation of the experimental witness S_{exp} , we ran 1000 different configurations of the experimental angles having errors of up to $\pm 1^\circ$. When we set the errors in the angles with a random sign, and a random, uniformly distributed, amplitude in $[0, 1]$, we reached a minimum value of $b = -0.029$; when the errors had a random sign but a fixed amplitude of 1° the minimum value was $b = -0.038$. Our measured value is still below these bounds.

4.4 Experiment

Our light source is a diagonally-polarised, 100 kHz linewidth (approximate coherence length of 1km) laser beam at 795 nm, in the lowest-order transverse spatial mode, the Hermite-Gaussian mode HG_{00} . We transform the beam into a HG_{10} spatial mode by first passing the beam through an element that adds a π -phase to half of the beam—a cover slip on a tip-tilt mount that spans half of the beam. The resulting spatial mode is a superposition of odd-order Hermite-Gaussian modes [188]. We then use spatial Fourier filtering to remove most of the higher-order spatial modes leaving just the HG_{10} mode. The qubit space of the target system consists of first-order spatial modes, where we define $|0\rangle = |\text{HG}_{10}\rangle$, and $|1\rangle = |\text{HG}_{01}\rangle$. The initial state of $|\psi\rangle_t$ (Fig. 5.1) is taken to be $|0\rangle$.

A polarising beamsplitter (PBS1) splits the beam into the top and bottom arms of an interferometer, see Fig. 4.3. The unitary operations in these arms, \hat{A} and \hat{B} , act on the transverse spatial mode, but should, ideally, not change the polarisation of the beam. The top and bottom arms are combined at the output polarising beamsplitter, PBS2, and the resulting mode is sent back to the other input of PBS1; this relay arm contains a telescope to ensure mode-matching, i.e., that the spatial mode that re-enters the interferometer is the same as the input spatial mode.

We realise the unitary operations \hat{A} and \hat{B} using a combination of inverting prisms [189] and cylindrical lenses [190, 191] as shown in Fig. 4.5. The inverting prisms rotate the incoming spatial mode. Unlike Dove prisms which act as poor polarisers [192], an inverting prism also acts approximately as a

\hat{A}	\hat{B}	$\gamma_{\hat{A},\hat{B}}$	$\langle \hat{X} \rangle_{\hat{A},\hat{B}}^{\text{theor.}}$	$\langle \hat{X} \rangle_{\hat{A},\hat{B}}^{\text{exp.}}$
\hat{I}	\hat{X}	-0.1967	1	0.9090 ± 0.0003
	\hat{Y}	-0.1967	1	0.8931 ± 0.0003
	\hat{Z}	-0.2775	1	0.9208 ± 0.0003
\hat{X}	\hat{I}	-0.1967	1	0.8801 ± 0.0004
	\hat{X}	-0.2572	1	0.8952 ± 0.0003
	\hat{Y}	0.2332	-1	-0.8365 ± 0.0004
\hat{Y}	\hat{Q}	0.5143	0	0.0553 ± 0.0007
	\hat{I}	-0.1967	1	0.9647 ± 0.0002
	\hat{X}	0.2332	-1	-0.8259 ± 0.0004
\hat{Z}	\hat{Y}	-0.2572	1	0.8346 ± 0.0004
	\hat{P}	0.5143	0	0.0107 ± 0.0008
	\hat{I}	-0.2775	1	0.9099 ± 0.0003
\hat{P}	\hat{Z}	-0.6131	1	0.8558 ± 0.0004
	\hat{P}	0.5143	0	0.0059 ± 0.0008
	\hat{Q}	0.5143	0	-0.0714 ± 0.0007
\hat{Q}	\hat{Y}	0.5143	0	-0.1217 ± 0.0007
	\hat{Z}	0.5143	0	-0.1539 ± 0.0007
	\hat{P}	-1.0286	1	0.9394 ± 0.0002
\hat{Q}	\hat{X}	0.5143	0	0.0177 ± 0.0007
	\hat{Z}	0.5143	0	-0.0209 ± 0.0008
	\hat{Q}	-1.0286	1	0.9659 ± 0.0002

Table 4.1: Data for the different combinations of unitary operations \hat{A} and \hat{B} . The coefficients $\gamma_{\hat{A},\hat{B}}$ were obtained by solving the semidefinite programming problem described in Eq.(S2) (for the combinations of \hat{A} and \hat{B} not shown in the table, the SDP returned a null coefficient $\gamma_{\hat{A},\hat{B}} = 0$, up to numerical precision). The expectation value $\langle \hat{X} \rangle_{\hat{A},\hat{B}}$ of the polarisation measurement on the control qubit after the unitaries \hat{A} and \hat{B} were applied is effectively a Stokes parameter. Together with $\gamma_{\hat{A},\hat{B}}$, the table lists both the theoretically expected values, $\langle \hat{X} \rangle_{\hat{A},\hat{B}}^{\text{theor.}}$, and the experimentally measured values, $\langle \hat{X} \rangle_{\hat{A},\hat{B}}^{\text{exp.}}$, of this Stokes parameter. Error bars (1σ) were calculated by propagation of error on the individual Stokes parameter with Poissonian counting statistics.

quarter-waveplate on polarisation [189], which we compensate using a combination of quarter- and half-waveplates.

Transformations for spatial modes require more optical elements than transformations for polarisation, hence in constructing the witness, we considered a tradeoff between its robustness to noise and the number of elements required to measure it in our setup. For this reason, in our experiment each operation \hat{A} and \hat{B} is chosen among one of the following six unitaries acting on the transverse spatial mode: the identity operation \hat{I} , the three Pauli operators \hat{X} , \hat{Y} and \hat{Z} , and the two linear combinations $\hat{P} = (\hat{Y} + \hat{Z})/\sqrt{2}$ and $\hat{Q} = (\hat{X} + \hat{Z})/\sqrt{2}$. These operations produce spatial modes that are either first-order Hermite-Gaussian or first-order Laguerre-Gaussian modes, thus keeping the spatial mode in the $\{|\text{HG}_{10}\rangle, |\text{HG}_{01}\rangle\}$ qubit subspace. Fig. 4.6 illustrates the resulting spatial modes for an input target qubit in the HG_{10} mode.

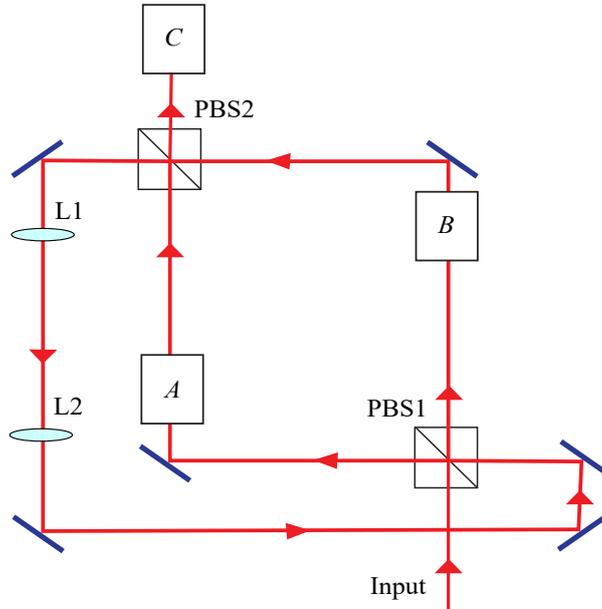


Figure 4.3: Experimental schematic. The control qubit is defined by polarisation. The polarising beamsplitter PBS1 routes the photon into either events A or B , which realise unitary operations \hat{A} or \hat{B} acting on the spatial mode of the photon. Event C is a \hat{X} polarisation measurement, determining the Stokes parameter of the photon in the diagonal/anti-diagonal basis. Lenses L1 and L2 are used as a telescope to ensure mode-matching.

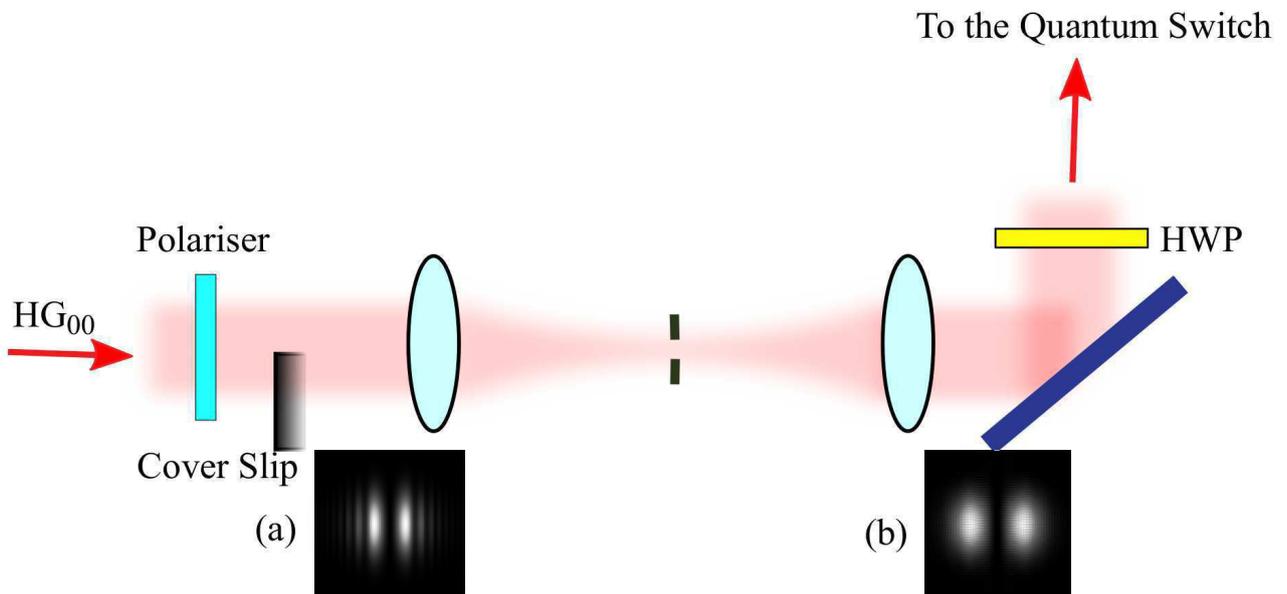


Figure 4.4: Preparation method for our input state. We started with a horizontally polarised fundamental Gaussian mode (HG_{00}). We placed a cover slip to introduce a π -phase shift to half of the beam. The resulting beam is a superposition of odd-order Hermite-Gaussian modes, as shown in (a). The higher order modes are eliminated by performing a spatial filtering on the beam, the resulting beam is in the HG_{10} mode, as shown in (b). The HWP is used to make the beam diagonally polarised. The spatial beam profiles in the figure are simulated.

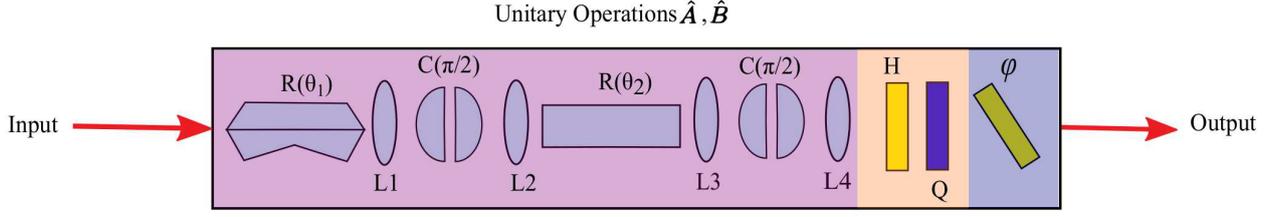


Figure 4.5: Top view of the setup for realising the unitary operations \hat{A} and \hat{B} using a set of special inverting prisms R , and pairs of cylindrical lenses C . The prisms rotate the incoming transverse mode, effectively implementing the rotation $R(\theta) = \begin{pmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{pmatrix}$ in the $\{|HG_{10}\rangle, |HG_{01}\rangle\}$ qubit subspace. The cylindrical lenses give a $\pi/2$ relative phase shift to Hermite-Gaussian components of the incoming photon, effectively implementing $C(\pi/2) = \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix}$. The spherical lenses (L) are used for mode-matching. The half-waveplates (H) and quarter-waveplates (Q) are used to correct polarisation changes caused by reflections in the prisms and φ represents a phase plate. The unitary operations of our interest are realised by varying the angles θ_1 and θ_2 . For example, in the figure $R(\theta_1)$ is rotated by 45° and for $R(\theta_2)$, the angle is set at 0° . With a 0° global phase, the above setup represents an \hat{X} operation which transforms an input Hermite-Gaussian HG_{10} beam to a Hermite-Gaussian HG_{01} .

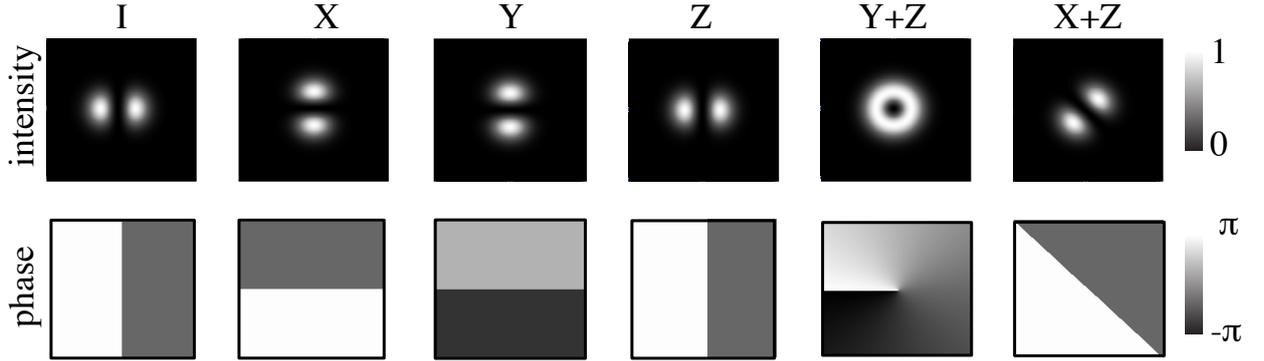


Figure 4.6: Spatial transformations. The result of the unitaries acting on an input spatial mode of HG_{10} are also first-order spatial modes.

4.4.1 Implementation of unitaries

The six unitaries necessary for the calculation of our causal witness are:

$$\hat{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \hat{X} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \hat{Y} = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \hat{Z} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \hat{P} = \frac{\hat{Y} + \hat{Z}}{\sqrt{2}} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -i \\ i & -1 \end{pmatrix} \text{ and } \hat{Q} = \frac{\hat{X} + \hat{Z}}{\sqrt{2}} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

To implement these unitaries we used the setup shown in Fig. 4.5. $C(\pi/2)$ is a mode-converter which shifts the Hermite-Gaussian mode HG_{01} by $\pi/2$ out of phase with respect to the Hermite-Gaussian HG_{10} component. $R(\theta)$ is a M-shaped rotating prism oriented at a physical angle θ , which reflects and rotates an incoming beam by 2θ . These are represented by the following matrices:

Unitary	φ	θ_1	θ_2
\hat{I}	$-\frac{\pi}{2}$	$\frac{\pi}{4}$	$\frac{\pi}{4}$
\hat{X}	0	$\frac{\pi}{4}$	0
\hat{Y}	0	0	$\frac{\pi}{4}$
\hat{Z}	0	0	0
\hat{P}	0	0	$\frac{\pi}{8}$
\hat{Q}	0	$\frac{\pi}{8}$	0

Table 4.2: Phases and angles for the unitary operations realised in our experiment, given by Eq. (S3).

$C(\pi/2) = \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix}$ and $R(\theta) = \begin{pmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{pmatrix}$. We started with an HG₁₀ mode prepared as in Fig. 4.4, using a cover slip to impart a π -phase shift to half of an incoming HG₀₀ beam, and encoded our target qubit in the spatial mode such that $|\text{HG}_{10}\rangle = |0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $|\text{HG}_{01}\rangle = |1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. After spatial Fourier filtering, a HWP was used to set the polarisation of the control.

The transformation performed by each box *A* and *B* (for a beam propagating from left to right in Fig. 4.5) can be written as

$$U(\varphi, \theta_1, \theta_2) = e^{i\varphi} C(\pi/2) R(\theta_2) C(\pi/2) R(\theta_1), \quad (4.10)$$

where the global phase φ can be imparted by a tilted phase plate which is placed only when \hat{I} is one of the operations. The rotating prisms act like a quarter-waveplate, shifting one linear polarisation by $\pi/2$ out of phase with respect to the other, hence a combination of a HWP and a QWP are used for polarisation correction. The cylindrical lens pairs in this implementation are fixed, and placed in between two spherical lenses to match the beam waists to that required by the mode-converters [193]. We only have two rotating components, the two inverting prisms. This is enough to implement the six unitaries as listed above given the following angles and phases as shown in Table 4.2.

4.4.2 Polarisation measurement at the output

At the interferometer output, after PBS2, event *C* corresponds to a polarisation measurement in the diagonal/antidiagonal basis—a measurement of the Stokes parameter corresponding to $\langle \hat{X} \rangle$ —selected using a half-waveplate and a third polarising beamsplitter. Due to experimental imperfections in the optical elements, the output mode has a marked transverse interference pattern, typically with two to three fringes. An iris is used to collect only light from one fringe, and this is then collected by a multimode fibre connected to a single-photon detector, thus tracing out the spatial mode of the photons.

For our witness, there are 21 combinations of \hat{A} and \hat{B} for which the coefficient $\gamma_{\hat{A},\hat{B}}$ is nonzero. Fig. 4.7 shows the measured Stokes values $\langle \hat{X} \rangle_{\hat{A},\hat{B}}$ for each of these combinations: the red bars are the theoretically expected values, which should all be +1, -1 or 0; the blue bars are the values measured in our experiment.

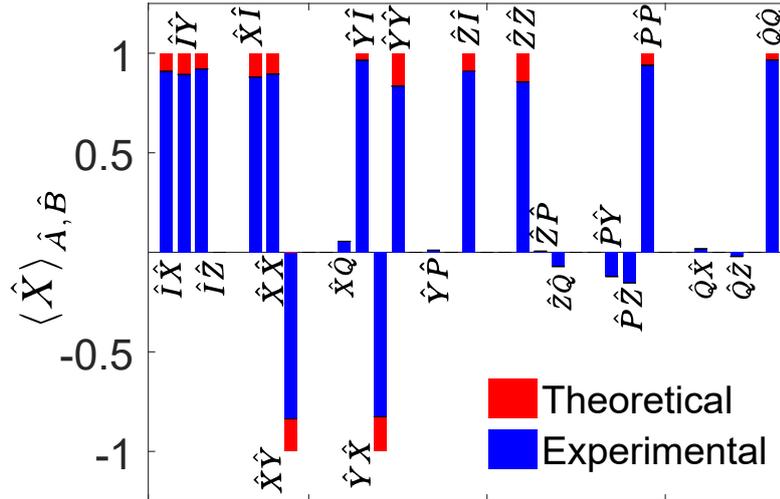


Figure 4.7: Stokes parameters $\langle \hat{X} \rangle_{\hat{A}, \hat{B}}$ obtained by measuring the polarisation of the output control qubit in the diagonal basis. The red bars show the ideal, theoretical values and the blue bars are the experimentally measured values. The unitary combinations are defined by combining the unitary operations at the top arm (\hat{A}) and the bottom (\hat{B}) arm, maintaining the order $\hat{I}, \hat{X}, \hat{Y}, \hat{Z}, \hat{P} = \frac{\hat{Y} + \hat{Z}}{\sqrt{2}}$ and $\hat{Q} = \frac{\hat{X} + \hat{Z}}{\sqrt{2}}$. A Stokes parameter of +1 means the output is diagonally polarised light, -1 means it is anti-diagonally polarised light. 1σ errors are too small to be visible in the plot.

4.4.3 Result

There are two main sources of errors in our experiment: rotational misalignments and imperfect mode matching. The inverting prisms are mounted on manual rotation stages with an uncertainty in angular position of 1° . Our witness is robust against these misalignments: accounting for these errors, one can derive a new corrected bound for causally separable processes, which we find to be close enough to zero that we still have room to obtain an experimental value below it. The imperfect mode-matching degrades the visibility of the interference of the spatial modes, which is then reflected in the values of the Stokes parameters that we obtain. We have modelled these imperfections and predict an expectation value for our causal witness within the range $-0.20 \lesssim \langle S \rangle \lesssim -0.14$, c.f. the ideal value of $\langle S \rangle \simeq -0.248$.

We measure $\langle S \rangle = -0.171 \pm 0.009$, within our expected range, and a value that is 18 standard deviations from the bound $\langle S \rangle \geq 0$ satisfied by all causally separable processes. Taking into account misalignment errors, the measured value is still 14 standard deviations below the—most conservative—corrected bound of $\langle S \rangle \geq -0.038$ for causally separable processes. This confirms that the measured process is causally nonseparable: it has no definite causal order.

The control and target systems in our experiment are encoded—as in previous experiments [183, 184, 194]—on different degrees of freedom of a single particle. As in all experiments, there are non-ideal aspects. Our experiment—and those of Refs. [183, 184]—do not overtly suffer from these

non-ideal aspects as can be seen by the high visibility observed in all implementations. The high visibility ensures the target operations are sufficiently similar for different control states, and that there is no net operation on the control.

4.5 Conclusion

Our architecture offers promising routes for further experimental investigations. Having polarisation as the control degree of freedom enables for instance using polarisation-entanglement—which can be of very high quality, e.g. reaching a tangle of $T \simeq 0.987$ [195]—as the control for entangling the causal order of different quantum switches [172, 194]. Having transverse spatial modes as the target degree of freedom enables encoding qudits—as opposed to qubits—for investigating quantum communication with indefinite causal order in larger Hilbert space dimensions [175, 180]. The benefits of using qudits for quantum information processing applications, such as improved security in quantum cryptography and higher information capacity in quantum communication, are well-known [196, 197]. Moreover, certain protocols demonstrating an advantage from indefinite causal order will require qudits for their implementation [175, 180]. Our implementation thus offers the possibility of exploring these advantages in the future. Other challenges include realising quantum switches that put more than two events in an indefinite causal order, and physically separating the control and target systems, so that the parties’ actions on the target system for different control states cannot be distinguished even in principle.

4.6 Author contribution

Contributor	Statement of contribution	%
K. Goswami	Writing of text	30
	Performing the experiment	45
	Analysing the experimental data	60
	preparation of figures	100
C. Giarmatzi	Writing of text	20
	Numerical simulation and optimisation	100
M. Kewming	Performing the experiment	5
F. Costa	proof-reading	25
	theoretical derivations	50
C. Branciard	proof-reading	25
	theoretical derivations	50
J. Romero	Writing of text	50
	Performing the experiment	50
	proof-reading	25
	supervision, guidance	50
	Analyzing the experimental data	40
	Experimental design	40
A. G. White	proof-reading	25
	supervision, guidance	50
	Experimental design	60

Chapter 5

Communicating via ignorance: Increasing communication capacity via superposition of order

The following publication has been incorporated as Chapter 5.

1. [14] **K. Goswami**, Y. Cao, G. A. Paz-Silva, J. Romero, and A. G. White, [Increasing communication capacity via superposition of order](#), *Phys. Rev. Res.* 2(3), 033292 (2020).

See Section 5.6 for a breakdown of author contributions.

5.1 Introduction

Noise is ubiquitous: communication protocols aim to optimize the amount of information that can be sent through a channel with a given amount of noise. In the limit of a completely noisy channel, no information can be transmitted [39]. This is true even with a single quantum channel [52]. Surprisingly, quantum physics offers strategies to transmit information in the scenario of two noisy channels, e.g., via superposition of path [198–200], or via superposition of order [180, 201, 202]. In these strategies the superposition in a control qubit determines the superposition in path or causal order. Placing two completely noisy channels in a superposition of paths—that is in different arms of an interferometer—allows some classical information to be communicated, at least 0.16 bits when the paths are equally weighted. Placing them instead in a superposition of causal order—that is the order in which the channels are applied is indefinite—allows up to 0.049 bits to be communicated when the orders are equally weighted [180]. Here we show that using superposition of causal order, a greater communication advantage can be achieved than superposition of paths [203], ideally up to 100% of information.

We use the quantum switch, a physically realizable process that simulates the superposition of causal order, which has been implemented in several photonic experiments [13, 176, 183, 184, 204]. We follow the experimental setup of from the previous chapter.

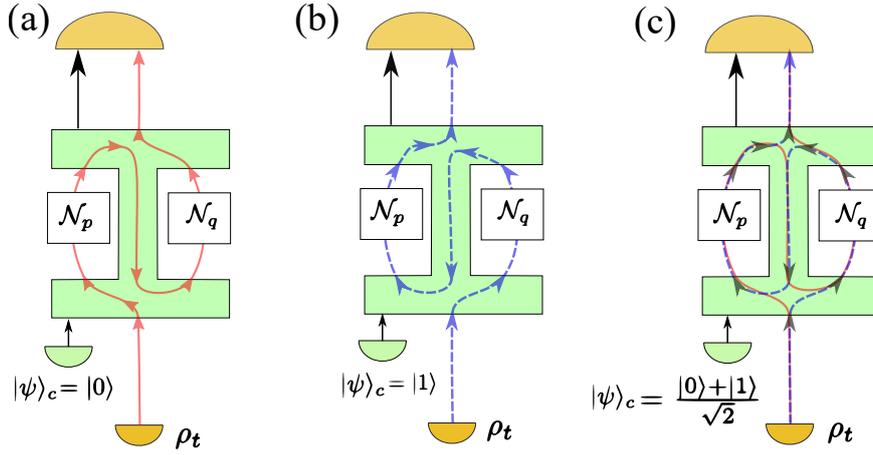


Figure 5.1: Classical communication through a quantum switch. The sender maps the classical message bit to a quantum state ρ_t —the target qubit. ρ_t passes through two noisy channels \mathcal{N}_p and \mathcal{N}_q . The order of the channels is controlled by a control qubit $|\psi_c\rangle$. (a) When the control qubit $\rho_c = |\psi_c\rangle\langle\psi_c|$ is off, i.e., $|\psi_c\rangle = |0\rangle$, the channels have a definite order $\mathcal{N}_q \circ \mathcal{N}_p$. (b) When the control qubit is on, $|\psi_c\rangle = |1\rangle$, the order is $\mathcal{N}_p \circ \mathcal{N}_q$. (c) When the control is in a superposition, $|\psi_c\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$, the channels have an indefinite order.

We analyze, and experimentally demonstrate, communication through various combinations of noisy and unitary channels in an indefinite causal order. We provide an example where perfect transmission is possible, given the freedom to choose the measurement basis of the control that determines the order. We outline the mathematical description of the channels and provide an experimental method to estimate the information-theoretic advantage quantified by the Holevo capacity, χ [59].

5.2 Background

We depict the communication protocol in Fig. 5.1. The sender maps a classical message into a quantum state $\rho_t = |\psi_t\rangle\langle\psi_t|$, which we will refer to as the target qubit. This state passes through two generalised Pauli channels \mathcal{N}_p and \mathcal{N}_q . We describe the noisy channels acting on the target qubit ρ_t as

$$\mathcal{N}_p(\rho_t) = \sum_{i=0}^3 p_i \sigma_i \rho_t \sigma_i^\dagger \quad (5.1)$$

$$\mathcal{N}_q(\rho_t) = \sum_{i=0}^3 q_i \sigma_i \rho_t \sigma_i^\dagger, \quad (5.2)$$

where $\sum_i p_i = \sum_i q_i = 1$. As the equation suggests, each Pauli channel is a probabilistic mixture of all three Pauli errors via the identifications: $\sigma_1 \equiv \sigma_x$ (bit flip); $\sigma_3 \equiv \sigma_z$ (phase flip); $\sigma_2 \equiv \sigma_y$ (combination of bit flip and phase flip); and $\sigma_0 \equiv I$. The order that these two channels are applied to ρ_t is selected by a control qubit ρ_c . If the control is off, $|0\rangle_c$, then the order is $\mathcal{N}_q \circ \mathcal{N}_p$, i.e., \mathcal{N}_p is before \mathcal{N}_q . If the control is on, $|1\rangle_c$, then $\mathcal{N}_p \circ \mathcal{N}_q$. However, if the control qubit is in a superposition state, $|\psi\rangle_c = (|0\rangle + |1\rangle)_c/\sqrt{2}$, the channels \mathcal{N}_p and \mathcal{N}_q have indefinite causal order. Specifically, given a

target qubit ρ_t and a control qubit initially in the state $\rho_c = |\psi_c\rangle\langle\psi_c|$, where $|\psi_c\rangle = \sqrt{\gamma}|0\rangle + \sqrt{1-\gamma}|1\rangle$, the total output state of the switch becomes

$$\mathfrak{s}[\mathcal{N}_p, \mathcal{N}_q](\rho_c \otimes \rho_t) = \sum_{i,j} K_{ij}(\rho_c \otimes \rho_t) K_{ij}^\dagger \quad (5.3)$$

with

$$K_{ij} = p_i q_j (|0\rangle\langle 0| \otimes \sigma_i \sigma_j + |1\rangle\langle 1| \otimes \sigma_j \sigma_i) \quad (5.4)$$

In matrix representation, the overall output state becomes

$$\mathfrak{s}[\mathcal{N}_p, \mathcal{N}_q](\rho_c \otimes \rho_t) = \begin{pmatrix} A & B \\ B & \tilde{A} \end{pmatrix}, \quad (5.5)$$

with,

$$\begin{aligned} A &= \gamma \mathcal{N}_q \circ \mathcal{N}_p(\rho_t), \\ \tilde{A} &= (1-\gamma) \mathcal{N}_p \circ \mathcal{N}_q(\rho_t), \\ B &= \sqrt{\gamma(1-\gamma)} (\varepsilon_+(\rho_t) - \varepsilon_-(\rho_t)), \end{aligned} \quad (5.6)$$

where (ε_-) ε_+ represents an auxiliary trace non preserving map composed of all the operators from \mathcal{N}_p and \mathcal{N}_q that (anti-commute) commute. That is,

$$\varepsilon_+(\rho_t) = \sum_{i=0}^3 p_i q_i \rho_t + \sum_{i=0}^3 r_{0i} \sigma_i \rho_t \sigma_i^\dagger \quad (5.7)$$

$$\varepsilon_-(\rho_t) = r_{12} \sigma_3 \rho_t \sigma_3^\dagger + r_{23} \sigma_1 \rho_t \sigma_1^\dagger + r_{31} \sigma_2 \rho_t \sigma_2^\dagger, \quad (5.8)$$

with $r_{ij} = p_i q_j + p_j q_i$. Note that $\mathcal{N}_p \circ \mathcal{N}_q(\rho_t) = \varepsilon_+(\rho_t) + \varepsilon_-(\rho_t) = \mathcal{N}_q \circ \mathcal{N}_p(\rho_t)$, any definite order of \mathcal{N}_p and \mathcal{N}_q will have the same effect on the target qubit. To take an extreme example, if \mathcal{N}_p and \mathcal{N}_q were both completely depolarizing, either of the definite orders $\mathcal{N}_p \circ \mathcal{N}_q(\rho_t)$ or $\mathcal{N}_q \circ \mathcal{N}_p(\rho_t)$ will completely scramble the target qubit.

Interestingly, the output of a quantum switch [Eqs. 5.5 and 5.6] implies that some information is contained in the control qubit. Depending on the outcome of a measurement in the σ_1^c basis of the control qubit, we obtain either of the conditional states $\varepsilon_+(\rho_t)$ or $\varepsilon_-(\rho_t)$. This means that a σ_1^c measurement allows us to estimate information encoded in the target qubit, and confirm whether there is a communication advantage when \mathcal{N}_p and \mathcal{N}_q are in an indefinite order.

In our quantum switch experiment, the control qubit is the polarization of light. A σ_1^c measurement in this case is equivalent to a Stokes measurement $S_2(\mathfrak{s}[\mathcal{N}_p, \mathcal{N}_q])$ at the output of the quantum switch $\mathfrak{s}[\mathcal{N}_p, \mathcal{N}_q]$. As Eq. 5.2 suggests, the channels \mathcal{N}_p and \mathcal{N}_q can be constructed from combinations of Pauli operations $\{\sigma_i\}$ acting on the target qubit. In the following subsection, we show how $S_2(\mathfrak{s}[\mathcal{N}_p, \mathcal{N}_q])$ can be calculated from Stokes measurements $S_2(\mathfrak{s}[\sigma_i, \sigma_j])$ at the output of a quantum switch of σ_i and σ_j , $\mathfrak{s}[\sigma_i, \sigma_j]$. This means that rather than physically implementing $\mathfrak{s}[\mathcal{N}_p, \mathcal{N}_q]$, we can simply use Stokes measurements from $\mathfrak{s}[\sigma_i, \sigma_j]$, so long as we keep track of the Pauli operations we perform, in effect using an additional memory.

5.2.1 Estimating the Stokes parameter of the quantum switch

With the Pauli decomposition discussed above, we now show how measurements of the control qubit, $\tilde{\rho}_c$, can be used to estimate the Holevo capacity. Since our control qubit is polarization, we express the output control in terms of the Stokes vector (S_1, S_2, S_3) [82]:

$$\tilde{\rho}_c = \frac{1}{2} \begin{pmatrix} 1+S_1 & S_2+iS_3 \\ S_2-iS_3 & 1-S_1 \end{pmatrix}. \quad (5.9)$$

From Eq. 5.5, if we measure the output control state, $\tilde{\rho}_c$ becomes,

$$\begin{aligned} \tilde{\rho}_c &= \begin{pmatrix} \text{Tr}(A) & \text{Tr}(B) \\ \text{Tr}(B) & \text{Tr}(\tilde{A}) \end{pmatrix} \\ &= \begin{pmatrix} \gamma & \text{Tr}(B) \\ \text{Tr}(B) & 1-\gamma \end{pmatrix}, \end{aligned} \quad (5.10)$$

where the second equality comes from the fact that, Pauli channels are trace preserving. It is easy to see that $\text{Tr}(B)$ is a real number, so comparing Eq. 5.9 and 5.10 at $\gamma=1/2$, we can see that $S_2=2\text{Tr}(B)$ and $S_1=S_3=0$. Thus we consider the effect of the quantum switch on the S_2 component of the control, note that to distinguish the operations on the control and target, we use $\sigma_0^c, \sigma_1^c, \sigma_2^c$ and σ_3^c for the control qubit and $\sigma_0, \sigma_1, \sigma_2$ and σ_3 for the target qubit,

$$\begin{aligned} S_2(\mathfrak{s}[\mathcal{N}_p, \mathcal{N}_q]) &= \text{Tr} \{ (\sigma_1^c \otimes \sigma_0) \cdot \mathfrak{s}[\mathcal{N}_p, \mathcal{N}_q] \} \\ &= \sum_{i,j} p_i q_j \text{Tr} \{ (\sigma_1^c \otimes \sigma_0) \cdot \mathfrak{s}[\sigma_i, \sigma_j] \} \\ &= \sum_{i,j} p_i q_j S_2(\mathfrak{s}[\sigma_i, \sigma_j]) \end{aligned} \quad (5.11)$$

The above equation shows that with knowledge of the control qubit for individual combinations of σ_i, σ_j we can get the S_2 of the switch with channels \mathcal{N}_p and \mathcal{N}_q .

From Eq. 5.11, we get $\tilde{\rho}_c$, and we can calculate the von Neumann entropy $H(\tilde{\rho}_c)$, which is necessary to obtain the Holevo capacity as shown in Eq. 5.13. The other quantity needed to evaluate 5.13 is the minimum entropy of the total output $H^{\min}(\mathfrak{s}[\mathcal{N}_p, \mathcal{N}_q])$. In order to calculate this, let us first write the action of the quantum switch, in terms of individual combinations of Pauli operations:

$$\begin{aligned} \mathfrak{s}[\mathcal{N}_p, \mathcal{N}_q] (\rho_c \otimes \rho_t) &= \sum_{i,j} p_i q_j (\gamma |0\rangle\langle 0|_c \otimes \sigma_i \sigma_j \rho_t \sigma_j^\dagger \sigma_i^\dagger \\ &\quad + (1-\gamma) |1\rangle\langle 1|_c \otimes \sigma_j \sigma_i \rho_t \sigma_i^\dagger \sigma_j^\dagger \\ &\quad + \sqrt{\gamma(1-\gamma)} |0\rangle\langle 1|_c \otimes \sigma_i \sigma_j \rho_t \sigma_i^\dagger \sigma_j^\dagger \\ &\quad + \sqrt{\gamma(1-\gamma)} |1\rangle\langle 0|_c \otimes \sigma_j \sigma_i \rho_t \sigma_j^\dagger \sigma_i^\dagger) \\ &= \sum_{i,j} p_i q_j \mathfrak{s}[\sigma_i, \sigma_j] (\rho_c \otimes \rho_t) \end{aligned} \quad (5.12)$$

From Eq. 5.12, we notice that from pairwise combinations of σ_i and σ_j operators we can find out the output $\mathfrak{s}[\mathcal{N}_p, \mathcal{N}_q](\rho_c \otimes \rho_t)$ and its minimum entropy $H^{\min}(\mathfrak{s}[\mathcal{N}_p, \mathcal{N}_q])$, given we use an optimized target state ρ_t . Note that, the operations in $\{\sigma_i\}$ either commute or anti-commute. We can construct $\mathfrak{s}[\sigma_i, \sigma_j](\rho_c \otimes \rho_t)$ by projecting the control qubit into the diagonal/antidiagonal basis which results to a product state. Denoting the anti-commutator by $\{\dots\}$ and the commutator by $[\dots]$, we have, with $\gamma=1/2$,

$$\mathfrak{s}[\sigma_i, \sigma_j](\rho_c \otimes \rho_t) = \begin{cases} |+\rangle\langle +|_c \otimes \{\sigma_i, \sigma_j\} \rho_t \{\sigma_i, \sigma_j\}^\dagger \\ \text{for } [\sigma_i, \sigma_j]=0 \\ \\ |-\rangle\langle -|_c \otimes [\sigma_i, \sigma_j] \rho_t [\sigma_i, \sigma_j]^\dagger \\ \text{for } \{\sigma_i, \sigma_j\}=0 \end{cases} .$$

The value of the Stokes parameter $S_2(\mathfrak{s}[\sigma_i, \sigma_j])$, which is 1 (−1) for commuting (anti-commuting) operations in the ideal case, can be experimentally obtained by noting that $S_2(\mathfrak{s}[\sigma_i, \sigma_j]) = \text{Tr}[(\sigma_1^c \otimes \sigma_0) \mathfrak{s}[\sigma_i, \sigma_j](\rho_c \otimes \rho_t)]$, i.e., expectation value of the operator $\sigma_1^c \otimes \sigma_0$.

The Stokes parameters $S_2(\mathfrak{s}[\sigma_i, \sigma_j])$ are important because we use these to obtain the output control qubit $\tilde{\rho}_c$ after tracing out the target qubit. With knowledge of $\tilde{\rho}_c$, we can then calculate the Holevo capacity $\chi(\mathfrak{s}[\mathcal{N}_p, \mathcal{N}_q])$ —a measure of the maximum amount of classical information that can be transferred through our arrangement of indefinitely ordered channels. This is given by [180]:

$$\chi(\mathfrak{s}[\mathcal{N}_p, \mathcal{N}_q]) = 1 + H(\tilde{\rho}_c) - H^{\min}(\mathfrak{s}[\mathcal{N}_p, \mathcal{N}_q]), \quad (5.13)$$

where $H^{\min}(\mathfrak{s}[\mathcal{N}_p, \mathcal{N}_q])$ is the von-Neumann entropy of the two-qubit output of the quantum switch, minimised over the input target state. On the other hand, $H(\tilde{\rho}_c)$ is the von-Neumann entropy of the output control qubit $\tilde{\rho}_c$.

5.3 Experiment

In our experiment, we implemented the quantum switch with all 16 pairs of σ_i and σ_j , as shown in Table 5.1. We measured $S_2(\mathfrak{s}[\sigma_i, \sigma_j])$, i.e., the diagonal and anti-diagonal components of the polarization of the output light, for each pair σ_i, σ_j , with the control qubit initialised to $|\psi\rangle_c = (|0\rangle + |1\rangle)_c / \sqrt{2}$ —diagonally polarized light. We measure over 10 s; the measured rate at the control output is around 100,000 counts per second. Table 5.1 summarizes the results. The first two columns are the ideal settings for the target operations on the transverse spatial mode, σ_i, σ_j , the third and fourth columns are the ideal output transformations, $\tilde{\rho}_t, \tilde{\rho}_c$ of the target and control qubits, respectively. Fig. 5.2 shows the action of the transformations from column 3 on our choice of input target qubit. The last two columns are the ideal and measured values of the Stokes parameter of the control output. We minimize the uncertainty in each measurement by accumulating a large number of counts. Computer-controlled waveplates, with angular orientation uncertainty of $\pm(2.5 \times 10^{-4})^\circ$, are used to measure in the diagonal/anti-diagonal basis. Our measurements are limited by the non ideal interferometric

σ_i	σ_j	$\tilde{\rho}_t$	$\tilde{\rho}_c$	$S_2(\mathfrak{s}[\sigma_i, \sigma_j])^{\text{theor.}}$	$S_2(\mathfrak{s}[\sigma_i, \sigma_j])^{\text{exp.}}$
σ_0	σ_0	$ 1\rangle$	D	1	0.8547 ± 0.0006
	σ_1	$ 0\rangle$	D	1	0.8718 ± 0.0005
	σ_2	$-i 0\rangle$	D	1	0.8792 ± 0.0005
	σ_3	$- 1\rangle$	D	1	0.8823 ± 0.0005
σ_1	σ_0	$ 0\rangle$	D	1	0.8459 ± 0.0006
	σ_1	$ 1\rangle$	D	1	0.8439 ± 0.0007
	σ_2	$-i 1\rangle$	A	-1	-0.8434 ± 0.0006
	σ_3	$- 0\rangle$	A	-1	-0.8540 ± 0.0007
σ_2	σ_0	$-i 0\rangle$	D	1	0.8473 ± 0.0007
	σ_1	$i 1\rangle$	A	-1	-0.8600 ± 0.0005
	σ_2	$ 1\rangle$	D	1	0.8447 ± 0.0006
	σ_3	$i 0\rangle$	A	-1	-0.8278 ± 0.0008
σ_3	σ_0	$- 1\rangle$	D	1	0.8316 ± 0.0006
	σ_1	$ 0\rangle$	A	-1	-0.8228 ± 0.0006
	σ_2	$-i 0\rangle$	A	-1	-0.8575 ± 0.0006
	σ_3	$ 1\rangle$	D	1	0.8780 ± 0.0005

Table 5.1: Data used to calculate the Holevo capacity using Eq. 5.13 for the different combinations of unitary operations σ_i and σ_j . $\tilde{\rho}_t$ and $\tilde{\rho}_c$ are the output target and control state respectively. $S_2(\mathfrak{s}[\sigma_i, \sigma_j])$ is the Stokes parameter obtained by measuring polarization of the control qubit in the diagonal/anti-diagonal basis after the unitaries σ_i and σ_j . We list both the theoretically expected values, $S_2(\mathfrak{s}[\sigma_i, \sigma_j])^{\text{theor.}}$, and the experimentally measured values, $S_2(\mathfrak{s}[\sigma_i, \sigma_j])^{\text{exp.}}$, of this Stokes parameter. Error bars (1σ) were calculated by propagation of error on the individual Stokes parameter with Poissonian counting statistics.

visibility in our switch. We calculate the average visibility from the values mentioned in Table 5.1, which is $85 \pm 2\%$. This is due to several factors. We use inverting prisms that have relatively coarse rotation accuracy $\pm 1^\circ$ to perform the unitary operations on the spatial mode. Moreover, each optical element is not perfectly flat, introducing wavefront distortions that limit the final visibility.

Our experimental setup is adapted from the quantum switch discussed in the previous chapter (Ref. in addition to the Ref. [13]), removing cylindrical lenses and using only inverting prisms, since only Pauli operations were necessary, Fig. 5.3. Our input light is diagonally polarized ($|\rho_c\rangle = (|0\rangle + |1\rangle)/\sqrt{2} = D$), and is in the first-order spatial mode ($|\rho_t\rangle = |1\rangle = |\text{HG}_{10}\rangle$).

We realize the unitary operations $\{\sigma_i\}$ using a pair of rotating prisms [189] as shown in Fig. 5.3. A mechanical rotation of the inverting prism results in a rotation of the incoming spatial mode of the photon, the outputs of the combined operation $\{\sigma_i \circ \sigma_j\}$ are shown in Fig. 5.2.

To implement σ_i, σ_j , we use up to two rotating prisms in each arm. Each prism is oriented at a physical angle θ , which reflects and rotates an incoming beam by 2θ . The action of the rotating prism on our target qubit subspace is represented by the following unitary operation:

$$R(\theta) = \begin{pmatrix} -\cos 2\theta & \sin 2\theta \\ \sin 2\theta & \cos 2\theta \end{pmatrix}. \quad (5.14)$$

We impart the global phase φ by a tilted phase plate. We write the transformation performed by the

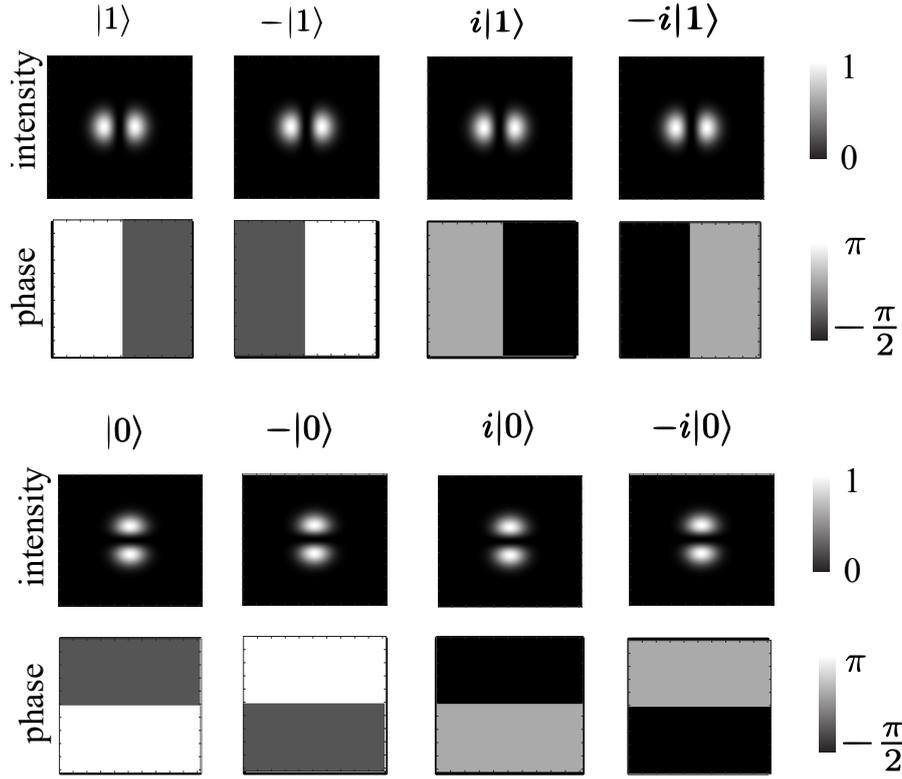


Figure 5.2: Predicted spatial mode of target qubit outputs, $\tilde{\rho}_t$, after the Pauli operations $\{\sigma_i, \sigma_j\}$ (see columns 1–3 of Table 5.1). The input target qubit ρ_t is $|1\rangle = \text{HG}_{10}$, a first-order Hermite-Gaussian mode.

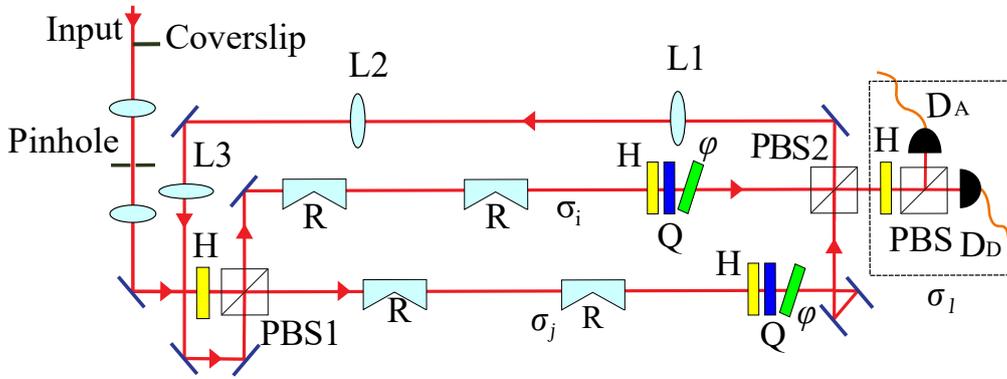


Figure 5.3: Schematic of quantum switch. R is the rotating prism 5.14, H and Q are the half and quarter waveplates respectively and φ are the phase plates. The control qubit is the polarization, $|\psi_c\rangle = D$. The polarization of the light controls order of Pauli operations $\{\sigma_i\}$ acting on the photonic spatial mode $|\psi_t\rangle = \text{HG}_{10}$, for horizontal polarization H, the order is $\sigma_i \circ \sigma_j$, for the vertical polarization V the order is $\sigma_j \circ \sigma_i$. The polarization D ensures superposition of the orders. X is a polarization measurement, determining the Stokes parameter of the measured photon in the diagonal/anti-diagonal basis. Lenses L1 to L3 form a telescope for mode-matching.

pair of prisms and the phase plate as:

$$U(\varphi, \theta_1, \theta_2) = e^{i\varphi} R(\theta_2) R(\theta_1). \quad (5.15)$$

We place the phase plate only when we are doing σ_2 operation. For operations, σ_1 and σ_3 we need

Unitary	φ	θ_1	θ_2
σ_0	0	$\frac{\pi}{2}$	$\frac{\pi}{2}$
σ_1	0	$\frac{\pi}{4}$	—
σ_2	$\frac{\pi}{2}$	$\frac{\pi}{2}$	$\frac{\pi}{4}$
σ_3	0	$\frac{\pi}{2}$	—

Table 5.2: Phases and angles for the unitary operations realized in our experiment, given by Eq. 5.15.

only one rotating prism, and for σ_0 and σ_2 we need both prisms. We achieve this by moving the second rotating prism via a translation stage. Table 5.2 are the angles and phases used to implement the Pauli operations. Since this rotation also changes the polarization of the field—the control qubit—which is not desired, this rotation is canceled after the two rotating prisms by the actions of the half- (H), and quarter- (Q) waveplates and the phase plate (ϕ), see Fig. 5.3.

We emphasize that in our architecture, the sender cannot access the individual channels—channels \mathcal{N}_p or \mathcal{N}_q —without using the other one. That is, when we consider combinations of channels where one of the channels is unitary, this unitary channel cannot be accessed without also going through the noisy channel. This is certainly not a unique realization of indefinite causal order. For example, one can use a Sagnac interferometer to achieve the same indefinite causal order as in Ref. [176].

5.3.1 Effects of experimental imperfections

We note, the Pauli operations we are implementing by the rotating prisms, can be non ideal due to uncertainty of the angles. This can affect the experiment in two ways. First, it leads to non zero S_1 and S_3 Stokes parameters in the output control qubit and second, these nonzero terms contribute to $H^{\min}(\mathfrak{s}[\mathcal{N}_p, \mathcal{N}_q])$. To account for this issue, we note that the uncertainty associated with our rotation mounts is $\pm 1^\circ$. We numerically introduce random uniformly distributed error of $\pm 1^\circ$ to the angles of the rotating prisms. We repeat the simulation for 500 iterations and measure the capacity in each run. We confirm that the contribution of the random errors are well within the range of our experimental visibilities, which is reflected in the orange shades of the graphs in Figs. 5.4, 5.5, 5.7, and 5.9.

5.4 Results

5.4.1 Two depolarizing channels

Now that we can calculate the output control qubit from Table 5.1, we can use Eq. 5.13 to estimate the Holevo capacity χ . We compare the experimental and predicted Holevo capacities for several combinations of channels. We first consider a quantum switch $\mathfrak{s}[\mathcal{N}, \mathcal{N}]$ of two depolarizing channels \mathcal{N} of identical strengths q , i.e., in Eq. 5.2 we set all coefficients to $q/4$:

$$\mathcal{N} = (1 - 3q/4)\rho_t + q/4(\sigma_1\rho_t\sigma_1 + \sigma_2\rho_t\sigma_2 + \sigma_3\rho_t\sigma_3), \quad (5.16)$$

This scenario was theoretically studied in Ref. [180]. Experimental and predicted Holevo capacities are shown in Fig. 5.4, which plots the logarithm of χ against increasing q . The red circles are our measured

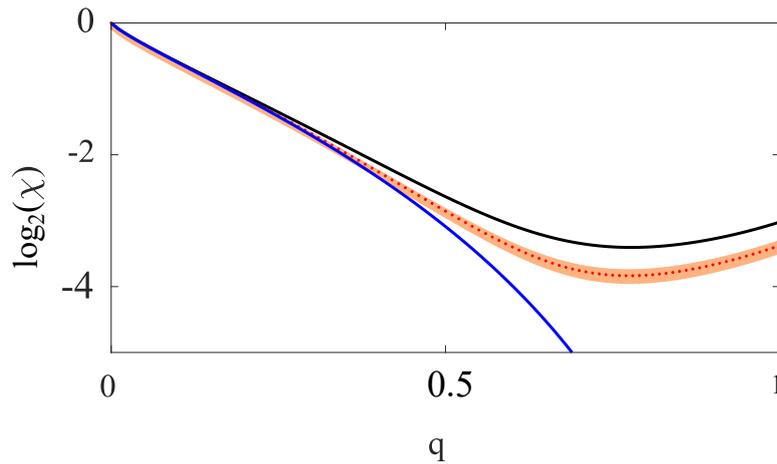


Figure 5.4: Logarithm of Holevo capacity χ , of two identical depolarizing channels, vs depolarizing channel strength, q . The blue line is the predicted Holevo capacity for the case of definite order. The black line is the predicted Holevo capacity for the case of indefinite order. The red dots are the measured values for Holevo capacity in our quantum switch; in excellent agreement with the predicted Holevo capacity for our experimental visibilities $85\pm 2\%$, shown by the orange shaded area. Note that the minimum measured Holevo capacity of $\chi = (2.1 \pm 0.2) \times 10^{-2}$ bits occurs at a depolarization $q = 0.78$, and that higher capacity of $\chi = (3.4 \pm 0.2) \times 10^{-2}$ bits occurs at full depolarization, $q = 1$.

values; the orange shaded region is the predicted Holevo capacity for visibilities of $85\pm 2\%$; the black curve is the Holevo capacity for an ideal quantum switch. The blue curve is the ideal Holevo capacity for two depolarizing channels in some definite order which—as expected—decreases monotonically with increasing depolarizing strength. In the limit of two fully depolarizing channels, $q = 1$, $\chi = 0$ bits are transmitted using definite order. In this limit there is a clear advantage in using quantum channels: ideally 4.9×10^{-2} bits can be transmitted, we measure $\chi = (3.4 \pm 0.2) \times 10^{-2}$ bits. This is a rather counterintuitive result as, individually none of the channels can transmit any information.

This nonzero Holevo capacity can be understood intuitively from the fact that the output of a quantum switch with two depolarizing channels is a statistical mixture of the output of a quantum switch with different Pauli operations $\{\sigma_i\}$. Some of these Pauli operations anti-commute, hence superpositions of the order of anti-commuting Pauli operations can preserve a finite amount of information in the target qubit.

This intuition helps us understand another striking prediction, which is that above some nonzero depolarization strength the Holevo capacity will increase. In the ideal case, we see that Holevo capacity attains a minimum value of 3.3×10^{-2} bits at $q = 0.78$ and then the capacity increases to the limit of 4.9×10^{-2} bits at $q = 1$, in stark contrast to the classical case of definite causal order which decreases monotonically to zero. Experimentally, we see this increase in information capacity from $\chi = (2.1 \pm 0.2) \times 10^{-2}$ bits at $q = 0.78$, i.e., at its worst absolute performance it is 13.5σ above the classical performance at that value of q .

We look at Eq. 5.13 to understand this behavior: for low depolarization strengths, the minimum entropy $H^{\min}(\mathfrak{s}[\mathcal{N}_p, \mathcal{N}_q])$ increases more rapidly than the von Neumann entropy $H(\tilde{\rho}_c)$. This means that for low q , the rate of depolarization of the composite system—target and control—is faster than

the rate of depolarization of the control qubit. However, at $q \approx 0.78$, $H(\tilde{\rho}_c)$ begins to increase more rapidly than $H^{\min}(\mathfrak{s}[\mathcal{N}_p, \mathcal{N}_q])$ so the depolarization rate of the control overtakes the depolarization rate of the composite system, and the information revival occurs.

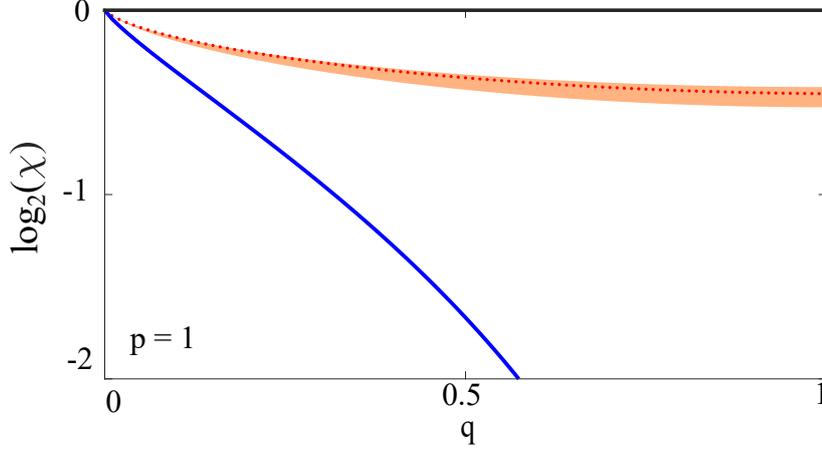


Figure 5.5: Logarithm of Holevo capacity χ , where one channel is a depolarizing channel of varying strength q , and the other is a σ_3 channel, the channel in Eq. (5.17) with $p=1$. The solid black line [$\log_2(\chi)=0$] is the theoretical predictions for Holevo capacity for indefinite causal order. The orange shade show the uncertainty due to non ideal visibility in our experiment. The red dots are the experimentally measured data points and the blue line is the theoretical predictions for Holevo capacity for definite order. We show the special case for $p=1$, i.e., the dephasing channel becomes a σ_3 channel. Note that in this situation, indefinite causal order allows us to get full 1 bit information. Experimentally we measure a Holevo capacity of 0.64 ± 0.02 bits at $q=1$.

5.4.2 Depolarizing and dephasing channel

As a second example we consider the quantum switch $\mathfrak{s}[\mathcal{N}, \mathcal{M}]$ composed of a depolarizing channel \mathcal{N} of strength q (Eq. 5.16) and a dephasing channel \mathcal{M} of strength p given by

$$\mathcal{M} = [(1-p)\rho_t + p\sigma_3\rho_t\sigma_3]. \quad (5.17)$$

When $q=1$, any definite order of these two channels results in a fully depolarizing channel, regardless of the value of p , thus $\chi=0$. However, with the additional freedom to prepare and measure the control qubit, the Holevo capacity of these two channels in indefinite order, $\chi(\mathfrak{s}[\mathcal{N}, \mathcal{M}]) \equiv \chi_i$ can be non zero and is a function of p . In fact, this combination achieves the maximum value of 1 bit when $p=1$ regardless of the depolarization strength q , as shown in Fig. 5.5 (black line). We emphasize the contrast with the zero capacity of the definitely ordered channels at $q=1$, Fig. 5.5 (blue line).

This unit Holevo capacity can be understood by noting that when $p=1$, \mathcal{M} is a unitary channel and the conditional states become:

$$\begin{aligned} \varepsilon_+(\rho_t) &= (q/4)\rho_t + (1-3q/4)\sigma_3\rho_t\sigma_3, \\ \varepsilon_-(\rho_t) &= q/4(\sigma_1\rho_t\sigma_1 + \sigma_2\rho_t\sigma_2). \end{aligned} \quad (5.18)$$

and thus from Eq. 5.13 $\chi_i = 1 + H(\tilde{\rho}_c) - H_{\rho_t}^{\min}[\varepsilon_+(\rho_t) \oplus \varepsilon_-(\rho_t)]$, where $H(\cdot)$ is again the von Neumann entropy and the minimisation is done over all possible target states. Notice that the states that

minimize the entropy of $\varepsilon_+(\rho_t) \oplus \varepsilon_-(\rho_t)$ are the eigenvectors of σ_3 . Take $\rho = |0\rangle\langle 0|$ as an example, $\varepsilon_+(\rho_t) = (1-q/2)|0\rangle\langle 0|$, while $\varepsilon_-(\rho_t)$, will be $(q/2)|1\rangle\langle 1|$ in which case, $H[\varepsilon_+(\rho) \oplus \varepsilon_-(\rho)] = H(q/2)$, and thus the Holevo-capacity of the quantum switch becomes 1 bit¹. We compare our experimental and predicted χ_i in Fig. 5.5, red dots are experimental values and the orange shaded regions accounts for non ideal visibility. For full depolarization at $q=1$, we show 0.64 ± 0.02 bits compared to the ideal case of $\chi_i=1$. Interestingly, this ideal capacity is strictly larger than the one achieved in Ref. [205] given the same channels in a path superposition. More strongly, in Sec. 5.4.5, we prove that these channels in a superposition of path cannot lead to unit capacity. We also compare predicted and experimental χ_i for other dephasing strengths p in Sec. 5.4.3.

We note that when [180] proposed communication advantage from indefinite causal order they did not claim that the advantage is unique to coherent superposition of order. Subsequently, there has been an active discussion on the origin of the communication advantage [198, 206, 207] with consensus yet to be reached. The experimental developments in the present work and [204] are further motivation to develop a well-defined resource communication theory featuring coherence.

5.4.3 More general combinations of depolarizing and dephasing channels

Previously we have introduced a combination of depolarizing channel of strength q and a dephasing channel of strength p as shown in Eq. 5.17. We have shown the special case for $p=1$ where it is possible to achieve unit Holevo capacity. In this section we show more general situations where $p \leq 1$. In Fig. 5.6, the black lines show the cases for either $p=0.5$, $p=0.8$, or $p=1$ combined with a depolarizing channel of strength q in an indefinite order. The blue line shows the capacity when these channels are in definite order, which regardless of the value of p monotonically decreases to zero when $q=1$. In contrast, with indefinite order arrangements the predicted Holevo capacities are 0.31 bits, 0.61 bits and 1 bit for $p=0.5$, $p=0.8$ and $p=1$ respectively. In Fig. 5.7 we plot the logarithm of Holevo capacities as a function of both depolarizing channel strength q for 2 different strengths of the dephasing channel, $p=0.5$, and $p=0.8$. The black curves on the graphs show the predicted Holevo capacity for the indefinite order, the orange shade is the uncertainty due to the fringe visibility and the red dots are experimental data points. In the case of $p=0$, the dephasing channel behaves as an identity channel and the input state ρ_t is equally scrambled in both definite and indefinite order cases, thus resulting no overall advantage. However, in other cases there is non zero information-theoretic advantage. At a full depolarizing strength of $q=1$: in the case of $p=0.5$, the predicted $\chi_i=0.31$ bits, whereas we experimentally show 0.179 ± 0.006 bits; at $p=0.8$, the predicted $\chi_i=0.61$ bits, we experimentally show 0.42 ± 0.01 bits.

¹This maximum advantage over definite order is also attained by having an indefinite order of one fully depolarizing channel and a unitary σ_2 or σ_3 channel, when the input states are the eigenstates of σ_2 and σ_3 , respectively.

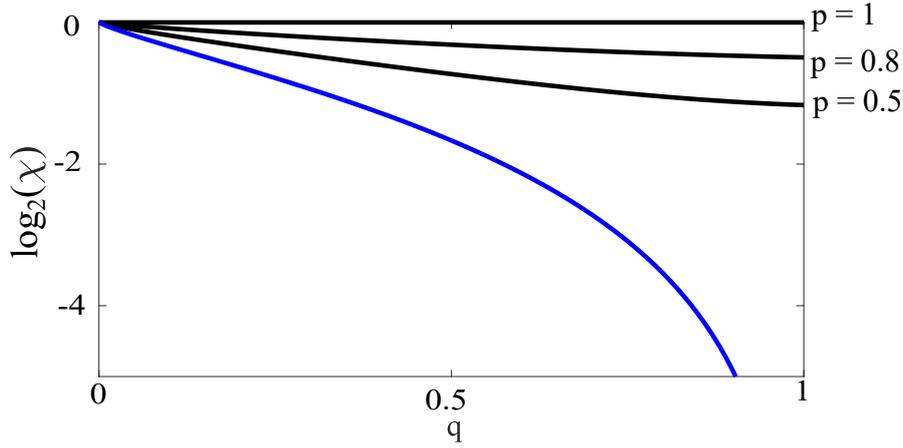


Figure 5.6: Predicted logarithm of Holevo capacity χ for the channels in Eq. 5.17 vs. depolarization noise strength, q . The solid black lines are for indefinite order in the cases of $p=0.5$, $p=0.8$, and $p=1$. The solid blue line are for definite order, regardless of the value of p , the combination of the dephasing and depolarizing channels scramble the information equally. At the full depolarization strength, $q=1$, $\chi_d=0$ whereas χ_i are 0.31 bits, 0.61 bits, and 1 bit in case of $p=0.5$, $p=0.8$, and $p=1$ respectively.

5.4.4 Bit flip and bit-phase-flip channels

Consider two noisy channels, a bit-flip and a bit-phase-flip channel with strength p given, correspondingly, by

$$\begin{aligned} N_p^1(\rho_t) &= (1-p)\rho_t + p\sigma_1\rho_t\sigma_1, \\ N_p^2(\rho_t) &= (1-p)\rho_t + p\sigma_2\rho_t\sigma_2. \end{aligned} \quad (5.19)$$

Then,

$$\begin{aligned} \varepsilon_+(\rho_t) &= (1-p)^2\rho_t + p(1-p)(\sigma_1\rho_t\sigma_1 + \sigma_2\rho_t\sigma_2) \\ \varepsilon_-(\rho_t) &= p^2\sigma_3\rho_t\sigma_3. \end{aligned} \quad (5.20)$$

It should be highlighted that the outcome corresponding to the $|-\rangle\langle-|$, namely the $\varepsilon_-(\rho_t)/p^2$ channel, is actually a purely unitary channel, capable of achieving perfect transmission despite both channels being noisy (as was also pointed out recently in Ref. [201]). However, here we are interested in the Holevo capacity resulting from considering both possible outcomes weighted by their corresponding probabilities, rather than the post-selected $\varepsilon_-(\rho_t)$. We solve the corresponding minimization problem to compute the Holevo capacity. Fig. 5.8 plots the Holevo capacity χ against different noise parameters p for both the cases of noisy channels in a definite order, and in an indefinite order. The black curve is the Holevo capacity for indefinite order, while the blue curve is for definite order. As shown by the difference of these two curves, indefinite order provides an advantage over the definite-order case. Indefinite order allows transmission of 0.27 bits or more over the whole range of p values. The minimum capacity of 0.27 bits which occurs at 0.37, is higher than the capacity of the definite-order case from $p=0.2$ to $p=0.8$, 60% of the range of p . The maximum advantage $\max[(\chi_i - \chi_d)] = 0.55$ bits occurs at $p=0.75$, where the Holevo capacity for the quantum switch and definite order cases are 0.75 and 0.19 bits, respectively. We can see there is a knee in the χ_i at noise parameter $p=0.5$. This is because when $p \leq 0.5$, the eigenvectors of σ_1 or σ_2 achieve the minimum entropy $H^{\min}(\mathfrak{s}[N_p^1, N_p^2])$, while for $p \geq 0.5$, the corresponding optimum state turns out to be the eigenvectors of σ_3 . When $p=0.5$,

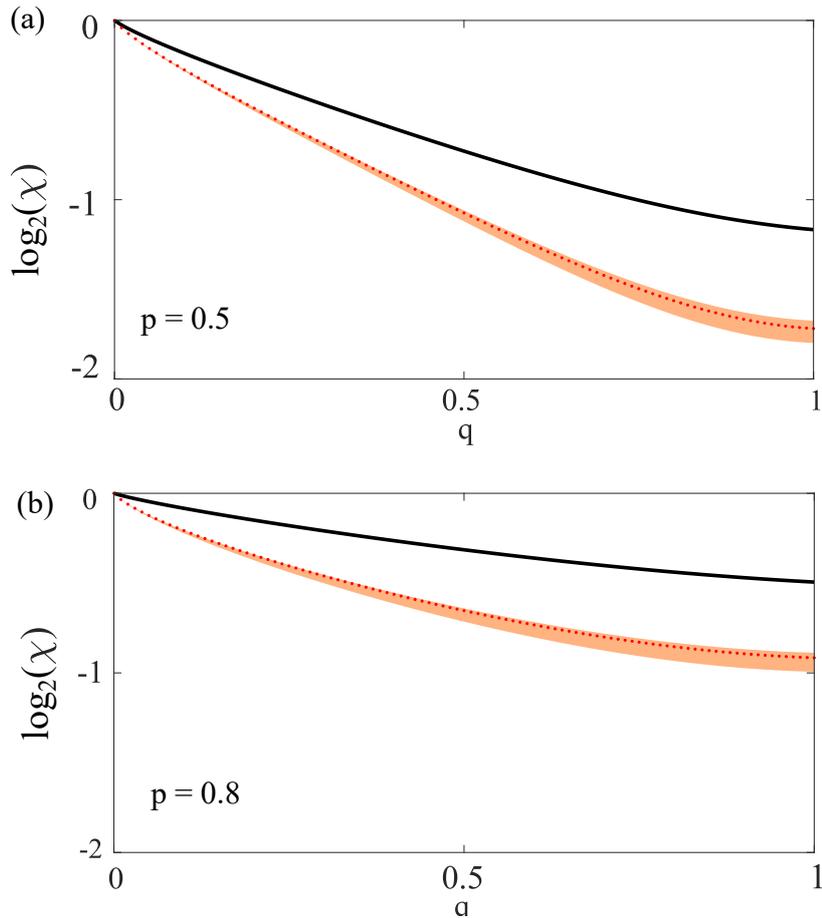


Figure 5.7: Logarithm of Holevo capacity χ , where one channel is a depolarizing channel of varying strength q , and the other one is a dephasing channel of strength p . The solid black lines are the theoretical predictions for Holevo capacity for indefinite causal order. The orange shades show the uncertainty due to non ideal visibility in our experiment. The red dots are the experimentally measured data points. (a) shows the case when $p=0.5$. In this case the ideal value is $\chi_i=0.31$ bits whereas we experimentally measure 0.179 ± 0.006 bits. (b) is the case when $p=0.8$. Here, the χ_i is 0.61 bits and we measure 0.42 ± 0.01 bits.

where the classical order results in full depolarization, the capacity resulting from indefinite order is $\chi_i \approx 0.31$ bits. In this case, any of the eigenstates of the Pauli operator achieves the maximum capacity thus increasing the domain over which we can optimally encode the input.

5.4.5 Holevo capacity from path superposition

Although the advantage in communication is not unique to superposition of order, there is strong reason to believe that the perfect transmission discussed above is not possible via a superposition of paths, where we place two noisy channels in two arms of an interferometer [205]. The optimized Holevo capacity that we obtain for a fully depolarizing channel and a unitary channel in path superposition is 0.75 bits. Moreover, we prove that reaching a value of 1 bit is impossible:

In our communication task, we consider the target system, the control system and two Pauli channels to transfer the target qubit. The control enables the target to go through a superposition of paths. In our

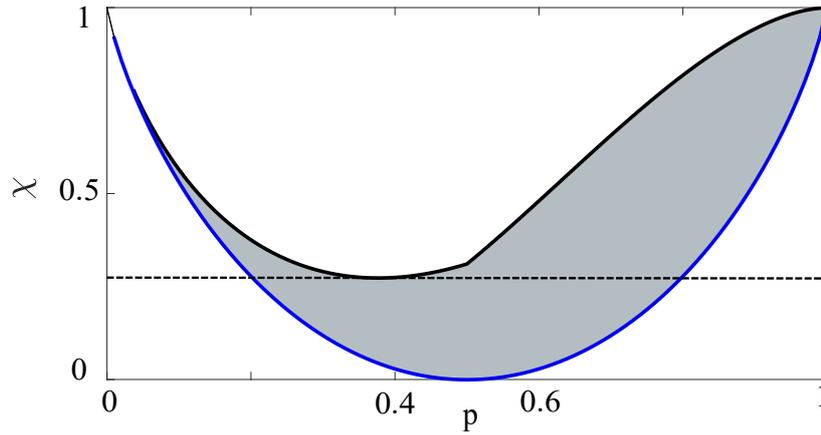


Figure 5.8: Predicted Holevo capacity χ for the channels in Eq. 5.19 vs noise parameter, p . The solid black line is the Holevo capacity of depolarizing channels in indefinite order and the solid blue line is for the case of definite order. We observe a knee at $p=0.5$, this is because the optimum state for $p \leq 0.5$ are the eigenvectors of the operator σ_1 and σ_2 . On the other hand, for $p \geq 0.5$ the optimum state is the eigenvector of σ_3 . At $p=0.5$, the definite-ordered channel becomes completely depolarizing, making $\chi_d = 0$, whereas at this point $\chi_i = 0.31$ bits. In the region of $p \geq 0.5$, the maximum difference between the two cases is 0.55 bits, occurring when $p=0.75$. The black horizontal dashed line denotes the minimum capacity of the indefinite-ordered case, 0.27 bits when $p=0.37$.

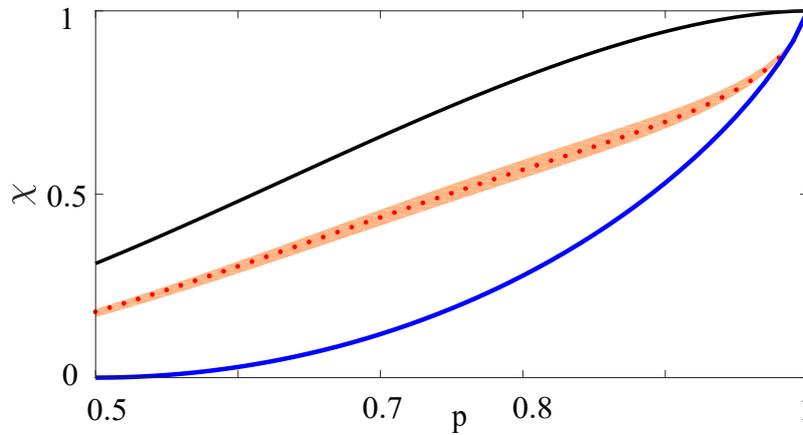


Figure 5.9: Experimental Holevo capacity χ for $p \geq 0.5$ region. where one channel is a bit-flip channel and the other one is a dephasing channel with varying strength p as in the Eq. 5.19. The solid black line is the theoretical predictions for Holevo capacity for indefinite causal order. The blue line is the theoretical prediction for Holevo capacity for definite causal order. The red dots are the experimental data points and the orange shade is the expected range due to non ideal visibility. At $p=0.5$, the experimentally measured Holevo capacity is 0.179 ± 0.006 bits, whereas $\chi_d=0$.

experiment, $|\psi_c\rangle$ is set to be $|+\rangle$, the target state is optimized to achieve the maximum communication capacity, and a generalized measurement on the control system is applied after noisy channels.

We consider our Pauli channels to be the depolarizing channel, \mathcal{N} , as shown in Eq. 5.16, and σ_3 . Following [205], after tracing out the environment, the output control and target state takes the

following form:

$$\begin{aligned} \rho_{tot}^{(ct)} = & \frac{1}{2} \left(|0\rangle\langle 0|^c \otimes \mathcal{N}(\rho_t) + |1\rangle\langle 1|^c \otimes \sigma_3 \rho_t \sigma_3 \right. \\ & \left. + |0\rangle\langle 1|^c \otimes T_0 \rho_t \sigma_3 + |1\rangle\langle 0|^c \otimes \sigma_3 \rho_t T_0^\dagger \right) \end{aligned} \quad (5.21)$$

where $T_0 = \sum_i e_i K_i$, K_i are arbitrary choices for Kraus operators for the depolarizing channel with the constraint $\sum_i |e_i|^2 = 1$. Exploiting the unitary freedom in the operator-sum representation (see for example Ref. [22]), one has that $e_i K_i = e_i \sum_j U_{i,j} \frac{\sigma_j}{2}$, with $U_{i,j}$ the entries of a unitary matrix, and thus $T_0 = \sum_j (\sum_i e_i U_{i,j}) \frac{\sigma_j}{2} \equiv \sum_j f_j \sigma_j / 2$, with $\sum_j |f_j|^2 = 1$. Now if we measure the control qubit in the $\{M_+, M_-\}$ basis ,

$$\begin{aligned} M_+ &= \cos \alpha |0\rangle + \sin \alpha |1\rangle, \\ M_- &= -\sin \alpha |0\rangle + \cos \alpha |1\rangle, \end{aligned} \quad (5.22)$$

the two possible output states, with corresponding probabilities p_+ and p_- , are given by

$$\begin{aligned} p_+ \rho_+^t &= \frac{1}{4} \left[\cos^2 \alpha \frac{\sigma_0}{2} + \sin^2 \alpha \sigma_3 \rho_t \sigma_3 \right. \\ & \quad \left. + \frac{\cos \alpha \sin \alpha}{2} \sum_i (f_i \sigma_i \rho_t \sigma_3 + f_i^* \sigma_3 \rho_t \sigma_i) \right], \\ p_- \rho_-^t &= \frac{1}{4} \left[\sin^2 \alpha \frac{\sigma_0}{2} + \cos^2 \alpha \sigma_3 \rho_t \sigma_3 \right. \\ & \quad \left. - \frac{\cos \alpha \sin \alpha}{2} \sum_i (f_i \sigma_i \rho_t \sigma_3 + f_i^* \sigma_3 \rho_t \sigma_i) \right]. \end{aligned} \quad (5.23)$$

where f_i^* is the complex conjugate of f_i .

Now, achieving a unit Holevo capacity implies that both the input and output state must be pure. In the Bloch representation, this implies that

$$\sum_i |r_i^{(+)}|^2 = 1, \quad (5.24)$$

$$\sum_i |r_i^{(-)}|^2 = 1, \quad (5.25)$$

$$\sum_i |r_i^{(in)}|^2 = 1 \quad (5.26)$$

where $\{r_i^{(+)}\}$, $\{r_i^{(-)}\}$ and $\{r_i^{(in)}\}$ are the Bloch vector components of the projected states ρ_+^t , ρ_-^t and the input target state, ρ_t respectively. We proceed by writing f_i into its real and imaginary parts, i.e., $f_j = f_{j,R} + i f_{j,I}$. Then, from $\sum |r_i^{(+)}|^2 = \sum |r_i^{(-)}|^2$ we find that

$$\sin 2\alpha (r_3^{(in)} f_{0,R} + f_{3,R} + r_1^{(in)} f_{2,I} - r_2^{(in)} f_{1,I}) + \cos 2\alpha = 0 \quad (5.27)$$

which can be replaced into $\sum |r_i^{(+)}|^2 + \sum |r_i^{(-)}|^2 = 2$ to obtain

$$\begin{aligned} \sin^2 2\alpha [& (r_2^{(in)} f_{0,I} - r_3^{(in)} f_{1,R} + r_1^{(in)} f_{3,R} + f_{2,I})^2 + (r_1^{(in)} f_{0,I} + r_3^{(in)} f_{2,R} + f_{1,I} - r_2^{(in)} f_{3,R})^2 \\ & + (f_{0,R} + r_1^{(in)} f_{1,R} + r_2^{(in)} f_{2,R} + r_3^{(in)} f_{3,R})^2] + \cos 4\alpha = 2 \end{aligned} \quad (5.28)$$

Next we proceed to show that, Eq. 5.28 cannot hold under conditions 5.26 and $\sum_i |f_i|^2 = \sum_i \sum_{k \in \{R, I\}} f_{i,k}^2 = 1$. Note that for this purpose, it is sufficient to prove that

$$\begin{aligned} F(r, f) &= (r_2^{(in)} f_{0,I} - r_3^{(in)} f_{1,R} + r_1^{(in)} f_{3,R} + f_{2,I})^2 \\ &\quad + (r_1^{(in)} f_{0,I} + r_3^{(in)} f_{2,R} + f_{1,I} - r_2^{(in)} f_{3,R})^2 \\ &\quad + (f_{0,R} + r_1^{(in)} f_{1,R} + r_2^{(in)} f_{2,R} + r_3^{(in)} f_{3,R})^2 < 3. \end{aligned} \quad (5.29)$$

Exploiting the relation that $2xy \leq x^2 + y^2$ for any real x and y , Eq. 5.26, and that for that for $|r_i| \leq 1$ one has that $r_i \leq |r_i|$ and $|r_i|^2 \leq |r_i|$, we have

$$\begin{aligned} F(r, f) &\leq (1 + r_1^{(in)} + r_2^{(in)} + |r_3^{(in)}|^2) f_{0,I}^2 + (1 + r_1^{(in)} + r_2^{(in)} + r_3^{(in)}) (f_{0,R}^2 + f_{1,I}^2 + f_{2,I}^2 + f_{3,R}^2) \\ &\quad + (1 + r_1^{(in)} + |r_2^{(in)}|^2 + r_3^{(in)}) f_{1,R}^2 + (1 + |r_1^{(in)}|^2 + r_2^{(in)} + r_3^{(in)}) f_{2,R}^2 \\ &\leq (1 + |r_1^{(in)}| + |r_2^{(in)}| + |r_3^{(in)}|) \sum_i |f_i|^2 \\ &\leq 1 + \sqrt{3 \sum_i |r_i^{(in)}|^2} \\ &\leq 1 + \sqrt{3} < 3, \end{aligned} \quad (5.30)$$

as desired. This supports the idea that \mathcal{N} and a unitary channel in a superposition of paths cannot lead to a pure output and thus cannot achieve unit Holevo capacity.

5.5 Conclusion

Quantum mechanics allows indefinite causal order, which allows us to communicate up to 1 (0.049) bits of information through one (two) fully depolarizing channels. This is possibly useful for communication through turbid media: there have been proposals for quantum communication protocols using a Sagnac interferometer (see Ref. [208] for a review) reminiscent of the quantum switch implementation of Wei *et al.* [176]. We can imagine a situation where one arm of the Sagnac loop consists of a unitary channel in the laboratory and the other arm goes through the atmosphere, e.g., ground-satellite-ground path. Placing these channels in indefinite causal order will enable to transmission of qudits in the transverse spatial mode even though the atmosphere is naturally noisy for these qudits. We can also use this idea for secret sharing, where a specific combination of channels, when connected indefinitely, can transfer information between two parties, but in a scenario of eavesdropping, any intervention will break the ‘‘indefiniteness,’’ and the message will remain scrambled. We can also think of imaging through scattering media where the information being communicated is in the modulation of the spatial mode, e.g., imaging through skin. In all these cases, placing the noisy channel in indefinite causal order with a clean channel will provide information that would normally be entirely lost.

5.6 Author contribution

Contributor	Statement of contribution	%
K. Goswami	writing of text	90
	proof-reading	10
	theoretical derivations	30
	numerical calculations	50
	performing the experiment	90
	preparation of figures	100
	Initial concept	80
	Responding to the reviewers	80
Y. Cao	writing of text	5
	numerical calculations	50
	theoretical derivations	35
G. A. Paz-Silva	writing of text	5
	theoretical derivations	35
J. Romero	proof-reading	45
	supervision, guidance	50
	Performing the experiment	10
	Initial concept	10
	Responding to the reviewers	10
A. G. White	proof-reading	45
	supervision, guidance	50
	Initial concept	10
	Responding to the reviewers	10

Chapter 6

Classical communication through quantum causal structures

The following publication has been incorporated as Chapter 6.

1. [15] **K. Goswami**, F. Costa, [Classical communication through quantum causal structures](#) *Phys. Rev. A*, **103**, 042606 (2021).

See Section 6.8 for a breakdown of the author contribution.

6.1 Introduction

One of the key questions in quantum information is the rate at which a quantum channel can transmit classical information, as quantified by the *classical capacity* of the channel [52, 209]. Holevo's seminal result [59], and following work [68, 210], provide upper bounds on the classical capacity, showing that each qubit can communicate at most one bit of classical information.

In a typical quantum communication protocol, the parties act in a fixed order. However, more general situations are possible, where causal order might be uncertain or even not defined. A practical example can be a distributed system, such as the internet, where different nodes communicate with each other. In such systems, local clocks can suffer from random errors and delays, leading to uncertainty in the ordering of the events [211]. Even more radically, recent developments have shown possibilities of *indefinite* causal structure, i.e., scenarios where the lack of order between the parties cannot be reduced to classical ignorance [5, 7]. From a foundational point of view, this is relevant, for example, in quantum gravity scenarios, where quantum superposition of spacetimes can result in an indefinite causal order of events [171, 172, 212]. Pragmatically, quantum control of causal order has been proposed as a resource for computation and communication [5, 173–175, 201, 202, 213–216], with several experimental implementations already performed [13, 14, 176, 177, 184, 217, 218].

In light of the foundational and applied relevance, it is important to understand how general quantum causal structure affects classical communication. In particular, given the possibility to violate causal inequalities [7, 203, 219, 220] and the advantage in certain communication tasks [174, 175, 214],

one may wonder whether an indefinite causal structure can augment the classical communication capacity and possibly exceed the Holevo bound [68, 210]. However, despite much work on various communication protocols, the notion of classical capacities in situations where the communicating parties themselves are indefinitely ordered has not yet been developed. One of the challenges is the fact that the definition of capacity involves asymptotically many independent uses of the communication resource but for indefinite causal order, the notion of repeated use turns out to be problematic [206, 221].

We address this gap through the process matrix formalism [7]. We introduce an asymptotic setting appropriate to a communication scenario without pre-defined causal order. This allows us to develop expressions for the asymptotic capacity of a process, under different encoding and decoding settings, reducing to analogue expressions for quantum channels. We find that, in each case, the classical communication from a sender to a receiver cannot exceed what is achievable in a definite causal order. In particular, the Holevo bound extends to general processes: the classical capacity cannot exceed one bit per received qubit. This despite the fact that, unlike in conventional channels, the receiver has access to an output system that can signal back to the sender, which constitutes an extra communication resource (for example, for the violation of causal inequalities). We also explore two-party communication protocols when causal order is definite but unknown (probabilistic). In such situations, the total bi-directional communication cannot exceed the maximum one-way communication—again, at most one bit per qubit in either direction. This extends to a similar bound for communication between multiple parties in a definite (but possibly probabilistic) causal order.

We present the work following way. In Section 6.2, we give an introduction to classical communication through quantum channels. In Section 6.3, we introduce the framework of the process matrix. In Section 6.4, we give a brief motivation behind our communication settings. In Section 6.5, we introduce asymptotic setting for processes, subsequently we define classical capacities of a process, and developed a bound for one way communication. In Section 6.6 we develop a bound for bi-directional communication protocol. We then generalise the bound for a multi-party broadcast communication protocol.

6.2 Classical communication through a quantum channel

Let us first review how one can use ordinary quantum channels to send classical information [52, 222]. In a one-way communication protocol, Alice has a classical message m , prepared according to some probability distribution $P(m)$, and encodes it into a quantum state ρ_m . She then sends it to Bob through a noisy quantum channel \mathcal{N} . Upon receiving the state, Bob extracts the message by using a positive operator valued measure (POVM) $\{E_{m'}\}_{m'}$, where $E_{m'} \geq 0$, $\sum_{m'} E_{m'} = \mathbb{1}$. Here m' denotes the measurement outcome. The conditional probability of Bob receiving a message m' given that Alice sends the message m is

$$p(m'|m) = \text{Tr}[E_{m'}\mathcal{N}(\rho_m)]. \quad (6.1)$$

The probability of error for a particular message m is

$$p_e(m) = p(m' \neq m | m) = 1 - p(m | m). \quad (6.2)$$

The goal of the protocol is to minimise the maximal probability of error $p_e^* := \max_m p_e(m)$. An *asymptotic setting* is a scenario where Alice can use n copies of the channel to send a k -bit message $m \in \{0, 1\}^k$, where both k and n can be arbitrarily large. In other words, she encodes k bits into an $n \geq k$ -bit message $X_m^{(n)} \in \{0, 1\}^n$ and subsequently an n -qubit state $\rho_m^{(n)}$ and then sends each qubit through an independent copy of the channel. The classical capacity of the quantum channel \mathcal{N} , is defined as the maximal rate $C = k/n$ such that asymptotically, $n \rightarrow \infty$, one can achieve noiseless communication, $p_e^* \rightarrow 0$ [39, 52].

Different encoding and decoding strategies can lead to different asymptotic settings resulting in different classical communication capacities for a channel \mathcal{N} , which we review below. A quantification of classical communication possible through a channel \mathcal{N} is given by the *Holevo quantity* [59], defined as

$$\begin{aligned} \chi(\mathcal{N}) := \max_{p(m), \rho_m} & S\left(\sum p(m) \mathcal{N}(\rho_m)\right) \\ & - \sum p(m) S(\mathcal{N}(\rho_m)). \end{aligned} \quad (6.3)$$

Here $S(\cdot)$ is the von Neumann entropy. Having introduced the Holevo quantity for a channel, it is interesting to see how this quantity is related to different classical capacities corresponding to different asymptotic configurations of channels. We discuss it below.

Product encoding - Product decoding: When the input quantum state is a product state of the form $\rho_m^{(n)} = \otimes_{i=1}^n \rho_i$ and the measurement operation is $E_{m'}^{(n)} = \otimes_{i=1}^n E_i$ with each E_i acting on the qubit $\mathcal{N}(\rho_i)$. Let us consider, the measurement result produces an n -bit string $Y_{m'}^{(n)} \in \{0, 1\}^n$ corresponding to the message m' . In the asymptotic setting, $n \rightarrow \infty$, the capacity in this setup is given by the conventional definition of classical capacity obtained by maximising the regularised mutual information, $I(Y_{m'}^{(n)} : X_m^{(n)})/n$, between Alice's input and Bob's output over the input probability distribution, the encoded quantum states and decoding measurement operators. The central idea of Shannon's capacity formula is that the mutual information $I(Y_{m'}^{(n)} : X_m^{(n)})$ is additive. Thus, the corresponding capacity is called *product capacity*, $C^{(1)}(\mathcal{N})$. This capacity is determined by the single use of the channel \mathcal{N} , with the optimised mutual information $I(Y_{m'}^{(1)} : X_m^{(1)})$, corresponding to the single-copy input and output variables $X_m^{(1)}$ and $Y_{m'}^{(1)}$ respectively, i.e.

$$\begin{aligned} C^{(1)}(\mathcal{N}) &= \lim_{n \rightarrow \infty} \max_{p(m), \rho_m^{(n)}, E_{m'}^{(n)}} \frac{I(Y_{m'}^{(n)} : X_m^{(n)})}{n} \\ &= \max_{p(m), \rho_m^{(1)}, E_{m'}^{(1)}} I(Y_{m'}^{(1)} : X_m^{(1)}). \end{aligned} \quad (6.4)$$

Holevo's theorem [59] states that this quantity is upper bounded by

$$C^{(1)}(\mathcal{N}) \leq \chi(\mathcal{N}), \quad (6.5)$$

with the $\chi(\mathcal{N})$ defined in Eq. (6.3). Hereinafter, for the sake of clarity, we are going to represent $I(Y_{m'}^{(1)}:X_m^{(1)})$ with $I(m':m)$.

Product encoding - Joint decoding: The difference with the previous case is that Bob can perform a joint measurement on the n -qubit system. The *Holevo-Schumacher-Westmoreland (HSW) capacity*, $C(\mathcal{N})$, associated with this strategy is simply equal to the Holevo quantity [68, 210]:

$$C(\mathcal{N}) = \chi(\mathcal{N}), \quad (6.6)$$

It is worth mentioning that the HSW capacity for some channels can be strictly greater than one shot capacity, i.e. $C(\mathcal{N}) > C^{(1)}(\mathcal{N})$, this is also known as *super-additivity*.

Joint encoding–Joint decoding: Here, Alice uses an entangled n -qubit state to encode the information and Bob performs a joint measurement on his output. The capacity associated with this strategy is given by *regularised Holevo quantity* [59, 68, 210]:

$$C^\infty(\mathcal{N}) = \chi_{\text{reg}}(\mathcal{N}), \quad (6.7)$$

with

$$\chi_{\text{reg}}(\mathcal{N}) = \lim_{n \rightarrow \infty} \frac{\chi(\mathcal{N}^{\otimes n})}{n}. \quad (6.8)$$

It has been shown in [72] that the capacity $C^\infty(\mathcal{N})$ can be strictly greater than the HSW capacity — $C^\infty(\mathcal{N}) > C(\mathcal{N})$. Thus in general, we can write $C^\infty(\mathcal{N}) \geq C(\mathcal{N}) \geq C^{(1)}(\mathcal{N})$.

The Holevo quantity $\chi(\mathcal{N})$, and consequently the regularised Holevo quantity $\chi_{\text{reg}}(\mathcal{N})$, are further upper bounded by $\log(d)$, where d is the output dimension of the channel \mathcal{N} . With this we summarise a sequence of inequalities:

$$\begin{aligned} I(m':m) &\leq C^{(1)}(\mathcal{N}) \leq C(\mathcal{N}) = \chi(\mathcal{N}) \\ &\leq C^\infty(\mathcal{N}) = \chi_{\text{reg}}(\mathcal{N}) \leq \log(d), \end{aligned} \quad (6.9)$$

with $I(m':m)$ being the unoptimized mutual information between Alice's input m and Bob's output m' . A consequence of this chain of inequalities is that, for any communication setting, a d dimensional quantum channel cannot transfer more than $\log(d)$ bits. In other words, quantum systems can carry at most one bit per qubit.

6.3 The Process framework

In conventional quantum communication protocols, the communicating parties act in a well-defined order. However, quantum mechanics allows possibilities, where the order between the communicating parties is unknown or even indefinite [5, 171]. This possibility can be modelled within the so-called

process matrix formalism [7, 154, 181], which we have described in Chapter 2 of this thesis. To maintain the flow, I will briefly recapitulate the process formalism here. Consider a situation involving two parties — Alice and Bob, each acting in a local laboratory. In each run of the experiment, each of them receives a quantum system in their respective laboratories, performs some operation on it and sends it out [223]; Alice’s (Bob’s) input and output systems will be denoted by A_I (B_I) and A_O (B_O), respectively. Most generally, each party’s operation consists in letting the system of interest interact with an additional “probe” system. The parties can use their probes to input information (by preparing them in arbitrary states) and to extract information, through arbitrary measurements. The most general operation is, therefore, a completely positive (CP) map $\mathcal{M} : X_I \otimes X'_I \rightarrow X_O \otimes X'_O$, where, for $X = A, B$; X'_I, X'_O denote the additional system and we use the system’s label to represent the corresponding state space.

It is convenient to represent CP maps as positive semidefinite matrices, $M \in X_I \otimes X'_I \otimes X_O \otimes X'_O$, using the Choi isomorphism [32]:

$$M^{X_I X'_I X_O X'_O} = \sum_{i,j=1}^{d_{X_I} d_{X'_I}} |i\rangle\langle j|^{X_I X'_I} \otimes \mathcal{M}(|i\rangle\langle j|^{X_I X'_I}). \quad (6.10)$$

Here, the set $\{|i\rangle\}$ represents an orthonormal basis in $X_I \otimes X'_I$ and d_X represents the dimension of X . If the map \mathcal{M} is completely positive and trace preserving (CPTP), then the Choi representation gives an additional constraint

$$\text{Tr}_{X_O X'_O} M^{X_I X'_I X_O X'_O} = \mathbb{1}_{X_I X'_I}. \quad (6.11)$$

A CPTP map (also called *channel*) represents an operation that can be performed with probability one, while a CP, trace non-increasing map generally is the conditional transformation corresponding to a particular outcome of a measurement.

The resource connecting the two communicating parties is described by the *process matrix* $W^{A_I A_O B_I B_O}$. This encodes the background process that governs how the systems on which the parties act relate to each other, be it a shared state, a channel from one to the other, or more general scenarios. We can view the process as a higher order map, transforming input quantum maps (the local operations) to output quantum maps (a channel acting on the probe system). The process matrix W has to satisfy a set of constraints:

$$W \geq 0, \quad (6.12)$$

$$\text{Tr} W = d_{A_O} d_{B_O}, \quad (6.13)$$

$${}_{B_I B_O} W = {}_{A_O B_I B_O} W, \quad (6.14)$$

$${}_{A_I A_O} W = {}_{B_O A_I A_O} W, \quad (6.15)$$

$$W = {}_{A_O} W + {}_{B_O} W - {}_{A_O B_O} W. \quad (6.16)$$

Here, ${}_x W := \mathbb{1}^x / d_x \otimes \text{Tr}_x W$ is the ‘trace-and-replace’ notation [181], which discards subsystem x and replaces it with the normalised identity. Here appropriate reordering of the tensor factors is implied.

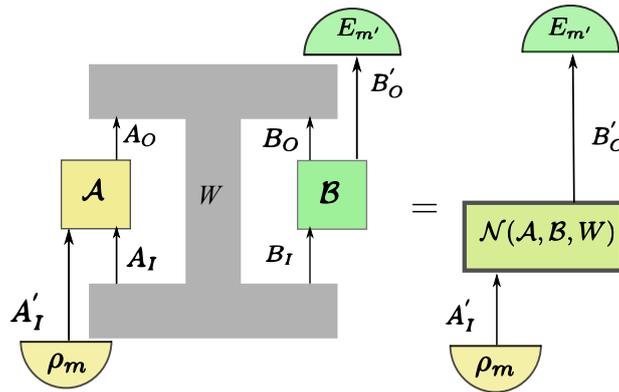


Figure 6.1: A process $W^{A_I A_O B_I B_O}$ with two CPTP maps $A^{A'_I A_I A_O}$ and $B^{B_I B_O B'_O}$ forms a new channel $\mathcal{N}(\mathcal{A}, \mathcal{B}, W)$, as in Eq. (5.2), with input system A'_I and the output system B'_O . Alice can use this channel to communicate to Bob by encoding the quantum state ρ_m at her input system and Bob performing a POVM measurement $E_{m'}$ at his output system.

To compose two CP maps, as well as a CP map and the process, we use the link product, denoted by $*$ [6]. For two positive semidefinite operators P and Q with the respective Hilbert spaces \mathcal{P} and \mathcal{Q} , the composition $P * Q$ is given as

$$P * Q := \text{Tr}_{\mathcal{P} \cap \mathcal{Q}}[(\mathbb{1}^{\mathcal{P} \setminus \mathcal{Q}} \otimes P^{T_{\mathcal{P} \cap \mathcal{Q}}})(Q \otimes \mathbb{1}^{\mathcal{Q} \setminus \mathcal{P}})]. \quad (6.17)$$

Here, \mathcal{P} and \mathcal{Q} are the Hilbert spaces associated with P and Q , the superscript $T_{\mathcal{P} \cap \mathcal{Q}}$ represents partial transpose on the shared Hilbert spaces. Note if the operators do not share Hilbert spaces, $\mathcal{P} \cap \mathcal{Q} = \emptyset$, then the link product reduces to the tensor product, $P * Q = P \otimes Q$. In contrast, when the operators do not have any non-overlapping spaces, $\mathcal{P} \setminus \mathcal{Q} \cup \mathcal{Q} \setminus \mathcal{P} = \emptyset$, the link product yields a scalar number: $P * Q = \text{Tr}(P^T Q)$.

The link product generalises the state transformation, composition of channels, and state-measurement; a state ρ transformed by a channel \mathcal{N} is given by $\mathcal{N}(\rho) = \rho * N$, where N is the Choi representation of \mathcal{N} . A sequential composition $\mathcal{M}_1 \circ \mathcal{M}_2 : X_1 \rightarrow X_3$, with the Choi representation $M_3^{X_1 X_3}$ is given by $M_3^{X_1 X_3} = M_1^{X_1 X_2} * M_2^{X_2 X_3}$. Here we assume the constituent channels to be $\mathcal{M}_1 : X_1 \rightarrow X_2$ and $\mathcal{M}_2 : X_2 \rightarrow X_3$ with the Choi representations M_1 and M_2 respectively. Finally, we can rewrite the state-measurement in Eq. (6.1) as $p(m' | m) = E_{m'}^T * \mathcal{N}(\rho_m) = E_{m'}^T * \rho_m * N$. Note that the Choi representation of a POVM-element $E_{m'}$ is its transpose $E_{m'}^T$.

The link product also captures the action of the process on the local operations. In our case, the action of W on Alice's and Bob's local operations \mathcal{M}^A , and \mathcal{M}^B is given by $(M^A * M^B) * W$. Here M^A (M^B) is the Choi representation of the map \mathcal{M}^A (\mathcal{M}^B).

6.4 Motivation

Before exploring communication protocols through the higher order process, it is important to distinguish between the communication setup introduced in Chapter 5 and the one we are using in the current chapter. We clarify the distinction in Fig. 6.2, where Fig. 6.2(a) demonstrates a scenario

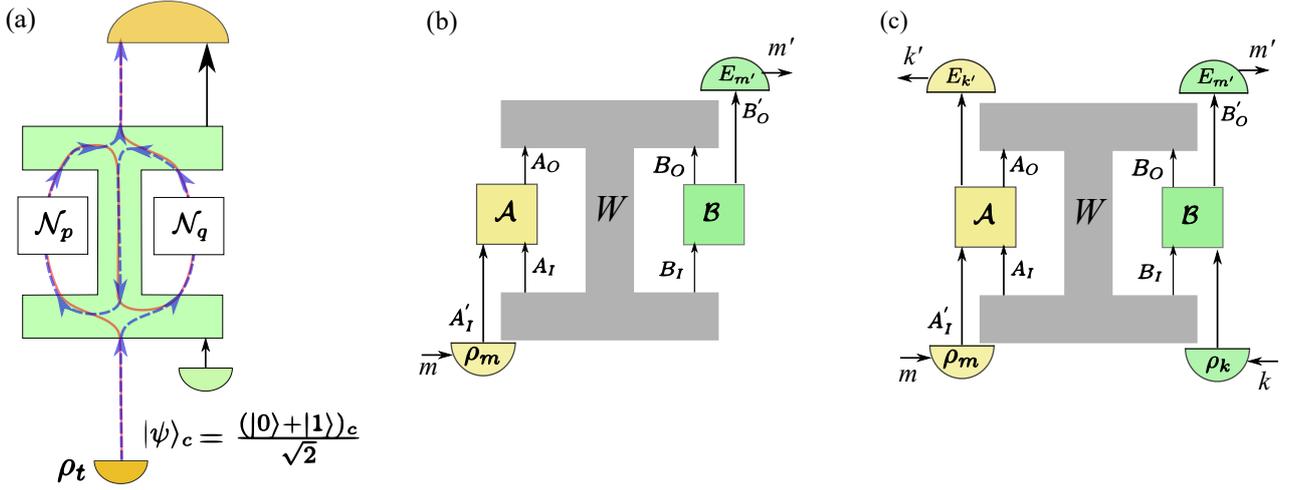


Figure 6.2: Different communication setups. Fig.(a) represents a communication through the quantum switch as discussed in Chapter 5. In this chapter, I take a different approach: Fig.(b); and (c). Here the communicating parties are in an indefinite order. I consider a bi-partite process W which lacks, in general, a fixed causal order. Fig.(b) is a one-way communication protocol (Sec.6.5). In a one-way communication, the sender encodes a classical message m in a quantum state ρ_m , and then through a CPTP map \mathcal{A} , sends it the process W . The receiver receives the quantum system through a CPTP map \mathcal{B} , and applies a POVM-element $E_{m'}$ to extract the message m' . Fig.(c) is the bi-directional communication protocol as described in Sec.6.6. The messages m and k are sent and messages m' and k' are received. In this case, the operations \mathcal{A} and \mathcal{B} are CP maps.

of communication through the quantum switch, with two noisy channels in indefinite causal order. However, In case of the quantum switch, the communicating parties have a fixed causal order, i.e. the sender causally precedes the receiver. In this chapter, we are not restricting ourselves to the quantum switch by considering the most general higher order process. More importantly, we allow the communicating parties to be in an indefinite causal order, see Fig. 6.2(b) for one-way communication and Fig. 6.2(c) for a bi-directional communication. The motivation behind this setup comes from the possibility to violate *causal inequalities* [7, 203, 219, 220] described as follows:

6.4.1 Causal inequality

Let us assume a scenario where two parties, sharing an indefinitely causal ordered process, are involved in a bi-directional communication [7]. We show this in Fig. 6.2(c) where the messages m and k are received and m', k' are sent. Additionally, Bob possess a classical bit b . If $b=0$, Alice has to guess Bob's message, and if $b=1$, Bob has to guess Alice's message. The goal of the game is to maximise the probability of success:

$$p_{\text{succ}} := \frac{1}{2} [p(k'=k|b=0) + p(m'=m|b=1)]. \quad (6.18)$$

In Ref. [7], it has been shown that when the communicating parties are causally ordered, the probability of success is bounded by

$$p_{\text{succ}} \leq \frac{3}{4}. \quad (6.19)$$

We refer this bound as *causal inequality*. However, when we consider a particular causally non-separable process given by

$$W^{A_I A_O B_I B_O} = \frac{1}{4} [\mathbb{1}^{A_I A_O B_I B_O} + \frac{1}{\sqrt{2}} (\sigma_z^{A_O} \sigma_z^{B_I} + \sigma_z^{A_I} \sigma_x^{B_I} \sigma_z^{B_O})]. \quad (6.20)$$

Here, for the sake of convenience, we drop the tensor products and identity operations. The local operations on the other hand are as follows: Bob adopts the following protocol. If he wants to read Alice's bit ($b=1$), he measures the incoming qubit in the σ_z basis and assigns $m' = 0$, $m' = 1$ to the outcomes $|0\rangle$, $|1\rangle$ respectively. On the other hand, if he wants to send his bit ($b = 0$), he measures his qubit in the σ_x basis, if the outcome is $|+\rangle$, he encodes k as $0 \rightarrow |0\rangle$, $1 \rightarrow |1\rangle$, whereas if the outcome is $|-\rangle$, he encodes it as $0 \rightarrow |1\rangle$, $1 \rightarrow |0\rangle$. Together with the local operations and the background process as in Eq. (6.20), we can violate the causal inequality with

$$p_{\text{succ}} = (2 + \sqrt{2})/4 > 3/4. \quad (6.21)$$

In Refs. [224, 225] it has been shown that the above probability of success is the maximum possible violation for all causally separable processes. To conclude, we observe violation of causal inequality, which is a classical communication task, is possible with the indefinite causal order. Hence, it is interesting to attribute different information theoretical models, e.g., classical capacity of different asymptotic settings, to such higher-order processes. Motivated by this goal, in the next section, we present one-way communication through processes.

6.5 One-directional communication through an indefinitely ordered process

In this section we introduce our classical communication protocol through an arbitrary process, as shown in Fig. 6.1. In this protocol, both Alice and Bob can use some quantum channels \mathcal{A} and \mathcal{B} . Alice's channel \mathcal{A} has $A'_I A_I$ as input and A_O as output, while Bob's channel \mathcal{B} has B_I as input and $B_O B'_O$ as output.

The process matrix $W^{A_I A_O B_I B_O}$ acting on these channels forms a new quantum channel $\mathcal{N}(\mathcal{A}, \mathcal{B}, W)$ with input quantum system A'_I and the output quantum system B'_O , as shown in Fig. 6.1. The Choi representation of this new channel is

$$\begin{aligned} \mathcal{N}(\mathcal{A}, \mathcal{B}, W)^{A'_I B'_O} & \\ & := (A^{A'_I A_I A_O} * B^{B_I B_O B'_O}) * W^{A_I A_O B_I B_O} \\ & = (A^{A'_I A_I A_O} \otimes B^{B_I B_O B'_O}) * W^{A_I A_O B_I B_O}. \end{aligned} \quad (6.22)$$

Here, A and B are the Choi representations of the quantum channels \mathcal{A} and \mathcal{B} respectively.

It is worth stressing that turning a process into a channel does not yet provide an estimate of the communication capacity. This is because the channel's input and output, A'_I and B'_O , can have arbitrary dimensions, which, in principle, could encode an arbitrary amount of information. For example, Bob could have a single-qubit B_I and B_O and a two-qubit B'_O . One of the two qubits is simply B_I , while the other can be a qubit entangled with B_O (that is to say, he sends through his output part of an entangled state while measuring the other part together with B_I). For a process with indefinite causal order, Bob's output is not discarded, and one could expect that the correlations between B_O and Alice's systems could allow her to encode two bits into B'_O , exceeding the Holevo bound. As we will see, this is in fact not possible, as the receiver's output turns out to be useless in a one-way communication scenario, even for processes with indefinite causal order. Before we can get to this, however, we have to define appropriate asymptotic settings for general processes.

6.5.1 Asymptotic setting

Similar to the conventional classical communication through quantum channels, we need to introduce a notion of asymptotic setting for process, namely to formalise the notion of repeated use of independent copies of a process. The goal turns out to be non-trivial as one can construct asymptotic settings by allowing joint local operations across different copies of processes [206, 221], resulting in non-trivial constraints on the admissible operations and processes [226, 227]. For example, Alice could feed the output of her first channel to her second one. This, however, would require extra knowledge about the causal relations between the different uses of the process and, for a process with bidirectional signalling, it would be incompatible with Bob sending his second output to the first input. As we are investigating causal structures as communication resources, we assume that all available causal relations are encoded in the process itself, which leads to the asymptotic setting, first introduced in Ref. [228], where only product operations across different parties are allowed.

Our choice of asymptotic setting results in a set of independent channels $\mathcal{N}_j = \mathcal{N}(\mathcal{A}_j, \mathcal{B}_j, W)$, as shown in Fig. 6.5. Here $\mathcal{A}_j, \mathcal{B}_j$ are the local operations performed by Alice and Bob respectively. Alice encodes her message m in a quantum state $\rho_m^{(n)} \in \otimes_{j=1}^n A_I^j$ and sends the state to Bob through the channels $\{\mathcal{N}_j\}$. After receiving the transformed state, Bob performs a POVM on his quantum system $\otimes_{j=1}^n B_O^j$. With this, we conceptualise a protocol for one way communication from Alice to Bob ($A \rightarrow B$) in the following way:

Definition 1. *Given a bipartite processes matrix W^{AB} , we define an $A \rightarrow B$ protocol with n uses of W as*

1. *A set of local operations $\{\mathcal{A}_j, \mathcal{B}_j\}_{j=1}^n$, where*

$$\begin{aligned} \mathcal{A}_j &: A_I^j \otimes A_I^j \rightarrow A_O^j, \\ \mathcal{B}_j &: B_I^j \rightarrow B_O^j \otimes B_O^j \end{aligned}$$

are CPTP maps;

2. A state encoding $m \mapsto \rho_m^{(n)} \in \bigotimes_{j=1}^n A_I^j$, where $\rho_m^{(n)} \geq 0$ and $\text{Tr} \rho_m^{(n)} = 1$;
3. A decoding POVM $\{E_{m'}^{(n)}\}_{m'}$, where $E_{m'}^{(n)} \geq 0$ and $\sum_{m'} E_{m'}^{(n)} = \mathbb{1}$.

Such a protocol produces a classical channel described by the conditional probabilities

$$P(m'|m) = \text{Tr} \left[E_{m'}^{(n)} \bigotimes_{j=1}^n \mathcal{N}_j(\rho_m^{(n)}) \right]. \quad (6.23)$$

We say that two protocols for the same process W are *equivalent* if they produce the same conditional probabilities $P(m'|m)$.

In general, the ancillary spaces A_I^j, B_O^j need not be isomorphic for different j . However, we can always embed each of them into a space isomorphic to one of the highest dimension. In the following, we assume that all spaces are of equal dimension and are identified through a choice of canonical basis.

Note that when we consider multiple copies of processes, in general, the local CPTP maps of Alice and Bob need not be identical. Thus the asymptotic setting of quantum process results in multiple copies of different channels, also known as *non-stationary asymptotic setting*. Formulating communication capacity of such a setup poses a non-trivial challenge [229, 230]. To alleviate this issue, we employ a scheme to make the channels stationary. Specifically, we replace the local operations $\{\mathcal{A}_j\}$ and $\{\mathcal{B}_j\}$ with fixed local operations $\{\mathcal{A}\}$ and $\{\mathcal{B}\}$ respectively with support of additional local CPTP maps \mathcal{E}_j and \mathcal{F}_j , where $\mathcal{A}_j = \mathcal{A} * \mathcal{E}_j$ and $\mathcal{B}_j = \mathcal{F}_j * \mathcal{B}$. Thus we have multiple independent and identical copies of the channel $\mathcal{N} = \mathcal{N}(\mathcal{A}, \mathcal{B}, W)$. Feasibility of this approach is due to the fact that, in an $A \rightarrow B$ protocol, Bob's output system B_O can be discarded, i.e., in such a protocol a process matrix W can be replaced by ${}_{B_O}W$, as shown in Fig. 6.3 and in Refs. [4, 5, 7, 231]. As we are going to use this fact multiple times, we formulate it as a lemma and prove it below for completeness:

Lemma 1. *If Alice has trivial ancillary output ($d_{A'_I} = 1$), i.e., $\mathcal{A} : A'_I \otimes A_I \rightarrow A_O$, we can replace W with ${}_{B_O}W$:*

$$\mathcal{N}(\mathcal{A}, \mathcal{B}, W) = \mathcal{N}(\mathcal{A}, \mathcal{B}, {}_{B_O}W). \quad (6.24)$$

Proof. It is sufficient to show $A * W = A * {}_{B_O}W$. Using condition (6.16), we can write $A * W = A * {}_{A_O}W + A * {}_{B_O}W - A * {}_{A_O B_O}W$. As the second and third terms are already in the desired form, we only need to look at the first term:

$$\begin{aligned} A * {}_{A_O}W &= \text{Tr}_{A_I A_O} \left[A^{A'_I A_I A_O} \cdot \left(\frac{\mathbb{1}^{A_O}}{d_O^A} \otimes \text{Tr}_{A_O} W \right) \right] \\ &= \frac{1}{d_O^A} \text{Tr}_{A_I} \left[\left(\text{Tr}_{A_O} A^{A'_I A_I A_O} \right) \cdot \left(\text{Tr}_{A_O} W \right) \right] \\ &= \frac{1}{d_O^A} \text{Tr}_{A_I} \left[\mathbb{1}^{A'_I A_I} \cdot \left(\text{Tr}_{A_O} W \right) \right] \\ &= \frac{\mathbb{1}^{A'_I}}{d_O^A} \otimes \text{Tr}_{A_I A_O} W = \frac{\mathbb{1}^{A'_I}}{d_O^A} \otimes \text{Tr}_{A_I A_O} ({}_{B_O}W), \end{aligned} \quad (6.25)$$

where we used $\text{Tr}_{A_O} A^{A'_I A_I A_O} = \mathbb{1}^{A'_I A_I}$ (because $A^{A'_I A_I A_O}$ is CPTP) in the third line and Eq. (6.15) in the last line. \square

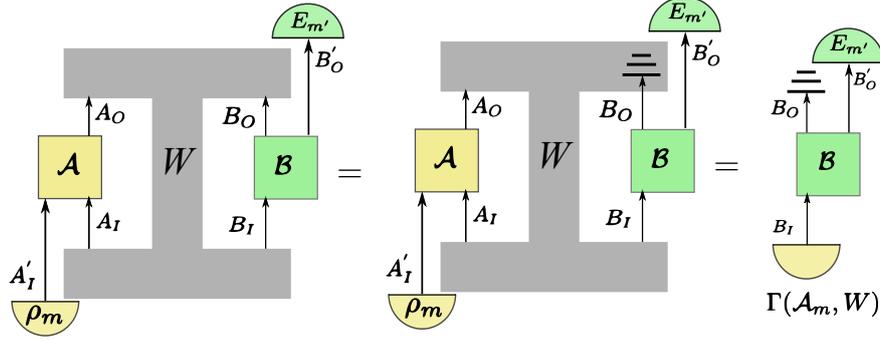


Figure 6.3: Pictorial depiction of Lemma 1. A one way communication from Alice to Bob through a channel $\mathcal{N}(\mathcal{A}, \mathcal{B}, W)$ simplifies to a channel $\mathcal{N}(\mathcal{A}, \mathcal{B}, B_O W)$. The rightmost picture shows further simplification to a quantum state $\Gamma(\mathcal{A}_m, W)$ with the system B_O being discarded, as in Eq. (6.43).

This lemma allows us to replace Bob's operation \mathcal{B}_j by $\sigma^{B_O} \otimes (\text{Tr}_{B_O} \mathcal{B}_j)^{B_I B'_O}$, with σ^{B_O} being an arbitrary state. For the encoding operation \mathcal{A}_j , on the other hand, we extend the input system to make it a controlled operation while treating the control state as Alice's extended encoded message. Thus we present the following theorem.

Theorem 1. *Every $A \rightarrow B$ protocol is equivalent to one with fixed local operations*

$$\bar{\mathcal{A}} : A_I''^j \otimes A_I^j \rightarrow A_O^j, \quad (6.26)$$

$$\bar{\mathcal{B}} : B_I^j \rightarrow B_O^j \otimes B_O''^j, \quad (6.27)$$

state encoding

$$\bar{\rho}_m^{(n)} = \bigotimes_{j=1}^n \mathcal{E}_j(\rho_m^{(n)}), \quad (6.28)$$

and decoding POVM

$$\bar{E}_{m'}^{(n)} = \bigotimes_{j=1}^n \mathcal{F}_j^\dagger(E_{m'}^{(n)}), \quad (6.29)$$

where $\mathcal{E}_j : A_I''^j \rightarrow A_I^j$, $\mathcal{F}_j : B_O^j \rightarrow B_O''^j$ are CPTP and \mathcal{F}^\dagger denotes the Hilbert-Schmidt adjoint, defined through $\text{Tr}[A^\dagger \mathcal{F}(B)] = \text{Tr}[\mathcal{F}^\dagger(A^\dagger) B]$.

Proof. (See Fig. 6.4 for a pictorial representation of the proof.) Let us start with the decoding. Since Alice only performs CPTP maps with no ancillary output, we can apply Lemma 1 and replace the process matrix W with $B_O W$, which is equal to identity on B_O . This implies that any $A \rightarrow B$ protocol is equivalent to one where we replace the local operations \mathcal{B}_j with $\sigma^{B'_O} \otimes \text{Tr}_{B'_O} \mathcal{B}_j$ for some arbitrary state σ . Choosing the space $B_O''^j$ isomorphic to B_I^j , we see that the original combination of local operations \mathcal{B}_j and decoding POVM is equivalent to performing the fixed operation $\bar{\mathcal{B}} = \sigma^{B'_O} \otimes \mathcal{J}^{B_I^j \rightarrow B_O''^j}$ in each lab, and decoding POVM as in Eq. (6.29), with $\mathcal{F}_j = (\text{Tr}_{B'_O} \mathcal{B}_j \circ \mathcal{J}^{B_I^j \rightarrow B_O''^j})^{B'_O}$.

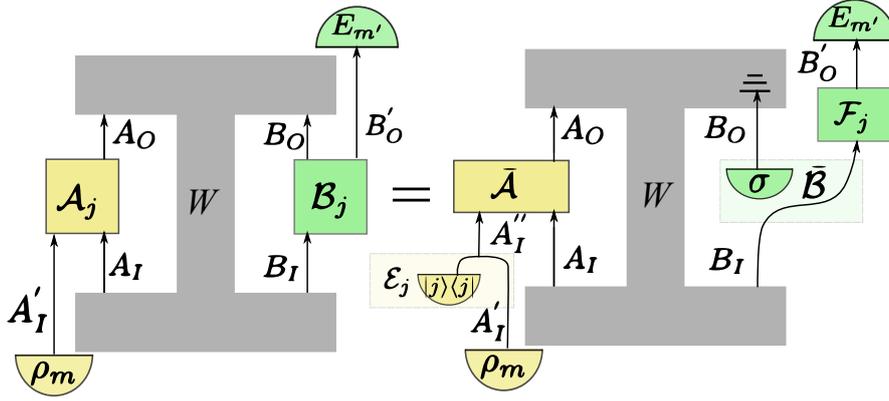


Figure 6.4: Pictorial depiction of Theorem 1. We convert a non-stationary channel $\mathcal{N}(\mathcal{A}_j, \mathcal{B}_j, W)$ to a stationary channel $\mathcal{N}(\bar{\mathcal{A}}, \bar{\mathcal{B}}, W)$. Due to Lemma 1, Bob's system B_O can be set to a fixed state σ and corresponding operation $\bar{\mathcal{B}} = \sigma^{B_O^j} \otimes \mathcal{J}^{B_I^j \rightarrow B_O''^j}$. Alice's operation, on the other hand, can be extended to a controlled CPTP map $\bar{\mathcal{A}}$ as described in Eq. (6.31).

Now for the encoding side: we set $A_I''^j = \mathcal{X}_j \otimes A_I'^j$ and define the controlled operation $\bar{\mathcal{A}} : \mathcal{X}_j \otimes A_I'^j \rightarrow A_O^j$ as

$$\bar{\mathcal{A}}(\sigma \otimes \rho) = \sum_{j=1}^n \langle j | \sigma | j \rangle \mathcal{A}_j(\rho)^{A_O^j}, \quad (6.30)$$

Here, each \mathcal{X}_j is a state space isomorphic to the operator space $\mathcal{L}(\mathbb{C}^n)$. The operation $\bar{\mathcal{A}}$ is manifestly CPTP. For canonical basis states in \mathcal{X}_j , this map gives

$$\bar{\mathcal{A}}(|j\rangle\langle j| \otimes \rho) = \mathcal{A}_j(\rho), \quad (6.31)$$

so the choice of local operation can be encoded into a choice of initial state, expanding the original encoding state as in Eq. (6.28), with the maps $\mathcal{E}_j : A_O'^j \rightarrow A_O''^j$ defined as

$$\mathcal{E}_j\left(\rho^{A_I'^j}\right) = (|j\rangle\langle j| \otimes \rho)^{A_I''^j}. \quad (6.32)$$

□

The relevance of this theorem is twofold. First, it shows that any protocol involving a different choice of local operations can be reproduced by fixing the local operations once and for all. This means that an asymptotic setting for processes can always be mapped to an asymptotic setting where the same channel is used n times, $\mathcal{N}(\bar{\mathcal{A}}, \bar{\mathcal{B}}, W)^{\otimes n}$. We call a protocol of this type *stationary*. Second, state encoding and decoding POVM of an arbitrary protocol transform into those of a stationary one through product maps, Eqs. (6.28) and (6.29). This means that the transformation preserves the nature of the asymptotic setting, viz. joint/product encoding or decoding. From now on, we will represent $\bar{\mathcal{A}}$, $\bar{\mathcal{B}}$, $\bar{\rho}_m^{(n)}$ and $\bar{E}_{m'}^{(n)}$ without the bar on top.

6.5.2 Classical capacities of a quantum process

Holevo quantity for a process: Having introduced a stationary protocol with an asymptotic setting of the channel $\mathcal{N}(\mathcal{A}, \mathcal{B}, W)$, as shown in Theorem 1, we can define the corresponding Holevo quantity for

a process W as

$$\chi(W)^{A \rightarrow B} := \max_{\mathcal{A}, \mathcal{B}} \chi[\mathcal{N}(\mathcal{A}, \mathcal{B}, W)]. \quad (6.33)$$

We also introduce the n -th extension $\chi(W^{\otimes n})^{A \rightarrow B}$ of the above quantity as

$$\chi(W^{\otimes n})^{A \rightarrow B} := \max_{\mathcal{A}, \mathcal{B}} \chi[\mathcal{N}(\mathcal{A}, \mathcal{B}, W)^{\otimes n}]. \quad (6.34)$$

Communication capacity for a process: We can associate different communication capacities to an arbitrary process as

$$C^\sharp(W)^{A \rightarrow B} = \max_{\mathcal{A}, \mathcal{B}} C^\sharp(\mathcal{N}(\mathcal{A}, \mathcal{B}, W)). \quad (6.35)$$

Where $C^\sharp = C^{(1)}, C, C^\infty$. Here $C^{(1)}(W)^{A \rightarrow B}$ represents product encoding-product decoding capacity, as in Eq. (6.4), $C(W)^{A \rightarrow B}$ represents product encoding-joint decoding capacity, as in Eq. (6.6) and finally, $C^\infty(W)^{A \rightarrow B}$ represents joint encoding-joint decoding capacity, as in Eq. (6.8).

We can relate the Holevo quantity for a process to different $C^\sharp(W)^{A \rightarrow B}$ capacities. We show this in the following lemma.

Lemma 2. *Different capacities associated with an arbitrary process W are related to the Holevo quantity $\chi(W)^{A \rightarrow B}$ in the following way.*

product encoding-product decoding:

$$C^{(1)}(W)^{A \rightarrow B} \leq \chi(W)^{A \rightarrow B}. \quad (6.36)$$

product encoding-joint decoding:

$$C(W)^{A \rightarrow B} = \chi(W)^{A \rightarrow B}. \quad (6.37)$$

Joint encoding-joint decoding:

$$C^\infty(W)^{A \rightarrow B} = \lim_{n \rightarrow \infty} \frac{\chi(W^{\otimes n})}{n}. \quad (6.38)$$

Proof. product encoding-product decoding:

Using Eqs. (6.35), (6.5) and (6.33), we can write

$$\begin{aligned} C^{(1)}(W)^{A \rightarrow B} &= \max_{\mathcal{A}, \mathcal{B}} C^{(1)}(\mathcal{N}(\mathcal{A}, \mathcal{B}, W)) \\ &\leq \max_{\mathcal{A}, \mathcal{B}} \chi(\mathcal{N}(\mathcal{A}, \mathcal{B}, W)) \\ &= \chi(W)^{A \rightarrow B}. \end{aligned} \quad (6.39)$$

product encoding-joint decoding:

Using Eqs. (6.35), (6.6) and (6.33) we can write

$$\begin{aligned} C(W)^{A \rightarrow B} &= \max_{\mathcal{A}, \mathcal{B}} C(\mathcal{N}(\mathcal{A}, \mathcal{B}, W)) \\ &= \max_{\mathcal{A}, \mathcal{B}} \chi(\mathcal{N}(\mathcal{A}, \mathcal{B}, W)) \\ &= \chi(W)^{A \rightarrow B}. \end{aligned} \quad (6.40)$$

Joint encoding-joint decoding:

Using Eqs. (6.35), (6.7), (6.8) and (6.34) we can write

$$\begin{aligned}
C^\infty(W)^{A \rightarrow B} &= \max_{\mathcal{A}, \mathcal{B}} C^\infty(\mathcal{N}(\mathcal{A}, \mathcal{B}, W)) \\
&= \max_{\mathcal{A}, \mathcal{B}} \chi_{\text{reg}}(\mathcal{N}(\mathcal{A}, \mathcal{B}, W)) \\
&= \max_{\mathcal{A}, \mathcal{B}} \lim_{n \rightarrow \infty} \frac{\chi[\mathcal{N}(\mathcal{A}, \mathcal{B}, W)^{\otimes n}]}{n} \\
&= \lim_{n \rightarrow \infty} \frac{\chi(W^{\otimes n})}{n}.
\end{aligned} \tag{6.41}$$

□

6.5.3 Bounds on the classical capacities of a quantum process

Although we have been able to reduce the classical capacities of processes to that of channels, our results so far do not provide an upper bound on how much information can be transmitted through a process. This is because the channel $\mathcal{N}(\mathcal{A}, \mathcal{B}, W)$ can have arbitrary input and output dimension.

To establish a bound, we first describe our protocol from a slightly different point of view. With Alice's input ensemble $\{p(m), \rho_m\}$, we can introduce a concatenation of ρ_m with the channel \mathcal{A} as $A_m^{A_I A_O} = \rho_m^{A_I'} * A^{A_I'' A_I A_O}$ with A_m being the Choi representation of the resulting CPTP map \mathcal{A}_m . Similarly, we can combine Bob's channel \mathcal{B} and POVM operation $\{E_m\}$ to describe a set of CP maps $\{B_{m'}^{B_I B_O} = B^{B_I B_O B_O'} * (E_{m'}^T)^{B_O'}\}_{m'}$, where $\sum_{m'} B_{m'}^{B_I B_O}$ is a CPTP map and $(E_{m'}^T)^{B_O'}$ is the Choi representation of $E_{m'}$. The superscript 'T', denoting the transpose operator, is due to definition (6.10), according to which the Choi of a measurement operator $E_{m'}$ is its transpose $E_{m'}^T$.

With this in mind, we present two theorems, that apply respectively to the product and joint encoding scenarios.

Theorem 2. *In a one-way communication scenario, the optimisation of the Holevo quantity of a process W can be simplified as*

$$\begin{aligned}
\chi(W)^{A \rightarrow B} &= \max_{\rho_m, \mathcal{A}, p(m)} S\left[\sum p(m) \Gamma(A_m, W)\right] - \\
&\quad \sum p(m) S[\Gamma(A_m, W)].
\end{aligned} \tag{6.42}$$

Here the $\Gamma(A_m, W) := A_m * \text{Tr}_{B_O} W / d_{B_O}$ is a map that transforms the Choi representation of the CPTP operation \mathcal{A}_m and the process matrix W , to a quantum state on Bob's input space B_I .

Proof. The reduced process [181] on which Bob applies his CPTP map $B_{m'}^{B_I B_O}$ is described by $A_m^{A_I A_O} * W^{A_I A_O B_I B_O}$. Now, using Lemma 1, we can write

$$\begin{aligned}
A_m^{A_I A_O} * W^{A_I A_O B_I B_O} &= \mathbb{1}^{B_O} \otimes \left(A_m^{A_I A_O} * \frac{\text{Tr}_{B_O} W^{A_I A_O B_I}}{d_{B_O}} \right) \\
&= \mathbb{1}^{B_O} \otimes \Gamma(A_m, W)^{B_I}.
\end{aligned} \tag{6.43}$$

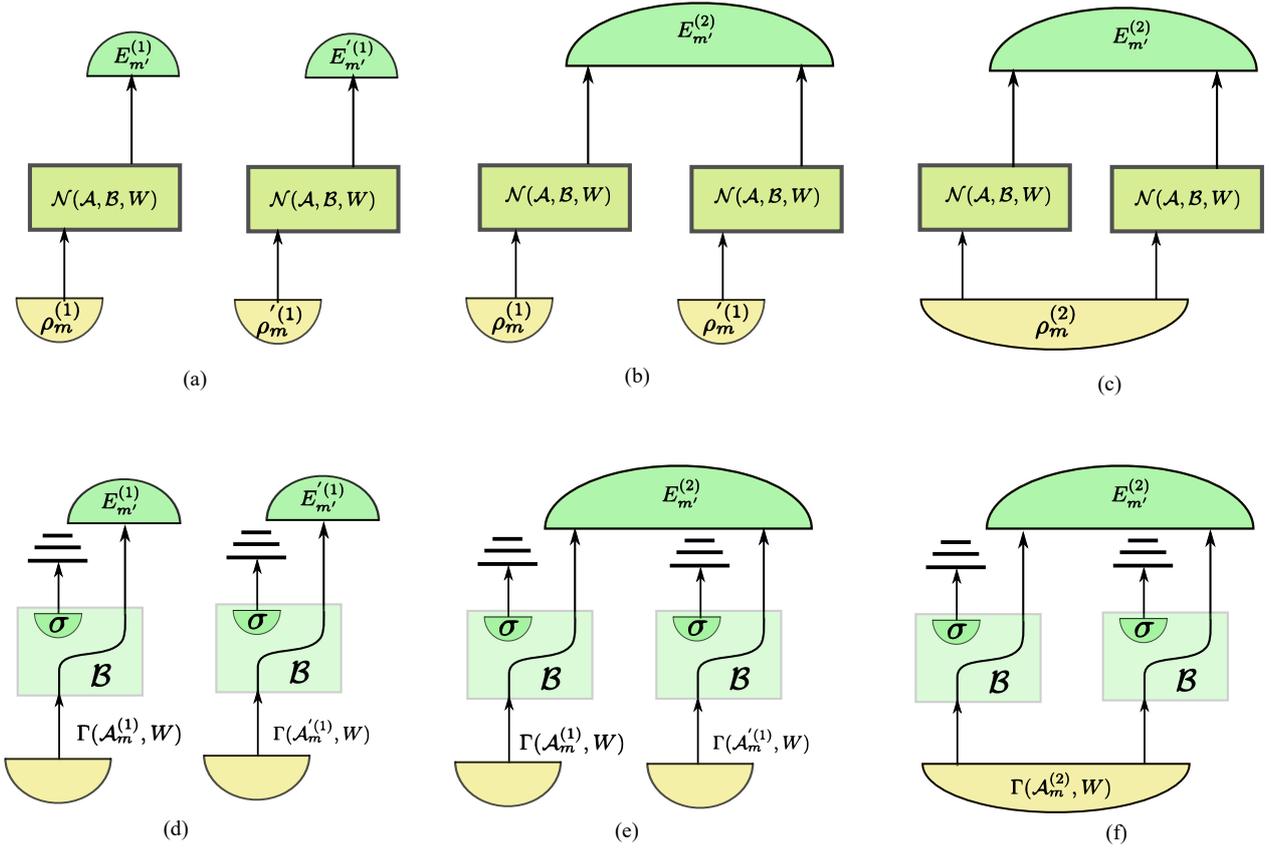


Figure 6.5: Encoding and decoding schemes. Here we show two copies of process W . Fig. (a), (b), (c) demonstrate product-encoding, product decoding, product encoding joint decoding and joint encoding-joint decoding respectively. Product encoding is achieved using the joint state $\rho_m^{(1)} \otimes \rho_m'^{(1)}$ and joint encoding is achieved using the entangled state $\rho_m^{(2)}$. Similarly, we use $E_m^{(1)} \otimes E_m'^{(1)}$ for product decoding and $E_m^{(2)}$ for joint decoding. Figs. (d), (e), (f) are the simplifications due to Theorem 1 and Theorem 2. Relevant labelling of the Hilbert spaces are described in the text.

Where $\Gamma(A_m, W) := A_m * \text{Tr}_{B_O} W / d_{B_O}$. In other words, as shown in Fig. 6.3, we can simplify Alice's CPTP map and the process to a quantum state $\Gamma(A_m, W)$ in the Hilbert space B_I , with Bob's output at B_O being discarded. The maximum classical information that can be encoded in the ensemble $\{p(m), \Gamma(A_m, W)\}$ is given by Eq. (6.42) [59] where we only need to optimise over the free parameters $p(m)$, ρ_m and A .

This implies one does not need to optimise over Bob's operation to obtain the Holevo quantity for the process. \square

A direct consequence of this theorem is that $\chi(W)^{A \rightarrow B} \leq \log(d_{B_I})$, because we have reduced the Holevo quantity of a process to that of an ensemble of states in B_I . In turn, this allows us to establish a bound on the product encoding capacities, i.e., $C^{(1)}(W)^{A \rightarrow B}$ and $C(W)^{A \rightarrow B}$, according to Eq. (6.39) and Eq. (6.40), respectively. However, for joint encoding schemes we need to evaluate the regularised Holevo quantity for the optimum channel $\mathcal{N}(A, B, W)$, as shown in Eq. (6.38). In light of this, we introduce the following theorem that bounds the capacity, $C^\infty(W)^{A \rightarrow B}$.

Theorem 3. *The joint encoding capacity for a process, $C^\infty(W)^{A \rightarrow B}$, is limited to Alice's joint CPTP map $\mathcal{A}_m^{(n)} = \rho_m^{(n)} * \mathcal{A}^{\otimes n}$ with $n \rightarrow \infty$, and the distribution $p(m)$.*

$$C^\infty(W)^{A \rightarrow B} = \lim_{n \rightarrow \infty} \max_{\rho_m^{(n)}, \mathcal{A}, p(m)} \frac{1}{n} \left(S \left[\sum p(m) \Gamma(\mathcal{A}_m^{(n)}, W) \right] - \sum p(m) S \left[\Gamma(\mathcal{A}_m^{(n)}, W) \right] \right). \quad (6.44)$$

Here the map Γ transforms the joint CPTP map $\mathcal{A}_m^{(n)}$ and the process W to an entangled state at Bob's input space $\otimes_{j=1}^n B_I^j$.

Proof. First, we apply Lemma 1 to each copy of the channel $\mathcal{N}(\mathcal{A}, \mathcal{B}, W)$ and replace it with $\mathcal{N}(\mathcal{A}, \mathcal{B}, B_O^j W)$. Then, in a joint encoding scheme, we apply $\otimes_j \mathcal{N}(\mathcal{A}, \mathcal{B}, B_O^j W)$ to a (possibly entangled) joint state $\rho_m^{(n)}$. Combining this joint state with the n copies of Alice's operation \mathcal{A} , we obtain a joint CPTP map $\mathcal{A}_m^{(n)} : \otimes_j A_I^j \rightarrow \otimes_j A_O^j$, with Choi representation $A_m^{(n)} = \rho_m^{(n)} * \mathcal{A}^{\otimes n}$. Plugging $A_m^{(n)}$ into the n copies $B_O^j W$, we get $A_m^{(n)} * \left(\otimes_{j=1}^n B_O^j W \right) = \mathbb{1}_{\otimes_{j=1}^n B_O^j} \otimes \Gamma(\mathcal{A}_m^{(n)}, W)$, where $\Gamma(\mathcal{A}_m^{(n)}, W) \in \otimes_{j=1}^n B_I^j$ is a (possibly entangled) state, defined as

$$\Gamma(\mathcal{A}_m^{(n)}, W) = \frac{A_m^{(n)} * \left(\text{Tr}_{\otimes_{j=1}^n B_O^j} W^{\otimes n} \right)}{\prod_{j=1}^n d_{B_O^j}}. \quad (6.45)$$

One can extend this setup to $n \rightarrow \infty$ and achieve a joint state at Bob's input Hilbert space $\otimes_{j=1}^\infty B_I^j$. Similar to Theorem 2, we calculate the maximum amount of classical information encoded in the ensemble $\{p(m), \Gamma(\mathcal{A}_m^{(n)}, W)\}$ and regularise it to obtain the joint encoding capacity $C^\infty(W)^{A \rightarrow B}$ where the free parameters are of course, $p(m)$, $\rho_m^{(n)}$ and \mathcal{A} . Thus we obtain Eq. (6.44). \square

Corollary 3.1. *The capacity $C^\infty(W)^{A \rightarrow B}$ is upper bounded by the logarithm of the dimension of Bob's input Hilbert space, i.e. $C^\infty(W)^{A \rightarrow B} \leq \log(d_{B_I})$.*

Proof. This is the consequence of Holevo's theorem [59]. The information content of the ensemble $\{p(m), \Gamma(\mathcal{A}_m^{(n)}, W)\}$ cannot exceed logarithm of the dimension of $\Gamma(\mathcal{A}_m^{(n)}, W)$, i.e., $n \log(d_{B_I})$. Regularising this quantity proves the corollary. \square

Now we summarise our results. If we consider Alice's input message m and Bob's output message m' , we can introduce a chain of inequalities for different classical capacities of the process W .

$$\begin{aligned} I(m' : m) &\leq C^{(1)}(W)^{A \rightarrow B} \leq C(W)^{A \rightarrow B} = \chi(W)^{A \rightarrow B} \\ &\leq C^\infty(W)^{A \rightarrow B} \leq \log(d_{B_I}). \end{aligned} \quad (6.46)$$

One can write down a similar chain of inequalities for a communication protocol from Bob to Alice. Note that this inequality holds even for a process W that contains shared entanglement between

Alice's and Bob's input Hilbert spaces. This does not contradict the higher capacity achievable in an entanglement assisted communication protocol, such as super-dense coding [19]. This is because when applying the inequalities in Eq. (6.46), one has to consider the total dimension of Bob's input Hilbert space, which consists of the part of the shared entangled state in Bob's possession and the quantum state that Alice communicates to him.

6.6 Broadcast communication

Having established the notion of one-way communication through a process, we proceed to explore scenarios where all communicating parties can transmit and receive information.

Two-party communication: Let us first consider the two-party situation, where Alice (Bob) sends the message $m(k)$ and Bob (Alice) receives the message $m'(k')$. The possibility to violate causal inequalities indicates that indefinite causal order can indeed provide an advantage in some two party games [7, 219]; however, it is unclear if this advantage results in a communication enhancement. To address this question, it is necessary to find limits on two-way communication for causally separable processes. In this section we address this question.

There are at least two ways a process can be used as a resource for bidirectional communication, depending on whether Alice's and Bob's instruments are fixed or if they are chosen depending on the direction of communication attempted. In the first case, the parties produce a single probability distribution $P(m', k' | m, k)$ from the process, and one looks for communication in the marginals $P(m' | m)$, $P(k' | k)$. In the second case, the parties can generate different probability distributions depending on who is sending and who is receiving. The one-directional capacities for the first case are upper bounded by those in the second case, as the best instrument to receive a message might differ from the best to send a message. In this section, we will be mostly concerned with the second case.

Let us then consider a scenario where the order between Alice's and Bob's local operations is determined based on a random outcome. We represent a process where Alice can signal to Bob, but not the other way around, by $W^{A \prec B} = {}_{B_O} W^{A \prec B}$ and the reversed direction of signalling by $W^{B \prec A} = {}_{A_O} W^{B \prec A}$. The process matrix W_{Sep} in this case is a convex combination of $W^{B \prec A}$ and $W^{A \prec B}$ [181]:

$$W_{\text{sep}} = \lambda W^{B \prec A} + (1 - \lambda) W^{A \prec B}, \quad (6.47)$$

where $0 \leq \lambda \leq 1$ is the probability for Bob to be first. We call such a process a *causally separable process* [7]. We investigate a scenario where both Alice and Bob are trying to send information to each other through the background process W_{Sep} . A reasonable attempt to quantify this bi-directional communication is to evaluate the sum of two product capacities, $C^{(1)}(W)^{A \rightarrow B}$ and $C^{(1)}(W)^{B \rightarrow A}$. We investigate this quantity and evaluate an operationally significant upper bound from the perspective of the classical capacity of the process.

Theorem 4. *For a bi-directional communication protocol through a causally separable process, defined in Eq. (6.47), the following inequality holds:*

$$\begin{aligned} C^{(1)}(W_{\text{sep}})^{A \rightarrow B} + C^{(1)}(W_{\text{sep}})^{B \rightarrow A} \\ \leq \lambda \log(d_{A_I}) + (1 - \lambda) \log(d_{B_I}). \end{aligned} \quad (6.48)$$

Proof. Considering a fixed input probability distribution $P(a)$, the following linear relationship among marginal conditional probabilities holds for a causally separable process [7].

$$P(a'|a)_{W_{\text{sep}}} = \lambda P(a'|a)_{W^{B \prec A}} + (1 - \lambda) P(a'|a)_{W^{A \prec B}}. \quad (6.49)$$

In our protocol, we consider $a \in \{m, k\}$ to be the inputs and $a' \in \{m', k'\}$ to be the outputs. Let us consider $A \rightarrow B$ communication. Consequently we can write:

$$\begin{aligned} C^{(1)}(W_{\text{Sep}})^{A \rightarrow B} &= \max I(m' : m) \\ &\leq \max [\lambda I(m' : m)_{W^{B \prec A}} + (1 - \lambda) I(m' : m)_{W^{A \prec B}}] \\ &= (1 - \lambda) \max I(m' : m)_{W^{A \prec B}} \\ &\leq (1 - \lambda) \log(d_{B_I}). \end{aligned} \quad (6.50)$$

Here, the first equation is due to Eqs. (6.4), and (6.35). The maximisation is taken over Alice's and Bob's local operations, their message ensembles and the POVM operations. The first inequality is due to the fact that mutual information $I(a' : a)$ is a convex function of $p(a'|a)$ for a fixed input probability distribution $p(a)$ [1]. We obtain the second equality because, for a definite ordered scenario $B \prec A$, output m' of Bob's local lab becomes independent of Alice's input. This makes $I(m : m')_{W^{B \prec A}} = 0$. The final inequality is due to Eq. (6.46). We apply a similar set of reasoning to obtain a bound for $B \rightarrow A$ communication to obtain,

$$C^{(1)}(W_{\text{Sep}})^{B \rightarrow A} \leq \lambda \log(d_{A_I}). \quad (6.51)$$

Adding Eq. (6.50) and (6.51), we find

$$\begin{aligned} C^{(1)}(W_{\text{sep}})^{A \rightarrow B} + C^{(1)}(W_{\text{sep}})^{B \rightarrow A} \\ \leq \lambda \log(d_{A_I}) + (1 - \lambda) \log(d_{B_I}). \end{aligned} \quad (6.52)$$

□

For the particular case $d_{A_I} = d_{B_I} = d$, we see that the sum of two product capacities is upper bounded by $\log(d)$. In other words, the total communication in causally separable processes can be no more than maximal one-way communication. We note that a weaker version of this inequality holds for the scenario where the parties' instruments are fixed regardless of the attempted direction of communication. In this case, the single-shot capacities coincide with the mutual information obtained from a single conditional probability distribution $P(m', k' | m, k)$, resulting in the inequality $I(m' : m) + I(k' : k) \leq \log(d)$. This is an example of an *entropic causal inequality*, first considered in

Ref. [232]. Remarkably, no violation of this inequality is known, and our own numerical search also did not reveal any violation of Eq. (6.48). This suggests that the bound on the total bidirectional communication might hold for general processes.

We note that the bound we established applies to all the quantum processes for which a physical interpretation is known. For example, in a process with coherent control of causal order, such as the quantum switch [6], tracing out the control leads to a separable bipartite process, to which the bound applies. More generally, it has been shown that any bipartite processes that admit a unitary extension is causally separable [233, 234].

Multi-party communication: The above-mentioned protocol can be extended to multiple parties. In this case, each party tries to communicate his/her information to the remaining parties. Similarly to above, we consider a process for N parties, $A^{(1)}, A^{(2)}, \dots, A^{(N)}$ that can be written as a probabilistic mixture of permutations of different causal order:

$$W_{\text{sep}}^N = \sum_{\sigma} q_{\sigma} W^{\sigma}. \quad (6.53)$$

Here, σ denotes the different permutations of the communicating parties and q_{σ} denotes the probability of occurrence of each permutation. Although this is not the most general process with definite causal order [154, 185], it is one of particular interest, as it represents a scenario where the order among parties can be set by external, random, variables, but is independent of the parties' actions.

Motivated by the previous section, we intend to find an upper bound to the quantity $\sum_{i,j} C^{(1)}(W_{\text{sep}}^N)^{i \rightarrow j}$. Here $i \rightarrow j$ refers to signalling from the party $A^{(i)}$ to the party $A^{(j)}$ and $C^{(1)}(W_{\text{sep}}^N)^{i \rightarrow j}$ is the one-way capacity defined for the bipartite reduced process for $A^{(i)}$ and $A^{(j)}$, obtained by fixing CPTP maps for all other parties, and maximised over the other parties' CPTP maps.

Theorem 5. *If the dimensions of all the input Hilbert spaces of the communicating parties are equal (d), then*

$$\sum_{i,j} C^{(1)}(W_{\text{sep}}^N)^{i \rightarrow j} \leq \frac{N(N-1)}{2} \log(d). \quad (6.54)$$

Proof. we can write the conditional probability $P(\vec{m}' | \vec{m}) = \sum_{\sigma} q_{\sigma} P_{\sigma}(\vec{m}' | \vec{m})$. We can write down the marginals $P(m'_j | m_i) = \sum_{\sigma} q_{\sigma} P_{\sigma}(m'_j | m_i) \forall i, j$. By the convexity of mutual information and the inequalities introduced in Eq. (6.46):

$$\begin{aligned} C^{(1)}(W_{\text{sep}}^N)^{i \rightarrow j} &\leq \max_{\sigma} \sum_{\sigma} q_{\sigma} I_{\sigma}(m'_j : m_i) \\ &= \max_{\forall \{i,j\} | \sigma(i) \prec \sigma(j)} \sum_{\sigma} q_{\sigma} I_{\sigma}(m'_j : m_i) \\ &\leq \sum_{\forall \{i,j\} | \sigma(i) \prec \sigma(j)} q_{\sigma} \log(d_{A_i^j}) \\ &= \sum_{\forall \{i,j\} | \sigma(i) \prec \sigma(j)} q_{\sigma} \log(d) \end{aligned} \quad (6.55)$$

The maximisation is taken over all communicating parties' local operations, their message ensembles and the POVM operations. The first inequality is due to the convexity of mutual information relative

to mixtures of conditional probabilities (as in the bipartite case). The first equality follows from the fact that if the party $\sigma(j)$ is in the causal past of the party $\sigma(i)$, then $I(m'_j:m_i) = 0$. d_{A_j} is the dimension of the input Hilbert space of the party A^j . The second equality follows because of our assumption of all the dimensions of the input Hilbert spaces being equal. Now, it is easy to see that the n -th party has total $n-1$ parties in his/her causal past. Therefore, considering each party trying to communicate with the remaining $N-1$ parties, the total number of available channels are $N(N-1) - \sum_{n=1}^N (n-1) = N(N-1)/2$. This results in

$$\sum_{i,j} C^{(1)}(W_{\text{sep}}^N)^{i \rightarrow j} \leq \frac{N(N-1)}{2} \log(d). \quad (6.56)$$

□

The key property that leads to the above bounds is the convexity of mutual information under probabilistic mixtures of classical channels. With this in mind, we see that the above results can be extended directly to the product encoding, joint decoding setting, replacing the product capacity $C^{(1)}$ with the HSW capacity C . Indeed, we have seen that C is given by the (maximised) Holevo quantity χ of the one-way channel generated by a process and, just like mutual information, χ is convex over the probabilistic mixture of channels. It remains an open question whether higher total transmission rates can be achieved in a joint encoding setting.

6.7 Conclusion

We have formalised classical communication through a general quantum causal structure, which may be either probabilistic or indefinite. We have defined the Holevo quantity as well as different classical capacities for an arbitrary process and established relationships among them. We have found that, for one-way communication through a general higher-order process with the communicating parties admitting an indefinite causal order, the various capacities can be reduced to those of ordinary channels, up to an optimisation over the operations performed in local laboratories. We have further shown that, for one-way communication, the classical capacity of a process cannot exceed the Holevo bound—at most one classical bit per received qubit—even in case of indefinite causal order.

Next, we have quantified bi-directional and more generally broadcast communication protocols for processes with definite but classical mixture of causal orders. We have demonstrated that the total amount of communication between two parties cannot exceed the maximal one-way capacity in a fixed causal order, with a similar bound extending to multipartite broadcast communication. One can ask whether a process with an indefinite causal structure can violate these bounds. We have answered this negatively for coherent control of causal order, as in the quantum switch [5]. It is an open question whether a more general process can violate the bounds. As we have not found any violation, it is an interesting possibility that the bounds we have found might constitute a universal limit to the total communication possible in any process.

6.8 Author contribution

Contributor	Statement of contribution	%
K. Goswami	writing of text	80
	proof-reading	20
	theoretical derivations	60
	preparation of figures	100
	initial concept	20
F. Costa	writing of text	20
	proof-reading	80
	theoretical derivations	40
	initial concept	80

Chapter 7

Conclusion

In this thesis I have used higher-order maps that transform input quantum maps to output quantum maps to explore two applications: (1) non-Markovian quantum dynamics and (2) indefinite causal structures.

Non-Markovian dynamics occur due to the inevitable system-environment interaction where the environment memorises the past information about the systems. The resulting temporal correlations among the operations that act on the system are particularly problematic: we need to estimate the non-Markovian noise to make error correction possible. In Chapter 3, I showed how we experimentally estimated such non-Markovianity by performing tomographically incomplete measurements and then using machine-learning models. Although a complete tomography of the process matrix achieves full characterisation, it would be intractable both computationally and experimentally.

I dedicated the subsequent chapters (Chapter 4, 5, and 6) to describe various aspects of indefinite causal structures. In Chapter 4, I described how we experimentally implemented the quantum switch—a device that demonstrates indefinite causal order. We verified the causal non-separability of the quantum switch by measuring a causal witness. In Chapter 5, I showed an information-theoretic advantage of the quantum switch. I discussed how we can transmit information when the causal order between two channels—one of which is depolarising—is indefinite; such information transmission is impossible in the case of channels with a definite order. Finally, in Chapter 6 I introduced different communication protocols through the most general bipartite processes. We developed notions of classical capacities of quantum processes. We also explored the one-way and broadcast communication protocols. In each case, we established relevant bounds for maximum possible information transmission.

Through the works presented in my thesis, I conveyed the utility of higher-order maps, which capture interesting quantum phenomena that are either problematic or completely impossible to characterise using conventional quantum techniques. There are several avenues of exploration to pursue from here.

Let us first explore further application in discrete-encoded quantum-logic protocols. In Chapter 6, I explored classical communication through higher order maps. This work will next need to be extended to quantum communication, e.g. quantum key distribution; secure multi-party voting, and so on.

Additionally, there are other interesting information-theoretic tasks worth exploring, most notably error-correction.

Very recently there have been initial forays into exploring quantum thermodynamics in the presence of indefinite causal structure [235–237]. For instance, using a quantum-switch, Guha *et. al.* [236] showed extraction of free-work; Felce and Vedral [237] achieved a non-classical cooling. I fully expect that extending these works to arbitrary higher-order maps will discover new quantum thermodynamic phenomena.

Finally, nearly all the work to date with higher-order maps has been with discrete-variable encodings: there remains much work to be done to apply higher-order maps to continuous-variable protocols [238–240]. It is an open question as to whether there will be an information-theoretic advantage in this case, and of course it may yield new capabilities and applications.

The results from this thesis, and the roadmap I have sketched above, suggests that this is only the beginning of using higher-order maps in quantum foundations and quantum technology.

Bibliography

- [1] T. M. Cover, J. A. Thomas, [Elements of information theory](#), Wiley-Interscience (2006).
URL <https://books.google.com.au/books?id=VWq5GG6ycxMC>
- [2] K. Kraus, [States, effects, and operations](#), SpringerVerlag, Berlin (1983).
URL <https://www.springer.com/gp/book/9783540127321>
- [3] G. Chiribella, G. M. D’Ariano, P. Perinotti, [Transforming quantum operations: Quantum supermaps](#), EPL (Europhysics Letters) 83 (3) (2008) 30004. doi:10.1209/0295-5075/83/30004.
URL <http://dx.doi.org/10.1209/0295-5075/83/30004>
- [4] G. Chiribella, G. M. D’Ariano, P. Perinotti, Quantum circuit architecture, Phys. Rev. Lett. 101 (2008) 060401. arXiv:0712.1325, doi:10.1103/PhysRevLett.101.060401.
- [5] G. Chiribella, G. M. D’Ariano, P. Perinotti, B. Valiron, Quantum computations without definite causal structure, Phys. Rev. A 88 (2) (2013) 022318. arXiv:0912.0195, doi:10.1103/PhysRevA.88.022318.
- [6] G. Chiribella, G. M. D’Ariano, P. Perinotti, Theoretical framework for quantum networks, Phys. Rev. A 80 (2) (2009) 022339. arXiv:0904.4483, doi:10.1103/PhysRevA.80.022339.
- [7] O. Oreshkov, F. Costa, Č. Brukner, Quantum correlations with no causal order, Nat. Commun. 3 (2012) 1092. arXiv:1105.4464, doi:10.1038/ncomms2076.
- [8] D. Kretschmann, R. F. Werner, [Quantum channels with memory](#), Phys. Rev. A 72 (2005) 062323. doi:10.1103/PhysRevA.72.062323.
URL <https://link.aps.org/doi/10.1103/PhysRevA.72.062323>
- [9] F. A. Pollock, C. Rodríguez-Rosario, T. Frauenheim, M. Paternostro, K. Modi, [Non-markovian quantum processes: Complete framework and efficient characterization](#), Phys. Rev. A 97 (2018) 012127. doi:10.1103/PhysRevA.97.012127.
URL <https://link.aps.org/doi/10.1103/PhysRevA.97.012127>

- [10] F. A. Pollock, C. Rodríguez-Rosario, T. Frauenheim, M. Paternostro, K. Modi, [Operational markov condition for quantum processes](#), Phys. Rev. Lett. 120 (2018) 040405. doi:10.1103/PhysRevLett.120.040405.
URL <https://link.aps.org/doi/10.1103/PhysRevLett.120.040405>
- [11] C. Giarmatzi, F. Costa, Witnessing quantum memory in non-markovian processes (2018). [arXiv:1811.03722](#).
- [12] K. Goswami, C. Giarmatzi, C. Monterola, S. Shrapnel, J. Romero, F. Costa, [Experimental characterization of a non-markovian quantum process](#), Phys. Rev. A 104 (2021) 022432. [arXiv:2102.01327](#), doi:10.1103/PhysRevA.104.022432.
URL <https://link.aps.org/doi/10.1103/PhysRevA.104.022432>
- [13] K. Goswami, C. Giarmatzi, M. Kewming, F. Costa, C. Branciard, J. Romero, A. White, [Indefinite causal order in a quantum switch](#), Phys. Rev. Lett. 121 (9) (Aug 2018). [arXiv:1803.04302](#).
URL <http://dx.doi.org/10.1103/PhysRevLett.121.090503>
- [14] K. Goswami, Y. Cao, G. A. Paz-Silva, J. Romero, A. G. White, [Increasing communication capacity via superposition of order](#), Physical Review Research 2 (3) (Aug 2020). doi:10.1103/physrevresearch.2.033292.
URL <http://dx.doi.org/10.1103/PhysRevResearch.2.033292>
- [15] K. Goswami, F. Costa, [Classical communication through quantum causal structures](#), Phys. Rev. A 103 (2021) 042606. doi:10.1103/PhysRevA.103.042606.
URL <https://link.aps.org/doi/10.1103/PhysRevA.103.042606>
- [16] D. J. Griffiths, D. F. Schroeter, Introduction to Quantum Mechanics, 3rd Edition, Cambridge University Press, 2018. doi:10.1017/9781316995433.
- [17] C. Gerry, P. Knight, Introductory Quantum Optics, Cambridge University Press, 2004. doi:10.1017/CBO9780511791239.
- [18] A. D. O’Connell, M. Hofheinz, M. Ansmann, R. C. Bialczak, M. Lenander, E. Lucero, M. Neeley, D. Sank, H. Wang, M. Weides, J. Wenner, J. M. Martinis, A. N. Cleland, [Quantum ground state and single-phonon control of a mechanical resonator](#), Nature 464 (2010) 697–703. doi:10.1038/nature08967.
URL <https://www.nature.com/articles/nature08967>
- [19] C. H. Bennett, S. J. Wiesner, [Communication via one- and two-particle operators on Einstein-Podolsky-Rosen states](#), Phys. Rev. Lett. 69 (1992) 2881–2884. doi:10.1103/PhysRevLett.69.2881.
URL <https://link.aps.org/doi/10.1103/PhysRevLett.69.2881>

- [20] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, W. K. Wootters, [Teleporting an unknown quantum state via dual classical and einstein-podolsky-rosen channels](#), *Phys. Rev. Lett.* 70 (1993) 1895–1899. doi:10.1103/PhysRevLett.70.1895.
URL <https://link.aps.org/doi/10.1103/PhysRevLett.70.1895>
- [21] C. H. Bennett, P. W. Shor, J. A. Smolin, A. V. Thapliyal, [Entanglement-assisted classical capacity of noisy quantum channels](#), *Physical Review Letters* 83 (15) (1999) 3081–3084. doi:10.1103/physrevlett.83.3081.
URL <http://dx.doi.org/10.1103/PhysRevLett.83.3081>
- [22] M. Nielsen, I. Chuang, *Quantum Computation and Quantum Information*, Cambridge University Press, 2000.
- [23] A. Einstein, B. Podolsky, N. Rosen, [Can quantum-mechanical description of physical reality be considered complete?](#), *Phys. Rev.* 47 (1935) 777–780. doi:10.1103/PhysRev.47.777.
- [24] J. S. Bell, [On the Einstein-Poldolsky-Rosen paradox](#), *Physics* 1 (3) (1964) 195–200.
- [25] J. S. Bell, [On the problem of hidden variables in quantum mechanics](#), *Rev. Mod. Phys.* 38 (1966) 447–452. doi:10.1103/RevModPhys.38.447.
- [26] A. Aspect, J. Dalibard, G. Roger, [Experimental test of bell’s inequalities using time- varying analyzers](#), *Phys. Rev. Lett.* 49 (25) (1982) 1804–1807. doi:10.1103/PhysRevLett.49.1804.
- [27] B. Hensen, H. Bernien, A. E. Dréau, A. Reiserer, N. Kalb, M. S. Blok, J. Ruitenber, R. F. L. Vermeulen, R. N. Schouten, C. Abellán, W. Amaya, V. Pruneri, M. W. Mitchell, M. Markham, D. J. Twitchen, D. Elkouss, S. Wehner, T. H. Taminiau, R. Hanson, [Loophole-free bell inequality violation using electron spins separated by 1.3 kilometres](#), *Nature* 526 (2015) 682–686.
URL <https://www.nature.com/articles/nature15759>
- [28] E. Schrödinger, [An undulatory theory of the mechanics of atoms and molecules](#), *Phys. Rev.* 28 (1926) 1049–1070. doi:10.1103/PhysRev.28.1049.
URL <https://link.aps.org/doi/10.1103/PhysRev.28.1049>
- [29] W. F. Stinespring, [Positive functions on \$C^*\$ -algebras](#), *Proceedings of the American Mathematical Society* 6 (2) (1955) 211–216.
URL <http://www.jstor.org/stable/2032342>
- [30] W. Heisenberg, [Über den anschaulichen inhalt der quantentheoretischen kinematik und mechanik](#), *Zeitschrift für Physik* 43 (1927) 172–198.
- [31] A. Jamiołkowski, [Linear transformations which preserve trace and positive semidefiniteness of operators](#), *Reports on Mathematical Physics* 3 (4) (1972) 275–278. doi:10.1016/0034-4877(72)90011-0.

- URL <https://linkinghub.elsevier.com/retrieve/pii/0034487772900110>
- [32] M.-D. Choi, **Positive semidefinite biquadratic forms**, *Linear Algebra Appl.* 12 (2) (1975) 95 – 100. doi:[https://doi.org/10.1016/0024-3795\(75\)90058-0](https://doi.org/10.1016/0024-3795(75)90058-0).
URL <http://www.sciencedirect.com/science/article/pii/0024379575900580>
- [33] M. Araújo, C. Branciard, F. Costa, A. Feix, C. Giarmatzi, Č. Brukner, **Witnessing causal nonseparability**, *New J. Phys.* 17 (10) (2015) 102001. doi:[10.1088/1367-2630/17/10/102001](https://doi.org/10.1088/1367-2630/17/10/102001).
URL <https://iopscience.iop.org/article/10.1088/1367-2630/17/10/102001>
- [34] B. Coecke, **Quantum picturalism**, *Contemporary Physics* 51 (1) (2010) 59–83. doi:[10.1080/00107510903257624](https://doi.org/10.1080/00107510903257624).
URL <http://www.tandfonline.com/doi/abs/10.1080/00107510903257624>
- [35] G. Chiribella, G. M. D’Ariano, P. Perinotti, **Optimal cloning of unitary transformation**, *Phys. Rev. Lett.* 101 (2008) 180504. doi:[10.1103/PhysRevLett.101.180504](https://doi.org/10.1103/PhysRevLett.101.180504).
URL <https://link.aps.org/doi/10.1103/PhysRevLett.101.180504>
- [36] M. Hillery, M. Ziman, V. Bužek, **Implementation of quantum maps by programmable quantum processors**, *Phys. Rev. A* 66 (2002) 042302. doi:[10.1103/PhysRevA.66.042302](https://doi.org/10.1103/PhysRevA.66.042302).
URL <https://link.aps.org/doi/10.1103/PhysRevA.66.042302>
- [37] G. Chiribella, G. M. D’Ariano, P. Perinotti, **Memory effects in quantum channel discrimination**, *Phys. Rev. Lett.* 101 (2008) 180501. doi:[10.1103/PhysRevLett.101.180501](https://doi.org/10.1103/PhysRevLett.101.180501).
URL <https://link.aps.org/doi/10.1103/PhysRevLett.101.180501>
- [38] S. Shrapnel, F. Costa, G. Milburn, **Updating the born rule**, *New Journal of Physics* 20 (5) (2018) 053010. doi:[10.1088/1367-2630/aabe12](https://doi.org/10.1088/1367-2630/aabe12).
URL <http://dx.doi.org/10.1088/1367-2630/aabe12>
- [39] C. E. Shannon, **A mathematical theory of communication**, *The Bell System Technical Journal* 27 (3) (1948) 379–423. doi:[10.1002/j.1538-7305.1948.tb01338.x](https://doi.org/10.1002/j.1538-7305.1948.tb01338.x).
- [40] S. Kullback, R. A. Leibler, **On Information and Sufficiency**, *The Annals of Mathematical Statistics* 22 (1) (1951) 79 – 86. doi:[10.1214/aoms/1177729694](https://doi.org/10.1214/aoms/1177729694).
URL <https://doi.org/10.1214/aoms/1177729694>
- [41] R. M. Fano, **Transmission of Information: A Statistical Theory of Communication**, Wiley, New York, 1961.

- [42] L. Brillouin, *Science and Information Theory*, Academic Press, New York, 1962.
- [43] E. T. Jaynes, *Papers on Probability, Statistics and Statistical Physics*, Reidel, Dordrecht, The Netherlands, 1982.
- [44] S. Kullback, *Information Theory and Statistics*, Dover Publications, 1997.
- [45] G. Chaitin, *Algorithmic Information Theory*, Cambridge Tracts in Theoretical Computer Science, 1987.
- [46] C. Adami, [The use of information theory in evolutionary biology](#), *Annals of the New York Academy of Sciences* 1256 (1) (2012) 49–65. doi:10.1111/j.1749-6632.2011.06422.x.
URL <http://dx.doi.org/10.1111/j.1749-6632.2011.06422.x>
- [47] R. Landauer, Irreversibility and heat generation in the computing process, *IBM Journal of Research and Development* 5 (3) (1961) 183–191. doi:10.1147/rd.53.0183.
- [48] A. Bérut, A. Petrosyan, S. Ciliberto, [Information and thermodynamics: experimental verification of landauer's erasure principle](#), *Journal of Statistical Mechanics: Theory and Experiment* 2015 (6) (2015) P06015. doi:10.1088/1742-5468/2015/06/p06015.
URL <http://dx.doi.org/10.1088/1742-5468/2015/06/P06015>
- [49] Y. Jun, M. Gavrilov, J. Bechhoefer, [High-precision test of landauer's principle in a feedback trap](#), *Physical Review Letters* 113 (19) (Nov 2014). doi:10.1103/physrevlett.113.190601.
URL <http://dx.doi.org/10.1103/PhysRevLett.113.190601>
- [50] J. Hong, B. Lambson, S. Dhuey, J. Bokor, [Experimental test of landauer's principle in single-bit operations on nanomagnetic memory bits](#), *Science Advances* 2 (3) (2016). arXiv:<https://advances.sciencemag.org/content/2/3/e1501492.full.pdf>, doi:10.1126/sciadv.1501492.
URL <https://advances.sciencemag.org/content/2/3/e1501492>
- [51] R. Gaudenzi, E. Burzurí, S. Maegawa, F. van der Zant, H. S. J. and Luis, [Quantum landauer erasure with a molecular nanomagnet](#), *Nat. Phys.* 14 (19) (2018) 565–568. doi:<https://doi.org/10.1038/s41567-018-0070-7>.
URL <https://www.nature.com/articles/s41567-018-0070-7#citeas>
- [52] M. M. Wilde, [From classical to quantum shannon theory](#), *Quantum Information Theory* (2013) xi–xii doi:10.1017/9781316809976.001.
URL <http://dx.doi.org/10.1017/9781316809976.001>
- [53] L. Szilard, [über die entropieverminderung in einem thermodynamischen system bei eingriffen intelligenter wesen](#), *Z. Physik* 53 (1929) 840–856. doi:<https://doi.org/10.1007/>

BF01341281.

URL <https://link.springer.com/article/10.1007%2F01341281>

- [54] C. H. Bennett, [Notes on Landauer's principle, reversible computation, and Maxwell's demon](#), *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* 34 (3) (2003) 501–510, quantum Information and Computation. doi:[https://doi.org/10.1016/S1355-2198\(03\)00039-X](https://doi.org/10.1016/S1355-2198(03)00039-X).
URL <https://www.sciencedirect.com/science/article/pii/S135521980300039X>
- [55] M. Tribus, E. McIrvine, [Energy and information](#), *Scientific American* 224 (1971) 179–188. doi:[10.1038/scientificamerican0971-179](https://doi.org/10.1038/scientificamerican0971-179).
URL <http://www.nature.com/scientificamerican/journal/v225/n3/pdf/scientificamerican0971-179.pdf>
- [56] P. Beale, [Statistical Mechanics](#), Elsevier Science, 1996.
URL <https://books.google.com.au/books?id=PIk9sF9j2oUC>
- [57] E. C. G. Sudarshan, [Equivalence of semiclassical and quantum mechanical descriptions of statistical light beams](#), *Phys. Rev. Lett.* 10 (1963) 277–279. doi:[10.1103/PhysRevLett.10.277](https://doi.org/10.1103/PhysRevLett.10.277).
URL <https://link.aps.org/doi/10.1103/PhysRevLett.10.277>
- [58] R. J. Glauber, [Coherent and incoherent states of the radiation field](#), *Phys. Rev.* 131 (1963) 2766–2788. doi:[10.1103/PhysRev.131.2766](https://doi.org/10.1103/PhysRev.131.2766).
URL <https://link.aps.org/doi/10.1103/PhysRev.131.2766>
- [59] A. Holevo, [Bounds for the quantity of information transmitted by a quantum communication channel](#), *Problems Inform. Transmission* 9 (1973) 177.
URL <http://mi.mathnet.ru/ppi903>
- [60] C. W. Helstrom, *Quantum detection and estimation theory*, Academic Press New York, 1976.
- [61] R. P. Feynman, [Simulating physics with computers](#), *Int. J. Theor. Phys.* 21 (1982) 467–488. doi:[10.1007/BF02650179](https://doi.org/10.1007/BF02650179).
URL <https://doi.org/10.1007/BF02650179>
- [62] W. Wootters, W. A. Zurek, [A single quantum cannot be cloned](#), *Nature* 299 (1982) 802–803. doi:<https://doi.org/10.1038/299802a0>.
URL <https://www.nature.com/articles/299802a0#citeas>
- [63] D. Dieks, [Communication by epr devices](#), *Physics Letters A* 92 (6) (1982) 271–272. doi:[https://doi.org/10.1016/0375-9601\(82\)90084-6](https://doi.org/10.1016/0375-9601(82)90084-6).
URL <https://www.sciencedirect.com/science/article/pii/0375960182900846>

- [64] C. H. Bennett, G. Brassard, [Quantum cryptography: Public key distribution and coin tossing](#), *Theoretical Computer Science* 560 (2014) 7–11, theoretical Aspects of Quantum Cryptography – celebrating 30 years of BB84. doi:<https://doi.org/10.1016/j.tcs.2014.05.025>.
URL <https://www.sciencedirect.com/science/article/pii/S0304397514004241>
- [65] A. K. Ekert, [Quantum cryptography based on bell's theorem](#), *Phys. Rev. Lett.* 67 (1991) 661–663. doi:[10.1103/PhysRevLett.67.661](https://doi.org/10.1103/PhysRevLett.67.661).
URL <https://link.aps.org/doi/10.1103/PhysRevLett.67.661>
- [66] C. H. Bennett, F. Bessette, G. Brassard, L. Salvail, J. Smolin, [Quantum cryptography based on bell's theorem](#), *Journal of Cryptology* 5 (1992) 3–28. doi:<https://doi.org/10.1007/BF00191318>.
URL <https://link.springer.com/article/10.1007%2FBF00191318>
- [67] P. Shor, Algorithms for quantum computation: discrete logarithms and factoring, *Foundations of Computer Science, 1994 Proceedings, 35th Annual Symposium on* (1994) 124–134 doi:[10.1109/SFCS.1994.365700](https://doi.org/10.1109/SFCS.1994.365700).
- [68] A. S. Holevo, The capacity of the quantum channel with general signal states, *IEEE Trans. Inf. Theory* 44 (1) (1998) 269–273. arXiv:[quant-ph/9611023](https://arxiv.org/abs/quant-ph/9611023), doi:[10.1109/18.651037](https://doi.org/10.1109/18.651037).
- [69] C. H. Bennett, P. W. Shor, J. A. Smolin, A. V. Thapliyal, [Entanglement-assisted classical capacity of noisy quantum channels](#), *Physical Review Letters* 83 (15) (1999) 3081–3084. doi:[10.1103/physrevlett.83.3081](https://doi.org/10.1103/physrevlett.83.3081).
URL <http://dx.doi.org/10.1103/PhysRevLett.83.3081>
- [70] S. Lloyd, [Capacity of the noisy quantum channel](#), *Physical Review A* 55 (3) (1997) 1613–1622. doi:[10.1103/physreva.55.1613](https://doi.org/10.1103/physreva.55.1613).
URL <http://dx.doi.org/10.1103/PhysRevA.55.1613>
- [71] M. Horodecki, J. Oppenheim, A. Winter, [Quantum state merging and negative information](#), *Communications in Mathematical Physics* 269 (1) (2006) 107–136. doi:[10.1007/s00220-006-0118-x](https://doi.org/10.1007/s00220-006-0118-x).
URL <http://dx.doi.org/10.1007/s00220-006-0118-x>
- [72] M. B. Hastings, [Superadditivity of communication capacity using entangled inputs](#), *Nat. Phys.* 5 (4) (2009) 255–257. arXiv:[0809.3972](https://arxiv.org/abs/0809.3972), doi:[10.1038/nphys1224](https://doi.org/10.1038/nphys1224).
URL <http://dx.doi.org/10.1038/nphys1224>
- [73] I. Devetak, A. W. Harrow, A. J. Winter, [A resource framework for quantum shannon theory](#), *IEEE Transactions on Information Theory* 54 (10) (2008) 4587–4618. doi:[10.1109/tit.](https://doi.org/10.1109/tit.)

2008.928980.

URL <http://dx.doi.org/10.1109/TIT.2008.928980>

- [74] M. Hayashi, K. Iwama, H. Nishimura, R. Raymond, S. Yamashita, **Quantum network coding**, Lecture Notes in Computer Science (2006) 610–621 [doi:10.1007/978-3-540-70918-3_52](https://doi.org/10.1007/978-3-540-70918-3_52).
URL http://dx.doi.org/10.1007/978-3-540-70918-3_52
- [75] A. Peres, **Quantum Theory: Concepts and Methods**, Springer, 1993.
- [76] J. L. O’Brien, A. Furusawa, J. Vučković, **Photonic quantum technologies**, Nature Photonics 3 (12) (2009) 687–695. [doi:10.1038/nphoton.2009.229](https://doi.org/10.1038/nphoton.2009.229).
URL <http://dx.doi.org/10.1038/nphoton.2009.229>
- [77] R. Miller, T. E. Northup, K. M. Birnbaum, A. Boca, A. D. Boozer, H. J. Kimble, **Trapped atoms in cavity QED: coupling quantized light and matter**, Journal of Physics B: Atomic, Molecular and Optical Physics 38 (9) (2005) S551–S565. [doi:10.1088/0953-4075/38/9/007](https://doi.org/10.1088/0953-4075/38/9/007).
URL <https://doi.org/10.1088/0953-4075/38/9/007>
- [78] I. Georgescu, **Trapped ion quantum computing turns 25**, Nat. Rev. Phys. 2 (278) (2020). [doi:https://doi.org/10.1038/s42254-020-0189-1](https://doi.org/10.1038/s42254-020-0189-1).
URL <https://www.nature.com/articles/s42254-020-0189-1#citeas>
- [79] P. Krantz, M. Kjaergaard, F. Yan, T. P. Orlando, S. Gustavsson, W. D. Oliver, **A quantum engineer’s guide to superconducting qubits**, Applied Physics Reviews 6 (2) (2019) 021318. [doi:10.1063/1.5089550](https://doi.org/10.1063/1.5089550).
URL <http://dx.doi.org/10.1063/1.5089550>
- [80] J. A. Jones, **Quantum computing and nuclear magnetic resonance**, PhysChemComm 4 (11) (2001) 49. [doi:10.1039/b103231n](https://doi.org/10.1039/b103231n).
URL <http://dx.doi.org/10.1039/b103231n>
- [81] D. Bohm, A suggested interpretation of the quantum theory in terms of “hidden” variables. i, Phys. Rev. 85 (1952) 166–179. [doi:10.1103/PhysRev.85.166](https://doi.org/10.1103/PhysRev.85.166).
- [82] D. Goldstein, **Polarized Light**, CRC Press, 2017.
URL <https://books.google.com.au/books?id=w6PMBQAAQBAJ>
- [83] R. Simon, N. Mukunda, **Minimal three-component su(2) gadget for polarization optics**, Phys. Lett. A 143 (4) (1990) 165–169. [doi:https://doi.org/10.1016/0375-9601\(90\)90732-4](https://doi.org/10.1016/0375-9601(90)90732-4).
URL <http://www.sciencedirect.com/science/article/pii/0375960190907324>

- [84] A. Forbes, I. Nape, [Quantum mechanics with patterns of light: Progress in high dimensional and multidimensional entanglement with structured light](#), *AVS Quantum Science* 1 (1) (2019) 011701. [arXiv:https://doi.org/10.1116/1.5112027](#), [doi:10.1116/1.5112027](#).
URL [https://doi.org/10.1116/1.5112027](#)
- [85] A. Rossi, G. Vallone, A. Chiuri, F. De Martini, P. Mataloni, [Multipath entanglement of two photons](#), *Phys. Rev. Lett.* 102 (2009) 153902. [doi:10.1103/PhysRevLett.102.153902](#).
URL [https://link.aps.org/doi/10.1103/PhysRevLett.102.153902](#)
- [86] J.-W. Pan, Z.-B. Chen, C.-Y. Lu, H. Weinfurter, A. Zeilinger, M. Żukowski, [Multiphoton entanglement and interferometry](#), *Rev. Mod. Phys.* 84 (2012) 777–838. [doi:10.1103/RevModPhys.84.777](#).
URL [https://link.aps.org/doi/10.1103/RevModPhys.84.777](#)
- [87] C. Reimer, M. Kues, P. Roztock, B. Wetzel, F. Grazioso, B. E. Little, S. T. Chu, T. Johnston, Y. Bromberg, L. Caspani, D. J. Moss, R. Morandotti, [Generation of multiphoton entangled quantum states by means of integrated frequency combs](#), *Science* 351 (6278) (2016) 1176–1180. [arXiv:https://science.sciencemag.org/content/351/6278/1176.full.pdf](#), [doi:10.1126/science.aad8532](#).
URL [https://science.sciencemag.org/content/351/6278/1176](#)
- [88] M. Kues, C. Reimer, P. Roztock, L. R. Cortés, S. Sciara, B. Wetzel, Y. Zhang, A. Cino, S. T. Chu, B. E. Little, D. J. Moss, L. Caspani, J. Azaña, R. Morandotti, [On-chip generation of high-dimensional entangled quantum states and their coherent control](#), *Nature* 546 (7660) (2017) 622–626. [doi:10.1038/nature22986](#).
URL [http://www.nature.com/articles/nature22986](#)
- [89] A. Mair, A. Vaziri, G. Weihs, A. Zeilinger, [Entanglement of the orbital angular momentum states of photons](#), *Nature* 412 (6844) (2001) 313–316. [doi:10.1038/35085529](#).
URL [http://www.nature.com/articles/35085529](#)
- [90] B. Ndagano, I. Nape, M. A. Cox, C. Rosales-Guzman, A. Forbes, [Creation and Detection of Vector Vortex Modes for Classical and Quantum Communication](#), *J. Lightwave Technol.* 36 (2) (2018) 292–301. [doi:10.1109/JLT.2017.2766760](#).
URL [http://ieeexplore.ieee.org/document/8085118/](#)
- [91] M. J. Padgett, [Orbital angular momentum 25 years on \[Invited\]](#), *Opt. Express* 25 (10) (2017) 11265. [doi:10.1364/OE.25.011265](#).
URL [https://www.osapublishing.org/abstract.cfm?URI=oe-25-10-11265](#)
- [92] M. Erhard, R. Fickler, M. Krenn, A. Zeilinger, [Twisted photons: new quantum perspectives in high dimensions](#), *Light Sci Appl* 7 (3) (2018) 17146–17146. [doi:10.1038/lsa.2017.](#)

146.

URL <http://www.nature.com/articles/lisa2017146>

- [93] M. Krenn, M. Malik, M. Erhard, A. Zeilinger, **Orbital angular momentum of photons and the entanglement of Laguerre–Gaussian modes**, *Phil. Trans. R. Soc. A.* 375 (2087) (2017) 20150442. doi:10.1098/rsta.2015.0442.
URL <https://royalsocietypublishing.org/doi/10.1098/rsta.2015.0442>
- [94] G. Molina-Terriza, J. P. Torres, L. Torner, **Twisted photons**, *Nature Phys* 3 (5) (2007) 305–310. doi:10.1038/nphys607.
URL <http://www.nature.com/articles/nphys607>
- [95] J. Pinnell, I. Nape, B. Sephton, M. A. Cox, V. Rodríguez-Fajardo, A. Forbes, **Modal analysis of structured light with spatial light modulators: a practical tutorial**, *J. Opt. Soc. Am. A* 37 (11) (2020) C146. doi:10.1364/JOSAA.398712.
URL <https://www.osapublishing.org/abstract.cfm?URI=josaa-37-11-C146>
- [96] H. Kogelnik, T. Li, **Laser Beams and Resonators**, *Appl. Opt.* 5 (10) (1966) 1550. doi:10.1364/AO.5.001550.
URL <https://www.osapublishing.org/abstract.cfm?URI=ao-5-10-1550>
- [97] M. J. Romero, **Orbital angular momentum entanglement**, Ph.D. Thesis, University of Glasgow (2012).
URL <https://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.564101>
- [98] N. R. Heckenberg, R. McDuff, C. P. Smith, A. G. White, **Generation of optical phase singularities by computer-generated holograms**, *Opt. Lett.* 17 (3) (1992) 221. doi:10.1364/OL.17.000221.
URL <https://www.osapublishing.org/abstract.cfm?URI=ol-17-3-221>
- [99] V. Bazhenov, M. Soskin, M. Vasnetsov, **Screw Dislocations in Light Wavefronts**, *Journal of Modern Optics* 39 (5) (1992) 985–990. doi:10.1080/09500349214551011.
URL <https://www.tandfonline.com/doi/full/10.1080/09500349214551011>
- [100] J. Leach, M. J. Padgett, S. M. Barnett, S. Franke-Arnold, J. Courtial, **Measuring the orbital angular momentum of a single photon**, *Phys. Rev. Lett.* 88 (2002) 257901. doi:10.1103/PhysRevLett.88.257901.
URL <https://link.aps.org/doi/10.1103/PhysRevLett.88.257901>

- [101] E. J. Galvez, P. R. Crawford, H. I. Sztul, M. J. Pysher, P. J. Haglin, R. E. Williams, **Geometric Phase Associated with Mode Transformations of Optical Beams Bearing Orbital Angular Momentum**, *Phys. Rev. Lett.* 90 (20) (2003) 203901. doi:10.1103/PhysRevLett.90.203901.
URL <https://link.aps.org/doi/10.1103/PhysRevLett.90.203901>
- [102] L. Allen, J. Courtial, M. J. Padgett, **Matrix formulation for the propagation of light beams with orbital and spin angular momenta**, *Phys. Rev. E* 60 (6) (1999) 7497–7503. doi:10.1103/PhysRevE.60.7497.
URL <https://link.aps.org/doi/10.1103/PhysRevE.60.7497>
- [103] N. González, G. Molina-Terriza, J. P. Torres, **How a Dove prism transforms the orbital angular momentum of a light beam**, *Opt. Express* 14 (20) (2006) 9093. doi:10.1364/OE.14.009093.
URL <https://www.osapublishing.org/oe/abstract.cfm?uri=oe-14-20-9093>
- [104] M. J. Padgett, J. P. Lesso, **Dove prisms and polarized light**, *Journal of Modern Optics* 46 (2) (1999) 175–179. doi:10.1080/09500349908231263.
URL <http://www.tandfonline.com/doi/abs/10.1080/09500349908231263>
- [105] M. J. Padgett, L. Allen, **The angular momentum of light: optical spanners and the rotational frequency shift**, *Optical and Quantum Electronics* 31 (1) (1999) 1–12. doi:10.1023/A:1006911428303.
URL <http://link.springer.com/10.1023/A:1006911428303>
- [106] A. Babazadeh, M. Erhard, F. Wang, M. Malik, R. Nouroozi, M. Krenn, A. Zeilinger, **High-dimensional single-photon quantum gates: Concepts and experiments**, *Phys. Rev. Lett.* 119 (2017) 180510. doi:10.1103/PhysRevLett.119.180510.
URL <https://link.aps.org/doi/10.1103/PhysRevLett.119.180510>
- [107] F. Schlederer, M. Krenn, R. Fickler, M. Malik, A. Zeilinger, **Cyclic transformation of orbital angular momentum modes**, *New J. Phys.* 18 (4) (2016) 043019. doi:10.1088/1367-2630/18/4/043019.
URL <https://iopscience.iop.org/article/10.1088/1367-2630/18/4/043019>
- [108] H. Rubinsztein-Dunlop, A. Forbes, M. V. Berry, M. R. Dennis, D. L. Andrews, M. Mansuripur, C. Denz, C. Alpmann, P. Banzer, T. Bauer, E. Karimi, L. Marrucci, M. Padgett, M. Ritsch-Marte, N. M. Litchinitser, N. P. Bigelow, C. Rosales-Guzmán, A. Belmonte, J. P. Torres, T. W. Neely, M. Baker, R. Gordon, A. B. Stilgoe, J. Romero, A. G. White, R. Fickler, A. E. Willner, G. Xie, B. McMorrán, A. M. Weiner, **Roadmap on structured light**, *J. Opt.* 19 (1) (2017) 013001.

[doi:10.1088/2040-8978/19/1/013001](https://doi.org/10.1088/2040-8978/19/1/013001).

URL <https://iopscience.iop.org/article/10.1088/2040-8978/19/1/013001>

- [109] D. S. Dahl, J. Romero, M. Ploschner, N. K. Fontaine, J. Carpenter, **High-dimensional quantum gates for azimuthal modes**, in: 14th Pacific Rim Conference on Lasers and Electro-Optics (CLEO PR 2020), OSA, Sydney, 2020, p. P5_2. [doi:10.1364/CLEOPR.2020.P5_2](https://doi.org/10.1364/CLEOPR.2020.P5_2).
URL https://www.osapublishing.org/abstract.cfm?URI=CLEOPR-2020-P5_2
- [110] C. Karlewski, M. Marthaler, **Time-local master equation connecting the Born and Markov approximations**, Phys. Rev. B 90 (10) (2014) 104302. [doi:10.1103/PhysRevB.90.104302](https://doi.org/10.1103/PhysRevB.90.104302).
URL <https://link.aps.org/doi/10.1103/PhysRevB.90.104302>
- [111] H. Carmichael, An open systems approach to quantum optics: lectures presented at the Université libre de Bruxelles, October 28 to November 4, 1991, no. m18 in Lecture notes in physics, Springer-Verlag, Berlin ; New York, 1993.
- [112] M. Schlosshauer, Decoherence and the Quantum-To-Classical Transition, Springer-Verlag, Berlin, Heidelberg, 2007, oCLC: 896254556.
- [113] P. Blanchard (Ed.), Decoherence: theoretical, experimental, and conceptual problems: proceedings of a workshop held at Bielefeld, Germany, 10-14 November 1998, no. 538 in Lecture notes in physics, Springer, Berlin ; New York, 2000.
- [114] R. Alicki, K. Lendi, **Quantum dynamical semigroups and applications**, Springer-Verlag, Berlin; New York, 2007, oCLC: 228402594.
URL <http://www.myilibrary.com?id=81696>
- [115] V. Gorini, A. Kossakowski, E. Sudarshan, **Completely positive dynamical semigroups of N-level systems**, J. Math. Phys. 17 (5) (1976) 821. [doi:10.1063/1.522979](https://doi.org/10.1063/1.522979).
URL <http://scitation.aip.org/content/aip/journal/jmp/17/5/10.1063/1.522979>
- [116] G. Lindblad, **On the generators of quantum dynamical semigroups**, Commun.Math. Phys. 48 (2) (1976) 119–130. [doi:10.1007/BF01608499](https://doi.org/10.1007/BF01608499).
URL <http://link.springer.com/10.1007/BF01608499>
- [117] S. Milz, M. S. Kim, F. A. Pollock, K. Modi, **Completely Positive Divisibility Does Not Mean Markovianity**, Phys. Rev. Lett. 123 (4) (2019) 040401, arXiv: 1901.05223. [doi:10.1103/PhysRevLett.123.040401](https://doi.org/10.1103/PhysRevLett.123.040401).
URL <http://arxiv.org/abs/1901.05223>

- [118] P. Pechukas, [Reduced Dynamics Need Not Be Completely Positive](#), Phys. Rev. Lett. 73 (8) (1994) 1060–1062. doi:10.1103/PhysRevLett.73.1060.
URL <https://link.aps.org/doi/10.1103/PhysRevLett.73.1060>
- [119] R. Alicki, [Comment on “Reduced Dynamics Need Not Be Completely Positive”](#), Phys. Rev. Lett. 75 (16) (1995) 3020–3020. doi:10.1103/PhysRevLett.75.3020.
URL <https://link.aps.org/doi/10.1103/PhysRevLett.75.3020>
- [120] P. Pechukas, [Pechukas Replies:](#), Phys. Rev. Lett. 75 (16) (1995) 3021–3021. doi:10.1103/PhysRevLett.75.3021.
URL <https://link.aps.org/doi/10.1103/PhysRevLett.75.3021>
- [121] S. Weinberg, [Precision Tests of Quantum Mechanics](#), Phys. Rev. Lett. 62 (5) (1989) 485–488. doi:10.1103/PhysRevLett.62.485.
URL <https://link.aps.org/doi/10.1103/PhysRevLett.62.485>
- [122] T. F. Jordan, [Why quantum dynamics is linear](#), J. Phys.: Conf. Ser. 196 (2009) 012010. doi:10.1088/1742-6596/196/1/012010.
URL <https://iopscience.iop.org/article/10.1088/1742-6596/196/1/012010>
- [123] C. A. Rodríguez-Rosario, K. Modi, A.-m. Kuah, A. Shaji, E. C. G. Sudarshan, [Completely positive maps and classical correlations](#), J. Phys. A: Math. Theor. 41 (20) (2008) 205301. doi:10.1088/1751-8113/41/20/205301.
URL <https://iopscience.iop.org/article/10.1088/1751-8113/41/20/205301>
- [124] A. Shabani, D. A. Lidar, [Vanishing Quantum Discord is Necessary and Sufficient for Completely Positive Maps](#), Phys. Rev. Lett. 102 (10) (2009) 100402. doi:10.1103/PhysRevLett.102.100402.
URL <https://link.aps.org/doi/10.1103/PhysRevLett.102.100402>
- [125] A. Brodutch, A. Datta, K. Modi, Á. Rivas, C. A. Rodríguez-Rosario, [Vanishing quantum discord is not necessary for completely positive maps](#), Phys. Rev. A 87 (4) (2013) 042301. doi:10.1103/PhysRevA.87.042301.
URL <https://link.aps.org/doi/10.1103/PhysRevA.87.042301>
- [126] A. Shabani, D. A. Lidar, [Erratum: Vanishing Quantum Discord is Necessary and Sufficient for Completely Positive Maps \[Phys. Rev. Lett. 102, 100402 \(2009\)\]](#), Phys. Rev. Lett. 116 (4) (2016) 049901. doi:10.1103/PhysRevLett.116.049901.
URL <https://link.aps.org/doi/10.1103/PhysRevLett.116.049901>
- [127] A. Shaji, E. Sudarshan, [Who’s afraid of not completely positive maps?](#), Physics Letters A 341 (1-4) (2005) 48–54. doi:10.1016/j.physleta.2005.04.029.

URL <https://linkinghub.elsevier.com/retrieve/pii/S0375960105005748>

- [128] C. J. Wood, **Non-completely positive maps: properties and applications**, arXiv:0911.3199 [quant-ph]ArXiv: 0911.3199 (Nov. 2009).
URL <http://arxiv.org/abs/0911.3199>
- [129] M. Ringbauer, C. J. Wood, K. Modi, A. Gilchrist, A. G. White, A. Fedrizzi, **Characterizing Quantum Dynamics with Initial System-Environment Correlations**, Phys. Rev. Lett. 114 (9) (2015) 090402. doi:10.1103/PhysRevLett.114.090402.
URL <https://link.aps.org/doi/10.1103/PhysRevLett.114.090402>
- [130] J. Morris, F. A. Pollock, K. Modi, **Non-markovian memory in ibmqx4** (2019). arXiv: 1902.07980.
- [131] G. A. L. White, C. D. Hill, F. A. Pollock, L. C. L. Hollenberg, K. Modi, **Demonstration of non-markovian process characterisation and control on a quantum processor**, Nat.Commun. 11 (1) (Dec 2020). doi:10.1038/s41467-020-20113-3.
URL <http://dx.doi.org/10.1038/s41467-020-20113-3>
- [132] L. Li, M. J. Hall, H. M. Wiseman, **Concepts of quantum non-markovianity: A hierarchy**, Phys. Rep. 759 (2018) 1–51. doi:10.1016/j.physrep.2018.07.001.
URL <http://dx.doi.org/10.1016/j.physrep.2018.07.001>
- [133] J. Emerson, R. Alicki, K. Życzkowski, **Scalable noise estimation with random unitary operators**, J. Opt. B: Quantum Semiclassical Opt. 7 (10) (2005) S347–S352. doi:10.1088/1464-4266/7/10/021.
URL <http://dx.doi.org/10.1088/1464-4266/7/10/021>
- [134] E. Knill, D. Leibfried, R. Reichle, J. Britton, R. B. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, D. J. Wineland, **Randomized benchmarking of quantum gates**, Phys. Rev. A 77 (2008) 012307. doi:10.1103/PhysRevA.77.012307.
URL <https://link.aps.org/doi/10.1103/PhysRevA.77.012307>
- [135] J. M. Epstein, A. W. Cross, E. Magesan, J. M. Gambetta, **Investigating the limits of randomized benchmarking protocols**, Phys. Rev. A 89 (2014) 062321. doi:10.1103/PhysRevA.89.062321.
URL <https://link.aps.org/doi/10.1103/PhysRevA.89.062321>
- [136] H. Ball, T. M. Stace, S. T. Flammia, M. J. Biercuk, **Effect of noise correlations on randomized benchmarking**, Phys. Rev. A 93 (2016) 022303. doi:10.1103/PhysRevA.93.022303.
URL <https://link.aps.org/doi/10.1103/PhysRevA.93.022303>

- [137] F. A. Pollock, C. Rodríguez-Rosario, T. Frauenheim, M. Paternostro, K. Modi, [Operational markov condition for quantum processes](#), Phys. Rev. Lett. 120 (2018) 040405. doi:10.1103/PhysRevLett.120.040405.
URL <https://link.aps.org/doi/10.1103/PhysRevLett.120.040405>
- [138] F. Costa, S. Shrapnel, Quantum causal modelling, New J. Phys. 18 (6) (2016) 063032. doi: <https://doi.org/10.1088/1367-2630/18/6/063032>.
- [139] F. A. Pollock, C. Rodríguez-Rosario, T. Frauenheim, M. Paternostro, K. Modi, [Non-markovian quantum processes: Complete framework and efficient characterization](#), Phys. Rev. A 97 (2018) 012127. doi:10.1103/PhysRevA.97.012127.
URL <https://link.aps.org/doi/10.1103/PhysRevA.97.012127>
- [140] S. Shrapnel, F. Costa, G. Milburn, Quantum markovianity as a supervised learning task, International Journal of Quantum Information 16 (08) (2018) 1840010. doi:10.1142/s0219749918400105.
- [141] M. Y. Niu, V. Smelyanskiy, P. Klimov, S. Boixo, R. Barends, J. Kelly, Y. Chen, K. Arya, B. Burkett, D. Bacon, Z. Chen, B. Chiaro, R. Collins, A. Dunsworth, B. Foxen, A. Fowler, C. Gidney, M. Giustina, R. Graff, T. Huang, E. Jeffrey, D. Landhuis, E. Lucero, A. Megrant, J. Mutus, X. Mi, O. Naaman, M. Neeley, C. Neill, C. Quintana, P. Roushan, J. M. Martinis, H. Neven, Learning non-markovian quantum noise from moiré-enhanced swap spectroscopy with deep evolutionary algorithm (2019). arXiv:1912.04368.
- [142] I. A. Luchnikov, S. V. Vintskevich, H. Ouerdane, S. N. Filippov, [Simulation complexity of open quantum dynamics: Connection with tensor networks](#), Phys. Rev. Lett. 122 (2019) 160401. doi:10.1103/PhysRevLett.122.160401.
URL <https://link.aps.org/doi/10.1103/PhysRevLett.122.160401>
- [143] I. A. Luchnikov, S. V. Vintskevich, D. A. Grigoriev, S. N. Filippov, [Machine learning non-markovian quantum dynamics](#), Phys. Rev. Lett. 124 (2020) 140502. doi:10.1103/PhysRevLett.124.140502.
URL <https://link.aps.org/doi/10.1103/PhysRevLett.124.140502>
- [144] F. F. Fanchini, G. Karpat, D. Z. Rossatto, A. Norambuena, R. Coto, Estimating the degree of non-markovianity using machine learning (2020). arXiv:2009.03946.
- [145] C. Guo, K. Modi, D. Poletti, [Tensor-network-based machine learning of non-markovian quantum processes](#), Phys. Rev. A 102 (6) (Dec 2020).
URL <http://dx.doi.org/10.1103/PhysRevA.102.062414>
- [146] B.-H. Liu, L. Li, Y.-F. Huang, C.-F. Li, G.-C. Guo, E.-M. Laine, H.-P. Breuer, J. Piilo, [Experimental control of the transition from markovian to non-markovian dynamics of open quantum](#)

- [systems](#), Nat. Phys. 7 (12) (2011) 931–934. doi:10.1038/nphys2085.
URL <https://doi.org/10.1038/nphys2085>
- [147] A. Chiuri, C. Greganti, L. Mazzola, M. Paternostro, P. Mataloni, [Linear optics simulation of quantum non-markovian dynamics](#), Sci. Rep. 2 (1) (2012) 968. doi:10.1038/srep00968.
URL <https://doi.org/10.1038/srep00968>
- [148] S. Yu, Y.-T. Wang, Z.-J. Ke, W. Liu, Y. Meng, Z.-P. Li, W.-H. Zhang, G. Chen, J.-S. Tang, C.-F. Li, G.-C. Guo, [Experimental investigation of spectra of dynamical maps and their relation to non-markovianity](#), Phys. Rev. Lett. 120 (2018) 060406. doi:10.1103/PhysRevLett.120.060406.
URL <https://link.aps.org/doi/10.1103/PhysRevLett.120.060406>
- [149] S.-J. Xiong, Q. Hu, Z. Sun, L. Yu, Q. Su, J.-M. Liu, C.-P. Yang, [Non-markovianity in experimentally simulated quantum channels: Role of counterrotating-wave terms](#), Phys. Rev. A 100 (2019) 032101. doi:10.1103/PhysRevA.100.032101.
URL <https://link.aps.org/doi/10.1103/PhysRevA.100.032101>
- [150] K.-D. Wu, Z. Hou, G.-Y. Xiang, C.-F. Li, G.-C. Guo, D. Dong, F. Nori, [Detecting non-markovianity via quantified coherence: theory and experiments](#), npj Quantum Information 6 (1) (2020) 55. doi:10.1038/s41534-020-0283-3.
URL <https://doi.org/10.1038/s41534-020-0283-3>
- [151] T. d. L. Silva, S. P. Walborn, M. F. Santos, G. H. Aguilar, A. A. Budini, [Detection of quantum non-markovianity close to the born-markov approximation](#), Phys. Rev. A 101 (2020) 042120. doi:10.1103/PhysRevA.101.042120.
URL <https://link.aps.org/doi/10.1103/PhysRevA.101.042120>
- [152] Á. Rivas, S. F. Huelga, M. B. Plenio, [Quantum non-markovianity: characterization, quantification and detection](#), Rep. Prog. Phys. 77 (9) (2014) 094001. doi:10.1088/0034-4885/77/9/094001.
URL <https://doi.org/10.1088%2F0034-4885%2F77%2F9%2F094001>
- [153] P. Štelmachovič, V. Bužek, [Dynamics of open quantum systems initially entangled with environment: Beyond the kraus representation](#), Phys. Rev. A 64 (2001) 062106. doi:10.1103/PhysRevA.64.062106.
URL <https://link.aps.org/doi/10.1103/PhysRevA.64.062106>
- [154] O. Oreshkov, C. Giarmatzi, [Causal and causally separable processes](#), New J. Phys. 18 (9) (2016) 093020. arXiv:1506.05449, doi:10.1088/1367-2630/18/9/093020.
- [155] G. Lindblad, [Non-markovian quantum stochastic processes and their entropy](#), Comm. Math. Phys. 65 (3) (1979) 281–294.
URL <https://projecteuclid.org:443/euclid.cmp/1103904877>

- [156] L. Accardi, A. Frigerio, J. T. Lewis, Quantum stochastic processes, *Publications of the Research Institute for Mathematical Sciences* 18 (1) (1982) 97–133. doi:[10.2977/prims/1195184017](https://doi.org/10.2977/prims/1195184017).
- [157] E. Davies, J. Lewis, An operational approach to quantum probability, *Comm. Math. Phys.* 17 (3) (1970) 239–260. doi:[10.1007/BF01647093](https://doi.org/10.1007/BF01647093).
- [158] M.-D. Choi, Positive semidefinite biquadratic forms, *Linear Algebra Appl.* 12 (2) (1975) 95–100. doi:[10.1016/0024-3795\(75\)90058-0](https://doi.org/10.1016/0024-3795(75)90058-0).
- [159] D. Kretschmann, R. F. Werner, Quantum channels with memory, *Phys. Rev. A* 72 (2005) 062323. doi:[10.1103/PhysRevA.72.062323](https://doi.org/10.1103/PhysRevA.72.062323).
- [160] F. A. Pollock, C. Rodríguez-Rosario, T. Frauenheim, M. Paternostro, K. Modi, **Operational markov condition for quantum processes**, *Phys. Rev. Lett.* 120 (4) (2018) 040405–. doi:[10.1103/PhysRevLett.120.040405](https://doi.org/10.1103/PhysRevLett.120.040405).
URL <https://link.aps.org/doi/10.1103/PhysRevLett.120.040405>
- [161] C. Giarmatzi, F. Costa, **A quantum causal discovery algorithm**, *npj Quantum Information* 4 (1) (2018) 17. doi:[10.1038/s41534-018-0062-6](https://doi.org/10.1038/s41534-018-0062-6).
URL <https://doi.org/10.1038/s41534-018-0062-6>
- [162] T. Hastie, R. Tibshirani, J. Friedman, **The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition**, Springer Series in Statistics, Springer New York, 2009.
URL <https://books.google.com.au/books?id=tVIjmNS3Ob8C>
- [163] C. Bishop, **Pattern Recognition and Machine Learning**, Information science and statistics, Springer (India) Private Limited, 2013.
URL <https://books.google.com.au/books?id=HL4HrgEACAAJ>
- [164] M. Stone, **Cross-validatory choice and assessment of statistical predictions**, *Journal of the Royal Statistical Society. Series B (Methodological)* 36 (2) (1974) 111–147.
URL <http://www.jstor.org/stable/2984809>
- [165] A. Erhard, J. J. Wallman, L. Postler, M. Meth, R. Stricker, E. A. Martinez, P. Schindler, T. Monz, J. Emerson, R. Blatt, **Characterizing large-scale quantum computers via cycle benchmarking**, *Nat. Commun.* 10 (1) (Nov 2019).
URL <http://dx.doi.org/10.1038/s41467-019-13068-7>
- [166] J. M. Gambetta, A. D. Córcoles, S. T. Merkel, B. R. Johnson, J. A. Smolin, J. M. Chow, C. A. Ryan, C. Rigetti, S. Poletto, T. A. Ohki, M. B. Ketchen, M. Steffen, **Characterization of addressability by simultaneous randomized benchmarking**, *Phys. Rev. Lett.* 109 (2012) 240504. doi:[10.1103/PhysRevLett.109.240504](https://doi.org/10.1103/PhysRevLett.109.240504).
URL <https://link.aps.org/doi/10.1103/PhysRevLett.109.240504>

- [167] J. P. Clemens, S. Siddiqui, J. Gea-Banacloche, **Quantum error correction against correlated noise**, Phys. Rev. A 69 (2004) 062313. doi:10.1103/PhysRevA.69.062313.
URL <https://link.aps.org/doi/10.1103/PhysRevA.69.062313>
- [168] C. Lupo, L. Memarzadeh, S. Mancini, **Removing correlations in signals transmitted over a quantum memory channel**, Phys. Rev. A 85 (2012) 012320. doi:10.1103/PhysRevA.85.012320.
URL <https://link.aps.org/doi/10.1103/PhysRevA.85.012320>
- [169] B. Karimi, J. P. Pekola, **Correlated versus uncorrelated noise acting on a quantum refrigerator**, Phys. Rev. B 96 (2017) 115408. doi:10.1103/PhysRevB.96.115408.
URL <https://link.aps.org/doi/10.1103/PhysRevB.96.115408>
- [170] M. A. Mohd Izhar, Z. Babar, H. V. Nguyen, P. Botsinis, D. Alanis, D. Chandra, S. X. Ng, L. Hanzo, Quantum turbo decoding for quantum channels exhibiting memory, IEEE Access 6 (2018) 12369–12381. doi:10.1109/ACCESS.2018.2808373.
- [171] L. Hardy, Towards quantum gravity: a framework for probabilistic theories with non-fixed causal structure, J. Phys. A: Math. Gen. 40 (2007) 3081–3099. arXiv:gr-qc/0608043, doi:10.1088/1751-8113/40/12/S12.
- [172] M. Zych, F. Costa, I. Pikovski, Č. Brukner, **Bell's theorem for temporal order**, Nat. Commun. 10 (1) (Aug 2019). arXiv:1708.00248.
URL <http://dx.doi.org/10.1038/s41467-019-11579-x>
- [173] M. Araújo, F. Costa, Č. Brukner, Computational Advantage from Quantum-Controlled Ordering of Gates, Phys. Rev. Lett. 113 (25) (2014) 250402. arXiv:1401.8127, doi:10.1103/PhysRevLett.113.250402.
- [174] A. Feix, M. Araújo, Č. Brukner, Quantum superposition of the order of parties as a communication resource, Phys. Rev. A 92 (2015) 052326. doi:10.1103/PhysRevA.92.052326.
- [175] P. A. Guérin, A. Feix, M. Araújo, Č. Brukner, **Exponential communication complexity advantage from quantum superposition of the direction of communication**, Phys. Rev. Lett. 117 (2016) 100502. arXiv:1605.07372, doi:10.1103/PhysRevLett.117.100502.
URL <https://link.aps.org/doi/10.1103/PhysRevLett.117.100502>
- [176] K. Wei, N. Tischler, S.-R. Zhao, Y.-H. Li, J. M. Arrazola, Y. Liu, W. Zhang, H. Li, L. You, Z. Wang, Y.-A. Chen, B. C. Sanders, Q. Zhang, G. J. Pryde, F. Xu, J.-W. Pan, **Experimental quantum switching for exponentially superior quantum communication complexity**, Phys. Rev. Lett. 122 (12) (Mar 2019). arXiv:1810.10238.
URL <http://dx.doi.org/10.1103/PhysRevLett.122.120504>

- [177] M. M. Taddei, J. Cariñe, D. Martínez, T. García, N. Guerrero, A. A. Abbott, M. Araújo, C. Branciard, E. S. Gómez, S. P. Walborn, L. Aolita, G. Lima, Experimental computational advantage from superposition of multiple temporal orders of quantum gates (2020). [arXiv:2002.07817](#).
- [178] G. Chiribella, Perfect discrimination of no-signalling channels via quantum superposition of causal structures, *Phys. Rev. A* 86 (4) (2012) 040301. [arXiv:1109.5154](#), [doi:10.1103/PhysRevA.86.040301](#).
- [179] T. Colnaghi, G. M. D'Ariano, S. Facchini, P. Perinotti, Quantum computation with programmable connections between gates, *Phys. Lett. A* 376 (2012) 2940–2943. [arXiv:1109.5987](#), [doi:10.1016/j.physleta.2012.08.028](#).
- [180] D. Ebler, S. Salek, G. Chiribella, [Enhanced communication with the assistance of indefinite causal order](#), *Phys. Rev. Lett.* 120 (2018) 120502. [arXiv:1711.10165](#), [doi:10.1103/PhysRevLett.120.120502](#).
URL <https://link.aps.org/doi/10.1103/PhysRevLett.120.120502>
- [181] M. Araújo, C. Branciard, F. Costa, A. Feix, C. Giarmatzi, Č. Brukner, Witnessing causal nonseparability, *New J. Phys.* 17 (10) (2015) 102001. [arXiv:1506.03776](#), [doi:10.1088/1367-2630/17/10/102001](#).
- [182] K. Goswami, J. Romero, [Experiments on quantum causality](#), *AVS Quantum Science* 2 (3) (2020) 037101. [doi:10.1116/5.0010747](#).
URL <http://dx.doi.org/10.1116/5.0010747>
- [183] L. M. Procopio, A. Moqanaki, M. Araújo, F. Costa, I. Alonso Calafell, E. G. Dowd, D. R. Hamel, L. A. Rozema, Č. Brukner, P. Walther, [Experimental superposition of orders of quantum gates](#), *Nature Communications* 6 (2015) 7913.
URL <http://dx.doi.org/10.1038/ncomms8913><http://10.0.4.14/ncomms8913><https://www.nature.com/articles/ncomms8913#supplementary-information>
- [184] G. Rubino, L. A. Rozema, A. Feix, M. Araújo, J. M. Zeuner, L. M. Procopio, Č. Brukner, P. Walther, [Experimental verification of an indefinite causal order](#), *Sci. Adv.* 3 (3) (2017) e1602589. [arXiv:1608.01683](#), [doi:10.1126/sciadv.1602589](#).
URL <http://dx.doi.org/10.1126/sciadv.1602589>
- [185] J. Wechs, A. A. Abbott, C. Branciard, [On the definition and characterisation of multipartite causal \(non\)separability](#), *New J. Phys.* 21 (1) (2019) 013027. [doi:10.1088/1367-2630/aaf352](#).
URL <https://iopscience.iop.org/article/10.1088/1367-2630/aaf352>

- [186] C. Branciard, Witnesses of causal nonseparability: an introduction and a few case studies, *Sci. Rep.* 6 (2016) 26018.
- [187] S. Boyd, L. Vandenberghe, [Convex Optimization](#), Cambridge University Press, 2004.
URL <http://www.stanford.edu/~boyd/cvxbook/>
- [188] J. Romero, D. Giovannini, M. G. McLaren, E. J. Galvez, A. Forbes, M. J. Padgett, Orbital angular momentum correlations with a phase-flipped gaussian mode pump beam, *J. of Opt.* 14 (8) (2012) 085401.
- [189] J. Leach, J. Courtial, K. Skeldon, S. M. Barnett, S. Franke-Arnold, M. J. Padgett, Interferometric methods to measure orbital and spin, or the total angular momentum of a single photon, *Phys. Rev. Lett.* 92 (1) (2004) 013601.
- [190] C. Tamm, C. Weiss, Bistability and optical switching of spatial patterns in a laser 7 (1990) 1034–1038. doi:10.1364/JOSAB.7.001034.
- [191] M. W. Beijersbergen, L. Allen, H. V. der Veen, J. P. Woerdman, Astigmatic laser mode converters and transfer of orbital angular momentum, *Opt. Commun.* 96 (1-3) (1993) 123–132.
- [192] D. L. Sullivan, Alignment of rotational prisms, *Appl. Opt.* 11 (9) (1972) 2028–2032.
- [193] M. J. Padgett, J. P. Lesso, Dove prisms and polarized light, *J. of Mod. Opt.* 46 (2) (1999) 175–179.
- [194] G. Rubino, L. A. Rozema, F. Massa, M. Araújo, M. Zych, Č. Brukner, P. Walther, Experimental entanglement of temporal orders (2017). arXiv:1712.06884.
- [195] A. Fedrizzi, T. Herbst, A. Poppe, T. Jennewein, A. Zeilinger, [A wavelength-tunable fiber-coupled source of narrowband entangled photons](#), *Opt. Express* 15 (23) (2007) 15377–15386. doi:10.1364/OE.15.015377.
URL <http://www.opticsexpress.org/abstract.cfm?URI=oe-15-23-15377>
- [196] M. Erhard, R. Fickler, M. Krenn, A. Zeilinger, Twisted photons: new quantum perspectives in high dimensions, *Light: Science & Applications* 7 (3) (2018) 17146.
- [197] N. J. Cerf, M. Bourennane, A. Karlsson, N. Gisin, Security of quantum key distribution using d-level systems, *Phys. Rev. Lett.* 88 (12) (2002) 127902.
- [198] A. A. Abbott, J. Wechs, D. Horsman, M. Mhalla, C. Branciard, [Communication through coherent control of quantum channels](#), *Quantum* 4 (2020) 333, arXiv: 1810.09826. doi:10.22331/q-2020-09-24-333.
URL <http://arxiv.org/abs/1810.09826>

- [199] N. Gisin, N. Linden, S. Massar, S. Popescu, [Error filtration and entanglement purification for quantum communication](#), *Phys. Rev. A* 72 (2005) 012338. doi:10.1103/PhysRevA.72.012338.
URL <https://link.aps.org/doi/10.1103/PhysRevA.72.012338>
- [200] G. Chiribella, H. Kristjánsson, Quantum Shannon theory with superpositions of trajectories, *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 475 (2225) (2019) 20180903. doi:10.1098/rspa.2018.0903.
- [201] S. Salek, D. Ebler, G. Chiribella, Quantum communication in a superposition of causal orders (2018). [arXiv:1809.06655](https://arxiv.org/abs/1809.06655).
- [202] G. Chiribella, M. Banik, S. S. Bhattacharya, T. Guha, M. Alimuddin, A. Roy, S. Saha, S. Agrawal, G. Kar, [Indefinite causal order enables perfect quantum communication with zero capacity channels](#), *New Journal of Physics* 23 (3) (2021) 033039. doi:10.1088/1367-2630/abe7a0.
URL <http://dx.doi.org/10.1088/1367-2630/abe7a0>
- [203] A. A. Abbott, C. Giarmatzi, F. Costa, C. Branciard, Multipartite causal correlations: Polytopes and inequalities, *Phys. Rev. A* 94 (2016) 032131. doi:10.1103/PhysRevA.94.032131.
- [204] Y. Guo, X.-M. Hu, Z.-B. Hou, H. Cao, J.-M. Cui, B.-H. Liu, Y.-F. Huang, C.-F. Li, G.-C. Guo, G. Chiribella, [Experimental transmission of quantum information using a superposition of causal orders](#), *Physical Review Letters* 124 (3) (Jan 2020). doi:10.1103/physrevlett.124.030502.
URL <http://dx.doi.org/10.1103/PhysRevLett.124.030502>
- [205] A. A. Abbott, J. Wechs, D. Horsman, M. Mhalla, C. Branciard, [Communication through coherent control of quantum channels](#), *Quantum* 4 (2020) 333. doi:10.22331/q-2020-09-24-333.
URL <https://quantum-journal.org/papers/q-2020-09-24-333/>
- [206] P. A. Guérin, M. Krumm, C. Budroni, Č. Brukner, [Composition rules for quantum processes: a no-go theorem](#), *New J. Phys.* 21 (1) (2019) 012001. [arXiv:1806.10374](https://arxiv.org/abs/1806.10374), doi:10.1088/1367-2630/aafef7.
URL <http://dx.doi.org/10.1088/1367-2630/aafef7>
- [207] H. Kristjánsson, G. Chiribella, S. Salek, D. Ebler, M. Wilson, [Resource theories of communication](#), *New Journal of Physics* 22 (7) (2020) 073014. doi:10.1088/1367-2630/ab8ef7.
URL <http://dx.doi.org/10.1088/1367-2630/ab8ef7>
- [208] W. A. de Brito, R. V. Ramos, [Quantum information technology with Sagnac interferometer: interaction-free measurement, quantum key distribution and quantum secret sharing](#), *Journal of*

- Modern Optics 55 (8) (2008) 1231–1241. doi:10.1080/09500340701633113.
URL <https://doi.org/10.1080/09500340701633113>
- [209] C. H. Bennet, P. Shor, [Quantum information theory](#), IEEE Trans. Inf. Theory 44 (6) (1998) 2724–2742.
URL <https://ieeexplore.ieee.org/document/720553>
- [210] B. Schumacher, M. D. Westmoreland, [Sending classical information via noisy quantum channels](#), Phys. Rev. A 56 (1997) 131–138. doi:10.1103/PhysRevA.56.131.
URL <https://link.aps.org/doi/10.1103/PhysRevA.56.131>
- [211] L. Lamport, [Time, clocks, and the ordering of events in a distributed system](#), Commun. ACM 21 (7) (1978) 558–565. doi:10.1145/359545.359563.
URL <https://doi.org/10.1145/359545.359563>
- [212] L. Hardy, [Quantum gravity computers: On the theory of computation with indefinite causal structure](#), Quantum Reality, Relativistic Causality, and Closing the Epistemic Circle (2009) 379–401 arXiv:quant-ph/0701019, doi:10.1007/978-1-4020-9107-0_21.
URL http://dx.doi.org/10.1007/978-1-4020-9107-0_21
- [213] M. Araújo, P. A. Guérin, Ä. Baumeler, [Quantum computation with indefinite causal structures](#), Phys. Rev. A 96 (5) (2017). arXiv:1706.09854.
URL <http://dx.doi.org/10.1103/PhysRevA.96.052315>
- [214] D. Ebler, S. Salek, G. Chiribella, [Enhanced communication with the assistance of indefinite causal order](#), Phys. Rev. Lett. 120 (2018) 120502. arXiv:1711.10165, doi:10.1103/PhysRevLett.120.120502.
URL <https://link.aps.org/doi/10.1103/PhysRevLett.120.120502>
- [215] D. Jia, F. Costa, [Causal order as a resource for quantum communication](#), Phys. Rev. A 100 (2019) 052319. arXiv:1901.09159.
URL <https://link.aps.org/doi/10.1103/PhysRevA.100.052319>
- [216] M. Caleffi, A. S. Cacciapuoti, [Quantum switch for the quantum internet: Noiseless communications through noisy channels](#), IEEE J Sel. Area Comm. 38 (3) (2020) 575–588. arXiv:1907.07432, doi:10.1109/jsac.2020.2969035.
URL <http://dx.doi.org/10.1109/JSAC.2020.2969035>
- [217] L. M. Procopio, A. Moqanaki, M. Araújo, F. Costa, I. Alonso Calafell, E. G. Dowd, D. R. Hamel, L. A. Rozema, Č. Brukner, P. Walther, [Experimental superposition of orders of quantum gates](#), Nat. Commun. 6 (1) (Aug 2015). arXiv:1412.4006.
URL <http://dx.doi.org/10.1038/ncomms8913>

- [218] Y. Guo, X.-M. Hu, Z.-B. Hou, H. Cao, J.-M. Cui, B.-H. Liu, Y.-F. Huang, C.-F. Li, G.-C. Guo, G. Chiribella, [Experimental transmission of quantum information using a superposition of causal orders](#), *Phys. Rev. Lett.* 124 (3) (Jan 2020). [arXiv:1811.07526](#).
URL <http://dx.doi.org/10.1103/PhysRevLett.124.030502>
- [219] C. Branciard, M. Araújo, A. Feix, F. Costa, Č. Brukner, [The simplest causal inequalities and their violation](#), *New J. Phys.* 18 (1) (2016) 013008. [arXiv:1508.01704](#), [doi:10.1088/1367-2630/18/1/013008](#).
- [220] A. A. Abbott, J. Wechs, F. Costa, C. Branciard, [Genuinely multipartite noncausality](#), *Quantum* 1 (2017) 39. [doi:10.22331/q-2017-12-14-39](#).
- [221] D. Jia, N. Sakharwade, [Tensor products of process matrices with indefinite causal structure](#), *Phys. Rev. A* 97 (3) (Mar 2018). [arXiv:1706.05532](#).
URL <http://dx.doi.org/10.1103/PhysRevA.97.032110>
- [222] L. Gyongyosi, S. Imre, H. V. Nguyen, [A survey on quantum channel capacities](#), *IEEE Commun. Surv.* 20 (2) (2018) 1149–1205. [arXiv:1801.02019](#), [doi:10.1109/COMST.2017.2786748](#).
- [223] M. Araújo, A. Feix, M. Navascués, Č. Brukner, [A purification postulate for quantum mechanics with indefinite causal order](#), *Quantum* 1 (2017) 10. [arXiv:1611.08535](#), [doi:10.22331/q-2017-04-26-10](#).
- [224] C. Brukner, [Bounding quantum correlations with indefinite causal order](#), *New Journal of Physics* 17 (8) (2015) 083034. [doi:10.1088/1367-2630/17/8/083034](#).
URL <https://doi.org/10.1088/1367-2630/17/8/083034>
- [225] G. Chiribella, D. Ebler, [Optimal quantum networks and one-shot entropies](#), *New Journal of Physics* 18 (9) (2016) 093053. [doi:10.1088/1367-2630/18/9/093053](#).
URL <https://doi.org/10.1088/1367-2630/18/9/093053>
- [226] P. Perinotti, [Causal structures and the classification of higher order quantum computations](#), *Tutorials, Schools, and Workshops in the Mathematical Sciences* (2017) 103–127 [arXiv:1612.05099](#), [doi:10.1007/978-3-319-68655-4_7](#).
URL http://dx.doi.org/10.1007/978-3-319-68655-4_7
- [227] A. Kissinger, S. Uijlen, [A categorical semantics for causal structure](#), *IEEE Symp. Logic Comput. Sci. (LICS)* 15 (Aug. 2019). [arXiv:1701.04732](#).
URL <https://lmcs.episciences.org/5681>
- [228] D. Jia, F. Costa, [Causal order as a resource for quantum communication](#), *Phys. Rev. A* 100 (5) (Nov 2019). [arXiv:1901.09159](#).
URL <http://dx.doi.org/10.1103/PhysRevA.100.052319>

- [229] S. Verdú, Te Sun Han, [A general formula for channel capacity](#), IEEE Trans. Inf. Theory 40 (4) (1994) 1147–1157.
URL <https://ieeexplore.ieee.org/document/335960>
- [230] M. Hayashi, H. Nagaoka, [A general formula for the classical capacity of a general quantum channel](#), Proc. IEEE Int. Symp. Inf. Theory (2002). [arXiv:quant-ph/0206186](#).
URL <http://dx.doi.org/10.1109/ISIT.2002.1023343>
- [231] T. Morimae, [The process matrix framework for a single-party system](#) (2014). [arXiv:1408.1464](#).
- [232] N. Miklin, A. A. Abbott, C. Branciard, R. Chaves, C. Budroni, [The entropic approach to causal correlations](#), New J. Phys. 19 (11) (2017) 113041. [arXiv:1706.10270](#), [doi:10.1088/1367-2630/aa8f9f](#).
URL <http://dx.doi.org/10.1088/1367-2630/aa8f9f>
- [233] J. Barrett, R. Lorenz, O. Oreshkov, [Cyclic quantum causal models](#) (2020). [arXiv:2002.12157](#).
- [234] W. Yokojima, M. T. Quintino, A. Soeda, M. Murao, [Consequences of preserving reversibility in quantum superchannels](#) (2020). [arXiv:2003.05682](#).
- [235] C. Mukhopadhyay, M. K. Gupta, A. K. Pati, [Superposition of causal order as a metrological resource for quantum thermometry](#) (2018). [arXiv:1812.07508](#).
- [236] T. Guha, M. Alimuddin, P. Parashar, [Thermodynamic advancement in the causally inseparable occurrence of thermal maps](#), Phys. Rev. A 102 (2020) 032215. [doi:10.1103/PhysRevA.102.032215](#).
URL <https://link.aps.org/doi/10.1103/PhysRevA.102.032215>
- [237] D. Felce, V. Vedral, [Quantum refrigeration with indefinite causal order](#), Phys. Rev. Lett. 125 (2020) 070603. [doi:10.1103/PhysRevLett.125.070603](#).
URL <https://link.aps.org/doi/10.1103/PhysRevLett.125.070603>
- [238] F. Giacomini, E. Castro-Ruiz, Č. Brukner, [Indefinite causal structures for continuous-variable systems](#), New J. Phys. 18 (11) (2016) 113026. [doi:10.1088/1367-2630/18/11/113026](#).
URL <https://iopscience.iop.org/article/10.1088/1367-2630/18/11/113026>
- [239] J. Foo, S. Onoe, M. Zych, [Unruh-dewitt detectors in quantum superpositions of trajectories](#), Phys. Rev. D 102 (2020) 085013. [doi:10.1103/PhysRevD.102.085013](#).
URL <https://link.aps.org/doi/10.1103/PhysRevD.102.085013>

- [240] L. C. Barbado, E. Castro-Ruiz, L. Apadula, i. c. v. Brukner, [Unruh effect for detectors in superposition of accelerations](#), Phys. Rev. D 102 (2020) 045002. doi:10.1103/PhysRevD.102.045002.
URL <https://link.aps.org/doi/10.1103/PhysRevD.102.045002>