acta physica slovaca vol. 61 No. 3, 273 – 390 **June 2011**

QUANTUM NETWORKS: GENERAL THEORY AND APPLICATIONS[1]

A. Bisio[2]*, G. Chiribella†, G. M. D'Ariano*, P. Perinotti*

**Quit group, Dipartimento di Fisica "A. Volta", via Bassi 6, 27100 Pavia, Italy*

†Perimeter Institute for Theoretical Physics, 31 Caroline St. North, Waterloo, Ontario N2L 2Y5, Canada

Received 29 August 2011, accepted 30 August 2011

In this work we present a general mathematical framework to deal with *Quantum Networks*, i.e. networks resulting from the interconnection of elementary quantum circuits. The cornerstone of our approach is a generalization of the Choi isomorphism that allows one to efficiently represent any given Quantum Network in terms of a single positive operator. Our formalism allows one to face and solve many quantum information processing problems that would be hardly manageable otherwise, the most relevant of which are reviewed in this work: quantum process tomography, quantum cloning and learning of transformations, inversion of a unitary gate, information-disturbance tradeoff in estimating a unitary transformation, cloning and learning of a measurement device.

DOI: 10.2478/v10155-011-0003-9

PACS: 03.65.-w, 03.65.Fd, 03.67.Ac

KEYWORDS: Quantum Information Processing, Quantum Circuits, Quantum Networks, Quantum Tomography, Quantum Cloning, Quantum Learning, Unitary Channels, Group theory in Quantum Mechanics

Contents

1 Introduction **276**

2 Quantum Networks: general theory **278**

2.1 Linear maps and linear operators 278

2.1.1 Choi isomorphism 279

2.1.2 The link product 282

2.2 Diagrammatic representation of linear maps 283

2.3 States, Channels and POVMs 284

[1] The material in this article was presented as a PhD thesis of A. Bisio at the University of Pavia. The work was conducted under the supervision of professor G. M. D'Ariano

[2] E-mail address: alessandro.bisio@unipv.it

2.4 Quantum Networks: constructive approach 287
2.5 Deterministic Quantum Networks 291
2.6 Probabilistic Quantum Network 295
2.7 Connection of Quantum Networks 298

3 Quantum Tomography 303
3.1 State tomography 303

4 Quantum Network Tomography 308
4.1 Optimal quantum tomography for states, effects and transformation 309
4.1.1 Optimization of data processing 309
4.1.2 Optimization of the setup 311
4.1.3 Realization scheme for the optimal tomography 317

5 Cloning a Unitary Transformation 319
5.1 Optimal cloning of a Unitary transformation 321
5.2 The optimal cloning network 325

6 Quantum learning of a unitary transformation 326
6.1 Optimization of quantum learning 329
6.1.1 Considered scenario: $M = 1$ 330
6.1.2 Optimization of the storing strategy 332
6.1.3 Optimization of the retrieving channel 334
6.1.4 Generalization to the M > 1 case 339
6.1.5 Optimal learning according to the single-copy fidelity 339
6.1.6 Optimal learning according to the global fidelity 339
6.2 Comparison with the cloning 340

7 Inversion of a unitary transformation 341
7.1 Learning scenario 341
7.2 Supermap scenario 343

8 Information-disturbance tradeoff in estimating a unitary transformation 345
8.1 Optimization of the tradeoff 346
8.2 Realization scheme for the optimal network 349

9 Learning and cloning of a measurement device 352
9.1 Formulation of the problem 353
9.2 Symmetries of the replicating network 355
9.3 Optimal learning 358
9.3.1 1 → 1 case 359
9.3.2 2 → 1 case 360
9.3.3 3 → 1 case 366
9.3.4 1 → 2 case 368
9.4 Optimal cloning 370

10 Conclusion 374

Acknowledgment 374

A Channel Fidelity 375

B Elements of Group Representation Theory 377

B.1 Basic definitions 377

B.2 Schur lemma and its applications 379

B.3 Relevant decompositions 380

B.3.1 The symmetric group $\mathbf{S}_n$ 381

B.3.2 Decomposition of $\mathbf{SU}(d)^{\otimes n}$ 381

B.3.3 $\mathbf{U} \otimes \mathbf{U}$ 383

B.3.4 $\mathbf{U} \otimes \mathbf{U} \otimes \mathbf{U}$ 384

B.3.5 $\mathbf{U} \otimes \mathbf{U}^*$ 385

B.3.6 $\mathbf{U} \otimes \mathbf{U} \otimes \mathbf{U}^*$ 386

B.3.7 $\mathbf{U} \otimes \mathbf{U} \otimes \mathbf{U} \otimes \mathbf{U}$ (2-dimensional case) 387

References 388

1 Introduction

In standard textbook Quantum Mechanics every physical system corresponds to a Hilbert space. The states of a system are unit rays in the corresponding Hilbert space, transformations of closed systems are described by unitary operators acting on the states, and measurements correspond to complete sets of orthogonal projectors, each projector corresponding to a measurement outcome. Born's statistical formula provides the outcome probability as the expectation of the corresponding projector in the state of the system. This formalism can be generalized to the case of open systems by including the environment in the dynamical description. It is then possible to describe any phenomenon in quantum mechanics in terms of unitary transformations and von Neumann or Lüders measurements. Despite this fact, a convenient formalism for the field of Quantum Information [1, 2] is rather provided by the notions of statistical operator, channel, and positive operator valued measure (POVM). One of the advantages in using such tools is that they provide an effective description of physical devices avoiding a detailed account of their implementation in terms of unitary interactions and von Neumann measurements. This concise description is extremely useful when dealing with optimization problems, like state estimation [3], where one can looks for the optimal measurement among all those allowed by quantum mechanics.

A recent trend in Quantum Information is to consider transformations, rather than states, as carriers of information e.g. in gate discrimination [4, 5, 6, 7, 8], programming [9], teleportation [10, 11, 12] and tomography [13, 14], along with multi-round quantum games [15], standard quantum algorithms [16, 17, 18] and cryptographic protocols [19, 20, 21]. This new perspective requires an appropriate description not only of state processing, but more generally of transformation processing. Such processing is obtained through more general physical devices—what we call *Quantum Networks*—that are made of composition of elementary circuits. A Quantum Network can be used to perform a huge variety of different tasks like transformations of states into channels, channels into channels, and even sequences of states/channels into channels. However, describing a large quantum network in terms of channels and POVMs is very inefficient. Indeed, if one needs to optimize a quantum network for some task, one is forced to carry out a cumbersome elementwise optimization. For this reason, having new notions that generalize those of channels and POVMs is crucial. Luckily enough, a general treatment of Quantum Networks on the same footing as states, channels, and POVMs is possible, both for the deterministic and the probabilistic case.

In this paper we review the aforementioned unified framework along with some of its most relevant applications. Our approach is based on a generalization of the Choi isomorphism that allows us to represent any Quantum Network in terms of a suitably normalized positive operator. This general theory is reviewed in Chapter 2, where we also provide some basic results of linear algebra that are needed in order to prove most results of the general theory. Following the exposition of Refs. [22, 23], we will introduce the notion of Quantum Network from a constructive point of view that consists in looking at networks as a result of composition of elementary circuits. We will begin by considering deterministic Quantum Networks, and then we will extend the results to the probabilistic case.

The Chapters from 3 to 9 are devoted to the applications of the developed formalism. The first application that we consider is the optimization of Quantum Tomography, where we will derive the optimal networks for tomographing states, transformations and measurements. The material of this Chapter was published in Refs [24, 25].

In Chapter 5, based on Ref. [26], we discuss the concept of *quantum cloning of a transformation*. While cloning of quantum states has been subject of many works, cloning of a transformation was never treated before. In particular, a general no-cloning theorem for transformations and the derivation of the optimal cloning network for a unitary transformation are shown.

Quantum Learning of a transformation is another task that is possible to analyse within the new theory of Quantum Networks. Suppose that a user is provided with N uses of an undisclosed transformation $\mathcal{T}$ today, and he needs to reproduce the same transformation on an unknown state provided tomorrow. The most general strategy the user can follow, is to exploit the N uses of $\mathcal{T}$ into a Quantum Network today, in order to store the transformation on a quantum memory. Tomorrow, the user will use the quantum memory to program a retrieving channel that reproduces $\mathcal{T}$. In Chapter 6 we will review Ref. [27], in which the optimal learning network for a unitary transformation is derived. The most relevant result here is that the optimal storing of a unitary can be achieved by making use only of a classical memory.

The optimal inversion of a unitary transformation is the subject of Chapter 7. We will derive the optimal network that realizes this task considering two different scenarios and we will prove that the ultimate performances in the inversion of a unitary are achieved by an estimate and prepare strategy. These results were published in Refs [27,28].

In Chapter 8, based on Ref. [29], we consider the tradeoff between information and disturbance in estimating a unitary transformation. We suppose that we have a black box implementing an unknown unitary transformation, with the restriction that the On the one hand, we may try to identify the unknown unitary on the other hand, we may want to use the black box on a variable input state. Since the two tasks are in general incompatible, there is a trade-off between the amount of information that can be extracted about a black box and the disturbance caused on its action: we cannot estimate an unknown quantum dynamics without perturbing it. In Chapter 8 we find the optimal scheme that introduces the minimum amount of disturbance for any given amount of extracted information.

The last application we consider regards Quantum Networks that replicate measurements. We will study the problem of optimal learning and cloning of von Neumann measurements. In particular we will show how the optimal learning from 3 uses can be achieved only by a sequential strategy. These results are the subject of Refs. [30,31], and are presented in Chapter 9.

Two appendices close this work: in the first one we introduce the notion of channel fidelity [32], which is frequently used in the applications and in the second one we review some basic results from group representation theory, with special emphasis on the decomposition of tensor product representations.

2 Quantum Networks: general theory

In this chapter we expose the general theory of Quantum Networks that was developed in [22, 23, 28]. We will start the presentation with some preliminary results of linear algebra, with emphasis on the Choi isomorphism. This theorem will allow us to represent quantum networks in terms of positive operators which are subject to a normalization constraint. A key point of the formalism is the notion of link product of operators that translates the physical link between quantum networks into the mathematical language.

After this fully mathematical section we recall some basic notions of ordinary quantum mechanics (states, quantum operations and POVMs) that we will use as a testbed for the mathematical tools previously introduced.

Section 2.4 is is focused on the definition of Quantum Network as a set of linear maps linked together; in the following sections the Choi representation of Quantum Networks is introduced first for the deterministic case and then for the probabilistic case.

In the final section the link product of operators will be used to express the link of quantum Networks.

2.1 Linear maps and linear operators

Let us start with some notational remarks: we denote as $\mathcal{L}(\mathcal{H})$ the set of linear operators A on $\mathcal{H}$ while $\mathcal{L}(\mathcal{H}_a, \mathcal{H}_b)$ denotes linear transformation from $\mathcal{H}_a$ to $\mathcal{H}_b$. The dimension of space $\mathcal{H}_a$ is denoted by d_a. We denote as $\mathcal{L}(\mathcal{L}(\mathcal{H}_a), \mathcal{L}(\mathcal{H}_b))$ the set of linear maps $\mathcal{M}$ from $\mathcal{L}(\mathcal{H}_a)$ to $\mathcal{L}(\mathcal{H}_b)$. Given a map $\mathcal{M} \in \mathcal{L}(\mathcal{L}(\mathcal{H}_a), \mathcal{L}(\mathcal{H}_b))$ we refer to $\mathcal{L}(\mathcal{H}_a)$ as the *input space* of $\mathcal{M}$ while $\mathcal{L}(\mathcal{H}_b)$ is called the *output space*. We make use of the following notation:

- $\mathsf{Supp}(A)$ denotes the support of A and $\mathrm{Rng}(A)$ denotes the range of A;
- ^{T} denotes transposition and * denotes complex conjugation;[3]
- A^{-1} denotes the inverse of an operator $A \in \mathcal{L}(\mathcal{H})$; if $\mathsf{Supp}(A)$ is not the whole $\mathcal{H}$, then A^{-1} will denote the inverse on its support.[4]

Within this presentation (unless explicitly mentioned) the Hilbert spaces are assumed to be finite dimensional. In order to avoid confusion when the number of Hilbert spaces proliferates we adopt this convention:

- $\mathcal{H}_{ab\ldots n} := \mathcal{H}_a \otimes \mathcal{H}_b \otimes \cdots \otimes \mathcal{H}_n$ where $a, b, \ldots, n$ are integer numbers;
- $A_{ab\ldots n}$ means $A \in \mathcal{L}(\mathcal{H}_{ab\ldots n})$;
- $|n\rangle_a$ means $|n\rangle \in \mathcal{H}_a$;
- Tr_a denotes partial trace over $\mathcal{H}_a$;
- ^{T_a} denotes partial transposition over $\mathcal{H}_a$.

[3] Both transposition and complex conjugation are meant with respect to a fixed orthonormal basis.

[4] More precisely A^{-1} denotes the Moore-Penrose generalized inverse.

Given an operator $A \in \mathcal{L}(\mathcal{H}_a)$ and a Hilbert space $\mathcal{H}_{a'}$ isomorphic to $\mathcal{H}_a$ $\mathcal{H}_{a'} \cong \mathcal{H}_a$, it is possible to define

$$A_{a'} := T_{a\to a'} A_a T_{a\to a'} \tag{2.1}$$

where $T_{a\to a'} = \sum_k |k\rangle_{a'} \langle k|_a$ and $\{|k\rangle_a\}$, $\{|k\rangle_{a'}\}$ are orthonormal bases for $\mathcal{H}_a$ and $\mathcal{H}_{a'}$ respectively. The above procedure is implicit whenever we make a change of label $A_{ab\ldots n} \to A_{a'b'\ldots n'}$

It is possible to define the following isomorphism between $\mathcal{L}(\mathcal{H}_b, \mathcal{H}_a)$ and $\mathcal{H}_a \otimes \mathcal{H}_b$

$$A = \sum_{n,m} \langle n| A |m\rangle |n\rangle \langle m| \leftrightarrow |A\rangle\rangle = \sum_{n,m} \langle n| A |m\rangle |n\rangle |m\rangle \tag{2.2}$$

where $\{|m\rangle\}(\{|n\rangle\})$ is a fixed orthonormal basis in $\mathcal{H}_b(\mathcal{H}_a)$. In the following we implicitly choose such a basis in every Hilbert space. The double-ket notation $|A\rangle\rangle$ is used to stress that the vector lives in a tensor product of Hilbert spaces (from a quantum mechanical perspective, $|A\rangle\rangle$ is proportional to a pure bipartite state) We will use the notation $|A\rangle\rangle_{ab}$ with the meaning $|A\rangle\rangle \in \mathcal{H}_{ab} = \mathcal{H}_a \otimes \mathcal{H}_b$.

By making use of Eq. (2.2) it is possible to prove that the following identities hold

$$A \otimes B|C\rangle\rangle = |ACB^T\rangle\rangle \tag{2.3}$$

$$A \in \mathcal{L}(\mathcal{H}_a, \mathcal{H}_c), \quad B \in \mathcal{L}(\mathcal{H}_b, \mathcal{H}_d), \quad C \in \mathcal{L}(\mathcal{H}_b, \mathcal{H}_a).$$

$$\mathrm{Tr}_b[|A\rangle\rangle\langle\langle A|_{ab}] = AA^\dagger \qquad \mathrm{Tr}_a[|A\rangle\rangle\langle\langle A|_{ab}] = A^T A^* \tag{2.4}$$

$$\mathrm{Tr}_a[A_{ab}(|I\rangle\rangle\langle\langle I|_{ac})] = A_{bc}^{T_c} \tag{2.5}$$

$$\langle\langle I|_{ac} A_{ab} |I\rangle\rangle_{ac} = \mathrm{Tr}_a[A_{ab}] \tag{2.6}$$

$$\text{for } d_a \leqslant d_c, \quad |I\rangle\rangle_{ac} = \sum_{n=1}^{d_a} |n\rangle_a |n\rangle_c$$

$$|I\rangle\rangle_{abcd} = |I\rangle\rangle_{ac} |I\rangle\rangle_{bd} \tag{2.7}$$

$$(\langle\langle I|_{ac} \otimes I_{bd})|A\rangle\rangle_{abcd} = (\langle\langle I|_{ac} \otimes I_{bd})(A_{ab} \otimes I_{cd})|I\rangle\rangle_{abcd} = |\,\mathrm{Tr}_a[A]\rangle\rangle_{bd} \tag{2.8}$$

Through this isomorphism it is possible to translate the inner product in $\mathcal{H} \otimes \mathcal{H}$ into the Hilbert-Schmidt product in $\mathcal{L}(\mathcal{H})$

$$\langle\langle A|B\rangle\rangle = \mathrm{Tr}[A^\dagger B] \tag{2.9}$$

2.1.1 Choi isomorphism

The following theorem, which is a generalization of the one in Refs. [33, 34, 35], introduces an isomorphism between linear maps and linear operators which is a a foundation stone of the theory of Quantum Networks.

Theorem 2.1 (Choi isomorphism) *Consider the map $\mathfrak{C} : \mathcal{L}(\mathcal{L}(\mathcal{H}_0), \mathcal{L}(\mathcal{H}_1)) \to \mathcal{L}(\mathcal{H}_0 \otimes \mathcal{H}_1)$ defined as*

$$\mathfrak{C} : \mathcal{M} \mapsto M_{10} \qquad M_{10} := \mathcal{M} \otimes \mathcal{I}_0(|I\rangle\rangle\langle\langle I|_{00}) \tag{2.10}$$

where $\mathcal{I}_0$ is the identity map on $\mathcal{L}(\mathcal{H}_0)$. Then $\mathfrak{C}$ defines an isomorphism between $\mathcal{L}(\mathcal{L}(\mathcal{H}_0)$, $\mathcal{L}(\mathcal{H}_1))$ and $\mathcal{L}(\mathcal{H}_0 \otimes \mathcal{H}_1)$. The operator $M = \mathfrak{C}(\mathcal{M})$ is called the Choi operator of $\mathcal{M}$.

Proof. To prove the thesis we will provide an explicit expression for the inverse map $\mathfrak{C}^{-1}$: $\mathcal{L}(\mathcal{H}_0 \otimes \mathcal{H}_1) \to \mathcal{L}(\mathcal{L}(\mathcal{H}_0), \mathcal{L}(\mathcal{H}_1))$. Let us define

$$[\mathfrak{C}^{-1}(M)](X) = \mathrm{Tr}_0[(I_1 \otimes (X_0)^T)M_{10}]; \tag{2.11}$$

it is easy to verify that $[\mathfrak{C}^{-1}(M)](X) = \mathcal{M}(X)$.

Suppose $X_0 = |i\rangle\,\langle j|_0$; we have

$$\begin{aligned} \mathrm{Tr}_0[(I_1 \otimes (|i\rangle\,\langle j|_0)^T)M_{10}] &= \mathrm{Tr}_0[(I_1 \otimes (|j\rangle\,\langle i|_0)\mathcal{M} \otimes \mathcal{I}_0(|I\rangle\!\rangle\langle\!\langle I|_{00})] = \\ &= \langle i|_0\,(\mathcal{M} \otimes \mathcal{I}_0(|I\rangle\!\rangle\langle\!\langle I|_{00}))\,|j\rangle_0 = \\ &= \langle i|_0\,(\sum_{m,n} \mathcal{M}(|n\rangle\,\langle m|_0)) \otimes |n\rangle\,\langle m|_0)\,|j\rangle_0 = \\ &= \mathcal{M}(|i\rangle\,\langle j|_0). \end{aligned} \tag{2.12}$$

From $[\mathfrak{C}^{-1}(M)](|i\rangle\,\langle j|) = \mathcal{M}(|i\rangle\,\langle j|)$ for any $|i\rangle\,\langle j|$ it follows $[\mathfrak{C}^{-1}(M)](X) = \mathcal{M}(X)$ for any X by linearity. ■

Corollary 2.1 (Operator-sum representation) *Let $\mathcal{M}$ be in $\mathcal{L}(\mathcal{L}(\mathcal{H}_0), \mathcal{L}(\mathcal{H}_1))$; then there exist $\{A_i | A_i \in \mathcal{L}(\mathcal{H}_0, \mathcal{H}_1)\}$ and $\{B_i | B_i \in \mathcal{L}(\mathcal{H}_0, \mathcal{H}_1)\}$ such that*

$$\mathcal{M}(X) = \sum_i A_i X B_i^\dagger \qquad \mathrm{Tr}[A_i^\dagger A_j] = \lambda_i \delta_{ij} \qquad \mathrm{Tr}[B_i^\dagger B_j] = \mu_i \delta_{ij}. \tag{2.13}$$

$\lambda_i, \mu_i \in \mathbb{R}$ and δ_{ij} is the Kronecker delta.

Proof. Exploiting Th. 2.1 we can write the action of $\mathcal{M}$ as $\mathcal{M}(X) = \mathrm{Tr}_0[(I_1 \otimes (X_0)^T)M_{10}]$ where M_{10} is the Choi operator of $\mathcal{M}$. Now consider the singular value decomposition of M_{10}, $M = \sum_i |A_i\rangle\!\rangle\langle\!\langle B_i|_{01}$, $\langle\!\langle A_i|A_j\rangle\!\rangle = \mathrm{Tr}[A_i^\dagger A_j] = \lambda_i\delta_{ij}$, $\langle\!\langle B_i|B_j\rangle\!\rangle = \mathrm{Tr}[B_i^\dagger B_j] = \mu_i\delta_{ij}$; If we insert this decomposition into Eq. (2.11) we have

$$\begin{aligned} \mathcal{M}(X) &= \mathrm{Tr}_0[(I_1 \otimes (X_0)^T) \sum_i |A_i\rangle\!\rangle\langle\!\langle B_i|] = \\ &= \sum_i \mathrm{Tr}_0[(I_1 \otimes (X_0)^T)(I_1 \otimes (A_i)^T)|I\rangle\!\rangle\langle\!\langle B_i|] = \\ &= \sum_i \mathrm{Tr}_0[(I_1 \otimes (A_i X)_0^T)|I\rangle\!\rangle\langle\!\langle B_i|] = \sum_i \mathrm{Tr}_0[|I\rangle\!\rangle\langle\!\langle B_i|I_1 \otimes (A_i X)_0^T] = \\ &= \sum_i \mathrm{Tr}_0[|I\rangle\!\rangle\langle\!\langle B_i X^\dagger A_i^\dagger|] = \sum_i \mathrm{Tr}_0[|I\rangle\!\rangle\langle\!\langle I|(A_i X B_i^\dagger)_1 \otimes I_0] \\ &= \sum_i A_i X B_i^\dagger \end{aligned} \tag{2.14}$$

where we used Eq. (2.2) and the cyclic property of the trace. ■

Th. 2.1 provides an extremely powerful representation of linear maps between operator spaces in terms of just one linear operator acting on a bigger Hilbert space. The following results will show how the properties of a linear map $\mathcal{M}$ translates into the properties of its Choi operator.

Lemma 2.1 (Trace preserving condition) *Let $\mathcal{M}$ be in $\mathcal{L}(\mathcal{L}(\mathcal{H}_0), \mathcal{L}(\mathcal{H}_1))$ and $M \in \mathcal{L}(\mathcal{H}_0 \otimes \mathcal{H}_1)$ be its Choi operator. Then we have*

$$\mathrm{Tr}[\mathcal{M}(X)] = \mathrm{Tr}[X] \quad \forall X \in \mathcal{L}(\mathcal{H}_0) \;\Leftrightarrow\; \mathrm{Tr}_1[M_{01}] = I_0 \tag{2.15}$$

Proof. If we insert Eq. (2.11) into Eq.(2.15) we get

$$\begin{aligned}\mathrm{Tr}[\mathcal{M}(X)] &= \mathrm{Tr}[(I_1 \otimes (X)^T)M_{10}] = \mathrm{Tr}_0[(X)^T\,\mathrm{Tr}_1[M_{01}]] = \\ &= \mathrm{Tr}[X] = \mathrm{Tr}[X^T] \qquad \forall X \in \mathcal{L}(\mathcal{H}_0).\end{aligned} \tag{2.16}$$

Since $\mathrm{Tr}[AB] = \mathrm{Tr}[A] \;\; \forall A \Leftrightarrow B = I$ we have that Eq. (2.16) holds if and only if $\mathrm{Tr}_1[M_{01}] = I_0$. ■

Lemma 2.2 (Hermitian preserving condition) *Let $\mathcal{M}$ be in $\mathcal{L}(\mathcal{L}(\mathcal{H}_0), \mathcal{L}(\mathcal{H}_1))$ and $M \in \mathcal{L}(\mathcal{H}_1 \otimes \mathcal{H}_1)$ be its Choi operator. Then we have*

$$\mathcal{M}(X)^\dagger = \mathcal{M}(X^\dagger) \;\Leftrightarrow\; M_{01}^\dagger = M_{01} \tag{2.17}$$

Proof. If we take the adjoint in Eq. (2.11) we have

$$\mathcal{M}(X)^\dagger = \mathrm{Tr}_0[(X)^{\dagger T} M_{01}^\dagger] = \tag{2.18}$$

If $M^\dagger = M$ clearly we have $\mathcal{M}(X)^\dagger = \mathcal{M}(X^\dagger)$. On the other hand if $\mathrm{Tr}_0[(X)^{\dagger T} M_{01}^\dagger] = \mathrm{Tr}_0[(X)^{\dagger T} M_{01}]$ for all X then $[\mathfrak{C}^{-1}(M_{01}^\dagger)](X) = [\mathfrak{C}^{-1}(M_{01})](X)$ for all X and so $\mathfrak{C}^{-1}(M_{01}) = \mathfrak{C}^{-1}(M_{01}^\dagger)$ that implies $M = M^\dagger$ ■

Lemma 2.3 (Completely-positive condition) *Let $\mathcal{M}$ be in $\mathcal{L}(\mathcal{L}(\mathcal{H}_0), \mathcal{L}(\mathcal{H}_1))$ and $M \in \mathcal{L}(\mathcal{H}_1 \otimes \mathcal{H}_1)$ be its Choi operator. Then we have*

$$\mathcal{M} \otimes \mathcal{I}_2(X) \geqslant 0 \quad \forall X \in \mathcal{L}(\mathcal{H}_0 \otimes \mathcal{H}_2) \;\Leftrightarrow\; M_{01} \geqslant 0 \tag{2.19}$$

Where $\mathcal{H}_2$ is an Hilbert space of arbitrary dimension. A linear map that satisfies condition (2.19) is called completely positive (CP).

Proof. If $\mathcal{M} \otimes \mathcal{I}_2(X) \geqslant 0$ for all $X \in \mathcal{L}(\mathcal{H}_0 \otimes \mathcal{H}_2)$ then clearly $\mathcal{M} \otimes \mathcal{I}_0(|I\rangle\rangle\langle\langle I|) = M \geqslant 0$. On the other hand, suppose that $M \geqslant 0$. Then M can be diagonalized in this way $M = \sum_i |A_i\rangle\rangle\langle\langle A_i|$ and the operator-sum representation of $\mathcal{M}$ becomes

$$\mathcal{M}(X) = \sum_i A_i X A_i^\dagger. \tag{2.20}$$

If we introduce an auxiliary Hilbert space $\mathcal{H}_2$ we have

$$\mathcal{M} \otimes \mathcal{I}_2(X) = \sum_i (A_i \otimes I_2) X (A_i^\dagger \otimes I_2) \geqslant 0 \Leftrightarrow X \geqslant 0 \tag{2.21}$$

The operator-sum decomposition in Eq. (2.20) is called canonical Kraus form. ■

2.1.2 The link product

Given two linear maps $\mathcal{M} \in \mathcal{L}(\mathcal{L}(\mathcal{H}_0), \mathcal{L}(\mathcal{H}_1))$ and $\mathcal{N} \in \mathcal{L}(\mathcal{L}(\mathcal{H}_1), \mathcal{L}(\mathcal{H}_2))$ it is possible to consider the composition

$$\mathcal{C} := \mathcal{N} \circ \mathcal{M} : \mathcal{L}(\mathcal{H}_0) \to \mathcal{L}(\mathcal{H}_2) \qquad \mathcal{N} \circ \mathcal{M}(X) := \mathcal{N}(\mathcal{M}(X)) \quad \forall X \in \mathcal{L}(\mathcal{H}_0).$$

Since we can represent $\mathcal{M}$ and $\mathcal{N}$ with the corresponding Choi operators M and N, it is reasonable to ask how the Choi operator C of the composition $\mathcal{C}$ can be expressed in terms of M and N. Consider the action of $\mathcal{C}$ on an operator $X \in \mathcal{L}(\mathcal{H}_0)$

$$\begin{aligned}\mathcal{C}(X) &= \mathrm{Tr}_1[(I_2 \otimes \mathrm{Tr}_0[(I_1 \otimes X_0^T)M]^T)N] = \\ &= \mathrm{Tr}_0[(I_2 \otimes X_0^T)\,\mathrm{Tr}_1[(I_2 \otimes M_{01}^{T_1})(I_0 \otimes N_{12})]];\end{aligned} \tag{2.22}$$

if we compare Eq. 2.22 with Eq. 2.11 we get

$$\mathfrak{C}(\mathcal{C}) = \mathrm{Tr}_1[(I_2 \otimes M_{01}^{T_1})(I_0 \otimes N_{12})] = N * M. \tag{2.23}$$

where we introduced the notation $N * M$ for the expression $\mathrm{Tr}_1[(I_2 \otimes M_{01}^{T_1})(I_0 \otimes N_{12})]$.

If we consider maps such that their input and output spaces are tensor product of Hilbert spaces it is possible to compose these maps only through some of these spaces. For example if we have $\mathcal{M} \in \mathcal{L}(\mathcal{L}(\mathcal{H}_0 \otimes \mathcal{H}_2), \mathcal{L}(\mathcal{H}_1 \otimes \mathcal{H}_3))$ and $\mathcal{N} \in \mathcal{L}(\mathcal{L}(\mathcal{H}_3 \otimes \mathcal{H}_5), \mathcal{L}(\mathcal{H}_4 \otimes \mathcal{H}_6))$ it is possible to define the composition

$$\mathcal{N} \star \mathcal{M} := (\mathcal{N} \otimes \mathcal{I}_1) \circ (\mathcal{M} \otimes \mathcal{I}_5). \tag{2.24}$$

Following the same steps as before we have that

$$\begin{aligned}\mathcal{M} \in \mathcal{L}(\mathcal{L}(\mathcal{H}_0 \otimes \mathcal{H}_2), \mathcal{L}(\mathcal{H}_1 \otimes \mathcal{H}_3)) &\leftrightarrow M \in \mathcal{L}(\mathcal{H}_0 \otimes \mathcal{H}_2 \otimes \mathcal{H}_1 \otimes \mathcal{H}_3) \\ \mathcal{N} \in \mathcal{L}(\mathcal{L}(\mathcal{H}_3 \otimes \mathcal{H}_5), \mathcal{L}(\mathcal{H}_4 \otimes \mathcal{H}_6)) &\leftrightarrow N \in \mathcal{L}(\mathcal{L}(\mathcal{H}_3 \otimes \mathcal{H}_5 \otimes \mathcal{H}_4 \otimes \mathcal{H}_6)) \\ \mathcal{N} \star \mathcal{M} &\leftrightarrow N * M = \mathrm{Tr}_3[(I_{456} \otimes M_{0123}^{T_3})(I_{012} \otimes N)].\end{aligned} \tag{2.25}$$

The above results suggest us the following definition

Definition 2.1 (Link product) *Let M be an operator in $\mathcal{L}(\bigotimes_{i \in \mathsf{I}} \mathcal{H}_i)$ and N be an operator in $\mathcal{L}(\bigotimes_{j \in \mathsf{J}} \mathcal{H}_j)$ where I and J are two finite set of indexes. Then the* link product *$N * M$ is an operator in $\mathcal{L}(\mathcal{H}_{\mathsf{I} \setminus \mathsf{J}} \otimes \mathcal{H}_{\mathsf{J} \setminus \mathsf{I}})$ defined as*

$$N * M := \mathrm{Tr}_{\mathsf{I} \cap \mathsf{J}}[(I_{\mathsf{J} \setminus \mathsf{I}} \otimes M^{T_{\mathsf{I} \cap \mathsf{J}}})(I_{\mathsf{I} \setminus \mathsf{J}} \otimes N)] \tag{2.26}$$

where $\mathsf{A} \setminus \mathsf{B} := \{i \in \mathsf{A} | i \notin \mathsf{B}\}$ and $\mathcal{H}_{\mathsf{A}} := \bigotimes_{i \in \mathsf{A}} \mathcal{H}_i$

Remark 2.1 *It is worth noting that if $\mathsf{I} \cap \mathsf{J} = \emptyset$ we have $N * M = N \otimes M$ while if $\mathsf{I} = \mathsf{J}$ $N * M = \mathrm{Tr}[M^T N]$;*

The previous discussion is summarized by the following theorem.

Theorem 2.2 (Composition of linear maps) *Let* $\mathsf{in}_\mathcal{M}, \mathsf{out}_\mathcal{M}, \mathsf{in}_\mathcal{N}, \mathsf{out}_\mathcal{N}$ *be four sets of indeces such that* $\mathsf{in}_\mathcal{M} \cap\ \mathsf{out}_\mathcal{N} = \emptyset$*.*
Let $\mathcal{M}$ *be map in* $\mathcal{L}(\mathcal{L}(\bigotimes_{i\in\mathsf{in}_\mathcal{M}}\mathcal{H}_i), \mathcal{L}(\bigotimes_{j\in\mathsf{out}_\mathcal{M}}\mathcal{H}_j)$*,* $\mathcal{N}$ *be map in* $\mathcal{L}(\mathcal{L}(\bigotimes_{i\in\mathsf{in}_\mathcal{N}}\mathcal{H}_i), \mathcal{L}(\bigotimes_{j\in\mathsf{out}_\mathcal{N}}\mathcal{H}_j)$ *and* $M \in \mathcal{L}(\bigotimes_{n\in\mathsf{in}_\mathcal{M}\cup\ \mathsf{out}_\mathcal{M}}\mathcal{H}_m)$*,* $N \in \mathcal{L}(\bigotimes_{n\in\mathsf{in}_\mathcal{N}\cup\ \mathsf{out}_\mathcal{N}}\mathcal{H}_n)$ *be their respective Choi operators. Then the Choi operator of the composition*

$$\mathcal{M}\star\mathcal{N} := (\mathcal{I}_{\mathsf{in}_\mathcal{N}\setminus(\mathsf{in}_\mathcal{N}\cap\ \mathsf{out}_\mathcal{M})}\otimes\mathcal{M})\star(\mathcal{I}_{\mathsf{out}_\mathcal{M}\setminus(\mathsf{out}_\mathcal{M}\cap\ \mathsf{in}_\mathcal{N})}\otimes\mathcal{N}) \tag{2.27}$$

is given by

$$\mathfrak{C}(\mathcal{M}\star\mathcal{N}) = M * N \tag{2.28}$$

We conclude this section with some properties of the link product

Lemma 2.4 (Properties of link product) *Let* $M_1, M_2, M_3\ M_4$ *be operators in* $\mathcal{L}(\bigotimes_{i\in\mathsf{I}_1}\mathcal{H}_i)$*,* $\mathcal{L}(\bigotimes_{i\in\mathsf{I}_2}\mathcal{H}_i)$*,* $\mathcal{L}(\bigotimes_{i\in\mathsf{I}_3}\mathcal{H}_i)$ *and* $\mathcal{L}(\bigotimes_{i\in\mathsf{I}_4}\mathcal{H}_i)$ *respectively. Then we have*

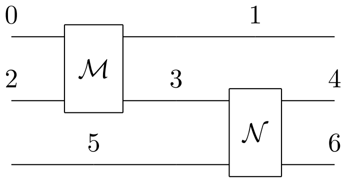
- *If* N_1 *is an operator on* $\mathcal{L}(\bigotimes_{i\in\mathsf{I}_1}\mathcal{H}_i)$ $(\alpha N_1 + \beta M_1) * M_3 = \alpha(N_1 * M_3) + \beta(M_1 * M_3)$ *for any* $\alpha, \beta \in \mathbb{C}$
- $M_1 * M_2 = M_2 * M_1$.
- *If* $M_1^\dagger = M_1$ *and* $M_2^\dagger = M_2$ *then* $(M_1 * M_2)^\dagger = M_1 * M_2$.
- *If* $\mathsf{I}_1 \cap \mathsf{I}_2 \cap \mathsf{I}_3 = \emptyset$ *then* $(M_1 * M_2) * M_3 = M_1 * (M_2 * M_3)$
- *If* $M \geqslant 0$ *and* $N \geqslant 0$ *then* $M * N \geqslant 0$.

Proof. The first four properties trivially follow from the definition. To prove the last property consider the maps $\mathfrak{C}^{-1}(M)$ and $\mathfrak{C}^{-1}(N)$. Since M and N are positive $\mathfrak{C}^{-1}(M)$ and $\mathfrak{C}^{-1}(N)$ are completely positive and also $\mathfrak{C}^{-1}(M)\star\mathfrak{C}^{-1}(N)$ is completely positive. Then $\mathfrak{C}(\mathfrak{C}^{-1}(M)\star\mathfrak{C}^{-1}(N)) = M * N$ is positive. ■

2.2 Diagrammatic representation of linear maps

It is useful to provide a pictorial representation of linear maps and their composition. We will sketch a linear map $\mathcal{M} : \mathcal{L}(\mathcal{H}_0 \otimes \cdots \otimes \mathcal{H}_n) \to \mathcal{L}(\mathcal{H}_{0'} \otimes \cdots \otimes \mathcal{H}_{n'})$ as box with n *input wires* on the left m *output wires* on the right as in Fig. 2.1.

Using this representation the composition in Eq. (2.24) can be sketched as follows



. (2.29)

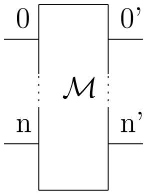
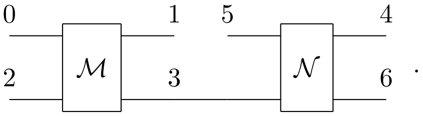


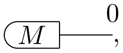
Figure 2.1. Pictorial representation of a linear map $\mathcal{M} : \mathcal{L}(\mathcal{H}_0 \otimes \cdots \otimes \mathcal{H}_n) \to \mathcal{L}(\mathcal{H}_{0'} \otimes \cdots \otimes \mathcal{H}_{n'})$; the input wires are labelled according to the labeling of the Hilbert spaces.

or equivalently



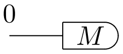
(2.30)

We do not draw wires corresponding to one dimensional Hilbert spaces. We will sketch a map $\mathcal{M} : \mathbb{C} \to \mathcal{L}(\mathcal{H}_0)$ with a one dimensional input as follows



(2.31)

where we use the label M instead of $\mathcal{M}$. In a similar way we represent maps $\mathcal{M} : \mathcal{L}(\mathcal{H}_0) \to \mathbb{C}$ that have one dimensional output space



(2.32)

2.3 States, Channels and POVMs

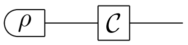
In the ordinary description of Quantum mechanics each physical system is associated with a Hilbert space $\mathcal{H}$ (that we will assume to be finite-dimensional) and the states of the system are represented by positive operators with unit trace $\rho \in \mathcal{L}(\mathcal{H})\rho \geqslant 0, \mathrm{Tr}[\rho] = 1$. Deterministic transformations of states are described by linear maps $\mathcal{C} : \mathcal{L}(\mathcal{H}_0) \to \mathcal{L}(\mathcal{H}_1)$ that have to be

- completely positive $\mathcal{C} \otimes \mathcal{I}_2(\rho) \geqslant 0$ for all $\rho \in \mathcal{L}(\mathcal{H}_0 \otimes \mathcal{H}_2)$;

- trace preserving $\mathrm{Tr}[\mathcal{C}(\rho)] = 1$ for all $\rho \in \mathcal{L}(\mathcal{H}_0), \mathrm{Tr}[\rho] = 1$

Deterministic transformations of states are called *quantum channels*. Thanks to Th. 2.1 and lemmas 2.1 and 2.3 of the previous section we know that a quantum channel $\mathcal{C} \in \mathcal{L}(\mathcal{L}(\mathcal{H}_0), \mathcal{L}(\mathcal{H}_1))$ can be represented by its Choi operator $C \in \mathcal{L}(\mathcal{H}_0 \otimes \mathcal{H}_1)$ that satisfies $C \geqslant 0$ and $\mathrm{Tr}_0[C] = I_1$ It is worth noting that the action $\mathcal{C}(\rho) = \mathrm{Tr}_0[(I_1 \otimes \rho^T)C]$ can be rewritten in terms of the link product as

$$\mathcal{C}(\rho) = C * \rho$$



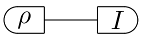
. (2.33)

where the state ρ is interpreted as the Choi operator of a preparation device, that is a channel $\tilde{\rho}$ from a one dimensional Hilbert space to $\mathcal{L}(\mathcal{H}_0)$

$$\begin{aligned}
&\tilde{\rho} : \mathbb{C} \to \mathcal{L}(\mathcal{H}_0) \qquad \tilde{\rho}(\lambda) = \lambda\rho \quad \forall\lambda \in \mathbb{C} \\
&\mathfrak{C}(\tilde{\rho}) = (\tilde{\rho} \otimes \mathcal{I}_{\mathbb{C}})(1 \otimes 1) = \tilde{\rho}(1) = \rho
\end{aligned} \tag{2.34}$$

Another relevant case is the one in which the output space is one dimensional. In this case we have a channel $\mathcal{C} : \mathcal{L}(\mathcal{H}_0) \to \mathbb{C}$ that receives a state ρ as an input and outputs the normalization $\mathcal{C}(\rho) = \mathrm{Tr}[\rho]$; it is easy to verify that its Choi operator is $\mathfrak{C}(\mathcal{C}) = I$ and so we have

$$\mathrm{Tr}[\rho] = \rho * I \tag{2.35}$$



.

We can then rewrite the normalization condition $\mathrm{Tr}_1[C_{01}] = I_0$ of the Choi operator of a quantum channel $\mathcal{C} \in \mathcal{L}(\mathcal{L}(\mathcal{H}_0), \mathcal{L}(\mathcal{H}_1))$ in the following way

$$C_{01} * I_1 = I_1 \tag{2.36}$$

A relevant class of channels are the *isometric channels*, that are defined as follows

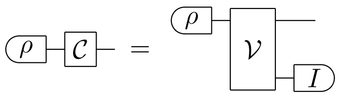
$$\mathcal{V} : \mathcal{L}(\mathcal{H}_0) \to \mathcal{L}(\mathcal{H}_1) \qquad \mathcal{V}(\rho) := V\rho V^\dagger \tag{2.37}$$

$$V \in \mathcal{L}(\mathcal{H}_0, \mathcal{H}_1), \quad V^\dagger V = I_0. \tag{2.38}$$

The following theorem [36,37] states that every quantum channel can be realized as an isometric channel on a larger system

Theorem 2.3 (Stinespring dilation theorem) *Let $\mathcal{C} : \mathcal{L}(\mathcal{H}_0) \to \mathcal{L}(\mathcal{H}_1)$ be completely positive trace preserving linear map. Then there exist an ancillary Hilbert space $\mathcal{H}_A$ and an isometry $V : \mathcal{L}(\mathcal{H}_0) \to \mathcal{L}(\mathcal{H}_1 \otimes \mathcal{H}_A)$, $V^\dagger V = I_0$ such that*

$$\mathcal{C}(\rho) = \mathrm{Tr}_A[\mathcal{V}(\rho)] = \mathrm{Tr}_A[V\rho V^\dagger] \tag{2.39}$$



V is called Stinespring dilation *of the channel $\mathcal{C}$*

Proof. Let C be the Choi Jamiołkowsky operator of $\mathcal{C}$ and define $\mathcal{H}_A = \mathsf{Supp}(C^*_{0'1'})$ (we introduced two auxiliary Hilbert spaces $\mathcal{H}_{0'}$ and $\mathcal{H}_{1'}$ and defined $C_{0'1'}$ according to Eq. (2.1)). Now consider the operator

$$V : \mathcal{H}_0 \to \mathcal{H}_1 \otimes \mathcal{H}_A \qquad V := I_1 \otimes C^{\frac{1}{2}*}_{0'1'} |I\rangle\!\rangle_{11'} T_{0\to 0'} \tag{2.40}$$

where $T_{0\to 0'} = \sum_k |k\rangle_{0'} \, |k\rangle_0$; By using Lemmas 2.1 and 2.3 together with Eqs. (2.3, 2.5, 2.6) it is easy to verify that V is an isometry

$$\begin{aligned}
V^\dagger V &= T_{0'\to 0} \langle\!\langle I|_{11'} (I_1 \otimes C^{\frac{1}{2}T}_{0'1'})(I_1 \otimes C^{\frac{1}{2}*}_{0'1'} |I\rangle\!\rangle_{11'} T_{0\to 0'} = \\
&= T_{0'\to 0} \, \mathrm{Tr}_{1'}[C^T_{0'1'}] T_{0\to 0'} = I_0,
\end{aligned} \tag{2.41}$$

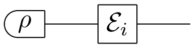
and that

$$
\begin{aligned}
\mathrm{Tr}_A[V\rho V^\dagger] &= \mathrm{Tr}_A[(I_1 \otimes C_{0'1'}^{\frac{1}{2}*})(|I\rangle\rangle_{11'}T_{0\to 0'})\rho(T_{0'\to 0}\langle\langle I|_{11'})(I_1 \otimes C_{0'1'}^{\frac{1}{2}T})] = \\
&= \mathrm{Tr}_{0'1'}[(I_1 \otimes C_{0'1'}^{T})(I_0' \otimes |I\rangle\rangle\langle\langle I|_{11'})(I_{11'} \otimes \rho_0')] = \\
&= \mathrm{Tr}_{0'}[\mathrm{Tr}_{1'}[(I_1 \otimes C_{0'1'}^{T})(I_0' \otimes |I\rangle\rangle\langle\langle I|_{11'})](I_1 \otimes \rho_0')] = \\
&= \mathrm{Tr}_{0'}[C_{0'1}^{T_{0'}}(I_1 \otimes \rho_0')] = C * \rho = \mathcal{C}(\rho) \qquad \blacksquare
\end{aligned}
$$

Remark 2.2 *The Stinespring dilation of a channel is generally non unique. We now prove that the isometric dilation given by Eq. (2.40) has minimum ancilla dimension. Suppose that there exists an isometric dilation* $W : \mathcal{H}_0 \to \mathcal{H}_1 \otimes \mathcal{H}_B$ *such that* $d_B = \dim(\mathcal{H}_B) < \dim(\mathcal{H}_A) = d_A$. *Each isometric dilation of a channel* $\mathcal{C}$ *provides operator-sum representation of* $\mathcal{C}$ *as follows* $\mathcal{C}(\rho) = \mathrm{Tr}_A[W\rho W^\dagger] = \sum_{n=1}^{d_B} \langle n| W\rho W^\dagger |n\rangle := \sum_{n=1}^{d_B} K_n \rho K_n^\dagger$ *where* $|n\rangle$ *is an orthonormal basis in* $\mathcal{H}_B$. *From the operator sum representation it is possible to recover the Choi operator* C *of* $\mathcal{C}$ *as follows* $\mathfrak{C}(\mathcal{C}) = \sum_{n=1}^{d_B}(K_n \otimes I)|I\rangle\rangle\langle\langle I|(K_n^\dagger \otimes I) = \sum_{n=1}^{d_B} |K_n\rangle\rangle\langle\langle K_n|$. *Since* $\langle\langle K_n|K_m\rangle\rangle = \langle n| V^\dagger V |m\rangle = \langle n|m\rangle = \delta_{nm}$, *the vectors* $|K_n\rangle\rangle$ *are linearly independent and this leads to the contradiction* $\dim(\mathsf{Supp}(C)) = \dim(\mathrm{Span}\{|K_n\rangle\rangle\}) = d_B < d_A = \dim(\mathsf{Supp}(C^*)) = \dim(\mathsf{Supp}(C))$. *The isometric dilation defined in Eq. (2.40) is called* minimal Stinespring dilation

The probabilistic counterpart of a quantum channel is the *quantum operation*. A quantum operation is a completely positive linear map $\mathcal{E} \in \mathcal{L}(\mathcal{L}(\mathcal{H}_0), \mathcal{L}(\mathcal{H}_1))$ which is *trace non-increasing* $\mathrm{Tr}[\mathcal{E}(\rho)] \leqslant 1$ for any state ρ. The Choi Jamiołkowski operator E of a quantum operation $\mathcal{E}$ satisfies the condition $E \leqslant \overline{E}$ where $\overline{E}$ is the Choi operator of a quantum channel. A set of quantum operation $\{\mathcal{E}_i\}$ that sum up to a channel $\mathcal{C}$ is called a *Quantum Instrument*[5] and it is represented by a set of positive operator E_i such that $\sum_i E_i = C$; the index i labels the possible classical outcomes of the instrument. The action of a Quantum Instrument is written as

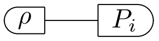
$$\sum_i \mathcal{E}_i(\rho) = \sum_i \rho * E_i$$



. (2.42)

and the probability that the Quantum Operation $\mathcal{E}_i$ takes place is $p_i = \mathrm{Tr}[(I \otimes \rho)E_i]$. A Quantum Instrument with one-dimensional output space is called *POVM* and is represented by a set of positive operator P_i such that $\sum_i P_i = I$; the elements P_i of a POVM are called *effects*. The link product

$$\mathrm{Tr}[\rho P_i^T] = \rho * P_i$$



. (2.43)

gives the probability p_i of the outcome i and coincides with the usual Born rule

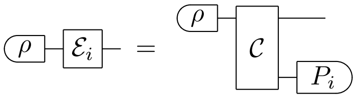
$$p_i = \mathrm{Tr}[\rho P_i] \tag{2.44}$$

if we make the substitution $P_i \leftrightarrow P_i^T$. We conclude this section with a theorem [38, 37] that provides a realization scheme for Quantum Instruments in terms of a deterministic evolution on a bigger system followed by a measurement on the ancilla.

[5] For simplicity we restricted ourselves to the case of a finite number of outcomes. The generalization to an arbitrary outcome space Ω can be obtained by defining a measure $\mathcal{E}_B$ that associate to any event $B \subseteq \Omega$ a quantum operation $\mathcal{E}_B$ such that $\mathcal{E}_\Omega$ is a Quantum channel.

Theorem 2.4 (Realization of Quantum Instruments) *Let $\{\mathcal{E}_i\}$, $\mathcal{E}_i \in \mathcal{L}(\mathcal{L}(\mathcal{H}_0), \mathcal{L}(\mathcal{H}_1))$ be a Quantum instrument. Then there exist an Hilbert space $\mathcal{H}_A$, a channel $\mathcal{C} \in \mathcal{L}(\mathcal{L}(\mathcal{H}_0), \mathcal{L}(\mathcal{H}_1 \otimes \mathcal{H}_A))$ and a POVM $\{P_i\}$, $P_i \in \mathcal{L}(\mathcal{H}_A)$ such that*

$$\mathcal{E}_i(\rho) = \mathrm{Tr}_A[\mathcal{C}(\rho)(I_1 \otimes P_i)]$$



(2.45)

Proof. Let us define $\mathcal{C} := \sum_i \mathcal{E}_i$ and let C be the Choi operator of $\mathcal{C}$ and E_i be the Choi operator of $\mathcal{E}_i$. Since $\mathcal{C}$ is a quantum channel, we can consider its minimal Stinespring dilation $V : \mathcal{H}_0 \rightarrow \mathcal{H}_1 \otimes \mathcal{H}_A$,
$\mathcal{H}_A = \mathsf{Supp}(C)$. Now we introduce the POVM $\{P_i \in \mathcal{L}(\mathcal{H}_A)\}$ $P_i = C^{-\frac{1}{2}T} E_i^T C^{-\frac{1}{2}T}$ (clearly $\sum_i E_i = I_A$ and $P_i^\dagger = P_i$). It is easy to verify that

$$\begin{aligned}\mathrm{Tr}_A[\mathcal{C}(\rho)(I_1 \otimes P_i)] &= \mathrm{Tr}_A[V\rho V^\dagger(I_1 \otimes P_i)] = \\ &= \mathrm{Tr}_A[(I_1 \otimes C^{\frac{1}{2}*}_{0'1'})(\rho_{0'}|I\rangle\rangle\langle\langle I|_{11'})(I_1 \otimes C^{\frac{1}{2}T}_{0'1'})(I_i \otimes P_i)] = \\ &= \mathrm{Tr}_A[(\rho_{0'}|I\rangle\rangle\langle\langle I|_{11'})(I_1 \otimes E_i^T)] = \mathcal{E}_i(\rho) \qquad (2.46)\end{aligned}$$

2.4 Quantum Networks: constructive approach

In this section we introduce the formal definition of Quantum Network. Within our approach a Quantum Network is obtained by assembling elementary circuits linking outputs of a circuit to inputs of another circuit; we consider "elementary circuits" channels, quantum operations, effects or state preparations each of them represented with the corresponding linear map. The restriction that we can connect only outputs with inputs and that we cannot have closed loops ensures causality (see Remark 2.4) and motivates the following definition

Definition 2.2 (Quantum Network) *A* quantum network $\mathcal{R}$ *is a linear map corresponding to directed acyclic graph (DAG) in which*

- *each arrow is labeled with a non negative integer number n (two different arrows cannot have the same label);*
- *an arrow with label n represents an Hilbert space $\mathcal{H}_n$;*
- *each vertex is labelled with a non negative integer number i (two different vertexes cannot have the same label);*
- *each vertex i represents a completely-positive trace non-increasing map $\mathcal{C}_i \in \mathcal{L}(\mathcal{H}_{\mathsf{in}_i} \otimes \mathcal{H}_{\mathsf{out}_i})$ ($\mathcal{H}_{\mathsf{A}} = \bigotimes_{k \in \mathsf{A}} \mathcal{H}_k$) where in_i is the set of incoming arrows at vertex i and out_i is the set of outgoing arrows at vertex i;*
- *an arrow between two vertices's i and j corresponds to the composition $\mathcal{C}_j \star \mathcal{C}_i$ of the linear maps $\mathcal{C}_i$ and $\mathcal{C}_j$*

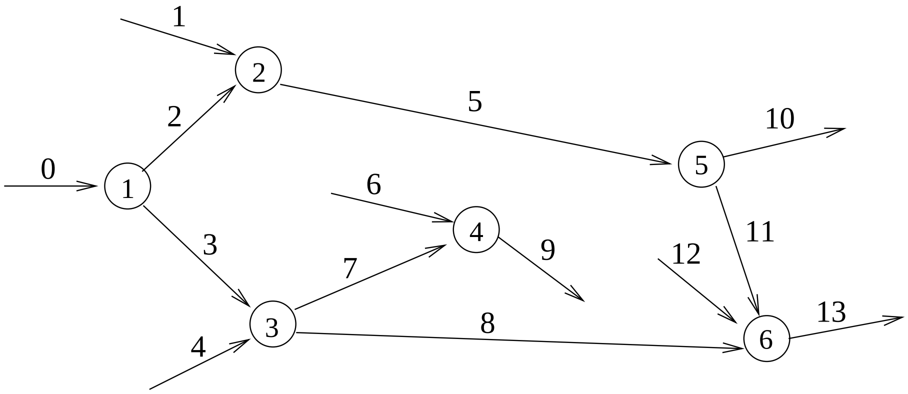


Figure 2.2. Graphical representation of a quantum network. The directions of the arrows represent the flow of quantum information in the network, that is quantum systems travelling from a vertex to another. Free incoming arrows represent input systems entering the network while free outgoing arrows represent output systems of the network.

- *we remove some vertices's with no incoming arrows (sources) and some vertices's with no outgoing arrows (sink). The free incoming arrows remaining represent input systems entering the network while the free outgoing arrows carry the output systems.*

If $\mathcal{C}_i$ is a channel for each vertex i $\mathcal{R}$ is called a deterministic quantum network. *If $\mathcal{C}_i$ is a trace decreasing for some vertex i, $\mathcal{R}$ is called a* probabilistic quantum network.

Fig. 2.2 provides a typical example of a quantum network.

Remark 2.3 *It is worth noting that the same Quantum Network can be realized in different ways as a sequence of maps*

$$\mathcal{R}^{(N)} = \mathcal{C}_1 \star \mathcal{C}_2 \star \cdots \star \mathcal{C}_N = \mathcal{C}'_1 \star \mathcal{C}'_2 \star \cdots \star \mathcal{C}'_N \tag{2.47}$$

and this fact reflects different possible physical implementation of the same network. In this work we are not interested in the inner structure of a network but only in its properties as a linear map from input spaces to output spaces. Because of this, whenever we introduce a Quantum Network $\mathcal{R}^{(N)}$, we actually mean an equivalence class of sequence of maps that give the same overall operator $\mathcal{R}^{(N)}$, i.e. we consider the two sequences of maps $\mathcal{C}_1 \star \mathcal{C}_2 \star \cdots \star \mathcal{C}_N$ and $\mathcal{C}'_1 \star \mathcal{C}'_2 \star \cdots \star \mathcal{C}'_N$ in Eq. 2.47 as the same object.

Remark 2.4 *The condition that the graph is acyclic means that no closed path is allowed. This requirement ensures that causality is preserved, since the flow of quantum information induces a causal order inside the network and a closed path would correspond to a time-loop. It is worth stressing that in our representation a physical closed loop in the lab, that is taking the output of a device and then sending it as an input to the same device, corresponds to many uses of the*

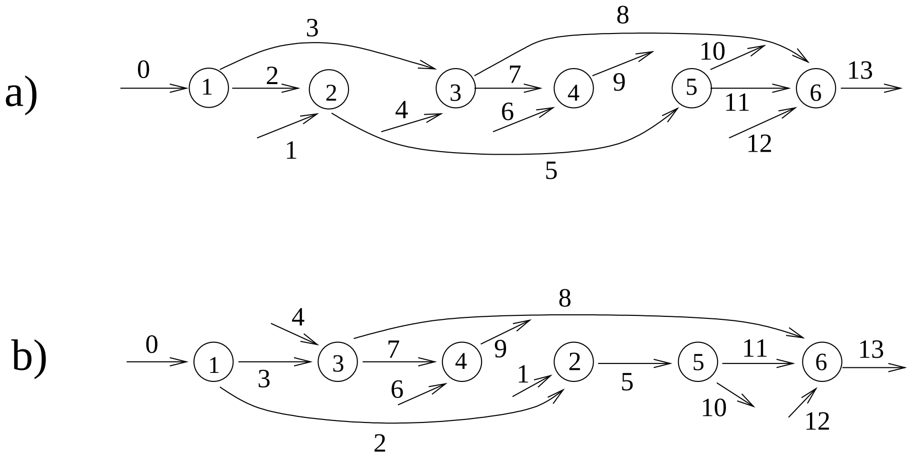
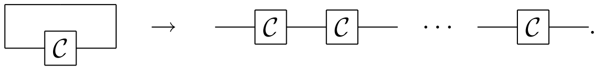


Figure 2.3. Two possible total orderings of the network in Fig. 2.2.

same transformation



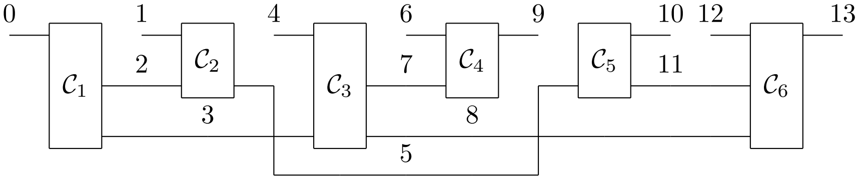
(2.48)

In this work we use the convention that a vertex in a network or a box in a circuit represents a single use of a physical device.

Any direct acyclic graph is naturally endowed with a partial ordering $\preceq$ among the vertices's, which is the causal ordering induced by the flow of quantum information (see Remark 2.4); we say that vertex i causally precedes vertex j ($i \preceq j$) if there exists a directed path from i to j. It is possible to prove that for a directed acyclic graph the partial ordering $\preceq$ can be extended, in a generally non unique way, to a total ordering $\leqslant$ (See Fig. 2.3).

Each vertex in the network corresponds to a step of a computation and the relation $i \preceq j$ means that step j cannot be performed before step i. If two vertexes are incomparable this means that the two steps can be run in parallel; extending the partial ordering to a total ordering consists in arbitrarily fixing an ordering among parallel computational steps that is compatible with the partial ordering $\preceq$.

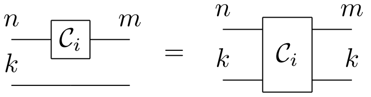
Since each vertex i in a quantum network corresponds to a linear map $\mathcal{C}_i$ and any arrow between two vertexes corresponds to a composition, we can exploit the diagrammatic representation that we introduced before and represent a quantum network in a circuit form



; (2.49)

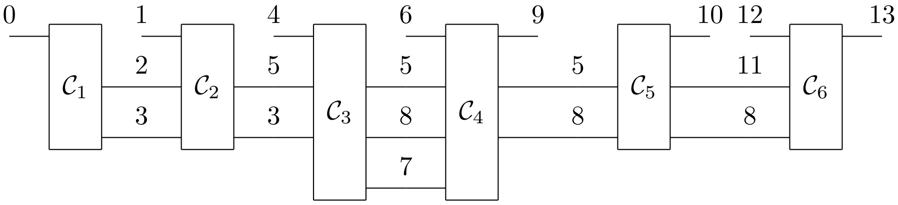
where the free incoming/outgoing arrows are now substituted by free input/output wires; The flow of quantum information is from left to right and the numbering of the boxes is chosen accordingly.

To avoid drawing crossing wires, it is possible to enlarge each box by tensoring with the identity map i.e.



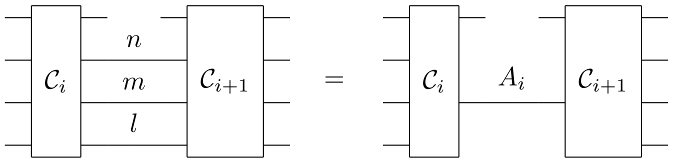
$$\mathcal{C}_i \to \mathcal{C}_i \otimes \mathcal{I}_k \tag{2.50}$$

in this way the network takes the shape of a chain



; (2.51)

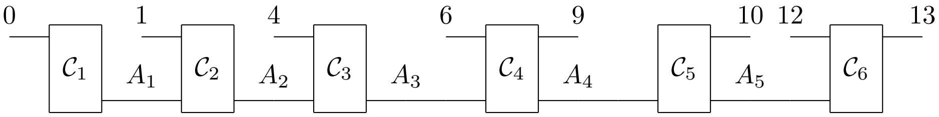
we can further lighten the diagram by merging the internal wires connecting two boxes



(2.52)

$$\mathcal{H}_{A_i} = \mathcal{H}_n \otimes \mathcal{H}_m \otimes \mathcal{H}_l.$$

In this way the circuit 2.51 becomes



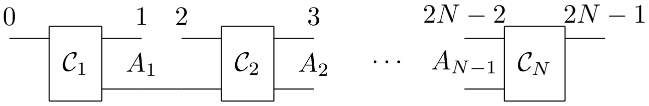
.

(2.53)

The previous considerations can be summarized in the following

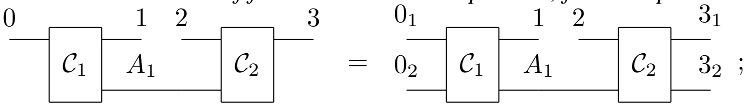
Lemma 2.5 (Circuit form for Quantum Networks) *Any quantum network $\mathcal{R}$ with N vertexes is equivalent to a concatenation of N completely positive trace non increasing linear maps*

$$\mathcal{R} = \mathcal{C}_1 \star \mathcal{C}_2 \star \cdots \star \mathcal{C}_N \tag{2.54}$$



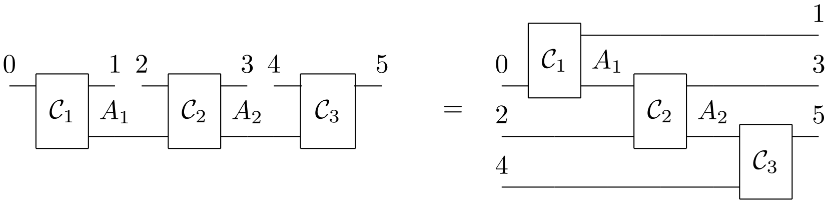
where $\mathcal{C}_i : \mathcal{L}(\mathcal{H}_a \otimes \mathcal{H}_{A_{i-1}}) \to \mathcal{L}(\mathcal{H}_b \otimes \mathcal{H}_{A_i})$.

Remark 2.5 *In Eq. (2.54) we chose to attach one free incoming and one free outgoing wire to each map $\mathcal{C}_i$. This is our standard representation of quantum network; we can without loss of generality sketch any quantum network in this way, since network in which some input/output wires are missing (like in 2.53) are just special cases. We can stress, if present, a tensor product structure $\mathcal{H}_a = \otimes_j \mathcal{H}_{a_j}$ of the Hilbert space carried by a free input/output wire a, by drawing as many wires as the number of factors in the tensor product, for example*



where $\mathcal{H}_0 = \mathcal{H}_{0_1} \otimes \mathcal{H}_{0_2}$ and $\mathcal{H}_3 = \mathcal{H}_{3_1} \otimes \mathcal{H}_{3_2}$. We also choosed to label the free input/output wires with increasing integer numbers; in this way the Hilbert spaces of the input wires are labeled with even numbers while the output ones correspond to odd numbers. We can define the overall input space of the network as $\mathcal{H}_{\mathrm{in}} = \bigotimes_{i=1}^{N} \mathcal{H}_{2i-2}$ and $\mathcal{H}_{\mathrm{out}} = \bigotimes_{j=1}^{N} \mathcal{H}_{2i-1}$

Lemma 2.5 reveals the equivalence between a Quantum Network and a sequence of N channels with memory; if we stretch and rearrange the input and the output wires



from a Quantum Network we get a sequence of memory channels from the left side to the right side. Since a Quantum Network is a sequence of linear maps, it can be considered as a linear map from $\mathcal{L}(\mathcal{H}_{\mathrm{in}})$ to $\mathcal{L}(\mathcal{H}_{\mathrm{out}})$. It is then possible to define the Choi operator of a Quantum Network

$$\mathfrak{C}(\mathcal{R}^{(N)}) = R^{(N)}, \tag{2.55}$$

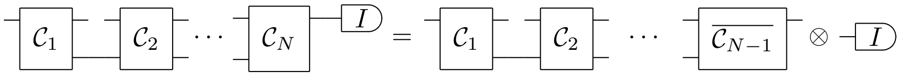
where we add the superscript $^{(N)}$ to record the number of vertex in the network i.e. $\mathcal{R}^{(N)}$ denotes a quantum network with N vertexes.

2.5 Deterministic Quantum Networks

The main aim of the following sections will be to inspect the structure of the Choi operator of a Quantum Network. In this section we consider the deterministic case while the probabilistic case will be discussed in the next section. Specializing Lemma 2.5 a deterministic Quantum Network $\mathcal{R}^{(N)}$ can be presented as a concatenation of N quantum channels $\mathcal{C}_i$; then the Choi operator of $\mathcal{R}^{(N)}$ is given by the link product of the C_i's. This structure leads to a peculiar normalization constraint for R^N.

Theorem 2.5 (Normalization Condition) *Let $\mathcal{R}^{(N)}$ be a deterministic Quantum Network and $R^N \in \mathcal{L}(\bigotimes_{i=0}^{2N-1} \mathcal{H}_i)$ (we use the labeling introduced in Lemma 2.5 and Remark 2.5) be its Choi operator. Then $R^{(N)} \geqslant 0$ and satisfies the following condition*

$$\mathrm{Tr}_{2N-1}\left[R^{(N)}\right] = I_{2N} \otimes R^{(N-1)} \tag{2.56}$$



where $R^{(N-1)} \in \mathcal{L}(\bigotimes_{i=0}^{2N-1}\mathcal{H}_i)$ *is the Choi operator of the reduced Quantum Network with* $N-1$ *vertexes and* $\overline{\mathcal{C}_{N-1}}$ *is a quantum channel such that* $\mathfrak{C}(\overline{\mathcal{C}_{N-1}}) := \overline{C_{N-1}} = \mathrm{Tr}_{A_{N-1}}[C_{N-1}]$.

Proof. Since $\mathcal{R}^{(N)}$ is a quantum Network with N vertexes, we can express it in terms of a concatenation of N channels

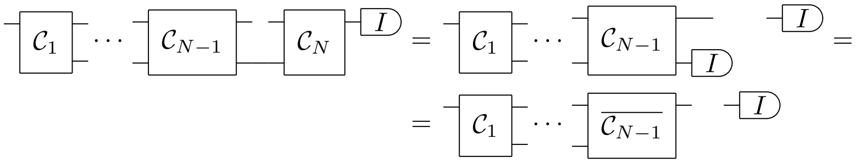
$$\begin{aligned}&\mathcal{R}^{(N)} = \mathcal{C}_1 \star \mathcal{C}_2 \star \cdots \star \mathcal{C}_N \\ &\mathcal{C}_i : \mathcal{L}(\mathcal{H}_{2i-2} \otimes \mathcal{H}_{A_{i-1}}) \to \mathcal{L}(\mathcal{H}_{2i-1} \otimes \mathcal{H}_{A_i}) \qquad \mathcal{H}_{A_0} \cong \mathcal{H}_{A_N} \cong \mathbb{C}.\end{aligned} \tag{2.57}$$

Let $C_i \in \mathcal{L}(\bigotimes_{k\in\mathsf{I}_i}\mathcal{H}_k$ be the Choi of $\mathcal{C}_i$ where we introduced the set $\mathsf{I}_i := \{2i-2, A_{i-1}, 2i-1, A_i\}$; we notice that $\mathsf{I}_i \cap \mathsf{I}_j \cap \mathsf{I}_k = \emptyset$ for all $i, j, k = 1, \dots, N$ and so, exploiting Lemma 2.4, we have

$$R^{(N)} = C_1 * C_2 * \cdots * C_N. \tag{2.58}$$

Since $\mathcal{C}_N$ is channel in $\mathcal{L}(\mathcal{L}(\mathcal{H}_{2N-2} \otimes \mathcal{H}_{A_{N-1}}), \mathcal{L}(\mathcal{H}_{2N-1}))$ its Choi-Jamiołkowsky operator satisfies $\mathrm{Tr}_{2N-1}[C_N] = I_{2N-2} \otimes I_{A_{N-1}}$ then we have

$$\begin{aligned}\mathrm{Tr}_{2N-1}[R^{(N)}] &= C_1 * C_2 * \cdots * C_{N-1} * \mathrm{Tr}_{2N-1}[C_N] = \\ &= C_1 * C_2 * \cdots * (C_{N-1} * I_{2N-2} \otimes I_{A_{N-1}}) = \\ &= C_1 * C_2 * \cdots * \mathrm{Tr}_{A_{N-1}}[C_{N-1}] \otimes I_{2N-2} = \\ &= C_1 * C_2 * \cdots * \overline{C_{N-1}} \otimes I_{2N-2} = \\ &= R^{N-1} \otimes I_{2N-2}\end{aligned} \tag{2.59}$$



■

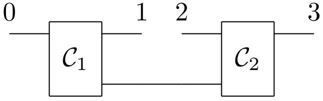
Corollary 2.2 *Let* $R^N \in \mathcal{L}(\mathcal{H}_{\mathrm{out}} \otimes \mathcal{H}_{\mathrm{in}})$ *(* $\mathcal{H}_{\mathrm{in}} = \bigotimes_{i=1}^N \mathcal{H}_{2i-2}$ *and* $\mathcal{H}_{\mathrm{out}} = \bigotimes_{j=1}^N \mathcal{H}_{2i-1}$*) be the Choi operator of a deterministic Quantum Network* $\mathcal{R}^{(N)}$ *. Then* $R^{(N)}$ *satisfies*

$$\mathrm{Tr}_{2k-1}[R^{(k)}] = I_{2k-2} \otimes R^{(k-1)}, \qquad 1 \leqslant k \leqslant N \tag{2.60}$$

where $R^{(0)} = 1$, $R^{(k)} \in \mathcal{L}(\mathcal{H}_{\mathrm{out}_k} \otimes \mathcal{H}_{\mathrm{in}_k})$, $\mathcal{H}_{\mathrm{in}_k} = \bigotimes_{i=0}^{k-1}\mathcal{H}_{2i}$, $\mathcal{H}_{\mathrm{out}_k} = \bigotimes_{i=0}^{k-1}\mathcal{H}_{2i+1}$.

Proof. Eq. (2.60) can be obtained by recursively applying Eq. (2.56). ■

Remark 2.6 *We want to stress that Eq. (2.60) reflects the causal ordering of the Quantum Network. This property translates the fact that information can be transmitted from system i to a system j if $i < j$ but not to a system $j' < i$. Consider the Network $\mathcal{R}^{(2)} \in \mathcal{L}(\mathcal{L}(\mathcal{H}_0 \otimes \mathcal{H}_2), \mathcal{L}(\mathcal{H}_0 \otimes \mathcal{H}_2))$*



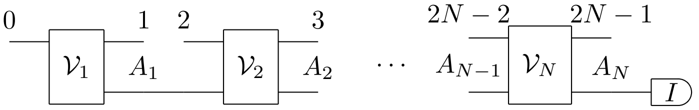
We will now prove that the condition that no information flows from 2 to 1 is equivalent to $\mathrm{Tr}_3[R^{(2)}_{0123}] = I_2 \otimes R^{(1)}_{01}$. The condition that there is no flow of information from 2 to 1 can be expressed by saying that upon application of the memory channel represented by $R^{(2)}$ to a general input state ρ_{02}, the partial state in 1 does not depend on the local state in 0 i.e. $\mathrm{Tr}_3[\sigma_{13}] = \mathrm{Tr}_3[\mathrm{Tr}_{02}[(\rho^T_{02} \otimes I_{13})R^{(2)}_{0123}]] = \mathcal{A}(\mathrm{Tr}_0[\rho_{02}])$ for a fixed channel $\mathcal{A}$. If $\mathrm{Tr}_3[R^{(2)}_{0123}] = I_2 \otimes R^{(1)}_{01}$ we have $\mathrm{Tr}_3[\sigma_{13}] = \mathrm{Tr}_3[\mathrm{Tr}_{02}[(\rho^T_{02} \otimes I_{13})R^{(2)}_{0123}]] = \mathrm{Tr}_{02}[(\rho^T_{02} \otimes I_{13})R^{(1)}_{013} \otimes I_2] = \mathrm{Tr}_0[(\mathrm{Tr}_2[\rho^T_{02}] \otimes I_1)R^{(1)}_{013}] = \mathcal{A}(\mathrm{Tr}_0[\rho_{02}])$ if we define $\mathcal{A} := \mathfrak{C}^{-1}(R^{(1)})$.

On the other hand let us suppose that $\mathrm{Tr}_3[\mathrm{Tr}_{02}[(\rho^T_{02} \otimes I_{13})R^{(2)}_{0123}]] = \mathcal{A}(\mathrm{Tr}_0[\rho_{02}])$ for a fixed $\mathcal{A}$. In particular if $\rho_{02} = \tau_0 \otimes \omega_2$ we have $\mathrm{Tr}_3[\mathrm{Tr}_{02}[(\tau^T_0 \otimes \omega^T_2) \otimes I_{13})R^{(2)}_{0123}]] = \mathrm{Tr}_0[(\tau^T_0 \otimes I_1)\,\mathrm{Tr}_2[(\omega^T_2 \otimes I_{10})\,\mathrm{Tr}_3[R^{(2)}_{0123}]]] = \mathrm{Tr}_0[(\tau^T_0 \otimes I_1)A_{01}] = \mathcal{A}(\mathrm{Tr}_0[\tau_0])$, where $\mathfrak{C}(\mathcal{A}) = A_{01} = \mathcal{S}(\omega_2))$ and $\mathcal{S} = \mathfrak{C}^{-1}(\mathrm{Tr}_3[R^{(2)}_{0123}])$. Since $\mathcal{A}$ is a constant we have $\mathcal{S}(\omega_2) = A_{01}$ for all ω, that implies $\mathfrak{C}(\mathcal{S}) = I_2 \otimes A_{01}$. The Quantum Network $\mathcal{R}^{(2)}$ when considered as channel from $\mathcal{L}(\mathcal{L}(\mathcal{H}_{02}))$ to $\mathcal{L}(\mathcal{L}(\mathcal{H}_{13}))$ has the properties of a semicausal channel *as discussed in Refs. [39, 40]*

The recursive normalization condition (2.60) and the positivity constraint characterize the Choi Operator of a deterministic Quantum Network. The following theorem tells us that a positive operator satisfying Eq. (2.60) is the Choi operator of a deterministic Quantum Network.

Theorem 2.6 (Realization of deterministic Quantum Networks) *Let $R^N \in \mathcal{L}(\mathcal{H}_{\mathrm{out}} \otimes \mathcal{H}_{\mathrm{in}})$ ($\mathcal{H}_{\mathrm{in}} = \bigotimes_{i=1}^N \mathcal{H}_{2i-2}$ and $\mathcal{H}_{\mathrm{out}} = \bigotimes_{j=1}^N \mathcal{H}_{2i-1}$) be a positive operator satisfying Eq. (2.60). Then $R^{(N)}$ is the Choi operator of a deterministic Quantum Network $\mathcal{R}^{(N)}$ given by the concatenation of N isometries followed by a trace on an ancillary space: for every state $\rho \in \mathcal{L}(\mathcal{H}_{\mathrm{in}})$ one has*

$$\mathcal{R}^{(N)}(\rho) = \mathrm{Tr}_{A_N}[V^{(N)} \cdots V^{(1)} \rho V^{(1)\dagger} \cdots V^{(N)\dagger}] \tag{2.61}$$



where $V^i \in \mathcal{L}(\mathcal{L}(\mathcal{H}_{2k-2} \otimes \mathcal{H}_{A_{k-1}}), \mathcal{L}(\mathcal{H}_{2k-1} \otimes \mathcal{H}_{A_k}))$ and $\mathcal{H}_{A_k}$ is an ancillary space, $\mathcal{H}_{A_0} = \mathbb{C}$ (in Eq. (2.61) we omitted the identity operators on the Hilbert spaces where the isometries do not act).

Proof. Define $\mathcal{H}_{A_k} = \mathsf{Supp}(R^{(k)*})$ and

$$V^{(k)} = I_{2k-1} \otimes R^{(k)\frac{1}{2}*} R^{(k-1)-\frac{1}{2}*} \,|I\rangle\!\rangle_{(2k-1)(2k-1)'} T_{(2k-2)\to(2k-2)'} \tag{2.62}$$

where $T_{n\to m} = \sum_i |i\rangle_m \langle i|_n$.
Using Eq. (2.60) one has $V^{(k)\dagger}V^{(k)} = \left(R^{(k-1)*}\right)^{-\frac{1}{2}} \mathrm{Tr}_{2k-1}[R^{(k)*}] \left(R^{(k-1)*}\right)^{-\frac{1}{2}} = I_{2k-2} \otimes I_{A_{k-1}}$ that is $V^{(k)}$ is an isometry. Now consider $W^{(N)} = V^{(N)} \cdots V^{(1)}$, which goes from $\mathcal{H}_{\mathrm{in}}$ to $\mathcal{H}_{\mathrm{out}} \otimes \mathcal{H}_{A_N}$; From Eq. 2.62 we have $W^{(N)} = (I_{\mathrm{out}} \otimes (R^{(N)*})^{\frac{1}{2}})|I\rangle\!\rangle_{(\mathrm{out})(\mathrm{out})'} \otimes T_{\mathrm{in}\to\mathrm{in}'}$ and Theorem 2.3 tells us that $W^{(N)}$ is an isometric dilation of $\mathcal{R}^N$ and so

$$\mathcal{R}^N(\rho) = \mathrm{Tr}_{A_N}[W^{(N)}\rho W^{(N)\dagger}] = \mathrm{Tr}_{A_N}[V^{(N)} \cdots V^{(1)}\rho V^{(1)\dagger} \cdots V^{(N)\dagger}]. \tag{2.63}$$

■

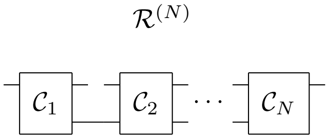
Corollary 2.3 *The minimal dimension of the ancilla space $\mathcal{H}_{A_k}$ is $\dim(\mathsf{Supp}(R^{(k)}))$ in Theorem 2.6*

Proof. Consider the isometries $W^{(k)} = V^{(k)} \cdots V^{(1)}$ where $V^{(i)}$ are defined according to Eq. 2.62. Theorem 2.3 tells us that $W^{(k)}$ is an isometric dilation of $\mathcal{R}^k$ with minimal ancillary space; then it is not possible to choose an ancillary space $\mathcal{H}_{B_k}$ with $\dim(\mathcal{H}_{B_k}) < \dim(\mathcal{H}_{A_k}) = \dim(\mathsf{Supp}(R^{(k)}))$. ■

Remark 2.7 *The maximum $d_{\max} := \max_{1\leqslant k\leqslant N} d_{A_k}$ provides an upper bound on the complexity of the Network in terms of quantum memory. Indeed, the Stinespring dilation theorem preserves coherence up to the last step; for example it can happen that some ancillary degrees of freedom are used only up to a step $k < N$ and then the isometries $V^{(k+1)}, \ldots, V^{(N)}$ act only trivially on them. In this case one can trace out some degrees of freedom before the last step. This deeper analysis of resources can be performed only by inspecting the structure of the isometries $V^{(k)}$.*

Remark 2.8 *We stress that the set of the Choi operators is a convex set; indeed, imposing linear constraints (like the one in Eq. 2.60) on a given convex set (like the set of positive operators) does not spoils the convexity.*

Theorems 2.5 and 2.6 provide a one to one correspondence between the set of deterministic Quantum Networks (considered as equivalence classes of different implementations as pointed out in Remark 2.3) and the set of positive operators satisfying the normalization (2.60)



$$\leftrightarrow \qquad \begin{array}{c} R^{(N)} \geqslant 0 \\ \text{such that} \\ \mathrm{Tr}_{2k-1}[R^{(k)}] = I_{2k-2} \otimes R^{(k-1)} \end{array} \qquad ;$$

following the same terminology introduced in Refs. [22, 23] we call the Choi operators of a Quantum Network *Quantum Combs*[6]. This result (and its generalization to the probabilistic case) allows to represent every Quantum Networks in terms of a single positive operator subjected to linear constraints. This is extremely relevant for applications. Indeed, optimizing a

[6] Whenever we want to stress the distinction between deterministic and probabilistic case we use the terms *deterministic Quantum Combs* and *probabilistic Quantum Combs* respectively.

Quantum Network by separately optimizing each device is extremely demanding. Thanks to this representation the optimization problem is reduced to an optimization problem over a convex set of suitably normalized positive operators. Moreover we notice that through Eq. (2.62) we are provided with an explicit expression of a Quantum Network that is represented by a given quantum comb $R^{(N)}$.

This allows us to formulate an algorithm for designing optimal Quantum Networks for a given task (e.g. cloning, discrimination, estimation):

1. Choose a suitable figure of merit F for the task of interest.
2. Find the positive operator $R^{(N)}$ satisfying constraint in Eq. (2.60) and optimizing F.
3. Set $R^{(0)} = 1$ and $I_{A_0} = 1$.
4. For $k = 1$ to $k = N$ do the following:
    (a) Calculate $I_{\overline{\mathrm{in_k}}} \otimes R^{(k)} = \mathrm{Tr}_{\overline{\mathrm{out_k}}}[C]$, where $I_{\overline{\mathcal{H}}}$ ($\mathrm{Tr}_{\overline{\mathcal{H}}}$) denotes the identity (partial trace) over all Hilbert spaces but $\mathcal{H}$;
    (b) define $V^{(k)}$ as in Eq. (2.62).
5. The optimal network is given by the concatenation of the $V^{(k)}$'s in Eq. (2.61)

2.6 Probabilistic Quantum Network

The aim of this section is to provide the equivalents of Theorems 2.5 and 2.6 for the case in which probabilistic Quantum Network are considered. We remind that a probabilistic Quantum Network $\mathcal{R}^{(N)}$ is equivalent to a concatenation of N completely positive trace non increasing linear maps[7]

$$\mathcal{R}^{(N)} = \mathcal{C}_1 \star \mathcal{C}_N \star \cdots \star \mathcal{C}_N .$$

Theorem 2.7 (Sub-normalization condition) *Let $\mathcal{R}^{(N)}$ be a probabilistic Quantum Network. and $R^{(N)} \in \mathcal{L}(\bigotimes_{i=0}^{2N-1} \mathcal{H}_i)$ be its Choi-Jamiolkowski operator; then there exists a Choi operator $S^{(N)}$ of a deterministic Quantum Network such that*

$$0 \leqslant R^{(N)} \leqslant S^{(N)}, \tag{2.64}$$

Proof. The proof is by induction. For $N = 1$ the probabilistic quantum Network is just a quantum operation and we know that its Choi operator $E^{(1)}$ is upper bounded by the Choi operator of a Quantum Channel, i.e. of a deterministic Quantum Network with 1 vertex. Now suppose that the statement holds for $N - 1$. Since $\mathcal{R}^{(N)} = \mathcal{C}_1 \star \mathcal{C}_N \star \cdots \star \mathcal{C}_N$ we have $R^N = C_1 * C_2 * \cdots * C_N$ where $C_i \leqslant \overline{C_i}$ for some $\overline{C_i}$ which is the Choi operator of a quantum channel. Exploiting the induction hypothesis we have that $C_1 * C_2 * \cdots * C_{N-1} := D \leqslant \overline{D}$ where is the Choi of a deterministic Quantum Network. Exploiting Lemma 2.4 we have

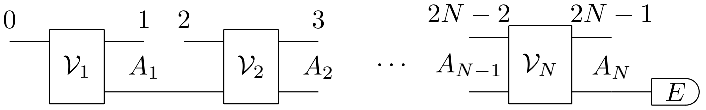
$$R^{(N)} = D * C_N \leqslant \overline{D} * C_N \leqslant \overline{D} * \overline{C_N} := S^{(N)} \tag{2.65}$$

that proves the statement. ■

[7] This definition includes deterministic networks as a special case.

Theorem 2.8 (Realization of probabilistic Quantum Networks) *Let $R^{(N)} \in \mathcal{L}(\bigotimes_{i=0}^{2N-1}\mathcal{H}_i)$ be a positive operator satisfying Eq. (2.64). Than this is the Choi-Jamiołkowsky operator of a probabilistic Quantum Network $\mathcal{R}^{(N)}$, consisting of N isometric channels followed by an effect on an ancillary space. For any $\rho \in \mathcal{L}(\mathcal{H}_{\mathrm{in}})$ we have*

$$\mathcal{R}^{(N)}(\rho) = \mathrm{Tr}_{A_N}[(V^{(N)}\cdots V^{(1)})\rho(V^{(1)\dagger}\cdots V^{(N)\dagger})E] \tag{2.66}$$



where $V^i \in \mathcal{L}(\mathcal{L}(\mathcal{H}_{2k-2}\otimes\mathcal{H}_{A_{k-1}}), \mathcal{L}(\mathcal{H}_{2k-1}\otimes\mathcal{H}_{A_k}))$ and $\mathcal{H}_{A_k}$ is an ancillary space, $\mathcal{H}_{A_0} = \mathbb{C}$ (in Eq. (2.66) we omitted the identity operators on the Hilbert spaces where the $V^{(k)}$'s and E do not act).

Proof. Let $S^{(N)}$ be the Choi operator of a deterministic Quantum Network such that $R^{(N)} \leqslant S^{(N)}$. Now we define $\mathcal{H}_{A_k}$ and $V^{(K)}$ for $S^{(N)}$ as in Eq. (2.62) and $E = S^{(N)*-\frac{1}{2}}R^{(N)*}S^{(N)*-\frac{1}{2}}$; It is easy to verify that

$$\begin{aligned}
&\mathrm{Tr}_{A_N}[(V^{(N)}\cdots V^{(1)})\rho(V^{(1)\dagger}\cdots V^{(N)\dagger})E] = \\
&= \mathrm{Tr}_{A_N}[(I_{\mathrm{out}}\otimes(S^{(N)*})^{\frac{1}{2}})(\rho_{\mathrm{in}'}\otimes|I\rangle\!\rangle\langle\!\langle I|_{(\mathrm{out})(\mathrm{out})'})(I_{\mathrm{out}}\otimes(S^{(N)*})^{\frac{1}{2}})\cdot \\
&\cdot(I_{\mathrm{out}}\otimes S^{(N)*-\frac{1}{2}}R^{(N)*}S^{(N)*-\frac{1}{2}})] = \\
&= \mathrm{Tr}_{A_N}[(I_{\mathrm{out}}\otimes\rho_{\mathrm{in}}^T)R^{(N)}] = \mathcal{R}^{(N)}(\rho)
\end{aligned}$$

Remark 2.9 *Theorem 2.8 says that any probabilistic Quantum Network can be split into a coherent part (sequence of isometries) and a final effect on an ancillary space.*

Thanks to theorems 2.7 and 2.8 we can represent any probabilistic Quantum Network in terms of a positive operator i.e. its probabilistic Quantum Comb. We now introduce the Quantum Network analogue of Quantum Instruments and POVMs; both of them will be exploited in the applications.

Definition 2.3 (Generalized Instrument) *A* Generalized Instrument *is a set of probabilistic Quantum Networks $\{\mathcal{R}_i^{(N)}\}$ whose sum is a deterministic Quantum Network $\mathcal{R}_\Omega^{(N)} = \sum_i \mathcal{R}_i^{(N)}$. The index i represents the classical outcome of the Network*[8].

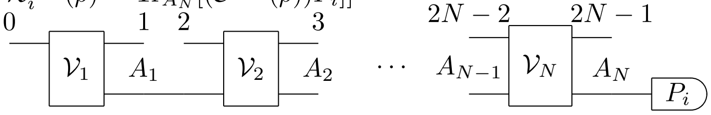
For Generalized Instruments the following analogue of Th. 2.4 holds:

Theorem 2.9 (realization of Generalized Instruments) *Let $\{\mathcal{R}_i^{(N)}, \mathcal{R}_i^{(N)} \in \mathcal{L}(\mathcal{L}(\mathcal{H}_{\mathrm{in}}), \mathcal{H}_{\mathrm{out}})\}$, $\mathcal{R}_\Omega^{(N)} = \sum_i \mathcal{R}_i^{(N)}$ be a Generalized Instrument. Then there*

[8] As we did when we introduced the concept of Quantum Instrument, we restrict ourselves to the case of finite number of outcomes. The generalization to an arbitrary outcome space Ω can be obtained by defining a measure $\mathcal{R}_B$ that associates to any event $B \subseteq \Omega$ a probabilistic Quantum Network $\mathcal{R}_B$ such that $\mathcal{R}_\Omega$ is a deterministic Quantum Network.

exist an Hilbert space $\mathcal{H}_{A_N}$*, a deterministic Quantum Network* $\mathcal{S}^{(N)} \in \mathcal{L}(\mathcal{L}(\mathcal{H}_{\text{in}}), \mathcal{L}(\mathcal{H}_{\text{out}} \otimes \mathcal{H}_{A_N}))$ *and a POVM* $\{P_i, P_i \in \mathcal{L}(\mathcal{H}_{A_N})\}$ *such that for any* $\rho \in \mathcal{L}(\mathcal{H}_{\text{in}})$ *we have*

$$\mathcal{R}_i^{(N)}(\rho) = \text{Tr}_{A_N}[(\mathcal{S}^{(N)}(\rho))P_i]] \tag{2.67}$$



Proof. The Proof is the same as in Th. 2.8; we just define $\mathcal{S}^{(N)} = \mathcal{V}_1 \star \cdots \star \mathcal{V}_N$, where the $\mathcal{H}_{A_k}$'s and $V^{(i)}$'s are defined as in Eq. (2.62) (now $R_\Omega^{(N)}$ plays the role of $R^{(N)}$) and

$$P_i = R_\Omega^{(N)-\frac{1}{2}*} R_i^{(N)} R_\Omega^{(N)-\frac{1}{2}*}. \tag{2.68}$$

It is easy to verify that

$$\begin{aligned} &\text{Tr}_{A_N}[(\mathcal{S}^{(N)}(\rho))P_i] = \text{Tr}_{A_N}[(V^{(N)} \cdots V^{(1)})\rho(V^{(1)\dagger} \cdots V^{(N)\dagger})P_i] = \\ &= \text{Tr}_{A_N}[(I_{\text{out}} \otimes (R_\Omega^{(N)*})^{\frac{1}{2}})(\rho_{\text{in}'} \otimes |I\rangle\rangle\langle\langle I|_{(\text{out})(\text{out})'})(I_{\text{out}} \otimes (R_\Omega^{(N)*})^{\frac{1}{2}}) \cdot \\ &\cdot (I_{\text{out}} \otimes R_\Omega^{(N)*-\frac{1}{2}} R_i^{(N)*} R_\Omega^{(N)*-\frac{1}{2}})] = \\ &= \text{Tr}_{A_N}[(I_{\text{out}} \otimes \rho_{\text{in}}^T) R_i^{(N)}] = \mathcal{R}_i^{(N)}(\rho) \end{aligned} \tag{2.69}$$

■

A relevant class of Generalized Instrument is the the following

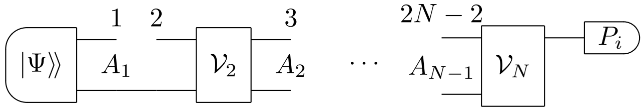
Definition 2.4 (Quantum Tester) *A* Quantum Tester *is a Generalized Instrument* $\{\mathcal{R}_i^N\}$ *such that* $\dim(\mathcal{H}_0) = \dim(\mathcal{H}_{2N-1}) = 1$.

Theorem 2.10 (normalization of Quantum Tester) *Let* $\{\mathcal{R}_i^{(N)}\}$ *be a quantum Tester. Then*

$$\begin{aligned} &\sum_i R_i^{(N)} := R_\Omega^{(N)} = R_\Omega^{(N-1)} \otimes I_{2N-2} \\ &\text{Tr}_{2k-1}[R_\Omega^{(k)}] = I_{2k-2} \otimes R_\Omega^{(k-1)}, \quad 2 \leqslant k \leqslant N-1 \\ &\text{Tr}_1[R_\Omega^{(1)}] = 1 \end{aligned} \tag{2.70}$$

Proof. Since $\dim(\mathcal{H}_{2N-1}) = 1$ and applying Theorem 2.5 to $R_\Omega^{(N)}$ we have $R^{(N)} = \text{Tr}_{2N-1}[R^{(N)}] = I_{2N-1} \otimes R_\Omega^{(N-1)}$. Clearly $\text{Tr}_1[R_\Omega^{(1)}] = I_0 = 1$ since $\dim(\mathcal{H}_0) = 1$.

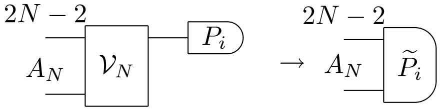
Theorem 2.11 (realization of Quantum Tester) *Let* $\{\mathcal{R}_i^{(N)}\}$ *be a quantum Tester. Then* $\{\mathcal{R}_i^{(N)}\}$ *can be realized by a deterministic Quantum Network* $\{\mathcal{S}^{(N)}\}$ *with* $\dim(\mathcal{H}_0) = 1$ *followed by a POVM on* $\mathcal{H}_{2N-1}$



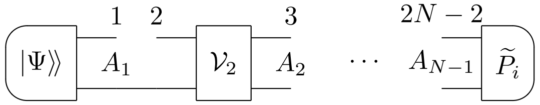
(2.71)

Proof. This result comes immediately from theorem 2.9 by relabeling $\mathcal{H}_{A_N} = \mathcal{H}_{2N-1}$. Since $\dim(\mathcal{H}_0) = 1$ the first isometry is just the preparation of the pure state $|\Psi\rangle\!\rangle := (I_1 \otimes R_\Omega^{(1)*\frac{1}{2}})|I\rangle\!\rangle_{11'} = |R_\Omega^{(1)*\frac{1}{2}}\rangle\!\rangle$. ■

Remark 2.10 *By making the substitution*



the realization scheme 2.71 can be rewritten as



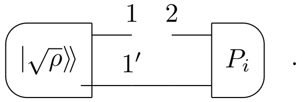
(2.72)

A special class of Quantum Testers is the one in which $N = 2$; this class has been independently introduced in Ref. [41] under the name Process-POVM.

Corollary 2.4 (characterization of Quantum 2-Testers) *Let* $\{\mathcal{R}_i^{(2)}, \mathcal{R}_i^{(2)} \in \mathcal{L}(\mathcal{L}(\mathcal{H}_1), \mathcal{L}(\mathcal{H}_2))\}$ *be a Quantum Tester with two vertexes. Then we have*

$$\sum_i R_i^{(2)} = \rho_1 \otimes I_2 \tag{2.73}$$

where ρ *is a state in* $\mathcal{L}(\mathcal{H}_1)$. $\{\mathcal{R}_i^{(2)}\}$ *can be split into a preparation of a pure state* $|\sqrt{\rho}\rangle\!\rangle \in \mathcal{H}_1 \otimes \mathcal{H}_{1'}$ *and a POVM* $\{P_i\}$ *on the space* $\mathcal{H}_2 \otimes \mathcal{H}_{1'}$ *(*$\mathcal{H}_{1'} = \mathsf{Supp}(\rho)$*)*



(2.74)

Proof. Eq. (2.73) comes from from Eq. (2.60) and Eq. (2.70). The realization (2.74) is just a special case of (2.72) with $A_1 = 1'$.

2.7 Connection of Quantum Networks

A Network of Quantum transformations can be used to achieve many different tasks. We can imagine to use it as a programmable device which implements different transformations on some inputs depending on the quantum state of the program (see Fig. 2.4). Moreover, the program itself of the Quantum Network can be a quantum channel rather then a state (Fig. 2.5): during the computation the network call a variable channel as a subroutine. More generally a Quantum Network can call several different channels at different times and even another Quantum Network. These kind of situation occur for example when multiple round Quantum games are considered; in this scenario the overall outcome of the game depends on the strategies chosen by the players that can be modeled as Quantum Networks (Fig. 2.6).

Another relevant case are Quantum Algorithms: they can be thought of as Quantum Networks calling N uses of a quantum oracle (Fig. 2.7). All the possible uses of a Quantum Network are

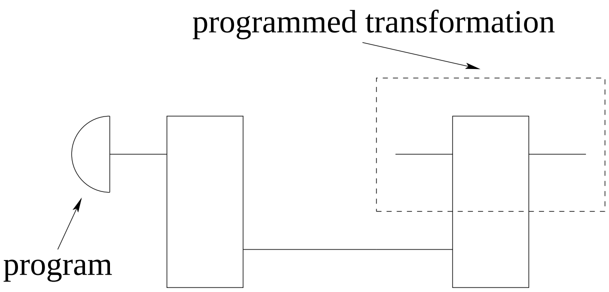


Figure 2.4. A Quantum Network with two vertexes used as a programmable device.

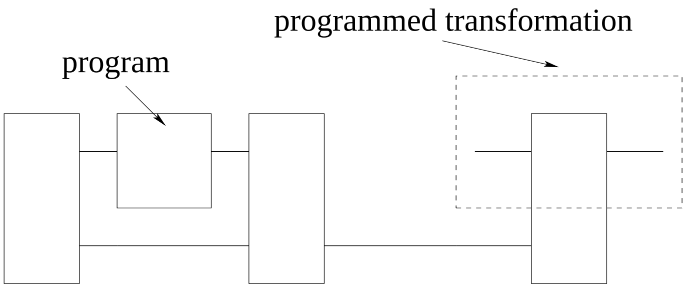
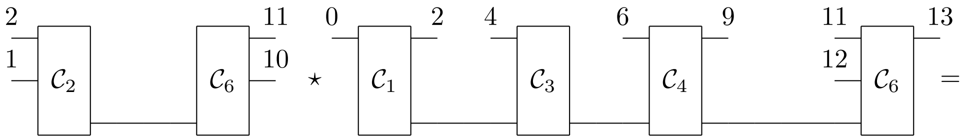


Figure 2.5. A Quantum Network calls a quantum channel as subroutine.

then equivalent to the connection of the network to another quantum network. Connecting two network $\mathcal{R}^{(N)}$ and $\mathcal{S}^{(M)}$ means composing the corresponding graphs by joining some of the free outgoing arrows of a network with free incoming arrows of the other in such a way that the final network $\mathcal{R}^{(N)} \star \mathcal{S}^{(M)}$ is still a directed acyclic graph[9]; we adopt the convention that if two vertexes $i \in \mathcal{R}^{(N)}$ and $j \in \mathcal{S}^{(M)}$ are connected by joining two arrows, the two arrows are identified with the same label (see Fig. 2.8). As we said in Section 2.4, a directed acyclic graph is endowed with a partial ordering among the vertexes that can be extended to a total ordering. Given two quantum networks $\mathcal{R}^N$ and $\mathcal{S}^M$ there is a priori no relative ordering between the vertexes of $\mathcal{R}^N$ and the vertexes of $\mathcal{S}^M$. However, since we require that the final network is still a directed acyclic graph, it is possible to define a total ordering among the vertexes in the union set $\mathcal{R}^N \cup \mathcal{S}^M$. This allows us to sketch the composition of two quantum networks in the circuit form



[9] As pointed out in Remark 2.4 this condition is necessary in order to avoid time loops

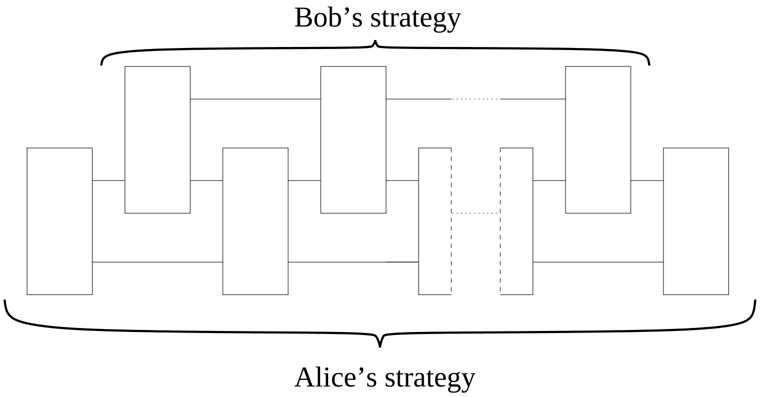


Figure 2.6. A multi-round two party game: Alice’s strategy is represented by the Quantum Network $\mathcal{A}$ and Bob’s strategy is represented by the Quantum Network $\mathcal{B}$. The outcome of the game can be seen as th interlinking of the two networks.

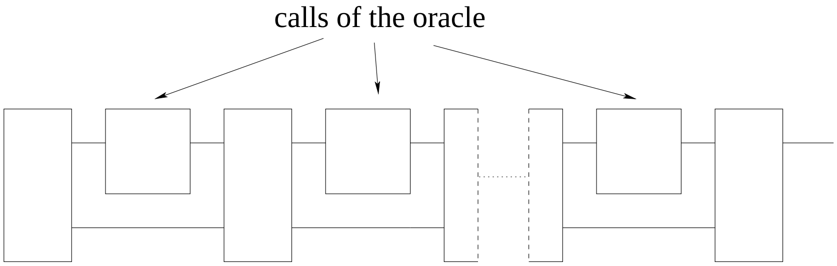
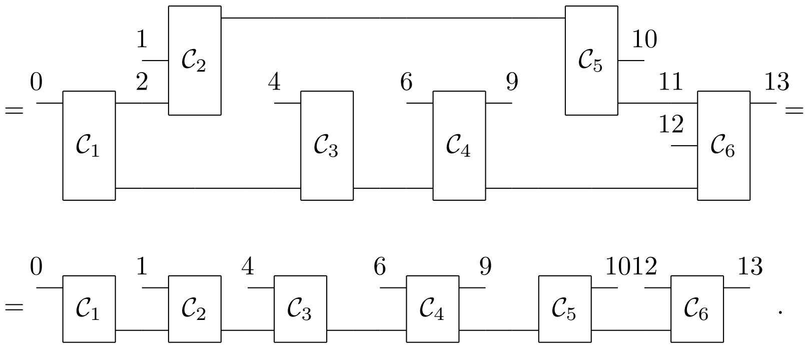


Figure 2.7. A Quantum algorithm realized by a Quantum Network in which N uses of the oracle are inserted.



(2.75)

We now want to determine the Choi operator of the composite network $\mathcal{R}^{(N)} \star \mathcal{S}^{(M)}$ in terms of the Choi operators $R^{(N)}$ and $S^{(M)}$ of the networks $\mathcal{R}^{(N)}$ and $\mathcal{S}^{(M)}$. From Eq. (2.75) it is clear that the combined network can be obtained by combining the linear maps $\mathcal{C}_i$, then its Choi operator will be the link product of all the C_i. We have then the following

Theorem 2.12 (Link of two Quantum Networks) *Let $\mathcal{R}^{(N)}$ and $\mathcal{S}^{(M)}$ be two Quantum Networks and $R^{(N)} \in \mathcal{L}(\bigotimes_{i\in\mathsf{R}} \mathcal{H}_i)$, $S^{(M)} \in \mathcal{L}(\bigotimes_{j\in\mathsf{S}} \mathcal{H}_j)$ be their Choi operators where we defined R and S the set of the free arrows of $\mathcal{R}^{(N)}$ and $\mathcal{S}^{(M)}$ respectively. If $\mathsf{R} \cap \mathsf{S}$ is the set of connected arrows then*

$$\mathfrak{C}(\mathcal{R}^{(N)} \star \mathcal{S}^{(M)}) = R^{(N)} * S^{(M)} \tag{2.76}$$

Proof. This result is an immediate consequence of Lemma 2.4.

Remark 2.11 *A relevant case of composition is the one in which we connect a quantum network*

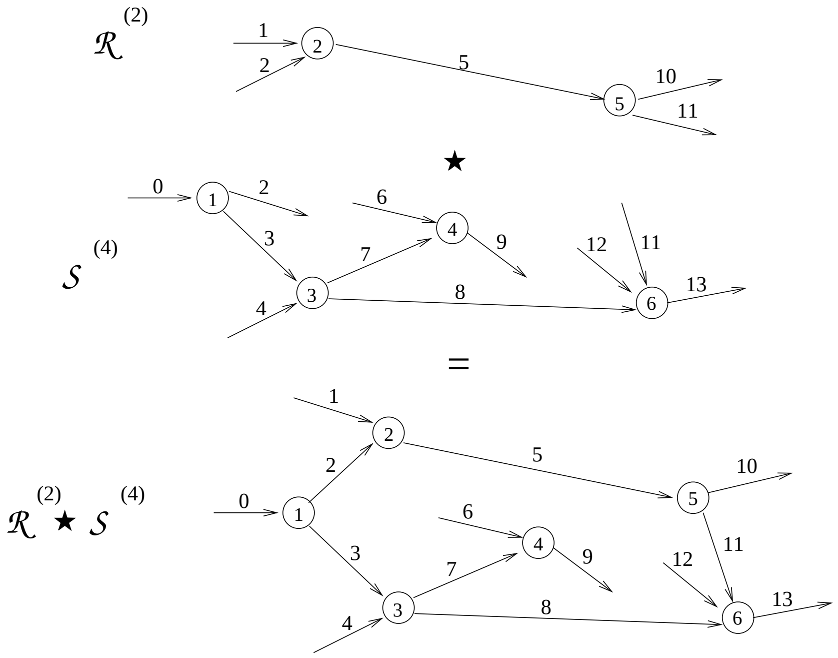
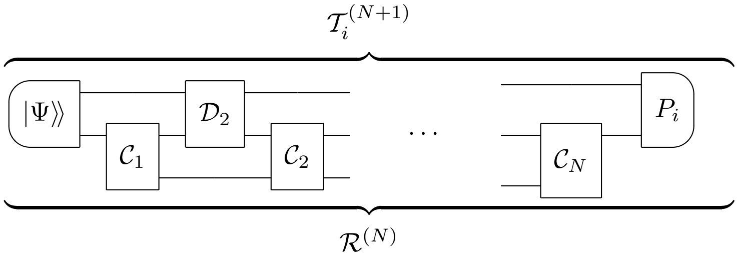


Figure 2.8. The scheme represents the connection of two quantum networks; the arrows that we are going to connect have the same labels.

$\mathcal{R}^{(N)}$ with a quantum tester $\{\mathcal{T}_i^{(N+1)}\}$ in this way:



(2.77)

The composite network $\mathcal{R}^{(N)} \star \mathcal{T}_i^{(N+1)}$ has only a classical outcome, i.e. the the index i. The link product $R^{(N)} * T_i^{(N+1)}$ gives the probability to obtain output i:

$$p(i|\mathcal{R}^{(N)}) = R^{(N)} * T_i^{(N+1)} = \mathrm{Tr}[R^{(N)} T_i^{(N+1)T}] \tag{2.78}$$

Eq. (2.78) can be interpreted as a generalized version of the Born rule: $R^{(N)}$ plays the role of a quantum state while the set $\{T_i^{(N+1)T}\}$ is the analogue of a POVM. A quantum tester represents the most general measurement process we can perform on a Quantum Network; Eq. (2.78) tells us that two Quantum Networks $\mathcal{R}^{(N)}$ and $\mathcal{S}^{(N)}$ that have the same Choi Jamiołkowski operator, give the same probability distribution for all testers $\mathcal{T}^{(N+1)}$: this means that $\mathcal{R}^{(N)}$ and $\mathcal{S}^{(N)}$ are experimentally indistinguishable.

3 Quantum Tomography

Calibration of physical devices is the basis of any experimental procedure, especially in quantum information, where the reliability of the processes involved in the computation is crucial. *Quantum Tomography* is the complete determination of physical devices in a purely experimental manner (by relying on some well established measurement instruments), without using detailed theoretical knowledge of its inner functioning. Originally introduced for determine the quantum state of radiation [42, 43, 44], Quantum Tomography soon became the standard technique in the measuring the fine details of any quantum device. In this chapter we will present a systematic theoretical approach to optimization of Quantum Tomography of finite dimensional systems, as it was introduced in [24, 25]. The optimization of a tomographic procedure involves two aspects: i) optimization of the experimental setup and ii) optimization of the data processing, that is the classical processing of the measurement outcomes. Our approach is based on the notion of *informationally complete measurement* [45]. The optimization of the data processing [46, 47] relies on the fact that the operators describing an informationally complete measurement are generally linearly dependent, thus allowing different expansions coefficients. For state tomography the optimization of the setup consists in finding the best informationally complete POVMs. However, when the more general scenario of quantum process tomography is considered, the optimization problem involves the choice of the input state as well (we are in the framework of the so called *ancilla assisted process tomography* [48, 49]); for this reason we will take advantage of the general theory of Quantum Networks that will allow us to optimize both the input state and final POVM at the same time.

We will begin by introducing Quantum Tomography of states and the key concepts that are needed in order to cope with the optimization. Then, thanks to the tools developed in Chapter 2 we will generalize this setting from quantum states to Quantum Networks. Finally, we will provide the optimal scheme for Quantum Tomography of states, channels and POVMs.

3.1 State tomography

Tomographing an unknown state ρ of a quantum system means performing a suitable POVM $\{P_i\}$ in such a way that ρ is completely determined by the probability distribution

$$p_i = \mathrm{Tr}[\rho P_i]. \tag{3.1}$$

Completely determining a quantum state means being able to predict the expectation value $\langle A \rangle = \mathrm{Tr}[\rho A]$ for any operator A, in terms of the probabilities p_i, i.e.

$$\langle A \rangle = \mathrm{Tr}[\rho A] = \sum_i p_i f(i, A) \qquad \forall A, \rho \in \mathcal{L}(\mathcal{H}) \tag{3.2}$$

where $f(i, A)$ denotes suitable expansion coefficients[10]. The function $f : (i, A) \mapsto f(i, A)$ is called *data processing* since it represents the processing of the outcomes i of the measurement $\{P_i\}$ in order to recover $\langle A \rangle$

[10] we assumed a linear reconstruction of the expectation value, that is we are considering *linear quantum tomography.*

From Eq. (3.2) we get:

$$\begin{aligned}\mathrm{Tr}[\rho A] \qquad &= \sum_i p_i f_i[A] = \sum_i \mathrm{Tr}[\rho P_i] f_i[A] = \\ &= \mathrm{Tr}\left[\rho \sum_i f_i[A] P_i\right] \qquad \forall A, \rho \quad \Leftrightarrow \quad A = \sum_i f_i[A] P_i \quad \forall A \end{aligned} \tag{3.3}$$

that is it is possible to expand any A over the used POVM $\{P_i\}$. When the expansion (3.3) holds for for all the operators in $\mathcal{L}(\mathcal{H})$, we have that $\mathrm{Span}\{P_i\} = \mathcal{L}(\mathcal{H})$ and we say that the POVM $\{P_i\}$ is *informationally complete*. Informationally completeness of the POVM is equivalent to the condition [50, 51]

$$\mathsf{Supp}(F) = \mathcal{H} \otimes \mathcal{H} \qquad F = \sum_i |P_i\rangle\!\rangle\langle\!\langle P_i| \tag{3.4}$$

where we exploit the isomorphism (2.2). A set of vectors $|v_i\rangle \in \mathcal{H}$ such that $\mathsf{Supp}(F) = \mathcal{H}$, $F = \sum_i |v_i\rangle \langle v_i|$ is called *frame*[11] and the operator F is called *frame operator*. Given a frame $\{|v_i\rangle\}$ it is possible to introduce a set of vectors $\{|u_i\rangle\}$, called *dual frame*, such that

$$\sum_i |v_i\rangle \langle u_i| = I. \tag{3.5}$$

If the $|v_i\rangle$ are linearly dependent the dual frame $\{|u_i\rangle\}$ is not unique. The expansion (3.3) can be rephrased in terms of the $|P_i\rangle\!\rangle$ in the following way:

$$|A\rangle\!\rangle = \sum_i f(i, A) |P_i\rangle\!\rangle. \tag{3.6}$$

and if we introduce a dual frame $|D_i\rangle\!\rangle$ ($\sum_i |P_i\rangle\!\rangle\langle\!\langle D_i| = I$) we have

$$\begin{aligned}&|A\rangle\!\rangle = \left(\sum_i |P_i\rangle\!\rangle\langle\!\langle D_i|\right)|A\rangle\!\rangle = \sum_i \langle\!\langle D_i|A\rangle\!\rangle |P_i\rangle\!\rangle \quad \Rightarrow \quad f_i[A] = \langle\!\langle D_i|A\rangle\!\rangle \\ &\langle A\rangle = \mathrm{Tr}[\rho A] = \langle\!\langle \rho|A\rangle\!\rangle = \sum_i \langle\!\langle D_i|A\rangle\!\rangle\langle\!\langle \rho|P_i\rangle\!\rangle\end{aligned} \tag{3.7}$$

We requested that the POVM has to be informationally complete because we have no prior information about the state ρ of the system, i.e. ρ can be an arbitrary normalized positive operator in $\mathcal{L}(\mathcal{H})$. However, we can suppose that the state ρ belongs to a given subspace $\mathcal{A} \subseteq \mathcal{L}(\mathcal{H})$; in this case the only operators we need to expand are the ones in $\mathcal{A}$ since $\mathrm{Tr}[A'\rho] = 0$ for all $A' \in \mathcal{A}^\perp$. Then the set $\{P_i\}$ is required to span only $\mathcal{A}$. Exploiting the isomorphism (2.2), if $\rho \in \mathcal{A} \subseteq \mathcal{L}(\mathcal{H})$ and $\mathrm{Span}\{P_i\} = \mathcal{A}$, we have that $|\rho\rangle\!\rangle \in \mathcal{V}_\mathcal{A}$, where we defined $\mathcal{H} \otimes \mathcal{H} \supseteq \mathcal{V}_\mathcal{A} := \mathrm{Span}\{|P_i\rangle\!\rangle\}$. If we denote with $Q_\mathcal{A}$ the projector on $\mathcal{V}_\mathcal{A}$ then Eq. (3.7) becomes

$$\begin{aligned}&|A\rangle\!\rangle = \left(\sum_i |P_i\rangle\!\rangle\langle\!\langle D_i|\right)|A\rangle\!\rangle = \sum_i \langle\!\langle D_i|A\rangle\!\rangle |P_i\rangle\!\rangle \quad \Rightarrow \quad f_i[A] = \langle\!\langle D_i|A\rangle\!\rangle \\ &\langle A\rangle = \sum_i \langle\!\langle D_i|Q_\mathcal{A}|A\rangle\!\rangle\langle\!\langle \rho|Q_\mathcal{A}|P_i\rangle\!\rangle.\end{aligned} \tag{3.8}$$

[11] in this presentation we are restricting ourselves to the finite dimensional case.

The condition that the POVM spans the subspace $\mathcal{A}$ can be rephrased in terms of the corresponding frame operator; it is possible to prove that

$$\mathrm{Span}\{P_i\} = \mathcal{A} \Leftrightarrow \mathsf{Supp}(F) = \mathcal{V}_{\mathcal{A}}. \tag{3.9}$$

First we notice that Eq. (3.9) can be rephrased as

$$\mathrm{Span}\{|P_i\rangle\!\rangle\} = \mathsf{Supp}(F); \tag{3.10}$$

we will verify both the inclusions $\mathrm{Span}\{|P_i\rangle\!\rangle\} \subseteq \mathsf{Supp}(F)$ and $\mathrm{Span}\{|P_i\rangle\!\rangle\} \supseteq \mathsf{Supp}(F)$. Since any vector $|X\rangle\!\rangle \in \mathcal{H} \otimes \mathcal{H}$ can be decomposed as $|X\rangle\!\rangle = |Y\rangle\!\rangle + |Z\rangle\!\rangle$ where $|Y\rangle\!\rangle \in \mathcal{V}_{\mathcal{A}}$ and $|Z\rangle\!\rangle \in \mathcal{V}_{\mathcal{A}}^{\perp}$ ($\mathcal{V}_{\mathcal{A}} = \mathrm{Span}\{|P_i\rangle\!\rangle\}$), we have

$$\begin{aligned} &F|X\rangle\!\rangle = \sum_i |P_i\rangle\!\rangle\langle\!\langle P_i|(|Y\rangle\!\rangle + |Z\rangle\!\rangle) = \sum_i |P_i\rangle\!\rangle\langle\!\langle P_i||Y\rangle\!\rangle = 0 \Rightarrow \\ &\Rightarrow \sum_i |\langle\!\langle P_i|Y\rangle\!\rangle|^2 = 0 \Rightarrow \langle\!\langle P_i|Y\rangle\!\rangle = 0 \ \ \forall i \Rightarrow |Y\rangle\!\rangle = 0 \Rightarrow \\ &\Rightarrow |X\rangle\!\rangle \in \mathcal{V}_{\mathcal{A}}^{\perp} \Rightarrow \mathrm{Ker}(F) \subseteq (\mathrm{Span}\{|P_i\rangle\!\rangle\})^{\perp} \Rightarrow \mathrm{Span}\{|P_i\rangle\!\rangle\} \subseteq \mathsf{Supp}(F). \end{aligned}$$

On the other hand, let F^{-1} be the inverse of F on its support; since $F^{\dagger} = F$ we have $F^{-1}F = I_{\mathsf{Supp}(F)} = FF^{-1}$; then it follows

$$\begin{aligned} &|X\rangle\!\rangle \in \mathsf{Supp}(F) \Rightarrow |X\rangle\!\rangle = F^{-1}F|X\rangle\!\rangle = FF^{-1}|X\rangle\!\rangle = \sum_i \langle\!\langle P_i|F^{-1}|X\rangle\!\rangle|P_i\rangle\!\rangle = \\ &= \sum_i c_i|P_i\rangle\!\rangle \Rightarrow |X\rangle\!\rangle \in \mathrm{Span}\,|P_i\rangle\!\rangle \Rightarrow \mathsf{Supp}(F) \subseteq \mathrm{Span}\{|P_i\rangle\!\rangle\}. \end{aligned}$$

We now need a criterion that quantifies how well our tomographic procedure estimates the expectation $\langle A\rangle$ of an observable A. As we have previously shown, a tomographic procedure involves two steps:

- the measurement process which is described by the infocomplete POVM $\{P_i\}$ or equivalently by the frame $|P_i\rangle\!\rangle$;
- the processing of the outcomes which is described by the dual $|D_i\rangle\!\rangle$.

That being so, the optimization problem consists in finding the best POVM $\{P_i\}$ and the best dual $|D_i\rangle\!\rangle$ according to a given figure of merit. Suppose now that the POVM is fixed and that every repetition of the experiment is independent; if the experimental frequencies are $\nu_i := \frac{n_i}{N}$ (n_i is the number of outcomes i occurred, and N is the total number of repetitions), the estimated expectation $\widetilde{\langle A\rangle}$ is then

$$\widetilde{\langle A\rangle} = \sum_i f(i,A)\nu_i \rightsquigarrow \langle A\rangle \tag{3.11}$$

where the symbol $\rightsquigarrow$ means that, by the law of large numbers, the left hand side converges in probability to the right hand side. A good figure of merit for the data processing strategy is the statistical error in the reconstruction of expectations, i.e. the variance of the random variable $\widetilde{\langle A\rangle}$. Since the variance of the mean is proportional to the variance of the distribution [52], the statistical error occurring when the processing in Eq. (3.11) is used, can be written as:

$$\delta(A) := \sum_i |f(i,A) - \langle A\rangle|^2 \nu_i \tag{3.12}$$

Averaging the statistical error over all possible experimental outcomes we have

$$\begin{aligned}\overline{\delta(A)} &:= \sum_k \left(\sum_i |f(i,A) - \langle A\rangle|^2 \nu_i^{(k)} \right) \mathbf{m}_k = \\ &= \sum_i |f(i,A) - \langle A\rangle|^2 \left(\sum_k \nu_i^{(k)} \mathbf{m}_k \right) = \sum_i |f(i,A) - \langle A\rangle|^2 p_i \end{aligned} \tag{3.13}$$

where the index k labels different experimental outcomes (i.e. a possible set of frequencies) and $\mathbf{m}_k$ is the multinomial distribution

$$\mathbf{m}_k = \frac{N!}{\prod_l n_l^{(k)}!} \prod_l p_l^{N\nu_l^{(k)}} \tag{3.14}$$

that gives the probability that the experiment gives the frequencies $\{\nu_l^{(k)}\}$ for each outcome l. In terms of ρ, P_i and D_i Eq. (3.13) becomes

$$\begin{aligned}\overline{\delta(A)} &= \sum_i |f(i,A) - \langle A\rangle|^2 p_i = \sum_i |\langle\langle D_i|A\rangle\rangle - \langle\langle \rho|A\rangle\rangle|^2 \langle\langle \rho|P_i\rangle\rangle = \\ &= \sum_i |\langle\langle D_i|A\rangle\rangle|^2 \langle\langle \rho|P_i\rangle\rangle - |\langle\langle \rho|A\rangle\rangle|^2; \end{aligned} \tag{3.15}$$

where we used Eq. (3.7) in the last equality. In a Bayesian scheme the state ρ is assumed to be randomly drawn from an ensemble $\mathcal{S} = \{\rho_n, p_n\}$ of state ρ_n with prior probability p_n. If we average the quantity $\overline{\delta(A)}$ over $\mathcal{S}$ we get

$$\begin{aligned}\overline{\delta(A)}_{\mathcal{S}} &:= \sum_n \left(\sum_i |\langle\langle D_i|A\rangle\rangle|^2 \langle\langle \rho_n|P_i\rangle\rangle - |\langle\langle \rho_n|A\rangle\rangle|^2 \right) p_n = \\ &= \sum_i |\langle\langle D_i|A\rangle\rangle|^2 \langle\langle \rho_{\mathcal{S}}|P_i\rangle\rangle - \sum_n |\langle\langle \rho_n|A\rangle\rangle|^2 p_n \end{aligned} \tag{3.16}$$

where $\rho_{\mathcal{S}} = \sum_n p_n \rho_n$. Moreover, a priori we can be interested in some observables more than other ones, and this can be specified in terms of a weighted set of observables $\mathcal{G} = \{A_m, q_m\}$, with weight $q_m > 0$ for the observables A_m. Averaging over $\mathcal{G}$ we have

$$\begin{aligned}\overline{\delta(A)}_{\mathcal{S},\mathcal{G}} &:= \sum_m \left(\sum_i |\langle\langle D_i|A_m\rangle\rangle|^2 \langle\langle \rho_{\mathcal{S}}|P_i\rangle\rangle - \sum_n |\langle\langle \rho_n|A_m\rangle\rangle|^2 p_n \right) q_m = \\ &= \sum_i \langle\langle D_i|G|D_i\rangle\rangle \langle\langle \rho_{\mathcal{S}}|P_i\rangle\rangle - \sum_{n,m} |\langle\langle \rho_n|A_m\rangle\rangle|^2 p_n q_m \end{aligned} \tag{3.17}$$

where $G = \sum_m q_m |A_m\rangle\!\rangle\langle\!\langle A_m|$. Since only the first term of Eq. (3.17) depends on P_i and D_i, the figure of merit is finally given by:

$$\eta := \sum_i \langle\!\langle D_i|G|D_i\rangle\!\rangle\langle\!\langle \rho_{\mathcal{S}}|P_i\rangle\!\rangle \tag{3.18}$$

If $\rho_n \in \mathcal{A}$ for all n then $Q_{\mathcal{A}}|\rho_n\rangle\!\rangle = |\rho_n\rangle\!\rangle$ for all n; then, reminding Eq. (3.8), Eq. (3.17) becomes

$$\eta = \sum_i \langle\!\langle D_i|Q_{\mathcal{A}} G Q_{\mathcal{A}}|D_i\rangle\!\rangle\langle\!\langle \rho_{\mathcal{S}}|Q_{\mathcal{A}}|P_i\rangle\!\rangle. \tag{3.19}$$

Then, the optimization problem consists in finding the POVM P_i and the dual D_i that minimize η. In the following section we generalize this scenario from quantum states to quantum networks.

4 Quantum Network Tomography

At the beginning of this chapter we said that Quantum Tomography consists in the determination of a physical device by means of experiments that produce classical information. If the physical device is a preparator of quantum system the experiments we can perform in order to determine its state are described by POVMs; on the other hand, if the physical device is a Quantum Network, the experiments are described by Quantum Testers, that are the generalization of the POVMs (see Remark 2.11). In analogy with what we did for the POVMs in the previous section it is possible to introduce *informationally complete tester*, that is a quantum tester $\{\Pi_i, \Pi_i \in \mathcal{L}(\otimes_{k=1}^{2N}\mathcal{H}_k)\}$ such that the probabilities $p_i = \mathrm{Tr}[\Pi_i^T R]$ are sufficient to completely characterize the (generally probabilistic) Quantum Network $\mathcal{R}$ (equivalently, to completely characterize its Choi operator $R \in \mathcal{L}(\otimes_{k=1}^{2N-2}\mathcal{H}_k)$). This condition can be rephrased by saying that the probabilities $p_i = \mathrm{Tr}[\Pi_i^T R]$ allow to evaluate $\mathrm{Tr}[TR]$ for all $T \in \mathcal{L}(\otimes_{k=1}^{2N-2}\mathcal{H}_k)$:

$$\mathrm{Tr}[TR] = \sum_i f(i,T) p_i = \sum_i f(i,T)\,\mathrm{Tr}[\Pi_i^T R]. \tag{4.1}$$

Following the same line as in Eq. (3.3) we can say that a tester $\{\Pi_i\}$ is informationally complete when

$$\mathrm{Span}\{\Pi_i^T\} = \mathcal{L}(\otimes_{k=1}^{2N-2}\mathcal{H}_k) \tag{4.2}$$

The following result proves that informationally complete testers actually exist

Theorem 4.1 (informationally complete quantum testers) *Let $\{P_i, P_i \in \mathcal{L}(\otimes_{k=1}^{2N-2}\mathcal{H}_k)\}$ be an informationally complete POVMs. Then the tester $\Pi_i = (d_1 d_2 \cdots d_{2N-2}^{-1} P_i^T$ is informationally complete.*

Proof. Since P_i is informationally complete we have $\mathrm{Span}\{P_i\} = \mathrm{Span}\{\Pi_i^T\} = \mathcal{L}(\otimes_{k=1}^{2N-2}\mathcal{H}_k)$. Then the set Π_i is informationally complete. Moreover $\sum_i \Pi_i = (d_1 d_2 \cdots d_{2N-2})^{-1} I$ and clearly $(d_1 d_2 \cdots d_{2N-2})^{-1} I$ satisfies Eq. (2.60). ■

The condition that $\{\Pi_i^T\}$ span the whole $\mathcal{L}(\otimes_{k=1}^{2N-2}\mathcal{H}_k)$ can be relaxed if we know that the Quantum Network $\mathcal{R}$ lies in a subspace $\mathcal{A}$ of $\mathcal{L}(\otimes_{i=k}^{2n}\mathcal{H}_k)$. A relevant case is the one in which we know that $\mathcal{R}$ is a deterministic network; in this case the set $\{\Pi_i\}$ is required to span only the subspace $\mathcal{D}$ spanned by deterministic combs $\mathcal{D} := \mathrm{Span}\{R | R \text{ satisfies Eq. (2.60)}\}$.

If $\{\Pi_i\}$ is an informationally complete tester the set $\{|\Pi_i\rangle\!\rangle\}$ is a frame and we can write the expansion

$$|T\rangle\!\rangle = \sum_i \langle\!\langle \Delta_i | T \rangle\!\rangle |\Pi_i\rangle\!\rangle \tag{4.3}$$

where we introduced the dual $|\Delta_i\rangle\!\rangle$. It is then straightforward to generalize Eq. (3.18)

$$\eta = \sum_i \langle\!\langle \Delta_i | G | \Delta_i \rangle\!\rangle \langle\!\langle R_{\mathcal{S}} | \Pi_i \rangle\!\rangle. \tag{4.4}$$

where we introduced an ensemble of quantum network $\mathcal{S} := \{R_n, p_n\}$ and a weighted set of observables $\mathcal{G} := \{T_m, q_n\}$, and we defined $R_{\mathcal{S}} = \sum_n p_n R_n$, $G = \sum_m q_m |T_m\rangle\!\rangle\langle\!\langle T_m|$.

If $R_n \in \mathcal{A}$ for all n it is possible to write an analogous of Eq. (3.19)

$$\eta = \sum_i \langle\langle \Delta_i | Q_{\mathcal{A}} G Q_{\mathcal{A}} | \Delta_i \rangle\rangle \langle\langle R_{\mathcal{S}} | Q_{\mathcal{A}} | \Pi_i \rangle\rangle \tag{4.5}$$

where $Q_{\mathcal{A}}$ is the projector on $\mathcal{V}_{\mathcal{A}}$ ($|R_n\rangle\rangle \in \mathcal{V}_{\mathcal{A}}$ for all n).

The analogy between Eqs. (4.4,4.5) and Eqs. (3.18,3.19) tells us that the optimization of Quantum state tomography and the optimization of Quantum network tomography consist in minimizing the same figure of merit; the only difference is that $\{\Pi_i\}$ is tester instead of a POVM and it must satisfy the constraint (2.70).

4.1 Optimal quantum tomography for states, effects and transformation

In this section we will show how to perform the optimization of quantum tomographic setups for (finite-dimensional) states, channels and effects, according to the figure of merit defined in Eqs. (4.4,4.5). As we pointed out in Section 3.1, optimizing quantum tomography can be divided in two main steps; the first optimization stage involves a fixed detector, and only regards the data processing, namely the choice of the dual Δ_i used to determine the expansion coefficients $f(i,T)$ for a fixed T. As we will prove in the following, the optimal dual Δ_i is independent of T, and only depends on the ensemble $\mathcal{S}$. The second stage consists in optimizing the detector, which is represented by a POVMs for the case of state tomography and by a Quantum 2-tester when the more general case of transformation is concerned.

Remark 4.1 *It is worth noting that the optimization of the* 2*-tester covers both the choice of the best input state for the transformation and the choice of the best final measurement. Even if at a first sight one could think to carry this two optimization separately, thanks to the general theory developed in Chapter 2, they can be rephrased as a single optimization problem over a set of suitably normalized positive operators.*

4.1.1 Optimization of data processing

In this section we provide the optimization of the dual frames (i.e. of the data processing) for the general case of quantum networks; this derivation is new and is a generalization of the one used in [47].

Let us fix the tomographing device, which is described by the frame $|\Pi_i\rangle\rangle$, and let us minimize Eq. (4.4) over the possible data processing strategies, i.e. over all the possible duals $\{|\Delta_i\rangle\rangle\}$. We notice that at this stage it is irrelevant whether Π_i is a quantum tester or a POVM. Let us introduce the operator

$$\mathcal{X} = \sum_i \frac{|\Pi_i\rangle\rangle\langle\langle \Pi_i|}{\langle\langle R_{\mathcal{S}} | \Pi_i \rangle\rangle} \tag{4.6}$$

Since $|\Pi_i\rangle\rangle$ is a frame, $F = \sum_i |\Pi_i\rangle\rangle\langle\langle\Pi_i|$ is invertible and then also $\mathcal{X}$ is invertible. We now introduce the set $\{|\widetilde{\Delta_i}\rangle\rangle\}$ defined as follows:

$$|\widetilde{\Delta_i}\rangle\rangle := \mathcal{X}^{-1} \frac{|\Pi_i\rangle\rangle}{\langle\langle R_{\mathcal{S}} | \Pi_i \rangle\rangle}. \tag{4.7}$$

It is easy to verify that $\{|\widetilde{\Delta_i}\rangle\!\rangle\}$ is a dual:

$$\sum_i |\widetilde{\Delta_i}\rangle\!\rangle\langle\!\langle \Pi_i| = \mathcal{X}^{-1}\sum_i \left(\frac{|\Pi_i\rangle\!\rangle\langle\!\langle \Pi_i|}{\langle\!\langle R_{\mathcal{S}}|\Pi_i\rangle\!\rangle}\right) = \mathcal{X}^{-1}\mathcal{X} = I. \tag{4.8}$$

Before proving that $\{|\widetilde{\Delta_i}\rangle\!\rangle\}$ is the optimal dual we need to prove the following lemma

Lemma 4.1 *Let $\{|\Pi_i\rangle\!\rangle\}$ be a frame and $\{|\widetilde{\Delta_i}\rangle\!\rangle\}$ be defined as in Eq. (4.7). Then, for any dual $\{|\Delta_i\rangle\!\rangle\}$ we have*

$$\sum_i \langle\!\langle R_{\mathcal{S}}|\Pi_i\rangle\!\rangle|\widetilde{\Delta_i}\rangle\!\rangle\langle\!\langle K_i| = 0 \tag{4.9}$$

where $|K_i\rangle\!\rangle = |\Delta_i\rangle\!\rangle - |\widetilde{\Delta_i}\rangle\!\rangle$.

Proof. For any dual $|\Delta_i\rangle\!\rangle$ we have $\sum_i |\Pi_i\rangle\!\rangle\langle\!\langle\Delta_i| = I$. Then, using Eq. (4.7) we have

$$\sum_i \langle\!\langle R_{\mathcal{S}}|\Pi_i\rangle\!\rangle|\widetilde{\Delta_i}\rangle\!\rangle\langle\!\langle K_i| = \sum_i \langle\!\langle R_{\mathcal{S}}|\Pi_i\rangle\!\rangle|\widetilde{\Delta_i}\rangle\!\rangle\langle\!\langle \Delta_i| - \sum_i \langle\!\langle R_{\mathcal{S}}|\Pi_i\rangle\!\rangle|\widetilde{\Delta_i}\rangle\!\rangle\langle\!\langle \widetilde{\Delta_i}| =$$
$$= \mathcal{X}^{-1}\sum_i |\Pi_i\rangle\!\rangle\langle\!\langle\Delta_i| - \mathcal{X}^{-1}\sum_i \frac{|\Pi_i\rangle\!\rangle\langle\!\langle\Pi_i|}{\langle\!\langle R_{\mathcal{S}}|\Pi_i\rangle\!\rangle}\mathcal{X}^{-1} = \mathcal{X}^{-1} - \mathcal{X}^{-1}\mathcal{X}\mathcal{X}^{-1} = 0$$

■

Theorem 4.2 (Optimal dual) *Let $\{|\Pi_i\rangle\!\rangle\}$ be a frame and $\{|\widetilde{\Delta_i}\rangle\!\rangle\}$ be defined as in Eq. (4.7). Then, for any dual $\{|\Delta\rangle\!\rangle\}$ we have*

$$\sum_i \langle\!\langle \Delta_i|G|\Delta_i\rangle\!\rangle\langle\!\langle R_{\mathcal{S}}|\Pi_i\rangle\!\rangle \geqslant \sum_i \langle\!\langle \widetilde{\Delta_i}|G|\widetilde{\Delta_i}\rangle\!\rangle\langle\!\langle R_{\mathcal{S}}|\Pi_i\rangle\!\rangle \tag{4.10}$$

i.e. the dual $\{|\widetilde{\Delta_i}\rangle\!\rangle\}$ minimizes Eq. (4.4)

Proof. *From Lemma 4.1 we have:*

$$0 = \mathrm{Tr}\left[G\left(\sum_i \langle\!\langle R_{\mathcal{S}}|\Pi_i\rangle\!\rangle|\widetilde{\Delta_i}\rangle\!\rangle\langle\!\langle K_i|\right)\right] = \sum_i \langle\!\langle K_i|G|\widetilde{\Delta_i}\rangle\!\rangle\langle\!\langle R_{\mathcal{S}}|\Pi_i\rangle\!\rangle =$$
$$= \sum_i \langle\!\langle \widetilde{\Delta_i}|G|K_i\rangle\!\rangle\langle\!\langle R_{\mathcal{S}}|\Pi_i\rangle\!\rangle.$$

It is now easy to verify that

$$\begin{aligned}\sum_i \langle\langle\Delta_i|G|\Delta_i\rangle\rangle\langle\langle R_{\mathcal{S}}|\Pi_i\rangle\rangle &= \sum_i (\langle\langle\widetilde{\Delta}_i| + \langle\langle K_i|)G(|\widetilde{\Delta}_i\rangle\rangle + |K_i\rangle\rangle)\langle\langle R_{\mathcal{S}}|\Pi_i\rangle\rangle = \\ &= \sum_i \langle\langle\widetilde{\Delta}_i|G|\widetilde{\Delta}_i\rangle\rangle\langle\langle R_{\mathcal{S}}|\Pi_i\rangle\rangle + \sum_i \langle\langle\widetilde{\Delta}_i|G|K_i\rangle\rangle\langle\langle R_{\mathcal{S}}|\Pi_i\rangle\rangle + \\ &+ \sum_i \langle\langle K_i|G|\widetilde{\Delta}_i\rangle\rangle\langle\langle R_{\mathcal{S}}|\Pi_i\rangle\rangle + \sum_i \langle\langle K_i|G|K_i\rangle\rangle\langle\langle R_{\mathcal{S}}|\Pi_i\rangle\rangle = \\ &= \sum_i \langle\langle\widetilde{\Delta}_i|G|\widetilde{\Delta}_i\rangle\rangle\langle\langle R_{\mathcal{S}}|\Pi_i\rangle\rangle + \sum_i \langle\langle K_i|G|K_i\rangle\rangle\langle\langle R_{\mathcal{S}}|\Pi_i\rangle\rangle \geqslant \\ &\geqslant \sum_i \langle\langle\widetilde{\Delta}_i|G|\widetilde{\Delta}_i\rangle\rangle\langle\langle R_{\mathcal{S}}|\Pi_i\rangle\rangle\end{aligned}$$

■

Corollary 4.1 *If $|\Delta_i\rangle\rangle$ is the optimal dual Eq. (4.4) can be rewritten as:*

$$\eta = \sum_i \langle\langle\Delta_i|G|\Delta_i\rangle\rangle\langle\langle R_{\mathcal{S}}|\Pi_i\rangle\rangle = \mathrm{Tr}[\mathcal{X}^{-1}G] \tag{4.11}$$

where $\mathcal{X}$ was defined in Eq. (4.6).

Proof. By making use of Eq. (4.7) we have:

$$\begin{aligned}&\sum_i \langle\langle\Delta_i|G|\Delta_i\rangle\rangle\langle\langle R_{\mathcal{S}}|\Pi_i\rangle\rangle = \mathrm{Tr}\left[\left(\sum_i |\Delta_i\rangle\rangle\langle\langle\Delta_i|\langle\langle R_{\mathcal{S}}|\Pi_i\rangle\rangle\right)G\right] = \\ &\mathrm{Tr}\left[\left(\mathcal{X}^{-1}\sum_i \frac{|\Pi_i\rangle\rangle\langle\langle\Pi_i|}{\langle\langle R_{\mathcal{S}}|\Pi_i\rangle\rangle}\mathcal{X}^{-1}\right)G\right] = \mathrm{Tr}\left[\mathcal{X}^{-1}G\right]\end{aligned}$$

Remark 4.2 *It is worth noting that the optimal dual does not depend on the set of the observables $\{T_m, q_m\}$. On the other hand, the optimal dual depends on the ensemble $\{R_n, p_n\}$ through $R_{\mathcal{S}}$ that appears in the definition of $\mathcal{X}$.*

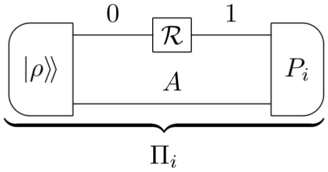
Remark 4.3 *We derived the optimal dual for the case in which the ensemble $\{R_n\}$ spans the whole $\mathcal{L}(\otimes_{k=1}^{2N-2}\mathcal{H}_k)$. When we consider the case $R_n \in \mathcal{A}$ for all n, the inverse of $\mathcal{X}$ becomes the inverse on its support and Eq. (4.11) becomes*

$$\eta = \mathrm{Tr}[\mathcal{X}^{-1}Q_{\mathcal{A}}GQ_{\mathcal{A}}]. \tag{4.12}$$

4.1.2 Optimization of the setup

In this section we address the problem of the optimization of the tester $\{\Pi_i\}$ that represents the experimental setup performing the measurement process on the unknown device we want to tomograph. We will analyze the case in which the unknown device is a Quantum Operation

$\mathcal{R} : \mathcal{L}(\mathcal{H}_0) \to \mathcal{L}(\mathcal{H}_1)$ represented by its Choi operator $R \in \mathcal{L}(\mathcal{H}_0 \otimes \mathcal{H}_1)$; within this framework the experimental setup is represented by a Quantum 2-tester $\Pi_i \in \mathcal{L}(\mathcal{H}_0 \otimes \mathcal{H}_1)$



$$. \tag{4.13}$$

We notice that the special case $\dim(\mathcal{H}_0) = 1$ corresponds to tomography of states while $\dim(\mathcal{H}_1) = 1$ corresponds to tomography of effects. In order to avoid a cumbersome notation we will perform the optimization for the case $\dim(\mathcal{H}_0) = \dim(\mathcal{H}_1) = d$; however, the generalization to the case $\dim(\mathcal{H}_0) \neq \dim(\mathcal{H}_1)$ is straightforward. We now need to make two assumptions about the ensemble of quantum operations $\{R_n, p_n\}$ and the weighted set of observables $\{T_n, q_n\}$:

- the average quantum operation is the maximally depolarizing channel $\mathcal{R}_{\mathcal{S}}(\rho) = I$ for any ρ, whose Choi operator is $R_{\mathcal{S}} = d^{-1} I_0 \otimes I_1$;
- the weighted set $\mathcal{G} = \{T_m, q_m\}$ of observables is such that $G = \sum_m q_m |T_m\rangle\!\rangle\langle\!\langle T_m| = I_{01}$; this happens for example when the set $\{T_m\}$ is an orthonormal basis, whose elements are equally weighted.

With this assumption Eq. (4.11) becomes

$$\eta = \mathrm{Tr}[\mathcal{X}^{-1}] = \mathrm{Tr}\left[\left(\sum_i \frac{d|\Pi_i\rangle\!\rangle\langle\!\langle\Pi_i|}{\mathrm{Tr}[\Pi_i]}\right)^{-1}\right] \tag{4.14}$$

We now prove that we can impose the covariance w.r.t. $\mathbf{SU}(d) \times \mathbf{SU}(d)$ on the tester. Let Π_i be the optimal quantum tester and Δ_i the corresponding optimal dual; we define

$$\begin{aligned} \Pi_{i,U,V} &:= (U_0 \otimes V_1)\Pi_i(U_0^\dagger \otimes V_1^\dagger) \\ \Delta_{i,U,V} &:= (U_0 \otimes V_1)\Delta_i(U_0^\dagger \otimes V_1^\dagger) \end{aligned} \tag{4.15}$$

where $U_0 \in \mathcal{L}(\mathcal{H}_0)$, $V_1 \in \mathcal{L}(\mathcal{H}_1)$ are unitary matrices with determinant equal to 1, i.e. they are two instances of the defining representation of $\mathbf{SU}(d)$. It is easy to check that $\Delta_{i,U,V}$ is a dual of $\Pi_{i,U,V}$; in fact we have

$$\begin{aligned} \sum_i \int \mathrm{d}U\, \mathrm{d}V\, |\Pi_{i,U,V}\rangle\!\rangle\langle\!\langle\Delta_{i,U,V}| &= \int \mathrm{d}g\, \mathrm{d}h\, W_{U,V} \left(\sum_i |\Pi_i\rangle\!\rangle\langle\!\langle\Delta_i|\right) W_{U,V}^\dagger \\ &= d^{-1} I \otimes I \end{aligned} \tag{4.16}$$

where we defined $W_{U,V} \in \mathcal{L}(\mathcal{H}_{010'1'})$, $W_{U,V} = U_0 \otimes V_1 \otimes U_{0'}^* \otimes V_{1'}^*$. We now prove that $\Pi_{i,U,V}$ and $\Delta_{i,U,V}$ give the same value of η as Π_i and Δ_i:

$$\int \mathrm{d}U\, \mathrm{d}V \sum_i d\langle\!\langle\Delta_{i,U,V}|\Delta_{i,U,V}\rangle\!\rangle\, \mathrm{Tr}[\Pi_{i,U,V}] = \sum_i d\langle\!\langle\Delta_i|\Delta_i\rangle\!\rangle\, \mathrm{Tr}[\Pi_i] = \eta$$

Because of this, we can w.l.o.g. optimize over the set of covariant testers; the condition that the covariant tester is informationally complete w.r.t. the subspace of transformations to be tomographed will be verified after the optimization.

Exploiting Theorem B.3 we have

$$\int \mathrm{d}U\,\mathrm{d}V \Pi_{i,U,V} = \int \mathrm{d}U\,\mathrm{d}V (U \otimes V)\Pi_i(U^\dagger \otimes V^\dagger) = I_{01}\frac{\mathrm{Tr}[\Pi_i]}{d^2}; \tag{4.17}$$

A generic covariant tester is then obtained by Eq. (4.15), with operators Π_i becoming “seeds” of the covariant tester and now being required to satisfy only the normalization condition[12]

$$\sum_i \mathrm{Tr}[\Pi_i] = d \tag{4.18}$$

in such a way that

$$\sum_i \int \mathrm{d}U\,\mathrm{d}V \Pi_{i,U,V} = d^{-1} I_{01} \tag{4.19}$$

satisfies the normalization (2.73). Because of the normalization (4.19) we have that $|\rho\rangle\rangle = \frac{1}{\sqrt{d}}|I\rangle\rangle$ in Eq. (2.74) that is, $\frac{1}{\sqrt{d}}|I\rangle\rangle$ is the optimal input state for the quantum operations R_n.

With the covariant tester Eq. (4.14) becomes

$$\eta = \mathrm{Tr}[\widetilde{\mathcal{X}}^{-1}], \tag{4.20}$$

where

$$\widetilde{\mathcal{X}} = \sum_i \int \mathrm{d}U\,\mathrm{d}V\; \frac{d|\Pi_{i,U,V}\rangle\rangle\langle\langle\Pi_{i,U,V}|}{\mathrm{Tr}[\Pi_{i,U,V}]} = \int \mathrm{d}U\,\mathrm{d}V\; W_{U,V}\mathcal{X}W_{U,V}^\dagger. \tag{4.21}$$

Applying Theorem B.3 and exploiting the decomposition of $U \otimes U^*$ (see Section B.3.5) we have

$$\widetilde{\mathcal{X}} = P^{pp} + AP^{qp} + BP^{pq} + CP^{qq}, \tag{4.22}$$

$$\begin{array}{ll} P^{pp} = P^p_{00'} \otimes P^p_{11'} & P^{qp} = P^q_{00'} \otimes P^p_{11'} \\ P^{pq} = P^p_{00'} \otimes P^q_{11'} & P^{qq} = P^q_{00'} \otimes P^q_{11'} \end{array} \tag{4.23}$$

having posed $P^p_{ab} = d^{-1}|I\rangle\rangle\langle\langle I|_{ab}$, $P^q_{ab} = I_{ab} - P^p_{ab}$ and

$$\begin{aligned} A &= \frac{\mathrm{Tr}[\mathcal{X}P^{qp}]}{\mathrm{Tr}[P^{qp}]} = \frac{1}{d^2-1}\left\{\sum_i \frac{\mathrm{Tr}[(\mathrm{Tr}_1[\Pi_i])^2]}{\mathrm{Tr}[\Pi_i]} - 1\right\}, \\ B &= \frac{\mathrm{Tr}[\mathcal{X}P^{pq}]}{\mathrm{Tr}[P^{pq}]} = \frac{1}{d^2-1}\left\{\sum_i \frac{\mathrm{Tr}[(\mathrm{Tr}_0[\Pi_i])^2]}{\mathrm{Tr}[\Pi_i]} - 1\right\}, \\ C &= \frac{\mathrm{Tr}[\mathcal{X}P^{qq}]}{\mathrm{Tr}[P^{qq}]} = \frac{1}{(d^2-1)^2}\left\{\sum_i \frac{d\,\mathrm{Tr}[\Pi_i^2]}{\mathrm{Tr}[\Pi_i]} - (d^2-1)(A+B) - 1\right\}. \end{aligned} \tag{4.24}$$

[12] this is the analogous of covariant POVM normalization in [1,53]

The identities in Eq. (4.24) can be obtained by making use of the identities (2.7) and (2.8)

We can now rewrite Eq. (4.20) as

$$\mathrm{Tr}[\widetilde{\mathcal{X}}^{-1}] = 1 + (d^2-1)\left(\frac{1}{A} + \frac{1}{B} + \frac{(d^2-1)}{C}\right). \tag{4.25}$$

Without loss of generality we can assume the operators $\{\Pi_i\}$ to be rank one. In fact, suppose that Π_i has rank higher than 1. Then it is possible to decompose it as $\Pi = \sum_j \Pi_{i,j}$ with $\Pi_{i,j}$ rank 1. The statistics of Π_i can be completely achieved by $\Pi_{i,j}$ through a suitable coarse graining. For the purpose of optimization it is then not restrictive to consider rank one Π_i, namely $\Pi_i = \alpha_i|\Psi_i\rangle\!\rangle\langle\!\langle\Psi_i|_{01}$, with $\sum_i \alpha_i = d$ and $||\Psi_i\rangle\!\rangle|^2 = 1$. Notice that all multiple seeds of this form lead to testers satisfying Eq. (4.19). Since $\Pi_i = \alpha_i|\Psi_i\rangle\!\rangle\langle\!\langle\Psi_i|$, exploiting Eq. (2.4) we have

$$\begin{aligned}
&\mathrm{Tr}[(\mathrm{Tr}_0[\alpha_i|\Psi_i\rangle\!\rangle\langle\!\langle\Psi_i|_{01}])^2] = \alpha_i^2\,\mathrm{Tr}[(\Psi_i\Psi_i^\dagger)^2] = \alpha_i^2\,\mathrm{Tr}[\Psi_i^\dagger\Psi_i\Psi_i^\dagger\Psi_i] = \\
&\alpha_i^2\,\mathrm{Tr}[(\Psi_i^\dagger\Psi_i\Psi_i^\dagger\Psi_i)^T] = \alpha_i^2\,\mathrm{Tr}[(\Psi_i^T\Psi_i^*)^2] = \mathrm{Tr}[(\mathrm{Tr}_1\,\alpha_i|\Psi_i\rangle\!\rangle\langle\!\langle\Psi_i|)^2] \quad \Rightarrow \quad A = B \\
&\mathrm{Tr}[(\alpha_i|\Psi_i\rangle\!\rangle\langle\!\langle\Psi_i|)^2] = \alpha_i^2\,\mathrm{Tr}[|\Psi_i\rangle\!\rangle\langle\!\langle\Psi_i|)^2] = \alpha_i^2 \quad \Rightarrow \quad C = \frac{d^2-1}{1-2A}
\end{aligned} \tag{4.26}$$

Eq. (4.25) becomes then

$$\eta = \mathrm{Tr}[\widetilde{\mathcal{X}}^{-1}] = 1 + (d^2-1)\left(\frac{2}{A} + \frac{(d^2-1)^2}{1-2A}\right) \tag{4.27}$$

where

$$0 \leqslant A = \frac{1}{d^2-1}\left(\sum_i \alpha_i\,\mathrm{Tr}[(\Psi_i\Psi_i^\dagger)^2] - 1\right) \leqslant \frac{1}{d+1} < \frac{1}{2} \quad . \tag{4.28}$$

Since η is a differentiable function of A, the minimum can be determined by deriving Eq. (4.27) with respect to A, obtaining

$$A = \frac{1}{d^2+1}; \tag{4.29}$$

the corresponding value of η is

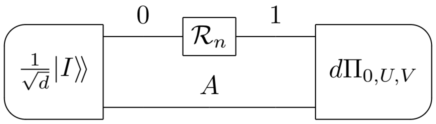
$$\eta = d^6 + d^4 - d^2. \tag{4.30}$$

This bound is achieved by a single seed $\Pi_0 = d|\Psi\rangle\!\rangle\langle\!\langle\Psi|$, with

$$\Psi = [d^{-1}(1-\beta)I + \beta|\psi\rangle\langle\psi|]^{\frac{1}{2}} \tag{4.31}$$

where $\beta = [(d+1)/(d^2+1)]^{1/2}$ and $|\psi\rangle$ is any pure state; the optimal tester is then

$$\Pi_{0,U,V} = (U\otimes V)d|\Psi\rangle\!\rangle\langle\!\langle\Psi|(U^\dagger\otimes V^\dagger) \tag{4.32}$$



We now have to verify that the set $\Pi_{0,U,V}$ is informationally complete. Exploiting Th. B.3 we have that

$$\begin{aligned} F &= \int \mathrm{d}U\,\mathrm{d}V\,|\Pi_{0,U,V}\rangle\!\rangle\langle\!\langle\Pi_{0,U,V}| = \int \mathrm{d}U\,\mathrm{d}V W_{U,V} d^2 |\Psi\rangle\!\rangle|\Psi\rangle\!\rangle\langle\!\langle\Psi|\langle\!\langle\Psi|W^\dagger_{U,V} = \\ &= \bigoplus_{\mu,\nu\in\{p,q\}} I_{\mu\nu}\otimes\frac{\mathrm{Tr}[|\Psi\rangle\!\rangle|\Psi\rangle\!\rangle\langle\!\langle\Psi|\langle\!\langle\Psi|P^{\mu\nu}]}{\mathrm{Tr}[P^{\mu\nu}]}. \end{aligned} \tag{4.33}$$

From the definition of Ψ given in Eq. (4.31) we have that $\mathrm{Tr}[|\Psi\rangle\!\rangle|\Psi\rangle\!\rangle\langle\!\langle\Psi|\langle\!\langle\Psi|P^{\mu\nu}] \neq 0$ for all ν,μ and thus F is invertible.

We now consider two relevant cases in which $R_n \in \mathcal{V} \subseteq \mathcal{L}(\mathcal{H}_{01})$:

- channels: $\mathcal{C} = \mathrm{Span}\{R\in\mathcal{L}(\mathcal{H}_{01})|\,\mathrm{Tr}_1[R] = I_0\} = \{R\in\mathcal{L}(\mathcal{H}_{01})|\,\mathrm{Tr}_1[R] = \lambda I_0, \lambda = d^{-1}\,\mathrm{Tr}[R]\in\mathbb{C}\}$;
- unital channels: $\mathcal{U} = \mathrm{Span}\{R\in\mathcal{C}|\,\mathrm{Tr}_0[R] = I_1\} = \{R\in\mathcal{L}(\mathcal{H}_{01})|\,\mathrm{Tr}_0[R] = \lambda I_1, \mathrm{Tr}_1[R] = \lambda I_0, \lambda = d^{-1}\,\mathrm{Tr}[R]\in\mathbb{C}\}$.

It is easy to prove that

$$\mathcal{V}_{\mathcal{C}} := \{|R\rangle\!\rangle | R\in\mathcal{C}\} = \mathrm{Ker}(P^{qp}) \tag{4.34}$$

exploiting Eq. (2.7) and Eq. (2.8):

$$\begin{aligned} &P^{qp}|R\rangle\!\rangle = \left(I_{00'} - \frac{|I\rangle\!\rangle\langle\!\langle I|_{00'}}{d}\right)\otimes\frac{|I\rangle\!\rangle\langle\!\langle I|_{11'}}{d}|R\rangle\!\rangle_{010'1'} = \\ &\frac{1}{d}|\,\mathrm{Tr}_1[R]\rangle\!\rangle_{00'}|I\rangle\!\rangle_{11'} - \frac{\mathrm{Tr}[R]}{d^2}|I\rangle\!\rangle_{00'}|I\rangle\!\rangle_{11'} = 0 \Leftrightarrow \mathrm{Tr}_1[R] = \frac{\mathrm{Tr}[R]}{d}I_0. \end{aligned} \tag{4.35}$$

In a similar way we have

$$\begin{aligned} &(P^{qp}+P^{pq})|R\rangle\!\rangle = \frac{1}{d}|I\rangle\!\rangle_{00'}|\,\mathrm{Tr}_0[R]\rangle\!\rangle_{11'} + \frac{1}{d}|\,\mathrm{Tr}_1[R]\rangle\!\rangle_{00'}|I\rangle\!\rangle_{11'} - 2\frac{\mathrm{Tr}[R]}{d^2}|I\rangle\!\rangle_{00'}|I\rangle\!\rangle_{11'} \\ &= \frac{1}{d}\left(\left(\mathrm{Tr}_1[R] - \frac{\mathrm{Tr}[R]}{d}I\right)_0\otimes I_1 + I_0\otimes\left(\mathrm{Tr}_0[R] - \frac{\mathrm{Tr}[R]}{d}I\right)_1\right)\otimes I_{0'1'}|I\rangle\!\rangle_{010'1'} = \\ &= 0 \Leftrightarrow \left(\mathrm{Tr}_1[R] - \frac{\mathrm{Tr}[R]}{d}I\right)_0\otimes I_1 + I_0\otimes\left(\mathrm{Tr}_0[R] - \frac{\mathrm{Tr}[R]}{d}I\right)_1 = 0 \Leftrightarrow \\ &\Leftrightarrow \mathrm{Tr}_1[R] = \frac{\mathrm{Tr}[R]}{d}I_0, \quad \mathrm{Tr}_0[R] = \frac{\mathrm{Tr}[R]}{d}I_1 \end{aligned} \tag{4.36}$$

that is

$$\mathcal{V}_{\mathcal{U}} = \mathrm{Ker}(P^{qp}+P^{pq}). \tag{4.37}$$

From Eq. (4.34) and Eq. (4.37) it follows

$$Q_{\mathcal{C}} = P^{pp}+P^{pq}+P^{qq} \qquad Q_{\mathcal{U}} = P^{pp}+P^{qq}. \tag{4.38}$$

Inserting Eq. (4.38) into Eq. (4.12) we have

$$\begin{aligned}\eta_{\mathcal{C}} &= \mathrm{Tr}[\widetilde{\mathcal{X}}^{-1}Q_{\mathcal{C}}] = \mathrm{Tr}[P^{pp} + B^{-1}P^{qp} + C^{-1}P^{qq}] \\ \eta_{\mathcal{U}} &= \mathrm{Tr}[\widetilde{\mathcal{X}}^{-1}Q_{\mathcal{U}}] = \mathrm{Tr}[P^{pp} + C^{-1}P^{qq}]\end{aligned} \tag{4.39}$$

($\widetilde{\mathcal{X}}^{-1}$ is the inverse on the support of $\mathcal{X}$). Reminding Eq. (4.26) the two figures of merit become

$$\begin{aligned}\eta_{\mathcal{C}} &= 1 + (d^2-1)\left(\frac{1}{A} + \frac{(d^2-1)^2}{1-2A}\right) \\ \eta_{\mathcal{U}} &= 1 + (d^2-1)\left(\frac{(d^2-1)^2}{1-2A}\right) \qquad .\end{aligned} \tag{4.40}$$

and the minima can be determined by derivation with respect to A thus leading to

$$\begin{aligned}\eta_{\mathcal{C}} &= d^6 + (2\sqrt{2}-3)d^4 + (5-4\sqrt{2})d^2 + 2(\sqrt{2}-1) \quad &&\text{for } A = \tfrac{1}{\sqrt{2}(d^2-1)+2} \\ \eta_{\mathcal{U}} &= (d^2-1)^3 + 1 &&\text{for } A = 0.\end{aligned}$$

The same results for quantum operation and for unital channels have been obtained in [54] in a different framework. The optimal tester for the two cases under examination have the same structure as in Eq. (4.32) where now in Eq. (4.31) we have $\beta = [(d+1)/(2+\sqrt{2}(d^2-1))]^{1/2}$ for channels and $\beta = 0$ for unital channels. Since $\beta = [(d+1)/(2+\sqrt{2}(d^2-1))]^{1/2}$ implies $\mathrm{Tr}[|\Psi\rangle\rangle|\Psi\rangle\rangle\langle\langle\Psi|\langle\langle\Psi|P^{\mu\nu}] \neq 0$, we have that optimal tester for channel tomography spans the whole $\mathcal{L}(\mathcal{H}_{01})$ (i.e. F is still invertible on the whole $\mathcal{H}_{010'1'}$).

In the case of unital channel we have $|\Psi\rangle\rangle = d^{-\frac{1}{2}}|I\rangle\rangle$ that leads to
$\mathrm{Tr}[|I\rangle\rangle|I\rangle\rangle\langle\langle I|\langle\langle I|P^{\mu\nu}] = 0$ if $\nu \neq \mu$. The frame operator becomes

$$\begin{aligned}F = \int\!\mathrm{d}U\,\mathrm{d}V\,|\Pi_{0,U,V}\rangle\rangle\langle\langle\Pi_{0,U,V}| &= \int\!\mathrm{d}U\,\mathrm{d}V W_{U,V}d^2|\Psi\rangle\rangle|\Psi\rangle\rangle\langle\langle\Psi|\langle\langle\Psi|W^{\dagger}_{U,V} = \\ = \int\!\mathrm{d}U\,\mathrm{d}V W_{U,V}|I\rangle\rangle|I\rangle\rangle\langle\langle I|\langle\langle I|W^{\dagger}_{U,V} &= P^{pp} + P^{qq}.\end{aligned} \tag{4.41}$$

Since $\mathsf{Supp}(F) = \mathsf{Supp}(P^{pp} + P^{qq}) = \mathcal{V}_{\mathcal{U}}$, the optimal tester spans the whole $\mathcal{U}$ as required.

The same procedure can be carried on when the operator G in Eqs. (4.11) (4.12) has the more general form $G = g_1P^{pp} + g_2P^{qp} + g_3P^{pq} + g_4P^{qq}$, where $P^{\nu\mu}$ are the projectors defined in (4.22). In this case Eq. (4.25) becomes

$$\mathrm{Tr}[\widetilde{\mathcal{X}}^{-1}G] = g_1 + (d^2-1)\left(\frac{g_2}{A} + \frac{g_3}{B} + \frac{(d^2-1)g_4}{C}\right), \tag{4.42}$$

which can be minimized along the same lines previously followed. G has this form when optimizing measuring procedures of this kind: *i*) preparing an input state randomly drawn from the set $\{U\rho U_g^{\dagger}\}$; *ii*) measuring an observable chosen from the set $\{U_hAU_h^{\dagger}\}$.

With the same derivation, but keeping $\dim(\mathcal{H}_0) \neq \dim(\mathcal{H}_1)$, one obtains the optimal tomography for general quantum operations. The special case of $\dim(\mathcal{H}_0) = 1$ (one has $P^q_{00'} = 0, P^p_{00'} = 1$ in Eq. (4.22)) corresponds to optimal tomography of states and gives

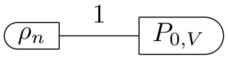
$$\eta = \frac{1}{d} + \frac{d^2-1}{A} \tag{4.43}$$

with $A = \frac{1}{d^2-1}\left(\sum_i \frac{d\,\mathrm{Tr}[P_i^2]}{\mathrm{Tr}[P_i]}\right)$. If we assume w.l.o.g that P_i is rank one we get $A = d(d-1)$ and the optimal value of η is

$$\eta = \frac{1}{d}\left(d^3 - d^2 + 1\right) \tag{4.44}$$

(compare with Ref. [47]). This bound is simply achieved by a covariant POVM

$$P_{0,V} = V d \left|\psi\right\rangle \left\langle\psi\right| V^\dagger \qquad \langle\psi|\psi\rangle = 1 \tag{4.45}$$



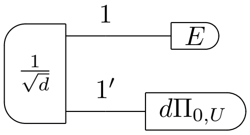
where $|\psi\rangle$ is any pure state.

On the other hand the case dim$(\mathcal{H}_1) = 1$ ($P^q_{11'} = 0, P^p_{11'} = 1$) gives the optimal tomography of effects. The optimal value of η turns out to be

$$\eta = \left(d^3 - d^2 + 1\right). \tag{4.46}$$

and is achieved by a covariant tester $\{\Pi_{0,U}\}$ through following scheme

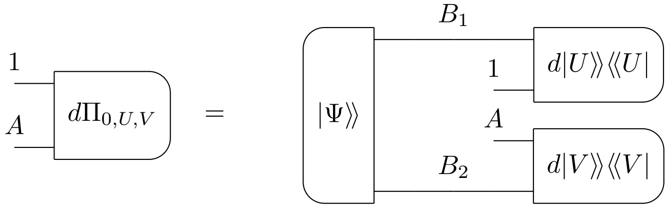
$$\Pi_{0,U} = U \left|\psi\right\rangle \left\langle\psi\right| U^\dagger \qquad \langle\psi|\psi\rangle = 1 \tag{4.47}$$



where $|\psi\rangle$ is any pure state. It is worth noting that both in the case of effects and in the case of states the derivation of the optimal tester is the same. The only difference is that for states we assume the average state $\rho_{\mathcal{S}}$ equal to I/d,while for effects we assumed $E_{\mathcal{S}} = I$. The assumption $E_{\mathcal{S}} = \sum_n p_n E_n = I$ can be interpreted by saying that we are considering a set of effects $\{\tilde{E}_n = p_n E_N\}$ that form a POVM; from this perspective the scheme (4.47) represents the optimal tomography of a POVM.

4.1.3 Realization scheme for the optimal tomography

In this section we illustrate a possible realization scheme for the optimal tomography of transformation in Eq. (4.32) that can be useful for an experimental realization. The first step is to prove the equality

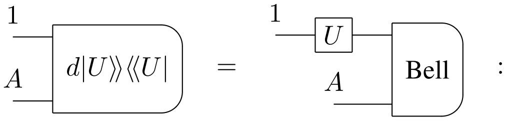

1
A
$d\Pi_{0,U,V}$
=
$|\Psi\rangle\rangle$
B_1
1
$d|U\rangle\rangle\langle\langle U|$
A
B_2
$d|V\rangle\rangle\langle\langle V|$

: (4.48)

we have

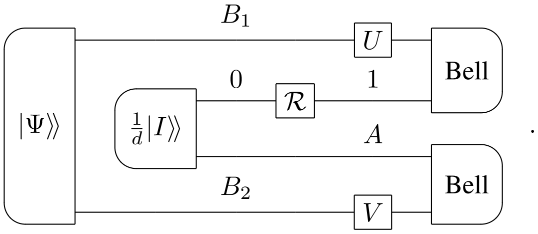
$$|\Psi\rangle\!\rangle\langle\!\langle\Psi|_{B_1B_2} * d|U\rangle\!\rangle\langle\!\langle U|_{B_11} * d|V\rangle\!\rangle\langle\!\langle V|_{B_2A} = U \otimes V d^2|\Psi\rangle\!\rangle\langle\!\langle\Psi|U^\dagger \otimes V^\dagger = d\Pi_{0,U,V} \tag{4.49}$$

Exploiting a result proved in [55] we also have that the continuous measurement $|U\rangle\!\rangle\langle\!\langle U|$ can be realized by applying a random unitary before a (discrete) Bell measurement, that is



(4.50)

Combining the scheme (4.50) with the scheme (4.48) we get:



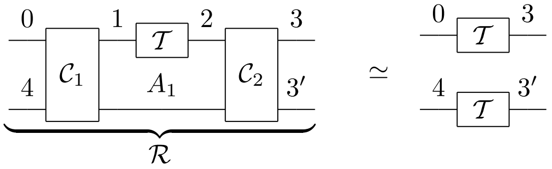
(4.51)

Referring to Eq. (4.51) the bipartite system carrying the Choi operator of the transformation is indicated with the labels 1 and A. We prepare a pair of ancillary systems B_1 and B_2 in the joint state $|\Psi\rangle\!\rangle\langle\!\langle\Psi|$, then we apply two random unitary transformations U and V to B_1 and B_2, finally we perform a Bell measurement on the pair $1 \quad B_1$ and another Bell measurement on the pair $A \;\; B_2$.

The scheme proposed is feasible using e. g. the Bell measurements experimentally realized in [56].

5 Cloning a Unitary Transformation

The no-cloning theorem [57] is one of the main results in Quantum Information, and it is the basis of the security of quantum cryptography. Although the cloning of quantum states has been extensively studied [58,59,60,61] the cloning of transformation is quite a new topic. This chapter reviews Ref. [26] where the cloning of a quantum transformation was introduced and the optimal network that clones a single use of a unitary transformation was derived. Cloning a single use of a transformation $\mathcal{T}$ means exploiting a single use of $\mathcal{T}$ inside a quantum network, in such a way that the overall transformation is as close as possible to two uses of $\mathcal{T}$



(5.1)

Cloning quantum transformations can be used for copying quantum software with a limited number of uses, and in other informational contexts, e.g. in the security analysis of multi-round cryptographic protocols with encoding on secret transformations. We can consider for example this alternative version of the BB84 [62] protocol. Bob prepares the maximally entangled state $2^{-\frac{1}{2}}|I\rangle\!\rangle$ of two qubits and sends one half of the system to Alice. Alice perform either a unitary from the set $A_1 = \{\sigma_\mu\}$ (where $\sigma_0 = I$ and $\sigma_{1,2,3}$ are the three Pauli matrices) or a unitary from the rotated set $A_2 = \{U\sigma_\mu\}$, where U is a unitary in $\mathbf{SU}(2)$. Then Alice sends back is portion of the system to Bob that finally measures either the Bell basis $\{2^{-\frac{1}{2}}|\sigma_\mu\rangle\!\rangle\}$ or the rotated basis $\{2^{-\frac{1}{2}}|U\sigma_\mu\rangle\!\rangle\}$. After they publicly announce their choice of basis and discarded the cases in which they took different choices. they use the values of μ as a secret key. A natural attack to this protocol is the quantum cloning (see Fig. 5.1).

Cloning an undisclosed transformation is a challenging task not only from a quantum-theory perspective but even classically. Indeed, the following result holds:

Theorem 5.1 (no-cloning for transformations) *Let $\mathcal{O}_1$ and $\mathcal{O}_2$ be two quantum or classical transformations and let $p \leqslant 1/2$ denote the minimum of the worst case error probability in discriminating between them. Then $\mathcal{O}_1$ and $\mathcal{O}_2$ cannot be perfectly cloned by a single use unless $p = 0$ (perfect discrimination) or $p = 1/2$ (i.e. $\mathcal{O}_1 = \mathcal{O}_2$)*

Proof. The proof is simple: if perfect cloning is possible, we can get three copies, perform three times the minimum error discrimination, and use majority voting to decide the most likely between $\mathcal{O}_1$ and $\mathcal{O}_2$ with worst case error probability $p' = p^2(3-2p)$. Since p is the minimum error probability, it must be $p \leq p'$, whose acceptable solutions are only $p = 0$ and $p = 1/2$. Viceversa, if $\mathcal{O}_1$ and $\mathcal{O}_2$ can be perfectly distinguished (i.e. $p = 0$), then they can be perfectly cloned by a classical strategy based on discrimination and subsequent re-preparation of the corresponding transformation. This result can be generalized to an arbitrary number of transformations:

Corollary 5.1 *Let $\{\mathcal{O}_i\}, i = 1, \ldots, N$ a set of transformations. Then perfect cloning is possible iff either $\mathcal{O}_i = \mathcal{O}_j$ for all i, j or $\{\mathcal{O}_i\}, i = 1, \ldots, N$ are perfectly discriminable by a single use*

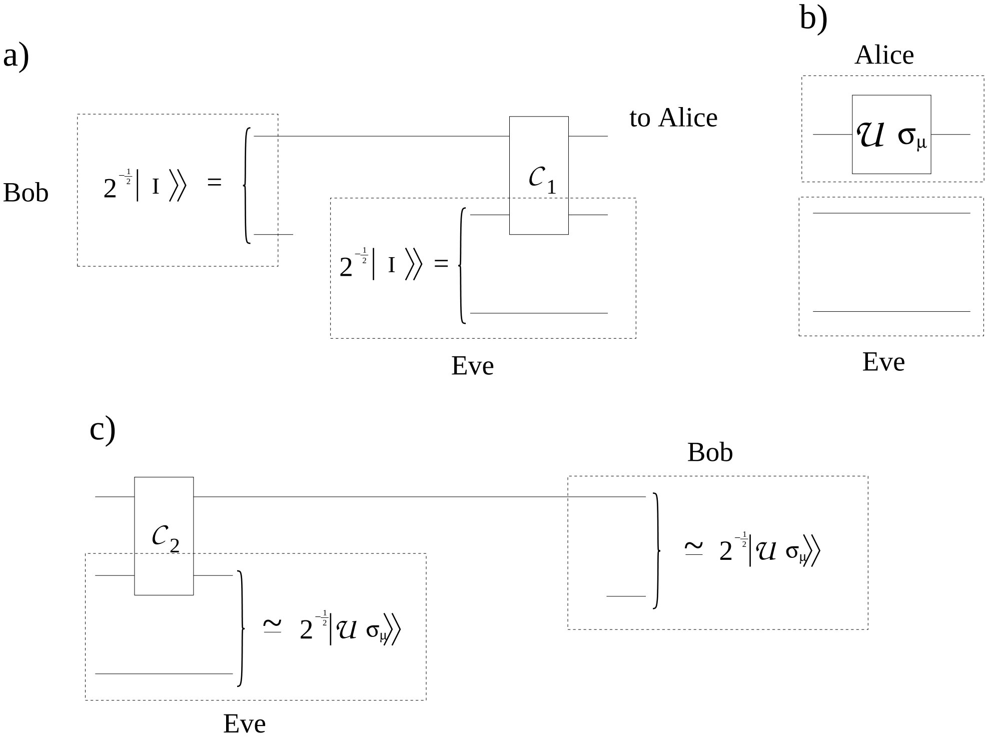

a)
Bob
$2^{-\frac{1}{2}}|I\rangle\rangle =$
$\mathcal{C}_1$
to Alice
$2^{-\frac{1}{2}}|I\rangle\rangle =$
Eve
b)
Alice
$\mathcal{U}\sigma_\mu$
Eve
c)
$\mathcal{C}_2$
$\simeq 2^{-\frac{1}{2}}|\mathcal{U}\sigma_\mu\rangle\rangle$
Eve
Bob
$\simeq 2^{-\frac{1}{2}}|\mathcal{U}\sigma_\mu\rangle\rangle$

Figure 5.1. Alternative version of the $BB84$ protocol with a possible eavesdropping. a) Bob prepares the state $2^{-\frac{1}{2}}|I\rangle\rangle$ and send one half of it to Alice. Eve prepares the same state as Bob, intercepts the portion of system addressed to Alice and performs the channel $\mathcal{C}_1$. Then Eve send one half of the outcome to Alice. b) Alice applies $\mathcal{U}\sigma_\mu$ to her portion of system and send it back to Bob. c) Eves intercepts the portion of system addressed to Bob and performs the channel $\mathcal{C}_2$; if the quantum network $\mathcal{C}_1 \star \mathcal{C}_2$ is a cloning network Eve obtains a state which is the same that Bob has and that is as close as possible to $2^{-\frac{1}{2}}|U\sigma_\mu\rangle\rangle$.

Remark 5.1 *It is worth noting that this result for $N > 2$ is non trivial also for classical transformations. Consider the following permutations of the set $\{1, 2, 3, 4\}$* [13]

$$\sigma_1 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 3 & 4 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 4 & 3 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \end{pmatrix}; \quad (5.2)$$

there is no way to perfectly discriminate among them by evaluating the permutations on a single element.

The existence of a no-cloning theorem immediately rises the problem of finding the optimal cloners: In the following section we will derive the optimal network which produces two approximate copies of a completely unknown unitary transformation $\mathcal{U} \in \mathbf{SU}(d)$.

[13] we use the following notation: the first row contains the elements $\{1, 2, 3, 4\}$, and the second row contains the images under the permutation of the elements above.

5.1 Optimal cloning of a Unitary transformation

Exploiting the general theory developed in chapter 2 the cloning network $\mathcal{R}$ (see Eq. (5.1)) can be represented by means of its Choi operator R that has to satisfy the constraint (2.60), that is

$$\mathrm{Tr}_{35}[R] = I_2 \otimes R^{(1)} \qquad \mathrm{Tr}_1[R^{(1)}] = I_{04}. \tag{5.3}$$

When we insert the unitary channel $\mathcal{U}$ in the network $\mathcal{R}$ we obtain the network

$$\mathcal{C}_U := \mathcal{R} \star \mathcal{U} \tag{5.4}$$

whose Choi operator is

$$C_U = R * |U\rangle\!\rangle\langle\!\langle U| = \langle\!\langle U^*|R|U^*\rangle\!\rangle \tag{5.5}$$

As a figure of merit we use the channel fidelity (see Appendix A) between $\mathcal{C}_U$ and and the two uses $\mathcal{U} \otimes \mathcal{U}$ of unitary channel, averaged over all the unitaries in $\mathbf{SU}(d)$:

$$\begin{aligned} F &:= \frac{1}{d^4}\int_{\mathbf{SU}(d)} \mathrm{d}U \mathcal{F}(\mathcal{C}_U, \mathcal{U}\otimes\mathcal{U}) = \frac{1}{d^4}\int_{\mathbf{SU}(d)} \mathrm{d}U \langle\!\langle U|\langle\!\langle U|C_U|U\rangle\!\rangle|U\rangle\!\rangle = \\ &= \frac{1}{d^4}\int_{\mathbf{SU}(d)} \mathrm{d}U \langle\!\langle U|\langle\!\langle U|\langle\!\langle U^*|R|U^*\rangle\!\rangle|U\rangle\!\rangle|U\rangle\!\rangle. \end{aligned} \tag{5.6}$$

The following Lemma exploits the symmetry of Eq. (5.6) and simplifies the structure of the optimal network:

Lemma 5.1 *The optimal cloning network maximizing the channel fidelity (5.6) can be assumed without loss of generality to satisfy the commutation relation*

$$[R, V_{04}^{\otimes 2} \otimes V_1^* \otimes W_2 \otimes W_{35}^{*\otimes 2}] = 0 \quad \forall V, W \in \mathbb{SU}(d) \; . \tag{5.7}$$

Proof. Let R be optimal. Then consider the average

$$\overline{R} = \int \mathrm{d}V\, \mathrm{d}W\; (V_{04}^{\otimes 2} \otimes V_1^* \otimes W_2 \otimes W_{35}^{*\otimes 2}) R (V_{04}^{\otimes 2} \otimes V_1^* \otimes W_2 \otimes W_{35}^{*\otimes 2})^\dagger; \tag{5.8}$$

exploiting the properties of the Haar measures (see Definition B.7) we have

$$\begin{aligned} F &= \frac{1}{d^4}\int \mathrm{d}U \langle\!\langle U|^{\otimes 2}\langle\!\langle U^*|\overline{R}|U^*\rangle\!\rangle|U\rangle\!\rangle^{\otimes 2} = \\ &= \frac{1}{d^4}\int \mathrm{d}U \langle\!\langle VUW^\dagger|^{\otimes 2}\langle\!\langle V^*U^*W^T|\overline{R}|V^*U^*W^T\rangle\!\rangle|VUW^\dagger\rangle\!\rangle^{\otimes 2} = \\ &= \frac{1}{d^4}\int \mathrm{d}U \langle\!\langle U|^{\otimes 2}\langle\!\langle U^*|R|U^*\rangle\!\rangle|U\rangle\!\rangle^{\otimes 2} \end{aligned} \tag{5.9}$$

that is $\overline{R}$ and R give the same value of F. From Theorem B.3 we have that $\overline{R}$ satisfies Eq. (5.7). Finally, we verify that $\overline{R}$ satisfies Eq. (5.3):

$$\begin{aligned}
&\mathrm{Tr}_{35}[\overline{R}] = \\
&= \mathrm{Tr}_{35}\left[\int \mathrm{d}V\,\mathrm{d}W\;(V_{04}^{\otimes 2}\otimes V_1^*\otimes W_2\otimes W_{35}^{*\otimes 2})R(V_{04}^{\otimes 2}\otimes V_1^*\otimes W_2\otimes W_{35}^{*\otimes 2})^\dagger\right] = \\
&= \int \mathrm{d}V\,\mathrm{d}W\;(V_{04}^{\otimes 2}\otimes V_1^*\otimes W_2)\,\mathrm{Tr}_{35}\,[R]\,(V_{04}^{\otimes 2}\otimes V_1^*\otimes W_2) = \\
&= \int \mathrm{d}V\,\mathrm{d}W\;(V_{04}^{\otimes 2}\otimes V_1^*\otimes W_2)I_2\otimes R^{(1)}(V_{04}^{\otimes 2}\otimes V_1^*\otimes W_2) = \\
&= I_2\otimes\int \mathrm{d}V(V_{04}^{\otimes 2}\otimes V_1^*)R^{(1)}(V_{04}^{\otimes 2}\otimes V_1^*) = I_2\otimes\overline{R}^{(1)}
\end{aligned}\tag{5.10}$$

$$\begin{aligned}
&\mathrm{Tr}_1[\overline{R}^{(1)}] = \mathrm{Tr}_1\left[\int \mathrm{d}V(V_{04}^{\otimes 2}\otimes V_1^*)R^{(1)}(V_{04}^{\otimes 2}\otimes V_1^*)\right] = \\
&= \int \mathrm{d}V V_{04}^{\otimes 2}\,\mathrm{Tr}_1\left[V_1^* R^{(1)} V_1^*\right]V_{04}^{\otimes 2} = \int \mathrm{d}V V_{04}^{\otimes 2}\,\mathrm{Tr}_1\left[R^{(1)}\right]V_{04}^{\otimes 2} = \\
&= \int \mathrm{d}V V_{04}^{\otimes 2} I_{04} V_{04}^{\otimes 2} = I_{04}
\end{aligned}\tag{5.11}$$

■

Exploiting Eq. (5.7) the figure of merit (5.6) becomes:

$$F = \frac{1}{d^4}\langle\!\langle I|\langle\!\langle I|\langle\!\langle I|R|I\rangle\!\rangle|I\rangle\!\rangle|I\rangle\!\rangle. \tag{5.12}$$

Thanks to the commutation relation (5.7), we can apply the decomposition (B.51) to the Choi operator R:

$$R = \sum_{\nu,\mu\in\mathsf{S}}\sum_{i,j,k,l=\pm} T^{\nu,i,j}\otimes T^{\mu,k,l}\otimes r_{\nu\mu}^{ik,jl} \tag{5.13}$$

where we notice that $(r_{\nu\mu}^{ik,jl})$ is a non negative matrix for any ν,μ.

The decomposition (B.44) induces the following identity

$$\mathcal{H}\otimes\mathcal{H}\otimes\mathcal{H} = \mathcal{H}_{\alpha+}\oplus\mathcal{H}_{\alpha-}\oplus\mathcal{H}_{\beta-}\oplus\mathcal{H}_{\gamma-} \tag{5.14}$$

that leads to

$$\begin{aligned}
&|I\rangle\!\rangle_{03}|I\rangle\!\rangle_{45}|I\rangle\!\rangle_{12} = |I\rangle\!\rangle_{(041)(352)} = \\
&= (T^{\alpha,+,+}\oplus T^{\alpha,-,-}\oplus P^{\beta-}\oplus P^{\gamma-})\otimes I_{352}|I\rangle\!\rangle_{(041)(352)} = \\
&= |T^{\alpha,+,+}\rangle\!\rangle + |T^{\alpha,-,-}\rangle\!\rangle + |T^{\beta,+,+}\rangle\!\rangle + |T^{\gamma,-,-}\rangle\!\rangle
\end{aligned}\tag{5.15}$$

Inserting Eq. (5.15) into Eq. (5.12) and reminding the decomposition (5.13) we have

$$\begin{aligned} F &= \frac{1}{d^4}\langle\!\langle I|\langle\!\langle I|\langle\!\langle I|R|I\rangle\!\rangle|I\rangle\!\rangle|I\rangle\!\rangle = \\ &= \frac{1}{d^4}\sum_{i'}\sum_{\nu'}\langle\!\langle T^{\nu',i',i'}|\left(\sum_{\nu,\mu}\sum_{i,j,k,l=\pm} T^{\nu,i,j}\otimes T^{\mu,k,l}\otimes r^{ik,jl}_{\nu\mu}\right)\sum_{j'}\sum_{\mu'}|T^{\mu',j',j'}\rangle\!\rangle = \\ &= \frac{1}{d^4}\sum_{\nu}\sum_{i,j} d_\nu r^{ii,jj}_{\nu\nu} \end{aligned} \tag{5.16}$$

where $d_\nu := \dim(\mathcal{H}_\nu)$.

We now express the normalization constraint in terms of the $r^{ik,jl}_{\nu\mu}$. Taking the trace over $\mathcal{H}_{35}$ in Eq. (5.7) we get:

$$\begin{aligned} &0 = [\mathrm{Tr}_{35}[R], V_{04}^{\otimes 2}\otimes V_1^*\otimes W_2] = [I_2\otimes R^{(1)}, V_{04}^{\otimes 2}\otimes V_1^*\otimes W_2] \Rightarrow \\ &\Rightarrow [R^{(1)}, V_{04}^{\otimes 2}\otimes V_1^*] \Rightarrow R^{(1)} := S = \sum_{\nu}\sum_{i,j} T^{\nu,i,j}s^{i,j}_{\nu} \end{aligned} \tag{5.17}$$

Reminding the decomposition (B.33) we have:

$$[\mathrm{Tr}_1[S], V_{04}^{\otimes 2}] = 0 \Rightarrow \mathrm{Tr}_1[S] = t_+P^+ \oplus t_-P^- \tag{5.18}$$

Comparing Eq. (5.17) and Eq. (5.18) we get:

$$\begin{aligned} &\mathrm{Tr}_1[S] = \mathrm{Tr}_1[\sum_{\nu}\sum_{i,j} T^{\nu,i,j}s^{i,j}_{\nu}] = t_+P^+ \oplus t_-P^- \Rightarrow \\ &\Rightarrow t_i d_i = \sum_{\nu} d_\nu s^{i,i}_{\nu} \qquad i = \pm. \end{aligned} \tag{5.19}$$

The normalization constraint $\mathrm{Tr}_1[S] = I_{04}$ becomes then

$$\mathrm{Tr}_1[S] = t_+P^+ \oplus t_-P^- = I_{04} \Rightarrow t_+ = t_- = 1 \qquad i = \pm. \tag{5.20}$$

Comparing now Eq. (5.17) with Eq. (5.13) we have

$$\begin{aligned} &\mathrm{Tr}_{35}[R] = I_2\otimes S \Rightarrow \mathrm{Tr}_{35}\left[R = \sum_{\nu,\mu}\sum_{i,j,k,l} T^{\nu,i,j}\otimes T^{\mu,k,l}\otimes r^{ik,jl}_{\nu\mu}\right] = \\ &= I_2\otimes\sum_{\nu}\sum_{i,j} T^{\nu,i,j}s^{i,j}_{\nu} \Rightarrow s^{i,i}_{\nu} = \sum_{k}\sum_{\mu}\frac{d_\mu}{d}r^{ik,ik}_{\nu\mu} \end{aligned} \tag{5.21}$$

Inserting Eq. (5.21) into Eq. (5.19) we obtain

$$t_i d_i = \sum_{\nu\mu}\sum_{k} d_\nu \frac{d_\mu}{d} r^{ik,ik}_{\nu\mu}. \tag{5.22}$$

The normalization (5.20) becomes then

$$dd_i = \sum_{\nu\mu} \sum_{k} d_\nu d_\mu r^{ik,ik}_{\nu\mu} \tag{5.23}$$

We are now ready to derive the optimal cloner:

Theorem 5.2 (optimal cloning network) *For the fidelity (5.16) the following bound holds:*

$$F \leqslant (d + \sqrt{d^2-1})/d^3. \tag{5.24}$$

The bound (5.24) can be achieved by a network as in Eq. (5.1) with:

- $\mathcal{C}_1 : \mathcal{L}(\mathcal{H}_{04}) \to \mathcal{L}(\mathcal{H}_{1A_1})$ *is given by:*

$$\mathcal{C}_1(\rho) = \sum_{i,j} \mathrm{Tr}_4[P_i \rho P_j] \otimes |i\rangle\langle j| \tag{5.25}$$

- $\mathcal{C}_2 : \mathcal{L}(\mathcal{H}_{2\,A_1}) \to \mathcal{L}(\mathcal{H}_{35})$ *is given by:*

$$\mathcal{C}_2(\sigma) = \sum_{i,j} \frac{d}{\sqrt{d_i d_j}}\, P_i \left[\langle i|\sigma|j\rangle \otimes I_5\right] P_j\ . \tag{5.26}$$

where $\mathcal{H}_{A_1} = \mathbb{C}^2$ *and* $\{|+\rangle, |-\rangle\}$ *is an orthonormal basis of* $\mathcal{H}_{A_1}$.

The resulting channel $\mathcal{C}_U = \mathcal{R} \star \mathcal{U}$ *is then given by*

$$\begin{aligned} \mathcal{C}'_U(\rho) &= \mathcal{C}_2 \star (\mathcal{U} \otimes \mathcal{I}_{A_1}) \star \mathcal{C}_1(\rho) \\ &= \sum_{i,j} \frac{d}{\sqrt{d_i d_j}}\, P_i \left[U\, \mathrm{Tr}_{0E}[P_i \rho P_j] U^\dagger \otimes I\right] P_j\ . \end{aligned} \tag{5.27}$$

Proof. Consider the matrix $(a_{i,k}) := \left(\sum_\nu r^{ik,ik}_{\nu\nu}\right)$: $(a_{i,k})$ is non negative and the bound $a_{i,k} \leqslant \sqrt{a_{i,i}a_{k,k}}$ holds. Then we have

$$F \leqslant \frac{1}{d^4}\left(\sum_i \sqrt{\sum_\nu d_\nu r^{\nu\nu}_{ii,ii}}\right)^2 = \frac{1}{d^4}\left(\sum_i \sqrt{\sum_\nu \frac{d_\nu^2}{d_\nu} r^{\nu\nu}_{ii,ii}}\right)^2 \tag{5.28}$$

where $\overline{\nu}$ labels the irreducible subspace of $U \otimes U \otimes U^*$ with minimum dimension, that is $\overline{\nu} = \alpha$. Exploiting the constraint (5.23) into Eq. (5.28) we get

$$\begin{aligned} F &\leqslant \frac{1}{d^4}\left(\sum_i \sqrt{\frac{1}{d_\alpha}\sum_\nu d_\nu^2 r^{\nu\nu}_{ii,ii}}\right)^2 \leqslant \\ &\leqslant \frac{1}{d^4}\left(\sum_i \sqrt{\frac{1}{d_\alpha} d_i d}\right)^2 = \frac{1}{d^4}(\sqrt{d_+} + \sqrt{d_-})^2 \end{aligned} \tag{5.29}$$

Direct computation of Eq. (5.6) with $\mathcal{C}_U$ as defined in Eq. (5.27) proves the achievability. ■

5.2 The optimal cloning network

In this section we discuss the inner structure of the optimal cloning network $\mathcal{R} = \mathcal{C}_2 \star \mathcal{C}_1$. We can extend $\mathcal{C}_1$ to a unitary interaction between the input systems $\mathcal{H}_0, \mathcal{H}_4$ and the memory $\mathcal{H}_{a_1}$: $\mathcal{C}_1(\rho) = \mathrm{Tr}_{0E}[V(\rho \otimes |0\rangle\langle 0|)V^\dagger]$, where $|0\rangle = (|+\rangle + |-\rangle)/\sqrt{2} \in \mathcal{M}$, and V is the controlled-swap $V = I \otimes |+\rangle\langle +| + S \otimes |-\rangle\langle -|$, $S|\phi\rangle|\psi\rangle = |\psi\rangle|\phi\rangle$. Such an extension has a very intuitive meaning in terms of quantum parallelism: for bipartite input $|\Psi\rangle_{04}$ the single-system unitary U is made to work on both B and E by applying it to the superposition $|\Psi\rangle_{04} + S|\Psi\rangle_{04}$ and discarding E.

On the other hand the channel $\mathcal{C}_2$ can be interpreted as an extension of optimal cloning of pure states [59]: if $\mathcal{C}_2$ receives the state $|\psi\rangle_2|+\rangle_{A_1}$ as an input, the output is $\mathcal{C}_2(|\psi\rangle\langle\psi| \otimes |+\rangle\langle +|) = d/d_+ \, [P_+(|\psi\rangle\langle\psi| \otimes I)P_+]$, which are indeed two optimal clones of $|\psi\rangle$. This means that realizing the optimal cloning of unitaries is a harder task than realizing the optimal cloning of states: an eavesdropper that is able to optimally clone unitaries must also be able to optimally clone pure states. This suggests that cryptographic protocols based on gates (such the alternative $BB84$ protocol described at the beginning of the chapter) might be harder to attack than protocols based on states.

Remark 5.2 *It is worth notice that the optimal cloning network that we derived in the previous sections, is not the optimal attack to the protocol in Fig. 5.1. We derived the optimal cloning network for an arbitrary unitary of* $\mathbf{SU}(d)$*; an optimization for the restricted set* $A_1 \cap A_2$ *of unitaries involved in the protocol could in principle achieve better performances.*

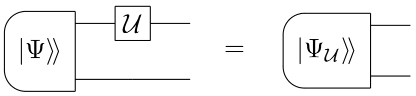
6 Quantum learning of a unitary transformation

Quantum memory is a key resource for quantum information and computation and great experimental efforts are in operation in order to realize it [63, 64, 65]. A quantum memory can be used to store an unknown transformation; in this way Alice can transmit the transformation to a distant Bob avoiding to send the physical device; Bob retrieves the transformation from the quantum memory.

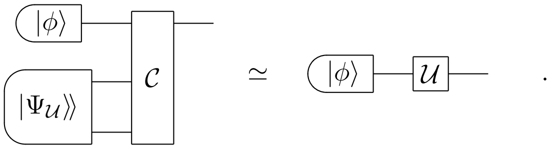
Quantum learning is an example of storing and retrieving of a transformation. Consider the scenario in which Alice puts at Bob's disposal N uses of a black box implementing an unknown unitary transformation $\mathcal{U} = U \cdot U^\dagger$. Today Bob is allowed to exploit such uses at his convenience, running an arbitrary quantum circuit that makes N calls to Alice's black box. Tomorrow, however, Alice will withdraw the black box and ask Bob to reproduce $\mathcal{U}$ on a new input state $|\psi\rangle$ unknown to him. Alice will then test the output produced by Bob, and assign a score that is as higher as the output is closer to $U|\psi\rangle$. More generally, Alice can ask Bob to reproduce $\mathcal{U}$ more than once, i.e. to produce $M \geqslant 1$ copies of $\mathcal{U}$.

Let us focus first on the case in which a single use of the black box is available today ($N = 1$) and a single copy has to be produced tomorrow ($M = 1$). The only thing Bob can do today is to apply the unknown unitary $\mathcal{U}$ to a known (generally entangled state) $|\Psi\rangle\rangle$ thus producing the state

$$|\Psi_\mathcal{U}\rangle\rangle := U \otimes I|\Psi\rangle\rangle$$

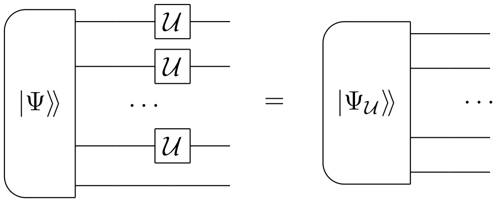


after that Bob can store the state $|\Psi_\mathcal{U}\rangle\rangle$ on a quantum memory. Tomorrow, when Alice will provide the unknown state $|\varphi\rangle$, Bob can send both $|\varphi\rangle$ and $|\Psi_\mathcal{U}\rangle\rangle$ as input to a channel $\mathcal{C}$ whose output state has to be as close as possible to $U\,|\phi\rangle$:



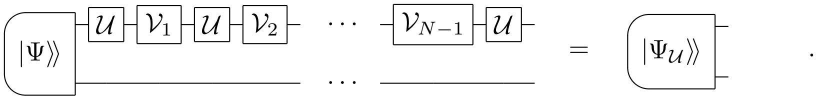
When $N > 1$ uses of the black box are available, Bob has several option to encode the unknown unitary into the state of the quantum memory: he can e.g. opt for a *parallel strategy* where $\mathcal{U}$ is applied on N different systems, yielding

$$|\Psi_\mathcal{U}\rangle\rangle = (U^{\otimes N} \otimes I)|\Psi\rangle\rangle$$



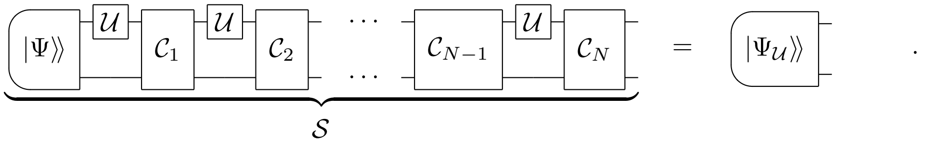
or for a *sequential strategy* where $\mathcal{U}$ is applied N times on the same system, generally alternated with other known unitaries, yielding

$$|\Psi_{\mathcal{U}}\rangle\!\rangle := (UV_{N-1}\dots V_2UV_1U \otimes I)|\Psi\rangle\!\rangle$$



The most general storing strategy is described by a quantum network in which the N uses of the transformation $\mathcal{U}$ are inserted (see also Fig. 2.7):

$$|\Psi_{\mathcal{U}}\rangle\!\rangle := S * |U\rangle\!\rangle\langle\!\langle U| * \dots * |U\rangle\!\rangle\langle\!\langle U|$$



Quantum learning of a transformation can be seen as an instance of *Quantum Programming* [9, 66, 67, 68, 69]: the retrieving channel is indeed an example of a programmable device that uses the state $|\Psi\rangle\!\rangle_{\mathcal{U}}$ as a program. The following result [9] tells us that a universal programmable quantum channel with a finite dimensional program state, does not exists.

Theorem 6.1 (No Programming) *There exists no universal programmable channel, that is a quantum channel* $\mathcal{C} : \mathcal{L}(\mathcal{H}_0 \otimes \mathcal{H}_P) \to \mathcal{L}(\mathcal{H}_1)$, *where* $\dim(\mathcal{H}_0) = \dim(\mathcal{H}_1) = d$ *and* $\dim(\mathcal{H}_P) < \infty$, *with the following property:*

$$\mathcal{C}(\rho \otimes \sigma_U) = U\rho U^\dagger \tag{6.1}$$

for all state $\rho \in \mathcal{L}(\mathcal{H}_0)$ *and all unitaries* $U \in \mathbf{SU}(d)$.

Proof. Consider an isometric dilation V of $\mathcal{C}$ and suppose that ρ is a pure state $|\psi\rangle$; we have

$$\mathrm{Tr}_A[V(|\psi\rangle\langle\psi| \otimes \sigma_U)V^\dagger] = U|\psi\rangle\langle\psi|U^\dagger \tag{6.2}$$

Adding an auxiliary Hilbert space $\mathcal{H}_{P'} \cong \mathcal{H}_P$ we have the identity

$$\mathrm{Tr}_{A'}[W|\psi\rangle\langle\psi| \otimes |\sigma_U^{\frac{1}{2}}\rangle\!\rangle\langle\!\langle\sigma_U^{\frac{1}{2}}|W^\dagger] = \mathrm{Tr}_{AP'}[V \otimes I_{P'}(|\psi\rangle\langle\psi| \otimes |\sigma_U^{\frac{1}{2}}\rangle\!\rangle\langle\!\langle\sigma_U^{\frac{1}{2}}|)V^\dagger \otimes I_{P'}] =$$
$$= \mathrm{Tr}_A[V(|\psi\rangle\langle\psi| \otimes \mathrm{Tr}_{P'}|\sigma_U^{\frac{1}{2}}\rangle\!\rangle\langle\!\langle\sigma_U^{\frac{1}{2}}|)V^\dagger] = \mathrm{Tr}_A[V(|\psi\rangle\langle\psi| \otimes \sigma_U)V^\dagger]$$

where we defined $\mathcal{H}_{A'} = \mathcal{H}_{P'} \otimes \mathcal{H}_A$ and $W = V \otimes I_{P'}$; then, w.l.o.g. we can consider a pure program state $|\tilde{\sigma}_U\rangle$. Since $U|\psi\rangle\langle\psi|U^\dagger$ is a pure state we must have $W(|\psi\rangle \otimes |\tilde{\sigma}_U\rangle) = U|\psi\rangle \otimes |\tau_U\rangle$ for some pure state $|\tau_U\rangle$. First we prove that the state $|\tau_U\rangle$ does not depend on $|\psi\rangle$; we have

$$\langle\tilde{\sigma}_U|\tilde{\sigma}_U\rangle\langle\psi|\psi'\rangle = (\langle\psi| \otimes \langle\tilde{\sigma}_U|)W^\dagger W(|\psi'\rangle \otimes |\tilde{\sigma}_U\rangle) =$$
$$= \langle\psi| \otimes \langle\tau_U|U^\dagger U|\psi'\rangle \otimes |\tau'_U\rangle = \langle\tau_U|\tau'_U\rangle\langle\psi|\psi'\rangle$$

and so $|\tau_U\rangle = |\tau'_U\rangle$ if if $\langle\psi|\psi'\rangle \neq 0$. On the other hand if $\langle\psi|\psi'\rangle = 0$ we have

$$\begin{aligned}
&U\tfrac{1}{\sqrt{2}}(|\psi'\rangle + |\psi\rangle) \otimes |\tau_U\rangle = W(\tfrac{1}{\sqrt{2}}(|\psi'\rangle + |\psi\rangle) \otimes |\tilde{\sigma}_U\rangle) = \\
&\quad = W(\tfrac{1}{\sqrt{2}}\,|\psi'\rangle \otimes |\tilde{\sigma}_U\rangle) + W(\tfrac{1}{\sqrt{2}}\,|\psi\rangle \otimes |\tilde{\sigma}_U\rangle) = \\
&\quad = U\tfrac{1}{\sqrt{2}}\,|\psi'\rangle \otimes |\tau'_U\rangle + U\tfrac{1}{\sqrt{2}}\,|\psi\rangle \otimes |\tau_U\rangle \Rightarrow |\tau'_U\rangle = |\tau_U\rangle\,.
\end{aligned}$$

Now let U_1 and U_2 be two unitaries different up to a global phase; for arbitrary $|\psi\rangle$ we have

$$\begin{aligned}
&W\,|\psi\rangle \otimes |\tilde{\sigma}_{U_1}\rangle = U_1\,|\psi\rangle \otimes |\tau_{U_1}\rangle \\
&W\,|\psi\rangle \otimes |\tilde{\sigma}_{U_2}\rangle = U_2\,|\psi\rangle \otimes |\tau_{U_2}\rangle\,;
\end{aligned}$$

if we take the scalar product of the previous two identities we get

$$\langle\tilde{\sigma}_{U_1}|\tilde{\sigma}_{U_2}\rangle = \langle\psi|\,U_1^\dagger U_2\,|\psi\rangle\,\langle\tau_{U_1}|\tau_{U_2}\rangle.$$

If $\langle\tau_{U_1}|\tau_{U_2}\rangle \neq 0$ we can write

$$\frac{\langle\tilde{\sigma}_{U_1}|\tilde{\sigma}_{U_2}\rangle}{\langle\tau_{U_1}|\tau_{U_2}\rangle} = \langle\psi|\,U_1^\dagger U_2\,|\psi\rangle$$

and since the left hand side of the equation does not depend on $|\psi\rangle$ we have that also $\langle\psi|\,U_1^\dagger U_2\,|\psi\rangle$ must not depend on $|\psi\rangle$. However, this is possible only if $U_1^\dagger U_2 = \lambda I$ for some $\lambda \in \mathbb{C}$ that is U_1 is equal to U_2 up to a global phase which is contrary to the hypothesis. Then it must be $\langle\tau_{U_1}|\tau_{U_2}\rangle = 0$ that implies $\langle\tilde{\sigma}_{U_1}|\tilde{\sigma}_{U_2}\rangle = 0$ that is, the programs of two distinct unitaries must be orthogonal states; since there are infinite distinct unitaries in $\mathbf{SU}(d)$ we cannot have $\dim(\mathcal{H}_P) < \infty$. ■

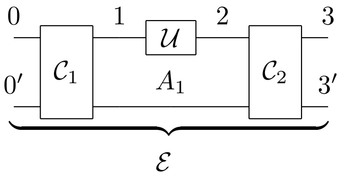
The case in which the program state σ_U is the output of a fixed quantum network in which N uses of the unknown unitary U are inserted, corresponds to the learning scenario; since Theorem 6.1 proved that perfect programming is not possible, quantum learning can be realized only approximately.[14] That being so, the search for the optimal learning protocol deserves interest.

Moreover, we can think of quantum learning as an instance of quantum cloning of a transformation as presented in the previous chapter[15]. In the learning case we have the additional constraint that the N uses are provided before than the input states on which we want to apply the M replicas. Let us focus on the $N = 1, M = 2$ case; the following two different scenarios

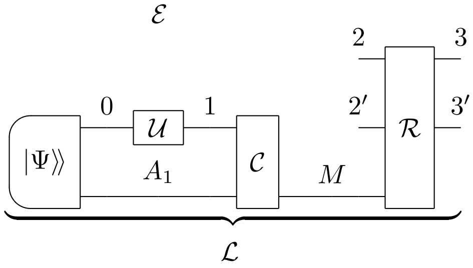
[14] Whether the optimal programming of unitaries coincides with the optimal Quantum Learning is still an open question.

[15] Clearly this interpretation make sense only if the number of uses N is greater than the number of replicas M

are possible:



(6.3)



(6.4)

The two networks in Eqs. (6.3) and (6.4) differ in their causal structure: in the learning scheme the input state $|\varphi\rangle$ cannot influence the state $|\Psi\rangle\rangle$ which the unitary is applied to; on the other hand, the general cloning scheme allows that the state $|\varphi\rangle$ can affect the input state of $\mathcal{U}$.

As pointed out in Remark 2.6, different causal structures reflect into different normalization of the Choi operators: for the the network $\mathcal{E}$ in Eq. (6.3) we have the constraint (see Eq. (5.3))

$$\mathrm{Tr}_{33'}[E] = I_2 \otimes E^{(1)} \qquad \mathrm{Tr}_1[E^{(1)}] = I_{00'}, \tag{6.5}$$

while for the learning scheme in Eq. (6.4) we have

$$\mathrm{Tr}_{33'}[L] = I_{22'} \otimes I_1 \otimes \rho \qquad \mathrm{Tr}_0[\rho] = 1, \tag{6.6}$$

(ρ is the partial state of $|\Psi\rangle\rangle$).

It is easy to prove that the constraint (6.6) is stronger than the constraint (6.5). Suppose that the operator E satisfies Eq. (6.6); then we have

$$\mathrm{Tr}_{33'}[E] = I_1 \otimes (\rho_0 \otimes I_{22'}) = I_1 \otimes E^{(1)} \qquad \mathrm{Tr}_0[E^{(1)}] = \mathrm{Tr}_0[\rho_0 \otimes I_{22'}] = I_{22'} \tag{6.7}$$

that coincides with Eq. (6.5) if we relabel 2 → 0, 1 → 2, $2' \rightarrow 0'$ and 0 → 1.

This proves that the cloning scheme is more general than the learning scheme and contains the latter as a special case. As a consequence we will show that the performances of the learning network are indeed worse than the performances of the cloning network.

6.1 Optimization of quantum learning

In this section, based on Ref. [27] we derive the optimal quantum learning of an unknown unitary randomly drawn from a group representation. The search of the optimal learning process can be divided into two steps:

- optimizing the *storing* network $\mathcal{S}$, that is the device that encodes the unknown transformation $\mathcal{U}$ into the state $|\Psi_{\mathcal{U}}\rangle\rangle$ of a quantum memory;

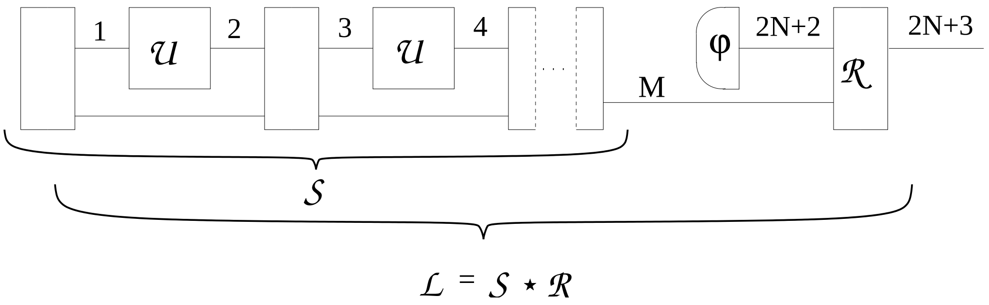


Figure 6.1. The learning process is described by a quantum network $\mathcal{L} = \mathcal{S} \star \mathcal{R}$ in which the N uses of $\mathcal{U}$ are plugged, along with the state $|\varphi\rangle$. The wires represent the input-output Hilbert spaces. The output of the network $\mathcal{S}$ is stored in a quantum memory M, later used by the retrieving channel $\mathcal{R}$

- finding the optimal *retrieving* channel $\mathcal{C}$, that receives $|\Psi_{\mathcal{U}}\rangle\rangle$ and an unknown state $|\phi\rangle$ as input and emulates $\mathcal{U}$ applied to $|\phi\rangle$.

An alternative to coherent retrieval is to estimate $\mathcal{U}$, to store the outcome in a classical memory, and to perform the estimated unitary on the new input state. This incoherent strategy has the double advantage of avoiding the expensive use of a quantum memory, and of allowing one to reproduce $\mathcal{U}$ an unlimited number of times with constant quality. However, incoherent strategies are typically suboptimal for the similar task of quantum cloning, and this would suggest that a coherent retrieval achieves better performances. Surprisingly enough, we find that the incoherent strategies already achieve the ultimate performance of quantum learning. We analyze the case in which U is a completely unknown unitary in a group $\mathbf{G}$, and we find that the performances of the optimal retrieving machine are equal to those of the optimal estimation.

We will show that the solution has the following structure:

- apply the N of the unknown unitary in parallel on a suitable entangled state;

- estimate the unknown unitary by measuring the state of the quantum memory

- produce the estimated unitary M times where M is the number of replicas that are required.

6.1.1 Considered scenario: $M = 1$

We tackle the optimization of learning starting from the case $M = 1$.

Referring to Fig. 6.1, we label the Hilbert spaces of quantum systems according to the following criterion: $\mathcal{H}_{2j-1}$ is the input of the j-th example of $\mathcal{U}$, and $\mathcal{H}_{2j}$ is the corresponding output. We denote by $\mathcal{H}_{\mathrm{i}} = \bigotimes_{j=1}^{N} \mathcal{H}_{2j-1}$ the Hilbert spaces of all inputs and by $\mathcal{H}_{\mathrm{o}} = \bigotimes_{j=1}^{N} \mathcal{H}_{2j}$ the Hilbert spaces of all outputs of the N examples. Alice's input state $|\varphi\rangle$ belongs to $\mathcal{H}_{2N+2}$, and the output state finally produced by Bob belongs to $\mathcal{H}_{2N+3}$. All spaces $\mathcal{H}_j$ considered here are $d-$dimensional, except the spaces $\mathcal{H}_0$ and $\mathcal{H}_{2N+1}$ which are one-dimensional, and are introduced just for notational convenience.

The Choi operator $L \in \mathcal{L}(\mathcal{H}_{\mathrm{i}} \otimes \mathcal{H}_{\mathrm{o}} \otimes \mathcal{H}_{2N+2} \otimes \mathcal{H}_{2N+3})$ of the learning network $\mathcal{L}$ satisfies the normalization condition (2.60), that becomes

$$\mathrm{Tr}_{2k-1}[L^{(k)}] = I_{2k-2} \otimes L^{(k-1)} \qquad k = 1, 2, \dots, N+2 \tag{6.8}$$

where $L^{(N+2)} = L$, $L^{(0)} = 1$, and $L^{(k)} \in \mathcal{L}(\mathcal{H}_j)_{j=0}^{2k-1}$.

When we insert the N example in the learning board we obtain the network

$$\mathcal{C}_U = \mathcal{L} \star \mathcal{U} \star \cdots \star \mathcal{U} \tag{6.9}$$

and then, according to Theorem 2.12, its Choi-Jamiołkowsky operator is given by

$$\begin{aligned} C_U &= L * |U\rangle\!\rangle\langle\!\langle U|_{12} * \cdots * |U\rangle\!\rangle\langle\!\langle U|_{(2N-1)(2N)} = \\ &= \mathrm{Tr}_{\mathrm{i,o}}\left[L\left(I_{2N+3} \otimes I_{2N+2} \otimes (|U\rangle\!\rangle\langle\!\langle U|^{\otimes N})^T\right)\right] = \langle\!\langle U^*|^{\otimes N} L |U^*\rangle\!\rangle^{\otimes N}. \end{aligned} \tag{6.10}$$

We now need to introduce a figure of merit that quantifies how close the resulting channel $\mathcal{C}_U$ is to the original unitary transformation $\mathcal{U}$. A reasonable choice is to maximize the channel fidelity $\mathcal{F}$ (see Definition A.1 and the following lemmas) between $\mathcal{C}_U$ and the target unitary averaged over U:

$$F := \int_{\mathbf{G}} \mathrm{d}U \mathcal{F}(\mathcal{C}_U, \mathcal{U}) = \frac{1}{d^2}\int_{\mathbf{G}} \mathrm{d}U \langle\!\langle U|C_U, |U\rangle\!\rangle) \tag{6.11}$$

Inserting Eq. (6.10) into Eq. (6.11) we have

$$F = \frac{1}{d^2}\int_{\mathbf{G}} \mathrm{d}U \langle\!\langle U|_{(2N+3)(2N+2)} \langle\!\langle U^*|_{\mathsf{o}\,\mathsf{i}}^{\otimes N} L |U^*\rangle\!\rangle_{\mathsf{o}\,\mathsf{i}}^{\otimes N} |U\rangle\!\rangle_{(2N+3)(2N+2)}. \tag{6.12}$$

The following lemma simplifies the search for the optimal learning network

Lemma 6.1 *The operator L maximizing the fidelity (6.12) can be assumed without loss of generality to satisfy the following commutation relation*

$$[L, U_{\mathrm{o}}^{*\otimes N} \otimes V_{\mathrm{i}}^{\otimes N} \otimes U_{2N+3} \otimes V^*_{2N+2}] = 0 \qquad \forall U, V \in \mathbf{G}. \tag{6.13}$$

Proof. Let L be the Choi operator of the optimal learning network; if we define

$$\overline{L} = \int \mathrm{d}Z\, \mathrm{d}W (Z_{\mathrm{o}}^{*\otimes N} \otimes W_{\mathrm{i}}^{\otimes N} \otimes Z_{2N+3} \otimes W^*_{2N+2})^\dagger L (Z^{*\otimes N} \otimes W^{\otimes N} \otimes Z \otimes W^*),$$

exploiting the properties of the Haar measure (see Definition B.7), we have

$$\begin{aligned} &\frac{1}{d^2}\int_{\mathbf{G}} \mathrm{d}U \langle\!\langle U|_{(2N+3)(2N+2)} \langle\!\langle U^*|_{\mathsf{o}\,\mathsf{i}}^{\otimes N} \overline{L} |U^*\rangle\!\rangle_{\mathsf{o}\,\mathsf{i}}^{\otimes N} |U\rangle\!\rangle_{(2N+3)(2N+2)} = \\ &= \frac{1}{d^2}\int_{\mathbf{G}} \mathrm{d}U \langle\!\langle ZUW^\dagger|_{(2N+3)(2N+2)} \langle\!\langle Z^*U^*W^T|_{\mathsf{o}\,\mathsf{i}}^{\otimes N} \overline{L} |Z^*U^*W^T\rangle\!\rangle^{\otimes N} |ZUW^\dagger\rangle\!\rangle = \\ &= \frac{1}{d^2}\int_{\mathbf{G}} \mathrm{d}U \langle\!\langle U|_{(2N+3)(2N+2)} \langle\!\langle U^*|_{\mathsf{o}\,\mathsf{i}}^{\otimes N} L |U^*\rangle\!\rangle_{\mathsf{o}\,\mathsf{i}}^{\otimes N} |U\rangle\!\rangle_{(2N+3)(2N+2)} \end{aligned} \tag{6.14}$$

that is $\overline{L}$ and L give the same value of F. $\overline{L}$, thanks to Theorem B.3, enjoys the property (6.13). Finally, it is easy to verify that $\overline{L}$ satisfies the constraint (6.8).

6.1.2 Optimization of the storing strategy

Lemma 6.1 allows us to look for the optimal learning network among the ones that satisfy Eq. (6.13).

Using Eq. (6.8) with $k = N + 2$ we have

$$\mathrm{Tr}_{2N+3}[L] = I_{2N+2} \otimes L^{(N+1)}. \tag{6.15}$$

The commutation (6.13) can be rewritten as

$$(U_{\mathrm{o}}^{*\otimes N} \otimes V_{\mathrm{i}}^{\otimes N} \otimes U_{2N+3} \otimes V^{*}_{2N+2})^{\dagger} L U_{\mathrm{o}}^{*\otimes N} \otimes V_{\mathrm{i}}^{\otimes N} \otimes U_{2N+3} \otimes V^{*}_{2N+2} = L \tag{6.16}$$

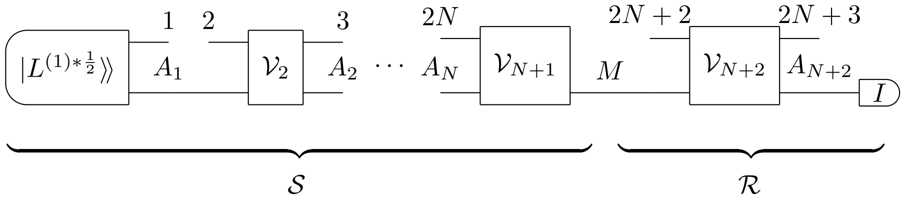
Taking the trace over $\mathcal{H}_{2N+3}$ in Eq. (6.16) and using Eq. (6.16) we get

$$\begin{aligned}
&\mathrm{Tr}_{2N+3}[U_{\mathrm{o}}^{*\otimes N} \otimes V_{\mathrm{i}}^{\otimes N} \otimes U_{2N+3} \otimes V^{*}_{2N+2})^{\dagger} L U_{\mathrm{o}}^{*\otimes N} \otimes V_{\mathrm{i}}^{\otimes N} \otimes U_{2N+3} \otimes V^{*}_{2N+2}] = \\
&= \mathrm{Tr}_{2N+3}[L] \Rightarrow U_{\mathrm{o}}^{*\otimes N} \otimes V_{\mathrm{i}}^{\otimes N})^{\dagger} L^{(N+1)} U_{\mathrm{o}}^{*\otimes N} \otimes V_{\mathrm{i}}^{\otimes N} = L^{(N+1)} \Rightarrow \\
&\Rightarrow [L^{(N+1)}, U_{\mathrm{o}}^{*\otimes N} \otimes V_{\mathrm{i}}^{\otimes N}] = 0.
\end{aligned} \tag{6.17}$$

We now prove that the commutation (6.17) implies that the parallel storage is optimal.

Lemma 6.2 (Optimality of parallel storage) *The optimal storage of U can be achieved by applying $U_{\mathsf{o}}^{\otimes N} \otimes I_{\mathsf{i}}^{\otimes N}$ on a suitable input state $|\Psi\rangle\!\rangle \in \mathcal{H}_{\mathsf{o}} \otimes \mathcal{H}_{\mathsf{i}}$.*

Proof. According to Th. 2.6 the learning Network $\mathcal{L}$ can be realized as a sequence of isometries, followed by a measurement on an ancillary space.



The storing network is then represented by the isometric channel $\mathcal{S} := \mathcal{W}^{(N+1)} := W^{(N+1)} \cdot W^{(N+1)\dagger}$ where $W^{(N+1)} = V^{(N)} \cdots V^{(1)} = I_{\mathsf{o}} \otimes L_{\mathsf{o}'\,\mathsf{i}'}^{(N+1)*\frac{1}{2}} |I\rangle\!\rangle_{\mathsf{o}\,\mathsf{o}'} \otimes T_{\mathsf{i}\rightarrow\mathsf{i}'}$ and $\mathcal{H}_M =$ $\mathsf{Supp}(L_{\mathsf{o}'\,\mathsf{i}'}^{(N+1)*\frac{1}{2}})$. The Choi Jamiolkowski operator of the storing network is then $S = W^{(N+1)}$ $|I\rangle\!\rangle\langle\!\langle I|_{\mathsf{i}\,\mathsf{i}'} W^{(N+1)\dagger} = |L^{(N+1)*\frac{1}{2}}\rangle\!\rangle\langle\!\langle L^{(N+1)*\frac{1}{2}}|_{\mathsf{o}\,\mathsf{i}\,\mathsf{o}'\,\mathsf{i}'}$ When we connect the storing board with the N copies of the unitary the final state on space $\mathcal{H}_M$ becomes

$$\begin{aligned}
&|\Psi_U\rangle\!\rangle\langle\!\langle\Psi_U| := S * |U\rangle\!\rangle\langle\!\langle U|_{12} * \cdots * |U\rangle\!\rangle\langle\!\langle U|_{(2N-1)(2N)} = \\
&= |L^{(N+1)*\frac{1}{2}}\rangle\!\rangle\langle\!\langle L^{(N+1)*\frac{1}{2}}|_{\mathsf{o}\,\mathsf{i}\,\mathsf{o}'\,\mathsf{i}'} * |U\rangle\!\rangle\langle\!\langle U|_{12} * \cdots * |U\rangle\!\rangle\langle\!\langle U|_{(2N-1)(2N)} =
\end{aligned}$$

and exploiting Eq. (6.17) we have

$$
\begin{aligned}
|\Psi_U\rangle\!\rangle &= \langle\!\langle U^*|_{\mathsf{o}\,\mathsf{i}}^{\otimes N}|L^{(N+1)*\frac{1}{2}}\rangle\!\rangle_{\mathsf{o}\,\mathsf{i}\,\mathsf{o}'\,\mathsf{i}'} = \langle\!\langle I|_{\mathsf{o}\,\mathsf{i}}^{\otimes N}(U^{T}{}_{\mathsf{o}}^{\otimes N}\otimes I_{\mathsf{i}})|L^{(N+1)*\frac{1}{2}}\rangle\!\rangle_{\mathsf{o}\,\mathsf{i}\,\mathsf{o}'\,\mathsf{i}'} = \\
&= \langle\!\langle I|_{\mathsf{o}\,\mathsf{i}}^{\otimes N}|(U^{T}{}_{\mathsf{o}}^{\otimes N}\otimes I_{\mathsf{i}})L^{(N+1)*\frac{1}{2}}\rangle\!\rangle_{\mathsf{o}\,\mathsf{i}\,\mathsf{o}'\,\mathsf{i}'} = \langle\!\langle I|_{\mathsf{o}\,\mathsf{i}}^{\otimes N}|L^{(N+1)*\frac{1}{2}}(U^{T}{}_{\mathsf{o}}^{\otimes N}\otimes I_{\mathsf{i}})\rangle\!\rangle = \\
&= \langle\!\langle I|_{\mathsf{o}\,\mathsf{i}}^{\otimes N}(U_{\mathsf{o}'}^{\otimes N}\otimes I_{\mathsf{i}'})|L^{(N+1)*\frac{1}{2}}\rangle\!\rangle_{\mathsf{o}\,\mathsf{i}\,\mathsf{o}'\,\mathsf{i}'} = (U_{\mathsf{o}'}^{\otimes N}\otimes I_{\mathsf{i}'})\langle\!\langle I|_{\mathsf{o}\,\mathsf{i}}^{\otimes N}|L^{(N+1)*\frac{1}{2}}\rangle\!\rangle_{\mathsf{o}\,\mathsf{i}\,\mathsf{o}'\,\mathsf{i}'} \\
&= (U_{\mathsf{o}'}^{\otimes N}\otimes I_{\mathsf{i}'})|\Psi\rangle\!\rangle_{\mathsf{o}'\,\mathsf{i}'}.
\end{aligned}
$$

where we defined $|\Psi\rangle\!\rangle_{\mathsf{o}\,\mathsf{i}\,\mathsf{o}'\,\mathsf{i}'} = \langle\!\langle I|_{\mathsf{o}\,\mathsf{i}}^{\otimes N}|L^{(N+1)*\frac{1}{2}}\rangle\!\rangle_{\mathsf{o}'\,\mathsf{i}'}$. Then every storing board can be realized applying $(U_{\mathsf{o}'}^{\otimes N}\otimes I_{\mathsf{i}'})$ to a suitable input state $|\Psi\rangle\!\rangle \in \mathcal{H}_{\mathsf{o}'\,\mathsf{i}'}$. ■

Remark 6.1 *It is possible to prove that the optimality of a parallel strategy is a common feature of all the problems involving estimation of group transformations. However, the only covariance (6.17) does not imply that the Quantum Network can be parallelized; a crucial aspect of the problem is that we have access to the physical transformation U and that the scheme $(U^{\otimes N}\otimes I)|\Psi\rangle\!\rangle$ is physically realizable. We will see (see Chapter 9) that there are cases in which the quantum storing network is covariant but it cannot be parallelized because the set transformations R_U we want to learn, even if they are orbit of a group representation (e.g. $R_U = UR_IU^\dagger$), do not form a group; In this case, an analogous of Eq. (6.17) holds but since we do not have physical access to the unitaries U, the optimal network cannot be assumed to be parallel.*

Optimization of learning is then reduced to finding the optimal input state $|\Psi\rangle$ and the optimal retrieving channel $\mathcal{R}$. The fidelity can be computed substituting $L = R * S$ in Eq. (6.12), and using the relation $\langle\!\langle U|\langle\!\langle U^*|^{\otimes N}(R*S)|U\rangle\!\rangle|U^*\rangle\!\rangle^{\otimes N} = \langle\!\langle U|R|U\rangle\!\rangle * \langle\!\langle U^*|^{\otimes N}S|U^*\rangle\!\rangle^{\otimes N} = \langle\!\langle U|R|U\rangle\!\rangle * |\Psi_U\rangle\!\rangle\langle\!\langle\Psi_U|$, which gives

$$
F = \frac{1}{d^2}\int_G \langle\!\langle U|\langle\Psi_U^*|R|U\rangle\!\rangle|\Psi_U^*\rangle \;\mathrm{d}U. \tag{6.18}
$$

The following lemma further simplifies the structure of the optimal input state for storage

Lemma 6.3 (Optimal states for storage) *The optimal input state for storage can be taken of the form*

$$
|\Psi\rangle\!\rangle = \bigoplus_j \sqrt{\frac{p_j}{d_j}}|I_j\rangle\!\rangle \in \widetilde{\mathcal{H}}\,, \tag{6.19}
$$

where p_j are probabilities, $\widetilde{\mathcal{H}} = \bigoplus_j(\mathcal{H}_j\otimes\mathcal{H}_j)$ is a subspace of $\mathcal{H}_{\mathsf{o}}\otimes\mathcal{H}_{\mathsf{i}}$ carrying the representation $\widetilde{U} = \bigoplus_j(U_j\otimes I_j)$, I_j being the identity in $\mathcal{H}_j$, and the index j labelling the irreducible representations U_j contained in the decomposition of $U^{\otimes N}$.

Proof. Let us consider the local state

$$
\begin{aligned}
\rho &:= \mathrm{Tr}_{\mathsf{i}'}[|\Psi\rangle\!\rangle\langle\!\langle\Psi|] = \mathrm{Tr}_{\mathsf{i}'}[\langle\!\langle I|_{\mathsf{o}\,\mathsf{i}}^{\otimes N}|L^{(N+1)*\frac{1}{2}}\rangle\!\rangle\langle\!\langle L^{(N+1)*\frac{1}{2}}|_{\mathsf{o}\,\mathsf{i}\,\mathsf{o}'\,\mathsf{i}'}|I\rangle\!\rangle_{\mathsf{o}\,\mathsf{i}}^{\otimes N}] = \\
&= \mathrm{Tr}_{\mathsf{i}'}[L_{\mathsf{o}'\,\mathsf{i}'}^{(N+1)\frac{1}{2}}\langle\!\langle I|_{\mathsf{o}\,\mathsf{i}}^{\otimes N}|I\rangle\!\rangle\langle\!\langle I|_{\mathsf{o}\,\mathsf{i}\,\mathsf{o}'\,\mathsf{i}'}|I\rangle\!\rangle_{\mathsf{o}\,\mathsf{i}}^{\otimes N}L_{\mathsf{o}'\,\mathsf{i}'}^{(N+1)\frac{1}{2}}] = \\
&= \mathrm{Tr}_{\mathsf{i}'}[L_{\mathsf{o}'\,\mathsf{i}'}^{(N+1)\frac{1}{2}}|I\rangle\!\rangle\langle\!\langle I|_{\mathsf{o}'\,\mathsf{i}'}L_{\mathsf{o}'\,\mathsf{i}'}^{(N+1)\frac{1}{2}}]
\end{aligned}
$$

It is easy to prove that $\rho \in \mathcal{L}(\mathcal{H}_{\mathsf{o}'})$ is invariant under $U^{\otimes N}$:

$$\begin{aligned}
U^{\otimes N} \rho U^{\dagger \otimes N} &= U^{\otimes N} \operatorname{Tr}_{\mathsf{i}'}[L_{\mathsf{o}'\,\mathsf{i}'}^{(N+1)\frac{1}{2}} |I\rangle\!\rangle\langle\!\langle I|_{\mathsf{o}'\,\mathsf{i}'} L_{\mathsf{o}'\,\mathsf{i}'}^{(N+1)\frac{1}{2}} U^{\dagger\otimes N} = \\
&= \operatorname{Tr}_{\mathsf{i}'}[(U^{\otimes N} \otimes I_{\mathsf{i}'}) L_{\mathsf{o}'\,\mathsf{i}'}^{(N+1)\frac{1}{2}} |I\rangle\!\rangle\langle\!\langle I|_{\mathsf{o}'\,\mathsf{i}'} L_{\mathsf{o}'\,\mathsf{i}'}^{(N+1)\frac{1}{2}} (U^{\dagger\otimes N} \otimes I_{\mathsf{i}'})] = \\
&= \operatorname{Tr}_{\mathsf{i}'}[L_{\mathsf{o}'\,\mathsf{i}'}^{(N+1)\frac{1}{2}} (U^{\otimes N} \otimes I_{\mathsf{i}'}) |I\rangle\!\rangle\langle\!\langle I|_{\mathsf{o}'\,\mathsf{i}'} (U^{\dagger\otimes N} \otimes I_{\mathsf{i}'}) L_{\mathsf{o}'\,\mathsf{i}'}^{(N+1)\frac{1}{2}}] = \\
&= \operatorname{Tr}_{\mathsf{i}'}[L_{\mathsf{o}'\,\mathsf{i}'}^{(N+1)\frac{1}{2}} (I_{\mathsf{o}'} \otimes U^{T\otimes N}) |I\rangle\!\rangle\langle\!\langle I|_{\mathsf{o}'\,\mathsf{i}'} (U^{\dagger\otimes N} \otimes I_{\mathsf{o}'}) L_{\mathsf{o}'\,\mathsf{i}'}^{(N+1)\frac{1}{2}}] = \\
&= \operatorname{Tr}_{\mathsf{i}'}[(I_{\mathsf{o}'} \otimes U^{T\otimes N}) L_{\mathsf{o}'\,\mathsf{i}'}^{(N+1)\frac{1}{2}} |I\rangle\!\rangle\langle\!\langle I|_{\mathsf{o}'\,\mathsf{i}'} L_{\mathsf{o}'\,\mathsf{i}'}^{(N+1)\frac{1}{2}} (U^{\dagger\otimes N} \otimes I_{\mathsf{o}'})] = \\
&= \operatorname{Tr}_{\mathsf{i}'}[L_{\mathsf{o}'\,\mathsf{i}'}^{(N+1)\frac{1}{2}} |I\rangle\!\rangle\langle\!\langle I|_{\mathsf{o}'\,\mathsf{i}'} L_{\mathsf{o}'\,\mathsf{i}'}^{(N+1)\frac{1}{2}}] = \rho
\end{aligned}$$

Decomposing $U^{\otimes N}$ into irreducible representations (irreps) we have $U^{\otimes N} = \bigoplus_j (U_j \otimes I_{m_j})$, where I_{m_j} is the identity on an m_j-dimensional multiplicity space $\mathbb{C}^{m_j}$. Reminding theorem B.2, ρ must have the form $\rho = \bigoplus_j p_j (I_j/d_j \otimes \rho_j)$, where ρ_j is an arbitrary state on the multiplicity space $\mathbb{C}^{m_j}$. Since $|\Psi\rangle\!\rangle$ is a purification of ρ, there exists a basis in which we have $|\Psi\rangle\!\rangle = |\rho^{\frac{1}{2}}\rangle\!\rangle = \bigoplus_j \sqrt{p_j/d_j}\; |I_j\rangle\!\rangle |\rho_j^{\frac{1}{2}}\rangle\!\rangle$, which after storage becomes $|\Psi_U\rangle\!\rangle = \bigoplus_j \sqrt{p_j/d_j} |U_j\rangle\!\rangle |\rho_j^{\frac{1}{2}}\rangle\!\rangle$. Hence, for every U the state $|\Psi_U\rangle\!\rangle$ belongs to the subspace $\widetilde{\mathcal{H}} = \bigoplus_j (\mathcal{H}_j^{\otimes 2} \otimes |\rho_j^{\frac{1}{2}}\rangle\!\rangle) \simeq \bigoplus_j \mathcal{H}_j^{\otimes 2}$. ■

6.1.3 Optimization of the retrieving channel

In this section we optimize the retrieving channel $\mathcal{R}$; exploiting some symmetries of $\mathcal{R}$ we can prove that the optimal retrieval is achieved by a measure and re-prepare strategy.

Thanks to Lemma 6.3 we can restrict our attention to the subspace $\widetilde{\mathcal{H}}$, and consider retrieving channels $\mathcal{R}$ from $(\mathcal{H}_{2N+2} \otimes \widetilde{\mathcal{H}})$ to $\mathcal{H}_{2N+3}$. The normalization of the Choi operator is then

$$\operatorname{Tr}_{2N+3}[R] = I_{2N+2} \otimes I_{\widetilde{\mathcal{H}}}\ . \tag{6.20}$$

The following lemma tells us that the optimal retrieving channel can be chosen among the covariant ones:

Lemma 6.4 *We can require without loss of generality that the operator R maximizing the fidelity (6.18) satisfies the commutation relation*

$$\left[R, U_{2N+3} \otimes V^*_{2N+2} \otimes \left(\bigoplus_j (U_j^* \otimes V_j)\right)\right] = 0 \qquad \forall U, V \in \mathbf{G}. \tag{6.21}$$

where $\bigoplus_j (U_j^* \otimes V_j)$ acts on $\widetilde{\mathcal{H}}$.

Proof. The proof consists in the same averaging argument that was used in the proof of Lemma 6.1. ■

According to Eq. (B.16), the representation $U \otimes U_j^*$ can be decomposed as

$$U_{2N+3} \otimes U_j^* = \bigoplus_{K \in \mathsf{irrepS}(U \otimes U_j^*)} \left(U_K \otimes I_{m_K^{(j)}}\right) \tag{6.22}$$

and in a similar way we have

$$V^*_{2N+2} \otimes V_j = \bigoplus_{L \in \mathsf{irrepS}(V^* \otimes V_j)} \left(V^*_L \otimes I_{m^{(j)}_L} \right) . \tag{6.23}$$

Combining Eq. (6.22) and Eq. (6.23) we have

$$\begin{aligned} U_{2N+3} \otimes V^*_{2N+2} \otimes \left(\bigoplus_j (U^*_j \otimes V_j) \right) &= \bigoplus_j \big((U \otimes U^*_j) \otimes (V^* \otimes V_j) \big) = \\ &= \bigoplus_{K,L} (U_K \otimes V_L \otimes I_{m_{KL}}) \end{aligned} \tag{6.24}$$

where $I_{m_{KL}}$ is given by $I_{m_{KL}} = \bigoplus_{j \in \mathsf{P}_{KL}} \left(I_{m^{(j)}_K} \otimes I_{m^{(j)}_L} \right)$, where P_{KL} is the set of values of j such that the irrep $U_K \otimes V^*_L$ is contained in the decomposition of $U_{2N+3} \otimes V^*_{2N+2} \otimes U^*_j \otimes V_j$.

Inserting the decomposition (6.24) into Eq. (6.21) we have

$$\left[R, \bigoplus_{K,L} (U_K \otimes V_L \otimes I_{m_{KL}}) \right] = 0 \tag{6.25}$$

that thanks to Theorem B.2, leads to the decomposition

$$R = \bigoplus_{K,L} (I_K \otimes I_L \otimes R_{KL}) \tag{6.26}$$

where R_{KL} is a positive operator on the multiplicity space $\mathbb{C}^{m_{KL}} = \bigoplus_{j \in \mathsf{P}_{KL}} \left(\mathbb{C}^{m^{(j)}_K} \otimes \mathbb{C}^{m^{(j)}_L} \right)$

The decomposition (6.22) induces the following decomposition of Hilbert spaces

$$\mathcal{H} \otimes \mathcal{H}_j = \bigoplus_{K \in \mathsf{irrepS}(U \otimes U^*_j)} \left(\mathcal{H}_K \otimes \mathbb{C}^{m^{(j)}_K} \right) \tag{6.27}$$

that allows us to write

$$I \otimes I_J = \bigoplus_{K \in \mathsf{irrepS}(U \otimes U^*_j)} \left(I_K \otimes I_{m^{(j)}_K} \right) . \tag{6.28}$$

From Eq. (6.28) we have

$$|I\rangle\!\rangle \otimes |I_J\rangle\!\rangle = \bigoplus_{K \in \mathsf{irrepS}(U \otimes U^*_j)} \left(|I_K\rangle\!\rangle \otimes |I\rangle\!\rangle_{m^{(j)}_K} \right) \tag{6.29}$$

that leads to the following identity:

$$\begin{aligned} |I\rangle\!\rangle |\Psi^*\rangle &= \bigoplus_j \sqrt{\frac{p_j}{d_j}} |I\rangle\!\rangle |I_j\rangle\!\rangle = \bigoplus_j \bigoplus_{K \in \mathsf{irrepS}(U \otimes U^*_j)} \sqrt{\frac{p_j}{d_j}} |I_K\rangle\!\rangle |I_{m^{(j)}_K}\rangle\!\rangle \\ &= \bigoplus_K \bigoplus_{j \in \mathsf{P}_{KK}} \sqrt{\frac{p_j}{d_j}} |I_K\rangle\!\rangle |I_{m^{(j)}_K}\rangle\!\rangle = \bigoplus_K |I_K\rangle\!\rangle |\alpha_K\rangle \ , \end{aligned} \tag{6.30}$$

where $|I_K\rangle\!\rangle \in \mathcal{H}_K^{\otimes 2}$ and $|\alpha_K\rangle \in \mathbb{C}^{m_{KK}}$ is given by

$$|\alpha_K\rangle = \bigoplus_{j \in \mathsf{P}_{KK}} \sqrt{p_j/d_j}\, |I_{m_K^{(j)}}\rangle\!\rangle. \tag{6.31}$$

Exploiting Eqs. (6.26) and (6.30) the fidelity (6.18) can be rewritten as

$$F = \sum_K \frac{d_K}{d^2} \langle \alpha_K | R_{KK} | \alpha_K \rangle \ . \tag{6.32}$$

We now prove that the optimal retrieving consists in a measure and re-prepare channel; we split the derivation into two parts.

Lemma 6.5 *For the fidelity in Eq. (6.32) the following bound holds*

$$F \leqslant \sum_K \frac{\left(\sum_{j \in \mathsf{P}_{\mathsf{KK}}} m_K^{(j)} \sqrt{p_j}\right)^2}{d^2} \tag{6.33}$$

where we remind that $m_K^{(j)}$ is the dimension of the multiplicity space $\mathcal{C}^{m_K^{(j)}}$ and that where P_{KK} is the set of values of j such that the irreducible representation $U_K \otimes V_K^*$ is contained in the decomposition of $U \otimes V^* \otimes U_j^* \otimes V_j$.

Proof. Taking the trace over $\mathcal{H}_{2N+3}$ into Eq. (6.21) gives

$$\left[\mathrm{Tr}_{2N+3}[R], V_{2N+2}^* \otimes \left(\bigoplus_j (U_j^* \otimes V_j) \right) \right] = 0; \tag{6.34}$$

reminding the decomposition (6.22) and exploiting Theorem B.2 we can write

$$\mathrm{Tr}_{2N+3}[R] = \bigoplus_j I_j \otimes \left(\bigoplus_L I_L \otimes r_L^{(j)} \right) \tag{6.35}$$

where $r_L^{(j)}$ is a positive operator acting on $\mathbb{C}^{m_L^{(j)}}$.

Comparing Eq. (6.26) traced over $\mathcal{H}_{2N+3}$ with Eq. (6.35) we have

$$\begin{aligned} &\bigoplus_L \left(\bigoplus_j I_j \otimes r_L^{(j)} \right) \otimes I_L = \bigoplus_L \left(\bigoplus_K \mathrm{Tr}_{2N+3}[I_K \otimes R_{KL}] \right) \otimes I_L \Rightarrow \\ &\Rightarrow \bigoplus_j I_j \otimes r_L^{(j)} = \bigoplus_K \mathrm{Tr}_{2N+3}[I_K \otimes R_{KL}] \end{aligned} \tag{6.36}$$

Let us now denote with P_j the projector on $\mathcal{H}_j$ with P_K the projector on $\mathcal{H}_K$ and with $P_K^{(j)}$ the projector on $\mathbb{C}^{m_K^{(j)}}$: we can then rewrite the decomposition (6.28) as $P_j \otimes I = \sum_K P_K \otimes P_K^{(j)}$

Projecting both the two sides of Eq. (6.36) on $\mathcal{H}_j \otimes \mathbb{C}^{m_L^{(j)}}$ we get

$$\begin{aligned}
I_j \otimes r_L^{(j)} &= (P_j \otimes P_L^{(j)}) \bigoplus_K \mathrm{Tr}_{2N+3}[I_K \otimes R_{KL}](P_j \otimes P_L^{(j)}) = \\
&= \bigoplus_K \mathrm{Tr}_{2N+3}[(P_j \otimes I_{2N+3} \otimes P_L^{(j)}) I_K \otimes R_{KL} (P_j \otimes I_{2N+3} \otimes P_L^{(j)})] \\
&= \bigoplus_K \mathrm{Tr}_{2N+3} \left[\sum_Q (P_Q \otimes P_Q^{(j)} \otimes P_L^{(j)}) I_K \otimes R_{KL} \sum_G (P_G \otimes P_G^{(j)} \otimes P_L^{(j)}) \right] \\
&= \bigoplus_K \mathrm{Tr}_{2N+3}[I_K \otimes R_{KL}^j]. \qquad (6.37)
\end{aligned}$$

where we used the notation $R_{KL}^j = (P_K^{(j)} \otimes P_L^{(j)}) R_{KL} (P_K^{(j)} \otimes P_L^{(j)})$. Taking the trace over $\mathcal{H}_j$ in Eq. (6.37) leads to

$$\begin{aligned}
&\mathrm{Tr}_j[I_j \otimes r_L^{(j)}] = \mathrm{Tr}_j \left[\bigoplus_K \mathrm{Tr}_{2N+3}[I_K \otimes R_{KL}^j] \right] \Rightarrow \\
&\Rightarrow d_j r_L^{(j)} = \mathrm{Tr}_{j\,2N+3} \left[\bigoplus_K I_K \otimes R_{KL}^j \right] = \\
&= \mathrm{Tr}_{(\bigoplus_K K \otimes m_K^{(j)})} \left[\bigoplus_K I_K \otimes R_{KL}^j \right] = \sum_K d_K \, \mathrm{Tr}_{m_K^{(j)}} \left[R_{KL}^j \right] \qquad (6.38)
\end{aligned}$$

where $\mathrm{Tr}_{(\bigoplus_K K \otimes m_K^{(j)})}$ denotes the trace over $\bigoplus_K \mathcal{H}_K \otimes C^{m_K^j} = \mathcal{H}_j \otimes \mathcal{H}_{2N+3}$.

Exploiting Eq. (6.35) into the normalization (6.20) we obtain

$$\mathrm{Tr}_{2N+3}[R] = \bigoplus_j I_j \otimes \left(\bigoplus_L I_L \otimes r_L^{(j)} \right) = I_{2N+2} \otimes I_{\widetilde{\mathcal{H}}} \Rightarrow r_L^{(j)} = I_{m_L^{(j)}} \qquad (6.39)$$

that together with Eq. (6.38) gives

$$I_{m_L^{(j)}} = \sum_K \frac{d_K}{d_j} \mathrm{Tr}_{m_K^{(j)}} \left[R_{KL}^j \right] \qquad (6.40)$$

that for $L = K$ implies the bound

$$\mathrm{Tr}[R_{KL}^j] \leqslant \frac{d_j m_K^{(j)}}{d_K} \qquad (6.41)$$

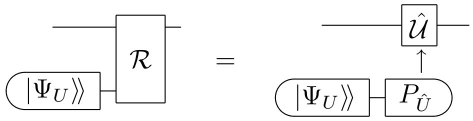
Reminding Eq. (6.31), for the fidelity (6.32) we then have the bound

$$\begin{aligned}
F &= \sum_K \frac{d_K}{d^2} \sum_{j,j' \in \mathsf{P}_{KK}} \sqrt{\frac{p_j p_{j'}}{d_j d_{j'}}} \langle\langle I_{m_K^{(j)}} | R_{KK} | I_{m_K^{(j')}} \rangle\rangle \leqslant \qquad (6.42) \\
&\leq \sum_K \frac{d_K}{d^2} \left(\sum_{j \in \mathsf{P}_{KK}} \sqrt{\frac{p_j \langle\langle I_{m_K^{(j)}} | R_{KK}^{(j)} | I_{m_K^{(j)}} \rangle\rangle}{d_j}} \right)^2 \leqslant \sum_K \frac{\left(\sum_{j \in \mathsf{P}_{\mathsf{KK}}} m_K^{(j)} \sqrt{p_j} \right)^2}{d^2}
\end{aligned}$$

having used the positivity of R_{KK} for the first bound and Eq. (6.41) the second. ■

It is now easy to prove the following

Theorem 6.2 (Optimal retrieving strategy) *The optimal retrieving of U from the memory state $|\Psi_U\rangle\!\rangle$ is achieved by measuring the ancilla with the POVM $P_{\hat{U}} = |\eta_{\hat{U}}\rangle\!\rangle\langle\!\langle\eta_{\hat{U}}|$ given by $|\eta_{\hat{U}}\rangle\!\rangle = \bigoplus_j \sqrt{d_j}|\hat{U}_j\rangle\!\rangle$, and, conditionally on outcome $\hat{U}$, by performing the unitary $\hat{U}$ on the new input system*



(6.43)

(the arrow represents the communication of the classical outcome of the measurement).

Proof. We now prove that the measure and prepare strategy described above achieves the bound (6.33). First, the Choi operator of the measure-and-prepare strategy has the form $R_{est} = \int_G |\hat{U}\rangle\!\rangle\langle\!\langle\hat{U}|_{(2N+3)(2N+2)} \otimes |\eta^*_{\hat{U}}\rangle\!\rangle\langle\!\langle\eta^*_{\hat{U}}|\,\mathrm{d}\hat{U}$. Using Eq. (6.30) with $|\Psi^*\rangle\!\rangle$ replaced by $|\eta^*_I\rangle\!\rangle$ and applying theorem B.3 we have

$$\begin{aligned}
R_{est} &= \int_G |\hat{U}\rangle\!\rangle\langle\!\langle\hat{U}| \otimes |\eta^*_{\hat{U}}\rangle\!\rangle\langle\!\langle\eta^*_{\hat{U}}|\,\mathrm{d}\hat{U} = \\
&= \int_G |\hat{U}V\rangle\!\rangle\langle\!\langle\hat{U}V| \otimes |\eta^*_{\hat{U}V}\rangle\!\rangle\langle\!\langle\eta^*_{\hat{U}V}|\,\mathrm{d}\hat{U}\,\mathrm{d}\hat{V} = \\
&= \int_G U \otimes V^T \otimes \widetilde{U^*} \otimes \widetilde{V^\dagger}|I\rangle\!\rangle\langle\!\langle I| \otimes |\eta^*_I\rangle\!\rangle\langle\!\langle\eta^*_I| U^\dagger \otimes V^* \otimes \widetilde{U^T} \otimes \widetilde{V}\,\mathrm{d}\hat{U}\,\mathrm{d}\hat{V} = \\
&= \int_G U \otimes V^T \otimes \widetilde{U^*} \otimes \widetilde{V^\dagger} \bigoplus_{KL} |I_k\rangle\!\rangle\langle\!\langle I_L|\,|\beta_K\rangle\,\langle\beta_L| U^\dagger \otimes V^* \otimes \widetilde{U^T} \otimes \widetilde{V}\,\mathrm{d}\hat{U}\,\mathrm{d}\hat{V} = \\
&= \bigoplus_K I_K \otimes I_K \otimes |\beta_K\rangle\,\langle\beta_K|
\end{aligned}$$

where $\widetilde{U^*} \otimes \widetilde{V^\dagger} = \bigoplus_j U_j \otimes V_j$ and $|\beta_K\rangle = \bigoplus_{j \in \mathsf{P}_{KK}} \sqrt{d_j}|I_{m^{(j)}_k}\rangle\!\rangle$. Eq. (6.32) then becomes

$$F_{est} = \sum_K \frac{|\langle\alpha_K|\beta_K\rangle|^2}{d^2} = \frac{\left(\sum_{j \in \mathsf{P}_{\mathsf{KK}}} m^{(j)}_K \sqrt{p_j}\right)^2}{d^2}. \tag{6.44}$$

■

By making use of the above result it is easy to optimize the input state for storing. In fact, such a state is just the optimal state for the estimation of the unknown unitary U [70], whose expression is known in most relevant cases. For example, when U is an unknown qubit unitary in $SU(2)$, learning becomes equivalent to optimal estimation of an unknown rotation in the Bloch sphere [71]. For large number of copies, the optimal input state is given by $|\Psi\rangle\!\rangle \approx \sqrt{4/N}\ \sum_{j=j_{\min}}^{N/2} \frac{\sin(2\pi j/N)}{\sqrt{2j+1}}\ |I_j\rangle\!\rangle$, with $j_{\min} = 0(1/2)$ for N even (odd), and the fidelity is

$F \approx 1-(2\pi^2)/N^2$. Remarkably, this asymptotic scaling can be achieved without using entanglement between the set of N qubits that are rotated and an auxiliary set of N rotationally invariant qubits: the optimal storing is achieved just by applying $U^{\otimes N}$ on a the optimal N-qubit state [71]. Another example is that of an unknown phase-shift $U = \exp[i\theta\sigma_z]$. In this case, for large number of copies the optimal input state is $|\Psi\rangle\!\rangle = \sqrt{2/(N+1)}\sum_{m=-N/2}^{N/2}\sin[\pi(m+1/2)/(N+1)]|m\rangle$ and the fidelity is $F \approx 1 - 2\pi^2/(N+1)^2$ [72]. Again, the optimal state can be prepared using only N qubits.

6.1.4 Generalization to the M > 1 case

Our result can be extended to the case where the user must reproduce $M > 1$ copies of the unknown unitary U. In this case, there are two different notions of optimality induced by two different figures of merit, namely the single-copy and the global fidelity. In the following we will examine both cases.

6.1.5 Optimal learning according to the single-copy fidelity

Let $\mathcal{C}_U$ be the M-partite channel obtained by the user, and $\mathcal{C}^{(i)}_{U,\Omega}$ be the local channel $\mathcal{C}^{(i)}_{U,\Omega}(\rho) = \mathrm{Tr}_{\bar{i}}[\mathcal{C}_U(\rho\otimes\Omega)]$, where ρ is the state of the i-th system, Ω is the state of the remaining $M-1$ systems, and $\mathrm{Tr}_{\bar{1}}$ denotes the trace over all systems except the i-th. The local channel $\mathcal{C}^{(i)}_{U,\Omega}$ describes the evolution of the i-th input of $\mathcal{C}_U$ when the remaining $(M-1)$ inputs are prepared in the state Ω. Since we can be interested in some replicas more than in other ones, we can imagine to associate a weight q_i ($\sum_i q_i = 1$) to each of the M copies; in this way the figure of merit becomes:

$$F^{(s)} = \int \mathrm{d}U \sum_i q_i \mathcal{F}(\mathcal{C}^{(i)}_{U,\Omega_i}, \mathcal{U}). \tag{6.45}$$

Of course, the fidelity between $\mathcal{C}^{(i)}_{U,\Omega_i}$ and the unitary U cannot be larger than the optimal fidelity of Eq. (6.33); moreover the optimal fidelity depends neither on q_i nor on Ω_i. Therefore, the measure-and-prepare strategy presented in Theorem 6.2 is optimal also for the maximization of Eq. (6.45), which do not decrease with increasing M.

6.1.6 Optimal learning according to the global fidelity

The optimization carried on for the case $M = 1$ can be extended to the maximization of the global fidelity between $\mathcal{C}_U$ and $U^{\otimes M}$

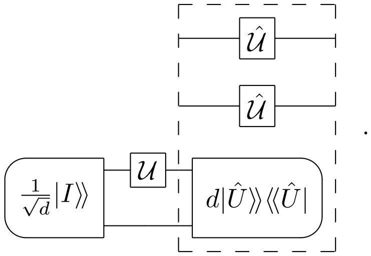
$$F^{(g)} = \int \mathrm{d}U \mathcal{F}(\mathcal{C}_U, \mathcal{U}^{\otimes M}) = \frac{1}{d^{2M}} \int \mathrm{d}U \langle\!\langle U|^{\otimes M} \langle\!\langle U^*|^{\otimes N} L |U\rangle\!\rangle^{\otimes M} |U^*\rangle\!\rangle^{\otimes N} \tag{6.46}$$

just by replacing U with $U^{\otimes M}$ in all derivations. Indeed, the role of the target unitary U in our derivations is completely generic: we never used the fact that the unitary emulated by the machine was equal to the unitaries provided in the examples. Therefore, following the same proofs for the case $M = 1$ it is immediate to see that also for the case of $M > 1$ copies with global fidelity the optimal strategy for storing consists in the parallel application of the examples on an input state of the form of Lemma 6.3 and that the optimal strategy for retrieving consists in measuring the optimal POVM $P_{\hat{U}}$ and in performing $\hat{U}^{\otimes M}$ conditionally on outcome $\hat{U}$. Note that in this case the coefficients $\{p_j\}$ in the optimal input state of Lemma 6.3) generally depend on M.

Remark 6.2 *Since we never used the fact that the N examples are identical, all the previous results hold even when the input (output) uses are not identical copies $U^{\otimes N}$ ($U^{\otimes M}$), but generally N (M) different unitaries, each of them belonging to a different representation of the group G. For example, if $G = \mathrm{SO}(3)$ the N examples may correspond to rotations (of the same angle and around the same axis) of N quantum particles with different angular momenta. Of course, the same remark also holds when the M output copies.*

6.2 Comparison with the cloning

Let us now focus on the optimal learning according to the global fidelity for the $N = 1$ and $M = 2$ case Specializing Eq. (6.19) the optimal state for storage becomes $\Psi = \frac{1}{\sqrt{d}}|I\rangle\rangle$ and the optimal learning board is


$\hat{\mathcal{U}}$
$\hat{\mathcal{U}}$
$\frac{1}{\sqrt{d}}|I\rangle\rangle$
$\mathcal{U}$
$d|\hat{U}\rangle\rangle\langle\langle\hat{U}|$

. (6.47)

The maximum value of the fidelity is given by replacing $\bigoplus_j (U_j^* \otimes V_j)$ with $U^* \otimes V$ and $U_{2N+3} \otimes V_{2N+2}^*$ with $U_{2N+3} \otimes V_{2N+2}^* \otimes U_{2N+5} \otimes V_{2N+4}^*$ in the previous derivation. From the decomposition (B.44) we have that $m_\alpha = 2, m_\beta = 1, m_\gamma = 1$,($m_\gamma = 0$ if $d = 2$); inserting these values into Eq. (6.44) we get

$$F^{(g)}_{N=1,M=2} = \frac{1}{d^4}\sum_\nu (m_\nu)^2 = \\ = \frac{6}{d^4} \text{ for } d > 2, \quad \text{or} \quad \frac{5}{d^4} \text{ for } d = 2 \tag{6.48}$$

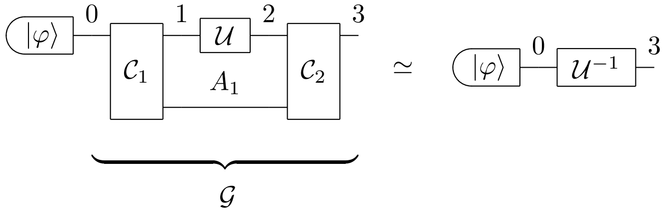
The learning with $N = 1$ and $M = 2$ can be compared with the optimal cloner $1 \rightarrow 2$ we derived in chapter 5. The maximum value of the fidelity was (see Eq: (5.24))

$$F^{(clon)}(d + \sqrt{d^2 - 1})/d^3 \tag{6.49}$$

which is much higher than $F^{(g)}$. This result stresses the difference between cloning and learning: since in the learning scenario we have to apply the unitary to a fix input state, we cannot exploit the full computational power of the unitary channel $\mathcal{U}$ and we cannot achieve the same performance of the optimal cloner.

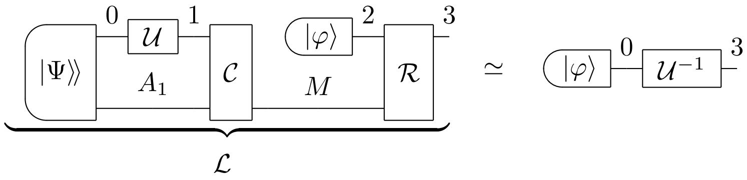
7 Inversion of a unitary transformation

In this chapter we consider the problem of finding the Quantum Network that realizes the optimal inversion of a unitary transformation. Let us suppose that we are provided with a single use of unitary transformation $\mathcal{U} = U \cdot U^\dagger$ but we need to apply its inverse $\mathcal{U}^{-1} = U^\dagger \cdot U$ on an unknown state $|\varphi\rangle$[16]. The most general strategy we can follow in order to achieve this task is to exploit the single use of $\mathcal{U}$ in a quantum network such that the resulting channel is as close as possible to target unitary $\mathcal{U}^{-1}$:



(7.1)

If the use of the unitary $\mathcal{U}$ is available only today while the state $|\varphi\rangle$ on which we need to apply the inverse $\mathcal{U}^{-1}$ will be provided tomorrow, we cannot apply the scheme in Eq. (7.1) and the best we can do is to apply a learning strategy (see Chapter 6):



(7.2)

We encountered the same situation when we compared the cloning and the learning of a unitary transformation. The Choi-Jamiołkowsky operators of $\mathcal{G}$ and $\mathcal{L}$ satisfy the conditions:

$$\mathrm{Tr}_3[G] = I_2 \otimes G^{(1)} \qquad \mathrm{Tr}_1[G^{(1)}] = I_0, \tag{7.3}$$

$$\mathrm{Tr}_3[L] = I_2 \otimes I_1 \otimes \rho \qquad \mathrm{Tr}_0[\rho] = 1, \tag{7.4}$$

that coincide with Eqs. (6.5) and (6.6) by defining $\mathcal{H}_0 \otimes \mathcal{H}_{0'} := \mathcal{H}_0$ and $\mathcal{H}_3 \otimes \mathcal{H}_{3'} := \mathcal{H}_3$.

As we noticed when we compared the learning and the cloning strategies, the constraint (7.4) is stronger than the constraint (7.3), and this means that the learning scheme in Eq. (7.2) can be interpreted as a special case of the scheme in Eq. (7.1).

In principle, one could expect that the strategy (7.1) allows to achieve better performances that the learning scheme (7.2). However, as we will see in the next sections, the optimal inversion is achieved by a measure and re-prepare strategy which is a special case of quantum learning.

7.1 Learning scenario

In this section we show that it is possible to extend the results of chapter 6 to the optimal learning of the inverse of an unknown unitary U. Then we can consider the more general scenario in which

[16] the generalization to the general case with N uses and and M replicas of the inverse is work in progress.

$N \geqslant 1$ uses of the unitary are available and $M \geqslant 1$ replicas have to be produced. The figure of merit is than the averaged channel fidelity between the inverses $\mathcal{U}^{-1^{\otimes M}}$ and the resulting replicas $\mathcal{L} \star \mathcal{U} \star \cdots \star \mathcal{U}$:

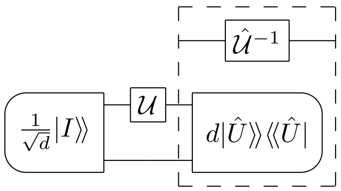
$$F = \frac{1}{d^{2M}} \int_G \langle\langle U^\dagger|^{\otimes M} \langle\langle U^*|^{\otimes N} \; L \; |U^\dagger\rangle\rangle^{\otimes M} |U^*\rangle\rangle^{\otimes N} \, \mathrm{d}U \tag{7.5}$$

as obtained by substituting U with $U^{\dagger \otimes M}$ in the target of Eq. (6.12). From this expression the commutation (6.13) becomes

$$[L, V^{\otimes M} \otimes U^{*\otimes M} \otimes U_o^{*\otimes N} \otimes V_i^{\otimes N}] = 0 \tag{7.6}$$

Therefore, the optimal inversion is obtained from our derivations by simply substituting $U_{2N+3} \rightarrow V^{\otimes M}$ and $V_{2N+2} \rightarrow U^{\otimes M}$. Accordingly, the optimal inversion is achieved by measuring the optimal POVM $P_{\hat{U}}$ on the optimal state $|\Psi_U\rangle\rangle$ and by performing $\hat{U}^{\dagger \otimes M}$ conditionally on outcome $\hat{U}$.

Focusing on the $N = 1, M = 1$ case the optimal network is:


$\hat{\mathcal{U}}^{-1}$
$\mathcal{U}$
$\frac{1}{\sqrt{d}}|I\rangle\rangle$
$d|\hat{U}\rangle\rangle\langle\langle\hat{U}|$

. (7.7)

The maximum value of F for this case is obtained by substituting $\bigoplus_j (U_j^* \otimes V_j)$ with $U^* \otimes V$ and $U_{2N+3} \otimes V_{2N+2}^*$ with $V_{2N+3} \otimes U_{2N+2}^*$ in the main derivation. Reminding the decomposition (B.41) we have that $m_p = m_q = 1$ and thus Eq. (6.44) gives

$$F = \frac{1}{d^4} \sum_\nu (m_\nu)^2 = \frac{2}{d^2} \tag{7.8}$$

Remark 7.1 *The optimal "learning of the inverse" of a unitary transformation provides the optimal approximate realignment of reference frames in the quantum communication scenario considered in Ref. [73], proving the optimality of the "measure-and-rotate" strategy conjectured therein. In that scenario, the storing state $|\Psi\rangle\rangle$ serves as a token of Alice's reference frame, and is sent to Bob along with a quantum message $|\phi\rangle$. Due to the mismatch of reference frames, Bob receives the decohered state $\sigma_\phi = \int_G |\Psi_U\rangle\rangle\langle\langle\Psi_U| \otimes U|\varphi\rangle\langle\varphi|U^\dagger \, \mathrm{d}U$, from which he tries to retrieve the message $|\varphi\rangle$ with maximum fidelity $f = \int \mathrm{d}\varphi \; \langle\varphi|\mathcal{R}'(\sigma_\varphi)|\varphi\rangle \, \mathrm{d}\varphi$, where $\mathcal{R}'$ is the retrieving channel and $\mathrm{d}\varphi$ denotes the uniform probability measure over pure states. The maximization of f is equivalent to the maximization of the channel fidelity $F' = \int_G \langle\langle U^\dagger| \langle\langle \Psi_U^*|R'|U^\dagger\rangle\rangle|\Psi_U^*\rangle\rangle \, \mathrm{d}U$, which is the figure of merit for optimal inversion. It is worth stressing that the state $|\Psi\rangle\rangle$ that maximizes the fidelity is not the state $|\Psi_{\mathrm{lik}}\rangle\rangle = \bigoplus_j \sqrt{d_j/L}|I_j\rangle\rangle$, $L = \sum_j d_j^2$ that maximizes the likelihood [74]. For $M = 1$ and $G = SU(2), U(1)$ the state $|\Psi\rangle\rangle$ gives an average fidelity that approaches 1 as $1/N^2$, while for $|\Psi_{\mathrm{lik}}\rangle\rangle$ the scaling is $1/N$. On the other hand, Ref. [73] shows that for $M = 1$ $|\Psi_{\mathrm{lik}}\rangle\rangle$ allows a perfect correction of the misalignment errors with probability of success $p = 1 - 3/(N+1)$, which is not possible for $|\Psi\rangle\rangle$. The determination of the best input state to maximize the probability of success, and the study of the probability/fidelity trade-off remain open interesting problems for future research.*

7.2 Supermap scenario

In this section we will review the derivation of Ref. [28] of the optimal inversion of a unitary transformation according to the scheme (7.1); since the quantum network $\mathcal{G}$ can be interpreted as a *supermap* $\mathcal{G} : \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2) \to \mathcal{L}(\mathcal{H}_0, \mathcal{H}_3)$ that maps the unknown unitary transformation $\mathcal{U} \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$ into another transformation $\mathcal{G}(\mathcal{U}) := \mathcal{G} \star \mathcal{U} \in \mathcal{L}(\mathcal{H}_0, \mathcal{H}_3)$ we call the scheme (7.1) the *supemap scenario* for the inversion of a unitary transformation.[17]

In order to make a meaningful comparison, we choose as figure of merit the averaged channel fidelity as we previously did in the learning scenario:

$$\begin{aligned} F &= \int_{SU(d)} dU \mathcal{F}(\mathcal{G}, \mathcal{U}) \\ &= \frac{1}{d^2} \int_{SU(d)} dU \langle\!\langle U^\dagger|_{30} \langle\!\langle U^*|_{21} G |U^\dagger\rangle\!\rangle_{30} |U^*\rangle\!\rangle_{21} \end{aligned} \tag{7.9}$$

The following lemma holds:

Lemma 7.1 *The operator G maximizing the fidelity (7.9) can be assumed without loss of generality to satisfy the commutation relation*

$$[G, U_3 \otimes W_2 \otimes U_1 \otimes W_0] = 0 \quad \forall V, W \in SU(d) \ . \tag{7.10}$$

Proof. The proof consists in the standard averaging argument (see e.g. Lemma 6.1): Let G be optimal. Then take its average $\overline{G} = \int \mathrm{d}U \, \mathrm{d}W \, (U_3 \otimes W_2 \otimes U_1 \otimes W_0) G (U_3 \otimes W_2 \otimes U_1 \otimes W_0)^\dagger$: it is immediate to see that $\overline{G}$ satisfies Eqs. (7.10) and (7.3) and has the same fidelity as G. ■

Thanks to Theorem B.3 and reminding the decomposition (B.33) G can be decomposed as

$$C = \sum_{\mu,\nu \in \mathsf{S}} a^{\mu\nu} P^{\mu}_{31} \otimes P^{\nu}_{20}, \tag{7.11}$$

where $\mathsf{S} = \{+, -\}$, $P^{\pm}_{ij}$ is the projector onto the symmetric/antisymmetric subspace of $\mathcal{H}_i \otimes \mathcal{H}_j$, and $a^{\mu\nu} \geqslant 0 \; \forall \mu, \nu$. Moreover, using Eq. (7.11) the fidelity (7.9) becomes

$$\begin{aligned} F &= \frac{1}{d^2} \langle\!\langle I|_{30} \langle\!\langle I|_{21} G |I\rangle\!\rangle_{30} |I\rangle\!\rangle_{21} \\ &= \frac{1}{d^2} \sum_{\nu \in \mathsf{S}} a^{\nu\nu} d_\nu, \qquad d_\nu = \mathrm{Tr}[P^\nu], \end{aligned} \tag{7.12}$$

while the normalization (7.3) can be rewritten as $\sum_{\mu \in \mathsf{S}} a^{\mu\nu} d_\mu = 1, \forall \nu \in \mathsf{S}$. The last equality implies the bound $F = \frac{1}{d^2} \sum_{\mu \in \mathsf{S}} a^{\mu\mu} d_\mu \leqslant 2/d^2$, which is achieved if and only if $a^{\mu\nu} = \frac{\delta_{\mu\nu}}{d_\mu}$, that is, if and only if

$$\begin{aligned} G &= \frac{P^+_{31} \otimes P^+_{20}}{d_+} + \frac{P^-_{31} \otimes P^-_{20}}{d_-} \\ &= \int_{SU(d)} d\hat{U} \; |\hat{U}^\dagger\rangle\!\rangle \langle\!\langle \hat{U}^\dagger|_{30} \otimes |\hat{U}^*\rangle\!\rangle \langle\!\langle \hat{U}^*|_{21}. \end{aligned} \tag{7.13}$$

[17] Clearly, also the learning network can be thought as a supermap; however, whenever in this chapter we use the term supermap, we refer to the scheme (7.1).

We have then proved that the learning and the supermap scenarios achieves the same value of F. Contrary to what one could expect there is no coherent strategy that achieves better performances than the measure and re-prepare learning scheme in Eq. (7.7).

8 Information-disturbance tradeoff in estimating a unitary transformation

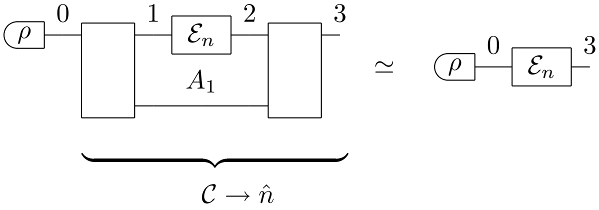
One of the key features of Quantum Mechanics is the impossibility of extracting information from a system without producing a disturbance on its state; this is the basis of the indeterminism of Quantum Mechanics and of quantum cryptography. However, a quantitative expression of the tradeoff between information and disturbance is generally a non trivial issue, and it has been the subject of numerous papers [76,77,78,79,80,81,82,83,84] since Heisenberg's γ-ray microscope thought experiment [75].

On the other hand, the case of extracting information from a black box without affecting the transformation it is expected to perform, has not been considered yet. More precisely, we consider the problem of both applying the black box to an arbitrary input state and estimating its transformation within the same use. Similarly to the case of state estimation, the information-disturbance tradeoff for channels is interesting for security analysis of two-way quantum cryptographic protocols [85, 86]. An information-disturbance problem in the estimation of the state of a quantum system can be split into two parts;

- making a measurement which supplies information about the state of the system;
- comparing the state of the system before the measurement with the state after the measurement.

Suppose we are provided with a system which is in an unknown state ρ_n randomly drawn from an ensemble $\{p_n \rho_n\}$ (p_n is the probability of getting the state ρ_n); we want to estimate the parameter n and compare the state after the measurement with the state before the measurement. The right tool for describing such a process which has both a classical (the result of the measurement) and a quantum (the final state) output is a quantum instrument $\{\mathcal{T}_{\hat{n}}\}$ (see Section 2.3). The quantum instrument $\{\mathcal{T}_{\hat{n}}\}$ with probability $p(\hat{n}|n) = \mathrm{Tr}[\mathcal{T}_{\hat{n}}(\rho_n)]$ outputs the classical outcome $\hat{n}$ (that is an estimate of n) and the quantum state $\rho'_n = \mathcal{T}_{\hat{n}}(\rho_n)/\,\mathrm{Tr}[\mathcal{T}_{\hat{n}}(\rho_n)]$: the closer $\hat{n}$ is to n the greater is the information and the closer ρ'_n is to ρ_n the less is the disturbance.

The previous framework can be generalized to the case of channels. Consider a quantum network $\mathcal{C}$ that can be linked with a single use of an unknown channel $\mathcal{E}_n$ randomly drawn from a set $\{\mathcal{E}_n\}$. We want the network $\mathcal{C}$ to provide us with an estimate $\hat{n}$ of n, but without affecting the output $\mathcal{E}(\rho)$ on the input state ρ



(8.1)

We notice that the resulting map $\mathcal{C} \star \mathcal{E}_n$ behaves like a quantum instrument; since $\mathcal{E}_n$ is a channel (i.e. a deterministic map) we have that $\mathcal{C}$ is actually a generalized instrument $\{\mathcal{C}_{\hat{n}}\}$ (see 2.3).

Obviously, if we are interested only in gathering information on the unknown channel, the optimal device is the one suggested by *channel estimation* [70]: we apply locally the channel to the best (according to some prior information) bi-partite state σ and then we perform a suitable measurement P_i. In this case we neglect the action of the channel on the input state of the circuit $\mathcal{E}(\rho)$. On the other hand, if we are not interested at all in gathering information about the channel $\mathcal{E}$, the best circuit board simply consists in applying $\mathcal{E}$ to ρ. Between these two extremal situations one can ask what is the maximum amount of information that is possible to gather without violating a disturbance threshold.

In this chapter we review Ref. [29] derived the best generalized instrument which achieves this task when the unknown channel is a unitary transformation, for any possible information-disturbance rate.

8.1 Optimization of the tradeoff

We now address the information-disturbance problem in the unitary case. Suppose we are provided with an unknown unitary gate $\mathcal{V} \in SU(d)$ picked randomly according to the Haar distribution; we now look for the best generalized instrument $\{\mathcal{R}_V \in \mathcal{L}(\mathcal{L}(\mathcal{H}_{02}) \bigotimes_{i=0}^{4} \mathcal{H}_i)\}$, $\int \mathrm{d}V \mathcal{R}_V = \mathcal{R}_\Omega$ $(V \in SU(d))$ which performs the best estimation of the group parameter V without affecting too much the performance of the unknown gate.

We now introduce two figures of merit in order to quantify the disturbance and the information gain. Minimization of the disturbance can be expressed by maximizing the channel fidelity $\mathcal{F}$ (defined in Eq. A.1) between the average resulting channel $\int \mathrm{d}V \mathcal{R}_V \star \mathcal{U} = \mathcal{R}_\Omega \star \mathcal{U}$ and the input unitary $\mathcal{U}$:

$$\mathcal{F}(\mathcal{R}_\Omega \star \mathcal{U}, \mathcal{U}) = \frac{1}{d^2} \langle\!\langle U|_{03} \langle\!\langle U^*|_{12} R_\Omega |U\rangle\!\rangle_{03} |U^*\rangle\!\rangle_{12}. \tag{8.2}$$

A reasonable choice for the figure of merit is the group average of the fidelity (8.2):

$$F(R_\Omega) := \int \mathrm{d}U \mathcal{F}(\mathcal{R}_\Omega \star \mathcal{U}, \mathcal{U}) = \frac{1}{d^2} \int \mathrm{d}U \langle\!\langle U|_{03} \langle\!\langle U^*|_{12} R_\Omega |U\rangle\!\rangle_{03} |U^*\rangle\!\rangle_{12}. \tag{8.3}$$

Now we need an expression to evaluate the amount of information gathered. The probability of outcome V when the input state of the network is $\rho \in \mathcal{B}(\mathcal{H}_0)$ has the following expression

$$p(V|U, \rho) = \mathrm{Tr}_3[\mathcal{R}_V \star \mathcal{U}(\rho_0)]. \tag{8.4}$$

In our derivation we assume $\rho_\mathcal{E} = d^{-1} I_0$ since this condition arises in two relevant cases:

- when the input system is prepared in a maximally entangled state with some ancillary system; this is the scenario in the protocols of Ref. [85]
- when the input system is prepared at random in one of the states of an ensemble (p_i, ρ_i), with the property $\rho_\mathcal{E} = \sum_i p_i \rho_i$. This is the case of the protocol of Ref. [86]

With this assumption Eq.(8.4) becomes

$$p(V|U) = \frac{1}{d} \mathrm{Tr}_3[(\mathcal{R}_V * \mathcal{U}) I_0] = \frac{1}{d} \mathrm{Tr}_{03}[\langle\!\langle U^*| R_V |U^*\rangle\!\rangle]. \tag{8.5}$$

Now we need a payoff function $c(U,V)$ which quantifies the error of estimating V when the unknown unitary is U: taking inspiration from the previous definition of disturbance, a good choice is again the channel fidelity, that is

$$c(U,V) := \mathcal{F}(\mathcal{U}\mathcal{V}) = \frac{1}{d^2}\langle\langle U|V\rangle\rangle = \frac{1}{d^2}|\operatorname{Tr}[UV^\dagger]|^2. \tag{8.6}$$

Then the information gain is given by

$$\begin{aligned} G(R_V) &:= \int dUdV p(V|U)c(U,V) = \\ &= \frac{1}{d^3}\int dUdV \operatorname{Tr}_{03}[\langle\langle U^*|_{12} R_V |U^*\rangle\rangle_{12}]|\langle\langle U|V\rangle\rangle|^2 \end{aligned} \tag{8.7}$$

The following lemma allows us to restrict to a specific class of generalized instruments.

Lemma 8.1 *For any generalized instrument* $\{\mathcal{R}_{V'}\}$, $\int \mathrm{d}V'\mathcal{R}_{V'} = \mathcal{R}_{\Omega'}$ *there exists another generalized instrument* $\{\mathcal{R}_V\}$, $\int \mathrm{d}V\mathcal{R}_V = \mathcal{R}_\Omega$ *such that*

$$\begin{aligned} R_V &= (V_0 \otimes V_1^* \otimes I_{23})R_I(V_0 \otimes V_1^* \otimes I_{23})^\dagger \\ &= (I_{01} \otimes V_2^T \otimes V_3^\dagger)R_I(I_{01} \otimes V_2^T \otimes V_3^\dagger)^\dagger. \end{aligned} \tag{8.8}$$

$$F(R_{\Omega'}) = F(\mathcal{R}_V) \tag{8.9}$$

$$G(R_{V'}) = G(\mathcal{R}_V) \tag{8.10}$$

Proof. This result is a straightforward application of the averaging argument for covariant POVMs [1]. Let R'_V be optimal; then let us consider

$$R_V := \int \mathrm{d}W(W_0 \otimes W_1^* \otimes I_{23})R_{W^\dagger V}(W_0^\dagger \otimes W_1^T \otimes I_{23}) \tag{8.11}$$

exploiting the properties of the Haar measure $\mathrm{d}W$ it is easy to verify that R_V enjoys the properties (8.8), (8.9) and (8.10). ■

Since $R_\Omega = \int dV R_V$, and reminding Eq. (8.8) we have

$$R_\Omega = \int dV(V_0 \otimes V_1^* \otimes I_{23})R_I(V_0 \otimes V_1^* \otimes I_{23})^\dagger. \tag{8.12}$$

Applying Theorem B.3 we get $[R_\Omega, W_0 \otimes W_1^* \otimes V_2 \otimes V_3^*] = 0$ and the normalization conditions $\operatorname{Tr}_3[R_\Omega] = R_{01}^{(1)} \otimes I_2$, $\operatorname{Tr}_1[R^{(1)}] = I_0$ become trivially

$$\operatorname{Tr}[R_I] = d^2. \tag{8.13}$$

Theorem B.3 and decomposition (B.42) allow us to rewrite the two figures of merit in the following way:

$$F = \operatorname{Tr}[R_F R_I], \qquad G = \operatorname{Tr}[R_G R_I] \tag{8.14}$$

$$R_F = \frac{1}{d^2(d^2-1)}(I_{0123} + d^2 P_{01}^p \otimes P_{23}^p - P_{01}^p \otimes I_{23} - I_{01} \otimes P_{23}^p)$$

$$R_G = \frac{1}{d^2(d^2-1)}\left(\left(1 - \frac{2}{d^2}\right) I_{03} \otimes I_{12} + I_{03} \otimes P_{12}^p\right)$$

where $P^p_{ij} = d^{-1}|I\rangle\rangle\langle\langle I|_{ij}$ is the projector on the one-dimensional invariant subspace of $V_i \otimes V_j^*$. Clearly we cannot independently optimize the two figures of merit. What we can do is to fix a value of G and then maximize F. We now prove that this approach is equivalent to fixing a disturbance-gain rate $0 \leqslant p \leqslant 1$ and maximize the convex combination:

$$pG + (1-p)F = \mathrm{Tr}[(pR_G + (1-p)R_F)R_I] \tag{8.15}$$

Let fix the value $G = \overline{G}$; now let us suppose that $R_I(p')$ maximize the combination $p'G + (1-p')F$ with p' such that $p' \,\mathrm{Tr}[R_I(p')G] = \overline{G}$. Clearly $R_I(p')$ achieves the maximum value of F since any other greater value of F would increase $pG + (1-p)F$. This explain why the optimal information disturbance tradeoff can be obtained by maximizing Eq. (8.15).

Since the only restrictions on R_I are positivity and the normalization given by Eq. (8.13), the optimal choice for the operator R_I is to take it proportional to the projector on the eigenspace of $pR_G + (1-p)R_F$ corresponding to the maximum eigenvalue; this projector can be shown [80] to be

$$R_I = |\chi\rangle\,\langle\chi| \tag{8.16}$$

$$|\chi\rangle = x|I\rangle\rangle_{03}|I\rangle\rangle_{12} + y|I\rangle\rangle_{01}|I\rangle\rangle_{23} \quad x, y \in \mathbb{R}^+$$

Reminding Eq. (8.8) we get $R_V = |\chi_V\rangle\,\langle\chi_V|$ with $|\chi_V\rangle = x|V\rangle\rangle_{03}|V^*\rangle\rangle_{12} + y|I\rangle\rangle_{01}|I\rangle\rangle_{23}$. Normalization condition (8.13) implies that x and y obey

$$d^2x^2 + d^2y^2 + 2xyd = d^2 \tag{8.17}$$

We notice that we correctly have just one free parameter which will depend on the tradeoff ratio p. Fidelity and gain can be calculated in terms of the parameters x and y, getting the following expressions

$$F = 1 - \frac{d^2-2}{d^2}x^2 \qquad G = \frac{2-y^2}{d^2} \tag{8.18}$$

We note that when $x = 0, y = 1$, we have $R_V = |I\rangle\rangle\langle\langle I|_{01} \otimes |I\rangle\rangle\langle\langle I|_{23}$ for all V, that is the generalized instrument is the identity map. In this case the performance of the unknown unitary is not affected at all and the channel fidelity reaches its maximum $F = 1$. On the other hand the information gain takes its minimum value $G = \frac{1}{d^2}$ which corresponds to random guessing U. The opposite case $x = 1, y = 0$ clearly gives the minimum value $F = \frac{2}{d^2}$ and the maximum $G = \frac{2}{d^2}$, which is the same given by the optimal estimation.

Using Eq. (8.17), we can easily express G as a function of x; then, upon eliminating x, we can express F as a function of G:

$$\sqrt{(d^2-2)(2-d^2G)} = \sqrt{(d^2-1)F-1} - \sqrt{1-F}. \tag{8.19}$$

It seems useful to introduce the variables $0 \leqslant I, D \leqslant 1$:

$$I = \frac{G - G_{min}}{G_{max} - G_{min}} \qquad D = \frac{F_{max} - F}{F_{max} - F_{min}} \tag{8.20}$$

where $G_{max} = 2d^{-2}$, $G_{min} = d^{-2}$, $F_{min} = 2d^{-2}$ and $F_{max} = 1$. Expression (8.19) can be rewritten in terms of D and I:

$$d^2(D-I)^2 - 4D(1-I) = 0; \tag{8.21}$$

the plot of Eq. (8.21) is reported in Figure 8.1.

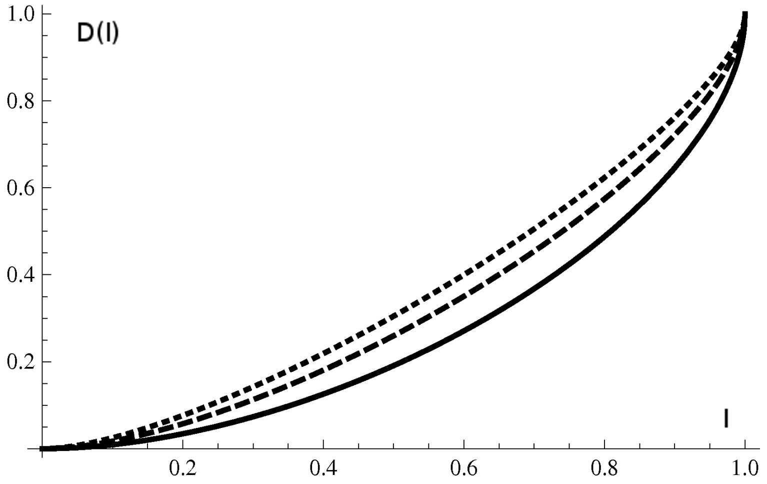


Figure 8.1. Plot of the lower bound $D(I)$ of the disturbance, corresponding to Eq. (8.21), for various value of d: solid line, $d = 2$; dashed line, $d = 3$;dotted line , $d = 4$.

8.2 Realization scheme for the optimal network

We now inspect the structure of the optimal network. Theorem 2.9 tells us that the generalized instrument can be realized by

- a deterministic network $\mathcal{S} : \mathcal{B}(\mathcal{H}_{02}) \to \mathcal{B}(\mathcal{H}_{13} \otimes \mathcal{H}_{A_2})$;
- a POVM $\{P_V = R_\Omega^{-\frac{1}{2}} R_V R_\Omega^{-\frac{1}{2}}\}$ on the ancilla space $\mathcal{H}_{0'1'2'3'}$.

The deterministic network $\mathcal{S}$ can be realized, according to Theorem 2.6, as a product of two isometries $W^{(1)} : \mathcal{H}_0 \to \mathcal{H}_{0\,A_1}$ and $W^{(2)} : \mathcal{H}_{2\,A_1} \to \mathcal{H}_{3\,A_2}$, $\mathcal{S} = Z \cdot Z^\dagger$, $Z = W^{(2)}W^{(1)}$.

Inserting Eq. (8.16) into Eq. (8.12) we have

$$R_\Omega = A\,P^p_{01} \otimes P^p_{23} + B\,P^q_{01} \otimes P^q_{23}, \qquad \frac{1}{d}\,\mathrm{Tr}_{23}[R_\Omega] = R^{(1)} = aP^p_{01} + bP^q_{01} \tag{8.22}$$

$$A = x^2 + d^2y^2 + 2dxy = d^2 - (d^2-1)x^2 \qquad B = \frac{x^2}{(d^2-1)} \tag{8.23}$$

$$a = \frac{A}{d}, b = \frac{d^2-1}{d}B \qquad a + (d^2-1)b = d \tag{8.24}$$

The explicit expression of $W^{(1)}$ is given by specializing Eq. (2.62)

$$W^{(1)} = (I_1 \otimes R^{(1)\frac{1}{2}*}_{1'0'})(|I\rangle\!\rangle_{11'} \otimes T_{0\to 0'}) = \tag{8.25}$$

$$= \frac{1}{\sqrt{d}}\,(y|I\rangle\!\rangle_{1'0'} \otimes T_{0\to 1} + x|I\rangle\!\rangle_{11'} \otimes T_{0\to 0'}) \tag{8.26}$$

If we input a pure state $|\psi\rangle$ in the first isometry, we will have as the output the superposition $\frac{1}{\sqrt{d}}\,(y|I\rangle\rangle_{1'0'}\otimes(\,|\psi\rangle\,)_1 + x|I\rangle\rangle_{11'}\otimes|\psi\rangle_{0'})$.

The explicit expression for the second isometry is given by:

$$W^{(2)} = I_3 \otimes R_\Omega{}^{\frac{1}{2}}_{0'1'2'3'} R^{(1)-\frac{1}{2}}_{1'0'} |I\rangle\rangle_{33'} T_{2\to 2'} \tag{8.27}$$

Thanks to Eq. (2.68) the POVM on the ancilla space ($0'1'2'3'$) can be written as

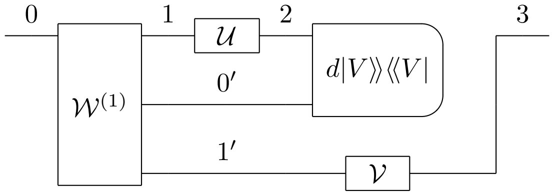
$$P_V = |\eta_V\rangle\langle\eta_V| \qquad |\eta_V\rangle = R_\Omega^{-\frac{1}{2}}|\chi_V\rangle \tag{8.28}$$

Isometry $W^{(2)}$ together with the POVM $\{|\eta_V\rangle\langle\eta_V|\}$ can be rewritten as a quantum instrument $\{\mathcal{T}_V : \mathcal{B}(\mathcal{H}_{21'}) \to \mathcal{B}(\mathcal{H}_3)$ where the maps $\{\mathcal{T}_V\}$ are defined as $\mathcal{T}_V(\rho) = \langle\eta_V|\,W^{(2)}\rho W^{(2)\dagger}\,|\eta_V\rangle\}$. Explicit calculation gives:

$$\langle\eta_V|\,W^{(2)} = \sqrt{d}V_{0'\to 3}\langle\langle V|_{21'} \tag{8.29}$$

we notice that the final instrument $\{\mathcal{T}_V\}$ does not depend on the parameters x and y.

Summarizing, the quantum network realizing the optimal information disturbance tradeoff in estimating a unitary transformation is as follows:



(8.30)

- The first isometry $W^{(1)}$ prepares a coherent superposition $\frac{1}{\sqrt{d}}\,(y|I\rangle\rangle_{1'0'}\otimes(\,|\psi\rangle\,)_1 + x|I\rangle\rangle_{11'}\otimes|\psi\rangle_{0'})$ which is tuned by the parameters x and y (that is by p in Eq. (8.15));
- the unitary U acts locally on system 1;
- at the end the instrument $\{\mathcal{T}_V\}$ is applied: $\{\mathcal{T}_V\}$ can realize either an estimate-and-reprapare strategy, or a teleportation protocol.

We now give a look to the complete action of the optimal circuit when the input is a pure state $|\psi\rangle$:

$$\begin{aligned} |\psi\rangle &\to \frac{1}{\sqrt{d}}\,(y|I\rangle\rangle_{1'0'}\otimes(\,|\psi\rangle\,)_1 + x|I\rangle\rangle_{11'}\otimes|\psi\rangle_{0'}) \to \\ &\to \frac{1}{\sqrt{d}}\,(y|I\rangle\rangle_{1'0'}\otimes(\,U\,|\psi\rangle\,)_2 + x|U\rangle\rangle_{21'}\otimes|\psi\rangle_{0'}) \to \\ &\to yU(|\psi\rangle)_3 + x\,\mathrm{Tr}[V^\dagger U]V(|\psi\rangle)_3. \end{aligned} \tag{8.31}$$

We remark that the optimal device essentially combines two strategies:

1. applying the unknown unitary U to the state $|I\rangle\rangle$, measuring the state $|U\rangle\rangle$, and then performing the estimated transformation V on the input state ψ. This is a measure and reprepare strategy which is optimal if $y = 0, x = 1$ (that is we are interested only in the information gain)

2. Applying U on the input state and then outputting $U|\psi\rangle$ (in our scheme this last step involves a teleportation protocol). This is clearly an optimal strategy if $x = 0, y = 1$, that is if we are interested only in leaving the action of U unaffected.

Surprisingly, the analytical expression of the tradeoff curve given in Eq. (8.21) is the same as the one for the estimation of a maximally entangled state [80]. It is worth noting that this is not a trivial consequence of the isomorphism $2^{-\frac{1}{2}}|U\rangle\rangle \leftrightarrow U$; indeed, this mathematical correspondence cannot be implemented by a physical invertible map. Once a unitary U is applied to the maximally entangled state $2^{-\frac{1}{2}}|I\rangle\rangle$ it is possible to retrieve the transformation U only probabilistically (this is the problem of the quantum learning discussed in chapter 6). Because of this reason there is no operational relation between the information disturbance tradeoff for unitary transformation and for maximally entangled states (the former is not a primitive of the latter and viceversa).

Besides its fundamental relevance, the information disturbance tradeoff for transformations is interesting as a possible eavesdropping for cryptographic protocol in which the secret key is encoded into a transformation. However this is not the case of the protocols [85, 86] where orthogonal unitaries are used and the security of the protocol is not based on the information disturbance tradeoff studied here. On the other hand the tradeoff we considered is an effective attack to the alternative $BB84$ protocol introduced in chapter 5. However, this alternative version of the $BB84$ protocol just involves two nonorthogonal unitaries; in principle, the tradeoff curve for a restricted of unitaries could be more favorable to the eavesdropper.

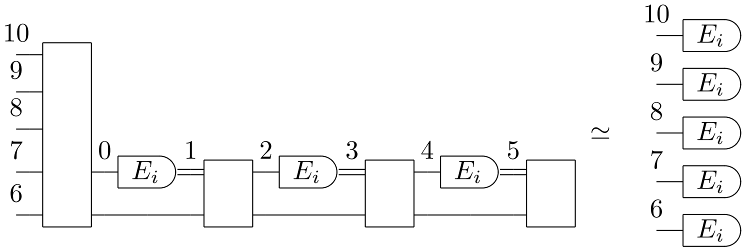
9 Learning and cloning of a measurement device

As we stressed in the introduction the recent trend of quantum information is to consider transformations as information carriers. Unlike what we did in all the previous chapters, in the present one we will not deal with unitary transformations but with measurements. We will consider quantum networks that, upon the insertion of N uses of an undisclosed measurement device, reproduce M approximate replicas of it.

When a measurement is an intermediate step of a quantum procedure its outcome can influence the following operations. This feed forward of the classical outcome can be conveniently described using a quantum system into which the outcome is encoded into perfectly distinguishable orthogonal states. In this sense a quantum measurement with only classical outcomes can be seen as a channel, which first measures the input system and based on the outcome prepares a state from a fixed orthogonal set.

In order to achieve this task different scenarios can be considered:

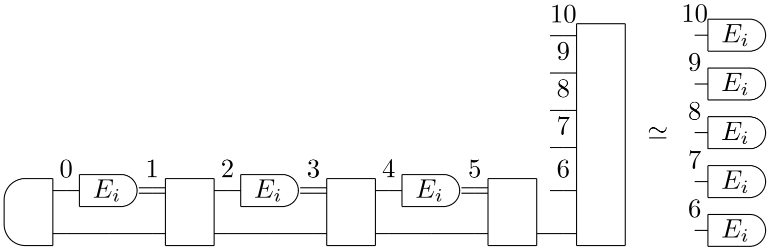
$N \to M$ **cloning**[18]: The measurement device and the states we want to measure are available at the same time;



(9.1)

(the double wire carries the classical outcome of the measurement).

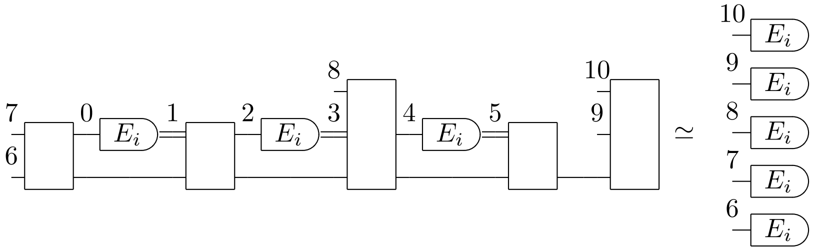
$N \to M$ **learning**: we can use the measurement device N times today and we want to replicate the same observables on M systems that will be provided tomorrow



(9.2)

[18] The term cloning of observables has been used in Ref. [87] referring to state cloning machines preserving the statistics of a class of observables.

$N \to M$ **hybrid**: we have to produce the replicas at different times



(9.3)

In the following we will consider some specific scenarios and compare their performances.

9.1 Formulation of the problem

In the following we will restrict ourselves to von Neumann measurement, i.e. sharp non degenerate POVMs:

$$E_i = |i\rangle \langle i| \tag{9.4}$$

where $\{|i\rangle\}_{i=1}^d$ is an o.n.b. of the Hilbert space $\mathcal{H}$. We notice that all the POVMs of this kind can be generated by rotating a reference POVM $\{|i\rangle \langle i|\}_{i=1}^d$ by arbitrary elements of the $\mathbf{SU}(d)$ group as follows

$$E_i^{(U)} = U\,|i\rangle \langle i|\, U^\dagger \qquad U \in \mathbf{SU}(d). \tag{9.5}$$

The classical outcome i of the POVM will be encoded into a quantum system by preparing the state $|i\rangle$ from a fixed orthonormal basis, which is the same for each POVM $\{E_i^{(U)}\}$. Within this framework the measurement device is modeled as the following measure-and-prepare quantum channel $\mathcal{E}^{(U)} : \mathcal{L}(\mathcal{H}) \to \mathcal{L}(\mathcal{H})$

$$\mathcal{E}^{(U)}(\rho) = \sum_i \mathrm{Tr}[E_i^{(U)} \rho]\, |i\rangle \langle i| \tag{9.6}$$

that measure the POVM $\{E_i^{(U)}\}$ on its input state and outputs the state $|i\rangle \langle i|$ if the outcome is i. The channel $\mathcal{E}^{(U)}$ is represented by its Choi operator

$$E^{(U)} = \sum_i {E_i^{(U)}}^T \otimes |i\rangle \langle i| = \sum_i U^*\, |i\rangle \langle i|\, U^T \otimes |i\rangle \langle i| \tag{9.7}$$

The N uses of the measurement device are then represented by the tensor product $E_{01}^{(U)} \otimes \cdots \otimes E_{2N-2\;2N-1}^{(U)}$ where the input and the output space of the k-th use of the measurement device are denoted by $2k-2$ and $2k-1$ respectively. We introduce the following notation:

$$\mathcal{H}_{\mathsf{or}} := \bigotimes_{k=1}^{N} \mathcal{H}_{2k-2}, \qquad \mathcal{H}_{\mathsf{cl}} := \bigotimes_{k=1}^{N} \mathcal{H}_{2k-1}. \tag{9.8}$$

Since we want the replicating network $\mathcal{R}$ to behave as M copies of the POVM $\{E_i^{(U)}\}$ upon insertion of the N uses $\mathcal{E}^{(U)}$, we have that $\mathcal{R}$ is actually a generalized instrument $\{\mathcal{R}_{\boldsymbol{i}}\}$ where $\boldsymbol{i}$ is the M-tuple of outcomes $(i_1, \dots, i_M)$. The overall resulting POVM is then

$$G_{\boldsymbol{i}}^{(U)} = (R_{\boldsymbol{i}} * E_{01}^{(U)} * \cdots E_{2N-2\,2N-1}^{(U)})^T \tag{9.9}$$

$$R_{\boldsymbol{i}} = \mathcal{L}(\mathcal{H}_{\mathsf{or}} \otimes \mathcal{H}_{\mathsf{cl}} \otimes \mathcal{H}_{\mathsf{re}}) \quad \mathcal{H}_{\mathsf{re}} = \bigotimes_{k=1}^{M} \mathcal{H}_{2N+k-1}$$

where $\mathcal{H}_{2N+k-1}$ denotes the input space of the k-th replica.

Our task is to find the network $\mathcal{R}_{\boldsymbol{i}}$ such that $G_{\boldsymbol{i}}^{(U)}$ is as close as possible to to M uses of $\{E_i^{(U)}\}$, i.e

$$\{G_{\boldsymbol{i}}^{(U)}\} \simeq \{E_{i_1}^{(U)} \otimes E_{i_2}^{(U)} \otimes \cdots \otimes E_{i_M}^{(U)}\} := \{E_{\boldsymbol{i}}^{(U)}\}. \tag{9.10}$$

In order to quantify the performances of the replicating network, we need to introduce a criterion which quantify the closeness between two POVMs. the following lemma provides such a tool:

Lemma 9.1 (distance criterion for POVM) *Let $\Sigma := \{1, \dots, d\}$ be a finite set of events and $\{P_i \in \mathcal{L}(\mathcal{H})\}$ and $\{Q_j \in \mathcal{L}(\mathcal{H})\}$ be two POVMs. Consider now the quantity*

$$\mathcal{F} := \frac{1}{d} \sum_i \mathrm{Tr}[P_i Q_i] \tag{9.11}$$

and suppose that either $\{P_i\}$ or $\{Q_j\}$ is a von Neumann measurement. Then $\mathcal{F} = 1 \Leftrightarrow P_i = Q_i \forall i$

Proof. If $\{P_i\}$ is a von Neumann measurement we have $P_i = |i\rangle\langle i|$ where $|i\rangle$ is an orthonormal basis of $\mathcal{H}$. Then we have $Q_i = P_i \Rightarrow Q_i = |i\rangle\langle i|$ and

$$\mathcal{F} = \frac{1}{d} \sum_i \mathrm{Tr}[P_i Q_i] = \frac{1}{d} \sum_i \mathrm{Tr}[|i\rangle\langle i|] = 1 \tag{9.12}$$

On the other hand if $\mathcal{F} = 1$ we have

$$d = \sum_i \mathrm{Tr}[P_i Q_i] = \sum_i \langle i| Q_i |i\rangle = \sum_{ij} \langle i| Q_j |i\rangle - \sum_{i \neq j} \langle i| Q_j |i\rangle =$$

$$= \mathrm{Tr}\left[\sum_j Q_j\right] - \sum_{i\neq j} \langle i| Q_j |i\rangle = d - \sum_{i\neq j} \langle i| Q_j |i\rangle \Rightarrow \sum_{i\neq j} \langle i| Q_j |i\rangle = 0$$

Since $Q_j \geqslant 0$ $\sum_{i\neq j} \langle i| Q_j |i\rangle = 0 \Rightarrow \langle i| Q_j |i\rangle \forall i \neq j$ which implies $Q_j = \alpha_j |j\rangle\langle j|$ with $\alpha_j \geqslant 0$. Finally the condition $\sum_j \alpha_j |j\rangle\langle j| = I$ implies $\alpha_j = 1$ and thus $Q_j = P_j$. ■

Assuming that the unknown POVM $\{E_i^{(U)}\}$ is randomly drawn according to the Haar distribution, we choose the quantity:

$$F := \int \mathrm{d}U \mathcal{F}(\{G_{\boldsymbol{i}}^{(U)}\}\{E_{\boldsymbol{i}}^{(U)}\}) \tag{9.13}$$

as a figure of merit.

After fixing one of the possible scenarios ($N \rightarrow M$ cloning, learning or hybrid) our task is to find the optimal generalized instrument $\mathcal{R}_{\boldsymbol{i}}$ maximizing the quantity $F{:=}\!\int\! \mathrm{d}U \mathcal{F}(\{G^{(U)}_{\boldsymbol{i}}\}\{E^{(U)}_{\boldsymbol{i}}\})$.

9.2 Symmetries of the replicating network

Here we exploit the symmetries of the figure of merit (9.13) to simplify the optimization problem. The first simplification relies on the fact that some wires of the network carry only classical information, representing the outcome of the measurement.

Lemma 9.2 (Restriction to diagonal networks) *The optimal generalized instrument $\{\mathcal{R}_{\boldsymbol{i}}\}$, with $\sum_{\boldsymbol{i}} \mathcal{R}_{\boldsymbol{i}} = \mathcal{R}_\Omega$ maximizing Eq. (9.13), can be chosen to satisfy:*

$$R_{\boldsymbol{i}} = \sum_{\boldsymbol{j}} R'_{\boldsymbol{i},\boldsymbol{j}} \otimes |\boldsymbol{j}\rangle \langle \boldsymbol{j}| , \tag{9.14}$$

where $\boldsymbol{j} = (j_1, \dots, j_N)$, $|\boldsymbol{j}\rangle := |j_1\rangle_1 \otimes \cdots \otimes |j_N\rangle_{2N-1} \in \mathcal{H}_{\mathsf{cl}}$, $0 \leqslant R'_{\boldsymbol{i},\boldsymbol{j}} \in \mathcal{L}(\mathcal{H}_{\mathsf{or}} \otimes \mathcal{H}_{\mathsf{re}})$ and $\sum_{\boldsymbol{j}}$ is a shorthand for $\sum_{j_1,\dots,j_N=1}^{d}$.

Proof. Let $\{R_{\boldsymbol{i}}\}$ be a generalized instrument. Let us define $\{\tilde{R}_{\boldsymbol{i}}\}$ as

$$\tilde{R}_{\boldsymbol{i}} := \sum_{\boldsymbol{j}} \langle \boldsymbol{j}| R_{\boldsymbol{i}} |\boldsymbol{j}\rangle \otimes |\boldsymbol{j}\rangle \langle \boldsymbol{j}| \qquad |j_1\rangle_1 \otimes \cdots \otimes |j_N\rangle_{2N-1} . \tag{9.15}$$

We now prove that $\{\tilde{R}_{\boldsymbol{i}}\}$ is a generalized instrument: reminding Eq. (9.7), we have

$$\begin{aligned}
\sum_{\boldsymbol{i}} \tilde{R}_{\boldsymbol{i}} &= \sum_{\boldsymbol{i}} \sum_{\boldsymbol{j}} \langle \boldsymbol{j}| R_{\boldsymbol{i}} |\boldsymbol{j}\rangle \otimes |\boldsymbol{j}\rangle \langle \boldsymbol{j}| = \sum_{\boldsymbol{j}} \langle \boldsymbol{j}| R_\Omega |\boldsymbol{j}\rangle \otimes |\boldsymbol{j}\rangle \langle \boldsymbol{j}| = \\
&= R_\Omega * \left(\sum_{j_1} |j_1\rangle \langle j_1| \otimes |j_1\rangle \langle j_1| \right) * \cdots * \left(\sum_{j_1} |j_N\rangle \langle j_N| \otimes |j_N\rangle \langle j_N| \right) = \\
&= R_\Omega * E^{(I)} * \cdots * E^{(I)}
\end{aligned} \tag{9.16}$$

where the link is performed on the space $\mathcal{H}_{\mathsf{cl}}$. The operator in Eq. (9.16) is the Choi-Jamiołkowsky of a deterministic quantum network with the same normalization of R_Ω. Finally we show that $\{R_{\boldsymbol{i}}\}$ and $\{\tilde{R}_{\boldsymbol{i}}\}$ when linked with the N uses of $E^{(U)}$ produce the same replicas $\{G^{(U)}_{\boldsymbol{i}}\}$:

$$\begin{aligned}
G^{(U)}_{\boldsymbol{i}} &= (R_{\boldsymbol{i}} * E^{(U)}_{01} * \cdots E^{(U)}_{2N-2\;2N-1})^T = \\
&= (\sum_{\boldsymbol{j}} (\langle \boldsymbol{j}|_{\mathsf{or}} U^{\dagger \otimes N} \langle \boldsymbol{j}|_{\mathsf{cl}}) R_{\boldsymbol{i}} (U^{\otimes N} |\boldsymbol{j}\rangle_{\mathsf{or}} |\boldsymbol{j}\rangle_{\mathsf{cl}})^T = \\
&= (\sum_{\boldsymbol{j}} (\langle \boldsymbol{j}|_{\mathsf{or}} U^{\dagger \otimes N} \langle \boldsymbol{j}|_{\mathsf{cl}}) \tilde{R}_{\boldsymbol{i}} (U^{\otimes N} |\boldsymbol{j}\rangle_{\mathsf{or}} |\boldsymbol{j}\rangle_{\mathsf{cl}})^T = \\
&= (\tilde{R}_{\boldsymbol{i}} * E^{(U)}_{01} * \cdots E^{(U)}_{2N-2\;2N-1})^T .
\end{aligned} \tag{9.17}$$

■

We now exploit the form of Eq. (9.14) to simplify the expression of the fidelity in Eq. (9.13) as follows:

$$F := \int \mathrm{d}U \mathfrak{F}(\{G_{\boldsymbol{i}}^{(U)}\}\{E_{\boldsymbol{i}}^{(U)}\}) = \\ = \frac{1}{d^M} \int \mathrm{d}U \sum_{\boldsymbol{i},\boldsymbol{j}} \langle \boldsymbol{i}|_{\mathsf{re}} U^{T^{\otimes N}} \langle \boldsymbol{j}|_{\mathsf{or}} U^{\dagger^{\otimes N}} R'_{\boldsymbol{i},\boldsymbol{j}} U^{*^{\otimes N}} |\boldsymbol{i}\rangle_{\mathsf{re}} U^{\otimes N} |\boldsymbol{j}\rangle_{\mathsf{or}} . \tag{9.18}$$

The following lemma exploits the symmetry properties of Eq. (9.18) and simplifies the structure of the $R'_{\boldsymbol{i},\boldsymbol{j}}$:

Lemma 9.3 (Restriction to covariant networks) *The operators $R'_{\boldsymbol{i},\boldsymbol{j}}$ that maximize Eq. (9.18) can be chosen to satisfy the commutation relation*

$$[R'_{\boldsymbol{i},\boldsymbol{j}}, U_{\mathsf{or}}^{\otimes N} \otimes U^{*}{}_{\mathsf{re}}^{\otimes M}] = 0 \tag{9.19}$$

Proof. The proof consists in the same averaging argument we used in proving lemmas 5.1 , 6.1 and 8.1 ■

The commutation relation (9.19) allows us to rewrite the figure of merit has:

$$F = \frac{1}{d^M} \int \mathrm{d}U \sum_{\boldsymbol{i},\boldsymbol{j}} \langle \boldsymbol{i}|_{\mathsf{re}} \langle \boldsymbol{j}|_{\mathsf{or}} R'_{\boldsymbol{i},\boldsymbol{j}} |\boldsymbol{i}\rangle_{\mathsf{re}} |\boldsymbol{j}\rangle_{\mathsf{or}} \tag{9.20}$$

Another symmetry of our figure of merit is related to the possibility of relabeling the outcomes of a POVM. We shall denote by σ the element of $\mathbf{S}_d$, the group of permutations of d elements as well as the linear operator that permutes the elements of basis $\{|i\rangle\}$ according to this permutation ($\sigma |i\rangle \equiv |\sigma(i)\rangle$).

Lemma 9.4 (Relabeling symmetry) *Without loss of generality we can assume that the operators $R'_{\boldsymbol{i},\boldsymbol{j}}$ that maximize Eq. (9.18) satisfy the relation*

$$R'_{\boldsymbol{i},\boldsymbol{j}} = R'_{\sigma(\boldsymbol{i}),\sigma(\boldsymbol{j})} \tag{9.21}$$

where we shortened $\sigma(\boldsymbol{i}) \equiv (\sigma(i_1), \dots, \sigma(i_M))$, $\sigma(\boldsymbol{j}) \equiv (\sigma(j_1), \dots, \sigma(j_N))$.

Proof. Without loss of generality we can suppose that the $R'_{\sigma(\boldsymbol{i}),\sigma(\boldsymbol{j})}$'s satisfy Eq. (9.19). Let us then define

$$\widetilde{R'}_{\boldsymbol{i},\boldsymbol{j}} = \frac{1}{d!} \sum_{\sigma \in \mathbf{S}_d} R'_{\sigma(\boldsymbol{i}),\sigma(\boldsymbol{j})} \tag{9.22}$$

This corresponds to a valid instrument $\{\widetilde{R'}_{\boldsymbol{i}}\}$, because it is a convex combination of instruments obtained from $R_{\sigma(\boldsymbol{i}),\sigma(\boldsymbol{j})}$ by relabeling the outcomes of the inserted and replicated measurements by permutation σ. Let us now evaluate the figure of merit for this new instrument:

$$F(\widetilde{R'}_{\boldsymbol{i},\boldsymbol{j}}) = \frac{1}{d^M}\sum_{\boldsymbol{i},\boldsymbol{j}} \langle \boldsymbol{i}| \langle \boldsymbol{j}| \widetilde{R'}_{\boldsymbol{i},\boldsymbol{j}} |\boldsymbol{i}\rangle |\boldsymbol{j}\rangle = \frac{1}{d^M d!}\sum_{\boldsymbol{i},\boldsymbol{j}} \langle \boldsymbol{i}| \langle \boldsymbol{j}| \left(\sum_{\sigma\in\mathbf{S}_d} R^{\sigma(\boldsymbol{i}),\sigma(\boldsymbol{j})}\right) |\boldsymbol{i}\rangle |\boldsymbol{j}\rangle =$$

$$= \frac{1}{d^M d!}\sum_{\boldsymbol{i},\boldsymbol{j}} \langle \boldsymbol{i}| \langle \boldsymbol{j}| \left(\sum_{\sigma\in\mathbf{S}_d} \sigma^{\otimes N}\otimes\sigma^{\otimes M} R'_{\sigma(\boldsymbol{i}),\sigma(\boldsymbol{j})} \sigma^{\otimes N}\otimes\sigma^{\otimes M}\right) |\boldsymbol{i}\rangle |\boldsymbol{j}\rangle = \tag{9.23}$$

$$= \sum_{\sigma\in\mathbf{S}_d}\sum_{\boldsymbol{i},\boldsymbol{j}} \langle \sigma(\boldsymbol{i})| \langle \sigma(\boldsymbol{j})| R'_{\sigma(\boldsymbol{i}),\sigma(\boldsymbol{j})} |\sigma(\boldsymbol{i})\rangle |\sigma(\boldsymbol{j})\rangle = F(R'_{\boldsymbol{i},\boldsymbol{j}}) \tag{9.24}$$

where the identity (9.23) follows from the commutation relation (9.19) with $U = U^* = \sigma$. It is easy to prove that $\widetilde{R'}_{\boldsymbol{i},\boldsymbol{j}}$ satisfies Eq. (9.21).

■

Remark 9.1 *It is worth notice that the properties (9.14), (9.19) and (9.21) induce the following structure of the replicated POVMs:*

$$G^{(U)}_{\sigma(\boldsymbol{i})} = (U\sigma)^{\otimes M} G^{(I)}_{\boldsymbol{i}} (\sigma U^\dagger)^{\otimes M} \tag{9.25}$$

The advantage of using the above symmetry is in the reduction the number of independent parts of the generalized instrument. Let us define the equivalence relation between strings $\boldsymbol{i}$ and $\boldsymbol{i}'$ as

$$\boldsymbol{i} \sim \boldsymbol{i}' \quad \Leftrightarrow \quad \boldsymbol{i} = \sigma(\boldsymbol{i}'), \tag{9.26}$$

for some permutation σ. Thanks to Eq. (9.21) there are only as many independent $R_{\boldsymbol{i},\boldsymbol{j}}$ as there are equivalence classes among sequences $\boldsymbol{i}, \boldsymbol{j}$. For the simplest case $M = N = 1$ and arbitrary dimension $d \geqslant 2$, there are only two classes, which we denote by xx and xy. The reason is that for any couple i', j' there is a permutation σ such that $\sigma(1) = i'$ and $\sigma(2) = j'$, thus the classes are defined by the conditions $i = j$ or $i \neq j$, respectively. For all the cases where $M + N = 3$ (e.g. $N = 1, M = 2$ or $N = 2, M = 1$), the vectors $\boldsymbol{i}$ and $\boldsymbol{j}$ have three components. Then, there are four or five equivalence classes depending on the dimension d being two or greater than two, respectively. We denote these equivalence classes by xxx, xxy, xyx, xyy, xyz and the set of these elements by C^3_d. In the general case, it is clear that the cardinality of classes is given by the number of disjoint partitions of a set with cardinality $M + N$, with number p of parts $p \leqslant d$. For $M + N \geqslant d$, this number is known as Bell number B_{M+N}, and is recursively defined as follows

$$B_{k+1} := \sum_{j=0}^{k} \binom{k}{j} B_j. \tag{9.27}$$

In the case $M + N < d$ the solution is provided by the sum for $k = 1, \ldots, d$ of numbers of disjoint partitions of a set with $N + M$ elements into k subsets, which is the sum of Stirling numbers of the second kind $S(M + N, k)$. The Stirling numbers are given by the following formula

$$S(n, k) := \frac{1}{k!}\sum_{j=0}^{k} (-1)^j \binom{k}{j} (k - j)^n, \tag{9.28}$$

thus providing the following expression for the cardinality of classes C_d^{M+N}

$$C_d^{M+N} = \sum_{k=1}^{d} \frac{1}{k!} \sum_{j=0}^{k} (-1)^j \binom{k}{j} (k-j)^n. \tag{9.29}$$

Exploiting Lemma 9.4 we can write the optimal generalized instrument as follows

$$S_{\boldsymbol{x},\boldsymbol{y}} := R'_{\boldsymbol{i},\boldsymbol{j}} = R'_{\sigma(\boldsymbol{i}),\sigma(\boldsymbol{j})}, \tag{9.30}$$

where $(\boldsymbol{x}, \boldsymbol{y})$ is a couple of strings of indices that represents one equivalence class. We will denote by L the set of equivalence classes $\mathsf{L} := \{(\boldsymbol{x}, \boldsymbol{y})\}$. The figure of merit can finally be written as follows

$$F = \frac{1}{d^M} \sum_{(\boldsymbol{x},\boldsymbol{y}) \in \mathsf{L}} n(\boldsymbol{x}, \boldsymbol{y}) \langle S_{\boldsymbol{x},\boldsymbol{y}} \rangle, \tag{9.31}$$

where $n(\boldsymbol{x}, \boldsymbol{y})$ is the cardinality of the equivalence class denoted by the couple $(\boldsymbol{x}, \boldsymbol{y})$, and $\langle S_{\boldsymbol{x},\boldsymbol{y}} \rangle = \langle \boldsymbol{i} | \langle \boldsymbol{j} | R'_{\boldsymbol{i},\boldsymbol{j}} | \boldsymbol{i} \rangle | \boldsymbol{j} \rangle$ for any string $\boldsymbol{i}, \boldsymbol{j}$ in the equivalence class denoted by $(\boldsymbol{x}, \boldsymbol{y})$. As a consequence of Schur's lemmas, the condition of Eq. (9.19) implies the following structure for the operators $S_{\boldsymbol{x},\boldsymbol{y}}$ (see Appendix B for the details)

$$S_{\boldsymbol{x},\boldsymbol{y}} = \bigoplus_{\nu} P^{\nu} \otimes r^{\nu}_{\boldsymbol{x},\boldsymbol{y}}, \tag{9.32}$$

where ν labels the irreducible representations in the Clebsch-Gordan series of $U_{\mathsf{out}}^{\otimes M} \otimes U^{*\otimes N}_{\mathsf{in}}$, and P^{ν} acts as the identity on the invariant subspaces of the representations ν, while $r^{\nu}_{\boldsymbol{x},\boldsymbol{y}}$ acts on the multiplicity space of the same representation. In the simplest case $M + N = 2$ we have

$$R_{a,b} = P^p r^p_{a,b} + P^q r^q_{a,b}, \tag{9.33}$$

where P^p and P^q are defined in Eq. (B.42). and $r^p_{a,b}$ and $r^q_{a,b}$ are non-negative numbers. In the case $M + N = 3$, with $M, N \neq 0$ we have two different decompositions, depending whether $d > 2$ or $d = 2$. When $d > 2$, we have (see Eq. (B.51))

$$R_{\boldsymbol{x},\boldsymbol{y}} = P^{\alpha} \otimes r^{\alpha}_{\boldsymbol{x},\boldsymbol{y}} + P^{\beta} r^{\beta}_{\boldsymbol{x},\boldsymbol{y}} + P^{\gamma} r^{\gamma}_{\boldsymbol{x},\boldsymbol{y}}. \tag{9.34}$$

When $d = 2$ we have that $\dim(\mathcal{H}_{\gamma,-}) = 0$ and the decomposition (9.34) becomes

$$R_{\boldsymbol{x},\boldsymbol{y}} = P^{\alpha} \otimes r^{\alpha}_{\boldsymbol{x},\boldsymbol{y}} + P^{\beta} r^{\beta}_{\boldsymbol{x},\boldsymbol{y}}. \tag{9.35}$$

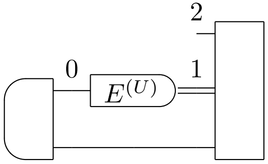
9.3 Optimal learning

In this section we derive the optimal quantum learning of a von Neumann measurement; we will consider the following scenarios:

- $1 \to 1$ learning
- $2 \to 1$ learning
- $3 \to 1$ learning
- $1 \to 2$ learning

9.3.1 $1 \to 1$ case

Consider the case in which today we are provided with a single use of a measurement device, and we need a replica to measure a state that will be prepared tomorrow; this scenario is described by the following scheme



(9.36)

Using the labeling as in Eq. (9.36) and exploiting the results of Section 9.2 for the case $M+N=2$, we have

$$\begin{aligned}
&\mathsf{L} = \{(x,x),(x,y)\},\\
&R_{i_{210}} = |i\rangle\langle i|_1 \otimes R_{x,x_{20}} + (I - |i\rangle\langle i|)_1 \otimes R_{x,y_{20}}\\
&R_{a,b} = P^p r^p_{a,b} + P^q r^q_{a,b}, \quad (a,b) \in \mathsf{L}
\end{aligned} \tag{9.37}$$

Exploiting the identity $\langle i| \langle j| P^p |i\rangle |j\rangle = \delta_{ij} 1/d$, and considering that $n(x,x) = d$ and $n(x,y) = d(d-1)$, the figure of merit in Eq. (9.31) for the can be rewritten as

$$\begin{aligned}
F =&\langle R_{x,x}\rangle + (d-1)\langle R_{x,y}\rangle =\\
&\sum_{\nu\in\{p,q\}} \left(r^\nu_{x,x}\Delta^\nu_{x,x} + (d-1) r^\nu_{x,y}\Delta^\nu_{x,y}\right),
\end{aligned} \tag{9.38}$$

where $\Delta^p_{x,x} = \frac{1}{d}$, $\Delta^p_{x,y} = 0$, and $\Delta^q_{a,b} = 1 - \Delta^p_{a,b}$. Let us now write the normalization conditions for the generalized instrument in terms of operators $R_{i,j}$. We have that that $R_\Omega := \sum_i R_i$ has to be the Choi operator of a deterministic quantum network and must satisfy Eq. (2.60), that is

$$R_\Omega = I_2 \otimes I_1 \otimes \rho \qquad \mathrm{Tr}[\rho] = 1, \quad \rho \geqslant 0. \tag{9.39}$$

The commutation relation (9.19) implies $[\rho, U^*] = 0$ that by Schur's lemmas gives

$$\rho = \frac{I}{d}. \tag{9.40}$$

Now, exploiting Eqs. (9.37) and (9.40), Eq. (9.39) becomes

$$I_1 \otimes R_{x,x} + (d-1) I_1 \otimes R_{x,y} = \frac{I}{d} \tag{9.41}$$

Substituting the expression of Eq. (9.33) in Eq. (9.41), we obtain

$$r^p_{x,x} + (d-1) r^p_{x,y} = r^q_{x,x} + (d-1) r^q_{x,y} = \frac{1}{d}. \tag{9.42}$$

From the constraint (9.42) the following bound follows

$$\begin{aligned}
F =&\sum_\nu \left(r^\nu_{x,x}\Delta^\nu_{x,x} + (d-1) r^\nu_{x,y}\Delta^\nu_{x,y}\right) \leq\\
&\sum_{\nu\in\{p,q\}} \overline{\Delta}^\nu \left(r^\nu_{x,x} + (d-1) r^\nu_{x,y}\right) = \frac{d+1}{d^2},
\end{aligned} \tag{9.43}$$

where $\overline{\Delta}^{\nu} := \max_{ij} \Delta^{\nu}_{i,j}$. The bound (9.43) is achieved by

$$r^q_{x,x} = r^p_{x,y} = 0, \quad r^p_{x,x} = \frac{1}{d}, \quad r^q_{x,y} = \frac{1}{d(d-1)},$$

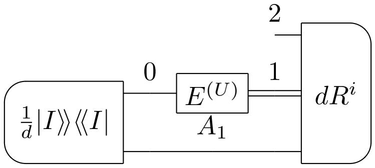
which corresponds to generalized instrument

$$R_i = |i\rangle\langle i|_1 \otimes \frac{1}{d} P^p + (I - |i\rangle\langle i|)_1 \otimes \frac{1}{d(d-1)} P^q, \tag{9.44}$$

that replicates the original Von Neuman measurement as follows

$$\begin{aligned} &G_i^{(U)} = R^i * E_{10}^{(U)^T} = \\ &\frac{1}{d(d-1)} U |i\rangle\langle i|_1 U^\dagger + \frac{d^2-d-1}{d^2(d-1)} I. \end{aligned} \tag{9.45}$$

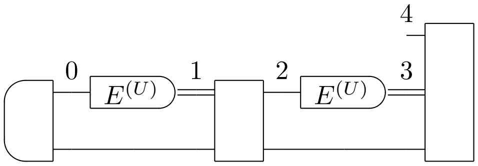
The optimal learning strategy can be realized by the following network



(9.46)

9.3.2 $2 \to 1$ case

We now consider the case in which we have two uses of $E^{(U)}$ at our disposal



(9.47)

Exploiting the symmetries introduced in Section 9.2 we have

$$\mathsf{L} = \{(x,xx),(x,xy),(x,yx),(x,yy),(x,yz)\}$$

$$R_i = \sum_{j,k} |j\rangle\langle j|_3 \otimes |k\rangle\langle k|_1 \otimes R'_{i,jk} \tag{9.48}$$

$$[R'_{i,jk}, U_4 \otimes U_2^* \otimes U_0^*] = 0 \tag{9.49}$$

$$R'_{i,jk} = \begin{cases} R_{x,xx} & \text{if} \quad i=j=k \\ R_{x,xy} & \text{if} \quad i=j\neq k \\ R_{x,yx} & \text{if} \quad i=k\neq j \\ R_{x,yy} & \text{if} \quad j=k\neq i \\ R_{x,yz} & \text{if} \quad i\neq j\neq k\neq i. \end{cases} \tag{9.50}$$

The figure of merit (9.20) becomes

$$F = \frac{1}{d} \sum_{(a,bc)\in \mathsf{L}} n(a,bc)\langle R_{a,bc}\rangle. \tag{9.51}$$

Let us now consider the normalization condition of the following generalized instrument $\boldsymbol{R}$

$$\sum_i R_i = I_4 \otimes I_3 \otimes S_{210} \qquad \mathrm{Tr}_2[S] = I_1 \otimes \rho_0. \tag{9.52}$$

Exploiting Eq. (9.48) we have

$$\begin{aligned}
&\sum_i R_i = \sum_{i,j,k} |j\rangle\,\langle j|_3 \otimes |k\rangle\,\langle k|_1 \otimes R'_{i,jk} = I_4 \otimes I_3 \otimes S_{210}\\
&\sum_{i,k} |k\rangle\,\langle k|_1 \otimes R'_{i,jk} = I_4 \otimes S_{210}, \quad \forall j\\
&\sum_i R'_{i,jk} = I_4 \otimes \langle k|\, S_{210}\, |k\rangle_1\,, \quad \forall j,k
\end{aligned} \tag{9.53}$$

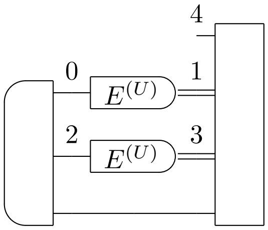
Exploiting the property (9.21) we have

$$\begin{aligned}
&I_4 \otimes \langle k|\, S_{210}\, |k\rangle_1 = \sum_i R'_{i,jk} = \sum_i R'_{\sigma(i),\sigma(j)\sigma(k)} =\\
&\quad = I_4 \otimes (\langle k|\, \sigma) S_{210} (\sigma\, |k\rangle_1) \quad \forall j,k.
\end{aligned} \tag{9.54}$$

This finally implies

$$\sum_i R'_{i,jk} = I_4 \otimes T_{20} \;\; \forall j,k \qquad \mathrm{Tr}_{20}[T] = 1. \tag{9.55}$$

Eq. (9.55) implies that the optimal strategy can be parallelized



(9.56)

Eq. (9.56) induces a further symmetry of the problem:

Lemma 9.5 *The operator $R'_{i,jk}$ in Eq. (9.48) can be chosen to satisfy:*

$$R'_{i,jk} = \mathsf{S} R'_{i,kj} \mathsf{S} \quad \forall k,j \tag{9.57}$$

where S is the swap operator $\mathsf{S}\, |k\rangle_2\, |j\rangle_0 = |j\rangle_2\, |k\rangle_0$.

Proof. The proof consists in the standard averaging argument. let us define $\overline{R}_{i,jk} := \frac{1}{2}(R'_{i,jk} + \mathsf{S}R'_{i,kj}\mathsf{S})$. It is easy to prove that $\{\overline{R}_{i,jk}\}$ satisfies the normalization (9.55) and that gives the same value of F as $R'_{i,kj}$. ■

Eq. (9.57) together with the decomposition (9.34) gives

$$\sigma_z r^{\alpha}_{a,bc}\sigma_z = r^{\alpha}_{a,cb} \quad r^{\beta}_{a,bc} = r^{\beta}_{a,cb} \quad r^{\gamma}_{a,bc} = r^{\gamma}_{a,cb} \tag{9.58}$$

where $\sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ and we used the property (B.47).

Considering that $n(x,xx) = d$, $n(x,xy) = n(x,yx) = n(x,yy) = d(d-1)$, and $n(x,yz) = d(d-1)(d-2)$, and that $\mathsf{S}R_{x,xy}\mathsf{S} = R_{x,yx}$, the figure of merit in Eq. (9.31) can be written as

$$\begin{aligned} F =&\langle R_{x,xx}\rangle + (d-1)\langle R_{x,yy}\rangle + 2(d-1)\langle R_{x,xy}\rangle + \\ &(d-1)(d-2)\langle R_{x,yz}\rangle = \\ &= \sum_{\nu} \mathrm{Tr}[\Delta^{\nu}_{x,xx} r^{\nu}_{x,xx} + (d-1)\Delta^{\nu}_{x,yy} r^{\nu}_{x,yy} + \\ &2(d-1)\Delta^{\nu}_{x,xy} r^{\nu}_{x,xy} + (d-1)(d-2)\Delta^{\nu}_{x,yz} r^{\nu}_{x,yz}] \end{aligned} \tag{9.59}$$

where

$$\Delta^{\nu}_{a,bc} := \mathrm{Tr}_{\mathcal{H}_\nu}[|ijk\rangle\langle ijk|], \tag{9.60}$$

and i, jk is any triple of indices in the class denoted by a, bc. Notice that in the case $d = 2$ the last term in the sum of Eq. (9.59) is 0. In particular, by direct calculation we have

$$\begin{aligned} &\Delta^{\alpha}_{x,xx} = \begin{pmatrix} \frac{2}{d+1} & 0 \\ 0 & 0 \end{pmatrix}, \quad \Delta^{\alpha}_{x,xy} = \frac{1}{2}\begin{pmatrix} \frac{1}{d+1} & \frac{1}{\sqrt{d^2-1}} \\ \frac{1}{\sqrt{d^2-1}} & \frac{1}{d-1} \end{pmatrix}, \quad \Delta^{\alpha}_{x,yx} = \sigma_z \Delta^{\alpha}_{x,xy}\sigma_z \\ &\Delta^{\alpha}_{x,yy} = \Delta^{\alpha}_{x,yz} = 0, \\ &\Delta^{\beta}_{x,xx} = \frac{d-1}{d+1}, \quad \Delta^{\beta}_{x,xy} = \frac{d}{2(d+1)}, \\ &\Delta^{\beta}_{x,yy} = 1, \quad \Delta^{\beta}_{x,yz} = \frac{1}{2}, \\ &\Delta^{\gamma}_{x,xx} = \Delta^{\gamma}_{x,yy} = 0, \quad \Delta^{\gamma}_{x,xy} = \frac{d-2}{2(d-1)}, \quad \Delta^{\gamma}_{x,yz} = \frac{1}{2}. \end{aligned} \tag{9.61}$$

The commutation relation (9.19) implies $[I_4 \otimes T_{20}, U_4^* \otimes U_2 \otimes U_0] = 0$ and taking the trace on $\mathcal{H}_4$ we get

$$[T_{20}, U_0 \otimes U_2] = 0, \tag{9.62}$$

which by theorem B.2 and the decomposition (B.33) implies $T_{20} = t_+ P^+ + t_- P^-$. The normalization $\mathrm{Tr}_{20}[T] = 1$ becomes $d_+ t_+ + d_- t_- = 1$ and Eq. (9.55) becomes

$$\begin{aligned} &\sum_{(a,bc)\in\mathsf{L}} \frac{n(a,bc)}{d^2}\left(r^{\alpha}_{a,bc} \otimes P^{\alpha} + r^{\beta}_{a,bc}P^{\beta} + r^{\gamma}_{a,bc}P^{\gamma}\right) = \\ &\qquad I_4 \otimes (t_+ P^+ + t_- P^-) = \\ &\qquad t_+(|+\rangle\langle+| \otimes P^{\alpha} + P^{\beta}) + t_-(|-\rangle\langle-| \otimes P^{\alpha} + P^{\gamma}), \end{aligned} \tag{9.63}$$

independently of j, k. This in turn implies that

$$\begin{aligned}
t_+ &= \sum_{(a,bc)\in \mathsf{L}} \frac{n(a,bc)}{d^2} \langle +|\, r^\alpha_{a,bc} \,|+\rangle = \sum_{(a,bc)\in \mathsf{L}} \frac{n(a,bc)}{d^2} r^\beta_{a,bc} \\
t_- &= \sum_{(a,bc)\in \mathsf{L}} \frac{n(a,bc)}{d^2} \langle -|\, r^\alpha_{a,bc} \,|-\rangle = \sum_{(a,bc)\in \mathsf{L}} \frac{n(a,bc)}{d^2} r^\gamma_{a,bc} \\
0 &= \sum_{(a,bc)\in \mathsf{L}} \frac{n(a,bc)}{d^2} \langle \pm|\, r^\alpha_{a,bc} \,|\mp\rangle
\end{aligned} \tag{9.64}$$

where we exploited the decomposition (B.44). Let us now introduce the notation

$$\begin{aligned}
s^\nu_{x,xx} &:= r^\nu_{x,xx} & s^\nu_{x,xy} &:= (d-1) r^\nu_{x,xy} \\
s^\nu_{x,yx} &:= (d-1) r^\nu_{x,yx} & s^\nu_{x,yy} &:= (d-1) r^\nu_{x,yy} \\
s^\nu_{x,yz} &:= (d-2)(d-1) r^\nu_{x,yz}.
\end{aligned} \tag{9.65}$$

Exploiting Eq. (9.50) and Eq. (9.58) the constraint (9.64) becomes

$$\begin{aligned}
& s^\alpha_{x,xx} + s^\alpha_{x,yy} = \begin{pmatrix} t_+ & 0 \\ 0 & t_- \end{pmatrix} \\
& s^\beta_{x,xx} + s^\beta_{x,yy} = t_+ \\
& s^\gamma_{x,xx} + s^\gamma_{x,yy} = t_- \\
& s^\alpha_{x,xy} + \sigma_z s^\alpha_{x,xy} \sigma_z + s^\alpha_{x,yz} = \begin{pmatrix} (d-1)t_+ & 0 \\ 0 & (d-1)t_- \end{pmatrix} \\
& 2 s^\beta_{x,xy} + s^\beta_{x,yz} = (d-1) t_+ \\
& 2 s^\gamma_{x,xy} + s^\gamma_{x,yz} = (d-1) t_-
\end{aligned} \tag{9.66}$$

and the figure of merit (9.59) becomes

$$F = \sum_\nu \sum_{(a,bc)\in \mathsf{L}} \mathrm{Tr}[\Delta^\nu_{a,bc} s^\nu_{a,bc}] \tag{9.67}$$

We are now ready to derive the optimal learning network; we will proceed as follows: i) first we will maximize the value of F for a fixed value of t_+ (remember that $t_- = (1 - d_+ t_+)/d_-$) and then ii) we will find the value of t_+ that maximize F. The figure of merit can be rewritten as:

$$F = F_\alpha + F_\beta + F_\gamma \tag{9.68}$$

where

$$F_\nu = \sum_{(a,bc)\in \mathsf{L}} \mathrm{Tr}[\Delta^\nu_{a,bc} s^\nu_{a,bc}]. \tag{9.69}$$

We now maximize F_β and F_γ for the case $d \geqslant 3$. Reminding the expressions (9.61) for the $\Delta^\nu_{i,jk}$ we have:

$$\begin{aligned} F_\beta = \sum_{(a,bc)\in \mathsf{L}} &\mathrm{Tr}[\Delta^\beta_{a,bc} s^\beta_{a,bc}] \leqslant \\ &\max(\Delta^\beta_{x,xx}, \Delta^\beta_{x,yy}) t_+ + \max(\Delta^\beta_{x,xy}, \Delta^\beta_{x,yz})(d-1)t_+ = \\ &\Delta^\beta_{x,yy} t_+ + \Delta^\beta_{x,yz}(d-1)t_+ = \\ &t_+ + \frac{d-1}{2} t_+ = \frac{d+1}{2} t_+ \end{aligned} \tag{9.70}$$

and

$$\begin{aligned} F_\gamma = \sum_{(a,bc)\in \mathsf{L}} &\mathrm{Tr}[\Delta^\gamma_{a,bc} s^\gamma_{a,bc}] \leqslant \\ &\max(\Delta^\gamma_{x,xx}, \Delta^\gamma_{x,yy}) t_- + \max(\Delta^\gamma_{x,xy}, \Delta^\gamma_{x,yz})(d-1)t_- = \\ &\Delta^\gamma_{x,yz} \frac{d-1}{2} t_- = \frac{d-1}{2} t_- . \end{aligned} \tag{9.71}$$

where we used the normalizations constraints (9.67). The upper bounds (9.70) and (9.71) can be achieved by taking

$$\begin{aligned} &s^\beta_{x,xx} = s^\beta_{x,xy} = s^\beta_{x,yx} = s^\gamma_{x,xx} = s^\gamma_{x,xy} = s^\gamma_{x,yx} = 0, \\ &s^\beta_{x,yy} = t_+, \quad s^\beta_{x,yz} = (d-1)t_+, \\ &s^\gamma_{x,yy} = t_-, \quad s^\gamma_{x,yz} = (d-1)t_-. \end{aligned}$$

For $d = 2$ the irreducible representation denoted by γ and the x, yz class do not exist and the optimization yields $s^\beta_{x,xy} = \frac{d-1}{2} t_+$.

Let us now consider F_α (in this case there is no difference between $d \geqslant 3$ and $d = 2$); reminding the expression of the $\Delta^\alpha_{i,jk}$ we have:

$$\begin{aligned} F_\alpha = \sum_{(a,bc)\in \mathsf{L}} &\mathrm{Tr}[\Delta^\alpha_{a,bc} s^\alpha_{a,bc}] = \\ &\mathrm{Tr}[\Delta^\alpha_{x,xx} s^\alpha_{x,xx}] + \mathrm{Tr}[2\Delta^\alpha_{x,xy} s^\alpha_{x,xy}] = \\ &\mathrm{Tr}\left[\begin{pmatrix} \frac{2}{d+1} & 0 \\ 0 & 0 \end{pmatrix} s^\alpha_{x,xx} + \begin{pmatrix} \frac{1}{d+1} & \frac{1}{\sqrt{d^2-1}} \\ \frac{1}{\sqrt{d^2-1}} & \frac{1}{d-1} \end{pmatrix} s^\alpha_{x,xy} \right] \leqslant \\ &\frac{2}{d+1} t_+ + \mathrm{Tr}\left[\begin{pmatrix} \frac{1}{d+1} & \frac{1}{\sqrt{d^2-1}} \\ \frac{1}{\sqrt{d^2-1}} & \frac{1}{d-1} \end{pmatrix} s^\alpha_{x,xy} \right], \end{aligned} \tag{9.72}$$

the bound can be achieved by taking

$$s^\alpha_{x,xx} = \begin{pmatrix} t_+ & 0 \\ 0 & t_- \end{pmatrix}. \tag{9.73}$$

Let us now focus on the expression $\mathrm{Tr}[\Delta^{\alpha}_{x,xy} s^{\alpha}_{x,xy}]$. The normalization constraint (9.66) for the operator $s^{\alpha}_{x,xy}$ can be rewritten as:

$$\begin{pmatrix} 2s^{\alpha,+,+}_{x,xy} + s^{\alpha,+,+}_{x,yz} & s^{\alpha,+,-}_{x,yz} \\ s^{\alpha,-,+}_{x,yz} & 2s^{\alpha,-,-}_{x,xy} + s^{\alpha,-,-}_{x,yz} \end{pmatrix} = (d-1)\begin{pmatrix} t_+ & 0 \\ 0 & t_- \end{pmatrix} \tag{9.74}$$

which implies

$$\begin{aligned} & s^{\alpha,+,-}_{x,yz} = s^{\alpha,-,+}_{x,yz} = 0 \\ & s^{\alpha,+,+}_{x,yz} + 2s^{\alpha,+,+}_{x,xy} = (d-1)t_+ \\ & s^{\alpha,-,-}_{x,yz} + 2s^{\alpha,-,-}_{x,xy} = (d-1)t_-. \end{aligned} \tag{9.75}$$

Then we have

$$\begin{aligned} \mathrm{Tr}[\Delta^{\alpha}_{x,xy} s^{\alpha}_{x,xy}] = & \frac{s^{\alpha,+,+}_{x,xy}}{d+1} + \frac{s^{\alpha,+,-}_{x,xy}}{\sqrt{d^2-1}} + \\ & \frac{s^{\alpha,-,+}_{x,xy}}{\sqrt{d^2-1}} + \frac{s^{\alpha,-,-}_{x,xy}}{d-1} \leqslant \\ & \frac{s^{\alpha,+,+}_{x,xy}}{d+1} + 2\frac{\sqrt{s^{\alpha,+,+}_{x,xy} s^{\alpha,-,-}_{x,xy}}}{\sqrt{d^2-1}} + \frac{s^{\alpha,-,-}_{x,xy}}{d-1} \leqslant \qquad (9.76) \\ & \frac{(d-1)t_+}{2(d+1)} + \frac{\sqrt{(d-1)t_+ t_-}}{\sqrt{d+1}} + \frac{t_-}{2} \qquad (9.77) \end{aligned}$$

where we used the positivity of the operator $s^{\alpha}_{x,xy}$ for the inequality (9.76) and the normalization (9.75) for the second inequality (9.77). The upper bound in Eq. (9.77) can be achieved by taking

$$s^{\alpha}_{x,xy} = \frac{(d-1)}{2}\begin{pmatrix} t_+ & \sqrt{t_+ t_-} \\ \sqrt{t_+ t_-} & t_- \end{pmatrix} \tag{9.78}$$

We can now write the figure of merit as:

$$\begin{aligned} F &= F_\alpha + F_\beta + F_\gamma = \\ &= \frac{(d-1)t_+}{2(d+1)} + \frac{\sqrt{(d-1)t_+ t_-}}{\sqrt{d+1}} + \frac{t_-}{2} + \frac{d+1}{2}t_+ + \frac{d-1}{2}t_- = \\ &= \frac{d^2+3d}{2(d+1)}t_+ + \frac{\sqrt{(d-1)t_+ t_-}}{\sqrt{d+1}} + \frac{d}{2}t_- \end{aligned} \tag{9.79}$$

The last step of the optimization can be easily done by making the substitution $t_- = d_-^{-1}(1 - d_+ t_+)$ in Eq. (9.79) and then maximizing $F = F(t_+)$. We will omit the details of the derivation and we rather show a plot (Fig. 9.1) representing the value of F depending on the dimension With the optimal learning network the replicated POVM has the following form:

$$G_i^{(U)} = \frac{dF-1}{d-1} U\,|i\rangle\,\langle i|\,U^\dagger + \frac{1-F}{d-1} \tag{9.80}$$

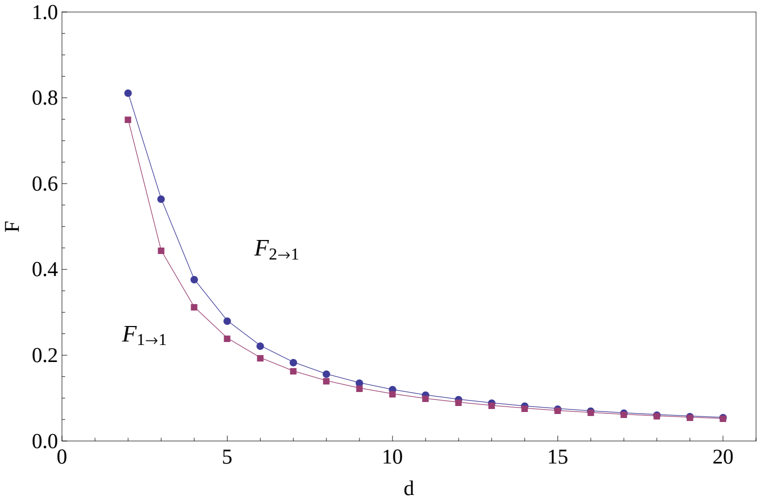
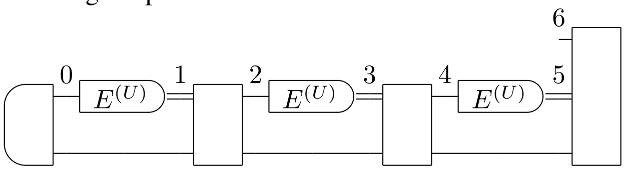


Figure 9.1. Optimal learning of a measurement device: we present the value of F for different values of the dimension d. The squared dots represent the optimal learning from a single use (1 ⟶ 1 learning) while the round dots represent the optimal learning from two uses (2 ⟶ 1 learning).

9.3.3 $3 \longrightarrow 1$ case

In this section we consider a learning network exploiting 3 uses of the measurement device and produces a single replica:



. (9.81)

In order to simplify the problem we restricy ourselves to the qubit case, that is we set $d = 2$. The derivation of the optimal learning network turns out to be very cumbersome althogh it follows the same lines as for the $2 \longrightarrow 1$ case. The $3 \longrightarrow 1$ scenario deserves interest because the optimal strategy does not allow for a strategy using the 3 uses of the measurement device in parallel.

Let us consider the normalization condition for the generalizd instrument $\{R_i\}$:

$$\begin{aligned}&\sum_{ijkl} |jkl\rangle \langle jkl|_{531} \otimes R'_{i,jkl} = I_{65} \otimes S_{43210}\\&\mathrm{Tr}_4[S] = I_3 \otimes T_{210}\end{aligned} \tag{9.82}$$

This implies

$$\begin{aligned}&\sum_{i} R'_{i,jkl} = I_6 \otimes \langle kl| \, S_{43210} \, |kl\rangle_{31} \quad \forall j,\\&\langle kl| \, \mathrm{Tr}_4[S] \, |kl\rangle = \langle l| \, T \, |l\rangle_1 \quad \forall k.\end{aligned} \tag{9.83}$$

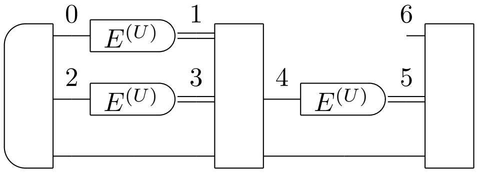
From the relabeling symmetry $R'_{i,jkl}=R'_{\sigma(i),\sigma(j)\sigma(k)\sigma(l)}$ we have $\langle kl|\, S\, |kl\rangle=\langle\sigma(k)\sigma(l)|\, S|\sigma(k)\sigma(l)\rangle$, and consequently

$$\langle kl|\, \mathrm{Tr}_4[S]\, |kl\rangle_{31} = \frac{1}{d^2}\,\mathrm{Tr}_{431}[S] =: \widetilde{T}_{20}, \quad \forall k,l. \tag{9.84}$$

This fact along with Eq. (9.82) allows us to conclude that

$$\begin{aligned}\mathrm{Tr}_4[S] = \mathrm{Tr}_4 \left[\sum_{kl} |kl\rangle\, \langle kl|_{31} \otimes \langle kl|\, S_{43210}\, |kl\rangle\right] = \\ \sum_{kl} |kl\rangle\, \langle kl|_{31} \otimes \widetilde{T}_{20} = I_{31} \otimes \widetilde{T}_{20}\end{aligned} \tag{9.85}$$

that implies that we can exploit the first two uses in parallel. We notice that in general $\langle kl|\, S\, |kl\rangle = \langle\sigma(k)\sigma(l)|\, S\, |\sigma(k)\sigma(l)\rangle$ does not imply that $\langle kl|\, S\, |kl\rangle = \widetilde{S}$ is independent of k,l, but only that $\langle kl|\, S\, |kl\rangle = \widetilde{S}_{ab}$, where a,b denotes the equivalence class of the couple (k,l). Consequently, we cannot in general assume that all the examples can be used in parallel. In fact, the optimal learning network has the following causal structure



$$. \tag{9.86}$$

where the state of system 4 depends on the classical outcome on system 3 and 1. The optimal fidelity achieves the value $F \simeq 0,87$ (we remind that for the $1 \to 1$ case we had $F = 0,75$ while for the $2 \to 1$ case we had $F = 0,81$).

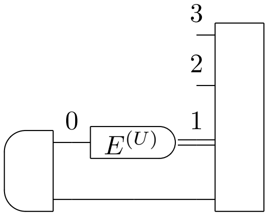
Remark 9.2 *One can wonder whether without assuming any symmetry it is possible to find a non-symmetric parallel strategy $\{R_i\}$ that achieves the optimal value of F. However we remind that for any strategy $\{R_i\}$ we can build a symmetric one with the same normalization, that is without spoiling the parallelism, and giving the same fidelity. Since the optimal symmetric network cannot be parallel, we have that any other optimal network has to be sequential as well.*

As we pointed out in Remark 6.1 the optimality of the parallel strategy is a common feature of the tasks involving group transformation. On the other hand, if the set of transformation considered is covariant under a group representation but does not form a group, the parallelism cannot be proven: the set of channels in Eq. (9.7) falls in the latter case. A similar situation arises in the Grover algorithm [17], that can be rephrased as the estimation of an unknown unitary from the set $\{U_n = I - 2\,|n\rangle\,\langle n|\}$; also in this case the unitaries $\{U_n\}$ do not a group and the optimal algorithm, as it was proved in Ref. [89], cannot be parallelized.

Quantum channel discrimination is a typical example of a task in which the optimality of sequential strategies easily arises. In Ref. [7] it was found that discrimination of unitary channels can be optimally performed in parallel, but as shown in Refs. [8, 90], there exist examples of non-unitary channels that can be better discriminate by sequential strategies.

9.3.4 $1 \to 2$ case

Our goal in this scenario is to create two replicas of the measurement after it was used once ($1 \to 2$ learning).



(9.87)

Using the symmetries we introduced in Section 9.2 we have

$$\mathsf{L} = \{(xx,x),(xx,y),(xy,x),(xy,y),(xy,z)\}$$

$$R_{ij} = \sum_k |k\rangle\langle k|_1 \otimes R'_{ij,k} \tag{9.88}$$

$$[R'_{ij,k}, U_3 \otimes U_2 \otimes U_0^*] = 0 \tag{9.89}$$

$$R'_{i,jk} = \begin{cases} R_{xx,x} & \text{if} \quad i=j=k \\ R_{xx,y} & \text{if} \quad i=j\neq k \\ R_{xy,x} & \text{if} \quad i=k\neq j \\ R_{xy,y} & \text{if} \quad j=k\neq i \\ R_{xy,z} & \text{if} \quad i\neq j\neq k\neq i. \end{cases} \tag{9.90}$$

and the figure of merit becomes

$$F = \frac{1}{d^2}\sum_{(ab,c)\in\mathsf{L}} n(ab,c)\langle R_{ab,c}\rangle \tag{9.91}$$

The commutation relations of $R_{ab,c}$ with $U_3 \otimes U_2 \otimes U_0^*$ is very similar to the one in Eq. (9.49) for the $2 \to 1$ case, because $U^* \otimes U \otimes U$ has same invariant subspaces as $U \otimes U^* \otimes U^*$. This enables us to write

$$R_{ab,c} = P^\alpha \otimes r^\alpha_{ab,c} + P^\beta r^\beta_{ab,c} + P^\gamma r^\gamma_{ab,c} \tag{9.92}$$

The following lemma introduces an additional symmetry property of the generalized instrument $\{R_{ij}\}$.

Lemma 9.6 *The operators $R_{ab,c}$ in Eq. (9.92) can be chosen to be satisfy*

$$R_{ab,c} = \mathsf{S}R_{ba,c}\mathsf{S} \quad \forall a,b,c \tag{9.93}$$

where S is the swap operator $\mathsf{S}\,|k\rangle_2\,|j\rangle_3 = |j\rangle_2\,|k\rangle_3$.

Proof. See Lemma 9.5. ■

Remark 9.3 *The symmetry (9.93) translates the possibility to exchange the inputs of the two replicas (Hilbert spaces $\mathcal{H}_2$ and $\mathcal{H}_3$) together with exchanging the measurement outcomes corresponding to these two replicas.*

Inserting the decomposition (9.92) into Eq. (9.93) and reminding Eq. (B.47) we have

$$\begin{aligned} &r^{\nu}_{ab,c} = r^{\nu}_{ba,c} \qquad \text{if } \nu = \beta, \gamma \\ &r^{\alpha}_{ab,c} = \sigma_z r^{\alpha}_{ba,c} \sigma_z \end{aligned} \tag{9.94}$$

Let us now consider the normalization constraint for the generalized instrument R_{ij}; since $\sum_{i,j} R_{ij}$ has to be a deterministic network we have

$$\sum_{ij} R_{ij} = I_{321} \otimes \rho_0, \quad \mathrm{Tr}[\rho] = 1 \tag{9.95}$$

where ρ has to be a positive operator. The commutation relation (9.89) implies $[\rho, U] = 0$ and so we have $\rho = \frac{1}{d} I_0$. Writing I_{3210} as $\sum_k |k\rangle\langle k|_1 \otimes (I_{m_\alpha} \otimes P^\alpha + P^\beta + P^\gamma)$ we can rewrite the normalization conditions as follows

$$\sum_{ij} R'^{\nu}_{ij,k} = \frac{1}{d} I_{m_\nu}. \tag{9.96}$$

If we use the following definitions

$$\begin{array}{ll} s^{\nu}_{xx,x} := r^{\nu}_{xx,x} & s^{\nu}_{xx,y} := (d-1) r^{\nu}_{xx,y} \\ s^{\nu}_{xy,x} := (d-1) r^{\nu}_{xy,x} & s^{\nu}_{xy,z} := (d-1)(d-2) r^{\nu}_{xy,z} \end{array} \tag{9.97}$$

the normalization becomes

$$\begin{aligned} s^{\nu}_{xx,x} + s^{\nu}_{xx,y} + 2 s^{\nu}_{xy,x} + s^{\nu}_{xy,z} &= \frac{1}{d}, \quad \text{if } \nu = \beta, \gamma \\ s^{\alpha}_{xx,x} + s^{\alpha}_{xx,y} + s^{\alpha}_{xy,x} + \sigma_z s^{\alpha}_{xy,x} \sigma_z + s^{\alpha}_{xy,z} &= \frac{1}{d} I_{m_\alpha} \end{aligned} \tag{9.98}$$

where we used the relabeling symmetry. Let us now express the figure of merit in terms of the $s^{\nu}_{ab,c}$:

$$\begin{aligned} F &= F_\alpha + F_\beta + F_\gamma \qquad (9.99) \\ F_\nu &= \frac{1}{d} \mathrm{Tr}[\Delta^{\nu}_{xx,x} s^{\nu}_{xx,x} + 2 \Delta^{\nu}_{xy,x} s^{\nu}_{xy,x} + \Delta^{\nu}_{xx,y} s^{\nu}_{xx,y} + \Delta^{\nu}_{xy,z} s^{\nu}_{xy,z}]] \end{aligned}$$

where $\Delta^{\nu}_{ab,c}$ are the same as the $\Delta^{\nu}_{a,bc}$ in Eq. (9.61) taking into account the change of Hilbert space labelling from $\mathcal{H}_0, \mathcal{H}_2, \mathcal{H}_4$ to $\mathcal{H}_2, \mathcal{H}_3, \mathcal{H}_0$. The maximization of F_β and F_γ is simple and yelds

$$F_\beta = \frac{1}{d^2} \qquad F_\gamma = \frac{1}{2d^2} \tag{9.100}$$

$$\begin{aligned} &s^{\beta}_{xx,x} = s^{\beta}_{xy,x} = s^{\beta}_{xy,y} = s^{\beta}_{xy,z} = 0 \\ &s^{\gamma}_{xx,x} = s^{\gamma}_{xx,y} = s^{\gamma}_{xy,x} = s^{\gamma}_{xy,y} = 0, \\ &s^{\beta}_{xx,y} = s^{\gamma}_{xy,z} = \frac{1}{d^2}. \end{aligned} \tag{9.101}$$

Let us now consider the maximization of F_α. Inserting the explicit expression of the $\Delta^\alpha_{ab,c}$ into Eq. (9.99) we have

$$\begin{aligned}
dF_\alpha =& \operatorname{Tr}\left[\begin{pmatrix} s^{\alpha,+,+}_{xx,x} & s^{\alpha,+,-}_{xx,x} \\ s^{\alpha,-,+}_{xx,x} & s^{\alpha,-,-}_{xx,x} \end{pmatrix}\begin{pmatrix} \frac{2}{d+1} & 0 \\ 0 & 0 \end{pmatrix}\right] + \\
&\operatorname{Tr}\left[\begin{pmatrix} s^{\alpha,+,+}_{xy,x} & s^{\alpha,+,-}_{xy,x} \\ s^{\alpha,-,+}_{xy,x} & s^{\alpha,-,-}_{xy,x} \end{pmatrix}\begin{pmatrix} \frac{1}{d+1} & \frac{1}{\sqrt{d^2-1}} \\ \frac{1}{\sqrt{d^2-1}} & \frac{1}{d-1} \end{pmatrix}\right] = \\
&= \frac{2s^{\alpha,+,+}_{xx,x}}{d+1} + \frac{s^{\alpha,+,+}_{xy,x}}{d+1} + \frac{s^{\alpha,-,-}_{xy,x}}{d-1} + \frac{2s^{\alpha,+,-}_{xy,x}}{\sqrt{d^2-1}} \leqslant \\
&\leqslant \frac{2}{(d+1)}\left(\frac{1}{d} - 2s^{\alpha,+,+}_{xy,x}\right) + \frac{s^{\alpha,+,+}_{xy,x}}{d+1} + \frac{s^{\alpha,-,-}_{xy,x}}{d-1} + 2\sqrt{\frac{s^{\alpha,+,+}_{xy,x}\, s^{\alpha,-,-}_{xy,x}}{d^2-1}} \leqslant \\
&\leqslant \frac{5d-3}{2d(d^2-1)} - \frac{3s^{\alpha,+,+}_{xy,x}}{d+1} + 2\sqrt{\frac{s^{\alpha,+,+}_{xy,x}}{2d(d^2-1)}}
\end{aligned} \tag{9.102}$$

where in the derivation of the bound (9.102) we used the positivity of $s^\alpha_{xy,x}$ and the constraints (9.98). The upper bound (9.102) can be achieved by taking

$$\begin{aligned}
&s^\alpha_{xx,x} = \begin{pmatrix} \frac{1}{d} - 2a & 0 \\ 0 & 0 \end{pmatrix} \quad s^\alpha_{xy,x} = \begin{pmatrix} a & \sqrt{\frac{1}{2d}a} \\ \sqrt{\frac{1}{2d}a} & \frac{1}{2d} \end{pmatrix}, \\
&s^\alpha_{xy,z} = s^\alpha_{xx,y} = 0
\end{aligned} \tag{9.103}$$

where we defined $a := s^{\alpha,+,+}_{xy,x}$. Eq. (9.102) gives the value of F_α as a function of a; the maximization of $F_\alpha(a)$ with the constraint $0 \leqslant a \leqslant \frac{1}{d}$ is easy and gives

$$F_\alpha = \frac{4(2d-1)}{3d^2(d^2-1)} \qquad \text{for } a = \frac{d+1}{18d(d-1)}. \tag{9.104}$$

and then for $d \geqslant 3$ we have

$$F = F_\alpha + F_\beta + F_\gamma = \frac{3d^2 + 4d + 4\sqrt{d^2-1} - 3}{2d^2(d^2-1)} \sim \frac{3}{2d^2}. \tag{9.105}$$

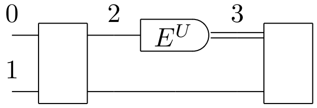
For $d = 2$ the invariant subspace $\mathcal{H}_\gamma$ does not appear and the fidelity becomes $F = F_\alpha + F_\beta = \frac{7+2\sqrt{3}}{12}$.

In the next section we consider a different scenario which is less restrictive than the learning scheme we have considered up to now. Similarly to what we had when comparing the optimal cloning and the optimal learning of a unitary, relaxing the constraints of the network allows to achieve better pefomances

9.4 Optimal cloning

In this section we turn our attention to the cloning scenario. As we previously discussed, this scheme is less restrictive than the learning one, since we allow both the M states to be measured and the N uses of the measurement device to be available at the same time.

We consider the case in which we are provided with a single use of the measurement device and we want to produce two replicas:



(9.106)

We can require for the optimal $1 \rightarrow 2$ cloning network the same symmetries we had for the $1 \rightarrow 2$ learning network. The set L in this case is $\mathsf{L} = \{(x, xx), (x, xy), (x, yx), (x, yy), (x, yz)\}$. Then the figure of merit becomes

$$F = \sum_{\nu} \sum_{(a,bc)\in\mathsf{L}} \mathrm{Tr}[\Delta^{\nu}_{a,bc} s^{\nu}_{a,bc}] \tag{9.107}$$

where $\Delta^{\nu}_{a,bc}$, $s^{\nu}_{a,bc}$ are the same as in section 9.3.4. The normalization condition for the $1 \rightarrow 2$ cloning scenario is different from the $1 \rightarrow 2$ learning. Instead of Eq. (9.96) we have

$$\sum_{i,jk} |i\rangle\langle i|_3 \otimes R'_{i,jk} = I_3 \otimes S_{210} \qquad \mathrm{Tr}_2[S] = I_{10} \tag{9.108}$$

which implies the following

$$\begin{aligned} I_{10} =& d\,\mathrm{Tr}_2[R_{x,xx} + R_{x,yy}] + \\ & d(d-1)\,\mathrm{Tr}_2[R_{x,xy} + R_{x,yx} + R_{x,yz}]. \end{aligned} \tag{9.109}$$

From the commutation $[R_{a,bc}, U_2^* \otimes U_1 \otimes U_0]$ it follows that $[\mathrm{Tr}_2[R_{a,bc}], U_1 \otimes U_0]$ and then, exploiting the decomposition (B.33) we have

$$\begin{aligned} t_+ P^+ + t_- P^- =& d\,\mathrm{Tr}_2[R_{x,xx} + R_{x,yy}] + \\ & d(d-1)\,\mathrm{Tr}_2[R_{x,xy} + R_{x,yx} + R_{x,yz}]. \end{aligned} \tag{9.110}$$

and finally by Eq. (9.109) $t_+ = t_- = 1$. Exploiting the decomposition $R_{a,bc} = \sum_\nu P^\nu \otimes r^\nu_{a,bc}$ along with Eq. (B.49), the normalization constraint (9.110) becomes

$$\begin{aligned} P^{\pm} =& P^{\pm} \sum_{\nu} \sum_{(a,bc)\in\mathsf{L}} \mathrm{Tr}_2[P^\nu \otimes s^\nu_{a,bc}] = \\ & \frac{1}{d_\pm} \sum_{(a,bc)\in\mathsf{L}} (d_\alpha s^{\alpha,\pm,\pm}_{a,bc} + d_{\delta_\pm} s^{\delta_\pm}_{a,bc}) P^{\pm}, \end{aligned} \tag{9.111}$$

where $\delta_+ = \beta$ and $\delta_- = \gamma$. Exploiting the relabeling symmetry (9.21) and the permutation symmetry (9.93) we have

$$d_+ = d_\alpha \sum_{(a,bc)\in\mathsf{L}} s^{\alpha,+,+}_{a,bc} + d_\beta \sum_{(a,bc)\in\mathsf{L}} s^{\beta}_{a,bc}, \tag{9.112}$$

$$d_- = d_\alpha \sum_{(a,bc)\in\mathsf{L}} s^{\alpha,-,-}_{a,bc} + d_\gamma \sum_{(a,bc)\in\mathsf{L}} s^{\gamma}_{a,bc}. \tag{9.113}$$

If we introduce the notation

$$s^{\beta}_{a,bc} := \begin{pmatrix} s^{\beta}_{a,bc} & 0 \\ 0 & 0 \end{pmatrix} \qquad s^{\gamma}_{a,bc} := \begin{pmatrix} 0 & 0 \\ 0 & s^{\gamma}_{a,bc} \end{pmatrix}$$

$$\Pi^{+} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \qquad \Pi^{-} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \tag{9.114}$$

the normalization constraints (9.112) and (9.113) can be rewritten as

$$\Pi^{+} \left(\sum_{\nu,(a,bc)\in \mathsf{L}} d_{\nu} s^{\nu}_{a,bc} \right) \Pi^{+} = d_{+}$$

$$\Pi^{-} \left(\sum_{\nu,(a,bc)\in \mathsf{L}} d_{\nu} s^{\nu}_{(a,bc)} \right) \Pi^{-} = d_{-}. \tag{9.115}$$

In order to solve the optimization problem we have to find the set $\mathsf{r} := \{r^{\nu}_{\ell}, \ell := (a,bc) \in \mathsf{L}, \nu \in \{\alpha, \beta\gamma\}\}$, $r^{\nu}_{\ell} \in \mathcal{L}(\mathbb{C}^2), r^{\nu}_{\ell} \geqslant 0$ subjected to the constraint (9.115) that maximizes the figure of merit (9.107); we will denote as M the set of all the r satisfying Eq. (9.115). Since the figure of merit (9.107) is linear and the set M is convex, a trivial result of convex analysis states that the maximum of a convex function over a convex set is achieved at an extremal point of the convex set. We now give two necessary conditions for a given r to be an extremal point of M. Let us start with the following

Definition 9.1 (Perturbation) *Let* s *be an element of* M*. A set of hermitian operators* $\mathsf{z} := \{z^{\nu}_{\ell}\}$ *is a* perturbation *of* s *if there exists* $\epsilon \geqslant 0$ *such that*

$$\mathsf{s} + h\mathsf{z} \in \mathsf{M} \qquad \forall h \in [-\epsilon, \epsilon] \tag{9.116}$$

where we defined $\mathsf{s} + h\mathsf{z} := \{s^{\nu}_{\ell} + hz^{\nu}_{\ell} | h \in [-\epsilon, \epsilon]\}$.

By the definition of perturbation it is easy to prove that an element s of M is extremal if and only if it admits only the trivial perturbation $z^{\nu}_{\ell} = 0 \, \forall \ell, \nu$. We now exploit this definition to prove two necessary conditions for extremality.

Lemma 9.7 *Let* s *be an extremal element of* M*. Then* s^{ν}_{ℓ} *has to be rank one for all* ℓ, ν*.*

Proof. Suppose that there is a $s^{\nu'}_{\ell'} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathsf{s}$ which is not rank one; then there exist ϵ such that $\mathsf{z} := \{0, \dots, 0, z^{\nu'}_{\ell'}, 0, \dots, 0\}$, $z^{\nu'}_{\ell'} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ is an admissible perturbation. ■

This lemma tells us that w.l.o.g. we can assume the optimal s to be a set of rank one matrices. Let us now consider a set s such that s^{ν}_{ℓ} is rank one for all ℓ, ν; any admissible perturbation z of s must satisfy

$$z^{\nu}_{\ell} = c^{\nu}_{\ell} s^{\nu}_{\ell} \qquad c^{\nu}_{\ell} \in \mathbb{R} \tag{9.117}$$

$$\Pi^{+} \left(\sum_{\nu,\ell} d_{\nu} c^{\nu}_{\ell} s^{\nu}_{\ell} \right) \Pi^{+} = \Pi^{-} \left(\sum_{\nu} d_{\nu} c^{\nu}_{\ell} s^{\nu}_{\ell} \right) \Pi^{-} = 0. \tag{9.118}$$

where the constraint (9.117) is required in order to have $s^\nu_\ell + hz^\nu_\ell \geqslant 0$, while Eq. (9.118) tells us that $\mathsf{s} + h\mathsf{z}$ satisfies the normalization (9.115). Let us now consider the map

$$f : \mathcal{L}(\mathbb{C}^2) \to \mathbb{C}^2 \qquad f(A) := \begin{pmatrix} \Pi^+ A \Pi^+ \\ \Pi^- A \Pi^- \end{pmatrix}$$
$$f\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a \\ d \end{pmatrix}$$

exploiting this definition Eq. (9.118) becomes

$$\sum_{\nu,\ell} c^\nu_\ell f(s^\nu_\ell) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{9.119}$$

Suppose now that the set $\bar{\mathsf{r}}$ has $N \geqslant 3$ elements; then $\{f(\bar{r}^\nu_\ell)\}$ is a set of $N \geqslant 3$ vectors of $\mathbb{C}^2$ that cannot be linearly independent. That being so, there exists a set of coefficients $\{c^\nu_\ell\}$ such that $\sum_{\nu,\ell} c^\nu_\ell f(s^\nu_\ell) = 0$ and then $z^\nu_\ell = c^\nu_\ell \bar{s}^\nu_\ell$ is a perturbation of $\bar{\mathsf{r}}$. We have then proved the following lemma

Lemma 9.8 *Let* s *be an extremal element of* M. *Then* s *cannot have more than* 2 *elements.*

Lemma 9.7 and Lemma 9.8 provide two sufficient conditions for extremality that allow us to restrict the search of the optimal s among the ones that satisfy

$$\mathsf{s} = \{s^{\nu'}_{\ell'}, s^{\nu''}_{\ell''}\} \qquad \mathrm{rank}(s^{\nu'}_{\ell'}) = \mathrm{rank}(s^{\nu''}_{\ell''}) = 1$$
$$\Pi^i \left(\sum_{\nu,\ell} d_\nu s^\nu_\ell \right) \Pi^i = d_i \quad i = +,- \tag{9.120}$$

The set of the admissible s is small enough to allow us to compute the value of F for all the possible cases. It turns out that the best choice is to take

$$\mathsf{s} = \{s^\alpha_{xx,x}, s^\alpha_{xy,x}\}$$
$$s^\alpha_{xx,x} = \begin{pmatrix} \frac{9d_+ -1}{9d} & 0 \\ 0 & 0 \end{pmatrix} \qquad s^\alpha_{xy,x} = \begin{pmatrix} \frac{1}{9d} & \frac{\sqrt{d_-}}{3d} \\ \frac{\sqrt{d_-}}{3d} & \frac{d_-}{d} \end{pmatrix};$$

the corresponding value of F is

$$F = \frac{4}{3d} \tag{9.121}$$

which is much higher then the maximum value (9.105) achieved by the $1 \to 2$ learning scheme.

10 Conclusion

The aim of this work was twofold. The first part was devoted to present a unified description of Quantum Networks in terms of their Choi operators. The core result of this approach are Theorem 2.5 and Theorem 2.6 that prove the isomorphism between the set of deterministic Quantum Networks and a set of suitably normalized positive operators. This result can be then generalized to probabilistic Networks. The second key ingredient of the theory is the notion of link product (see Definition 2.1) that allows us to express the composition of quantum networks in terms of their Choi operators (Theorem 2.12).

In the second part of the work, we made use of this formalism to solve some relevant optimization problems. The representation of Quantum Networks as positive operators is extremely efficient in handling tasks that involve manipulation of transformations like process tomography (Chapter 3) and cloning, learning and inversion of transformations (Chapters 5, 6 and 7).

Even if the tools provided by the general theory of Quantum Networks simplify a lot many scenarios, it is also true that in order to analytically carry on the optimization we had to make a clever use of the symmetries of the various problems. The full power of the general theory reveals itself when combined with the techniques provided by the group representation theory (Appendix B): this happy marriage lies at the core of the results achieved in the optimization problems involving unitary transformations.

However, in many problems in quantum information theory like for example in channel discrimination [6, 7, 8], we cannot exploit such strong symmetry properties; the general theory of quantum network is still powerful [7] but the results from group theory cannot be applied. A possible way out (in some cases the only one) is the numerical approach. The set of the admissible Choi operators of Quantum network with fixed causal structure, is a convex set of positive operator. It is then possible to implement computer routines [96, 97] that solve the semidefinite program corresponding to the optimization problem in exam.

Acknowledgements

We thank M. Sedlak for his productive collaboration on Quantum Networks theory, which provided some of the results reviewed here. This work is supported by Italian Ministry of Education through grant PRIN 2008 and the EC through project COQUIT. Research at Perimeter Institute for Theoretical Physics is supported in part by the Government of Canada through NSERC and by the Province of Ontario through MRI.

A Channel Fidelity

This short appendix has the purpose to introduce the *channel fidelity* as a notion of distance between quantum channels. This definition was introduced in [32] and discussed in [91] In the following we will review the definition of channel fidelity and some of its most relevant properties.

Definition A.1 (Channel Fidelity) *Let $\mathcal{C} \in \mathcal{L}(\mathcal{L}(\mathcal{H})_a, \mathcal{L}(\mathcal{H})_b)$ and $\mathcal{D} \in \mathcal{L}(\mathcal{L}(\mathcal{H})_a, \mathcal{L}(\mathcal{H})_b)$ be two quantum channels and C and D be their Choi-Jamiołkowsky operators. We call* channel fidelity *the following expression*

$$\mathcal{F}(\mathcal{C},\mathcal{D}) := f\left(\frac{C}{d_a}, \frac{D}{d_a}\right) \tag{A.1}$$

where f is the state fidelity $f(\rho,\sigma) := |\operatorname{Tr}[\sqrt{\sigma^{\frac{1}{2}}\rho\sigma^{\frac{1}{2}}}]|^2$.

The channel fidelity enjoys many properties inherited by the state fidelity:

Lemma A.1 (Properties of channel fidelity) *The channel fidelity $\mathcal{F}$ defined in definition A.1 enjoys the following properties:*

- $0 \leq \mathcal{F}(\mathcal{C},\mathcal{D}) \leq 1$*, and* $\mathcal{F}(\mathcal{C},\mathcal{D}) = 1$ *if and only if* $\mathcal{C} = \mathcal{D}$.
- $\mathcal{F}(\mathcal{C},\mathcal{D}) = \mathcal{F}(\mathcal{D},\mathcal{C})$ *(symmetry).*
- *For any two isometric channels $\mathcal{V}$ and $\mathcal{W}$ (i.e., $\mathcal{V}(\rho) = V\rho V^\dagger$ and $\mathcal{W}(\rho) = W\rho W^\dagger$ with isometry V and W), $\mathcal{F}(\mathcal{V},\mathcal{W}) = (1/d^2)|\operatorname{Tr}(U^\dagger V)|^2$.*
- *For any $0 < \lambda < 1$, $\mathcal{F}(\mathcal{C}, \lambda\mathcal{D}_1 + (1-\lambda)\mathcal{D}_2) \geq \lambda\mathcal{F}(\mathcal{C},\mathcal{D}_1) + (1-\lambda)\mathcal{F}(\mathcal{C},\mathcal{D}_2)$ (concavity).*
- $\mathcal{F}(\mathcal{C}_1 \otimes \mathcal{C}_2, \mathcal{D}_1 \otimes \mathcal{D}_2) = \mathcal{F}(\mathcal{C}_1,\mathcal{D}_1)\mathcal{F}(\mathcal{C}_2,\mathcal{D}_2)$ *(multiplicativity with respect to tensoring).*
- *$\mathcal{F}$ is invariant under composition with unitary channels, i.e., for any unitary channel $\mathcal{U}$, $\mathcal{F}(\mathcal{U} \star \mathcal{C}, \mathcal{U} \star \mathcal{D}) = \mathcal{F}(\mathcal{C},\mathcal{D})$.*
- *$\mathcal{F}$ does not decrease under composition with arbitrary channels, i.e., for any channel $\mathcal{R}$, $\mathcal{F}(\mathcal{R} \star \mathcal{C}, \mathcal{R} \star \mathcal{D}) \geq \mathcal{F}(\mathcal{C},\mathcal{D})$.*

Proof. See Ref. [32] ■

The following lemma provides a physical interpretation of the channel fidelity $\mathcal{F}(\mathcal{A},\mathcal{B})$ between two channels $\mathcal{A}$ and $\mathcal{B}$ (one of them unitary) as the fidelity between the output states of $\mathcal{A}$ and $\mathcal{B}$ uniformly averaged over all input pure states.

Lemma A.2 *Let $\mathcal{A} \in \mathcal{L}(\mathcal{L}(\mathcal{H})_a, \mathcal{L}(\mathcal{H})_b)$ and $\mathcal{B} \in \mathcal{L}(\mathcal{L}(\mathcal{H})_a, \mathcal{L}(\mathcal{H})_b)$ be two channels and let us define $d = \dim(\mathcal{H}_a)$. If either $\mathcal{A}$ or $\mathcal{B}$ is a unitary channel we have*

$$\mathsf{F} := \int \mathrm{d}\varphi f(\mathcal{A}(|\varphi\rangle\langle\varphi|), \mathcal{B}(|\varphi\rangle\langle\varphi|)) = \frac{d}{d+1}\mathcal{F}(\mathcal{A},\mathcal{B}) + \frac{1}{d+1} \tag{A.2}$$

where $|\varphi\rangle \in \mathcal{H}_a$, $\mathrm{d}\varphi$ is the normalized ($\int \mathrm{d}\varphi = 1$) Haar measure over the set of pure states and f is the state fidelity.

Proof. First we notice that we can parametrize each vector $|\psi\rangle \in \mathcal{H}_a$ as $U|0\rangle$ where $|0\rangle$ is a fixed vector and U is a unitary operator on $\mathcal{H}_a$; with this parametrization the measure $\mathrm{d}\varphi$ becomes the usual Haar measure $\mathrm{d}U$ of $\mathbf{SU}(d)$. The left hand side of Eq. (A.2) now becomes:

$$\mathsf{F} = \int \mathrm{d}U f(\mathcal{A}(U|0\rangle\langle 0|U^\dagger), \mathcal{B}(U|0\rangle\langle 0|U^\dagger)). \tag{A.3}$$

Now suppose that $\mathcal{B}$ is a unitary channel $\mathcal{B} = V \cdot V^\dagger$. Eq. (A.3) becomes:

$$\begin{aligned}\mathsf{F} &= \int \mathrm{d}U f(\mathcal{A}(U|0\rangle\langle 0|U^\dagger), \mathcal{B}(U|0\rangle\langle 0|U^\dagger)) = \\ &= \int \mathrm{d}U f(\mathcal{A}(U|0\rangle\langle 0|U^\dagger), VU|0\rangle\langle 0|U^\dagger V^\dagger) = \\ &= \int \mathrm{d}U \langle 0|U^\dagger V^\dagger(\ I \otimes \langle 0|U^T)A(I \otimes U^*|0\rangle)VU|0\rangle = \\ &= \mathrm{Tr}\left[(V^\dagger \otimes I)A(V \otimes I)\left(\int \mathrm{d}U U \otimes U^*(|0\rangle|0\rangle\langle 0|\langle 0|)U^\dagger\langle 0|U^T\right)\right]\end{aligned} \tag{A.4}$$

Reminding Theorem B.3 and the decomposition (B.42) we have

$$\int \mathrm{d}U U \otimes U^*(|0\rangle|0\rangle\langle 0|\langle 0|)U^\dagger\langle 0|U^T = \frac{1}{d(d+1)}|I\rangle\!\rangle\langle\!\langle I| + \frac{1}{d(d+1)}I \tag{A.5}$$

that leads to

$$\begin{aligned}\mathsf{F} &= \frac{1}{d(d+1)}\mathrm{Tr}\left[(V^\dagger \otimes I)A(V \otimes I)|I\rangle\!\rangle\langle\!\langle I|\right] + \frac{1}{d(d+1)}\mathrm{Tr}\left[(V^\dagger \otimes I)A(V \otimes I)\right] = \\ &= \frac{1}{d(d+1)}\mathrm{Tr}\left[A|V\rangle\!\rangle\langle\!\langle V|\right] + \frac{1}{d+1} = \frac{d}{d+1}\mathcal{F}(\mathcal{A},\mathcal{B}) + \frac{1}{d+1}\end{aligned} \tag{A.6}$$

■

B Elements of Group Representation Theory

This Appendix is an introduction to the basic tools of group representation theory that are needed in this work. The key results of the appendix are the Schur's lemma B.2 and the Theorem B.2 that allow us to decompose an operator that commutes with a unitary representation of a group. The last section of this appendix is devoted to the decomposition of some relevant tensor product representations. All the results in this appendix are presented without proofs; a more exhaustive presentation can be found for example in [92, 93, 94, 95].

B.1 Basic definitions

Definition B.1 (Group) *A group* $\mathbf{G}$ *is a set of elements with a law of composition that assigns each ordered couple of elements* $g_1, g_2 \in \mathbf{G}$ *another element* $g_1 g_2$ *of* $\mathbf{G}$*. This composition law has to satisfy the following requirements:*

$$g_1(g_2g_3) = (g_1g_2)g_3 \quad \forall g_1, g_2, g_3 \in \mathbf{G} \tag{B.1}$$

$$\exists e \in \mathbf{G} : ge = eg = g \quad \forall g \in \mathbf{G} \tag{B.2}$$

$$\forall g \in \mathbf{G} \exists g^{-1} \in \mathbf{G} : gg^{-1} = g^{-1}g = e. \tag{B.3}$$

If $\mathbf{G}$ *has a finite number of elements we say that* $\mathbf{G}$ *is a finite group.*

Typical examples of groups are

- $\mathbf{GL}(n, \mathbb{R})$: the set of $n \times n$ real invertible matrices with matrix multiplication;
- $\mathbf{S}_n$: the group of permutation of n objects (the composition is the successive operation of permutations);
- $\mathbf{U}(1)$: the set 1×1 unitary matrices with matrix multiplication;
- $\mathbf{SU}(d)$: the set of $d \times d$ unitary matrices with determinant 1 with matrix multiplication;

A relevant class of groups are Lie groups

Definition B.2 (Lie Group) *A group* $\mathbf{G}$ *which is a differentiable manifold and such that the maps*

$$(g_1, g_2) \rightarrow g_1 g_2, \qquad g_1 \rightarrow g_1^{-1} \tag{B.4}$$

are smooth, is a Lie group. If $\mathbf{G}$ *(as a manifold) is compact, we say that* $\mathbf{G}$ *is a compact Lie group.*

$\mathbf{GL}(n, \mathbb{R})$, $\mathbf{U}(1)$, $\mathbf{SU}(d)$ are Lie groups but only $\mathbf{U}(1)$ and $\mathbf{SU}(d)$ are compact. From now on we restrict to the case of finite group and compact Lie groups.

Definition B.3 (Unitary Representation) *Let* $\mathbf{G}$ *be a group and* $\mathcal{H}$ *a Hilbert space. A unitary representation of* $\mathbf{G}$ *on* $\mathcal{H}$ *is a map* $g \rightarrow U_g$ *from* $\mathbf{G}$ *to set of bounded linear operator* $\mathcal{B}(\mathcal{H})$ *such that:*

$$U_g \textit{ is unitary } \forall g \in \mathbf{G} \tag{B.5}$$

$$U_g U_h = U_{gh} \forall g, h \in \mathbf{G} \tag{B.6}$$

$$U_e = I. \tag{B.7}$$

Definition B.4 (Equivalent Representation) *Let $\{U_g \mid g \in \mathbf{G}\}$ be a unitary representation of $\mathbf{G}$ on $\mathcal{H}$ and $\{V_g \mid g \in \mathbf{G}\}$be a unitary representation of $\mathbf{G}$ on $\mathcal{K}$. We say that $\{U_g\}$ is equivalent to $\{V_g\}$ if there exists an isomorphism $T : \mathcal{H} \to \mathcal{K}$ such that*

$$TU_g = V_g T \qquad \forall g \in \mathbf{G} \tag{B.8}$$

$$T^\dagger T = I_{\mathcal{H}} \qquad TT^\dagger = I_{\mathcal{K}} \tag{B.9}$$

The isomorphism T is often called intertwiner.

Remark B.1 *The notion of representation makes a bridge between group theory and quantum physics. Indeed, the action of a group on an Hilbert space induces a transformation on the set of quantum states $\mathcal{S}(\mathcal{H})$*

$$\rho \to U_g \rho U_g^\dagger \qquad \rho \in \mathcal{S}(\mathcal{H}). \tag{B.10}$$

Definition B.5 (Invariant Subspace) *Let $\{U_g \mid g \in \mathbf{G}\}$ a unitary representation of $\mathbf{G}$ on $\mathcal{H}$ and let $\mathcal{K} \subseteq \mathcal{H}$, be a subspace of $\mathcal{H}$. We say that $\mathcal{K}$ is invariant with respect to $\mathbf{G}$ if*

$$U_g(\mathcal{K}) \subseteq \mathcal{H} \qquad \forall g \in \mathbf{G} \tag{B.11}$$

Definition B.6 (Irreducible Representation) *Let $\{U_g \mid g \in \mathbf{G}\}$ a unitary representation of $\mathbf{G}$ on $\mathcal{H}$ and let $\mathcal{K} \subseteq \mathcal{H}$, be an invariant subspace. We say that $\{U_g\}$ is irreducible in $\mathcal{K}$ if there exists no proper subspace $\mathcal{V}$ of $\mathcal{K}$ that is invariant with respect to $\mathbf{G}$. A subspace carrying an irreducible representation is called irreducible subspace.*

Lemma B.1 (Subrepresentation) *Let $\{U_g\}$ be a unitary representation of $\mathbf{G}$ on $\mathcal{H}$ and $\mathcal{K}$ be an invariant subspace of $\mathcal{H}$. The restriction $\{U_g|_{\mathcal{K}}\}$ of $\{U_g\}$ on $\mathcal{K}$ is still a representation and it is called a subrepresentation of $\{U_g\}$.*

Finite groups and compact lie groups share a very relevant feature that is called complete reducibility, that is, any representation can be decomposed as a discrete sum of irreducible representations.

Theorem B.1 (Complete Reducibility) *Let $\mathbf{G}$ be a finite group or a compact Lie group and $\{U_g\}$ a unitary representation of $\mathbf{G}$ on a Hilbert space $\mathcal{H}$. Then there exists a discrete set of irreducible unitary subrepresentations $\{U_g|_{\mathcal{H}_k}\}$ such that*

$$U_g = \bigoplus_k U_g|_{\mathcal{H}_k}, \qquad \bigoplus_k \mathcal{H}_k = \mathcal{H} \tag{B.12}$$

Let $\{U_g\}$ be a reducible, as opposed to irreducible, representation of a group $\mathbf{G}$ on a Hilbert space $\mathcal{H}$. Suppose now that there are only two invariant subspaces $\mathcal{H}_1$ and $\mathcal{H}_2$ ($\mathcal{H} = \mathcal{H}_1 \oplus \mathcal{H}_2$) with dimensions n and m respectively. Then Theorem B.1 says that for all $g \in \mathbf{G}$, U_g can be written in a block diagonal form

$$U_g = \begin{pmatrix} U_g^{(1)} & 0 \\ 0 & U_g^{(2)} \end{pmatrix} \tag{B.13}$$

where $U_g^{(1)}$ is a $n \times n$ submatrix and $U_g^{(2)}$ is a $m \times m$ submatrix.

It can happen that in the decomposition $U = \bigoplus_k U_k$ (we omit the index of the group element) U_k is equivalent to U_l for some $k \neq l$; that being so, it is usual to rewrite the decomposition in this way:

$$U = \bigoplus_{\mu \in \mathrm{irrepS(U)}} \bigoplus_{i=1}^{m_\mu} U_{\mu,i} \tag{B.14}$$

where irrepS(U) represents the set of equivalence classes of irreducible representations contained in the decomposition of (U) and i labels different representations in the same class; m_μ is the number of different equivalent irreducible representations in the same class and it is called multiplicity. Likewise we write:

$$\bigoplus_k \mathcal{H}_k = \mathcal{H} \qquad \bigoplus_{\mu \in \mathrm{irrepS(U)}} \bigoplus_{i=1}^{m_\mu} \mathcal{H}_{\mu,i} \tag{B.15}$$

There is an isomorphism between the spaces $\bigoplus_{i=1}^{m_\mu} \mathcal{H}_{\mu,i}$ and $\mathcal{H}_\mu \otimes \mathbb{C}^{m_\mu}$ where $\mathcal{H}_\mu$ is an abstract Hilbert space of dimension d_μ $(\dim(\mathcal{H}_{\mu,\mathrm{i}}) = \dim(\mathcal{H}_{\mu,\mathrm{j}})$ for all i and j). If we denote with T^μ_{ij} the intertwiner connecting the equivalent representation $U_{\mu,i}$ and $U_{\mu,j}$ it can be written in the simple form $T^\mu_{ij} = I_{d_\mu} \otimes |i\rangle \langle j|$ where $\{|i\rangle\}$ is an o.n.b. for the space $\mathbb{C}^{m_\mu}$ and I_{d_μ} is the identity on the abstract space $\mathcal{H}_\mu$. Thanks to this isomorphism it is possible to rewrite the decomposition B.12 in this way

$$U_g = \bigoplus_{\mu \in \mathrm{irrepS(U)}} U_g^\mu \otimes I_{m_\mu}, \qquad \mathcal{H} = \bigoplus_{\mu \in \mathrm{irrepS(U)}} \mathcal{H}_\mu \otimes \mathbb{C}^{m_\mu} \tag{B.16}$$

It is customary to call $\mathcal{H}_\mu$ representation space and $\mathbb{C}^{m_\mu}$ multiplicity space.

B.2 Schur lemma and its applications

Lemma B.2 (Schur) *Let $\{U_g\}$ and $\{V_g\}$ two irreducible representations of the same group $\mathbf{G}$ on Hilbert spaces $\mathcal{H}$ and $\mathcal{K}$ respectively. Let $O : \mathcal{H} \to \mathcal{K}$ an operator such that such that $OU_g = V_g O$ for all $g \in \mathbf{G}$. If $\{U_g\}$ and $\{V_g\}$ are equivalent then $O = \lambda T$, where T is the isomorphism defined in Definition B.4 and $\lambda \in \mathbb{C}$. If $\{U_g\}$ and $\{V_g\}$ are not equivalent, then $O = 0$*

The Schur lemma is a powerful tool for inspecting the structure of operators commuting with a group representation.

Theorem B.2 (Characterization of the Commutant) *Let $\{U_g\}$ be a unitary representation of a group $\mathbf{G}$ and $O \in \mathcal{B}(\mathcal{H})$ an operator such that $[O, U_g] = 0$ for all $g \in \mathbf{G}$. Then*

$$O = \bigoplus_{\mu \in \mathrm{irrepS(U)}} I_{d_\mu} \otimes O_\mu \tag{B.17}$$

A typical example of operator in the commutant of a representation is the group average of an operator. Suppose that $\{U_g\}$ is a unitary representation of a finite group $\mathbf{G}$ on an Hilbert space $\mathcal{H}$ and $O \in \mathcal{B}(\mathcal{H})$. Then we can define

$$\overline{O} = \frac{1}{|\mathbf{G}|} \sum_{g \in \mathbf{G}} U_g O U_g^\dagger \tag{B.18}$$

where $|\mathbf{G}|$ is the cardinality of $\mathbf{G}$. Eq. (B.18) can be generalized to the case of Lie groups but to do this we need a preliminary definition

Definition B.7 (Invariant measure) *Let $\mathbf{G}$ be a Lie group. A measure $\mu_L(dg)$ on $\mathbf{G}$ is called left invariant if $\mu_L(gB) = \mu_L(B)$ for any $g \in \mathbf{G}$ and any region $B \subseteq \mathbf{G}$. A measure $\mu_r(dg)$ on $\mathbf{G}$ is called right invariant if $\mu_L(Bg) = \mu_L(B)$ for any $g \in \mathbf{G}$ and any region $B \subseteq \mathbf{G}$.*

Any Lie group can be endowed with a right invariant measure and a left invariant measure. When this to measures coincide the group is called unimodular; in this work we consider only unimodular group so we can talk about invariant measure without any misunderstanding. When the Lie group is compact (as it is always the case in this presentation) the invariant measure can be normalized in this way

$$\int_{\mathbf{G}} dg = 1. \tag{B.19}$$

Now we can define the group average for the case of (compact unimodular) Lie groups:

$$\overline{O} = \int_{\mathbf{G}} dg U_g O U_g^\dagger \tag{B.20}$$

As a consequence of Theorem B.2 we have

Theorem B.3 (Group average of an operator) *Let $\{U_g\}$ be a unitary representation of a finite (compact) group on an Hilbert space $\mathcal{H}$. Let O be an operator in $\mathcal{B}(\mathcal{H})$ and $\overline{O}$ its group average (as defined in Eq. B.18 for the finite case and in Eq. B.20 for the compact case). Then we have*

$$[\overline{O}, U_g] = 0 \qquad \forall g \in \mathbf{G} \tag{B.21}$$

$$\overline{O} = \bigoplus I_{d_\mu} \otimes \frac{\mathrm{Tr}_{\mathcal{H}_\mu}[P^\mu O P^\mu]}{d_\mu} \tag{B.22}$$

where P^μ is the projector on $\mathcal{H}_\mu \otimes \mathbb{C}^{m_\mu}$ and $\mathrm{Tr}_{\mathcal{H}_\mu}$ denotes the partial trace over $\mathcal{H}_\mu$.

B.3 Relevant decompositions

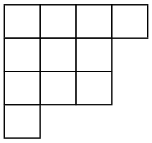
In this section we will give some results about the decomposition into irreducible representations for

B.3.1 The symmetric group S_n

$\mathbf{S}_n$ is the group of permutation of n objects. It can be proved that the number of inequivalent irreducible representation of $\mathbf{S}_n$ is given by the number of partition of n.[19] It is useful to associate each partition $\nu = (\nu_{\mathrm{i}})$ of n with a *Young diagram*. A Young diagram is a collection of boxed arranged in left aligned rows, the row lengths not increasing from the top to the bottom; as an example consider the partition

$$n = 11 \qquad \nu = (4, 3, 3, 1) \tag{B.23}$$

The corresponding Young diagram has the following shape



(B.24)

The usefulness of this pictorial representation will be more evident in the following section

B.3.2 Decomposition of $\mathbf{SU}(d)^{\otimes n}$

At the beginning of this chapter $\mathbf{SU}(d)$ was defined as the group of $d \times d$ unitary matrices U with determinant equal to 1. This definition identifies $\mathbf{SU}(d)$ with its smallest-dimensional faithful irreducible representation: this representation is usually called the *defining representation*. . Then $\mathbf{SU}(d)^{\otimes n}$ will denote the unitary representation $\{U^{\otimes n}\}$ over the Hilbert space $\mathcal{H}^{\otimes n}$ where $\dim(\mathcal{H}) = d$. In this section we will use both $\mathbf{SU}(d)^{\otimes n}$ and $U^{\otimes n}$ with the same meaning.

Let now consider the action of $\mathbf{S}_n$ on factorized vectors:

$$s \cdot (|\psi\rangle_1 \otimes \cdots \otimes |\psi\rangle_n) = |\psi\rangle_{s^{-1}(1)} \otimes \cdots \otimes |\psi\rangle_{s^{-1}(n)} \qquad s \in \mathbf{S}_n; \tag{B.25}$$

this action can be extended by linearity to the whole $\mathcal{H}^{\otimes n}$ leading to a representation of $\mathbf{S}_n$ over $\mathcal{H}^{\otimes n}$. This representation of $\mathbf{S}_n$ commutes with the representation $\mathbf{SU}(d)^{\otimes n}$ and it can be proved[20] that the irreducible subspaces of these two representations are the same. Each irreducible representation U_ν in the decomposition $U^{\otimes n} = \bigoplus_\nu U_\nu \otimes I_{m_\nu}$ is then in correspondence with a Young diagram ν.

From a Young diagram one can obtain a *Young tableaux* filling the empty boxes with the integers numbers from 1 to n; a *standard Young tableau* is a tableau in which the numbers in each row grow from left to right and the numbers in each column grow from top to bottom e.g.

$$\begin{array}{|c|c|c|c|}\hline 1 & 2 & 5 & 7 \\ \hline 3 & 4 & 8 \\ \cline{1-3} 6 & 9 & 11 \\ \cline{1-3} 10 \\ \cline{1-1} \end{array} \qquad \begin{array}{|c|c|c|c|}\hline 1 & 3 & 4 & 7 \\ \hline 2 & 5 & 6 \\ \cline{1-3} 8 & 10 & 11 \\ \cline{1-3} 9 \\ \cline{1-1} \end{array}$$

[19] A partition of an integer n is a way of writing n as a sum of positive integers.

[20] This result is the *Schur-Weyl duality*. The aim of this section is to introduce (without claiming to be rigorous) some consequence of this theorem that are exploited for proving many results of Quantum Information Theory.

Given an irreducible representation U_ν in the decomposition of $\mathbf{SU}(d)^{\otimes n}$, the dimension m_ν of the corresponding multiplicity space is given by the number of admissible standard Young tableaux associated to the Young diagram corresponding to U_ν.

The following combinatorial procedure gives the dimension of $\mathcal{H}_\nu$:

- Given the Young diagram ν number the rows and the columns with integer numbers $1, 2, \ldots, n$ from top to bottom for the rows and $1, 2, \ldots, m$ from left to right for the columns;
- associate each box b of ν with the expression $\frac{l_\mathrm{b}}{h_\mathrm{b}}$ where:

  $$l_\mathrm{b} = d + j - i \tag{B.26}$$

  $d = \dim(\mathcal{H})$, j is the column which the box b belongs to and i is the row which i belongs to;

  $$h_\mathrm{b} = 1 + r + s \tag{B.27}$$

  r is the number of boxes to the right of b in the same row and s is the number of boxes below it in the same column;
- Finally we have

  $$\dim(\mathcal{H}_\nu) = \prod_\mathrm{b} \frac{l_\mathrm{b}}{h_\mathrm{b}} \tag{B.28}$$

We notice that when the number of rows is greater than n there will be at least one box b for which we have $h_\mathrm{b} = 0$; such diagrams correspond to the mapping $g \rightarrow 0$ for all $g \in \mathbf{SU}(d)$ and can be discarded in the decomposition of $\mathbf{SU}(d)^{\otimes n}$.

Since each irreducible representation of $SU(d)$ appears in the decomposition of $U^{\otimes n}$ for some n, then it is possible to establish the correspondence

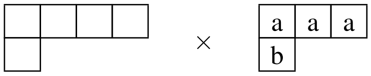
Irreducible representations of $\mathbf{SU}(d)$ ↔ Young diagrams
with at most $d-1$ columns

A relevant example is the defining representation which corresponds to the Young diagram made of a single box □

This 1 to 1 correspondence allows to deal with decomposition of tensor product of irreducible representation of $\mathbf{SU}(d)$ in a diagrammatic way. If $\{U_\alpha\}$ and $\{U_\beta\}$ are two irreducible representations of $\mathbf{SU}(d)$ we associate their tensor product representation $\{U_\alpha \otimes U_\beta\}$ with the product $\alpha \times \beta$ of the corresponding Young diagrams α and β. The following procedure provides the expansion of the product of two Young diagrams as a sum of Young diagrams.

Expansion algorithm for Young diagram

- Write the product of two Young diagrams α and β labelling successive rows of β with indexes $a, b, \dots$ as follows:



(B.29)

- At each stage add boxes $\boxed{\text{a}}\dots\boxed{\text{a}}, \boxed{\text{b}}\dots\boxed{\text{b}}$ from β to α one at a time checking that:
  - the created diagrams ν have no more than d columns
  - boxes with the same label must not appear in the same column
  - when the the created tableaux is read from left to right and from top to bottom the sequence of letters $a, b, \dots$ must be such that any any point of the sequence the number of b's occurred is not bigger than the number of a's, the number of c's occurred is not bigger than the number of b's etc.
- Two diagrams ν, μ of the shame shape are considered different only if the labeling is different.

Finally we can write the expansion

$$\alpha \times \beta = \sum_\nu \sum_i \nu_i \tag{B.30}$$

where ν labels diagram with different shape and i labels different diagrams with the same shape. It is worth noting that the product of Young diagrams, as defined by means of the previous expansion, enjoys the following properties:

$$\alpha \times \beta = \beta \times \alpha \qquad (\alpha \times \beta) \times \gamma = \alpha \times (\beta \times \gamma) \tag{B.31}$$

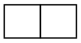
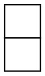
Each diagram ν_i in the expansion (B.30) corresponds to an irreducible representation in the decomposition $U_\alpha \otimes U_\beta = \sum_\nu \sum_i U_{\nu,i}$; diagrams with the same shape represent equivalent representations and the number of diagrams with the same shape but with different labeling gives the dimension m_ν of the multiplicity space. Finally we have the following correspondence

$$\alpha \times \beta = \sum_\nu \sum_i \nu_i \quad \leftrightarrow \quad U_\alpha \otimes U_\beta = \sum_\nu U_\nu \otimes \mathbb{C}^{m_\nu} \tag{B.32}$$

The following examples will clarify the previous discussion

B.3.3 $\mathbf{U} \otimes \mathbf{U}$

The admissible Young diagrams for $\mathbf{SU}(d)^{\otimes 2}$ are

$\nu_+ =$  $\nu_=$  with $\dim(\mathcal{H}_+) = \frac{d(d+1)}{2}$ and $\dim(\mathcal{H}_-) = \frac{d(d-1)}{2}$. The admissible standard Young tableaux are

$$\nu_+ = \begin{array}{|c|c|}\hline 1 & 2 \\ \hline \end{array} \qquad \nu_- = \begin{array}{|c|}\hline 1 \\ \hline 2 \\ \hline \end{array}$$

thus we have $m_+ = m_- = 1$ and the decomposition becomes:

$$U \otimes U = U_+ \otimes U_- \qquad U_+ \in \mathcal{B}(\mathcal{H}_+), \;\; U_- \in \mathcal{B}(\mathcal{H}_-). \tag{B.33}$$

$\mathcal{H}_+$ and $\mathcal{H}_-$ can be proved to be the symmetric and the anti-symmetric subspace of $\mathcal{H} \otimes \mathcal{H}$ respectively. If $\{|i\rangle \;\; i = 1, \ldots, d\}$ is a basis for $\mathcal{H}$ it is possible to find a basis for $\mathcal{H}_+$ and $\mathcal{H}_-$; we have

$$\mathcal{H}_+ = \mathrm{Span}\left\{|n_+\rangle := \frac{1}{\sqrt{2}}(|i\rangle\,|j\rangle + |j\rangle\,|i\rangle), \; i, j = 1 \ldots d\right\} \tag{B.34}$$

$$\mathcal{H}_- = \mathrm{Span}\left\{|n_-\rangle := \frac{1}{\sqrt{2}}(|i\rangle\,|j\rangle - |j\rangle\,|i\rangle), \; i, j = 1 \ldots d\right\}. \tag{B.35}$$

Exploiting Eqs. (B.34, B.35) it is easy to check that $\mathcal{H}_+$ and $\mathcal{H}_-$ are invariant subspaces of $\mathbf{SU}(d)^{\otimes 2}$. We introduce

$$P^+ = \sum_n |n_+\rangle\,\langle n_+| \qquad P^- = \sum_n |n_-\rangle\,\langle n_-| \tag{B.36}$$

P^+ is the projector on the symmetric subspace and P^- is the projector on the antisymmetric subspace.

We notice that the expansion of the product □ × □ would lead to the same decomposition for $U \otimes U$.

B.3.4 $\mathbf{U} \otimes \mathbf{U} \otimes \mathbf{U}$

The admissible Young diagrams for $\mathbf{SU}(d)^{\otimes 3}$ are

$$\alpha = \begin{array}{|c|c|}\hline & \\ \hline & \multicolumn{1}{c}{} \\ \cline{1-1} \end{array} \qquad \beta = \begin{array}{|c|c|c|}\hline & & \\ \hline \end{array} \qquad \gamma = \begin{array}{|c|}\hline \\ \hline \\ \hline \\ \hline \end{array}.$$

with $\dim(\mathcal{H}_\alpha) = \frac{d(d+1)(d-1)}{3}$, $\dim(\mathcal{H}_\beta) = \frac{d(d+1)(d+2)}{6}$ $\dim(\mathcal{H}_\gamma) = \frac{d(d-1)(d-2)}{6}$. We notice that for $d = 2$ $\dim(\mathcal{H}_\gamma) = 0$ and the representation labelled by γ does not appear in the decomposition. The admissible standard Young tableaux are

$$\alpha_1 = \begin{array}{|c|c|}\hline 1 & 2 \\ \hline 3 & \multicolumn{1}{c}{} \\ \cline{1-1} \end{array} \qquad \alpha_2 = \begin{array}{|c|c|}\hline 1 & 3 \\ \hline 2 & \multicolumn{1}{c}{} \\ \cline{1-1} \end{array} \qquad \beta = \begin{array}{|c|c|c|}\hline 1 & 2 & 3 \\ \hline \end{array} \qquad \gamma = \begin{array}{|c|}\hline 1 \\ \hline 2 \\ \hline 3 \\ \hline \end{array}$$

thus we have $m_\alpha = 2$, $m_\beta = m_\gamma = 1$ and the decomposition becomes:

$$\begin{aligned}&U \otimes U \otimes U = U_\alpha \otimes I_{m_\alpha} \oplus U_\beta \oplus U_\gamma \\ &U_\alpha \in \mathcal{B}(\mathcal{H}_\alpha),\;\; U_\beta \in \mathcal{B}(\mathcal{H}_\beta),\;\; U_\gamma \in \mathcal{B}(\mathcal{H}_\gamma),\;\; I_{m_\alpha} \in \mathbb{C}^{m_\alpha} = \mathbb{C}^2.\end{aligned} \tag{B.37}$$

An equivalent way to decompose $U \otimes U \otimes U$ is through the expansion of the product $\square \times \square \times \square$

B.3.5 $\mathbf{U} \otimes \mathbf{U}^*$

Let us start with a preliminary definition

Definition B.8 (conjugate representation) *Let $\{U_g\}$ be a unitary representation of a group $\mathbf{G}$. Then it is possible to define its conjugate representation $\{U_g^*\}$ in this way:*

$$U_g^* = U_{g^{-1}}^T \qquad \forall g \in \mathbf{G} \tag{B.38}$$

It is straightforward to notice that the conjugate of the defining representation U of $\mathbf{SU}(d)$ is the one formed by the complex conjugate matrices U^*. The Young diagram corresponding to the representation U^* is the one corresponding to a column of $d-1$ boxes

$$U^* \quad \leftrightarrow \quad \left.\begin{array}{c}\square\\ \square\\ \vdots\\ \square\end{array}\right\} d-1 \text{ boxes} \tag{B.39}$$

It is worth noting that for $d = 2$ both U and U^* are represented by the Young diagram made of a single box. This agree with the fact that the defining representation U of $\mathbf{SU}(2)$ and its conjugate U^* are equivalent; Indeed for all $U \in \mathbf{SU}(2)$ we have

$$CU = U^*C \qquad C = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \tag{B.40}$$

The easiest way to the decompose of $U \otimes U^*$ is exploiting the Young diagrams formalism and the expansion algorithm:

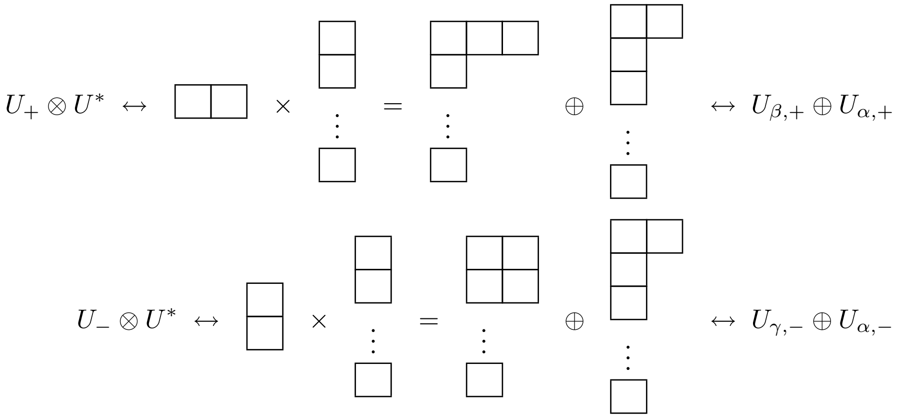
$$U \otimes U^* \quad \leftrightarrow \quad \square \quad \times \quad \begin{array}{c}\square\\ \square\\ \vdots\\ \square\end{array} \quad = \quad \begin{array}{c}\square\\ \square\\ \square\\ \vdots\\ \square\end{array} \quad \oplus \quad \begin{array}{ll}\square & \square\\ \square & \\ \vdots & \\ \square & \end{array} \quad \leftrightarrow \quad U_p \oplus U_q \tag{B.41}$$

where $U_p \in \mathcal{B}(\mathcal{H}_p)$ and $U_q \in \mathcal{B}(\mathcal{H}_q)$ $\dim(\mathcal{H}_p) = 1$, $\dim(\mathcal{H}_q) = d^2 - 1$. An explicit form for the projectors on $\mathcal{H}_p$ and $\mathcal{H}_q$ can be given:

$$P^p = d^{-1}|I\rangle\!\rangle\langle\!\langle I| \qquad P^q = I \otimes I - P^p. \tag{B.42}$$

B.3.6 $\mathbf{U} \otimes \mathbf{U} \otimes \mathbf{U}^*$

We can decompose the representation $U \otimes U \otimes U$ as follows. First, as we showed previously, $U \otimes U$ can be decomposed as $U_+ \oplus U_-$ and so we have $U \otimes U \otimes U^* = (U_+ \oplus U_-) \otimes U^* = (U_+ \otimes U^*) \oplus (U_- \oplus U^*)$. We now further decompose $U_+ \otimes U^*$ and $U_- \oplus U^*$:


$U_+ \otimes U^* \leftrightarrow$
$\times$
$=$
$\oplus$
$\leftrightarrow U_{\beta,+} \oplus U_{\alpha,+}$
$U_- \otimes U^* \leftrightarrow$
$\times$
$=$
$\oplus$
$\leftrightarrow U_{\gamma,-} \oplus U_{\alpha,-}$

(B.43)

Then the following decomposition holds:

$$U \otimes U \otimes U^* = U_{\alpha,+} \oplus U_{\alpha,-} \oplus U_{\gamma,-} \oplus U_{\beta,+} \tag{B.44}$$

$$\dim(\mathcal{H}_{\alpha,+}) = \dim(\mathcal{H}_{\alpha,-}) = d,$$
$$\dim(\mathcal{H}_{\beta,+}) = d\frac{d^2+d-2}{2}, \qquad \dim(\mathcal{H}_{\gamma,-}) = d\frac{d^2-d-2}{2}$$

We notice that for $d = 2$ the subspace $\dim(\mathcal{H}_{\gamma,-}) = 0$ Since $U_{\alpha,+}$ and $U_{\alpha,-}$ are equivalent representations the decomposition (B.44) can be rewritten as

$$U \otimes U \otimes U^* = U_\alpha \otimes I_{m\alpha} \oplus U_\gamma \oplus U_\beta \qquad I_{m_\alpha} \in \mathcal{B}(\mathbb{C}^2) \tag{B.45}$$

where we relabeled $\mathcal{H}_{\beta,+} = \mathcal{H}_\beta$, $\mathcal{H}_{\gamma,-} = \mathcal{H}_\gamma$ and $\mathcal{H}_{\alpha,+} \oplus \mathcal{H}_{\alpha,-} = \mathcal{H}_\alpha \otimes \mathbb{C}^2$. We now provide two basis for $\mathcal{H}_{\alpha,+}$ and $\mathcal{H}_{\alpha,-}$

$$\mathcal{H}_{\alpha,+} = \text{Span}\left\{ |k_{\alpha,+}\rangle := \frac{1}{\sqrt{2(d+1)}} \left(|I\rangle\!\rangle_{02} \, |k\rangle_1 + |I\rangle\!\rangle_{12} \, |k\rangle_0 \right) \right\}$$
$$\mathcal{H}_{\alpha,-} = \text{Span}\left\{ |k_{\alpha,-}\rangle := \frac{1}{\sqrt{2(d-1)}} \left(|I\rangle\!\rangle_{02} \, |k\rangle_1 - |I\rangle\!\rangle_{12} \, |k\rangle_0 \right) \right\}. \tag{B.46}$$

where we introduced the labeling $\mathcal{H} \otimes \mathcal{H} \otimes \mathcal{H} := \mathcal{H}_0 \otimes \mathcal{H}_1 \otimes \mathcal{H}_2$ and $|k\rangle_i$ means $|k\rangle \in \mathcal{H}_i$. We notice the properties

$$\mathsf{S} \, |k_{\alpha,+}\rangle = |k_{\alpha,+}\rangle \qquad \mathsf{S} \, |k_{\alpha,-}\rangle = - \, |k_{\alpha,-}\rangle \tag{B.47}$$

where S is the swap operator $\mathsf{S}\left|\psi\right\rangle_0\left|\phi\right\rangle_1=\left|\phi\right\rangle_0\left|\psi\right\rangle_1$ In terms of these two basis the isomorphism between $\mathcal{H}_{\alpha,+}$ and $\mathcal{H}_{\alpha,-}$ has the following form:

$$T^{\alpha,+,-}=\sum_k\left|k_{\alpha,+}\right\rangle\left\langle k_{\alpha,-}\right|. \tag{B.48}$$

From Eqs. (B.43) and (B.46) we can derive the expression for the projectors on $\mathcal{H}_\beta$ and $\mathcal{H}_\gamma$:

$$P^\beta=P^+\otimes I_2-T^{\alpha,+,+} \qquad P^\gamma=P^-\otimes I_2-T^{\alpha,-,-} \tag{B.49}$$

$$T^{\alpha,+,+}=\sum_k\left|k_{\alpha,+}\right\rangle\left\langle k_{\alpha,+}\right| \qquad T^{\alpha,-,-}=\sum_k\left|k_{\alpha,-}\right\rangle\left\langle k_{\alpha,-}\right|. \tag{B.50}$$

$T^{\alpha,+,+}$ is the projector on $\mathcal{H}_{\alpha,+}$ and $T^{\alpha,-,-}$ is the projector on $\mathcal{H}_{\alpha,-}$.

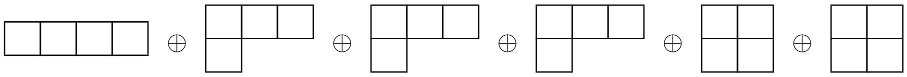
Exploiting Theorem B.2 any operator O satisfying the commutation $[O,U\otimes U\otimes U^*]$ can be decomposed as

$$O=\sum_{\nu\in\mathsf{S}}\sum_{i,j=\pm}T^{\nu,i,j}o_\nu^{i,j} \qquad o_\nu^{i,j}\in\mathbb{R} \tag{B.51}$$

where $\mathsf{S}=\{\alpha,\beta,\gamma\}$, $T^{\beta,+,-}=T^{\beta,-,+}=T^{\beta,-,-}=0$, $T^{\gamma,+,-}=T^{\gamma,-,+}=T^{\gamma,+,+}=0$, $T^{\beta,+,+}=P^\beta$ and $T^{\beta,+,+}=P^\beta$.

B.3.7 $\mathbf{U}\otimes\mathbf{U}\otimes\mathbf{U}\otimes\mathbf{U}$ (2-dimensional case)

Expanding the product □ × □ × □ × □ × □ leads to the decomposition



$$U^{\otimes 4}=U_a\oplus U_b\otimes I_{m_b}\oplus U_c\otimes I_{m_c} \tag{B.52}$$

$$\dim(\mathcal{H}_a)=5 \quad \dim(\mathcal{H}_b)=3 \quad \dim(\mathcal{H}_c)=1$$
$$\mathbb{C}^{m_b}=\mathbb{C}^3 \quad \mathbb{C}^{m_c}=\mathbb{C}^2$$

Since we are considering the case $d=2$, U and U^* are equivalent and the decomposition (B.52) can be generalized to the cases in which one or more U is replaced with U^*. For example we have:

$$U^*\otimes U\otimes U\otimes U=(C\otimes I^{\otimes 3})\,U_a\oplus U_b\otimes I_{m_b}\oplus U_c\otimes I_{m_c}\,(C\otimes I^{\otimes 3}) \tag{B.53}$$

where C was defined in Eq. (B.40).

References

[1] A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory*, North Holland, Amsterdam (1982)

[2] M. A. Nielsen, I. L. Chuang, *Quantum Computation and Quantum Information*, Cambridge University Press, Cambridge (2000)

[3] C. W. Helstrom, *Quantum Detection and Estimation Theory*, Academic Press, New York (1976)

[4] A. Acín, Phys. Rev. Lett. **87**, 177901 (2001)

[5] G. M. D'Ariano, P. Lo Presti, and M. G. A. Paris, Phys. Rev. Lett. **87**, 270404 (2001)

[6] M. F. Sacchi J. Opt. B **7**, S333 (2005)

[7] G. Chiribella, G. M. D'Ariano, P. Perinotti, Phys. Rev. Lett. **101**, 180501 (2008)

[8] A. W. Harrow, A. Hassidim, D. W. Leung, J. Watrous, Phys. Rev. A **81**, 032339 (2010)

[9] M. A. Nielsen, I. L. Chuang Phys. Rev. Lett. **79**, 321 (1997)

[10] S. F. Huelga, J. A. Vaccaro, A. Chefles, and M. B. Plenio, Phys. Rev. A **63**, 042303 (2001)

[11] S. D. Bartlett, W. J. Munro Phys. Rev. Lett. **90**, 117901 (2003)

[12] Y.-F. Huang, X.-F. Ren, Y.-S. Zhang, L.-M. Duan, G.-C. Guo, Phys. Rev. Lett. **93**, 240501 (2004)

[13] Y. S. Weinstein, T. F. Havel, J. Emerson, N. Boulant, M. Saraceno, S. Lloyd, D. G. Cory, J. Chem. Phys. **121(13)**, 6117-6133 (2004)

[14] J. L. O'Brien, G. J. Pryde, A. Gilchrist, D. F. V. James, N. K. Langford, T. C. Ralph, A. G. White, Phys. Rev. Lett. **93**, 080502 (2004)

[15] G. Gutoski and J. Watrous, Proc. of the 39th Annual ACM Symposium on Theory of Computation, 565 (2007).

[16] Proc. R. Soc. Lond. A **439**, 553-558 (1992).

[17] Grover L.K. Proceedings of the 28th Annual ACM Symposium on the Theory of Computing, 212 (1996)

[18] P. Shor, SIAM Rev. **41**, pp. 303-332 (1999).

[19] H. P. Yuen, quant-ph/0207089.

[20] G. M. D'Ariano, D. Kretschmann, D. M. Schlingemann, R. F. Werner, Phys. Rev. A **76** 032328 (2007).

[21] S. Pirandola, S. Mancini, S. Lloyd, and S. L. Braunstein, Nature Physics **4**, 726 - 730 (2008).

[22] G. Chiribella, G. M. D'Ariano, P. Perinotti, Phys. Rev. Lett. **101**, 060401 (2008)

[23] G. Chiribella, G. M. D'Ariano, P. Perinotti, Phys. Rev. A **80**, 022339 (2009)

[24] A. Bisio, G. Chiribella, G. M. D'Ariano, S. Facchini, and P. Perinotti, Phys. Rev. Lett. **102**, 010404 (2009).

[25] A. Bisio, G. Chiribella, G. M. D'Ariano, S. Facchini, and P. Perinotti, IEEE Journal of Selected Topics in Quantum Electronics **15** 1646 (2009)

[26] G. Chiribella, G. M. D'Ariano, P. Perinotti, Phys. Rev. Lett. **101**, 180504 (2008)

[27] A. Bisio, G. Chiribella, G. M. D'Ariano, S. Facchini, P. Perinotti Phys. Rev. A **81**, 032324 (2010)

[28] A. Bisio, G. Chiribella, G. M. D'Ariano, P. Perinotti, Phys. Rev. A **83**, 022325 (2011)

[29] A. Bisio, G. Chiribella, G. M. D'Ariano, P. Perinotti, Phys. Rev. A **82**, 062305 (2010).

[30] A. Bisio, G. M. D'Ariano, P. Perinotti, M. Sedlák, Physics Letters A **375**, 3425-3434 (2011).

[31] A. Bisio, G. M. D'Ariano, P. Perinotti, M. Sedlák, (accepted in Phys. Rev. A). arXiv:1103.5709

[32] M. Raginsky Phys. Lett. A **290**, 11 (2001)

[33] J. de Pillis, Linear Transformations Which Preserve Hermitian and Positive Semidefinite Operators, Pacific J. of Math. **23**, 129 (1967)

[34] M.-D. Choi, Lin. Alg. and Appl. **10**, 285 (1975)

[35] A. Jamiołkowski Rep. Mod. Phys. **3**, 275 (1972)

[36] W. F. Stinespring, Proc. Amer. Math. Soc. **6**, 211 (1955)

[37] G. Chiribella, G. M. D'Ariano, P. Perinotti J. Math. Phys. **50**, 042101 (2009)

[38] M. Ozawa. J. Math. Phys. **25**, 79 (1984)

[39] T. Eggeling, D. Schlingemann, R. F. Werner, Europhys. Lett. **57**, 782-788 (2002).

[40] M. Piani, M. Horodecki, P. Horodecki, R. Horodecki Phys. Rev. A **74**, 012305 (2006)

[41] M. Ziman, Phys. Rev. A **77**, 062112 (2008).

[42] D. T. Smithey, M. Beck, M. G. Raymer, and A. Faridani, Phys. Rev. Lett. **70**, 1244 (1993).

[43] K. Vogel and H. Risken, Phys. Rev. A **40**, 2847 (1989).

[44] G. M. D'Ariano, C. Macchiavello, and M. G. A. Paris, Phys. Rev. A **50**, 4298 (1994)

[45] P. Busch, Int. J. Theor. Phys. **30**, 1217 (1991).

[46] G. M. D'Ariano and P. Perinotti, Phys. Rev. Lett. **98**, 020403 (2007).

[47] A. J. Scott, Phys. A **39**, 13507 (2006).

[48] G. M. D'Ariano, P. Lo Presti, Phys. Rev. Lett. **86**, 4195 (2001).

[49] W. Dür and J. I. Cirac, Phys. Rev. A **64**, 012317 (2001).

[50] R. J. Duffin, A. C. Schaeffer, Trans. Am. Math. Soc. **72**, 341 (1952).

[51] P. G. Casazza, Taiw. J. Math. **4**, 129 (2000)

[52] G .Casella, R. L. Berger, *Statistical Inference*, Duxbury Press (2001).

[53] G. M. D'Ariano. P. Perinotti, M. F. Sacchi, J. Opt.B: Quantum and Semicl. Optics **6**, S487 (2004)

[54] A. J. Scott, J. Phys. A **39**, 13507 (2006)

[55] G. Chiribella, G. M.D'Ariano, D. M. Schlingemann, Phys. Rev. Lett. **98**, 020403 (2007)

[56] P. Walther, A. Zeilinger, Phys. Rev. A **72**, 010302(R) (2005)

[57] W. K. Wootters, W.H.Zurek, Nature **299**, 802 (1982)

[58] V. Buzek, M. Hillery, Physics World **14**, 25 (2001).

[59] R. Werner, Phys. Rev. A **58**, 1827 (1998)

[60] J. Fiurasek, R. Filip, N. J. Cerf Quant. Inform. Comp. **5**, 583 (2005).

[61] V. Scarani, S. Iblisdir, N. Gisin, and A. Acín, Rev. Mod. Phys. **77**, 1225 (2005)

[62] C. H. Bennett, G. Brassard, Proceedings IEEE Int. Conf. on Computers, Systems and Signal Processing, Bangalore, India (IEEE New York, 1984), pp. 175-179

[63] R. Zhao, Y. O. Dudin, S. D. Jenkins, C. J. Campbell, D. N. Matsukevich, T. A. B. Kennedy, A. Kuzmich, Nature Physics **5**, 100 (2009)

[64] A. I. Lvovsky, B. C. Sanders, W. Tittel Nature Photonics **3**, 706 - 714 (2009)

[65] B. Julsgaard, J. Sherson, J. I. Cirac, J. Fiurasek, E. S. Polzik Nature **432**, 482 - 486 (2004)

[66] G. Vidal, L. Masanes, J. I. Cirac Phys. Rev. Lett. **88**, 047905 (2002)

[67] M. Ziman, V. Buzek Phys. Rev. A **72**, 022343 (2005)

[68] G. M. D'Ariano, P. Perinotti Phys. Rev. Lett. **94**, 090401 (2005)

[69] M. Micuda, M. Jezek, M. Dusek, J. Fiurasek Phys. Rev. A **78**, 062311 (2008)

[70] G. Chiribella, G. M. D'Ariano, M. F. Sacchi, Phys. Rev. A **72** 042338 (2005)

[71] G. Chiribella, G. M. D'Ariano, P. Perinotti, M. F. Sacchi, Phys. Rev. Lett. **93** 18053 (2004)

[72] V. Buzek, R. Derka, S. Massar, Phys. Rev. Lett. **82**, 2207 (1999)

[73] S. D. Bartlett, T. Rudolph, R. W. Spekkens, P. S. Turner, New J. Phys. **11**, 063013 (2009)

[74] G. Chiribella, G. M. D'Ariano, P. Perinotti, M. F. Sacchi, Phys. Rev. A **70**, 062105 (2004)

[75] W. Heisenberg, Zeitsch. Phys. **43**, 172 (1927).

[76] M. O. Scully, B.-G. Englert, and H. Walther, Nature **351**, 111 (1991).

[77] C. A. Fuchs and A. Peres, Phys. Rev. A **53**, 2038 (1996).

[78] K Banaszek, Phys. Rev. Lett. **86**, 1366 (2001) .

[79] M. Ozawa, Ann. Phys. **311**, 350 (2004).

[80] M. F. Sacchi, Phys. Rev. Lett. **96**, 220502 (2006).

[81] F. Sciarrino, M. Ricci, F. De Martini, R. Filip, and L. Mišta Jr., Phys. Rev. Lett. **96**, 020408 (2006).

[82] L. Maccone, Phys. Rev. A **73**, 042307 (2006).

[83] D. Kretschmann, D. Schlingemann and R. F. Werner, IEEE Trans. Inf. Theory **4**, 1708 (2008).

[84] F. Buscemi, M. Hayashi, M. Horodecki, Phys. Rev. Lett. **100**, 210504 (2008).

[85] K. Boström, T. Felbinger, Phys. Rev. Lett. **89**, 187902 (2002).

[86] M. Lucamarini, S. Mancini, Phys. Rev. Lett. **94**, 14051 (2005).

[87] A. Ferraro, M. Galbiati, M. G. A. Paris, J. Phys. A **39**, L219-L228 (2006).

[88] S. Boyd, L. Vanderberghe, *Convex Optimization* Cambridge University Press, Cambridge (2004)

[89] C. Zalka Phys. Rev. A **60**, 2746 (1999)

[90] G.Wang, M. Ying. Phys. Rev. A **73**, 042301 (2006)

[91] V. P. Belavkin, G. M. D'Ariano, M. Raginsky J. Math. Phys. **46**, 062106 (2005)

[92] W. Fulton and J. Harris, *Representation theory: a first course*, Springer, (1996)

[93] H. F. Jones, *Groups, Representations and Physics* Taylor and Francis (1990)

[94] W. Fulton, *Young tableaux : with applications to representation theory and geometry* Cambridge University Press, Cambridge (1997)

[95] A. O. Barut, R. Raczka, *Theory of group representations and applications* World Scientific, Singapore (1986)

[96] M. Grant, S. Boyd, http://cvxr.com/cvx/

[97] J. Watrous, private communication (2010)

Giulio Chiribella received the M.Sc. degree and the PhD in physics from the University of Pavia in 2003 and 2006 respectively. From 2006 to 2009 he was post-doc fellow of Dipartimento di Fisica "A. Volta" of Università degli Studi di Pavia. Since 2009 he is senior postdoc at Perimeter Institute of Theoretical Physics. He was one of the founders of the theory of quantum combs. Recently he proposed an operational axiomatization of Quantum Mechanics, which has received much interest at numerous international conferences. His research interests are: Quantum information processing, quantum optics, quantum estimation, foundations of quantum Mechanics, algebraic and group theoretical methods.

Alessandro Bisio received the M.Sc. degree and the PhD in physics from the University of Pavia in 2007 and 2010 respectively. Since 2010 he is a post-doc fellow of Dipartimento di Fisica "A. Volta" of Università degli Studi di Pavia. His research interests are: Quantum Information Processing, Quantum Optics, algebraic and group theoretical methods and foundations of Quantum Mechanics.

Giacomo Mauro D'Ariano is full professor of Quantum Information and Quantum Optics at the University of Pavia. Fellow of the Optical Society of America, Member of the Lombard Academy of Science and Letters, member of the Center for Photonic Communication and Computing of the Department of Electrical and Computer Engineering of Northwestern University (Evanston IL), with which he regularly collaborates since 1994. In Pavia he created the research group QUIT (Quantum Information Theory), which is scientifically very active at the international level. He conceived and developed the method of quantum homodyne tomography as the first quantitative technique to determine experimentally the state of radiation, technique now very popular. He then generalized the method to arbitrary quantum system and arbitrary ensemble average, achieving a universal measurement method. He conceived and developed the first experimental technique for the complete quantum characterization of a measuring apparatus or of the transformation of a device. He introduced a novel theoretical method to deal with covariant measurements and transformations, which has recently lead him and his research team to the solution of the long-standing problems of phase-estimation and broadcasting of mixed states of qubits. He was one of the founders of the theory of quantum combs. Recently he proposed an operational axiomatization of Quantum Mechanics, which has received much interest at numerous international conferences.

Paolo Perinotti received the M.Sc. degree in physics from the University of Pavia in 1999 and the PhD in physics from the University of Milan in 2002. From 2002 to 2006 he was INFM (Istituto Nazionale di Fisica della Materia) post-doc fellow. From 2006 to 2011 he was post-doc fellow of Dipartimento di Fisica "A. Volta" of Università degli Studi di Pavia. He is presently Research Associate at University of Pavia. He was one of the founders of the theory of quantum combs. Recently he proposed an operational axiomatization of Quantum Mechanics, which has received much interest at numerous international conferences. His research interests are: Quantum Information and Quantum Mechanics of Measurements and Open Systems, Quantum estimation, discrimination and tomography of states and devices, and logical foundations of quantum mechanics.