Quantum channels and memory effects

Filippo Caruso
QSTAR, LENS, Dipartimento di Fisica e Astronomia, University of Florence, I-50019 Sesto Fiorentino, Italy and Institut für Theoretische Physik, Universität Ulm, D-89069 Ulm, Germany

Vittorio Giovannetti
NEST, Scuola Normale Superiore e Istituto Nanoscienze-CNR, Piazza dei Cavalieri 7, I-56126 Pisa, Italy

Cosmo Lupo
MIT, Research Laboratory of Electronics, 77 Massachusetts Avenue, Cambridge MA 02139, USA, and School of Science and Technology, University of Camerino, I-62032 Camerino, Italy

Stefano Mancini
School of Science and Technology, University of Camerino, I-62032 Camerino, Italy and INFN-Sezione di Perugia, Via A. Pascoli, I-06123 Perugia, Italy

Any physical process can be represented as a quantum channel mapping an initial state into a final state. Hence it can be characterized from the point of view of communication theory, i.e., in terms of its ability in transferring information. Quantum information provides the theoretical framework and the proper mathematical tools to accomplish this. In this context the notion of codes and communication capacities have been introduced by generalizing them from the classical Shannon theory of information transmission and error correction. The underlying assumption of this approach is to consider the channel not as acting on a single system, but on sequences of systems, which, when properly initialized allows one to overcome the noisy effects induced by the physical process under consideration. While most of the work produced so far have been focused on the case where a given channel transformation acts identically and independently on the various elements of the sequence (memoryless configuration in jargon), correlated error models appears to be a more realistic way to approach the problem. A slightly different, yet conceptually related, notion of correlated errors applies to a single quantum system which evolves continuously in time under the influence of an external disturbance which acts on it in a non-Markovian fashion. This leads to the study of memory effects in quantum channels: a fertile ground where interesting novel phenomena emerge at the intersection of quantum information theory with other branches of physics. We survey the field of quantum channels theory on a wide scenario that embraces also these specific and complex settings.

PACS numbers: 03.67.-a, 89.70.-a, 02.50.-r

Contents

I. Introduction 2

II. Quantum channels: basic definitions and properties 4
   A. CPTP transformations 5
   B. Composition rules and structural properties 5
   C. Stinespring representation, Kraus representation, and Choi-Jamiolkowski isomorphism 6
   D. Heisenberg picture: the dual channel 6
   E. cq- and qc-channels 7
   F. Entanglement Breaking and PPT channels 7
   G. Complementary channels and Degradability 8
   H. Causal, Localizable, LOCC and Separable channels 8
   I. Examples 9
      1. Qubit channels 9
      2. Erasure channels 10

III. From Memoryless to Memory Quantum Channels 14
   A. Memoryless quantum channels 15
      1. Compound and averaged quantum channels 15
   B. Non-anticipatory memory quantum channels 16
      1. Input-output and entanglement fidelity of a quantum channel 16
      2. Distance measures for channels 16
   C. Quasi-local algebras approach 17
   D. Taxonomy of Non-Anticipatory Quantum Memory Channels 18
      1. Localizable memory quantum channels 18
      2. Finite-memory channels 18
      3. Perfect memory channels 19
      4. Markovian channels 20
      5. Fixed–point channels 21
      6. Indecomposable and forgetful channels 22
      7. Decaying input memory cq-channels 22
      8. Long-term memory channels 22

IV. Quantum Codes 23
   A. Standard quantum coding theory 23

*Electronic address: filippo.caruso@lens.unifi.it
†Electronic address: v.giovannetti@sns.it
‡Electronic address: cosmo.lupo@unicam.it
§Electronic address: stefano.mancini@unicam.it
In his seminal 1948 work “A Mathematical Theory of Communication”, C. E. Shannon set the basis of modern communication technology (Verdú, 1998). Neglecting all the semantic aspects (which are irrelevant at the level of engineering) he stressed that “the fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point.” In particular “the system must be designed to operate for each possible selection, not just the one which will actually chosen since this is unknown at the time of design.” (Shannon, 1948). At the heart of this view is what one may call the “channel” formalism, where any noisy communication line is depicted as a stochastic map connecting input signals selected by the sender of the message (Alice), who is operating at one end of the line, to their corresponding output counterparts accessible to the receiver of the messages (Bob), who is operating at the other end. In the same article Shannon also proved that the performance of a transmission line can be gauged by a single quantity, the capacity of the channel, which measures the maximum rate at which information can be reliably transferred when Alice and Bob, operating on long sequences of transmitted signals, follow a pre-established protocol (error correcting code procedure) aimed to nullify the detrimental effects of the communication noise. The rational behind this approach (which is typical to communication theory) being that communication is expensive while local operations are somehow free (unless external constraints are explicitly imposed by the selected implementation).

Rolph Landauer was the first to put on firm ground the fact that information is not just an abstract, mathematical notion but has instead an intrinsic physical nature which poses limits on the possibility of processing and transferring it (Landauer, 1961). That is why quantum mechanics, the most advanced physical theory, comes into play in the study of communication processes (Ben- nett and Shor, 1998). In this context (quantum information theory) it is recognized that any message two parties wish to exchange must be written into the states of some quantum system, say a photonic pulse propagating along an optical fiber (Caves and Drummond, 1994), and that the processing, the transmission, and the reading of such data must be carried out following the rather unconventional prescriptions established by quantum mechanics. As in the classical setting this scenario is properly formalized by introducing the notion of quantum channels as those mappings which, generalizing the notion of channel of the Shannon theory, link the initial states of the quantum information carriers (controlled by Alice) to the their output states (controlled by Bob). Interestingly enough, to evaluate the quality of these exotic communication lines several non-equivalent notions of coding procedures, as well as of corresponding capacities, must be introduced. Indeed while Alice and Bob might still be willing to use a quantum channel to exchange purely classical messages (say their social security numbers), new forms of communication can now be envisioned. For instance Alice can be interested in transferring to Bob purely quantum messages, e.g., the unknown quantum state of a quantum memory that is located in her lab, or half of a maximally entangled...
the quantum capacity that she has locally produced. The ability in sustaining state (Gühne and Tóth, 2009; Horodecki et al. 1992) have shown that quantum entanglement (Gühne and Tóth 2009; Horodecki et al. 2009) is a catalytic resource for communication. Indeed, even though entanglement alone does not constitute a communication link between distant parties (Werner 2000), allowing Alice and Bob to use pre-shared entanglement in the design of their communication protocols can boost the performance of basically any communication line they have access to (even in terms of the quantum capacity). This fact naturally brings in the notion of entanglement assisted capacities of a quantum channel (Bennett et al. 1999 b), which is yet a different way of gauging the performances of a communication line.

The vast majority of the work on quantum channels has been concerned with the study of memoryless configurations where sequences of exchanged quantum carriers are supposed to undergo the action of noisy transformations which affect them independently and identically. In this scenario coding theorems have been derived which allow one to express the various capacities of the communication line in terms of rather compact entropic formulas. For instance the classical (resp. quantum) capacity of a memoryless quantum channel is known to be computable in terms of the Holevo (resp. coherent) information (Devetak 2005; Holevo 1998 a; Lloyd 1997; Schumacher and Westmoreland 1997; Short 2002). The memoryless assumption is indeed a useful hypothesis which permits to simplify the input-output mapping induced by the noise. It also provides a realistic description for those communication schemes where the temporal rates at which signals are fed into the communication line are sufficiently low to allow for a resetting of the channel environment and to prevent signals cross-talking. Nonetheless it is not always justified. For instance, with increasing signal feeding rates, successive transmissions happen so rapidly that the environment may retain a “memory” of past events. Optical fibers are an example in which such effects can occur and can be explored experimentally (Ball, Dragan, and Banaszek 2004; Banaszek et al. 2004). Similarly, in quantum information processors, especially in solid state implementations, qubits may be so closely spaced that the same environmental degree of freedom will interact jointly with several of them (even if non-nearest-neighboring) leading to cross-talks and correlations in the noise (Duan and Guo 1998; Hu, Zhou and Guo 2007). The consideration of spatial and temporal memory effects is therefore becoming increasingly pressing with the continuing miniaturization of information processing devices and with increasing communication rates through channels. Moreover, from a fundamental point of view, quantum memory channels provide a general framework which encompasses the memoryless ones as a special case.

Apparently, the interest towards information transmission through quantum channels with memory spread after a model introduced by Macchiavello and Palma (2002). Here an example of a qubit channel with Markovian correlated noise was analyzed in which the encoding of information by means of entangled input states may increase the transmission rate of classical information. Subsequently, the study of quantum channels with memory has largely been confined to channels with Markovian correlated noise with the aim of deriving bounds on the classical capacity, see, e.g., (Bowen and Mancini 2004; Bowen, Devetak and Mancini 2005; Hamada 2002). Then, coding theorems have been devised for a class of quantum memory channels having structural properties that guarantee “regular” asymptotic behavior (Bjelaković and Boche 2008; Datta and Dorlas 2007; Kretschmann and Werner 2005). This approach can be traced back to the work of Dobrushin (1963), who considered a wide class of channels exhibiting stationary or ergodic behavior. A completely different path was taken by Hayashi and Nagaoka (Hayashi and Nagaoka 2003), that applied the “information-spectrum” method to obtain a coding theorem for the classical capacity, following the work by Verdú and Han (1994) on classical channels with memory. In the context of continuous variable systems (Braunstein and van Loock 2005; Weedbrook et al. 2012), generalizing results obtained in the classical setting (Shannon 1949) for the capacity of power-constrained Gaussian channels, quantum “water-filling” formulae have been recently derived (Pilyavets, Lupo and Mancini 2009; Schäfer, Karpov and Cerf 2010) and exact expressions for classical and quantum capacities have been computed (Lupo, Giovannetti and Mancini 2009) using an “unravelling” technique based on Toeplitz distribution theorem (Gray 1972), which allows one to map memory correlations into effective memoryless models.

Beyond quantum communication, a detailed study of quantum channels and of the mechanisms which are responsible for the introduction of memory correlations, has implications in the broader research field of quantum open system dynamics (Breuer and Petruccione 2002). As a matter of fact the input-output scheme that underlines the channel formalism is reminding us of what in physics is conventionally described as a series of scattering events, the scattered particles playing the role of the messages, while the scattering matrix playing the role of the communication line. More generally quantum channels can be used to mimic all those physical processes (temporal evolution, data-processing, etc.) which imply a state change of a system of interest from an initial to a final configuration under the influence of an external agent (the system environment). Exploiting this connection, insight on the system evolution can then be gained by analyzing its quality as a communication line. Along this direction models have been introduced in which
memory effects of communication line are described as arising from the interaction with a multi-partite environment initialized in a correlated state (Giovannetti and Mancini, 2005), allowing remarkable links between information theoretical quantities, like capacities, and statistical properties, like phase-transitions, of the underlying many-body environment (Plenio and Virmani, 2007, 2008).

When studying open system dynamics one shall not only deal with the problem of the input-output evolution of a sequence of otherwise independent carriers. Indeed it is also interesting to address the problem on the evolution of a single carrier in time, to see whether or not the associated trajectory can be described as a collection of quantum channels which are applied sequentially on that system. This is intimately related to the semigroup structure of the set of quantum channels, hence with dynamical maps and master equations (Alicki and Lendi, 1987). It turns out that dynamical evolutions which can be split into infinitesimal pieces correspond to the set of solutions of (possibly time dependent) master equations standardly used to describe open systems dynamics (Wolf and Cirac, 2008). A more general approach to open quantum systems uses the Nakajima-Zwanzig projection operator technique (Nakajima, 1958; Zwanzig, 1960) which shows, that under fairly general conditions, the master equation for the reduced density operator takes the form of a nonlocal equation in which memory effects are taken into account through the introduction of a memory kernel. Then, the problem first put forward in (Daffer, Wódkiewicz and McIver, 2003; Daffer et al., 2004) becomes to find those conditions on the memory kernel ensuring that the time evolution map is a bona fide quantum channel (Chruściński and Kossakowski, 2012).

This work aims to provide an overview of the field of quantum channels by adopting an approach which, to a large extent, is self-contained. Starting from basics results obtained in the case of memoryless configuration, it allows one to highlight also the recent developments concerning memory effects. For this purpose all the technical aspects which are directly related to quantum channels are presented and explicitly discussed in the manuscript. By necessity however, those quantum information notions which have been already reviewed elsewhere are not discussed here. In particular, in the present article the reader will not find an introduction to the structure of quantum states, entanglement, and measurement: a rather complete report on these topics can indeed be found elsewhere, e.g. in the books (Bengtsson and Zyczkowski, 2006; Holevo 2011; Nielsen and Chuang, 2000; Petz, 2008; Preskill 1998; Schumacher and Westmoreland, 2010) or in the review articles (Gühne and Töth, 2009; Horodecki et al., 2009) which provide a full account on entanglement theory. It is also worth mentioning a series of papers which, at various level, deal with quantum communication on memoryless channels. Of fundamental importance in this respect is the article (Bennett and Shor, 1998) which can be seen as a sort of manifesto for quantum information theory. The review (Caves and Drummond, 1994) presents instead a rather detailed account of the mathematical and technological issues one faces when dealing with quantum communication with photonic sources (even if some of the open problems discussed there were solved in more recent years, this article remains a useful guidance to the field). Reference (Galindo and Martín-Delgado, 2002) provides a rather compact overview on quantum information theory and discusses in a simple but clear form the basics aspects of the Shannon approach. A more mathematically oriented point of view on quantum channels is presented in Ref. (Keyl, 2002). Reference (Holevo and Giovannetti, 2012) focuses on channel capacities and their entropic characterization, while finally (Weedbrook et al., 2012) is a detailed introduction to the field of Gaussian bosonic channels.

The paper is organized as follows. The basics definitions and properties of quantum channels are presented in Sec. II. Section III discusses the transition from the memoryless setting to a more general scenario which allows for correlations in quantum communication. The various classes of memory channels are then reviewed in Sec. III.D. Section IV discusses quantum error correction codes and achievable information transmission rates. Coding theorems and capacities are presented in Sec. V. Section VI concerns models of memoryless as well as memory channels that can be solved for their communication capacities. Then, the divisibility of quantum channels and their relation to dynamical maps are discussed in Sec. VII. Finally, conclusions and an outlook on physical realizations are given in Sec. VIII.

Appendices A and B provide elementary material about distance measures for states and quasi-local algebras, respectively. In Appendix C an alternative proof of the structure decomposition theorem for non-anticipatory channels is presented, while in Appendix D an explicit derivation of capacities upper bounds is provided.

II. QUANTUM CHANNELS: BASIC DEFINITIONS AND PROPERTIES

In a typical communication scenario two parties (Alice the sender of the message and Bob the receiver) aim to exchange (classical or quantum) information by encoding it into (possibly arbitrarily long) sequences of signals which propagate through the medium that separate them. A train of transmitted signals defines a sequence of independent uses of the communication line (channel uses), and their input-output evolution from Alice to Bob is determined by the noise which tampers with the transmission process. In classical information theory (Gal- lager 1968) this is schematized by assigning an input alphabet \(X\) and an output alphabet \(Y\) whose elements \(x\) and \(y\) represent respectively the individual signals at the input and at the output of the transmission line. The
noise instead is assigned in terms of a stochastic process characterized by conditional probabilities that, given an input sequence \((x_1, x_2, \cdots)\) of elements of \(X\) transmitted by Alice, Bob will receive the sequence \((y_1, y_2, \cdots)\) of elements of \(Y\).

In quantum information the channel uses are represented by the degrees of freedom (e.g., polarization, spins) of a collection \(\{q_1, q_2, \cdots\}\) of identical information carrying objects (e.g., optical pulses, flying atoms or ions) which are locally produced by Alice and organized in a time-ordered sequence. In this setting the noise can then be described by assigning a proper mapping which acts on the (global) input states of the information carriers to produce the associated (global) output states received by Bob. Sec. III will show in which way memory effects can arise in these processes focusing on the features that are responsible for introducing correlations among the various carriers. Before doing so, however, it is useful to remind that quantum mechanics poses some fundamental structural constraints on the transformations describing the evolution of quantum systems, which must apply independently from the underlying physical mechanisms that governs the process and independently from the composite nature of the input system.

A. CPTP transformations

Let \(\Phi\) be a mapping describing the input-output relations of a generic quantum system \(Q\) (e.g., the carriers \(\{q_1, q_2, \cdots\}\) introduced in the previous paragraph) evolving under the action of some physical process

\[\rho_Q \in \mathcal{S}(\mathcal{H}_Q) \mapsto \rho_Q' = \Phi(\rho_Q) \in \mathcal{S}(\mathcal{H}_{Q'}) \, .\] (1)

Here \(\mathcal{S}(\mathcal{H}_Q)\) and \(\mathcal{S}(\mathcal{H}_{Q'})\) stand for the sets of density operators (non-negative operators with unit trace) defined on the Hilbert spaces \(\mathcal{H}_Q, \mathcal{H}_{Q'}\) (the latter may be different in general) associated, respectively, to the input and output system (unless explicitly stated in what follows it is assumed that these spaces are finite dimensional). To comply with the prescription of quantum mechanics the map \(\Phi\) must be \(\mathcal{S}(\mathcal{H}_Q)\) linear when extended to the set \(\mathcal{T}(\mathcal{H}_Q)\) of trace-class linear operators of \(\mathcal{H}_Q\). This follows from the requirement that \(\Phi\) must transform mixtures of input density operators into mixtures of the associated outputs, i.e., \(\sum_i p_i \rho_Q(i) \mapsto \sum_i p_i \Phi(\rho_Q(i))\), with \(p_i\) the probability associated with the input state \(\rho_Q(i)\) (the linearity requirement ensures that the extension of \(\Phi\) from \(\mathcal{S}(\mathcal{H}_Q)\) to \(\mathcal{T}(\mathcal{H}_Q)\) is unique).

i) linear when extended to the set \(\mathcal{T}(\mathcal{H}_Q)\) of trace-class linear operators of \(\mathcal{H}_Q\). This follows from the requirement that \(\Phi\) must transform mixtures of input density operators into mixtures of the associated outputs, i.e., \(\sum_i p_i \rho_Q(i) \mapsto \sum_i p_i \Phi(\rho_Q(i))\), with \(p_i\) the probability associated with the input state \(\rho_Q(i)\) (the linearity requirement ensures that the extension of \(\Phi\) from \(\mathcal{S}(\mathcal{H}_Q)\) to \(\mathcal{T}(\mathcal{H}_Q)\) is unique).

ii) completely positive (i.e., when acting on \(Q\) it must preserve the positivity of any density operator, including those describing a joint state \(\rho_{QA}\) of \(Q\) and an arbitrary ancillary system \(A\));

iii) trace-preserving, (i.e., it must preserve the normalization of all input states).

A violation of any of these three conditions implies the possibility of maintaining the statistical interpretation of the theory.

A transformation fulfilling the requirement i) is often said to be a super-operator. If the transformation also fulfills condition ii) and is trace-contracting (i.e., \(|\text{Tr}[\Phi(\rho)]| \leq |\text{Tr}[\rho]|\) for all \(\rho \in \mathcal{T}(\mathcal{H}_Q)\)), then it identifies a quantum operation. Finally any transformation fulfilling all three conditions given above is said CPTP (completely positive and trace preserving) and identifies a quantum channel. Special instances of these last maps are provided by the isometric channels

\[\rho_Q \mapsto U(\rho_Q) := U \rho_Q U^\dagger \, ,\] (2)

which are induced by the action of an isometric transformation \(U\) connecting \(\mathcal{H}_Q\) to \(\mathcal{H}_{Q'}\), i.e., \(U^\dagger U = \mathbb{1}\) with \(U^\dagger\) being the adjoint of \(U\) and \(\mathbb{1}\) being the identity on \(\mathcal{H}_Q\). In particular when \(Q = Q'\) and \(U\) corresponds to a unitary operator, Eq. (2) defines a unitary channel on \(Q\) which admits the channel \(U^{-1} \cdots := U^{-1} \cdots U\) as CPTP inverse. Furthermore if \(U\) is the identity operator \(\mathbb{1}\) on \(\mathcal{H}_Q\), the resulting transformation is the identity channel, denoted \(\text{id}\), which maps any state into itself, i.e., \(\text{id}(\rho_Q) = \rho_Q\) for all \(\rho_Q\).

B. Composition rules and structural properties

While referring the reader to (Bengtsson and Życzkowski 2006; Breuer and Petruccione 2002; Holevo 2011, 2012; Keyl 2002; Petz 2008) for an exhaustive characterization, here the most relevant structural properties of the set \(\mathcal{P}(Q \mapsto Q')\), formed by the CPTP maps connecting system \(Q\) to system \(Q'\), are reviewed.

Convexity: given \(\Phi, \Lambda \in \mathcal{P}(Q \mapsto Q')\) and \(p \in [0, 1]\), the transformation

\[\rho_Q \mapsto p \Phi(\rho_Q) + (1 - p) \Lambda(\rho_Q) \, ,\] (3)

is still an element of \(\mathcal{P}(Q \mapsto Q')\). For instance convex combinations of unitary channels on \(Q\) define the random unitary channel subset of \(\mathcal{P}(Q \mapsto Q)\).

Concatenation of channels: given two CPTP channels, \(\Phi\) from \(Q\) to \(Q'\) and \(\Omega\) from \(Q'\) to \(Q''\), one can define their concatenation \(\Phi \circ \Omega\) as the following CPTP transformation from \(Q\) to \(Q''\),

\[\rho_Q \mapsto (\Omega \circ \Phi)(\rho_Q) := \Omega(\Phi(\rho_Q)) \, .\] (4)

Channels concatenation allows one to introduce a relation of equivalence between CPTP maps. In particular two maps \(\Phi, \Lambda \in \mathcal{P}(Q \mapsto Q')\) are said to be unitary equivalent if exists unitary channels \(U \in \mathcal{P}(Q \mapsto Q)\) and \(U' \in \mathcal{P}(Q' \mapsto Q')\) such that

\[\Phi = U' \circ \Lambda \circ U \, ,\] (5)

the relation being reversible in the from \(\Lambda = U^{-1} \circ \Phi \circ U'^{-1}\). From the above properties it follows also that
Ψ(Q → Q) equipped with the concatenation rule possesses a non-Abelian semigroup structure, the channel id being the identity element of the set and the unitary channels U being the only invertible elements.

Tensor product of channels: – given two CPTP channels, Φ from Q to Q′ and Ω from R to R′, one can define their tensor product Φ ⊗ Ω as the CPTP transformation from the composite system QR to the composite system Q′R′, which given an arbitrary operator tensor NQ ⊗ MR ∈ T(HQ ⊗ HR) transforms it into

\[(Φ ⊗ Ω)(NQ ⊗ MR) := Φ(NQ) ⊗ Ω(MR).\]  

Special instances are provided by the transformation Φ ⊗ id obtained by tensoring Φ with the identity channel acting on an external system: these maps are called extension of Φ and represent the action of such channel when the system Q (where Φ was originally defined), is described as part of an enlarged composite system – notice that this structure was implicitly assumed when stating point ii) of the previous section.

Concatenations and tensor products of quantum channels represent two alternative ways of composing CPTP maps which, to some extent, mimic respectively the in- and in-parallel composition rules of electrical circuit elements. In particular, as discussed in Sec. VII channel concatenation is naturally suited to characterize the temporal correlations of a single quantum system (the sequential applications of CPTP maps corresponding to different stage of the system evolution). On the contrary the tensor product allows to describe spatial correlations which might be present in the evolution of composite quantum systems. Also, as discussed in Sec. [1] tensor products can be employed to describe the transformations that a sequence of information carriers encounters when transmitted through a communication line.

C. Stinespring representation, Kraus representation, and Choi-Jamiołkowski isomorphism

It can be shown [Stinespring 1955] that a mapping satisfies the CPTP conditions detailed in the previous section, if and only if it admits dilations that allow one to represent it in terms of a unitary coupling with an external environment E (which is possibly fictitious and may not correspond to the actual physical environment responsible for the system evolution). For instance, taking for simplicity Q = Q′, one can write

\[Φ(ρ_Q) = \text{Tr}_E \left[ U_{QE}(ρ_Q ⊗ ω_E)U_{QE}^† \right],\]  

where ω_E is a fixed state of E, U_{QE} is the unitary transformation coupling the latter to the input system Q, and Tr_E denotes the partial trace over the environment. The representation is not unique. Nonetheless by enlarging the environment E to describe the environment state as a pure state, ω_E = |ω_E⟩⟨ω_E| (see Sec. [1] for a proper definition of this purification mechanism), the choice of U_{QE} can be shown to be unique up to a local isometric transformation on E. Under this condition the dilation provides what is generally known a Stinespring representation for Φ.

The CPTP conditions are also equivalent to the possibility of expressing Φ in operator sum (or Kraus) representation [Choi 1975, Kraus 1971].

\[Φ(ρ_Q) = \sum_j K_j ρ_Q K_j^†,\]  

with \{K_j\} being operators on H_Q satisfying the normalization condition \(\sum_j K_j^† K_j = \text{id}\). As in the case of the unitary dilation, their choice is in general not unique. One can however guarantee that a Kraus representation exists with no more of \(d^2\) elements (\(d\) being the dimension of H_Q). Kraus and the Stinespring representations can also be put in mutual correspondence by identifying the operator K_j with the linear operator \(E(e_j)U_{QE}|ω_E⟩\) of H_Q, where \{e_j\} is an orthonormal basis of E.

It is finally worth reminding that there exists a fundamental relation, known as the Choi-Jamiołkowski (CJ) isomorphism [Choi 1975, Jamiołkowski 1972], which permits to describe any CPTP Φ as a density operator of a composite system QA with A being an auxiliary system having the same dimension d of Q. The explicitly connection is obtained by applying the map Φ to half of a maximally entangled state Gühne and Tóth 2009 Horodecki et al. 2009 \(|β⟩_Q = \sum_j |e_j⟩_Q \otimes |e_j⟩_A/\sqrt{d}\) of QA to create the so called CJ state of the channel

\[ρ_{QA}^{Φ} := (Φ ⊗ \text{id})(|β⟩_Q A ⟨β|),\]  

where id stands for the identity map on A and \{|e_j⟩_Q\} and \{|e_j⟩_A\} represent orthonormal basis of Q and A respectively.

D. Heisenberg picture: the dual channel

Equation [1] implicitly assumes the Schrödinger picture in which the states of the system are evolved while the observables are kept fixed. In the Heisenberg picture, in which instead the states are fixed and the observables evolve in time, the CPTP transformation Φ ∈ Ψ(Q → Q)
$Q'$ is replaced by its dual map

$$O \in \mathcal{B}(\mathcal{H}_Q) \mapsto \Phi^*(O) \in \mathcal{B}(\mathcal{H}_Q),$$

(10)

operating on the bounded operator algebra $\mathcal{B}(\mathcal{H}_Q)$ of the receiver observable and defined through the identity

$$\text{Tr} [\Phi(\rho_O) O] = \text{Tr} [\rho_O \Phi^*(O)],$$

(11)

which holds for all $O \in \mathcal{B}(\mathcal{H}_Q)$ and for all $\rho_O \in \mathcal{S}(\mathcal{H}_Q)$. As its Schrödinger counterpart, the transformation $\Phi^*$ is linear and completely positive, but in general it is not trace preserving. On the other hand it is always 

unital, i.e., it maps the identity operator into itself. Operator sum representations for $\Phi^*$ can be easily constructed from those of $\Phi$ (8), yielding

$$\Phi^*(O) = \sum_j K_j^i O K_j.$$

(12)

Notice also that in the dual picture the concatenation of channels goes in reverse order with respect to the Schrödinger picture, i.e., given $\Phi$ and $\Omega$ CPTP maps,

$$(\Omega \circ \Phi)^* = \Phi^* \circ \Omega^*.$$

(13)

E. cq- and qc-channels

Besides considering physical transformations which represent the evolution of quantum carriers, in quantum information it is useful to describe processes which map classical inputs into quantum states (cq-channels) or, vice-versa, quantum states into classical inputs (qc-channels). Specifically the former define state preparation procedures where a symbol $x$ extracted from a classical alphabet $X$ with probability $p_x$ is encoded into a state $\rho_Q^{(x)}$ of the quantum system $Q$, thus producing an average density operator $\sum_{x \in X} p_x \rho_Q^{(x)}$. qc-channels instead define measurement procedures which, given $\rho_Q \in \mathcal{S}(\mathcal{H}_Q)$, produce a classical outcomes $x \in X$ with conditional probabilities

$$p_x(\rho_Q) = \text{Tr}[E_x \rho_Q],$$

(14)

with $\{E_x\}_{x \in X}$ being a set of positive operators on $\mathcal{H}(Q)$, satisfying the normalization condition $\sum_x E_x = 1$, which defines the statistic of the measurement in the POVM (Positive-Operator Valued Measure) representation (Breuer and Petruccione 2002, Holevo 2011, Petz 2008). cq- and qc-channels can both be extended to CPTP maps (1) by introducing an ancillary quantum system $Q'$ of dimension equal to the cardinality of $X$, and characterized by an orthonormal set $\{ |e_x\rangle_{Q'}\}_{x \in X}$ (Holevo 1998). For instance taking $p_x = Q' \langle e_x | \rho_Q Q' | e_x \rangle_{Q'}$ with $\rho_Q \in \mathcal{S}(\mathcal{H}_Q')$, the cq-channel defined above induces the following CPTP mapping from $Q'$ to $Q$,

$$\rho_Q' \mapsto \Phi_{cq}(\rho_Q') := \sum_{x \in X} Q' \langle e_x | \rho_Q Q' | e_x \rangle_{Q'} \rho_Q^{(x)}.$$

(15)

Analogously the qc-channel induces the following CPTP mapping from $Q$ to $Q'$,

$$\rho_Q \mapsto \Phi_{qc}(\rho_Q) := \sum_{x \in X} | e_x \rangle_{Q} Q' \langle e_x | \text{Tr}[E_x \rho_Q].$$

(16)

The concatenation $\Phi_{qc} \circ \Phi \circ \Phi_{qc}$, with $\Phi$ being a generic quantum channel on from $Q$ to $Q'$, can be also represented as a CPTP map (from $Q'$ to $Q'$) and describes the typical scenario where a collection of classical messages (represented by elements of the set $X$) are transferred to Bob via a quantum link (represented by the $\Phi$) who “reads” them through the POVM $\{ E_x \}_{x \in X}$. In particular when applied to elements of the orthonormal set $\{|e_x\rangle_{Q'}\}_{x \in X}$, $\Phi_{cq} \circ \Phi \circ \Phi_{qc}$, it induces a classical stochastic process where $x \in X$ is mapped into $x' \in X$ with conditional probability

$$p(x' | x) = \text{Tr}[E_{x'} \Phi(\rho \langle e_x | e_x \rangle_{Q'})].$$

(17)

F. Entanglement Breaking and PPT channels

cq- and qc-channels defined in the previous section are particular instances of a larger group of CPTP transformations, called Entanglement Breaking (EB). As the name suggests, a channel $\Phi$ is EB if, when operating on half of joint input state $\rho_{QA}$ of $Q$ and of an ancillary system $A$, produces outputs states $\rho_{Q'A} = (\Phi \otimes \text{id})(\rho_{QA})$ that are separable (i.e., not entangled) (Gühne and Tóth 2009, Horodecki et al. 2009). These maps are closed under convex combination and channel concatenations, that is, given $\Phi_1$ and $\Phi_2$ EB, then $p\Phi_1 + (1-p)\Phi_2$ and $\Phi_1 \circ \Phi_2$ are also EB for all $p \in [0,1]$ (more generally concatenating an EB map with a generic CPTP map produces a EB channel). Necessary and sufficient condition for being EB can be found in Refs. (Horodecki, Shor and Ruskai 2003, Ruskai 2003) and, for the special case of infinite dimensional system, in Ref. (Holevo 2008). In particular, $\Phi$ is EB if and only if its associated CJ state is separable. Alternatively $\Phi$ is EB if and only if it is possible to identify a POVM $\{F_k\}$ on $Q$ and collection of states $\{\rho_{Q,k}\}_k$ in the output space $Q'$ such that

$$\Phi(\rho_Q) = \sum_k \rho_{Q,k}^\Phi \text{Tr}[F_k \rho_Q],$$

(18)

for all input $\rho_Q$ (this last condition immediately shows that maps (15), (16) are indeed EB).

EB channels form a proper subset of PPT channels. The latter are defined as those channels which produce output states $\rho_{Q'A} = (\Phi \otimes \text{id})(\rho_{QA})$ with positive partial transpose (PPT) (Horodecki et al. 1996, Peres 1996, Rains 2001). A necessary and sufficient condition for such a property is to have CJ state with PPT. Channels which are PPT but not EB are called entanglement binding maps. Generalizations of EB channels have been presented in (De Pasquale and Giovannetti 2012) to describe those CPTP maps which become EB only after
a certain number of concatenations, and in [Filippov, Rybar, and Ziman 2012] [Moravčíková and Ziman 2010] to describe maps which, when acting jointly on a composite system break entanglement among the subsystems that compose it.

G. Complementary channels and Degradability

Associated with the Stinespring representation \( \Phi \) is the notion of complementary channel of \( \Phi \). The latter is the CPTP map \( \tilde{\Phi} \in \mathcal{P}(Q \mapsto E) \) which sends the initial states of the system \( Q \) into the states of the environment \( E \) through the transformation

\[
\rho_Q \mapsto \tilde{\Phi}(\rho_Q) := \text{Tr}_Q \left[ U_{QE} \left( \rho_Q \otimes |\omega\rangle_E \langle \omega| \right) U_{QE}^\dagger \right],
\]

where \( \text{Tr}_Q \) denotes the partial trace over the system Hilbert space. The purity of the environmental state \( \omega_E = |\omega\rangle_E \langle \omega| \) ensures the unicity of \( \tilde{\Phi} \) up to an isometric transformation on \( E \). Channels \( \Phi \) defined in terms of unitary dilations \( \Phi = \Omega \circ \Phi \) with non pure states \( \omega \) are called weak-complementaries of \( \Phi \) and in general don’t enjoy such symmetry [Caruso and Giovannetti 2006] [Caruso, Giovannetti and Holevo 2006].

The definition of \( \Phi \) allows us to introduce another property of quantum channels, which is called degradability [Devetak and Shor 2005]. A map \( \Phi \) is degradable when one can recover the final environment state \( \Phi(\rho_Q) \) just applying a third CPTP map to the output system state. More formally, a degradable map is such that there exists a CPTP map \( \Omega \in \mathcal{P}(Q \mapsto E) \) satisfying the relation:

\[
\tilde{\Phi} = \Omega \circ \Phi.
\]

Similarly, a channel is called anti-degradable when the opposite relation holds, i.e.,

\[
\Phi = \Omega \circ \tilde{\Phi},
\]

for some \( \Omega \in \mathcal{P}(E \mapsto Q) \). Special example of anti-degradable channels are the symmetric channels introduced in [Smith, Smolin and Winter 2008]; these are CPTP maps for which \( \Phi \) and \( \tilde{\Phi} \) coincide. Structural properties of degradable and anti-degradable channels have been extensively analyzed in Ref. [Cubitt, Ruskai and Smith 2005], showing for instance that EB channels are always anti-degradable. Analogous definitions can be obtained for weak-complementary channels: in this case one says that \( \Phi \) is weakly-degradable if Eq. (20) holds – no need instead to define a weakly-anti-degradability condition as the latter can be shown to be equivalent to the anti-degradability condition [Caruso, Giovannetti and Holevo 2006].

H. Causal, Localizable, LOCC and Separable channels

Additional structures arise when a quantum channel \( \Phi \) acts on a multipartite system, e.g., a bipartite one, \( Q = Q_1Q_2 \). It is useful to imagine that the two subsystems are associated with spatially or temporally separated laboratories where local CPTP maps can be applied and can exchange classical of quantum information. In particular, bipartite channels can be characterized in terms of: 1) how the output of one subsystem changes if a local transformation is applied on the input of the other subsystem; 2) which resources (e.g., a pre-shared quantum state, local operations on the subsystems, classical or quantum communication) are needed to simulate the bipartite channel.

The notions of causal and semi-causal channels developed in Refs. [Eggeling, Schlingemann and Werner 2001] [Piani et al. 2006] provide means of characterizing how the output of one subsystem depends on the input of the other. In this context a quantum channel \( \Phi \in \mathcal{P}(Q \mapsto Q) \) acting on a bipartite system \( Q = Q_1Q_2 \) is said to be \( Q_1 \rightarrow Q_2 \) semicausal [Beckman et al. 2001] if for any local CPTP map \( \Psi \in \mathcal{P}(Q_1 \mapsto Q_1) \) applied to \( Q_1 \) before the action of \( \Phi \), there is no detectable effect in the subsystem \( Q_2 \), i.e.,

\[
\text{Tr}_{Q_1} \left[ \Phi(\rho_Q) \right] = \text{Tr}_{Q_1} \left[ \Phi \left( \left( \Psi \otimes \text{id} \right) (\rho_Q) \right) \right], \tag{22}
\]

where \( \rho_Q \) is a generic (possibly entangled) input state of the two carriers and where \( \text{Tr}_{Q_1} \) denotes the partial trace with respect to \( Q_1 \). In other words, for \( Q_1 \rightarrow Q_2 \) semicausal maps cross-talking from \( Q_1 \) to \( Q_2 \) is prevented. Similarly, one introduces the notion of \( Q_2 \rightarrow Q_1 \) semicausal map. When both properties are satisfied, the map is called causal or non-signaling. Special examples of non-signaling channels are the tensor product channels \( \Phi = \Phi_1 \otimes \Phi_2 \) with \( \Phi_{1,2} \) being CPTP maps operating locally on \( Q_1 \) and \( Q_2 \) respectively.

Another way of characterizing a bipartite quantum channel is in terms of the physical resources which are needed to simulate it. A bipartite channel is said to be localizable if it can be implemented by applying local CPTP maps on the subsystems with the assistance of a pre-shared bipartite quantum state [Beckman et al. 2001]. Notice that the simulation of localizable channels does not require classical nor quantum communication between the two laboratories. Formally, this is the case when \( \Phi \) can be represented as

\[
\Phi(\rho_{Q_1Q_2}) = \text{Tr}_{A_1A_2} \left[ (\Psi \otimes \Omega) \left( \rho_{Q_1Q_2} \otimes \omega_{A_1A_2} \right) \right], \tag{23}
\]

where \( \omega_{A_1A_2} \) is a shared bipartite state, and \( \Psi \) and \( \Omega \) are quantum channels acting locally on subsystems \( Q_1A_1 \) and \( Q_2A_2 \), respectively. Otherwise, \( \Phi \) is called \( Q_1 \rightarrow Q_2 \) semi-localizable if also one-way quantum communication from \( Q_1 \) to \( Q_2 \) is required to simulate the channel. Accordingly in this case Eq. (23) is replaced by

\[
\Phi(\rho_{Q_1Q_2}) = \text{Tr}_A \left[ (\Psi \otimes \Omega) \left( \rho_{Q_1Q_2} \otimes \omega_A \right) \right], \tag{24}
\]

with \( \omega_A \) being the state of an ancillary system \( A \) which is transmitted from one laboratory to the other and acts as the mediator between \( Q_1 \) and \( Q_2 \), and \( \Psi \) and \( \Omega \) are quantum channels acting on the systems \( Q_2A \) and \( Q_1A \).
respectively. Analogously, one defines a $Q_2 \rightarrow Q_1$ semilocalizable map.

By comparison of (22) and (24) it follows that all $Q_1 \rightarrow Q_2$ semilocalizable maps are $Q_2 \rightarrow Q_1$ semicausal, which in turn implies that all localizable maps are causal. Moreover, it can be proven that semicausality implies semilocalizability, hence semicausal and semilocalizable maps coincide, although causal and localizable maps do not. 

An important class of bipartite quantum channels are finally those that can be simulated only with local operations and classical communication (LOCC), see (Chitambar et al., 2012) for a recent survey. These channels are hence termed LOCC channels. Classical communication is generally allowed in both directions between the two laboratories (otherwise, one defines 1-way LOCC channels). The most general LOCC channel is obtained by concatenating local CPTP maps and one-way classical communication. A LOCC transformation can hence be simulated by a finite number of iterations of the following sequence of operations: 1) on one of the two subsystems, say $Q_1$, a local CPTP map $\Psi_1$ is applied; 2) classical information is sent from $Q_1$ to $Q_2$, possibly conditioned on the local output of the map $\Psi_1$; 3) conditioned on the received classical information, a local CPTP map $\Psi_2$ is applied on subsystem $Q_2$; 4) the sequence of operations is repeated with the roles of $Q_1$ and $Q_2$ exchanged. A closely related class of bipartite channels is that of separable channels, defined as those channels admitting a Kraus representation in which all the Kraus operators are in the form of a direct product of operators acting on the local subsystems. It is easy to see that all the LOCC channels are separable. Interesting enough, there exist separable channels which are not LOCC (Bennett et al., 1999a).

I. Examples

Here some enlightening examples of quantum channels are presented.

1. Qubit channels

Qubit channels are the simplest, yet non trivial, example of quantum channels: they are CPTP transformations $\Phi \in \mathcal{B}(Q \rightarrow Q)$ that map the states of a two-dimensional quantum system (qubit) into states of the same system (in this case $\mathcal{H}_Q = \mathbb{C}^2$). A compact characterization of these channels can be obtained by adopting the Bloch ball representation according to which any density operator $\rho$ of the system is uniquely identified with the corresponding (Bloch) vector $r = (r_x, r_y, r_z) \in \mathbb{R}^3$ of length $|r| \leq 1$ via the correspondence

$$\rho = \rho(r) := \frac{1}{2}(\mathbb{1} + r \cdot \sigma),$$

(25)

where $s = (\sigma_x, \sigma_y, \sigma_z)^T$ is the column vector formed by the Pauli matrices. In this framework any qubit channel $\Phi$ induces affine transformations of the form

$$r \rightarrow r' = Mr + t,$$

(26)

with $M$ and $t$ being respectively a fixed $3 \times 3$ real matrix and a fixed three dimensional real vector satisfying certain consistency requirements – see Refs. (King and Ruskai 2001; Ruskai 2003 Ruskai Szarek, and Werner 2002). In particular qubit unital channels are obtained for $t = 0$ and $M^T M \leq I$, the inequality being saturated if and only if $\Phi$ describes a unitary transformation (the latter case corresponds to have $M \in \text{SO}(3, \mathbb{R})$). Exploiting this fact and the matrix singular value decomposition (Horn and Johnson 1990) one can use the unitary equivalence of Eq. (4) to identify a canonical form for the qubit channel $\Phi$ where the matrix $M$ of Eq. (26) writes as $O'DO'$ with $D$ being a real diagonal $3 \times 3$ matrix, and with $O'$ and $O$ being elements of $\text{SO}(3, \mathbb{R})$.

An important class of qubit channels which have been extensively analyzed in the literature are those which admit a representation (8) with only two Kraus operators $K_0$ and $K_1$. In the canonical basis formed by the eigenvectors $\{|0\rangle, |1\rangle\}$ of the Pauli operator $\sigma_z$ they can be parametrized as

$$K_0 = \begin{pmatrix} \cos \theta & 0 \\ 0 & \cos \phi \end{pmatrix}, \quad K_1 = \begin{pmatrix} 0 & \sin \phi \\ \sin \theta & 0 \end{pmatrix},$$

(27)

with $\theta, \phi \in [0, \pi]$, up to unitary rotation, the corresponding affine mapping (26) being obtained with $M = \text{diag}(\cos(\phi - \theta), \cos(\phi + \theta), (\cos(2\theta) + \cos(2\phi))/2)$ and $t = (0, 0, \cos(2\theta) - \cos(2\phi))/2$. In the Stinespring representation (7) these maps describe situations where the qubit system interacts with the smallest non-trivial environment (i.e., another qubit initialized in a pure state) and can be shown to be degradable for $\cos(2\theta)/\cos(2\phi) \geq 0$ and anti-degradable otherwise (Caruso and Giovannetti, 2007 Giovannetti and Fazio, 2005 Wolf and Pérez-Garcia, 2007). In particular setting $\cos(2\theta) = 1/2$ and $\cos(2\phi) = \eta - 1/2$, Eq. (27) defines the amplitude damping channel with damping rate $\eta$. For $\sin \theta = \pm \sin \phi$ instead one gets unital maps. Specifically for $\phi = 0$, Eq. (27) describes the bit-flip channel that exchanges the states $|0\rangle$ and $|1\rangle$ with probability $p_x = \sin^2 \phi$. By applying the unitary matrix (Hadamard transform) $\frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ to $K_1, K_2$ in Eq. (27) it is obtained a unitarily equivalent channel called phase-flip channel (or phase damping channel) that introduces a $\pi$ shift between the states $|0\rangle$ and $|1\rangle$ with probability $p_z = \sin^2 \phi$. Still Eq. (27) for $\phi = -\theta$ describes the bit-phase flip channel which with probability $p_y = \sin^2 \phi$ exchanges the states $|0\rangle$ and $|1\rangle$ and also add a relative $\pi$ shift to them. Convex combinations of these three maps plus the identity channel define the class of Pauli channels: i.e.,

$$\Phi(\rho) = p_0 \rho + p_1 \sigma_x \rho \sigma_x + p_2 \sigma_y \rho \sigma_y + p_3 \sigma_z \rho \sigma_z,$$

(28)
with non-negative parameters $p_0 + p_1 + p_2 + p_3 = 1$. Via the canonical representation detailed in the previous paragraphs any other unital qubit map can be obtained from Eq. (28) through the concatenation Eq. (5).

2. Erasure channels

Erasure channels describe those communication scenarios where errors are somehow heralded (i.e., the receiver Bob can determine whether or not something bad has happened to Alice’s original message): accordingly they provide the simplest examples of CPTP maps operating among spaces of different dimensionality. Given a system $Q$ described by the Hilbert space $\mathcal{H}_Q$, an erasure map is a stochastic transformation connecting $\mathcal{H}(\mathcal{H}_Q)$ with $\mathcal{H}(\mathcal{H}_{Q'})$, where $\mathcal{H}_{Q'} = \mathcal{H}_Q \oplus |e\rangle$ and $\oplus$ denotes the direct sum of the input Hilbert space with an extra “erasure” state $|e\rangle$ (the error “flag”) which is orthogonal to each of the vectors of $Q$. In particular as described in Ref. [Bennett, DiVincenzo, and Smolin, 1997] [Grassl, Beth, and Pellizzari, 1997], the channel will send the input $\rho_Q$ into itself with probability $1 - p$ and into $|e\rangle$ with probability $p$.

3. Weyl covariant channels

Given a quantum system of finite dimension $d$ and the canonical basis $\{|e_k\rangle\}_{k=0,...,d-1}$, consider the group $\mathbb{Z}_d \times \mathbb{Z}_d$ as a discrete phase space and take the unitary representation of such a group in the Hilbert space $\mathcal{H}$ of the system as

$$z = (x, y) \mapsto W_z = U^x V^y,$$

where $x, y \in \mathbb{Z}_d$ and $U, V$ are unitary operators on $\mathcal{H}$ generalizing the Pauli operators $\sigma_x$ and $\sigma_z$ in the following way [Gottesman, 1999]

$$U|e_k\rangle = |e_{k+1(\text{mod} d)}\rangle, \quad V|e_k\rangle = \exp\left(\frac{2\pi i k}{d}\right)|e_k\rangle.$$

The operators $W_z$ are the discrete Weyl operators and satisfy the canonical commutation relations

$$W_z W_{z'} = e^{(2\pi i/d)\Sigma z z'^\top} W_{z'} W_z,$$

where

$$\Sigma := \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$

is the matrix representation of the symplectic form.

The algebra $\mathcal{B}(\mathcal{H})$ of operators on $\mathcal{H}$ can be considered as a Hilbert space supplied by the Hilbert-Schmidt inner product. There, the Weyl operators form an orthogonal basis, $\text{Tr}\left(W_z W_{z'}^\dagger\right) = d\delta_{z, z'}$, hence for all $O \in \mathcal{B}(\mathcal{H})$ one has

$$O = \sum_z f_O(z) W_z, \quad f_O(z) = \frac{1}{d} \text{Tr}\left(O W_z^\dagger\right).$$

In this scenario a CPTP map $\Phi$ is called Weyl covariant [Fukuda and Holevo, 2005] when for all $z \in \mathbb{Z}_d \times \mathbb{Z}_d$ it fulfills the identity

$$\Phi \circ W_z = W_z \circ \Phi,$$

with $W_z$ representing the quantum channel $z$ associated with the unitary $W_z$. Applying Eq. (34) to the operator $W_z$, from Eq. (31) it follows that $\Phi(W_z)$ commutes with $W_z$. Hence, by means of Eq. (33), one can write

$$\Phi(W_z) = \phi(z) W_z,$$

with $\phi(z)$ a complex-valued function, termed ‘characteristic function of the channel’ (w.l.o.g. it can be assumed $\phi(0) = 1$).

Special examples of Weyl covariant channels are provided by Weyl channels [Amosov, 2007] defined as those CPTP maps which admit Kraus decomposition in terms of random Weyl operators, i.e., $\Phi = \sum_{z} \rho_{z} W_z$, with $\rho_z$ a probability distribution over $\mathbb{Z}_d \times \mathbb{Z}_d$. These are unital maps and their associated characteristic function is given by $\phi(z) = \sum_{e} e^{(2\pi i/d)\Sigma z e} \rho_z$. In particular any $d$–depolymerizing channel

$$\Phi(\rho) = \lambda \rho + (1 - \lambda) \frac{1}{d} \mathbb{1},$$

with $\lambda \in [0, 1]$ is a Weyl channel having $\phi(z) = \lambda$ for $z \neq 0$. One notices also that since for $d = 2$ the Weyl operators reduce to the standard Pauli operators including identity, any unital qubit ($d = 2$) channel which is unitarily equivalent to the Pauli channels Eq. (28) is a Weyl channel. It is also possible to define the transpose $d$-depolymerizing transformation,

$$\Phi(\rho) = \lambda \rho^\top + (1 - \lambda) \frac{1}{d} \mathbb{1},$$

which defines a CPTP map for $\lambda \in \left[1 - \frac{d}{\pi+1}, 1 - \frac{d}{\pi-1}\right]$ [Fannes et al., 2004].

4. Continuous variable quantum channels

Up to now mainly finite dimensional Hilbert spaces have been considered. This has been done to avoid technicalities related with the proper definition of the domains of the functionals. It is true however that the most common implementations of quantum communication lines are typically realized with continuous-variable (CV) systems [Braunstein and van Loock, 2005] [Weedbrook et al., 2012] which at quantum level are associated with infinite dimensional space – consider for instance the transferring of classical signals encoded into light pulses propagating along optical fibers or in free-space [Caves and Drummond, 1994].

One remark that CV systems admit a description in terms of a discrete set of (say $n$) bosonic oscillators, typically a set of normal modes of the electromagnetic
tors \(\rho\) obeying canonical commutation relations, \([a_k, a_k^\dagger] = \delta_{kk}\). Introducing the generalized “positions” and “momenta” coordinates \(\{x_k, y_k\}_{k=1,\ldots,n}\), the density operators \(\rho\) of the CV system can be represented in terms of the (symmetrically ordered) characteristic functions \(\chi(z) = \text{Tr}[\rho V(z)]\), where \(z := (x_1, y_1, \ldots, x_n, y_n)^T\) defines the vector of phase-space variables and \(V(z) := \exp \left[ \frac{i}{\sqrt{2}} \sum_k (x_k + i y_k) a_k^\dagger - \text{h.c.} \right]\) denote \(n\)-mode Weyl operators. The function \(\chi(z)\) and the operator \(V(z)\) represent the infinite dimensional counterparts of \(\rho(f)\) and \(\mathcal{W}_2\) introduced in Sec. 11.1.3. In particular, \(\chi(z)\) fulfill commutation relations analogous to Eq. (31), i.e.,

\[
V(z)V(z') = e^{2\pi i z^\top \Sigma(n) z} V(z')V(z),
\]

where now \(\Sigma(n)\) is the \(2n \times 2n\) block matrix defined by \(\Sigma(n) = \bigoplus_{k=1}^n \Sigma\) with \(\Sigma\) as in Eq. (32) representing the single-mode phase-space canonical symplectic form.

Due to their physical relevance and to the relative simplicity of their mathematical description, a remarkable class of states is that of the so-called Gaussian states (Eisert and Plenio 2003; Ferraro, Olivares and Paris 2005; Holevo 2011). They correspond to multi-mode (thermal) Gibbs states of Hamiltonians which are quadratic in the ladder operators of the system and are formally identified by property of possessing Gaussian characteristic functions, i.e.,

\[
\chi(z) = \exp \left( i m^\top z - \frac{1}{2} z^\top C(n) z \right).
\]

In this expression \(m\) is the vector of first moments

\[
m_i = \text{Tr}(x_i \rho),
\]

where

\[
\sqrt{2}x := (a_1 + a_1^\dagger), i(a_1 - a_1^\dagger), (a_2 + a_2^\dagger), i(a_2 - a_2^\dagger), \ldots
\]

Furthermore, \(C(n)\) is the covariance matrix (CM)

\[
C(n)_{ij} = \frac{1}{2} \text{Tr}(x_i x_j \rho) + \frac{1}{2} \text{Tr}(x_j x_i \rho) - \text{Tr}(x_i \rho)\text{Tr}(x_j \rho),
\]

obeying the generalized uncertainty relation (Simon et al. 1994)

\[
C(n) - i\Sigma(n)/2 \geq 0.
\]

Concerning quantum channels in CV systems, attention has been mainly devoted to the study of Gaussian channels, i.e., CPTP maps that map Gaussian input states into Gaussian output states (Holevo and Werner 2001; Weedbrook et al. 2012). A part from attenuation and thermalization events arising from linear interactions with bosonic baths, they include also squeezing and linear amplification processes. When applied to a (non-necessarily Gaussian) density operator \(\rho\) with characteristic function \(\chi(z)\), a Gaussian channel \(\Phi\) will transform it into an output density operator \(\rho' = \Phi(\rho)\) having characteristic function

\[
\chi'(z) = \chi(X(n)^\top z) f(z),
\]

where \(X(n)\) is a matrix inducing a linear transformation on the \(2n\)-dimensional phase-space vector \(z\), and the function \(f(z)\) is Gaussian, i.e., \(f(z) = \exp (i d(n)^\top z - \frac{1}{2} z^\top Y(n) z)\). The linear term proportional to the vector \(d(n)\) accounts for a translation (displacement) of the mean \(m\), while the quadratic term proportional to the matrix \(Y(n)\) adds a term to the CM. CPTP conditions are ensured if and only if

\[
Y(n) - i(\Sigma(n) - \Sigma(n)X(n)X(n)^\top)/2 \geq 0.
\]

Gaussian transformations which are also unitary are characterized by the property that \(X(n)\) is a symplectic matrix (i.e., \(X(n)\Sigma(n)X(n)^\top = \Sigma(n)\)), and \(Y(n) = 0\). A \(n\)-mode Gaussian channel is hence characterized by the triad \((d(n), X(n), Y(n))\) under the constraint (45) and the concatenation of two Gaussian channels with associated triads \((d_1(n), X_1(n), Y_1(n))\) and \((d_2(n), X_2(n), Y_2(n))\) yields \((X_2(n)d_1(n) + d_2(n), X_2(n)X_1(n), X_2(n)Y_1(n)X_2(n)^\top + Y_2(n))\). It follows that, by applying suitable Gaussian unitaries at the input and output of the channel, one can always reduce the channel in a canonical form, in which \(d(n) = 0\), and the matrices \(X(n), Y(n)\) take a particular symmetric form. For the case of channels acting on one or two modes, the reduction to canonical forms allows the classification of Gaussian quantum channels according to invariance under unitary transformations (Caruso and Giovannetti 2005; Caruso et al. 2008; Holevo 2007b; Serafini, Eisert and Wolf 2005).

The basic processes of linear attenuation and amplification are modeled by single-mode Gaussian channels with \(X^{(1)} = \sqrt{\eta}\) and \(Y^{(1)} = (1 - \eta)/2\). For \(\eta \leq 1\) these channels describe linear losses (with attenuation factor \(\eta\)), while for \(\eta > 1\) they model the process of parametric amplification (with gain \(\eta\)). If extra Gaussian noise affects the attenuation or amplification process, one gets the noisy versions of the lossy and amplifier channel. In particular, the lossy and noisy Gaussian channel is defined by \(X^{(1)} = \sqrt{\eta}\) and \(Y^{(1)} = (1 - \eta)(N_{\text{th}} + 1/2)\) (\(\eta \in [0,1]\) and \(N_{\text{th}} \geq 0\)), and the additive noise Gaussian channel by \(X^{(1)} = 1\) and \(Y^{(1)} = N_{\text{add}}\) (\(N_{\text{add}} \geq 0\)). Notice that the additive noise channel can be obtained from the lossy and noisy channel by taking the limit of \(\eta \to 1\) and \(N_{\text{th}} \to \infty\) with \((1 - \eta)(N_{\text{th}} + 1/2) = N_{\text{add}}\).

### J. Transfer fidelities and channel distances

In quantum information distance measures are of fundamental importance: by determining how far apart two states or two transformations are from each other, they are an essential guidance in the optimization of the data-processing.
1. Input-output and entanglement fidelity of a quantum channel

A proper way to determine how much a system $Q$ is modified by the action of a channel $\Phi \in \mathcal{P}(Q \mapsto Q)$, can be obtained by considering the fidelity functional $F(\rho_1, \rho_2)$ (Jozsa 1994, Uhlmann 1976), which, given $\rho_1$, $\rho_2$ two states of a quantum system, gauges how close they are, the maximum value 1 being reached only when $\rho_1 = \rho_2$ (see Appendix A for a brief summary of the other relevant properties). Accordingly, for each input $\rho_Q$ one defines the input-output (or transfer) fidelity associated with the map $\Phi$, as

$$F(\rho_Q; \Phi) := F(\rho_Q, \Phi(\rho_Q)),$$

which for a pure state $|\psi\rangle_Q$ is linked to the error probability $P_e(|\psi\rangle_Q; \Phi)$ of not getting the right state at the channel output, via the identity

$$P_e(|\psi\rangle_Q; \Phi) = 1 - F(|\psi\rangle_Q; \Phi).$$

An overall estimate of the disturbance introduced by the channel can then be obtained by looking at how different from unity is the minimum or (alternatively) the average of $F(\rho_Q; \Phi)$ evaluated with respect to all possible pure input states of $Q$, i.e., the quantities

$$F_{\text{min}}(\Phi) := \min_{|\psi\rangle_Q} F(|\psi\rangle_Q; \Phi),$$

$$\bar{F}(\Phi) := \int d\mu(\psi) F(|\psi\rangle_Q; \Phi),$$

the rational being that $F_{\text{min}}(\Phi) = 1$, as well as $\bar{F}(\Phi) = 1$, can occur if and only if $\Phi$ coincides with the identity channel $\text{id}$. The average in Eq. (48) is performed with respect to the Haar measure $d\mu(\psi)$ of the group whose action on a vector $|\psi\rangle$ is able to generate the entire space of states (Bengtsson and Życzkowski 2006); the minimization in Eq. (48) instead can be generalized to include also mixed states by exploiting the concavity property of the fidelity (Nielsen and Chuang 2000, Wilde 2013), i.e., $F_{\text{min}}(\Phi) = \min_{\rho_Q} F(\rho_Q; \Phi)$.

To gauge the disturbance of the channel $\Phi$, one may also consider its entanglement fidelities (Schumacher 1996), defined as the input-output fidelities of the extended map $\Phi \otimes \text{id}$ when operating on purifications of the density matrices $\rho_Q$. One reminds that a purification of a density matrix $\rho_Q \in \mathcal{S}(\mathcal{H}_Q)$ is any pure state $|\psi\rangle_{QR} \in \mathcal{H}_Q \otimes \mathcal{H}_R$ of the enlarged system formed by $Q$ and by an ancillary system $R$, which fulfills the property $\rho_Q = \text{Tr}_R(|\psi\rangle_{QR}\langle\psi|_{QR})$ (Gühne and Tóth 2009, Horodecki et al. 2009). The entanglement fidelity is then written as

$$F_e(\rho_Q; \Phi) := F(|\psi\rangle_{QR}; \Phi \otimes \text{id}),$$

where $\text{id}$ is the identity map on $R$. It is important to stress that $F_e(\rho_Q; \Phi)$ is independent on the way $|\psi\rangle_{QR}$ is constructed and on the choice of the ancillary system: as a matter of fact, given $\{K_j\}_{j}$ a set of Kraus operators of $\Phi$, it can be expressed as

$$F_e(\rho_Q; \Phi) = \sum_j |\text{Tr}(\rho_Q K_j)|^2.$$  

Operationally the entanglement fidelity functional (50) can be used to detect the detrimental effects on the transmission of half of the entangled state $|\psi\rangle_{QR}$ through the channel $\Phi$. This quantity is related to the input-output fidelity (46) via the inequality

$$F_e(\rho_Q; \Phi) \leq F(\rho_Q; \Phi),$$

implying that values of $F_e(\rho_Q; \Phi)$ close to one force $F(\rho_Q; \Phi)$ to approach unity too. Slightly weaker versions of the opposite implication can also be proven — see e.g. (Holevo 2012, Kretschmann and Werner 2004). In particular given $\epsilon > 0$, if $F(|\psi\rangle_Q; \Phi) \geq 1 - \epsilon$ for all input states $|\psi\rangle_Q$ which belongs to the support of the density matrix $\rho_Q$, then (Barnum, Knill, and Nielsen 2000)

$$F(\rho_Q; \Phi) \geq 1 - 3\epsilon/2.$$  

Furthermore, taking $\rho_Q$ to be the completely mixed state of $Q$, i.e., the density matrix $\mathbb{1}/d$ ($d$ being the dimension of $\mathcal{H}_Q$) whose purification is a maximally entangled state of $QR$, from Eq. (51) it follows that

$$F_e(\mathbb{1}/d; \Phi) = \frac{1}{d^2} \sum_j |\text{Tr}(K_j)|^2,$$

which can be put in correspondence with the average fidelity (49) through the identity (Horodecki et al. 1999, Nielsen 2002)

$$F(\Phi) = \frac{d F_e(\mathbb{1}/d; \Phi) + 1}{d + 1}.$$  

2. Distance measures for channels

Distance measures for quantum channels (and in general for quantum operations) are typically written as $||\Phi - \Psi||$ where $||A||$ denotes a proper norm of the superoperator $\Lambda$. Suitable choices are

$$||A||_k := \sup_{||\Omega||_k \leq 1} ||\Lambda(\Omega)||_k,$$

where the index $k$ identifies a norm for the operators of the system. Specifically for $k = 1$, $||\Omega||_1 = \text{Tr}\sqrt{\Omega^\dagger \Omega}$ is the trace norm; for $k = 2$, $||\Omega||_2 = \sqrt{\text{Tr}\Omega^\dagger \Omega}$ is the Hilbert-Schmidt norm; and finally for $k = \infty$, $||\Omega||_\infty = \sup_{|\psi\rangle} \langle\psi| \Omega |\psi\rangle$ is the standard operator norm — remind that they obey the following ordering $||\Omega||_\infty \leq ||\Omega||_2 \leq ||\Omega||_1$ (Horn and Johnson 1990). While correctly defined, when applied to CPTP maps, the norms (56) result to be unstable under channel extension. In particular $||A \otimes \text{id}||_k$ can explicitly depend upon the dimensionality of ancillary system where the $\text{id}$ channel is defined. In
order to amend this, regularizations have been proposed. Of particular relevance are the norm of complete bound-
ness, or cb-norm (Paulsen 2003), and the diamond-norm
(Kitaev 1997). Given a generic (not necessarily CPTP)
map \( \Lambda : \mathcal{S}(C^n) \mapsto \mathcal{S}(C^k) \) they are defined respectively as

\[
|||\Lambda|||_{cb} := \sup_m |||\Lambda \otimes id_m|||_\infty , \quad (57)
|||\Lambda|||_\diamond := |||\Lambda \otimes id_n|||_1 , \quad (58)
\]

where \( id_n \) denotes the identity channel on \( \mathcal{S}(C^n) \). While not obvious at least in the case of \( |||\cdots||| \), both these norms are stable under channel extension. Furthermore they are related through the identity \( |||\Lambda|||_\diamond = |||\Lambda^*|||_{cb} \), where \( \Lambda^* \) is the dual of \( \Lambda \) (Johnston, Kribs, and Paulsen 2009).

In the context of quantum communication, the properties of the cb-norm have been extensively reviewed in Refs. (Belavkin et al. 2005; Holevo and Werner 2001; Johnston, Kribs, and Paulsen 2009; Keyl 2002; Kretschmann 2003; Kretschmann and Werner 2004, 2005). Here one reminds only that it is well behaved under tensor product composition rule [6], since it fulfills the property

\[
|||\Lambda_1 \otimes \Lambda_2|||_{cb} = |||\Lambda_1|||_{cb} |||\Lambda_2|||_{cb} . \quad (59)
\]

Furthermore if \( \Lambda \) is completely positive then \( |||\Lambda|||_{cb} = |||\Lambda(1)|||_\infty \). Accordingly if \( \Lambda \) is CPTP and \( \Lambda^* \) its dual channel \( \Lambda^* \), one has that \( |||\Lambda|||_{cb} \) can take any value up to \( d \) (the dimension of the channel input space) while \( |||\Lambda^*|||_{cb} \) is always 1.

Finally, another useful distance measure for quantum channels is the one introduced by (Grace et al. 2010) as a distance between unitary operations acting on a bipartite quantum system, where only the effect of the operations on one component (the subsystem of interest) is relevant in the measure, while the effect on the other component (environment) can be arbitrary.

K. Channels and entropies

In the study of quantum communication, entropic quantities play a fundamental role in characterizing quantum channels in terms of their efficiency as communication lines (Barnum, Nielsen, and Schumacher 1998). A comprehensive characterization of these functionals can be obtained moving in the so called “Church of the Larger Hilbert Space”, a construction based on the Stone-spring dilation form [7] where also the input state \( \rho_Q \) of the system \( Q \) is represented as a reduced density operator of a pure state \( |\psi_p\rangle_Q \) of a larger system \( QR \) via a purification – Fig. 1. Let us denote by

\[
S(\rho) := -\mathrm{Tr}(\rho \log_2 \rho) , \quad (60)
\]

the von Neumann entropy of the density operator \( \rho \) (Ohya and Petz 1993; Petz 2008; Wehrl 1978) which generalizes to quantum mechanical systems the Shannon entropy of a classical random variable \( X \), defined as

\[
H(X) := -\sum_x p(x) \log_2 p(x) , \quad (61)
\]

with \( p(x) \) being the probability that \( X \) acquires the value \( x \) (Cover and Thomas 1991; Gallager 1968). Then, given a quantum channel \( \Phi \in \mathcal{F}(Q \mapsto Q') \) and an input state \( \rho_Q \), there are three important entropic quantities related to the pair \((\rho_Q, \Phi)\). First comes the entropy of the input state \( S[\Phi := S(\rho_Q)] \) (input entropy), which exploiting the fact that the purification \( |\psi_p\rangle_Q \) is pure can also be expressed as the entropy of the ancillary system \( R \), i.e., \( S[\Phi] = S[Q] = S[\rho_Q] \); second comes the entropy of the output state \( \Phi(\rho_Q), \) i.e., \( S[\Phi(\rho_Q)] \) (output entropy); finally there is the entropy of exchange (Barnum, Nielsen, and Schumacher 1998; Schumacher 1996) computed as the von Neumann entropy of the environment \( E \) after the interaction with \( Q \), i.e., the entropy measured at the output of the complementary channel \( \Phi \) defined in Eq. (19).

\[
S[E'] = S(\rho_Q; \Phi) := S(\Phi(\rho_Q)) = S[\Phi(\rho_Q)] = S(\rho_Q) \quad (62)
\]

\[
= S[\Phi(\rho_Q)] = S(Q' \mapsto \Phi(\rho_Q)) \quad (63)
\]

\[
\geq S(\rho_Q) \quad (64)
\]

\[
|S(\Phi(\rho_Q)) - S(\rho_Q; \Phi)| \leq S(\rho_Q) \quad (65)
\]

\[
S(\rho_Q; \Phi) \leq h[F_e(\rho_Q; \Phi)] + [1 - F_e(\rho_Q; \Phi)] \log_2 (d^2 - 1) \quad (66)
\]

with \( d \) the dimension of the channel input,

the Shannon binary entropy function (Cover and Thomas 1991; Gallager 1968), and \( F_e(\rho_Q; \Phi) \) the entanglement fidelity introduced in Eq. (50).

The input, output and exchange entropies are the building blocks for constructing several information quantities. For instance one defines the quantum mutual information \( I(\rho_Q; \Phi) \) between the system \( Q \) at the output of the map \( \Phi \) and the system \( R \) which enters in the purification \( |\psi_p\rangle_Q \), i.e.,

\[
I(\rho_Q; \Phi) := S(\rho_Q) + S[\Phi(\rho_Q)] - S(\rho_Q; \Phi) . \quad (67)
\]

This is a non-negative quantity which is known to be concave and sub-additive with respect to \( Q \) (Bennett et al. 2002; Cerf and Adami 1997).
FIG. 1 Pictorial representation of a quantum channel $Φ \in \Psi(Q \rightarrow Q')$ as unitary interaction $U_{QE}$ between the system state $ρ_Q$ and the environmental one $ρ_E$. The action of $Φ \otimes id$ on the purification of $|ψ_ρ⟩_{QR}$ and the complementary map $\tilde{Φ}$ are also shown.

Subtracting $S(ρ_Q)$ from $I(ρ_Q; Φ)$ one also defines the channel **coherent information** (Barannik, Nielsen, and Schumacher 1998; Schumacher and Nielsen 1996).

$$J(ρ_Q; Φ) := S[Φ(ρ)] - S(ρ_Q; Φ) = S[Φ(ρ)] - S[Φ(ρ_Q)] ,$$

where in the last expression it is enlightened that $J(ρ_Q; Φ)$ can also be expressed as the difference between the output entropy of $Φ$ and of its complementary counterpart $Φ$. Differently from $I(ρ_Q; Φ)$ the function $J(ρ_Q; Φ)$ is in general neither non-negative, nor convex or sub-additive (DiVincenzo, Shor and Smolin 1998; Smith and Smolin 2007). Both the quantum mutual and coherent information however satisfy data-processing inequalities. In particular given $Φ_1$ and $Φ_2$ CPTP channels, one has (Holevo 2012)

$$I(ρ_Q; Φ_2 \circ Φ_1) \leq \min\{I(ρ_Q; Φ_1), I(Φ_1(ρ_Q); Φ_2)\} ,$$

while

$$J(ρ_Q; Φ_2 \circ Φ_1) \leq J(ρ_Q; Φ_1) .$$

A further entropic quantity useful for characterizing the channel $Φ$ is the channel Holevo information. At variance with the previous expressions this is a functional of the input ensemble $E := \{p_j; ρ_Q^{(j)}\}_j$ Alice feeds into the channel (here $\{p_j\}_j$ is a probability distribution while $\{ρ_Q^{(j)}\}_j$ is a collection of inputs). Accordingly one has

$$χ(E; Φ) := S[Φ(ρ_Q)] - \sum_j p_j S[Φ(ρ_Q^{(j)})] = \sum_j p_j S(Φ(ρ_Q))∥Φ(ρ_Q^{(j)}) ,$$

where $ρ_Q = \sum_j p_j ρ_Q^{(j)}$ is the average state associated with $E$ and where in the last identity the quantum relative entropy $S(ρ_1∥ρ_2) := \text{Tr}[ρ_1 \log ρ_1 - \log ρ_2]$ (Lindblad 1975; Schumacher and Westmoreland 2000) has been used. Via the Holevo Bound (Holevo 1973 a,b) the quantity $χ(E; Φ)$ provides an upper bound on the information one could retrieve on the random variable $X$ associated with index $j$ of the ensemble $E$ if allowed to measure the corresponding states at the output of the channel $Φ$. Specifically, indicating with $Y$ the random variables associated with the estimation of $j$ after a POVM has been performed on the density matrix $Φ(ρ_Q^{(j)})$, one has

$$I(X : Y) \leq χ(E; Φ) ,$$

with $I(X : Y) := H(Y) + H(X) - H(X, Y)$ being the (Shannon) mutual information associated with the couple $X$ and $Y$ (Cover and Thomas 1991; Gallager 1968). The functional $χ(E; Φ)$ is non-negative, non-subadditive (Hastings 2009), and fulfills the data-processing inequality

$$χ(E; Φ_2 \circ Φ_1) ≤ χ(E; Φ_1) ,$$

for all $Φ_1, Φ_2$ CPTP maps and for all ensemble $E$. A connection with the coherent information can be established via the identity (Devetak 2005; Holevo 2012)

$$χ(E; Φ) - χ(E; \tilde{Φ}) = J(ρ_Q; Φ) - \sum_j p_j J(ρ_Q^{(j)}; Φ) ,$$

with $ρ_Q$ being the average density matrix of the ensemble $E = \{ρ_Q^{(j)}\}_j$.

III. FROM MEMORYLESS TO MEMORY QUANTUM CHANNELS

Having in mind the multi-uses communication scenario detailed at the beginning of Sec. 1 in which a time-ordered sequence of carriers $Q := \{q_1, q_2, \ldots\}$ propagates from Alice to Bob along a noisy channel, this section starts by discussing the simplest case where they are affected by uncorrelated identical maps and then move on to consider correlations among uses, i.e., memory effects.

A. Memoryless quantum channels

**Memoryless quantum channels** describe those scenarios in which the noise acts *identically and independently* on each element of the sequence $Q$. Under this assumption the multi-use map associated with the communication line is expressed as a tensor product of a CPTP map $Φ : \mathcal{E}(\mathcal{H}_q) \rightarrow \mathcal{E}(\mathcal{H}_q)$ that acts on the states of a single carrier $q$. Therefore, indicating as $\mathcal{H}_Q^{(n)} := \mathcal{H}_q^{\otimes n}$, the Hilbert space of the first $n$ carriers of the system, its input density operators $ρ_Q^{(n)} \in \mathcal{E}(\mathcal{H}_Q^{(n)})$ will be mapped into

$$Φ^{(n)}(ρ_Q^{(n)}) = Φ^{\otimes n}(ρ_Q^{(n)}) ,$$

where $Φ$ is a CPTP map.
with $\Phi^{n} := \Phi \otimes \cdots \otimes \Phi$. Equivalently, one can say that the Kraus operators of the memoryless map $\Phi^{(n)}$ can be expressed as a tensor products $K_{q_{1}} \otimes \cdots \otimes K_{q_{n}}$ formed by independent and identically distributed sequences extracted from the Kraus set $\{K_{q}\}_{q}$ associated with the single carrier channel $\Phi$. For an explicit example consider for instance the model discussed in Ref. [Giovannetti 2005]. Here the carriers $Q$ are assumed to propagate from Alice to Bob, one by one and at constant speed, while interacting with an external environmental system via a constant coupling described by the unitary operator $U_{qe} \in B(H_{q} \otimes H_{e})$. The environment $e$ is also assumed to undergo a dissipative process which, on a time-scale $\tau$ tends to reset it into a stable configuration $\omega_{e}$. The memoryless regime is achieved in the limit in which the rate $\nu$ at which the carriers propagate from Alice to Bob is much lower than the inverse of the relaxation time $\tau$, i.e., $\nu \ll 1/\tau$. In this limit in fact each carrier couples with identical and independent environmental states. Defining then $\omega_{E}^{n} := \omega_{e_{1}} \otimes \cdots \otimes \omega_{e_{n}}$, this allows one to write

$$\rho_{Q}^{(n)} \mapsto \text{Tr}_{E} \left[ U_{q_{n},e_{n}} \otimes \cdots \otimes U_{q_{1},e_{1}} \left( \rho_{Q}^{(n)} \otimes \omega_{E}^{n} \right) U_{q_{1},e_{1}}^{\dagger} \otimes \cdots \otimes U_{q_{n},e_{n}}^{\dagger} \right],$$

which reduces to Eq. (75) when identifying $\text{Tr}_{e}[U_{qe}(\cdots \otimes \omega_{e})U_{qe}^{\dagger}]$ with unitary dilation of the single-use channel $\Phi$.

1. Compound and averaged quantum channels

Before entering into the subject of memory quantum channels, let us briefly discuss the situation in which the channel map, though intended as acting like (75), is not perfectly known to the sender and receiver. Such a situation can be modeled by considering not a single CPTP map, but rather a set $\{\Phi_{i}\}_{i}$ of them. Here $\Phi_{i} : \mathcal{S}(H_{q}) \mapsto \mathcal{S}(H_{q})$ and the set $\{\Phi_{i}\}_{i}$ can in principle contain a finite or infinite (countable or not) number of CPTP maps. This leads to the notion of memoryless compound quantum channel, i.e., the family $\{\Phi_{n}^{\otimes n} : \mathcal{S}(H_{q}^{\otimes n}) \mapsto \mathcal{S}(H_{q}^{\otimes n})\}_{n,i}$. Averaged channels are closely related to compound channels. The difference is that in the former the sender and receiver know an a priori probability distribution $\{p_{i}\}_{i}$, governing the appearance of the members of compound channel. It means that for any $n \in \mathbb{N}$ one can write the averaged channel map as

$$\Phi^{(n)}(\rho_{Q}^{(n)}) = \sum_{i} p_{i} \Phi_{i}^{\otimes n}(\rho_{Q}^{(n)}).$$

Equation (77) describes a scenario in which, with some probability $p_{i}$ all the carriers of the system are operated by the same identical local transformation $\Phi_{i}$. The index $i$ can be interpreted as a “switch” selecting different memoryless channels, and (77) as the average channel over different values of the switch. Classical counterparts of compound and averaged channels were studied since long time ago [Ahluwalia 1968 Blackwell 1959 Jacobs 1962 Wolfowitz 1960]. Compound and averaged quantum channels were introduced only recently [Bjelakovic and Boche 2009 Hayashi 2008]. In Sec. III.D.8 one shall see that these channels are closely related to a special set of memory channels having long-term memory.

B. Non-anticipatory memory quantum channels

Whenever the tensorial decomposition of Eq. (75) doesn’t apply, one can speak of memory channels or correlated noise channels. Among the plethora of possibilities, the following will focus only on those configurations which have physical relevance and which have attracted some interest in the recent literature. In particular, one shall treat those models in which the noise respects the time-ordering of the carriers $Q$ so that at a given channel use, the output cannot be influenced by successive inputs. This property generalizes the notion of semicausality discussed in Sec. III.D.9 to the case of a multiple (ordered) subsystems. Inspired by the classical theory of communication [Gallager 1968] one can name the quantum communication lines which fulfills such condition, non-anticipatory quantum channels (notice however that in the approach of Kretschmann and Werner 2005 these maps are called just causal – more on this in Sec. III.C).

Under non-anticipatory condition there must exist a family of CPTP maps $F := \{\Phi^{(n)}; n = 1,2,\ldots\}$ with $\Phi^{(n)} : \mathcal{S}(H_{Q}^{(n)}) \mapsto \mathcal{S}(H_{Q}^{(n)})$ which allows one to express the output states of the first $n$ carriers in terms of the density matrices of their associated inputs, i.e.,

$$\rho_{Q}^{(n)} \mapsto \Phi^{(n)}(\rho_{Q}^{(n)}).$$

Clearly the property (78) requires that the family $F$ must fulfill the minimal consistency requirement that for all $m < n$ the element $\Phi^{(m)}$ should be obtained as a restriction of $\Phi^{(n)}$ over the degree of freedom of the first $m$ carriers. That is, given $\rho_{Q}^{(n)} \in \mathcal{S}(H_{Q}^{(n)})$ and $\rho_{Q}^{(m)} \in \mathcal{S}(H_{Q}^{(m)})$ one must have

$$\Phi^{(m)}(\rho_{Q}^{(m)}) = \text{Tr}^{(m)} \left[ \Phi^{(n)}(\rho_{Q}^{(n)}) \right],$$

whenever $\rho_{Q}^{(m)} = \text{Tr}^{(m)}[\rho_{Q}^{(n)}]$, where $\text{Tr}^{(m)}$ stands for the partial trace over all the carriers but the first $m$.

As already noticed, in the language introduced in Sec. III.D.9 non-anticipatory channels can be classified as semicausal with respect to the natural ordering of the channels uses. The representation of semicausal channels given in Eq. (24) can hence be applied, yielding a representation of non-anticipatory quantum channels where each carrier couples sequentially with a common memory system $M$. The back-action of $M$ on the message state simulates the memory effects of the transmission. Accordingly, all the non-anticipatory CPTP maps can be
expressed as
\[
\Phi^{(n)}(\rho_Q^{(n)}) = \text{Tr}_M \left[ U_{q,M} \cdots U_{q,M} \left( \rho_Q^{(n)} \otimes \omega_M^j \right) \right] U_{q,M}^\dagger \cdots U_{q,M}^\dagger ,
\] (80)
where for all \( j = 1, 2, \ldots, n \), \( U_{q,M} \) is a unitary transformation which describes the coupling of the \( j \)-th carrier with the memory system \( M \), and where \( \omega_M \) is some given state of \( M \), see Fig. 2 a). The unitary transformations \( U_{q_i,M} \) may in general depend on the carrier label \( j \). Otherwise, if they are independent on \( j \) the memory channel has the additional property of being invariant under translation of the carrier labels. An explicit proof of Eq. (80) was first given in Kretschmann and Werner 2005 in the context of quasilocal algebras (see also Appendix B), under the assumption of translational invariance of the noise (more on this will be provided in Sec. III). An alternative proof which doesn’t make use of this hypothesis can be found in Appendix C.

In Eq. (80) \( M \) is in general a large system whose dimension \( d_M \) is an explicit function of \( n \) (in any case it can always be chosen to be less than or equal to \( d^n \) with \( d \) being the dimension of a single carrier). As a matter of fact, as explained in Appendix C one can take \( M \) to be a composite system of components \( m_1, m_2, \ldots, m_n \) whose dimensions can always be chosen to be not larger than \( d^2 \). In this configuration then one can assume \( \omega_M \) to be a pure tensor product state of local terms \( |0\rangle_{m_1} \otimes \cdots \otimes |0\rangle_{m_n} \), and write \( U_{q,M} \) as a transformation which couples the \( j \)-th carrier only with the first \( j \) elements of \( M \), i.e.,
\[
U_{q,M} = 1_{m_n} \otimes \cdots \otimes 1_{m_{j+1}} \otimes U_{q_j M} m_{j-1} \cdots m_1 ,
\] (81)
with \( 1_{m} \) being the identity operator on the \( m \)-component of the environment, see Fig. 2 b).

An alternative, but fully equivalent, representation for non-anticipatory channels is obtained by adding to Eq. (80) a collection of local environments which individually couples with the carriers, i.e.,
\[
\Phi^{(n)}(\rho_Q^{(n)}) = \text{Tr}_{ME} \left[ U_{q,Mn} \cdot \cdots U_{q,Mn} \left( \rho_Q^{(n)} \otimes \omega_M^{\otimes n} \right) \right] U_{q,Mn}^\dagger \cdots U_{q,Mn}^\dagger ,
\] (82)
where for all \( j = 1, 2, \ldots, n \), \( U_{q_j,M} \) is now the unitary transformation which describes the coupling of the \( j \)-th carrier with its own local environment \( e_j \) and with the memory system \( M \), where \( \omega_M^{\otimes n} := \omega_{e_1} \otimes \cdots \otimes \omega_{e_n} \) as in the memoryless case, and \( \omega_M \) is some given state of \( M \), see Fig. 2 c). In principle one can distinguish different setups where Alice, Bob or Eve (third party) has the control of the initial/final states of the memory system \( M \) (Kretschmann and Werner 2005). Equation (82) was first introduced by Bowen and Mancini 2004 as a model for representing correlated channels: from Eq. (80) it follows that it provides a general unitary dilation for every non-anticipatory quantum maps. It can also be expressed in terms of a \( n \)-fold concatenation of a sequence of CPTP maps acting on a single carrier and the memory system \( M \) (Bowen and Mancini 2004 Kretschmann and Werner 2005). Such concatenation is shown pictorially in Fig. 2 c) and results in the following identity
\[
\Phi^{(n)}(\rho_Q^{(n)}) = \text{Tr}_M \left[ \Phi_Q^{(n)}(\rho_Q^{(n)} \otimes \omega_M^j) \right] ,
\] (83)
with
\[
\Phi_Q^{(n)} := \Phi_{q,M} \circ \Phi_{Q,M}^{(n-1)} = \Phi_{q,M} \circ \cdots \circ \Phi_{q,M} ,
\] (84)
where for \( j = 1, 2, \ldots, n \), \( \Phi_{q,M} : \mathcal{S}(\mathcal{H}_q \otimes \mathcal{H}_M) \rightarrow \mathcal{S}(\mathcal{H}_q \otimes \mathcal{H}_M) \) is a CPTP map that operates on the \( j \)-th carrier and on the memory ancilla \( M \) and is defined by the unitary dilation
\[
\Phi_{q,M}(\cdot \cdots \cdot) = \text{Tr}_{e_j} \left[ U_{q,M} e_j \left( \cdots \otimes \omega_{e_j} \right) U_{q,M)^\dagger e_j} \right] .
\] (85)
In this representation the evolution of \( M \) after the interaction with the carriers is provided by the transformation
\[
\omega_M \mapsto \Psi^{(n)}(\rho_Q^{(n)}; \omega_M) := \text{Tr}_Q \left[ \Phi^{(n)}_Q \rho_Q^{(n)} \otimes \omega_M \right] ,
\] (86)
which explicitly depends upon the input state of \( Q \).

Cases of special interest (Kretschmann and Werner 2005) are those in which, for all \( j \), the \( \Phi_{q,M} \) describes the same mapping \( \Phi = \Phi_{q,M} \) on \( \mathcal{S}(\mathcal{H}_q \otimes \mathcal{H}_M) \) which, according to Eq. (84) becomes the generator of the \( n \)-fold concatenation. That characterizes memory channels which are non-anticipatory and translation invariant (i.e., invariant under translation of the information carriers, \( q_j \rightarrow q_{j+1} \)). Memoryless channels can then be included in this class as a limiting case in which the generator \( \Phi \) can be expressed as a tensor product channel that acts independently on the carrier \( q \) and on the memory system \( M \). In terms of the unitary dilation (82) this is equivalent to assume that the unitaries \( U_{q_j,M} \) in Eq. (76) factorize in a tensor product \( U_{q_j e_j} \otimes V_M \), where \( V_M \) is a unitary operator on the memory system and \( U_{q_j e_j} \) acts only on the degree of freedom of the \( j \)-th carrier and on its local environment \( e_j \).

A special subset of non-anticipatory channels is formed by symbol independent (SI) maps (Bowen, Devetak and Mancini 2005). They are communication lines where previous input states do not affect the action of the channel on the current input state. In other words the symbol independent maps are non-anticipatory (or semicausal) with respect to all possible ordering of the carriers (in this sense they are hence fully non-anticipatory). Accordingly, given a generic subsets of the carrier set \( Q \), its output state is uniquely determined by the corresponding input state via a proper CPTP mapping. Following the terminology introduced in Sec III (H) (Beckman et al. 2001 Eggeling, Schlingemann and Werner 2001 Piani et al. 2006), they can be said non-signaling (or causal) channels, meaning that the output states of any subset
where memory quantum channels were always thought of as concatenations of smaller units which, starting from an official “first carrier” element, process one quantum signal at a time. An alternative view where the communication lines are treated as mappings applied on an infinitely long message strings, is proposed in [Bjelaković and Boche 2008, Kretschmann and Werner 2005, Richter and Werner 1996]. This approach requires some advanced mathematical tools that are briefly reviewed in the Appendix E.

To set the stage, suppose to have a quantum channel which transforms input states of an infinitely extended quantum lattice system (representing the infinite message string) into output states on the same system. In [Kretschmann and Werner 2005] this map is formally assigned by working in the Heisenberg picture (see Sec. II.D) via the introduction of a completely positive and unital map $\Phi^*: \mathcal{B}^Z \to \mathcal{A}^Z$, operating on the quasi-local algebras $\mathcal{B}^Z$ and $\mathcal{A}^Z$ [Bratteli and Robinson 1979], that define the observable quantities on the lattice as described by the receiver Bob and the sender Alice, respectively. In this context one says that the channel is translational invariant or (borrowing from [Bjelaković and Boche 2008]) stationary if $\Phi^*$ commutes with the shift operator on the lattice, i.e.,

$$\Phi^* \circ T_B = T_A \circ \Phi^*, \quad (87)$$

($T_B$ and $T_A$ being the representation of the shift operator on $\mathcal{B}^Z$ and $\mathcal{A}^Z$ respectively). Furthermore, $\Phi^*$ is said to be ergodic if it is extremal in the convex set of stationary channels.

Requiring then that future inputs should not affect past measurements, i.e., the non-anticipatory property [79], Ref. [Kretschmann and Werner 2005] introduces the definition of causal channel as a completely positive and unital translational invariant map $\Phi^*$ that fulfills the constraint

$$\Phi^* \left( O^{(-\infty,z]} \otimes 1^{[1+z,\infty]} \right) = \Phi^* \left( O^{(-\infty,z]} \right) \otimes 1^{[1+z,\infty)}, \quad (88)$$

for all $z \in \mathbb{Z}$ and for all $O^{(-\infty,z]} \in \mathcal{B}^{(-\infty,z]}$, where $\mathcal{B}^{(-\infty,z]}$ denotes the set of bounded operators defined on lattice elements up to that associated with the label $z$. In particular, memoryless configurations are obtained when also the condition

$$\Phi^* \left( 1^{(-\infty,z]} \otimes O^{[1+z,\infty)} \right) = 1^{(-\infty,z]} \otimes \Phi^* \left( O^{[1+z,\infty)} \right), \quad (89)$$

applies for all $O^{(-\infty,z]} \in \mathcal{B}^{[1+z,\infty]}$.

Example of causal (non necessarily memoryless) maps [58] are provided by concatenated memory channels [Kretschmann and Werner 2005] which can be easily constructed by adapting the concatenation scheme of

2 It should be noticed that this notion of ergodicity refers to in-series composition of quantum channels and differs from ergodicity of in-series concatenation discussed in [Burgarth and Giovannetti 2007, Burgarth et al. 2013, Raginsky 2002, Richter and Werner 1996] and references therein.

C. Quasi-local algebras approach

Till now one has followed a constructive approach where memory quantum channels were always thought of

FIG. 2 Unitary dilations for a non-anticipatory quantum memory channels. a) graphical sketch of the representations of Eq. (80): here the noise correlations among the $n$ channel uses can be described via a series of concatenated unitary interactions with a common reservoir $M$ whose dimension in general depends (exponentially) upon $n$ ($n = 3$ in the example). Notice that while the carrier $q_1$ might influence the outcome of $q_2q_3$ via their common interaction with $M$, $q_2q_3$ cannot influence the output of the first carrier; b) the environment $M$ can be also represented as a collection of smaller systems $M_1, M_2, \ldots$ initially prepared into a separable state while, as shown in Eq. (81), the unitary transformation operating on the $j$-th channel use couples it with the first $j$ subsystems only; c) unitary dilation [82] where apart form $M$ a series of local environment $e_1, e_2, \ldots$ are also present. In all the diagrams the unitary operators (represented by the white boxes) are applied sequentially on the input states of the global system (i.e., the carriers and the environment) starting for the one on the top of the figure. The carriers and the environmental states evolve, respectively, from-left-to-right and from-top-to-bottom while interacting meeting at a white box. The trash-bin symbol stands for the partial trace operation on the environment.

of the carriers cannot be influenced by the input state of the remaining carriers.

Channels which are not SI are said to exhibit inter-symbol interference (ISI) [Bowen, Devetak and Mancini 2005], that is, the input states of previous carriers affect the action of the channel on the current input. From a physical point of view, in ISI channels there is a non negligible back action of the carrier onto the memory during their interaction. So the carrier’s state (symbol) influences the subsequent actions of the channel. On the contrary, in SI channels the carrier does not influence the memory during their interaction. Usually this happens because the memory is much larger (in terms of degrees of freedom) of the single carrier. A pedagogical example of ISI channels is the quantum shift channel, where each input state is replaced by the previous input state, i.e., given the $j$-th carrier $q_j$ whose state is $\rho_j$, then $\Phi_{q_j}(\rho_j) = \rho_{j-1}$.
Eqs. [83] - [84] to the quantum lattice formalism. Within this context Ref. (Kretschmann and Werner 2005) proves a structure theorem which shows that any map obeying Eq. [88] can always be represented as concatenated memory channels produced by an assigned generator (see previous section).

Although cq-channels can be easily included in the above formalism by expressing them as CPTP maps via the embedding [15], it is worth reviewing the approach adopted in Bjelaković and Boche 2008 to address these special set of maps. Here a cq-channel taking values on the classical alphabet is described as a mapping which to each $x \in X^\mathbb{Z}$ (the set of doubly infinite sequences with components from alphabet $X$) associates a complex value linear functional $W(x, \cdots)$ on $B^\mathbb{Z}$, i.e.,

$$W(x, \cdots) .$$

Ultimately, via the Gelfand-Naimark-Segal correspondence (Bratteli and Robinson 1979), the functional $W(x, \cdots)$ can be identified with a density operator $\rho_x$ defined on the Hilbert space $H$ carrying a representation $\pi$ of the quasi-local algebra (see Appendix B), through the identification $W(x, \cdots) = \text{Tr}[\rho_x \pi(\cdots)]$. In this form the stationary condition (87) of the cq-channel writes as $W(T_{in}x, b) = W(x, T_{b}b)$ for all $x \in X^\mathbb{Z}$ and all $b \in B^\mathbb{Z}$ (here $T_{in}$ and denote the shift operator on $X^\mathbb{Z}$). The causality condition (88) writes instead

$$W(x, b) = W(\bar{x}, b),$$

for $z \in \mathbb{Z}, b \in B^{(-\infty, z]}$ and all $x, \bar{x} \in X^\mathbb{Z}$ ($x_i = \bar{x}_i ; \forall i \leq z$). Similarly, memoryless configurations (89) are recovered when Eq. [91] applies also for all $b \in B^{(-\infty, \infty)}$ and all $x, \bar{x} \in X^\mathbb{Z}$ ($x_i = \bar{x}_i ; \forall i \geq z$).

D. Taxonomy of Non-Anticipatory Quantum Memory Channels

Here one reviews those classes of non-anticipatory quantum channels which have been discussed in the literature.

1. Localizable memory quantum channels

A subset of non-anticipatory quantum channels which represent the natural multi-partite generalization of the localizable maps of Refs. (Beckman et al. 2001; Egelging, Schlingemann and Werner 2001; Piani et al. 2006) reviewed in Sec. II.H has been introduced in Refs. (Giovannetti and Mancini 2005; Plenio andVirmani 2007, 2008). For such models, the mapping [78] is expressed in terms of (not necessarily identical) local unitary couplings with a correlated many-body environmental system $E := \{e_1, e_2, \cdots \}$ — see Fig. 3. These transformations are clearly SI: memory effects appear because, differently from the memoryless case [76], the many-body environment is initialized in a state $\omega^{(n)}_E$ which does not factorize, i.e.,

$$\Phi^{(n)}(\rho_Q^{(n)}) = \text{Tr}_E \left[ U_{q_n e_n} \otimes \cdots \otimes U_{q_1 e_1} \left( \rho_Q^{(n)} \otimes \omega^{(n)}_E \right) U_{q_1 e_1}^\dagger \otimes \cdots \otimes U_{q_n e_n}^\dagger \right].$$

It is worth mentioning that a variant of this model (Rossini et al. 2008) where the local unitary interaction $U_{q_n e_n} \otimes \cdots \otimes U_{q_1 e_1}$ is replaced by a local Hamiltonian coupling between carriers and environments, is neither SI nor non-anticipatory.

An alternative representation for the localizable mappings described by Eq. [92] has been also provided in Ref. (Caruso, Giovannetti and Palma 2010) by generalizing a model presented in Ref. (Ban, Sasaki and Takeoka, 2002; Bowen and Bose, 2004) for memoryless channels. In this approach the channel noise is effectively described as a quantum teleportation protocol (Bennett et al. 1993; Braunstein and Kimble, 1998; Vaidman, 1994) that went wrong because the communicating parties used non optimal resources (e.g., the state they shared was not maximally entangled). In the case of [92] each of the carriers gets teleported independently using the same procedure, the correlations arising from the fact that the communicating parties use as shared resource a correlated many-body quantum state.

2. Finite-memory channels

The expression finite-memory channels (Bowen and Mancini 2004) is used to indicate those non-anticipatory channels which admit a representation of the form [82] with $M$ being finite dimensional. The dimension of the memory is determined by the number of Kraus operators in the single channel expansion. Within the representation [80] examples of finite-memory channels are obtained by assuming that the unitary transformations [81] couple the carriers with no more than a fixed number $k$ of environmental subsystems, the parameter $k$ playing the role of the correlation length of the channel. More precisely for all $j \geq k$ one has,

$$U_{q_j M} = \mathbb{I}_{m_{j-1}} \otimes \cdots \otimes \mathbb{I}_{m_{j+1}} \otimes U_{q_j M_{j-1} \cdots m_{j-k-1} \otimes \cdots \otimes \mathbb{I}_{m_j}},$$

(see Fig. 4 for a graphical representation of the case with $k = 2$). Notice that the case of a memoryless channel can be considered as an extreme example of finite-memory channels, where $k = 1$ and each carrier interacts with a devoted component of the multipartite environment $M$ (specifically, for each $j$, the carrier $q_j$ interacts with $m_j$ only).
FIG. 3 Model for a localizable, fully non-anticipatory quantum memory channel. Here the correlations are introduced by allowing the state of the environment (gray element) to be initially entangled. As in the previous figures white boxes represent unitary couplings while the trash-bin indicates partial trace over the corresponding degree of freedom. These maps are SI and hence non-anticipatory (therefore they also admit unitary dilations of the form described in Fig. 2).

3. Perfect memory channels

Memoryless channels have unitary dilations in which the environment has a dimension which is at least exponentially growing in $n$ (i.e., $\log[\dim\mathcal{H}_E(n)] = n \log d_e$) or, equivalently, by possessing a (minimal) operator sum representations whose Kraus sets contains a number of elements which is exponentially growing in $n$. The same property typically holds also for memory channels with the important exception of the perfect memory channels (Giovannetti, Burgarth and Mancini, 2009; Kretschmann and Werner, 2005). Perfect memory channels are those admitting a representation as in Eq. (82) where the carriers only interact with the memory system, that is, $U_{q_jM_e} = U_{q_jM}U_{e_j}$. The simplest example of such communication lines is obtained by assuming that the memory system $M$ in Eq. (80) does not scale with $n$ and it is finite dimensional. Under this hypothesis the maps $\Phi^{(n)}$ explicitly admit a unitary dilation with an environment (the system $M$) of constant size. A comparison with the dimension of the Hilbert space $\mathcal{H}_Q(n)$ of the information carriers, which grows exponentially with $n$, shows that information cannot be stuck in the channel environment for a long time. As a consequence, in the asymptotic limit of long carrier sequences, no information is expected to be lost to the environment, yielding optimal communication capacity (see Sec. V). A typical example is provided by the shift channel (see Sec. III.B.) which can be described as in Eq. (80) by assuming $M$ to have the same dimension of a single carrier and by taking $U_{q_jM}$ as the SWAP operator. It is worth noticing that in Ref. (Bowen, Devetak and Mancini, 2005) it was conjecture that the types of memory channels that, analogously to the shift channel, display only intersymbol interference may be represented as perfect memory channels.

More generally the class of perfect memory channels can be extended to include all the CPTP maps (78) that admit unitary dilations (80) in which the dimension $d_M$ of the environmental system $M$ is sub-exponential in $n$, i.e.,

$$\lim_{n \to \infty} \frac{1}{n} \log[d_M] = 0.$$  \hspace{1cm} (94)

As discussed explicitly in Sec. V.C.2 also in this case the channel is asymptotically noiseless (Giovannetti, Burgarth and Mancini, 2009; Kretschmann and Werner, 2005).
4. Markovian channels

An important class of non-anticipatory quantum channels is given by the channels with Markovian correlated noise. They describe noise models in which the carriers are transformed via the applications of strings of local CPTP maps whose elements are randomly generated by a classical Markov process. Explicitly, Markovian channels admit the following representation:

\[
\Phi^{(n)}(\rho_Q^{(n)}) = \sum_{i_1, \ldots, i_n} p_{i_1, \ldots, i_n}^{(n)} \rho_Q^{(n-1)}_{i_1} \cdots \rho_Q^{(2)}_{i_2} \cdots \rho_Q^{(1)}_{i_1} \times \Phi_{i_1}^{(i_1)}(\cdots \times \Phi_{i_n}^{(i_n)})
\]

where \( \{ \Phi_{i_j}^{(j)} \}_{i_j} \) is a set of CPTP maps operating on the \( j \)-th carrier, where \( p_{i_1}^{(1)} \) is an initial probability distribution, and where finally for \( j \geq 2 \), the \( p_{i_j}^{(j)} \) are conditional probabilities.

The mapping (95) is SI (see Sec. III.D) since modifying the input state of previous (or subsequence) channels use does not have any effect on the output states of the carriers that follow (or precede). A unitary dilution of the form \( \frac{1}{\sqrt{2}} \) can be obtained by identifying the initial state \( \omega_M \) of the memory \( M \) with the pure vector \( \sum_i \sqrt{p_i^{(1)}} |i\rangle_M \), and by taking the unitary \( U_{q,M,e} \) in such a way that for all vectors \( |\psi\rangle_{q_j} \) of the \( j \)-th carriers one has

\[
U_{q,M,e} |\psi, i', 0\rangle = K_{j}^{(i)}(\ell) |\psi, i, i'\rangle,
\]

and

\[
U_{q,M,e} |\psi, 0\rangle = \sum_i \sqrt{p_i^{(1)}} K_{j}^{(i)}(\ell) |\psi, i, i'\rangle,
\]

for \( j \geq 2 \) (in the above expressions \( \{ K_{j}^{(i)}(\ell) \}_{i,j} \) is a set of Kraus operators for \( \Phi_{i_j}^{(j)} \); \( |\psi, i, i'\rangle \) stands for the state \( |\psi\rangle_{q_j} |i\rangle_M |\psi, i, i'\rangle_e_j \), while \( |i\rangle_M \) and \( |\psi, i, i'\rangle_e_j \) are orthonormal basis for the memory systems \( M \) and \( e_j \) respectively).

Most of the analysis conducted so far focused on the special case of homogeneous Markovian processes in which both the \( p_{i_j}^{(j)} \) and the \( \Phi_{i_j}^{(j)} \) do not depend upon the carrier label \( j \) (i.e., \( p_{i_j}^{(j)} := p_i^{(i')} \)). Under these conditions one also says that the quantum Markov process is regular if the corresponding classical Markov process \( p_i^{(i')} \) is regular, i.e., if some power of the transition matrix \( \Gamma \) (whose entries are the transition probabilities \( p_i^{(i')} \)) has only strictly positive elements. In this case, for \( j \to \infty \) the statistical distribution of the local noise converges to a stationary distribution \( p_i^{(\infty)} := \lim_{n \to \infty} p_i^{(n)} \), with

\[
p_i^{(n)} := \sum_{i'} (\Gamma^{n-1})_{i,i'} p_i^{(1)}
\]

being the probability of getting \( \Phi_{i_j}^{(j)} \) on the \( j \)-th carrier. The initial probability \( p_i^{(1)} \) is said to be stationary if it satisfies the eigenvector equation \( \sum_{i'} \Gamma_{i,i'} p_i^{(1)} = p_i^{(1)} \) (when this happens \( p_i^{(j)} = p_i^{(1)} \) and the local statistical distribution of \( \Phi_{i_j}^{(j)} \) is identical for all the carriers).

The first example of a regular Markov process has been analyzed by [Macchiavello and Palma 2002]. Here the carriers are assumed to be qubits and the CPTP transformations \( \Phi_q^{(j)} \) entering in Eq. 95 are unitary rotations \( \Phi_q^{(j)}(\cdots) := |\sigma_{i,q}\rangle \langle \sigma_{i,q}| \) where \( \sigma_{0,q} = I \) is the identity operator while for \( i = x, y, z \), \( \sigma_{i,q} \) is the Pauli matrix. The conditional probability \( p_{i,i'} \) which describes the associated classical Markov process was finally written as

\[
p_{i,i'} = (1 - \mu) p_i^{(1)} + \mu \delta_{i,i'},
\]

where \( \mu \in [0,1] \) is a correlation parameter (notice that for \( \mu = 0 \) the model describes a memoryless channel while for \( \mu = 1 \) it describes a long-term memory channel – see Sec. III.D.8). This model of Markovian correlated Pauli channel shows a remarkable feature when it is used for the transmission of classical information (see Sec. V). That is, when two successive uses of the channel are considered, classical information is optimally encoded in either separable states or maximally entangled states, depending whether the correlation parameter \( \mu \) is below or above a certain threshold value. This feature was first conjectured in [Macchiavello and Palma 2002], then proven for certain instances of the model in [Macchiavello, Palma and Virmani 2004], and finally proven for general Markovian correlated Pauli channels in [Daems 2006]. Remarkably, this effect is at the root of the superadditivity property of memoryless quantum channels for transmitting classical information [Hastings 2009] (see Sec. V.B.3).

An experimental demonstration of the optimality of entangled qubit pairs for encoding classical information through a correlated Pauli channel was provided by [Banaszek et al. 2004] for mechanically induced correlated birefringence fluctuations, which in turn induce correlated depolarization [Ball, Dragan, and Banaszek 2004].

A generalized model of \( d \)-dimensional Markovian correlated Pauli channel was considered by [Shadman et al. 2011] for the problem of sending classical information using a dense-coding protocol. An alternative model of two-qubit correlated channel was characterized by [Caruso et al. 2008b] in terms of the minimum output entropy.

Going beyond the case of two uses of a qubit channel, Markovian correlated depolarization over an arbitrary number of channel uses was studied in [Demkowicz-Dobrzański, Kolenderski and Banaszek 2007; Karimipour and Memarzadeh 2006b], and the case of Markovian correlated noise in higher dimensional quantum systems was considered in [Karimipour and Memarzadeh 2006b; Karpov, Daems and Cerf 2006b]. Generally speaking, the optimality of entangled states for encoding classical information can be interpreted in terms of a decoherence-free subspace (see Sec. IV) associated to the correlated noise model: this has been considered for the Hilbert space defined by multiple uses of a qubit channel.
5. Fixed–point channels

Within the representation \[83\] a channel is said to be a fixed–point memory channel (Bowen, Devetak and Mancini 2005) if the initial memory state \(\omega_M\) of the representation is left invariant after each interaction with the carriers. Specifically, recalling the definition \[80\] this notion is formalized by the following identity

\[\Psi(n)(\rho_Q^{(n)};\omega_M) = \omega_M, \quad \forall \rho_Q^{(n)} \in \mathcal{G}(\mathcal{H}_Q^{(n)}).\]  

(100)

Fixed–point channels can be easily shown to be symbol independent while the opposite is not necessarily true. Indeed from Eq. \[84\] one has that the output state of \(n\)-th carrier \(\rho_Q^{(n)} := \text{Tr}_{Q(n-1)}[\Phi^{(n)}(\rho_Q^{(n)})]\) can be expressed as

\[\rho_Q^{(n)} = \text{Tr}_{Q(n-1)M}[\Phi_{q,n,M} \circ \Phi^{(n-1)}_{Q,M}](\rho_Q^{(n)} \otimes \omega_M)\]

\[= \text{Tr}_M[\Phi_{q,n,M} \circ \Psi^{(n-1)}(\rho_Q \otimes \omega_M)]\]

\[= \text{Tr}_M[\Phi_{q,n,M}(\rho_Q \otimes \omega_M)],\]  

(101)

which only depends upon the reduced density operator \(\rho_Q\), and not on the previous information carriers (in these expressions \(\text{Tr}_{Q(n-1)M}\) indicates the partial trace with respect to \(M\) and the first \((n-1)\) carriers).

Markovian memory channels are examples of fixed–point memory channels, where the memory system can be represented by the classical variable of the underlying Markov chain. Being classical, the memory system can be chosen in such a way that is unaffected by the back-action of the input system. This representation can be made explicitly by choosing a unitary dilation of the form \[86\]. An other example is provided by (Plenio and Virmani 2007 2008), in which the input system interacts with the memory system by a controlled-unitary transformation, where the memory is the control and the system is the target. In this setting, the resulting memory channel is a fixed–point one if the initial state of the memory is diagonal in the control basis. While in the model considered in (Plenio and Virmani 2007 2008) the channel is both fixed–point and symbol independent, one can easily build models of memory channels which are fixed–point but not symbol independent. For instance, that can be obtained if the controlled unitary mixes different information carriers.

6. Indecomposable and forgetful channels

An indecomposable channel is one where, for each channel input, the long-term behavior of the channel is independent of the initial memory state (Bowen, Devetak and Mancini 2005). Such independence can be quantified by evaluating the distance between different trajectories \(\Psi(n)(\rho_Q^{(n)};\omega_M)\) and \(\Psi(n)(\rho_Q^{(n)};\omega_M')\) associated through Eq. \[86\] to two different initial memory configurations \(\omega_M\) and \(\omega_M'\). Specifically a finite–memory quantum channel is said to be indecomposable if for any input state \(\rho_Q^{(n)}\) and \(\epsilon > 0\) there exists an \(N(\epsilon)\) such that for \(n \geq N(\epsilon)\),

\[D\left[\Psi(n)(\rho_Q^{(n)};\omega_M), \Psi(n)(\rho_Q^{(n)};\omega_M')\right] \leq \epsilon,\]  

(102)

for any pair of initial state of the memory \(\omega_M, \omega_M'\) (here \(D\) is the trace distance – see Eq. \[A1\]). Equivalently Eq. \[102\] can be stated by saying that for large \(n\), \(\Psi(n)(\rho_Q^{(n)};\omega_M)\) converge to a state of \(M\) which depends on \(\rho_Q^{(n)}\) but not on \(\omega_M\) (compare this with the behavior \[100\] of the fixed-point memory channels). Here one notices that for finite dimensional system this implies that exists a family of CPTP channels \(\Theta(n) : \mathcal{G}(\mathcal{H}_Q^{(n)}) \to \mathcal{G}(\mathcal{H}_M)\) which fulfills the identity

\[\text{Tr}_Q[\Phi_{Q,M}^{(n)}(O_{Q,M}^{(n)})] \to \Theta^{(n)}(O_Q^{(n)})\]  

(103)

in the limit \(n \to \infty\) for all the operators \(O_{Q,M}^{(n)}\) on \(Q\) and \(M\), with \(O_Q^{(n)} = \text{Tr}_M[O_{Q,M}^{(n)}]\). In the Heisenberg picture – see Sec. 3D – this also can be stated as

\[\Phi_{Q,M}^{(n)}(1_Q \otimes \ldots) \to \Theta^{(n)}(\cdot\ldots) \otimes 1_M,\]  

(104)

with \(\Phi_{Q,M}^{(n)}\) and \(\Theta^{(n)}\) being the dual of \(\Phi_{Q,M}^{(n)}\) and \(\Theta^{(n)}\), respectively.

The main features of indecomposable channels have been revisited through the notion of forgetful channels (Kretschmann and Werner 2005). The latter has been originally introduced in the quasi-local algebra approach detailed in Sec. 3C, where the quantum memory channels are assumed to be translation invariant and non-anticipatory. In the representation \[83\] this definition coincides with the limiting condition \[103\], which in Ref. (Kretschmann and Werner 2005) is written in terms of the cb-norm distance (see Sec. 3I.2). In this context a memory channel is said to be strictly forgetful if there exists a finite integer \(m\) such that the rhs and the lhs of Eq. \[103\] exactly coincides for all \(n \geq m\). A simple example of forgetful channels can be obtained if
\( \Phi \) is defined by the concatenation of a generator map \( \Phi_{QM} := p \text{id} + (1 - p) \text{SWAP} \), where \( p \in [0,1] \), and SWAP denotes the swap channel which exchanges \( q \) and \( M \). In this case the only way for \( \Phi^{(n)} \) to not be forgetful is to choose the ideal channel in every step of the concatenation. However, the probability for this event vanishes as \( p^n \), implying that Eq. (103) holds.

Several criteria for a quantum memory channel to be forgetful have been proposed \cite{KretschmannWerner}. For instance, a sufficient condition is that the cb-norm distance between the rhs and the lhs of Eq. (103) falls below 1 for some finite \( n \). From a physical point of view, one could expect a generic quantum memory channel to be forgetful. Indeed, it can be proven that the subset of forgetful channels is dense and open (according to the topology induced by the cb-norm) \cite{KretschmannWerner}.

In the case of Markovian channels, the forgetfulness is determined by the asymptotic properties of the underlying Markov chain: in particular, for a discrete variable memory system, the channel is forgetful if and only if the underlying Markov chain converges to a unique stationary state \cite{DattaDorlas}. On the other hand, if the memory system is described by continuous variables, one could have situations in which the Markov chain has a unique stationary state, yet the convergence property in cb-norm is not verified. To overcome this limitation, a weaker notion of forgetfulness, named weak forgetfulness, was introduced in \cite{LupoMemarzadeh} for Markovian channel. Although restricted to this setting, its definition coincides with that of indecomposability \cite{Lupo}, and is equivalent to forgetfulness for discrete variable Markov chain. Beyond this setting, the model of Gaussian memory channel in \cite{Lupo} was proven to be indecomposable under restricted conditions on the memory initialization, e.g., if the initial state of the memory is a Gaussian state with finite first and second moments. Finally, the relation between the forgetfulness of the channel and the chaotic quantum evolution of the memory system was studied in \cite{BarretoLemosBjelakovic} for a model of dephasing channel with memory.

7. Decaying input memory cq-channels

Indecomposable/forgetful channels represent configurations where the effect of the far past inputs do not affect much present and future outputs. Within the quasi-algebra approach a similar notion has been developed in \cite{BjelakovicBoche} for the special class of cq-channels \cite[Sec. III.C]{Mancini}. Specifically a cq-channel defined by Eq. (90) is said to have decaying input memory if for each \( \epsilon > 0 \) there exists a non-negative integer \( m(\epsilon) \) such that

\[
|W(x, b) - W(x', b)| \leq \epsilon, \quad (105)
\]

for all \( b \in \mathcal{B}^{[n,\infty]} \), \( n \in \mathbb{Z} \), whenever \( x_i = x'_i \) for \( n - m \leq i \) and \( m \geq m(\epsilon) \). Notice that \( \sum_n |W(x, b) - W(x', b)| \) is a distance (between quantum states). Then (105) says that, starting from the \( n \)-th use, the outputs of two identical channels \( W \) are almost (i.e., within a distance \( \epsilon \)) the same provided that the inputs have started to be different from the \( n - m(\epsilon) \)-th use. Hence \( m(\epsilon) \) gives an estimation of the memory length.

This provides a ‘continuity’ property of the channel which plays a crucial role in establishing coding theorems, an idea that will also appear in Section IV.D and goes back to the classic paper by McMillan \cite{McMillan}.

8. Long-term memory channels

Long-term quantum memory channels describe those communication lines in which the effect of the memory does not decay with the channel uses. These channels are defined as those memory channels which are not forgetful.

Extreme examples are provided by statistical mixture of memoryless channels \cite{Norris} of Sec. III.A.1 as pointed out in \cite{DattaDorlas} \cite{DattaSuhov}. The memory correlations of this class of channels can be considered to be given by a Markov chain which is aperiodic but not irreducible \cite{Norris}. This can be easily seen by noticing that Eq. (95) reduces to Eq. (77) by setting \( p_{ij}^{(j)} = \delta_{ij} \) for all \( j = 2, \ldots n \). Hence, once a particular branch, \( i = 1, \ldots, M \), has been chosen, the successive inputs are sent through this branch (aperiodicity) and transition between the different branches (which correspond to the different states of the Markov chain) is not permitted (reducibility).

It is worth stressing that the transformation (77) is fully non-anticipatory (i.e., the input state of any subset of carriers cannot influence the output state of the remaining ones): as a consequence, fixing an ordering, it can always be represented as in Fig. 2 with a proper choice of the unitary couplings.

IV. QUANTUM CODES

Coding theory is the branch of information science which studies how to use software strategies to contrast the effect of an assigned noise source which is affecting a communication line or the components (memory elements) of a database. In a sense it can be described as the last resort which can be exploited once no further improvements can be obtained at the level of hardware engineering.

Both in the classical and in the quantum setting, the key idea to prevent the corruption of information, is to use redundancy: by properly spreading a given message over many information carriers instead of a single one, one can take advantage from the structural properties of noise source. Consider for instance the paradigmatic case where one wishes to store the information contained
in (say) \( k \) information carriers, affected by an assigned error model described by the channel \( \Phi^{(k)} \), into a larger set of \( n \geq k \) carriers, affected by the noise \( \Phi^{(n)} \). Then a generic coding strategy consists in identifying an encoding CPTP map \( \Phi^{(k \to n)}_E : \mathcal{S}(\mathcal{H}^{\otimes k}) \to \mathcal{S}(\mathcal{H}^{\otimes n}) \), shuffling a generic state from the smaller space of the \( k \) carriers to the larger space of the \( n \) carriers, and a decoding CPTP map \( \Phi^{(n \to k)}_D : \mathcal{S}(\mathcal{H}^{\otimes n}) \to \mathcal{S}(\mathcal{H}^{\otimes k}) \), moving the information back to the original set, under the requirement that the resulting channel \( \Phi^{(n \to k)}_D \circ \Phi^{(n)} \circ \Phi^{(k \to n)}_E \) is somehow “less” noisy than the original transformation \( \Phi^{(k)} \) – see Fig. 2. The image \( \mathcal{Q} \) of \( \Phi^{(k \to n)}_E \) is called quantum error correcting code (QECC): it represents the information vault where messages are supposed to be stored/transmitted and, therefore, to prevent the noise from affecting them. The ratio \( R = k/n \) represents the communication rate of the code whose inverse measures how much information spread is involved in the procedure. It worth noticing that if \( \Phi^{(k \to n)}_E \) is taken to be an isometry (an option which often implicitly assumed in QECC) the space \( \mathcal{Q} \) becomes a proper vector subspace of \( \mathcal{H}^{\otimes n} \) of dimension \( d^k \), its elements being the codewords of the code (here \( d \) is the local dimension a single carrier). When this happens \( \Phi^{(n)} \circ \Phi^{(k \to n)}_E \) is just a restriction of \( \Phi^{(n)} \) on \( \mathcal{Q} \) and the encoding mapping can be fully specified by simply assign the latter. This also justifies the consideration of a recovery map \( \Phi^{(n)}_R \) in place of the decoding map \( \Phi^{(n \to k)}_D \) (in the following this simplification will be assumed).

Different equivalent ways have been devised to evaluate the quality of a given coding procedure. The most commonly used is by means of the input-output fidelities introduced in Sec. [4]. In particular good choices are the minimum or average fidelity functionals, i.e.,
\[
F_{\min}(\Phi^{(k \to n)}_D \circ \Phi^{(n)}_E \circ \Phi^{(k \to n)}_E) \quad \text{and} \quad F(\Phi^{(n \to k)}_D \circ \Phi^{(n)}_E \circ \Phi^{(k \to n)}_E).
\]
It is important however to distinguish between two different scenarios: the case where the messages to be stored/transmitted are classical, and the case where instead they are purely quantum. In the first scenario the minimization (resp. average) involved in Eq. (48) (resp. Eq. (49)) needs not to be performed over the entire input space of the \( k \) carriers but only with respect to the orthogonal set of states which in \( \mathcal{H}^{\otimes k} \) encode the classical messages one wishes to protect. Vice-versa in the second scenario, which imply the possibility of producing arbitrary superposition of the input signals, the minimization (resp. average) is performed over the whole input space \( \mathcal{H}^{\otimes k} \). In this last configuration the effectiveness of a correcting code can also be quantified by other distance measures, like the entanglement fidelity.

The second scenario, which imply the possibility of producing arbitrary superposition of the input signals, the effectiveness of a correcting code can also be quantified by other distance measures, like the entanglement fidelity. It can be shown that in the nondegenerate case different error operators map the code into orthogonal subspaces, while for degenerate codes it may happen that distinct error operators transform the code into non-orthogonal subspaces.

A. Standard quantum coding theory

Consider a memoryless quantum channel \( \Phi^{(n)} = \Phi^{(\otimes n)} \), characterized by a set of Kraus operators \( \{K_i\} \), on the Hilbert space \( \mathbb{C}^{2 \otimes n} \) of \( n \) qubits
\[
\rho \mapsto \Phi^{(n)}(\rho) = \sum_i K_i^\dagger \rho K_i, \tag{106}
\]
where \( K_i = K_{i_1} \otimes \ldots \otimes K_{i_n} \) describes i.i.d. errors on single qubits. It is possible to show that \( \mathcal{Q} \) is able to correct errors belonging to a subset \( \mathcal{Q} \subseteq \{K_i\} \) iff there exists a Hermitian matrix \( S \) such that
\[
P_Q K_i^\dagger S K_m P_Q = S m K_i, \tag{107}
\]
for any pair of error operators \( K_i, K_m \in \mathcal{Q} \) [Knill et al. 2002]. \( P_Q \) denoting the projector into \( \mathcal{Q} \). The pair \( (\mathcal{Q}, \Omega) \), consisting of a quantum code \( \mathcal{Q} \) and a set of error operators \( \Omega \), is called degenerate if the matrix \( S \) in (107) is singular; otherwise, \( (\mathcal{Q}, \Omega) \) is said to be nondegenerate. It can be shown that in the nondegenerate case different error operators map the code into orthogonal subspaces, while for degenerate codes it may happen that distinct error operators transform the code into non-orthogonal subspaces [Knill et al. 2002].

For the sake of presentation, here it is assumed that the information carriers are qubits. In this context a \([n,k]\) quantum correcting code \( \mathcal{Q} \) is given by a 2-dimensional subspace of \( \mathbb{C}^{2 \otimes n} \) encoding \( k \) logical qubits into \( n \) physical qubits \( (n \geq k) \). One can assume the error operators to be proportional to the Pauli operators acting on the \( j \)-th qubit and corresponding to no-error, bit-flip error, bit-phase-flip error and phase-flip error, i.e., \( K_0 \propto I, K_{x,j} \propto \sigma_{x,j}, K_{y,j} \propto \sigma_{y,j}, K_{z,j} \propto \sigma_{z,j} \). This restriction to Pauli errors represents no loss of generality. Indeed, it can be easily shown that if a code corrects a given set of errors it can also correct any linear combination (by complex coefficients) [Nielsen and Chuang 2000]. It is
hence sufficient to restrict to Pauli operators since they are a basis on the space of qubit operators. Given a subset $\Omega \subset \{ K_i \}$, that can be corrected one says that the code $Q$ has $\Omega$-correcting ability. To each $K_i$ one can assign a weight $t$, an integer $0 \leq t \leq n$ denoting the number of qubit where operators $K_{ij}$ ($j = 1, \ldots, n$) act differently from identity. Then, the correction ability of $Q$ can be also expressed by specifying the value of the distance $d = 2t + 1$ of the code, meaning that $Q$ corrects all errors affecting at most $t$ qubits.

Suppose that errors are i.i.d. with probability $p_e$ on each qubit, then for any of the $\binom{n}{t+1}$ ways of choosing $t+1$ locations, the probability that errors occur at every one of those locations results $p_e^{t+1}$ (disregarding whether additional errors occur in the remaining $n-t-1$ locations). Therefore one has the following upper bound on the probability that at least $t+1$ errors occur in the block of $n$ qubits $\binom{n}{t+1}p_e^{t+1}$. This means that for $p_e$ small the performance of the code $1-F \approx O(p_e^{t+1})$ is substantially improved over the unprotected data $1-F \approx O(p_e)$.

An upper bound on the rates achievable by non-degenerate quantum codes is given by the quantum version of the (classical) Hamming bound [Ekt and Macchiavello (1996)]

$$2^k \sum_{i=0}^t 3^i \binom{n}{i} \leq 2^n,$$

which for large $n$ and $\frac{d}{n}$ fixed yields the approximate bound $R \leq 1 - \frac{d}{n} \log_2 3 - h(\frac{d}{n+1})$ with $h$ the binary entropy [60].

There are also upper bounds that apply to all quantum codes, not just non-degenerate ones, like the quantum Singleton bound [Knill and Laflamme (1997)]

$$n \geq 4t + k.$$  (109)

On the other hand a lower bound on the rates, confirming that good codes indeed exists [Calderbank et al. (1997)], comes from the quantum version of the Gilbert-Varshamov theorem stating that a $[n,k]$ quantum code of distance $d = 2t + 1$ exists with

$$k \geq \max \left\{ k' \left| 2^{k'} \sum_{i=0}^{2t} 3^i \binom{n}{i} \leq 2^n \right. \right\}.  \quad (110)$$

For large $n$ and $d/n$ fixed one gets the approximate bound $R \geq 1 - \frac{d}{n} \log_2 3 - h(\frac{d}{n})$.

Unfortunately the explicit construction of quantum codes is not an easy task. Historically the first quantum code that appeared was a $[9,1]$ code with $d = 3$ [Shor (1995)] whose basis codewords read

$$\left[ \frac{1}{\sqrt{2}} (|000\rangle + |111\rangle) \right] \otimes^3, \quad \left[ \frac{1}{\sqrt{2}} (|000\rangle - |111\rangle) \right] \otimes^3.$$  (111)

Its construction relies on simple argument. A three qubit code would suffice to protect against single bit flip. The reason the three qubit clusters are repeated three times is to protect against phase errors as well.

Then attempts were made following classical linear codes [Hill (1985)]. A classical $[n,k]$ linear code $C$ is a subspace of the $n$-times product of finite binary field $F_2$ representing a bit, i.e., $C \subseteq F_2^n$. The subspace is characterized by a generator $k \times n$ matrix $G$ (whose rows generate the codewords), or equivalently by a parity check $(n-k) \times n$ matrix $H$ such that $Hv^\top = 0$, $\forall v \in C$. Errors $e_i$, taking $v \in C$ into $v + e_i$ can be detected by applying the parity check $H(v + e_i)^\top = H e_i^\top = \text{synd}(e_i)$ (error syndrome) and corrected iff they give rise to distinct syndromes, i.e., $H(e_i + e_j)^\top \neq 0$ for $i \neq j$. If $C$ is with distance $d = 2t + 1$ it means that it is able of correcting up to $t$ errors, i.e., bitflip errors in at most $t$ bits. The set of errors correctable by $C$ will be denoted by $\mathcal{E}$.

A great advantage of linear codes over general error correcting codes is their compact specification. Using them, quasi classical [or CSS (Calderbank, Shor, Steane)] codes were constructed in the following way [Calderbank and Shor (1996), Steane (1996)]. Consider two classical linear codes $C_1$ and $C_2$ such that $C_2^\perp \subseteq C_1$, where $C_i^\perp$ is the dual code to $C_2$, i.e., consisting of those bit strings that are orthogonal to the codewords of $C_2$. If $C_1$ is a $[n, k_1]$ code with distance $d_1$ and $C_2$ is a $[n, k_2]$ code with distance $d_2$, then the corresponding CSS quantum code is a $[n, k_1 + k_2]$ code with distance $\min(d_1, d_2)$. Its basis codewords are

$$\frac{1}{\sqrt{|C_1^\perp|}} \sum_{u \in C_1^\perp} |u + w\rangle, \quad u \in C_1.$$  (112)

Performing the Hadamard transform on each qubit of the code one can switch from the $\sigma_z$ to $\sigma_x$ basis, hence from $C_1$ to $C_2$, so that the basis codewords become

$$\frac{1}{\sqrt{|C_2^\perp|}} \sum_{w \in C_2^\perp} |u + w\rangle, \quad u \in C_2.$$  (113)

Then to correct errors one can implement the parity check of $C_1$ by measuring in the $\sigma_z$ basis and that of $C_2$ by measuring in the $\sigma_x$ basis. An example along this line is provided by the $[[7,1]]$ code with $d = 3$ [Steane (1996)].

Another, more general, way to construct quantum codes is to exploit the group structure of the set of errors as it is done for stabilizer codes [Gottesman (1997)]. The method can be summarized as follows. First notice that the set of all possible errors on $n$ qubit, can be written
as
\[
P_n := \left\{ \pm \prod_{j=1}^{n} \sigma_{x,j}^{X(j)} \prod_{j=1}^{n} \sigma_{z,j}^{Z(j)} | X(j), Z(j) \in \mathbb{F}_2 \right\} ,
\]
and forms a multiplicative group known as Pauli group. Getting rid of the ± sign in front of any error operator, one can relate the elements of \( P_n \) to 2\( n \)-dimensional binary vectors
\[
e \in P_n \iff (e_X | e_Z) \equiv e \in \mathbb{F}_2^n \times \mathbb{F}_2^n ,
\]
where \( e_X \) (resp. \( e_Z \)) is the \( n \)-bits vector of components \( X^{(j)} \) (resp. \( Z^{(j)} \)) specifying on which qubits the \( \sigma_x \) (resp. \( \sigma_z \)) error occurs and \((e_X | e_Z)\) is the joint \( e_X \), \( e_Z \) vector. Then, one considers an Abelian subgroup \( G \subseteq P_n \)
\[
G = \text{span} \{ g_i \in P_n | 1 \leq i \leq n - k \} ,
\]
where \( g_1, g_2, \ldots, g_{n-k} \) are independent of each other. The set of vectors stabilized by \( G \) forms the quantum code
\[
Q = \{ | x \rangle \in \mathbb{C}^{2^{2n}} | g(x) = | x \rangle, \forall g \in G \} .
\]
By using the vectors in \( \mathbb{F}_2^n \times \mathbb{F}_2^n \) corresponding to the generators \( g_1, g_2, \ldots, g_{n-k} \) it is possible to write down the following \((n-k) \times 2n\) matrix
\[
H := \begin{pmatrix}
g_1.x & g_1.z \\
\vdots & \vdots \\
g_{n-k}.x & g_{n-k}.z
\end{pmatrix} ,
\]
whose \( j \)-th row is given by the vector \((g_j, x | g_j, z)\).

The subspace \( Q \) is a \([n, k]\) code. Provided \( H \) is totally singular, i.e., \( g_i.x \cdot g_j.z + g_j.x \cdot g_i.z = 0, \forall i, j = 1, \ldots, n-k \), there is a \([2n, k]\) classical linear code \( C \) with \( H \) its parity check matrix. Letting \( \Omega \subseteq P_n \) be a subset of quantum errors and \( \mathcal{C} \subseteq \mathbb{F}_2^n \times \mathbb{F}_2^n \) the corresponding subset according to (115), then the analysis of \( \Omega \)-correction ability of \( Q \) can be traced back to that of \( \mathcal{C} \) for \( \mathcal{C} \) in the classical framework.

In this way it results (Gaitan 2008) that \( Q \) has \( \Omega \)-correcting ability iff for every \((e_{1.X} | e_{1.Z}), (e_{2.X} | e_{2.Z}) \in \mathcal{C} \) it is
\[
H(e_{1.X} + e_{2.X} | e_{1.Z} + e_{2.Z})^\top \neq 0 .
\]
The above condition states that \( \Omega \) is correctable by \( Q \) iff \( \text{synd}(e_1) \neq \text{synd}(e_j) \) for all \( e_1, e_j \in \mathcal{C} \) (which corresponds to (107)).

A stabilizer code of distance \( d \) has the property that each element of \( P_d \) of weight \( t \) less than \( d \) either lies in the stabilizer or anti commute with some element of the stabilizer. An example is provided by the \([5, 1]\) code with \( d = 3 \) introduced in (Laflamme et al. 2009) and saturating the quantum Hamming bound (108). It is worth remarking that a systematic method to find stabilizer generators exists based on the connection with vectors over Galois field \( GF(4) \) (Calderbank et al. 1998).

If the subgroup \( S \subseteq P_n \) is not Abelian, it can be as well to construct a QECC provided that entanglement between encoder and decoder is available (Brum, Devetak, and Hsieh 2006). The trick consists in extending the generators of \( S \) (by attaching extra Pauli operators at their end) in order to generate a new group \( S' \) that results Abelian and for which the above theory can be applied. Because entanglement is used, these are entanglement-assisted QECC and the notation \([n, k; m]\) with \( m \) denoting the number of entangled ancilla quits \((n - m - k\) giving the number of unentangled ancilla quits). Entangled-assisted codes are better than their non entangled counterparts and may lead to higher rates. The reason is that entanglement allows to increase the dimension of the decoding Hilbert space to \( 2^{n+m} \) compared to \( 2^n \) for unentangled ancillas. This also leads to revise the Hamming bound (108) with \( 2^{n+m} \) on the r.h.s. (Bennett et al. 2002).

Finally, notice that in constructing codes, besides pursuing the highest possible rate, one should also take into account the complexity of the encoding/decoding procedures. It can be evaluated by means of the number of elementary steps, i.e., number of elementary gate operations, needed. Efficient encoding/decoding requires polynomial scaling of complexity vs block code length \( n \). Luckily, stabilizer codes are efficiently encodable/decodable (Gottesman 1997), but usually do not achieve the channel capacity (see Sec. V).

In classical information theory it has been very recently discovered a class of codes, polar codes, that conjugate the two requests of achieving capacity and having simple encoding/decoding (Arikan 2009). Specifically they are encodable/decodable with a complexity scaling like \( n \log n \) and they attain the Shannon limit in the case of equiprobable inputs (symmetric capacity). The idea underlying these codes is the channel polarization effect. Suppose to have \( n \) identical uncorrelated channels with equiprobable inputs. Then one can create, through a suitable (channel adapted) transformation, a set of \( n \) logical channels such that each logical channel is either noiseless or completely noisy. The logical channels are thus polarized into “good” and “bad” channels. The messages can then be transmitted via the former ones, while the inputs to the latter are fixed to values known to the decoder. When \( n \rightarrow \infty \) it turns out that the fraction of good channels tends to the (symmetric) capacity of the original channel.

A similar result has been pursued in the quantum realm. However the quantum version of polar codes do not admit an efficient decoding algorithm (Wilde and Guha 2011a,b). Quite generally they can be efficiently decoded when they are turned into entanglement assisted codes, i.e., when using pre-shared entanglement between

\footnote{Actually \( \sigma_y = i \sigma_x \sigma_z \), however the imaginary unit is irrelevant to the end of error correction.}
bounded by (Gaitan, 2008) and the quantum version of Reed-Muller codes (Steane 2002). Aside them, a particularly simple way to construct codes that can correct multiple errors is to concatenate several single-error correcting codes, i.e., codes with \( d = 3 \). For the sake of simplicity, one illustrates the case of two layers of concatenation and consider single qubit encoding. Assume the inner code (first layer) is a \([n_1, k_1]\) stabilizer code \( Q_1 \) with distance \( d_1 \), and the outer code (second layer) is a \([n_2, 1]\) stabilizer code \( Q_2 \) with distance \( d_2 \). The concatenated code \( Q = Q_1 \circ Q_2 \) maps \( k_1 \) qubits into \( n = n_1n_2 \) qubits, with code construction parsing the \( n \) qubits into \( n_2 \) blocks \( B(b) \) (\( b = 1, ..., n_2 \)) each containing \( n_1 \) qubits. Explicitly, the concatenated code \( Q \) is constructed as follows. For any codeword \( |c_{\text{out}}\rangle \) of the outer code \( Q_2 \),

\[
|c_{\text{out}}\rangle = \sum_{i_1...i_{n_2}} \alpha_{i_1...i_{n_2}} |i_1...i_{n_2}\rangle, \quad (120)
\]

with \( |i_1...i_{n_2}\rangle = |i_1\rangle \otimes ... \otimes |i_{n_2}\rangle \), replace each basis vector \( |i_j\rangle \) by a basis vector \( |\phi_{i_j}\rangle \) of the inner code \( Q_1 \), so that

\[
|c_{\text{conc}}\rangle := \sum_{i_1...i_{n_2}} \alpha_{i_1...i_{n_2}} |\phi_{i_1}\rangle \otimes ... \otimes |\phi_{i_{n_2}}\rangle. \quad (121)
\]

Notice that the above mentioned construction produces a \([n_1n_2, k_1]\) code with distance \( d = d_1d_2 \).

If there are \( L \) levels of concatenation of the same single qubit code, and \( p_e \) is the error probability on single qubit, it is possible to show that the code failure probability is bounded by (Gaitan 2008)

\[
p_e^{(L)} \leq p_0 \left( \frac{p_e}{p_0} \right)^{2L}; \quad (122)
\]

where \( p_0 \) is an estimate of the threshold error probability that can be tolerated and depends on the chosen single qubit code. Hence, provided that \( p_e < p_0 \), one can make the code failure probability as small as one wish by adding enough levels to the code.

C. Performance of standard quantum codes in the presence of correlated errors

Although the idea of i.i.d. errors was underlying the standard theory of quantum error correcting codes, recent studies aim at characterizing the effect of correlated noise on the performance of most relevant QECCs.

A case study is provided by a regular Markovian channel (see Sec. III.D.4) where the CPTP maps \( \Phi_{Q_1}^{(i)} \) in Eq. (99) are of the form \( \Phi_{Q_1}^{(i)}(\ldots) = \Phi^{(i)}(\ldots) = K_i(\ldots)K_i^\dagger \), with unitary Kraus operators \( K_i \) and \( p_{i}^{(j)} = p_{i} = (1 - \mu)p_i + \mu \delta_{i,i'} \). A sequence of \( n \) uses of the memory channel are hence represented by the map

\[
\Phi_{\mu} \left( \rho_Q^{(n)} \right) = \sum_{i_1,...,i_n} p_{i_{n-1}i_{n-2}...i_1} \rho_{i_{n-1}i_{n-2}...i_1} \times (K_{i_1} \otimes ... \otimes K_{i_n}) \rho_Q^{(n)} (K_{i_1} \otimes ... \otimes K_{i_n})^\dagger. \quad (123)
\]

The correlation parameter \( \mu \) roughly quantifies the degree of memory of the considered channel. For \( \mu = 0 \) one obtains the case of i.i.d. (memoryless) noise, while the limit \( \mu = 1 \) describes completely correlated errors.

In Refs. (Cafaro and Mancini 2010a,b) it has been shown that the performance of stabilizer codes (evaluated by means of entanglement fidelity [54]) is lowered by increasing \( \mu \). The same relation between the fidelity and the correlation parameter has been observed for a model of long term memory channel (see Sec. III.D.8) obtained by convex combination of uncorrelated and completely correlated quantum channels

\[
\Phi(\rho) = (1 - \mu) \Phi_{\mu=0}(\rho) + \mu \Phi_{\mu=1}(\rho), \quad (124)
\]

where \( \Phi_{\mu=0} \) and \( \Phi_{\mu=1} \) are given by (123) in the limiting cases of \( \mu = 0 \) and \( \mu = 1 \) (corresponding to uncorrelated and completely correlated errors respectively). Actually, the effect of the memory is to take the code probability of error back to a linear dependence on the single error probability.

From the above results it seems that noise correlations are always detrimental for standard QECC (see also (Klesse and Frank 2005)). However, this is not generally true. For instance, D’Arrigo, Benenti and Facchetti 2009 considered Kraus operators on the \( l \)-th qubit given by \( K_l = \exp (-i \phi_l \sigma_z) \), with \( \phi_l \)’s being a random phase shifts distributed according to a zero mean Gaussian, and introduced memory effects through a non-diagonal covariance matrix of such a distribution. Within this model the code probability of error keeps a quadratic dependence on the single error probability (typical of memoryless setting as results from Sec. IV.A). Furthermore (Shabani 2008) observed that, when qubits all interact with the same dephasing environment, better performance can be obtained in the presence of correlated errors, depending on the timing of the recovery.

In order to better understand the different behaviour of standard QECC in the presence of memory, it should be noted that the condition of independent errors is not equivalent to the condition of memoryless quantum channel. Indeed, one can say that independent errors are those for which the probability of \( k \) errors is of the order of \( e^k \), given that the single error has a (small enough)
probability $\epsilon$. On the other hand, in the setting of memoryless channels each qubit independently interacts with its own environment, and different environments do not interact among themselves. While memoryless channels give rise to independent errors, the converse is not necessarily true. As a matter of fact, there are situations where although qubits do not interact independently with their environments, the generated errors still satisfy the independence condition — provided that qubits do not directly interact with each other (Hwang, Ahn and Hwang 2001). Thus, in such cases standard QECCs work enough well.

On the other hand, whenever input qubits also interact with each other the standard QECC are easily jeopardized. This may happen for quantum channels with ISI (see Sec. II.B, e.g., when different input qubits directly interact via a controlled-unitary transformation. To show that, assume to have a quantum code $Q$, with basis codewords $\{|\phi_i\rangle\}_{i=1}^{|Q|}$, that can correct any error operator belonging to $Q$ when applied to any vector $|\phi\rangle \in Q$. Specifically, the code works in such a way that for any noise operator $K \in Q$, a corrupted basis codeword $|\phi_i\rangle' = K|\phi_i\rangle$ is mapped, through the recovery map to a product state $|\phi_i\rangle \otimes |\text{synd}(K)\rangle$, where $\text{synd}(K)$ stands for the error-syndrome of $K$. Since $\text{synd}(K)$ does not depend on $|\phi_i\rangle$, one can correct errors applied to any linear combination $\sum_i c_i|\phi_i\rangle$ of the basis vectors $\{|\phi_i\rangle\}_{i=1}^{|Q|}$. However, this is no longer true if different input qubits interact each other, e.g., via a controlled-unitary transformation. In such a case the error may depend on the specific basis codeword $|\phi_i\rangle$. In that case the corrupted codeword reads $\sum_i c_iK_i|\phi_i\rangle$. Then, upon decoding one gets $\sum_i c_i|\phi_i\rangle \otimes |\text{synd}(K_i)\rangle$ which, by tracing out the syndrome register, gives a state different of the original one $\sum_i c_i|\phi_i\rangle$. Actually, for controlled bit-flip errors, it is like the environment gets information about the codeword and thus corrupt it. This kind of errors is particularly relevant for modeling adversarial noise, where an adversary is allowed to decide which error operator to apply based on the specific codeword it acts upon. In Ref. [Ben-Aroya and Ta-Shma 2009] it has shown that no QECC can perfectly correct controlled bit-flip errors, although a QECC of arbitrarily high dimension can approximately correct them (i.e., it can give back a codeword ‘close’ to the original one once an error acted on it).

**D. Decoherence free subspaces**

A suitable strategy to deal with completely correlated errors is represented by Noiseless Codes, also known as Decoherence Free Subspaces (DFS). This is a ‘passive’ quantum error correction method where the key idea is that of avoiding decoherence by encoding quantum information into special subspaces that are protected from the interaction with the environment by virtue of some specific dynamical symmetry (Duan and Guo 1997; Lidar, Chiang and Whaley 1998; Lidar and Whaley 2003; Palma, Suominen and Ekert 1996; Zanardi and Rasetti 1997).

It turns out that a subspace $H_{DFS}$ of $H_Q$ is a DFS if and only if all Kraus operators, when restricted to $H_{DFS}$, are equal, up to a multiplicative constant, to a given unitary transformation $U^\text{DFS}_Q$. In the case of imperfect initialization, i.e., state not initialized inside a DFS, in a suitable basis the matrix representation of the Kraus operators that describe the action of the channel is given by

$$K_k = \left( \begin{array}{cc} s_kU^\text{DFS}_Q & 0 \\ 0 & M_k \end{array} \right),$$

(125)

where $M_k$ is an arbitrary matrix that acts on the orthogonal complement $H_{DFS}$ (with $H_Q = H_{DFS} \oplus H_{DFS}$) and may cause decoherence there (Shabani and Lidar 2005). Equation (125) implies

$$K_k^\dagger K_l = \left( \begin{array}{cc} S_{kl} & 0 \\ 0 & M_k^\dagger M_l \end{array} \right),$$

(126)

where $S_{kl} = s_k s_l$. Applying condition (107) to the present setting, it follows that DFS can be viewed as a special class of QECCs, where upon restriction to the code space $Q = H_{DFS}$, all recovery operators, i.e., Kraus operators of recovery map, are proportional to the inverse of the unitary $U^\text{DFS}_Q$. It is worth noticing that in the DFSs case the matrix $S$ has rank 1. Hence, a DFS is an example of maximally degenerated quantum error correcting code. In the case of perfect DFS encoding, the necessary and sufficiency conditions are less restrictive than Eq. (125) — see (Shabani and Lidar 2005).

An example of effective application of DFS encoding can be obtained for the case of the ‘completely correlated’ channel (see Sec. III.D.8). By putting $n = 2, \ K_0 = 1, K_1 = \sigma_z,$ and $p_0 = 1 - p, p_1 = p$ in Eq. (123) one obtains a quantum channel with Kraus operators

$$K_{00} = \sqrt{p}1 \otimes 1, \ K_{01} = 0, \ K_{10} = 0, \ K_{11} = \sqrt{(1 - p)}\sigma_z \otimes \sigma_z.$$

(127)

A DFS is given by span$\{|01\rangle, |10\rangle\}$ where one can safely encode a qubit

$$|0_L\rangle = |01\rangle, \ |1_L\rangle = |10\rangle.$$

(128)

By extending this argument, one can say that in the case of completely correlated errors it is possible to exploit the invariance of a subspace to safely encode information.

In Ref. (Chiribella et al. 2011), it has been provided a generalized quantum Hamming bound for nondegenerate codes, which depends on the rank of the CI state (see Sec. III.A) associated to the noise process and holds for any kind of (possibly correlated) channel model. The original Hamming bound (108), which was formulated for the case of independent noise on the encoding systems is then recovered as a particular case. On the other hand, for completely correlated noise it has been shown how
to exploit degeneracy to violate the generalized quantum Hamming bound and achieve perfect quantum error correction with fewer resources than those needed for non-degenerate codes. As an example consider the following channel

$$\Phi(\rho^{(n)}) = p \rho + \sum_{i=1,\ldots,n,j>i} \left( p_{X,ij} \sigma_{x,i} \sigma_{x,j} \rho_Q \sigma_{x,i} \sigma_{x,j} + p_{Y,ij} \sigma_{y,i} \sigma_{y,j} \rho \sigma_{y,i} \sigma_{y,j} + p_{Z,ij} \sigma_{z,i} \sigma_{z,j} \rho \sigma_{z,i} \sigma_{z,j} \right),$$

(129)

where the input state is left unchanged with probability \( p = 1 - \sum_{i=1,\ldots,n,j>i} (p_{X,ij} + p_{Y,ij} + p_{Z,ij}) \), while it undergoes Pauli errors \( \sigma_{x,i} \), \( \sigma_{y,i} \) and \( \sigma_{z,i} \) on qubits \( i \) and \( j \) with probabilities \( p_{X,ij} \), \( p_{Y,ij} \) and \( p_{Z,ij} \) respectively. By evaluating the rank of the CJ state associated with the map (129) one obtains the generalized quantum Hamming bound of Ref. (Chiribella et al., 2011):

$$2^k \left[ 1 + 3 \left( \frac{n}{2} \right) \right] \leq 2^n.$$

(130)

Then, by considering for instance \( k = 1 \) one gets \( n = 7 \) as smallest integer satisfying the bound. However, one can also construct codes with lower values for \( n \). For instance, this is the case of the code

$$|0_L\rangle = |000\rangle, \quad |1_L\rangle = |111\rangle.$$

(131)

Now notice that these basis codewords are not affected by the action of \( \sigma_z \) on any pair of qubits. Consequently the action of \( \sigma_x \) on a pair of qubits is identical to the action of \( \sigma_y \) on the same pair of qubits. In other words the code is degenerate. Therefore, one has only to correct errors due to \( \sigma_x \) operators. This can be realized through a projective measurement onto the subspaces \( S_{00} = \text{span}\{|000\rangle, |111\rangle\}, \ S_{01} = \text{span}\{|100\rangle, |011\rangle\}, \ S_{10} = \text{span}\{|010\rangle, |101\rangle\} \) and \( S_{11} = \text{span}\{|001\rangle, |110\rangle\} \). If the measurement outcome is "00", no errors have affected the qubits; on the contrary, if the measurement outcome is "01", errors have affected qubits 2 and 3 and can be corrected by applying there \( \sigma_x \). Similarly, all the other possible errors can be detected and corrected. It is hence clear that this code violates the quantum Hamming bound (130) thanks to the invariance of the coding subspace under the action of pair of \( \sigma_z \) which allows for perfect error correction.

Having seen that DFSs are suitable to encode information in the presence of completely correlated errors, it is natural to expect that their performance decreases by reducing the degree of errors’ correlation (Demkowicz-Dobrzański, Kolenderski and Banaszek, 2007). Actually Refs. (Cafaro and Mancini, 2010a) and (Cafaro and Mancini, 2011) confirm this fact for the Markovian model of (123) and for the model of (124) respectively.

These results are opposite to those reviewed in the previous Section IV.C. Then, one may argue that, for the memory channel models (123), (124), where the memory effects are described by a single parameter \( \mu \), there must be a threshold value \( \mu^* \) that allows one to select the best code between the standard and the noiseless ones (Cafaro and Mancini, 2010a, 2011; D’Arrigo et al., 2008).

### E. Designing quantum codes for correlated errors

The specific features of error models can be used to design new quantum codes that better cope with correlated errors.

The results of the previous Subsection suggest that it might be convenient to concatenate decoherence-free subspaces with standard quantum error correcting codes in order to achieve higher entanglement fidelity values in both low and high correlations regimes. This kind of concatenation was first introduced in Ref. (Lidar, Bacon and Whaley, 1999), and it was investigated in the context of memory channels in Ref. (Clements, Siddiqui and Gea-Banacloche, 2004), and subsequently in Ref. (Cafaro and Mancini, 2011).

As an illustrative example, consider to encode one logical qubit into a decoherence free subspace (Q\textsubscript{DFS} = Q\textsubscript{outer}) spanned by the basis codewords

$$|0_L\rangle = |+\rangle, \quad |1_L\rangle = |-\rangle,$$

(132)

and then encode each qubit of this basis into a three-qubit bit repetition code (133) (Q\textsubscript{bit} = Q\textsubscript{inner}). One obtains that the basis codewords of the concatenated code \( \mathcal{Q} = \mathcal{Q}_{\text{bit}} \otimes \mathcal{Q}_{\text{DFS}} \) are given by

$$|0_L\rangle = \frac{1}{2} \left( |000000\rangle - |000111\rangle - |111000\rangle - |111111\rangle \right),$$

$$|1_L\rangle = \frac{1}{2} \left( |000000\rangle + |000111\rangle - |111000\rangle - |111111\rangle \right).$$

The entanglement fidelity for the concatenation of a repetition code and a noiseless code for the models of (123) and (124) with \( K_0 = 1, K_1 = \sigma_z, p_0 = 1 - p \) and \( p_1 = p \) is reported in Fig. 3. It turns out that in the first case the concatenated code does not work well for partially correlated errors. It is always better to use either the outer or the inner code alone depending on whether one is below or above the threshold value \( \mu^*(p) \). On the contrary, in the second case the concatenated code works optimally almost everywhere. Hence, one may argue that for the model of (124) the concatenation procedure is particularly advantageous in the presence of partially correlated errors.

Another error model often employed is that of burst errors. Such errors can be considered as affecting a continuous sequence of quits as opposed to random single qubit errors. They are well studied in the classical framework where corresponding error correcting codes have been developed (Peterson and Weldon, 1972). In (Vatan, Roychowdhury and Anantram, 1997), quantum analog of burst-error correcting codes have been considered. Hamming and Gilbert-Varshamov type bounds have been derived showing that these codes are more efficient than...
codes protecting against random errors. In fact, to protect against burst errors of width $b$ (that is, errors occurring on a number $n$ of consecutive qubits with $b$ a fixed constant), it is enough to map $n - \log_2 n - O(b)$ qubits to $n$ qubits, while in the case of $t$ random errors at least $n - t \log n$ qubits should be mapped to $n$ qubits. Furthermore, in [Vatan, Roychowdhury and Anantaram 1997] an explicit construction of almost optimal quantum codes has been presented starting from classical binary cyclic codes.

A classical binary cyclic code $C$ is a linear subspace of $\mathbb{F}_2^n$ closed under the cyclic shift operator, i.e., if $(c^{(1)}, c^{(2)}, \ldots, c^{(n)})$ is in $C$, then so is $(c^{(n)}, c^{(1)}, \ldots, c^{(n-1)})$ [Hill 1985]. A burst of width $b$ is a vector in $\mathbb{F}_2^n$ whose only nonzero components are among $b$ consecutive components. Then, a linear code $C$ has burst-correcting ability $b$ iff, for every burst $w_1$ and $w_2$ of width $\leq b$ it is $H(w_1 + w_2)^\top \neq 0$, with $H$ the parity check matrix of $C$.

The definition of quantum burst-correcting codes straightforwardly follows from [115]. Consider the set $\Omega$ of quantum errors (hence the corresponding set $\mathcal{C}$ of classical error) such that both

$$\begin{align*}
\mathcal{C}_X &= \{ e_X \in \mathbb{F}_2^m \mid \exists e_Z \in \mathbb{F}_2^2 \Rightarrow (e_X | e_Z) \in \mathcal{C} \}, \quad \text{(133)} \\
\mathcal{C}_Z &= \{ e_Z \in \mathbb{F}_2^m \mid \exists e_X \in \mathbb{F}_2^2 \Rightarrow (e_X | e_Z) \in \mathcal{C} \}, \quad \text{(134)}
\end{align*}$$

are bursts of width $\leq b$. Then any quantum code $Q$ having $\Omega$-correcting ability is called a $b$-burst quantum correcting code.

Now, suppose to have a $(3b+1)$-burst-correcting binary $[n,k]$ cyclic code $C$ that is weakly self-dual, i.e., such that $C^\perp \subseteq C$. Then, to construct the $b$-burst-correcting $[n,k]$ quantum code one can proceed as follows.

Let the $(n-k) \times n$ matrix $H$ be a parity check matrix for the classical code $C$. Let $H_{-m}$ denote the matrix that is obtained from $H$ by cyclically shifting the columns $m$ times to the right. Since $C$ is cyclic, also $H_{-m}$ is a parity check matrix of $C$. Consider the stabilizer quantum code $[[n,k]]$ defined by the parity check matrix (see [118])

$$H = \left( \tilde{H} + \tilde{H}_{-b} \mid \tilde{H} + \tilde{H}_{-2b+1} \right). \quad \text{(135)}$$

The weak self duality of the code $C$ guarantees the matrix $H$ to be totally singular.

Let $e = (e_X | e_Z)$ and $e' = (e'_X | e'_Z)$ be bursts of width $\leq b$, with $e \neq e'$ and take

$$w = e_X + e'_X + (e_X + e'_X)_{-b} + e_Z + e'_Z + (e_Z + e'_Z)_{-2b+1}, \quad \text{(136)}$$

where $e_{-b}$ denotes the vector obtained by cyclically shifting $e$ to the right $b$ times. Then, it is easy to check that $w \neq 0$ and $w$ is the sum of two bursts of width $\leq 3b+1$. Hence $w \notin C$ and

$$H(e + e')^\top = Hw^\top \neq 0, \quad \text{(137)}$$

which guarantees that the ability of the quantum code $[[n,k]]$ to correct $b$-burst errors.

The existence of classical $(3b+1)$-burst-correcting binary cyclic codes that are weakly self-dual with length $n = 2^m - 1$ and dimension $k = n - m - (3b+1)$ is known [Peterson and Weldon 1972]. For them, $m$ satisfies $m = 0 \pmod{m_b}$ with a fixed integer $m_b$ only depending on $b$. Hence these classical codes lead to almost optimal quantum codes (compared with the bound $n - \log_2 n - O(b)$ given above).

By increasing the length of the bursts, one should increase the length of the burst code as well. Alternatively it might be possible to resort to the interleaving technique. By using this method, the codewords can be distributed amongst the qubit stream so that consecutive words are never next to each other. On de-interleaving they are returned to their original positions so that any errors that have occurred become widespread. This ensures that any burst (long) errors now appear as random (short) errors.

Classically the interleaving of $m$ codewords $(c_1, c_2, \ldots, c_m)$ of an $[n,k]$ code is achieved by permuting the positions of bits in codewords as follows...
\[ (c_1, c_2, \ldots, c_m) = \left( \left( c_1^{(1)}, c_1^{(2)}, \ldots, c_1^{(n)} \right), \left( c_2^{(1)}, c_2^{(2)}, \ldots, c_2^{(n)} \right), \ldots, \left( c_m^{(1)}, c_m^{(2)}, \ldots, c_m^{(n)} \right) \right) \rightarrow \left( \left( c_1^{(1)}, c_1^{(2)}, \ldots, c_1^{(n)} \right), \left( c_2^{(1)}, c_2^{(2)}, \ldots, c_2^{(n)} \right), \ldots, \left( c_m^{(1)}, c_m^{(2)}, \ldots, c_m^{(n)} \right) \right). \] (138)

The procedure is equivalent to constructing the code as an \( m \times n \) array where every row is a codeword of the original \([n, k]\) code \((c_1, c_2, \ldots, c_m)\). Now, a burst of length \( \leq bm \) can have at most \( b \) symbols in any row of this array. Since each row can correct a burst of length \( \leq b \), the code can correct all bursts of length \( \leq bm \) (the parameter \( m \) is the interleaving degree).

Therefore, given an \([n, k]\) classical code correcting bursts of length \( \leq b \), then interleaving this code to the degree \( m \) produces an \([nm, km]\) classical code correcting bursts of length \( \leq bm \) \cite{Peterson and Weldon, 1972}.

Moving to the quantum framework, in order to interleave quantum codes, one needs to exchange the qubits one by one, following \( (138) \). Therefore, the basic step of the quantum interleaving simply is a swapping operation between two qubits. Then, the classical result can be extended as follows \cite{Kawataba, 2000}; interleaving an \([n, k]\) quantum code correcting bursts of length \( \leq b \) to the degree \( m \) produces an \([nm, km]\) quantum code correcting bursts of length \( \leq bm \).

F. Convolutional Codes

It is possible to extend the notion of stabilizer codes introduced in Sec. \( \ref{sec:convolutional} \) to codes which allow for an overlap between the individual steps of the encoding operation \cite{Chau, 1998, 1999, Ollivier and Tillich, 2003, 2004}. From classical coding theory, codes with these properties are called convolutional codes. Although not specifically designed for memory channels, they are intimately related to them.

A quantum convolutional stabilizer code is defined by the generators of its stabilizer group just like a block stabilizer code (see Eq. \( \ref{eq:convolutional} \)). Consider a convolutional code encoding \( k \) logical qubits per \( n \) physical qubits, such that every block has an output of \( n+m \) qubits. \( n \) of those are output qubits, while \( m \) are passed on to the next step. Then, an \([n, k]\) \( m \)-convolutional stabilizer code is given by the Abelian stabilizer group

\[ \mathcal{G} = \text{span}\{g_{j,i} = 1^\otimes j \otimes g_{0,i} | 1 \leq i \leq n-k, 0 \leq j \}, \] (139)

where \( g_{0,i} \in \mathcal{P}_{n+m} \) and all \( g_{i,j} \) are independent of each other.

Notice that the total number of physical qubits - length of the code - is left unspecified. Actually it is useful to set it to infinity by considering a Pauli group \( \mathcal{P}_\infty \) with elements defined on a semi-infinite chain of qubits, but acting nontrivially only on a bounded number of them.

Then the generators are considered to be padded from the right with identities \( \mathbb{I} \).

The structure of the stabilizer group generators can be summarized, following \( \ref{eq:convolutional} \), by a semi-infinite matrix

\[ H = \begin{pmatrix} \mathbb{I} & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & \mathbb{I} & 0 \\ 0 & \cdots & 0 & \mathbb{I} \end{pmatrix} \] (140)

Each line of the matrix represents one of the \( g_{j,i} \) and each column a different qubit. Thus, any given entry of \( H \) is a Pauli matrix for the corresponding qubit and generator. The rectangles represent which qubits are potentially affected by the action of the generators. Clearly, when \( m = 0 \) one has each block separately and obtain a block code.

Actually, it is possible to use the invariance by \( n \) qubit translation of the generators to find a shorter description. One defines the shift (delay) operator \( D \) acting on any element \( A \in \mathcal{P}_\infty \) by

\[ D[A] = \mathbb{I}^\otimes n \otimes A. \] (141)

Then, the generators of the code can be written as:

\[ g_{j,i} = D^j[g_{0,i}], \ 0 \leq j, \ 1 \leq i \leq n-k. \] (142)

Using this, one only needs to consider the first \( n-k \) generators. All the others are obtainable by repeated applications of \( D \).

In addition to applying \( D^j \) to an element of the Pauli group \( A \) (with bounded support), it is also possible to consider a polynomial \( P(D) = \sum_j \alpha_j D^j \) and apply it as \( P(D)[A] = \prod_j \alpha_j D^j[A] \). That critically relies on the fact that all copies of \( A \) shifted by \( D^j \) commute (see e.g., \cite{Ollivier and Tillich, 2004}).

Now, it is worth noticing that the encoding operations for convolutional codes can be described as quantum memory channels, due to the fact that some of the output qubits of the \( n+1 \)th encoding step will be used as inputs in the \( n \)th encoding step, thus the blocks overlap. These qubits correspond to the memory system of the memory channel describing the encoding.
The encoding operation is the same in every step (neglecting initialization and finalization), thus every step is described by the same channel (see Fig. 6). This can be stated more precisely saying that for every \([n, k]\) convolutional stabilizer code one can find an encoding operation which is described by the concatenation of a Weyl covariant memory channel with uni-modular characteristic function (see Section II.I.3). The channel has \(n\) input and output qubits and uses \(m\) qubits of memory. One use of the channel corresponds to one block in the encoding (Gütschow 2010).

Unfortunately convolutional codes also carry disadvantages. Because information is transmitted from one block to the next, errors can spread as well. Depending on the encoding algorithm errors that spread without bound on the output side can occur. These are called catastrophic errors and have to be avoided by employing non-catastrophic convolutional codes (Grassl and Roetteler 2006). The link between memory channels and convolutional stabilizer codes allows to prove that a convolutional encoder is non-catastrophic iff the memory channel representing it is strictly forgetful (Gütschow 2010). Lacking the boundaries between code blocks, convolutional codes exhibit the same "continuous structure" as channels with memories. As such, they are not only describable by memory channels, but could also result particularly suited for memory channels.

V. CAPACITIES OF QUANTUM CHANNELS

A natural question that arises after having examined correcting codes is what are the maxima communication rates achievable in quantum channels. For classical channels the highest rate (number of bits per channel uses) of reliable information transmission attainable via the application of encoding and decoding error correcting procedures defines the capacity (Cover and Thomas 1991; Gallager 1968). In this context reliability refers to the requirement that the transferred messages have to be received without possibility of misunderstanding, i.e., the communication errors have to be removed by the selected coding strategy. One speaks of zero-error capacity when imposing this constraint for codes of finite length (i.e., codes which operate on a finite number of information carries or channel uses) (Körner and Orlitsky 1998; Shannon 1956). However in many cases of physical and technological interest, it is more reasonable and mathematically more convenient to enforce such condition only in the asymptotic limit of infinitely long messages. Under this paradigm in fact explicit expressions for the channel capacity are available as a function of the noise model which is tampering the communication line. For instance in the case of a memoryless classical channel characterized by the conditional probability \(p(y|x)\) of producing the output symbol \(y\) when fed with input \(x\), the associated capacity can be expressed as (Shannon 1948)

\[
C_{SH} = \max_{p(x)} I(X : Y) , \tag{143}
\]

where the maximization is performed over all probability distributions on \(x\), and where \(I(X : Y)\) is the corresponding Shannon mutual information – see Sec. II.K. The proof leading to (143) relies on the notion of typical sequences (Cover and Thomas 1991; Gallager 1968) and it does not provide an explicit recipe for determining the optimal coding and decoding strategies (this is why error correcting codes and capacities are often treated as distinct subjects with no exception for the quantum realm). Yet Shannon’s result establishes a fundamental benchmark that is useful to test the effectiveness of any coding procedure – an informal and clear introduction to this topics can be found in (Preskill 1998) or (Galindo and Martín-Delgado 2002). Strong versions of the converse Shannon theorem have been proved (Arimoto 1973; Wolfowitz 1964), which establish that if the rate of communication of a (memoryless) classical channel exceeds \(C_{SH}\) then the error probability of any coding scheme converges to one in the limit of many channel uses.

The notion of capacity based on asymptotic reliability has also an important operational meaning stated by the Reverse Shannon Theorem, a results which, strangely enough, was only formulated and proved only recently within the context of quantum communication (Bennett et al. 1999 b; 2002). According to it, for any classical noisy channel of capacity \(C_{SH}\), if the sender and receiver share an unlimited supply of random bits, an expected \(nC_{SH} + o(n)\) uses of a noiseless binary channel are sufficient to exactly simulate \(n\) uses of the original channel.

As anticipated in Sec. II the generalization of the above ideas to the quantum setting leads to the introduction of a plethora of channel capacities, depending on whether classical or quantum information has to be transmitted, and whether additional resources, as pre-shared entanglement, are exploited. An unification of these quantities under a common formalism based on resource inequalities has been presented in (Abeyesinghe and Hayden...
A. Operational definitions

1. Sending bits or qubits on a quantum channel

The classical (resp. quantum) capacity $C$ (resp. $Q$) of a quantum channel defined by the CPTP maps $\Phi^{(n)}$ of Eq. (75) is the maximum rate $R$ at which classical (resp. quantum) information, encoded on a set of quantum carriers, can be sent reliably from the sender Alice to the receiver Bob [Shor, 1995]. As in the classical setting [Shannon, 1948], the rate is measured as the ratio $R = k/n$ among the number $k$ of bits (resp. qubits) transmitted and the number $n$ of carriers employed (the “redundancy” of the code according to Sec. IV). Similarly the reliability condition is introduced by requiring that in the asymptotic limit of $k \to \infty$ the error probability of the procedure will asymptotically vanish (or, equivalently, the fidelity of the transmission will approach unity), while keeping $R$ constant. In view of these operational definitions, $C$ and $Q$ can be expressed as the following limit [Bennett et al., 2002; Bennett and Shor, 1998]

$$\lim_{\epsilon \to 0} \lim_{n \to \infty} \sup \left\{ \frac{k}{n} : \exists \Phi_E^{(k-n)}, \exists \Phi_D^{(n-k)} \right\}$$

$$\min_{m \in M} F \left( \left| m \right|, \Phi_D^{(n-k)} \circ \Phi^{(n)} \circ \Phi_E^{(k-n)} \right) > 1 - \epsilon \right\}, \quad (144)$$

where, analogously to the notation introduced at the beginning of Sec. IV and as sketched in Fig. 7, $\Phi_E^{(k-n)}$ and $\Phi_D^{(n-k)}$ are respectively encoding and decoding channels mapping elements $\left| m \right|$ from a reference input space $M$ (the messages Alice wishes to send to Bob) to states $\Phi_E^{(k-n)}(\left| m \right|) := \Phi_E^{(k-n)}(\left| m \right|, \left| m \right|)$ of $n$ carriers (the codewords of the procedure), and $F$ is the input-output fidelity function introduced in Sec. IV.1. In particular to fix the units properly, the expression for $C$ is obtained by taking $M$ to be a collection of $2^k$ orthogonal vectors which, without loss of generality can be identified with the elements $\left\{ \left| 0 \right|, \left| 1 \right| \right\} \otimes k$ of the computational basis of $k$ qubits. On the other hand, for the quantum capacity $Q$ the set $M$ coincides with the whole $C^{2^n}$ (as the latter includes the elements of the canonical basis, it trivially follows that for a given communication line one has $Q \leq C$). The limits in Eq. (144) is finally computed by first taking a supremum limit in $k \to \infty$ (which always exists) and then sending the error parameter $\epsilon$ to zero to enforce the transmission fidelity to approach unity for all input messages. In the above expression this is explicitly enforced by requiring $1 - \epsilon$ to lower bound the minimum value achieved on $M$ by the transmission fidelity. Such rather strong requirement however can be relaxed by replacing it with a similar constraint that only applies on the average transmission fidelity: this doesn’t affect the limit in Eq. (144) and hence the definitions of $C$ and $Q$ [Keyl, 2002; Kretschmann and Werner, 2004]. Similarly the same value of $Q$ one gets from (144) can also be obtained by substituting the minimum in $F$ with the entanglement fidelity introduced in Eq. (52) [Barrett, Knill, and Nielsen, 2000; Kretschmann and Werner, 2004].

Finally, one observes that the definition (144) yields a natural data-processing inequality for the capacities. For instance, given $C$ the classical capacity of a quantum channel whose CPTP mapping $\Phi^{(n)} = \Phi^{(n)} \circ \Phi^{(n)}$ is obtained by concatenating other two CPTP maps, one has

$$C \leq \min\{C', C''\}, \quad (145)$$

where $C'$ and $C''$ are respectively the classical capacities of channels described by $\Phi^{(n)}$ and $\Phi^{(n)}$ (the same relation applying also for the quantum capacities $Q$, $Q'$ and $Q''$ as well as for all the other capacities reviewed in the following sections with the notable exception of those discussed in Sec. V.A.4 where the constraints may introduce spurious effects in the optimization, see e.g., Ref. [Giovannetti and Mancini, 2005]). The proof of this rather intuitive fact follows by observing that when passing from $C$ to $C'$ one can interpret $\Phi^{(n)}$ as part of the decoding procedure of map $\Phi^{(n)}$. Since the $C'$ is obtained by optimizing the transmission rate with respect to all possible decodings, including those which do not use $\Phi^{(n)}$ as a preliminary stage, it follows that $C'$ is certainly not smaller than $C$. Similarly when passing from $C$ to $C''$, one can interpret $\Phi^{(n)}$ as part of the encoding stage for $\Phi^{(n)}$: again since $C''$ is the optimal rate with respect to all encoding maps one has that it is certainly not smaller than $C$. An application of the above analysis to two channels $\Phi_1^{(n)}$ and $\Phi_2^{(n)}$ which are unitarily equivalent [El] shows that they must possess the same capacities: in this case indeed the CPTP concatenation which links the two maps can always be reversed, producing both the inequality $C(\Phi_1^{(n)}) \leq C(\Phi_2^{(n)})$ and its counterpart $C(\Phi_2^{(n)}) \leq C(\Phi_1^{(n)})$. 

2. Capacities assisted by ancillary resources

Entanglement is the fundamental resource for quantum information. In the context of quantum commu-
Private classical capacity of a quantum channel

The private classical capacity \( C_p \) of a quantum channel is defined as the maximum rate at which classical information can be transmitted privately from the sender to the receiver. Formally this is enforced by requiring that a third party (Eve) who has access to channel environment and who is trying to recover Alice messages to Bob, will get it with an error probability that is approaching unity in asymptotic limit of infinitely long messages (Devetak, 2005). Again \( C_p \) can be expressed as the limit in Eq. (144) by further constraining the coding and decoding procedures to satisfy an entropic inequality that implement the privacy requirement. Specifically this is obtained by upper bounding with \( \epsilon \) the Holevo information (71) of the complementary channel (19) of \( \Phi(n) \) and associated with the uniform ensemble \( \mathcal{E} = \{ p_m = 2^{-k}, \Phi_E^{(k-n)}(m) \}_{m \in \mathcal{M}} \) generated by the encoding mapping selected by Alice. As the complementary map is the transformation that links the channel inputs to the images they produce on the environment (see Sec. II.C), this choice, via the Holevo Bound (72), ensures that the information Eve can recover on Alice’s messages vanishes when taking the limit \( \epsilon \to 0 \). By construction \( C_p \) is always smaller or equal to the corresponding \( C \) or greater equal to \( Q \), i.e.,

\[
C \geq C_p \geq Q, \tag{146}
\]

(the last inequality being associated with the fact that the ability of sending all vectors of \( \mathbb{C}^{2 \otimes k} \) with unit fidelity ensures that no information on the transferred states is passing from Alice to the environment).
4. Constrained capacities

The operational definitions of capacities can be modified in order to account for possible constraints on the input (or output) states on the channel. For instance, three weaker versions of the classical capacity of a quantum channel have been identified. Specifically, one defines the one-shot (or product-state, or classical-quantum, or finally Holevo) classical capacity $C_{cq}$ (which following a rather universal conventional hereafter will be indicated with the symbol $C_1$) by requiring that the employed coding maps entering in Eq. (144) produce only separable codewords, that is, $\Phi_E^{k \to n}(m)$ is a separable state of the $n$ carriers for all messages $|m\rangle$ in $\mathcal{M}$ (one notes incidentally that the analogous of $C_1$ for the classical private capacity $C_p$, i.e., the one-shot private classical capacity $C_{p1}$, has been defined in Holevo and Werner (2001) – see Sec. V.B.2). Similarly one defines a quantum-classical capacity $C_{qc}$ by leaving the encoding channel unconstrained but imposing the decoding channels $\Phi_D^{(n-k)}$ to be LOCC. Finally assuming LOCC operations for $\Phi_D^{(n-k)}$ and separability for the codewords $\Phi_E^{(k \to n)}(m)$ one defines the classical-classical capacity $C_{cc}$. The unconstrained capacity $C$ (often identified in this context also as the quantum-quantum or analogous of (10) of the maps $\Phi$) is bounded on the one-shot private classical capacity $\Phi_{qc}$, the ordering between the last two being at present unknown.

Of special interest are also a class of physically motivated constrained capacities obtained by introducing a family of observables $\{A(n)\}_{n=1,\ldots,\infty}$ and by imposing that for any $n$ the mean value of $A(n)$ is bounded on the ensemble of states at the input of $n$ uses of the quantum channel. The capacity of a quantum channel under such a constraint can be defined as in Eq. (144) under the additional requirement that for any $k$ and $n$

$$\text{Tr} \left( A(n) \rho(n) \right) \leq a, \quad (147)$$

with $\rho(n) = \Phi_E^{k \to n}(1/2^k)$, $1/2^k$ being the average state over the set $\mathcal{M}$. In particular, a relevant role is played by additive observables, for which one can put $A(n) = n^{-1} \sum_{k=1}^n A_k$, where $A_k \equiv A$ is the observable for a single input quantum system at the input of the $k$-th use of the channel.

The notion of constrained capacity naturally applies in the context of CV channels (see Sec. II.1.4). Indeed, due to the fact that the carrier Hilbert space is infinite-dimensional it turns out that the capacity of a CV channel can be infinite (Holevo and Werner 2001). As a matter of fact, an infinite value for the capacity corresponds to the encoding of information into larger and larger sectors of the Hilbert space. Clearly, that is in contradiction with the finiteness of the resources employed in physical realizations, e.g., the finiteness of the mean energy. It is hence meaningful to introduce a notion of capacity under a physically motivated constraint. Most natural choices are to impose a constraint on the mean value of the energy or the number of bosonic excitation per mode. In the latter case one has $A_n = n^{-1} \sum_{k=1}^n a^T_k a_k$, where $\{a_k, a_k^T\}$ are the canonical ladder operators at the channel input. From a technical point of view, these choices, besides being physically sound, guarantee that the set of states satisfying the constraint form a compact set. This is a crucial feature to ensure that the the coding theorems for Gaussian channels under constrained mean excitation number (or energy) yields to expressions formally analogous to the unconstrained case, with the optimization being performed over input ensembles satisfying the constraint (Holevo 1999, 2003; Holevo and Shirokov 2005; Holevo 2006). Since the excitation number and the energy are quadratic in the canonical variables, their mean values can be expressed in terms of the first and second moments of having no more than $N$ mean excitations per mode is expressed in terms of the first and second moments

$$\frac{\text{Tr}(C(n)) + |m|^2}{2n} \leq N + 1/2. \quad (148)$$

5. A superoperator norm approach to quantum capacities

The limit which defines $C$ and $Q$ in Eq. (144) indicates that for sufficiently large $k$ there exists $\Phi_D^{(n-k)}$ and $\Phi_E^{(k-n)}$ which makes $\Phi_D^{(n-k)} \circ \Phi(n) \circ \Phi_E^{(k-n)}$ close to the identity transformation $\text{id}_\mathcal{M}$ on $\mathcal{M}$. Specifically for the quantum capacity $Q$, $\text{id}_\mathcal{M}$ is the identity super-operator on $\mathbb{C}^{2^k}$, while for the classical capacity $C$, the map $\text{id}_\mathcal{M}$ is the fully depolarizing channel on $\mathbb{C}^{2^k}$ which leaves the elements of its computational basis $\{|0\rangle, |1\rangle\}^{\otimes k}$ unchanged.

Based on this observation a definition of capacities which is fully equivalent to the approach of Sec. II.1.4 can be given in terms of the cb-norm superoperator distance defined in Sec. II.1.2. In this approach Kretschmann (2003), a positive quantity $R$ is said to be an achievable rate for the channel $\Phi$ if for all sequence $\{k_i, n_i\}_{i \in N}$ with $\lim_{i \to \infty} k_i = \infty$ and $\limsup_{i \to \infty} k_i/n_i < R$ one has

$$\lim_{i \to \infty} \inf_{\Phi_D, \Phi_E} \|\Phi_E^{(k_i-n_i)} \circ \Phi(n_i) \circ \Phi_D^{(n_i-k_i)} - \text{id}_\mathcal{M}\|_{cb} = 0, \quad (149)$$

where $\Phi_E^{(k_i-n_i)}$, $\Phi^{(n_i)}$ and $\Phi_D^{(n_i-k_i)}$ are the duals of the the maps $\Phi_E^{(k_i-n_i)}$, $\Phi(n_i)$ and $\Phi_D^{(n_i-k_i)}$ defined in (144), while $\text{id}_\mathcal{M}$ is the dual of the identity map on $\mathcal{M}$. With these prescriptions the values of $Q$ and $C$, are then identified as the supremum of the corresponding achievable rates. A similar construction was presented also in Refs. Holevo and Werner (2001) Kretschmann.
and Werner [2004]: here however the cb-norm distance was used directly in the Schrödinger channel representation.

6. Zero-error capacities for quantum channels

All the definitions introduced so far assume a notion of capacity in which the error probability is required to nullify only in the asymptotic limit of large enough \( n \). As in classical communication theory (Körner and Orlitsky 1998; Shannon 1956) however, a more stringent reliability requirement can be enforced, i.e., imposing that the min-fidelity on \( \mathcal{M} \) should equal 1 for a finite number \( n \) of channel uses. Under this condition one is led to the definition of zero-error classical and quantum capacities (Medeiros and de Assis 2005) which, with reference to the notation introduced in Sec. V.A, can be formally expressed as

\[
\sup_k \left\{ \frac{k}{n} : \exists \Phi_E^{(k \to n)}, \exists \Phi_D^{(n \to k)} \right\}. \quad (150)
\]

\[
\min_{|\psi\rangle \in \mathcal{M}} F \left( |\psi\rangle; \Phi_D^{(n \to k)} \circ \Phi^{(n)} \circ \Phi_E^{(k \to n)} \right) = 1
\]

This corresponds to the maximal communication rate achievable by perfect codes, as introduced in Sec. IV.

B. Coding theorems for memoryless channels

Coding theorems provide expressions for the communication capacities of memoryless quantum channels \( C(\Phi) \) in terms of suitable entropic functions of the input and output states of the channel. While referring the reader to Holevo 2012; Wilde 2013; Winter 1999 a for a detailed review of the subject, here are reported the main results concerning the entanglement-assisted classical capacity, the classical capacity and its one-shot version, the private classical capacity, and the quantum capacity. Limitations and applicability of these expressions for the case of memory channels will be discussed in the following sections.

1. The Holevo-Schumacher-Westmoreland coding theorem

Preliminary attempts to compute the classical capacity of quantum channels were presented in Ref. (Hausladen et al. 1995; 1996). A closed expression for the one-shot classical capacity introduced in Sec. V.A can be formally expressed as

\[
C_1(\Phi) = \max_{E} \chi(E; \Phi), \quad (151)
\]

(owing to the concavity of von Neumann entropy \( \chi(p; \Phi) \)) for input state ensembles \( E = \{p_j, \rho_j\} \) (possibly satisfying some additional input constraints), i.e.,

\[
C_1(\Phi) = \max_{E} \chi(E; \Phi), \quad (151)
\]

This corresponds to the maximal communication rate achievable by perfect codes, as introduced in Sec. IV.

B. Coding theorems for memoryless channels

Coding theorems provide expressions for the communication capacities of memoryless quantum channels \( C(\Phi) \) in terms of suitable entropic functions of the input and output states of the channel. While referring the reader to Holevo 2012; Wilde 2013; Winter 1999 a for a detailed review of the subject, here are reported the main results concerning the entanglement-assisted classical capacity, the classical capacity and its one-shot version, the private classical capacity, and the quantum capacity. Limitations and applicability of these expressions for the case of memory channels will be discussed in the following sections.

1. The Holevo-Schumacher-Westmoreland coding theorem

Preliminary attempts to compute the classical capacity of quantum channels were presented in Ref. (Hausladen et al. 1995; 1996). A closed expression for the one-shot classical capacity introduced in Sec. V.A can be formally expressed as

\[
C_1(\Phi) = \max_{E} \chi(E; \Phi), \quad (151)
\]

This corresponds to the maximal communication rate achievable by perfect codes, as introduced in Sec. IV.

B. Coding theorems for memoryless channels

Coding theorems provide expressions for the communication capacities of memoryless quantum channels \( C(\Phi) \) in terms of suitable entropic functions of the input and output states of the channel. While referring the reader to Holevo 2012; Wilde 2013; Winter 1999 a for a detailed review of the subject, here are reported the main results concerning the entanglement-assisted classical capacity, the classical capacity and its one-shot version, the private classical capacity, and the quantum capacity. Limitations and applicability of these expressions for the case of memory channels will be discussed in the following sections.
for large enough \( n \), which approach \[151\] when taking the supremum over \( \mathcal{E} \).

It is important to notice that the POVM which comes with the proof of the HSW theorem is explicitly a joint one: this is the reason why the rhs of Eq. \[151\] coincides with \( C_1(\Phi) \) and not with the \( C_{cc}(\Phi) \) capacity of Sec. V.A.3 (the latter indeed is the maximum rate attainable when allowing only LOCC operations among the various channels outputs). As a matter of fact, a close scrutiny reveals that the upper bound on Eq. \[151\] instead can be obtained as a simple generalization of Eq. \[151\]. This is done by adopting a block-coding strategy which, for all \( n \), allows one to represent the density matrices produced by \( \Phi \otimes n \) acts as a single carrier map with an associated single-shot capacity \( C_1(\Phi^\otimes n) \). The resulting capacity of \( \Phi \) can then be obtained by taking the limit over \( n \) of the associated rates, i.e.,

\[
C(\Phi) = \lim_{n \to \infty} \frac{1}{n} C_1(\Phi^\otimes n). \tag{152}
\]

2. The private classical capacity theorem

The capacity formula for the private classical capacity \( C_p(\Phi) \) introduced in Sec. V.A.3 was given in Ref. \[151\]. As for the HSW theorem discussed in the previous section it is derived by first providing a closed expression for its single-shot version \( C_{p,1}(\Phi) \) (i.e., the private classical capacity attainable by using only separable codewords), and then using block coding to compute \( C_p(\Phi) \) via the identity

\[
C_p(\Phi) = \lim_{n \to \infty} \frac{1}{n} C_{p,1}(\Phi^\otimes n). \tag{153}
\]

Analogously to Eq. \[151\], \( C_{p,1}(\Phi) \) is a functional of the Holevo information \[71\]. In this case however one has

\[
C_{p,1}(\Phi) = \max_{\hat{\rho}} \left[ \chi(\mathcal{E}; \Phi) - \chi(\mathcal{E}; \hat{\Phi}) \right], \tag{154}
\]

where \( \hat{\Phi} \) is the complementary channel of \( \Phi \) as defined in Sec. \[11.5\].

The converse part of the proof which leads to Eq. \[154\] is obtained by joining the classical Fano inequality and the Holevo Bound at the output of the map \( \Phi \) with the privacy requirement imposed on the Holevo information of the complementary channel \( \hat{\Phi} \). Vice-versa the direct part of the coding theorem is based on the fact that, using a typical subspace argument, for each \( \epsilon > 0 \) and for each ensemble \( \mathcal{E} = \{p_j, \rho_j\} \) satisfying \( \chi(\mathcal{E}; \Phi) > \chi(\mathcal{E}; \hat{\Phi}) \), one can identify up to \( N \approx 2^n (\chi(\mathcal{E}; \Phi) - \chi(\mathcal{E}; \hat{\Phi})) \) distinct messages to Bob without Eve being able to read them.

3. The Lloyd-Shor-Devetak theorem

The expression for the quantum capacity \( Q(\Phi) \) of a memoryless channel \( \Phi \) is provided by the Lloyd-Shor-Devetak (LSD) theorem \[151\], according to which one has

\[
Q(\Phi) = \lim_{n \to \infty} \frac{1}{n} Q_{1}(\Phi^\otimes n), \tag{155}
\]

where for a generic channel \( \Psi \) one has

\[
Q_1(\Psi) = \max_{\rho} J(\rho; \Psi), \tag{156}
\]

with \( J(\rho; \Psi) \) is the associated coherent information \[68\] and with the maximization being performed over all input states – see also \[151\]. It is worth stressing that differently from the case of the classical capacities, the functional \( Q_{1} \) introduced above has no operational meaning. In particular \( Q_{1}(\Phi) \) doesn’t correspond to the single-shot version of \( Q(\Phi) \), this quantity having no meaning when discussing the transfer of quantum information where superposition among all the possible codewords are allowed.

The converse part of the LSD theorem can be obtained as an application of the quantum Fano inequality \[65\] which, when imposing a lower limit to the entanglement fidelity, forces the dimensionality of \( M \) to be bounded in terms of the channel coherent information \[151\]. A detailed derivation of this relation is presented for the general case of non-necessarily memoryless channels. Following \[151\] a relatively simple proof of the direct part of the theorem instead can be obtained as a modification of the coding theorem for the private classical capacity. The idea here is to extract from the codes which lead to privacy of the classical messages, those which allow also the preservation of the coherent superpositions among the various codewords. It turns out that this can be enforced by restricting the maximum in \[154\] to only those ensemble \( \mathcal{E} \) formed by pure states elements: a condition which, thanks to \[74\], allows one to identify as achievable rates for quantum communication those obtained from Eq. \[154\] in which \( \chi(\mathcal{E}; \Phi) - \chi(\mathcal{E}; \hat{\Phi}) \) gets replaced by \( J(\rho; \Psi) \).

4. The Bennett-Shor-Smolin-Thapliyal theorem and the Quantum Reverse Shannon theorem

The entanglement-assisted classical capacity of a memoryless channel is given in terms of the quantum mutual
information defined in Eq. (67): i.e.,
\[ C_{ea}(\Phi) = \max_{\rho} I(\rho; \Phi), \]  
(157)
where the maximization is over any input \( \rho \) (the corresponding quantum capacity version \( Q_{ea}(\Phi) \) being half of \( C_{ea}(\Phi) \) as already anticipated in Sec. VIII A.2). This result was proven in [Bennett et al., 1999b, 2002] by generalizing the dense-coding protocol [Bennett and Wiesner, 1992] to the case of noisy memoryless channel. In dense-coding, the sender and the receiver share a maximally entangled state in a Hilbert space of finite dimension, say \( d^2 \). The sender encodes classical information by applying \( d^2 \) generalized \( d \)-dimensional Pauli unitaries to one half of the maximally entangled states, which is then sent through the channel. These transformations maps the given states into \( d^2 \) orthogonal states, which the receiver can reliably distinguish. More generally, the sender may encode classical information by applying generic CPTP maps on her half of the maximally entangled state and then sending it through the channel. However, Bennett et al., 2002 proved that the use of encoding by generalized Pauli unitaries is optimal. This is obtained using the results of [Holevo, 1998a; Schumacher and Westmoreland, 1997] about encoding classical information into quantum states, which apply also in this settings. It is worth noticing that the encoding by generalized Pauli unitaries constrains \( \rho \) to be the maximally mixed state, \( \rho = 1/d \). This limitation is circumvented by considering the typical input states in the asymptotic limit of many uses of the channel, in which the average input state is always the maximally mixed one on the typical subspace. Remarkably, due to the subadditivity of the quantum mutual information, the expression for \( C_{ea}(\Phi) \) of a memoryless channel does not require the regularization over the channel uses.

An important property of \( C_{ea}(\Phi) \) is provided by the Quantum Reverse Shannon theorem [Bennett et al., 2009; Berta, Christandl and Renner, 2011] which generalizes the Reverse Shannon theorem discussed in the introductory paragraphs of the Sec. V. It establishes that the input-output mapping of \( n \) uses of a memoryless quantum channel \( \Phi \) can be simulated using \( nC_{ea}(\Phi) + o(n) \) uses of a noiseless (qubit or bit) channel by providing the sender and the receiver an unlimited supply of prior shared entanglement.

5. Superadditivity and superactivation

The expressions in (152), (153) and (155) for the classical, private and quantum capacity of a memoryless quantum channel require the computation of the regularized limit over the number of uses of the channel, \( n \to \infty \). This is a characteristic trait of quantum mechanics which has no counterpart in the classical Shannon theory where the input of the communication line can always be reduced to a collection of (possibly coordinated) individual channel uses. Such simplification doesn’t hold in the quantum framework due to the presence of entanglement among the various information carriers which requires a truly multipartite description of the communication process. This holistic feature of quantum mechanics is hence expressed by the inequalities \( C(\Phi) \geq C_1(\Phi) \), \( C_2(\Phi) \geq C_{p,1}(\Phi) \) and \( Q(\Phi) \geq Q_1(\Phi) \). For a given memoryless channel \( \Phi \), if the first inequality is strict, that is, \( C(\Phi) > C_1(\Phi) \], one says that the Holevo information of \( \Phi \) is superadditive, otherwise it is said to be additive. Similarly, one says that the coherent information is superadditive whenever \( Q(\Phi) > Q_1(\Phi) \), and additive otherwise – see e.g. [Smith, 2010].

At a higher level of complexity when two different channels, \( \Phi_1 \) and \( \Phi_2 \), are used in parallel the inequalities \( C(\Phi_1 \otimes \Phi_2) \geq C(\Phi_1) + C(\Phi_2) \), \( C_p(\Phi_1 \otimes \Phi_2) \geq C_p(\Phi_1) + C_p(\Phi_2) \), and \( Q(\Phi_1 \otimes \Phi_2) \geq Q(\Phi_1) + Q(\Phi_2) \) follows from simple coding arguments (for instance the rate \( C(\Phi_1) + C(\Phi_2) \) can always be attained by feeding the inputs of \( \Phi_1 \) and \( \Phi_2 \) independently with their corresponding optimal codes). If it happens that one of these inequalities is strict, then one says that the classical, resp. private, resp. quantum capacity is superadditive under tensor product of the channels \( \Phi_1 \) and \( \Phi_2 \) (notice that the additivity of (say) the Holevo capacity of \( \Phi_1 \) and \( \Phi_2 \) doesn’t necessarily guarantee the non-superadditivity of \( C \) under the tensor product \( \Phi_1 \otimes \Phi_2 \).

These (super)additivity issues are instances of a general (super)additivity problem in quantum information theory [Holevo, 2007 c]. While it was early known that the coherent information can be superadditive [DiVincenzo, Shor and Smolin, 1998; Shor and Smolin, 1999], hence the regularization over \( n \) in Eq. (155) is in general necessary, the problem of determining whether the Holevo information is additive or superadditive under tensor product of quantum channels has been for long time an open problem. The additivity of the Holevo information was shown to be equivalent to the additivity of other quantities in quantum information theory [Shor, 2004], most notably the entanglement of formation [Bennett et al., 1996] and the minimum output entropy [King and Ruskai, 2001], and was put in connection with the behaviour of a family of operator norms under composition of quantum channels [Amosov and Holevo, 2000; Hayden and Winter, 2008]. Only recently it has been established that the Holevo information can indeed be superadditive for certain quantum channels [Hastings, 2009], also implying the superadditivity of the minimum output entropy and the entanglement of formation. Extensions of the of the results presented in [Hastings, 2009] can be found in Refs. Aubrun, Szarek and Werner, 2010; Brandão and Horodecki, 2010; Fukuda, King, and Moser, 2010.

However, notwithstanding the fact that the Holevo information is generally superadditive, it has been proven to be additive for certain classes of quantum channels [Amosov and Mancini, 2009; Hiroshima, 2006], among which, qubit unital channels [King, 2002]; entanglement breaking channels [Shor, 2002a]. In particular it is worth
observing that in the case of qc-channels (16) the classical capacity formula (152) reduces to the Shannon capacity (143) for a classical channel with conditional probability \( p(y|x) = \langle e_y | E_x | e_y \rangle \). Moreover, as observed in (Holevo 2007a) if a quantum channel is (super)additive for the Holevo information, so is its complementary channel. The coherent information has been proven to be additive for degradable channels (see in Sec. [1]), for anti-degradable channels (for which the quantum capacity is always zero (Bennett, DiVincenzo, and Smolin 1997)), and for entanglement-breaking channels (Cubitt, Ruskai and Smith 2008) and PPT maps (see Sec. II.G) (Horodecki et al. 1996; Peres 1996). Regarding the private classical capacity \( C_p(\Phi) \) it is known that it coincides with its one-shot version \( C_{p,1}(\Phi) \) for degradable and anti-degradable maps (Devetak 2005): in particular in both cases one has \( C_p(\Phi) = C_{p,1}(\Phi) = Q(\Phi) = Q_1(\Phi) \), which for anti-degradable maps implies \( C_p(\Phi) = 0 \).

A remarkable example of superadditivity for the quantum capacity, called superactivation, has been provided by Smith and Yard 2008 by building up from previous results on the quantum assisted capacities (Smith, Smolin and Winter 2008). In particular, it has been shown that it is possible to find channels \( \Phi_1 \) and \( \Phi_2 \) with zero quantum capacity, i.e., \( Q(\Phi_1) = Q(\Phi_2) = 0 \), for which, by parallel use of the two communication lines \( \Phi_1 \) and \( \Phi_2 \) in \( \Phi_1 \otimes \Phi_2 \), it becomes possible to transmit quantum information, i.e., \( Q(\Phi_1 \otimes \Phi_2) > 0 \). Specifically, \( \Phi_1 \) and \( \Phi_2 \) are given by an anti-degradable channel and a PPT channel which possesses a non-zero private classical capacity (Horodecki et al. 2005; 2008). Notice that the two channels have zero quantum capacity for different reasons: the first as a consequence of the no-cloning theorem and the second due to the fact that entanglement cannot be distilled from PPT state. However, while it cannot be used to distill entanglement, there exist PPT channels that can still be used to establish a secret key between the sender and the receiver. See (Brandão, Öpenheim and Strelchuk 2012) for other examples in terms of depolarizing maps and for a more general construction.

It is finally worth noticing that in the context of zero-error classical capacity discussed in Sec. V.A.6 superactivation effects have been observed in Refs. (Duan 2009; Duan and Shi 2008).

C. Coding theorems for memory channels

The operational definitions of channel capacities, introduced in Sec. V.A apply to both memoryless and memory quantum channels. Indeed, they express the optimal classical and quantum information transmission rates between two parties, no matter how complex is the internal structure of the communication line. However, the memory setting is often more complicated than the memoryless one. One notices in particular that when dealing with non-anticipatory channels, introduced in Sec. III.B, different notions of coding procedures and capacities can be defined depending on whom, among Alice, Bob or a third party (Eve), controls (or uses for the encoding and decoding procedures) the initial and final states of the memory system \( M \). For instance for the same communication line one can introduce the classical capacities \( C_{AB} \), \( C_{AE} \), \( C_{EB,\mu} \), \( C_{EE,\mu} \), with the first (second) index representing the party controlling the initial (final) memory state and \( \mu \) being the Eve’s choice for the initial state of the memory \( M \) (when considered) – same classification holds for the other forms of capacity, i.e., quantum, private classical, etc.. The differences between these various choices have been analyzed in Ref. (Kretschmann and Werner 2005): below, for the sake of simplicity, only the situation in which the third party (Eve) has full control of the memory \( M \) will be considered.

1. Entropic bounds

The presence of correlations introduced by noise makes more remote the possibility of formalizing capacities in terms of entropic quantities when residing in the memory setting. A useful strategy is to derive bounds on the capacities, with particular attention to upper bounds, and then show whenever possible their achievability (thus providing coding theorems). The first attempts in this direction was presented in Ref. (Bowen and Mancini 2004), where for the case of finite-memory channels, i.e., maps with memory of finite dimension (described in Sec. III.D.2), the bounds of Eqs. (158) and (159) discussed below (as well as an analogous inequality for the entanglement assisted capacity) were derived and shown to be achievable for a class of Markovian channels.

Simple geometric considerations allow one to conclude that both \( C \) and \( Q \), independently from the noise model can never be larger than \( \log_2 d \), where \( d \) is the dimension of the Hilbert space of an individual carrier (the rational being that in the space of \( n \) carriers one cannot fit more than \( d^n \) orthogonal states). This threshold however is not particularly informative as it does not depend upon the CPTP mapping which describes the action of the channel (its value being achieved only by noiseless channels, i.e., by identity or unitary maps). Tighter upper limits can be derived by following the same derivation instead from the Holevo Bound (72) and from the quantum Fano inequality (65) respectively. Specifically as explicitly shown in Appendix D.1 for the classical capacity one gets

\[
C \leq \lim_{n \to \infty} \frac{1}{n} \max_{\mathcal{E}} \chi(\mathcal{E}; \Phi^{(n)}),
\]

(158)

where \( \chi(\ldots) \) is the Holevo information defined as in Eq. (151) and the maximization is performed over the ensembles of the first \( n \) input carriers. Similarly for the quantum capacity one has

\[
Q \leq \lim_{n \to \infty} \frac{1}{n} \max_{\rho} J(\rho, \Phi^{(n)}),
\]

(159)

where the maximization is now performed over the set of density matrices of the first \( n \) carriers, and where \( J(\ldots) \)
is the coherent information defined in Eq. (68) – see Appendix D.2.

By direct comparison with Eqs. (152) and (155), one notices that for memoryless channels [i.e., when \( \Phi^{(n)} = \Phi^\otimes n \)] the bounds given above coincide with the exact values of the corresponding capacities. If the channel has memory correlations however, this feature is typically lost apart from some special configurations that will be analyzed in the following (the noise correlations preventing the possibility of using block coding strategies to saturate the gap).

It is worth stressing that the inequalities (158) and (159) refer to the limit of infinite number of channel uses, i.e., they are asymptotic. Besides them, one could also consider bounds referring to a single channel use. Notice that any situation in which a channel is used a finite number of times with arbitrarily correlated noise can be equivalently described as a single use of a larger channel. Bounds on the one shot classical capacity have been found in (Wang and Renner 2012) by using relative entropy type measures defined via hypothesis testing, while bounds on the one shot quantum capacity have been derived in (Buscemi and Datta 2010) in terms of a generalization of relative Renyi entropy of zero order.

2. Perfect memory channels

Perfect memory channels admit a Kraus representation with a number of Kraus operators growing sub-exponentially with the number of channel uses \( n \) – see Sec. III.D.3. In other terms, the size of the environment is not large enough to ‘contain’ the information sent from Alice to Bob, which is exponentially increasing in \( n \), and so asymptotically the loss of information into the environment is negligible. This intuitively explains that perfect memory channels are asymptotically noiseless and have maximal capacities, i.e., \( C = Q = \log_2 d \).

Specifically, one can verify that (Giovannetti, Burgarth and Mancini 2009; Kretschmann and Werner 2005) given a perfect memory channel \( \Phi^{(n)} \) [see Eq. (80)], for sufficiently large \( n \) there exists a coding procedure which allows zero-error classical communication for reference set \( \mathcal{M} \) of size

\[
|\mathcal{M}| \geq \frac{d^n}{d_M^2},
\]

(160)
d\( M \) satisfying Eq. (94) and \( d^n \) being the size of the \( n \) carriers. The corresponding rate is hence \( R \geq \log_2 d - \frac{1}{n} \log_2 d_M \), that for \( n \to \infty \) converges to the optimal value \( \log_2 d \), implying hence \( C = \log_2 d \). Analogously, for sufficiently large \( n \) there exists a zero-error quantum communication with a reference set \( \mathcal{M} \) of dimension

\[
|\mathcal{M}| \geq \frac{d^n}{d_M^2 + d_M^2},
\]

(161)
with rate \( R \geq \log_2 d - \frac{1}{n} \log_2 [d_M^4 + d_M^2] \) which, again, for \( n \to \infty \) converges to the optimal value \( \log_2 d \).

3. Forgetful channels

Forgetful channels are characterized by the property that the effects of the initial memory state become negligible with time, i.e., memory effects die away exponentially fast, as discussed in Sec. III.D.6. This feature allows one to prove that the upper bounds of Eqs. (158) and (159) can be actually asymptotically achieved (Kretschmann and Werner 2005). This important result can be demonstrated by invoking a double-blocking encoding procedure which effectively maps forgetful channels into memoryless ones.

Consider a \( n \) fold concatenation of a memory channel \( \Phi^{(n)} \). If the channel is strictly forgetful (see Sec. III.D.6) there exists a finite integer \( m \) such that for all \( n \geq m \) the final state of the memory system does not depend on its initial state. In such a case, it is possible to group the channels \( \Phi^{(n)} \) into blocks of length \( m + l \), encoding the input in the \( l \) channels and ignoring the intermediate \( m \) ones. In such a way, the memory channel is reduced to a memoryless one defined on the larger Hilbert spaces \( \mathcal{H}_Q \otimes \mathcal{H}_M^{n+m} \), hence allowing one to extend the coding theorems for memoryless channels. Remarkably, the double-block strategy can be applied even if the channel is forgetful although not strictly forgetful. Therefore, the memoryless expressions for classical and quantum channel capacities in Eqs. (152)-(155) can be applied also in the memory setting for forgetful maps, and the entropic upper bounds in Eqs. (158) and (159) are exactly achieved (the same holds true for entanglement-assisted capacity).

Forgetful channels have been proven to constitute a dense set with the topology induced by the cb-norm distance (Kretschmann and Werner 2005). That implies that any non-forgetful channel can be approximated by a forgetful one. Notwithstanding, their capacities may be different. An example can be given in the context of Markovian channels. A long-term memory channel, Sec. III.D.8 can be approximated by a forgetful Markovian channel, Sec. III.D.4. However, according to the coding theorem for long-term memory channel discussed in the following section, the capacity of the latter does not approximate the capacity of the former (Datta and Dorlas 2009).

4. Long-term memory channels

An example of memory channels for which the bounds (158) and (159) are not tight, is provided by the long-term quantum memory channels of the form (72) – see Sec. III.D.8. In that context it is worth noticing that once the set \( \{\Phi_i\} \) contains a finite number of elements, the determination of capacities of averaged channel is equivalent to the determination of capacities of the associated compound channel (see Sec. III.A.1), since for finite sums one can always bound the error probability of the individual (memoryless) branches by the error prob-
ability of the averaged channel and vice-versa. Then, under the circumstance of \( \{ \Phi_i \}_{i=1}^{N_s} \), it has been shown that the one-shot classical capacity is given by the expression

\[
C_1 = \sup_\mathcal{E} \left( \min_i \chi(\mathcal{E} ; \Phi_i) \right),
\]

where the supremum is taken over all finite ensembles \( \mathcal{E} \) of input states \( \text{Datta and Dorlas, 2007} \). This result has been derived by employing a quantum version of Feinstein’s Fundamental Lemma \( \text{Feinstein, 1954} \) \( \text{Khinchin, 1957} \) and a generalization of Helstrom’s theorem \( \text{Helstrom, 1976} \). The basic idea is to allow Alice and Bob to use the first channels uses of the communication line to determine which, among the various possible channels \( \Phi_i \) happens to be assigned to statistical process \( p_i > 0 \) which defines the communication line via Eq. \( (177) \), and then to use proper HSW encoding to optimize the communication rate. Accordingly it is clear that the maximum rate for which reliable transmission can be guaranteed is the lowest one among those allowed by the \( \Phi_i \). Indeed by operating the channel to the highest rate allowed by the collection of maps \( \{ \Phi_i \} \) will introduce errors with finite probability.

The one-shot capacity can be generalized to give the classical capacity of the channel in the usual manner, that is, by considering inputs which are product states over uses of blocks of \( n \) channels, but which may be entangled across different uses within the same block. This yields the value

\[
C = \lim_{n \to \infty} \frac{1}{n} C_1(\Phi^{(n)}),
\]

which, in general, is explicitly smaller than the bound \( \text{Bjelaković and Boche, 2008} \). Similarly, the entanglement-assisted classical capacity has been proven to be expressed as \( \text{Datta, Sukov and Dorlas, 2008} \)

\[
C_{ea} = \sup_\rho \left( \min_i I(\rho ; \Phi_i) \right),
\]

where \( I(\rho ; \Phi_i) \) is the quantum mutual information \( \text{Datta, Sukov and Dorlas, 2008} \). Finally, \( \text{Bjelaković, Boche and Nötzel, 2009} \) provided the expression for the quantum capacity

\[
Q = \lim_{n \to \infty} \frac{1}{n} \max_\rho \left( \inf_i J(\rho ; \Phi_i^{(n)}) \right).
\]

Actually it has been shown \( \text{Bjelaković, Boche and Nötzel, 2009} \), by means of a discretization technique based on \( \tau \)-nets, that this result holds true for compound channel associated to an arbitrary set \( \{ \Phi_i \} \) (not only a finite one). Finding the best rate for quantum communication over an arbitrary set of channels can be viewed as universal coding problem. As such this result looks like a quantum channel counterpart of the universal quantum data compression result discovered in \( \text{Jozsa et al., 1998} \).

5. Ergodic cq-channels with decaying input memory

For cq-channels \( \text{Sec. II.E.} W : A^\mathbb{Z} \times B^\mathbb{Z} \to C \) which are stationary ergodic \( \text{Sec. II.C} \) and have decaying input memory \( \text{Sec. II.D.7} \), a coding theorem has been derived \( \text{Bjelaković and Boche, 2008} \) such that the classical capacity is given by

\[
C(W) = \sup_{\text{p stationary ergodic}} i(p, W),
\]

where

\[
i(p, W) := \lim_{n \to \infty} \frac{1}{n} \left( S(\rho_p^n) + S(\rho_W^n) - S(\rho_{p,W}^n) \right),
\]

with

\[
\rho_p^n = \sum_{x^n \in A^n} p^n(x^n) |x^n\rangle\langle x^n|,
\]

\[
\rho_W^n = \sum_{x^n \in A^n} p^n(x^n) \rho_{x^n},
\]

\[
\rho_{p,W}^n = \sum_{x^n \in A^n} p^n(x^n) |x^n\rangle\langle x^n| \otimes \rho_{x^n}.
\]

Here \( \rho_{x^n} \) denotes the density operator of the output state \( W^n(x^n, \cdot) \), \( x^n \in A^n \) and \( |x^n\rangle = |e_{x_1}\rangle \otimes \ldots \otimes |e_{x_n}\rangle \) for some orthonormal basis \( \{ |e_i\rangle \}_{i=1}^{A} \) of \( C^{|A|} \).

The sup in Eq. \( (166) \) is calculated over all stationary ergodic probability measures \( p \) on \( A^\mathbb{Z} \). That is, consider a shift \( T : A^\mathbb{Z} \to A^\mathbb{Z} \) of double infinite sequences of \( A \), then \( p \) is stationary if \( p(Ta) = p(a) \) for all \( a \in A^\mathbb{Z} \). Moreover, it is ergodic if for all \( a \in A^\mathbb{Z} \) such that \( Ta = a \) it is \( p(a) = 0 \) or \( 1 \).

The above theorem results as an extension of coding theorem for input memoryless cq-channel whose proof combines Wolfowitz’s code construction \( \text{Wolfowitz, 1957} \) and a version of the Feinstein’s lemma \( \text{Blackwell, Breiman and Thomasian, 1958} \) based on the notion of the joint input output probability distribution.

VI. SOLVABLE MODELS

A. Examples of solvable models for memoryless channels

This section collects examples of discrete and continuous memoryless quantum channels for which classical or quantum capacities can be analytically calculated. For most of them the calculation is made feasible by the fact that the Holevo information or the coherent information is additive. Hence the regularization in the limit of infinite \( n \) of Eqs. \( (152) \) and \( (155) \) is not necessary as the capacities equal their one-shot version.

1. Discrete variable memoryless channels

A closed expression for the classical capacity can be obtained for unital qubit channels (mapping the two-dimensional identity operator into itself – see Sec. II.1.1)
and for the depolarizing channel acting on a finite dimensional Hilbert space of arbitrary dimension. For these channels the Holevo information has been proven to be additive (King 2002, 2003).

Since any unital qubit channel is unitary equivalent to a Pauli channel (28) and as discussed at the end of Sec. V.3, capacities are invariant under unitary transformations, it is sufficient to consider the latter. As anticipated the classical capacity for these maps equals its one-shot version. The fundamental ingredient to achieve this goal is the inequality (D9) derived in Appendix D.1 (which for this special channel can be shown to be achievable), and the fact that $S_{\text{min}}(\Phi^{\otimes n})$ happens to be additive, i.e., $S_{\text{min}}(\Phi^{\otimes n}) = nS_{\text{min}}(\Phi)$. The resulting expression for the classical capacity is then computed as

$$C(\Phi) = C(\Phi) = 1 - h\left(\frac{1 + \xi}{2}\right), \quad (171)$$

where $h$ is the binary Shannon entropy (66) and $\xi$ is the maximum among $|p_0 + p_1 - p_2 - p_3|$, $|p_0 - p_1 + p_2 - p_3|$ and $|p_0 - p_1 - p_2 + p_3|$. One may notice that the capacity reaches its maximum value 1 if and only if $\xi = 1$, i.e., when one of the probabilities $p_i$ is different from zero (in this case the channel is indeed unital). Vice-versa the capacity nullify for $\xi = 0$, i.e., when $p_i = 1/4$ which correspond to a fully depolarizing qubit map sending $\rho$ into the completely mixed state $I/2$. In a similar way the classical capacity of the depolarizing channel of Eq. (36) can be shown to be

$$C(\Phi) = C(\Phi) = \log_2 d + \left[1 - \lambda - \frac{1 - \lambda}{d}\right] \log_2 \left[\frac{1 - \lambda}{d}\right] + \left[\lambda + \frac{1 - \lambda}{d}\right] \log_2 \left[\lambda + \frac{1 - \lambda}{d}\right], \quad (172)$$

and the maximum in Eq. (151) is achieved by a set of $d$ equiprobable orthogonal pure states (King 2003). Moreover, the entanglement-assisted classical capacity $C_{ea}(\Phi)$ is given by the same expression as for $C(\Phi)$ but replacing $d$ with $d^2$. A similar expression can be derived for the transpose depolarizing channel of Eq. (37) which has also been proven to have additive Holevo information (Datta, Holevo and Suhov 2006, Fannes et al. 2004).

Concerning the large class of qubit maps in Eq. (27), the full quantum capacity, corresponding to the single-shot one for the degradable case (since the coherent information is additive) and vanishing for the anti-degradable regime, is given by (Giovannetti and Fazio 2005, Wolf and Pérez-García 2007)

$$Q(\Phi) = \begin{cases} f(\theta, \phi) & \text{for } \cos(2\theta)/\cos(2\phi) > 0 \\ 0 & \text{for } \cos(2\theta)/\cos(2\phi) \leq 0 \end{cases}, \quad (173)$$

with

$$f(\theta, \phi) = \max_{q \in [0, 1]} \left[ h(q \cos^2 \theta + (1 - q) \sin^2 \phi) - h(q \sin^2 \theta + (1 - q) \sin^2 \phi) \right]. \quad (174)$$

Finally, the erasure channel introduced in Sec. II.1.2 is one of the few examples for which one can compute the whole set of capacities. This map is degradable for $p \leq 1/2$ and has a quantum capacity $Q(\Phi) = (1 - 2p) \log_2 d$ with $d$ being the dimension of the input carrier; for $p \geq 1/2$ it is, instead, anti-degradable, hence with vanishing $Q$. Its Holevo information is also additive, yielding a classical capacity equal to $C(\Phi) = (1 - p) \log_2 d$ which can also be shown to coincide with the two-way classically assisted quantum capacity $Q_2(\Phi)$. Finally the entanglement-assisted classical capacity is $C_{ea}(\Phi) = 2C(\Phi) = 2(1 - p) \log_2 d$ \cite{Ben}.

2. Continuous variable memoryless channels

The very first example of a non trivial CV memoryless channel for which the capacity has been explicitly computed was provided by \cite{Holevo et al. 1999b} who considered a cq-channel where classical messages are mapped into Gaussian states obtained by continuously displacing an assigned Gibbs reference state. This is a special example of one-mode Gaussian channels – see Sec. II.1.4. Under the memoryless condition the latter can be identified by a triad $(d^{(1)}, X^{(1)}, Y^{(1)})$, where $d^{(1)}$ is a two-component displacement vector, and $X^{(1)}, Y^{(1)}$ are $2 \times 2$ matrices. In this memoryless case, a sequence of $n$ consecutive channel uses is hence described by a triad $(d^{(n)}, X^{(n)}, Y^{(n)})$ where $d^{(n)} = \bigoplus_{k=1}^{n} d^{(1)}$, $X^{(n)} = \bigoplus_{k=1}^{n} X^{(1)}$, $Y^{(n)} = \bigoplus_{k=1}^{n} Y^{(1)}$.

Only a restricted set of Gaussian channels have been solved for their capacities. First of all, the property of (anti)degradability holds for the single-mode channels describing the process of linear attenuation and amplifica-
the Holevo information has been proven to be lossy and noisy. This is in particular true for the Gaussian channel and for the additive noise Gaussian channel, see Sec. II.I.4).

\[ Q(\Phi) = \log \eta - \log |1-\eta| \,. \] (175)

Analogously, if the mean number of bosonic excitation at the channel input is constrained to be less than \(N\) (see Sec. II.I.4), for \(\eta > 1/2\) the constrained quantum capacity reads (Holevo and Werner, 2001; Wolf, Pérez-García and Giedke, 2007)

\[ Q(\Phi, N) = Q_1(\Phi, N) = g(\eta N) - g((1-\eta)N) \,, \] (176)

where \(g(x) := (x + 1) \log_2 (x + 1) - x \log_2 x \) for \(x > 0\) and \(g(x) := 0\) for \(x \leq 0\). Concerning the classical capacity, the Holevo information has been proven to be additive for the lossy bosonic channel (\(\eta \in [0,1]\)). Under a constraint of \(N\) mean input excitation, the constrained classical capacity (Giovannetti, et al., 2004) and the constraint entanglement assisted classical capacity (Giovannetti, et al., 2003 a,b) read

\[ C(\Phi;N) = C_1(\Phi;N) = g(\eta N) \,, \] (177)
\[ C_{ca}(\Phi;N) = g(N) + g(\eta N) - g((1-\eta)N) \,, \] (178)

while their unconstrained counterparts, differently from the quantum one, are unbounded.

For the lossy channel (and for the linear amplifier channel in the case of the quantum capacity) the additivity property allows a single-letter expression for the capacity. Moreover, the optimal input states, maximizing the Holevo information or the coherent information are Gaussian states, a property which was first conjectured by Holevo and Werner (2001). It is generally believed that the same property should hold also for all Gaussian channels. However, although great efforts have been devoted to this problem (García-Patrón et al., 2012; Giovannetti, et al., 2004c, 2010; Guha, Erkmen and Shapiro, 2007; Lloyd, et al., 2009) the conjecture is still awaiting to be proved (or disproved) – see however Refs. (Giovannetti, et al., 2013; König and Smith, 2013 a,b) for a series of bounds which show that violations of the conjecture are bounded to be negligible in many case of physical interest. The validity of this conjecture is of utmost importance for the calculation of the classical capacity of certain Gaussian channels, for which it has been proven that the Holevo information is additive over Gaussian input states (Hiroshima, 2006; Serafini, Eisert and Wolf, 2005). This is in particular true for the lossy and noisy Gaussian channel and for the additive noise Gaussian channel, see Sec. II.I.4.

B. Examples of solvable models for memory channels

The main difficulty in the evaluation of the capacities of quantum channels with memory relies on the requirement of the regularization of the corresponding entropic quantities in the limit of infinite uses of the channel. For the case of forgetful channels this gives the exact expression for the capacities, while in general it provides an upper bound for non-forgetful channels.

For this reason, up to date, only few models of memory quantum channels have been fully solved in terms of capacities. One is the dephasing channel (in the discrete variable setting), and the other is the lossy bosonic channel (in the continuous variable setting), with different types of correlations.

1. Discrete memory channels

Referring to the model discussed in Sec. II.I.4 consider a sequence of qubit carriers propagating at rate \(\nu\) and interacting each one with a single qubit environment subject in turn to relaxation process described by amplitude damping with a rate characterized by \(\tau\). Then assume the carrier-environment interaction as a control-unitary such that when the carrier is in \(|0\rangle_q\) nothing happens to the environment, while when \(q_j\) is in \(|1\rangle_q\) the environment undergoes to the unitary transformation described by the operator \(\gamma \sigma_z + \sqrt{1-\gamma^2} \sigma_z\). One hence has a memory channel whenever the condition \(\nu \tau < 1\) is not satisfied. However, it is possible (Giovannetti 2005) to trace this model back to a memoryless phase damping channel \(\Phi_\tau\) with \(p_z = (1-\gamma)/2\) the probability of \(\sigma_z\) error. Here \(\gamma\) is a complicated function of several parameters including \(\nu\) and \(\tau\) and it reduces to \(\gamma\) for \(\nu \tau \ll 1\) (the memoryless limit of Section II), while it can be \(\tau > \gamma\) for \(\nu \tau \geq 1\), thus making \(\Phi_\tau\) less noisy than \(\Phi_\gamma\).

In the case of the phase damping channels (see Sec. II.I.1) the capacities can be explicitly computed. For instance, since the noise does not affect the populations associated with the computational basis, the classical capacity of the phase damping channel \(\Phi_\gamma\) is \(C(\Phi_\gamma) = 1\).

On the other hand the quantum capacity of a phase damping channel \(\Phi_\gamma\) is \(Q(\Phi_\gamma) = 1 - h(p_z)\) (Devetak and Shor, 2005; Wolf and Pérez-García, 2007) where \(h\) is the binary entropy. One hence has \(Q(\Phi_\gamma) \geq Q(\Phi_\tau)\) for \(\gamma > \gamma\), i.e., enhanced quantum capacity by memory effects.

Markovian correlated dephasing has been considered by D’Arrigo, Benenti and Falci (2007) where the degree of correlations is expressed by a correlation parameter \(\mu \in [0,1]\) which characterize the Markovian transition probabilities as in Eq. (99). Likewise in the above model, the quantum capacity increases when considering an higher degree of memory. In particular, the memoryless dephasing channel capacity is recovered for \(\mu = 0\), while for \(\mu = 1\) (perfect memory) the channel is asymptotically noiseless, i.e., \(Q(\Phi) = 1\) (Bowen and Mancini).
2004. D’Arrigo, Benenti and Falci 2007 also considered a microscopic model for correlated dephasing defined in terms of a spin-boson model, where quantum information is encoded in a train of qubits and a single bosonic mode represents the memory system. Lower bounds for the quantum capacity of a qubit memory channel with both correlated dephasing and damping have been evaluated numerically starting from a microscopic spin-boson model with Jaynes-Cummings interaction in the presence of strong dephasing noise (Benenti, D’Arrigo, and Falci 2009) [Benenti, D’Arrigo and Falci 2012].

Plenio and Virmani 2007, 2008 have considered another model of dephasing memory channel for qubits. It can be traced back to the scenario introduced in Giovannetti and Mancini 2005 and schematized in Fig. 3, where each individual information carrier (a qubit in this case) interacts with a corresponding environment particle, the correlations being established by the environment multi-particle state. Specifically they have considered the case where the two-particle (two-qubit) interaction is defined by a controlled-phase gate, the environmental particle being the controller qubit that determines which unitary transformation will be applied to the carrier. As a consequence the join state of the carriers gets transformed through mixtures of random sequences of identity and $\sigma_z$ operators, each sequence being characterized by a (correlated) probability which depends upon the diagonal elements of the environment initial state.

The interesting feature of this model is that it allows to write down explicit formulae for the associated capacities for the channel in terms of properties of the many-body environment that share a close relationship with thermodynamical quantities. In particular, the CJ state of their family of correlated states is a maximally correlated state (i.e., state of the form $\sum_{i,j} \alpha_{i,j} |i\rangle \langle j|$) [Rains 1999a, 2001], and, combining this feature with the forgetfulness of such maps, one can show (Plenio and Virmani 2007, 2008) that the quantum capacity can be expressed in terms of the regularized diagonal entropy of the system environment, i.e.,

$$Q(\Phi) = 1 - \lim_{n \to \infty} \frac{S(\text{diag}(\rho_{env}))}{n}, \quad (179)$$

where $\text{diag}(\rho_{env})$ is the environmental state in the computational basis after eliminating all off-diagonal elements [note that the coding argument used in order to arrive to Eq. (179) has been also independently shown by Hamada 2002]. For the special case in which the initial state of the environment is described by a classically correlated many-body system (i.e., diagonal in the computational basis), the last term on the right hand side of Eq. (179) coincides with the thermodynamical entropy of the environment. Hence, the capacity is given by

$$Q(\Phi) = 1 - \left(1 - \beta \frac{\partial}{\partial \beta}\right) \lim_{n \to \infty} \frac{1}{n} \log_2 Z_n, \quad (180)$$

where $Z_n$ is the partition function for $n$ environment spins, and $\beta$ is the associated inverse temperature. In other words, one can exploit results from classical statistical physics in order to compute the capacity, as shown by Eq. (180).

The calculation of the entropy of the associated many-body system, and hence of the quantum capacity of the memory channel, can be done exactly in certain relevant cases. One of them is the case of many-body systems described by Matrix Product States (MPS) involving only rank-1 matrices. For a sake of simplicity, one does focus on a translationally invariant MPS for a 1D system of 2-level particles, with periodic boundary conditions. This environmental state is characterized by two matrices $A_0$ and $A_1$ and is given by the following expression $|\psi\rangle = \sum_{i_1...i_n} \text{Tr}(A_{i_1} ... A_{i_n}) |i_1 ... i_n\rangle$. Then, by dephasing each qubit, the resulting unnormalized state is

$$\rho = \sum_{i_1...i_n} \text{Tr} \left[ \prod_{k=1}^{n} (A_{i_k} \otimes \bar{A}_{i_k}) \right] |i_1 ... i_n\rangle \langle i_1 ... i_n|,$$

where $\bar{A}$ is the complex conjugate matrix of $A$. It is possible to show that, if $|i_1 ... i_n\rangle$ has $l$ occurrences of 0 and $n-l$ of 1, and $k$ boundaries between 0s and 1s blocks, then the corresponding diagonal elements of $\rho$ are proportional to $a^l b^{n-l} c^k$, with $a$ (resp. $b$) being the eigenvalue of $A_0 \otimes A_0$ (resp. $A_1 \otimes A_1$), and $c$ being the eigenvalue of $(A_0 \otimes A_0)(A_1 \otimes A_1)/(ab)$.

Finally, it is worth remarking that Wolf et al. 2006 showed the existence of Hamiltonians exhibiting quantum phase transitions and with ground states being Matrix Product States (MPS) involving only matrices of rank-1. Hence, it can be shown that the diagonal elements of such MPSs are equal to the probability $\varphi$ of microstates in corresponding classical Ising chains. Therefore, by exploiting this connection, one can easily compute the limit in Eq. (179) by using well known many-body physics methods. Fig. 9 shows the case of the following Hamiltonian

$$\sum_i 2(g^2-1)\sigma_{z,i}\sigma_{z,i+1}-(1+g)^2\sigma_{x,i}+(g-1)^2\sigma_{x,i}\sigma_{x,i}\sigma_{z,i+1}.$$  

In this case, one knows that the ground state is a rank-1 MPS which possesses a non-standard ‘phase transition’ at $g = 0$, where indeed some correlation functions are non-differentiable (though continuous) and the ground state energy is analytic (Wolf et al. 2006).

2. Continuous memory channels

Among Gaussian memory channels, one can identify a subclass of channels for which the memory effects can be unraveled. That is, by applying suitable unitary encoding and decoding transformations, $n$ uses of such channels are unitary equivalent to $n$ independent single-mode
channels used in parallel. By applying known results for the memoryless setting one may then compute the capacities of the memory channel or estimate lower bounds in terms of the Gaussian capacities (see Sec. VI.A.2).

Such a unitary mapping from \( n \) uses of a Gaussian memory channel to \( n \) parallel uses of independent single-mode channels was first considered in (Cerf et al., 2005; Giovannetti and Mancini, 2005), and then applied for estimating the communication capacities of Gaussian memory channels in several settings (Lupo, Memarzadeh, and Mancini, 2009; Lupo, Pilyavets and Mancini, 2009; Lupo, Giovannetti and Mancini, 2010). A formal definition of the class of memory channels that can unraveled first appeared in (Lupo and Mancini, 2010).

If one takes the one-mode channel as a reference point, representing a single use of the channel, \( n \) uses of the quantum memory channel are characterized by the triads \((d^{(n)}, X^{(n)}, Y^{(n)})\) such that either \(d^{(n)} \neq \bigoplus_{k=1}^{n} J^{(1)}\) or \(X^{(n)} \neq \bigoplus_{k=1}^{n} X^{(1)}\). A memory channel can be unraveled if there exist unitary transformations \(\Phi^{(n)}_E, \Phi^{(n)}_D\) acting on \(n\) modes, such that \(\Phi^{(n)}_D \phi^{(n)}_E \Phi^{(n)}_E = \bigotimes_{k=1}^{n} \phi^{(1)}_k\), that is, \(n\) uses of the memory channel are unitary equivalent to the tensor product of \(n\) independent – but not necessary identical – single-mode Gaussian channel (this mapping is depicted in Fig. 10). Since the application of unitary transformations cannot change the capacities of the channel, they can be equivalently computed for the unraveled channel, in which each input mode is transformed independently (although in general not-identically). If one is interested in the calculation of constrained capacities, then one has to take in account how the constraint changes under the action of the encoding and decoding unitaries. A relevant setting is that of encoding and decoding transformations preserving the constraint. For the case of constrained mean input excitation-number, the constraint is preserved if \(\sum_{k=1}^{n} a_k^\dagger a_k = \Phi^{(n)}_E \sum_{k=1}^{n} a_k^\dagger a_k \Phi^{(n)}_E\). This is the case when the encoding unitary is a linear passive transformation. For optical realization, such transformations are implemented by a network of beam-splitters and phase-shifters (see e.g., (Ferraro, Olivares and Paris, 2005)).

As discussed in Sec. VI.A.2 the only non-trivial models of single-mode memoryless Gaussian channels that have been solved for the capacity are the lossy channel (for both the classical and quantum capacities) and the linear amplifier channel (for the quantum capacity). A solution is hence suitable for memory channels that are mapped to tensor product of lossy or amplifier channels upon unraveling. This is the case for the model of lossy channel with memory introduced in (Lupo, Giovannetti and Mancini, 2010). In this model, the action of the channel upon \( n \) uses is defined by the concatenation of \( n \) identical unitary transformations coupling the input modes \( a_1, a_2, \ldots, a_n \) with a collection of local environments \( e_1, e_2, \ldots, e_n \) and the memory system \( m \), where both the local environments and the memory system are represented by bosonic modes. Specifically the evolution of the \( k \)-th input mode is obtained by a concatenation of two beam-splitter transformations, the first with transmissivity \( \epsilon \) and the second with transmissivity \( \eta \), see Fig. 11. This results in a non-stochastic channel with ISI (see Sec. III.D) having the same structure depicted in Fig. 2 c). By varying the transmissivity parameters, the model may be reduced to a memoryless lossy bosonic channel (Giovannetti, et al., 2004) (the input \( a_k \) only influences the output \( b_k \)) or to a channel with perfect memory (all \( a_k \) interacts only with the memory mode \( m_1 \)) (see Sec. III.D.3). For specific values of the parameters (that is, \( \eta = 0, \epsilon = 1 \)) \(\Phi^{(n)}\) describes a quantum shift channel (Bowen and Mancini, 2004), where each input state is replaced by the previous one. Such a model was extended in (Lupo, Giovannetti, and Mancini, 2010) to encompass memory effects in linear amplification processes.

Both the lossy memory channel of (Lupo, Giovannetti, et al., 2004)

FIG. 10 Unraveling of \( n \) uses of a memory channel. Each horizontal line indicates one bosonic mode, propagating from the left to the right. \(\phi^{(n)}_k\) denotes \( n \) uses of the memory channel. \(E^{(n)}\) and \(D^{(n)}\) are pre-processing and post-processing Gaussian unitaries. \(\phi^{(1)}_k\)'s are one-mode Gaussian channels.
and Mancini, 2010) and the amplifier memory channel of (Lupo, Giovannetti and Mancini, 2010b) can be unraveled into the tensor product of one-mode lossy or amplifier channels. The capacities of these memory channels can hence be computed following four steps: first the memory channel is unraveled into the direct product of the single-mode Gaussian channels; second the optimization of the relevant entropic function is performed mode-wise under constrained mean input excitation number; then the distribution of the mean excitation number over the input modes is optimized; finally the asymptotic limit of infinite channel uses is considered. The optimization of the distribution of the mean excitation number leads to a quantum water filling solution for the capacity of the memory channel, where the way the mean excitation number is distributed over input modes is analogous to the way water distributes into a vessel (Cover and Thomas, 1991). While algorithms for the optimization were presented in (Pilyavets, Lupo and Mancini, 2009; Schäfer, Karpov and Cerf, 2011), the most delicate point is the consideration of the asymptotic limit (Lupo, Memarzadeh and Mancini, 2009; Lupo, Giovannetti and Mancini, 2010).

The same approach can be pursued for other memory channels that can be unraveled into the direct product of Gaussian channels different that the lossy or amplifier channel. Models of Gaussian memory channels with additive noise and with loss and noise have been considered. In these cases, the procedure outlined above has been applied to evaluate the Gaussian capacity of the memory channel (see Sec. VI.A.2).

Memory channels with additive noise are characterized by having $X^n = 1$. It has been pointed out that these kind of correlated channel can be realized by means of multimode CV teleportation protocol (Ban, Sasaki, and Takeoka, 2002; Braunstein and Kimble, 1998; Vaidman, 1994), where the teleportation resource is a multimode state (Caruso, Giovannetti and Palma, 2010). The memory channel considered in (Cerf et al., 2005, 2006) belongs to this class. The latter was defined for two channel uses, represented by two bosonic modes, which are affected by correlated additive noise. A generalization of this model to the case of more than two channel uses was first introduced in (Ruggeri and Mancini, 2007a) and subsequently in (Lupo, Memarzadeh and Mancini, 2009; Schäfer, Karpov and Cerf, 2009), where the additive noise, characterized by the matrix $Y^{(n)}$, constitutes a Markov process.

A model of lossy and noisy Gaussian memory channels was first introduced in (Giovannetti and Mancini, 2005). In this model, upon $n$ uses of the memory channel, $n$ input modes interact mode-wise with $n$ environmental modes by a beam-splitter transformation with given transmissivity and phase. This model is characterized by the beam-splitter transmissivity $\eta$ and by the matrix $Y^{(n)} = (1 - \eta)C^{(n)}$, where $C^{(n)}$ is the CM of the environmental state.

It is easy to recognize that these models are SI memory channels, which are instances of the general scheme depicted in Fig 3. These Gaussian memory channels can be unraveled whenever the matrix $Y^{(n)}$ have a suitable form. Under certain conditions they can be unraveled with the use of energy-preserving unitary pre-processing transformation (Lupo, Pilyavets and Mancini, 2009; Pilyavets, Zborovskii and Mancini, 2008) (see also (Ruggeri et al., 2005)).

It worth noticing that, differently from the case of discrete-variable memory channels (see Sec. III.D.4), there is no transitional behavior in these models of Gaussian memory channels: the optimal input states are either separable or entangled according to the model symmetries (Cerf et al., 2005, 2006; Lupo and Mancini, 2010). As entangled states cannot be prepared locally, it is crucial to identify suboptimal input states that can be prepared efficiently. This issue was considered in (Schäfer, Karpov and Cerf, 2011), where it was shown that encoding classical information via Gaussian matrix-product states (Adesso and Ericsson, 2006; Schuch, Cirac and Wolf, 2008), which can be efficiently prepared, may allow to achieve a reliable communication rate close to the channel capacity. An analysis of correlated additive Gaussian channels beyond the case of Markovian correlations was presented in (Schäfer, Karpov and Cerf, 2011).

Finally, it is worth remark that the study of models of Gaussian memory channels that can be unraveled has also stimulated and motivated a deep analysis of the communication capacities of the single-mode memoryless Gaussian channel (Lupo, et al., 2011; Pilyavets, Lupo and Mancini, 2009; Schäfer, Karpov and Cerf, 2010). In particular (Pilyavets, Lupo and Mancini, 2009) and (Schäfer, Karpov and Cerf, 2010) provided a complete characterization of one-mode Gaussian channels, respectively for the case of lossy channels and additive noise, in terms of the solutions of the optimization problem of computing the Gaussian classical capacity.
As discussed in Sec. III.D the concatenation of CPTP maps define a new quantum channel. It is also worth considering whether the converse is also true, that is, under which conditions a quantum channel \( \Phi \in \mathbb{P} \) acting on a system \( Q \mapsto Q \) can be expressed as a concatenation of other elements of \( \mathbb{P} \). This is intimately related to the semigroup structure of the set of quantum channels, hence with dynamical maps and master equations.

### VII. QUANTUM CHANNELS DIVISIBILITY AND DYNAMICAL MAPS

#### A. Divisible and indivisible quantum channels

Loosely speaking, by divisibility of a quantum channel \( \Lambda \in \mathbb{P} \) one refers to the possibility of decomposing it in terms of concatenation of other channels, i.e., to the possibility of writing \( \Lambda = \Lambda_1 \circ \Lambda_2 \), with \( \Lambda_i \in \mathbb{P} \). Obviously, every channel \( \Lambda \in \mathbb{P} \) is divisible in the following way: \( \Lambda = (\Lambda \circ \mathcal{U}^{-1}) \circ \mathcal{U} \), with \( \mathcal{U} \) any unitary map. A non trivial definition of (in)divisibility has been introduced by [Wolf and Cirac 2008]. According to that, a quantum channel \( \Lambda \in \mathbb{P} \) is indivisible if every decomposition of the form \( \Lambda = \Lambda_1 \circ \Lambda_2 \), with \( \Lambda_i \in \mathbb{P} \), implies that either \( \Lambda_1 \) or \( \Lambda_2 \) is a unitary conjugation. Otherwise \( \Lambda \) is said to be divisible. It happens that quantum channels with full Kraus rank \((d^2)\) are all divisible [Wolf and Cirac 2008].

The subset of \( \mathbb{P} \) given by the divisible quantum channels is denoted below by \( \mathcal{D} \). The notion of divisibility can then be refined by considering different kinds of divisible quantum channels. First, one introduces a notion of Markovianity for quantum channels related to their decomposability, rather than to their composability as done in Sec. III.D. According to [Wolf and Cirac 2008] a quantum channel is called Markovian if it is an element of a completely positive continuous one-parameter semigroup.

In such a case there exists a (Liouvillean) generator \( \mathcal{L} \) such that the quantum channel can be written as \( \Lambda(t) = e^{t\mathcal{L}} \in \mathbb{P} \) for all \( t \geq 0 \). A standard form for such generators was derived in [Gorini, Kossakowski and Sudarshan 1976; Lindblad 1976]:

\[
\mathcal{L} \rho = i [\rho, H] + \sum_{\alpha, \beta} G_{\alpha, \beta} \left( F_{\alpha} \rho F_{\beta}^\dagger - \frac{1}{2} \{ F_{\beta}^\dagger F_{\alpha}, \rho \} \right)
\] (183)

where \( G \geq 0 \), \( \{ , \} \) denotes the anti-commutator and the operators \( H \) and \( F_{\alpha} \) respectively describe the Hamiltonian and non-Hamiltonian dynamical terms.

Through the (Liouvillean) generator \( \mathcal{L} \) one can write down the dynamical (master) equation for the system density operator \( \rho \) [Breuer and Petruccione 2002]

\[
\frac{d}{dt} \rho(t) = \mathcal{L} \rho(t).
\] (184)

Its solution, for given initial condition \( \rho(t_0) \), reads

\[
\rho(t) = \Lambda(t-t_0) \rho(t_0) \quad \text{with} \quad \Lambda(t-t_0) = e^{(t-t_0)\mathcal{L}}
\]

obeying the homogeneous composition law

\[
\Lambda(t_1) \circ \Lambda(t_2) = \Lambda(t_1 + t_2),
\] (185)

for \( t_1, t_2 \geq 0 \), hence defining a one-parameter semigroup of CPTP maps. As consequence, Eq. (184) is called Markovian master equation.

A class of Markovian master equations of this kind can be obtained as the continuous-time limit of a concatenation of identical system-bath interactions. These models, known as collision models [Alicki and Lendi 1987; Rau 1963; Scarani et al. 2002; Terhal and DiVincenzo 2000; Ziman and Bužek 2005; Ziman et al. 2002, 2003], are defined by the iterated unitary interactions of the system \( Q \) with \( n \) identical reservoirs \( \mathbb{E} = (e_1, \ldots, e_n) \). This cascade process, depicted in Fig. 12 defines a quantum channel of the form

\[
\Phi^n(\rho_Q) = \text{Tr}_E \left[ U_{Q_{e_1}} \cdots U_{Q_{e_n}} (\rho_Q \otimes \omega_E^\otimes n) U_{Q_{e_n}}^\dagger \cdots U_{Q_{e_1}}^\dagger \right],
\] (186)

where \( U_{Q_{e_j}} \)’s indicate the identical unitary transformations coupling the system \( Q \) with the environmental systems. A comparison with Fig. 2 is useful to enlighten the relations between this model and the unitary dilation of memory channels introduced in Eq. (50): basically, in passing from the latter to Eq. (186), the environment and the carriers have exchanged their roles transforming the spatial correlations of Eq. (80) into temporal correlations. An hybrid approach which includes both effects has been recently introduced in [Giovannetti and Palma 2012]: as shown in Fig. 13 the scheme has the same structure of Fig. 12 for each row, and the same of Fig. 2a for each column. This model provides a link between memory channels and time-continuous dynamical evolutions.

The set of Markovian quantum channels is denoted below by \( \mathfrak{M} \). Clearly \( \mathfrak{M} \subset \mathcal{D} \) because any Markovian quantum channel can be divided into a large number of equal infinitesimal channels being it the solution of the (time-independent) master equation (184).

Then one can attempt to single out the class of quantum channels that can be split into infinitesimal pieces, i.e., into channels arbitrary close to the identity. Clearly it would contain \( \mathfrak{M} \). Actually the set \( \mathfrak{F} \) of infinitesimal divisible quantum channels can be defined [Wolf and Cirac 2008] as the closure of the set of all families \( \{ \Lambda(t_2, t_1) \in \mathbb{P} | t_2, t_1 \in [0, t] \} \) of quantum channels for which there exists a continuous mapping \([0, t] \times [0, t] \rightarrow \mathbb{P}\) onto \( \{ \Lambda(t_2, t_1) \} \) such that

![FIG. 12 The cascade structure of a collision model, defined by the concatenation of identical unitaries (Scarani et al. 2002).](image-url)
FIG. 13 The cascade structure leading to the master equation for correlated quantum channels discussed in Giovannetti and Palma 2012, described by Eq. (186). Each row corresponds to a single collision model (see Fig. 12), and each column correspond to a memory channel (see Fig. 2a).

1) \( \Lambda(t_3, t_2) \circ \Lambda(t_2, t_1) = \Lambda(t_3, t_1) \), for all \( 0 \leq t_1 \leq t_2 \leq t_3 \leq t \),

2) \( \lim_{\tau \to 0} |||\Lambda_{\tau + \varepsilon, \tau} - \text{id}|||_2 = 0 \), for all \( \tau \in [0, t) \),

where \(||| \cdot |||_2\) is the superoperator norm defined in Eq. (56) of Appendix 11.2 – the closure being intended with respect to the associated distance. For a given family it is like to say that there is a continuous path in \( \mathcal{P} \) (where one can move by concatenating quantum channels) connecting any element of the family with the identity.

Actually one could consider in the above definition a set \( \mathcal{J}_M \) analogous to \( \mathcal{J} \) with the restriction \( \Lambda \in \mathcal{M} \), i.e., of the form \( \Lambda(t) = e^t \mathcal{L} \). It will obviously be \( \mathcal{J}_M \subseteq \mathcal{J} \). Intuitively also the converse should be true since any quantum channel close to the identity is ‘almost Markovian’ according to the definition of Markovian quantum channel. In fact, it has been proven in Wolf and Cirac 2008 that any infinitesimally divisible quantum channel can be (arbitrary well) approximated by a product of Markovian quantum channels.

In summary one has the following chain of inclusion \( \mathcal{M} \subset \mathcal{J} \subset \mathcal{D} \subset \mathcal{P} \). The complement of \( \mathcal{D} \) to \( \mathcal{P} \) is given by the indivisible quantum channels.

**B. Non-Markovian master equations**

The simplest generalization of the dynamical equation (184) is obtained by introducing a time-dependent Liouvillian \( \mathcal{L}(t) \) admitting the representation (183), but with time-dependent operators, \( H(t) \) and \( F_\alpha(t) \). Hence, the time-dependent equation for the dynamical map \( \Lambda(t, t_0) \)

\[
\frac{d}{dt} \Lambda_{t, t_0} = \mathcal{L}(t) \circ \Lambda(t, t_0) \quad \Lambda(t_0, t_0) = \text{id} \quad (187)
\]

has formal solution

\[
\Lambda(t, t_0) = \mathcal{T} \exp \left( \int_{t_0}^{t} \mathcal{L}(\tau) d\tau \right) \quad (188)
\]

where \( \mathcal{T} \) denotes time-ordering. Differently from the time-homogeneous case (185), the explicit dependence on time implies that the dynamical map \( \Lambda(t, t_0) \) is no more a function of ‘\( t - t_0 \)’ only. Notwithstanding, it still satisfies the \textit{inhomogeneous} composition law

\[
\Lambda(t, s) \circ \Lambda(s, t_0) = \Lambda(t, t_0) \quad (189)
\]

for any \( t \geq s \geq t_0 \). The Markovian character is hence preserved by the time-dependent dynamical equation (187) and it implies the infinitesimal divisibility discussed in Section VII.A. This is obviously true if one intends the Markovian character simply expressed by an associative binary operation like (189) (a quantum version of the Chapman-Kolmogorov equation). However it results that the Chapman-Kolmogorov equation is a necessary but not sufficient condition for having Markov chains (processes) Vacchini et al. 2011.

On the other hand, from the fact that any infinitesimal divisible quantum channel can be (arbitrarily well) approximated by a product of Markovian quantum channels (as discussed at the end of Sec. VII.A), it follows that every infinitesimally divisible quantum channel can be written as a solution of time-dependent master equation Wolf and Cirac 2008 proved this fact for \( d = 2 \) and argued the same for \( d > 2 \). Hence, loosely speaking one can say that the class of infinitesimal divisible channels corresponds to the set of solutions of time-dependent master equations.

A more general dynamical equation comes from the Nakajima-Zwanzig projection operator technique Breuer and Petruccione 2002 Nakajima 1958 Zwanzig 1960 and reads as follows:

\[
\frac{d}{dt} \rho(t) = \int_{t_0}^{t} \mathcal{K}(t - u) \rho(u) du \quad \rho(t_0) = \rho_0 \quad (190)
\]

Here one has memory effects modeled by the \textit{memory kernel} super-operator \( \mathcal{K}(t) \). Hence, the rate of change of the state at time also depends on its history, and the Markovian setting (184) is recovered when \( \mathcal{K}(\tau) = 2b(\tau) \mathcal{L} \).

The dynamical map \( \Lambda(t, t_0) \) associated to the non-Markovian evolution (190) is a solution of

\[
\frac{d}{dt} \Lambda(t, t_0) = \int_{t_0}^{t} d\tau \mathcal{K}(t - \tau) \circ \Lambda(\tau, t_0) \quad \Lambda(t_0, t_0) = \text{id} \quad (191)
\]

It appears to be a function of both \( t_0 \) and \( t \). However, one can notice that the dynamics of an open quantum system can be always understood as the reduced dynamics of its unitary dilation (see Sec. II) which includes the environment. Being the unitary dynamics of an isolated system homogeneous in time, it follows that, once the degrees of freedom of the environment are taken into account, the
dynamical map will be only a function of the difference \( t-t_0 \), that is, \( \Lambda(t, t_0) \equiv \Lambda(t-t_0) \). This mirrors the fact that any solution of (191) is also a solution of the time-dependent equation (Chruściński and Kossakowski, 2010)

\[
\frac{d}{dt} \Lambda(t-t_0) = \mathcal{L}(t, t_0) \circ \Lambda(t, t_0), \quad \Lambda(t_0, t_0) = \text{id}, \quad (192)
\]

with a time-dependent Liouvillian defined by the logarithmic derivative of the dynamical map \( \mathcal{L}(t-t_0) := \left( \frac{d}{dt} \Lambda(t-t_0) \right) \circ \Lambda^{-1}(t-t_0) \). Nevertheless, the explicit dependence of the generator on the initial time \( t_0 \) implies that \( \mathcal{L} \) is effectively non-local in time. Although the formal solution of (192) is analogous to (188), it does not satisfy the composition law (189), a fact which represents a signature of memory effects.

Then, a fundamental problem is to find those conditions on the memory kernel \( K(t) \) that ensure that the time evolution map \( \Lambda(t, t_0) \) is CPTP, i.e., a quantum channel.Contrary to the Markovian case, a full characterization of legitimate memory kernels is still missing.

In (Chruściński and Kossakowski, 2012), a class of memory kernels giving rise to legitimate quantum dynamics (quantum channels) has been provided. The construction is based on a simple idea of normalization: starting from a family of (possibly non-trace-preserving) CPTP maps satisfying a certain additional condition one is able to ‘normalize’ it in order to obtain a legitimate dynamics, i.e., a CPTP map. Non-Markovian master equations have been also described in Ref. (Ciccarello et al., 2013, Rybar et al., 2012) by generalizing the collision models discussed in the previous section and in (Shabani and Lidar, 2005) exploiting adaptive strategies that involve the measurements of the system environment followed by local transformations.

C. Markovian vs non-Markovian dynamics

Given a CPTP map, the problem of determining whether or not it admits an infinitesimal generator of the form (183), has been proven to be computationally hard (Cubitt, Eisert and Wolf, 2012, Wolf et al., 2008).

For CPTP maps that do not belong to \( \mathcal{M} \), a measure of non-Markovianity has been introduced in (Wolf et al., 2008) in terms of the minimal amount of white noise \( \mathcal{L}_\mu \) that has to be added in order to make \( \log \Lambda + \mathcal{L}_\mu \) of the form (183).

Besides the Markovianity definition given in Section VII.A and the above mentioned quantifier of (non)Markovianity other proposals have been put forward, see e.g. Refs. (Breuer, Laine and Piilo, 2009, Lu, Wang and Sun, 2010, Luo, Fu and Song, 2012, Rivas, Huelga and Plenio, 2010).

On one hand, Rivas, Huelga and Plenio (2010) considered the equivalence between Markovian dynamics and infinitesimal divisibility and introduced a measure of deviation from it. Given a maximally entangled state \( \beta \) of the system of interest and a suitable ancillary system, due to the Choi-Jamiolkowski isomorphism (9), \( \Lambda(t+\epsilon, t) \) is a CPTP map iff \( \langle \beta | \Lambda(t+\epsilon, t) \otimes \text{id} | \beta \rangle \geq 0 \). Then, one can consider \( \| \langle \Lambda(t+\epsilon, t) \otimes \text{id} | \beta \rangle_1 \| \) as a measure of the non-CPTP character of \( \Lambda(t+\epsilon, t) \). In fact, due to the trace preserving property, this quantity equals 1 if \( \Lambda(t+\epsilon, t) \) is CPTP, otherwise it is greater than 1. Actually, the derivative of this quantity has been considered

\[
g(t) := \lim_{\epsilon \to 0} \frac{\| \langle \Lambda(t+\epsilon, t) \otimes \text{id} | \beta \rangle_1 \| - 1}{\epsilon} . \quad (193)
\]

It happens that \( g(t) > 0 \) iff the map \( \Lambda \) is indivisible.

On the other hand (Breuer, Laine and Piilo, 2009) used a fundamental property of CPTP maps, namely the fact that they cannot increase the trace-distance

\[
D(\Lambda(t, 0)(\rho_1), \Lambda(t, 0)(\rho_2)) \leq D(\rho_1, \rho_2) , \quad (194)
\]

for any pair of states \( \rho_1, \rho_2 \). If a family of CPTP maps is infinitesimally divisible, the monotonicity of the trace distance holds true locally, that is,

\[
\frac{d}{dt} D(\Lambda(t, 0)(\rho_1), \Lambda(t, 0)(\rho_1)) \leq 0 . \quad (195)
\]

According to that the dynamical map \( \Lambda(t, 0) \) is said to be non-Markovian if there exists a value of \( t \) such that Eq. (195) is violated, for some initial states \( \rho_1, \rho_2 \). Physically, this implies a temporal increase in the distinguishability of the two quantum states, a consequence the backflow of information from the surrounding environment.

The criteria relying on (193) and (195) allow one to define computable measure of non-Markovianity. A natural quantifier derived from the criterion of (Rivas, Huelga and Plenio, 2010) reads

\[
N_{\text{RHP}}(\Lambda) = \frac{\int_0^\infty g(t)dt}{1 + \int_0^\infty g(t)dt} , \quad (196)
\]

where \( g(t) \) is as in Eq. (193). From the criterion of (Breuer, Laine and Piilo, 2009) one defines the non-Markovianity quantifier

\[
N_{\text{BLP}}(\Lambda) = \sup_{\rho_1, \rho_2} \int dt \left[ \frac{d}{dt} D(\Lambda(t', 0)(\rho_1), \Lambda(t', 0)(\rho_1)) \right]_{t'=t} dt , \quad (197)
\]

where the integral is performed only for those \( t \) such that Eq. (195) is violated.

It has been pointed out that the relation these two criteria resembles that between separable and PPT states in entanglement theory (Chruściński, Kossakowski and Rivas, 2011). Indeed, any family of CPTP maps which is Markovian according to the first criterion is as well Markovian according to the second one, that is, \( N_{\text{RHP}}(\Lambda) = 0 \) implies \( N_{\text{BLP}}(\Lambda) = 0 \), while the converse is in general not true. An example comparing non-divisibility and non-Markovianity, for the case of Gaussian channels, has been recently discussed in Ref. (Benatti, Floreanini, and Olivares, 2012) while a test of
VIII. SUMMARY AND OUTLOOK

In the last decades, the subject of quantum channels has become prominent for its usefulness in foundational issues (Kraus, 1983) as well as in technological applications (see the latest striking experiments in quantum communication (Arshed, Toor, and Lidar, 2010). Its quantum capacity behavior as function of time is strongly dependent on the couplings parameters and on the temperature of the bath. For generic values of these parameters, recurrence in the quantum capacity as function of time is of small amplitude and quickly vanishes. On the contrary, for commensurable values of these parameters the quantum capacity becomes a periodic function of time. This feature indicates the backflow of information from the environment to the central spin: a signature of non-Markovian dynamics. This is also related to the increased distinguishability of states pointed out by the non-Markovianity criterion introduced by (Breuer, Laine, and Piilo, 2009).

Examples of quantum channels showing memory effects are abundant in quantum information processing. An unmodulated spin chain has been proposed as a model for short distance quantum communication (Bose, 2003). In such a scheme, the state to be communicated over the channel is placed on one of the spins of the chain, propagates for a specific amount of time, and is then received at a distant spin of the chain. When viewed as a model for quantum communication, it is generally assumed that a reset of the spin chain occurs after each signal, for instance by applying an external magnetic field, resulting in a memoryless channel. However, a continuous operation without resetting corresponds to a quantum channel with memory (Bayat et al., 2008). Another model of a quantum channel with memory is the so-called one-atom maser or micromaser (Benenti, D’Arrigo, and Falci, 2009). In such a device, excited atoms interact with the photon field inside a high-quality optical cavity. If the photons inside the cavity have sufficiently long lifetime, atoms entering the cavity will feel the effect of the preceding atoms, introducing ISI correlations (see Sec. III.B) among consecutive signal states.

Another source of correlated noise in the propagation of the electromagnetic field is due to atmospheric turbulence, whose effects on the signal propagation can be modeled as random changes of the channel’s characteristics (Semenov and Vogel, 2009, 2010). Moreover, the decoherence induced by atmospheric turbulence introduces cross talks (Boyd et al., 2011; Tyler and Boyd, 2009), i.e., ISI correlations (see Sec. III.B), when information is encoded in the transverse degrees of freedom of the electromagnetic field, e.g., the orbital angular momentum.

Furthermore, the propagation of the quantum electromagnetic field in linear dispersive media, including the free-space propagation and through linear optical systems, can be described by a quantum channel with memory (Giovannetti, et al., 2004b; Lupo, et al., 2011b; Shapiro, 2009), where wave diffraction introduces memory effects.
Memory effects also arise in the context of quantum cryptography. Quite generally, one can categorize the collective attacks within the framework of memoryless channels, while the coherent attacks within the memory channels framework (Gisin et al., 2002; Scarani et al., 2009). However, this link has been subjected to limited attention and probably needs further explorations. Actually, in one-way quantum key distribution memory effects that introduce correlations among transmitted symbols can advantage the eavesdropper (Ruggeri and Mancini, 2007b). Only if the legitimate users have the control of the noise correlations, by properly tuning them, they can reduce eavesdropper information (Vasile et al., 2011a). Instead, in two-way quantum key distribution checking the presence or absence of noise correlations can help in countering eavesdropper attacks (Pirandola et al., 2008). Then, an analysis of memory effects in other channel uses-configurations, like zero error channel capacity, channels with feedback, channels with unknown parameters and multiuser channels, should be pursued.

Finally, moving to the framework of time-continuous quantum evolution, non-Markovian effects are relevant in several physical systems characterized by the interaction with a structured environment. Examples are in the framework of solid state physics, as quantum dots in photonic crystals (Madsen et al., 2011; Vats, John and Busch, 2002), and in the soft matter framework as the case of exciton dynamics surrounding by their protein environment (Caruso et al., 2009; Caruso, Huelga and Plenio, 2010; Plenio and Huelga, 2008; Reubentrost, Chakraborty and Aspuru-Guzik, 2009; Thorwart et al., 2009).

In summary, more efforts are needed to gain a full understanding of quantum channels, however the presented work constitutes a rather general frame where the still missing dowels of the puzzle could be settled.

### Acknowledgments

The authors would like to thank K. Banaszek, N. J. Cerf, D. Chruscinski, N. Datta, S. F. Huelga, C. Macchiavello, O. Pilyavets, M. B. Plenio, A. Shabani, B. Vacchini and S. Virmani, for useful discussions and for careful reading the manuscript. S.M. acknowledges fruitful and enjoyable collaboration with Garry Andrew Bowen though whom he became acquainted with the subject of quantum channels. The work of F.C. has been supported by EU FP7 Marie-Curie Programme (Intra-European Fellowship and Career Integration Grant) and by MIUR-FIRB (Project No. RBFR10M3SB). The QSTAR is the MPQ, LENS, IIT, UniFi Joint Center for Quantum Science and Technology in Arectri. The work of C.L. and S.M. has been supported by the European Commission, under the FET-Open grant agreement CORNER, number FP7-ICT-213681.

### Appendix A: Distance measures

A proper way to measure the distance between two quantum states \( \rho_1, \rho_2 \in \mathcal{S}(\mathcal{H}_Q) \) of a quantum system \( Q \), is provided by the trace distance defined as:

\[
D(\rho_1, \rho_2) := \frac{1}{2} \| \rho_1 - \rho_2 \|_1 ,
\]

with \( \|O\|_1 := \text{Tr}\sqrt{O^\dagger O} \) being the trace-norm of the operator \( O \) [Nielsen and Chuang, 2000; Wilde, 2013]. While fulfilling all the conditions of a regular distance (i.e., positivity, symmetry and triangular inequality) the trace distance possesses other interesting properties which makes it operationally well defined. For instance it is bounded between 0 and 1 (reaching the latter value only when \( \rho_1 \) and \( \rho_2 \) have orthogonal support). Furthermore the trace distance is preserved under unitary transformations, i.e., \( D(U\rho_1 U^\dagger, U\rho_2 U^\dagger) = D(\rho_1, \rho_2) \) (implying that the distance between physical states does not depend upon the coordinate system used to describe them) but it is contractive under CPTP maps \( \Phi \), i.e.,

\[
D(\Phi(\rho_1), \Phi(\rho_2)) \leq D(\rho_1, \rho_2) ,
\]

(implying that the action of noise tends to blur the difference among states). Finally, \( D(\rho_1, \rho_2) \) can be identified with the maximum distance between the statistical distributions \( \{p_x(\rho_1) = \text{Tr}\{E_x \rho_1\}\}_{x \in X} \) and \( \{p_x(\rho_2) = \text{Tr}\{E_x \rho_2\}\}_{x \in X} \) obtained by performing the same POVM measurement \( \{E_x\}_{x \in X} \) on \( \rho_1 \) and \( \rho_2 \).

Another quantity which is useful to gauge how close two density matrices \( \rho_1 \) and \( \rho_2 \) are, is the fidelity defined as [Jozsa, 1994; Uhlmann, 1976]

\[
F(\rho_1, \rho_2) := \|\rho_1^{1/2} \rho_2^{1/2}\|_1^2 ,
\]

which for \( \rho_1 \) being rank one, i.e., \( \rho_1 = |\psi_1\rangle\langle \psi_1| \), coincides with the probability of finding \( \rho_2 \) in the vector \( |\psi_1| \), i.e.,

\[
F(|\psi_1|, \rho_2) = \langle \psi_1 | \rho_2 | \psi_1 \rangle .
\]

The function \( F(\rho_1, \rho_2) \) is symmetric (i.e., \( F(\rho_1, \rho_2) = F(\rho_2, \rho_1) \)) and always in the range \([0,1]\) (equal to 1 if and only if \( \rho_1 = \rho_2 \) and vanishing for density operators with orthogonal supports, e.g., for orthogonal pure states). Furthermore, \( F \) is invariant under the action of a unitary evolution, \( F(U\rho_1 U^\dagger, U\rho_2 U^\dagger) = F(\rho_1, \rho_2) \), and increasing under CPTP map,

\[
F(\Phi(\rho_1), \Phi(\rho_2)) \geq F(\rho_1, \rho_2) .
\]

While not a distance itself, the fidelity is directly linked to the Bures distance (Bures, 1969) via the identity \( D_B(\rho_1, \rho_2) := [2 - 2F(\rho_1, \rho_2)]^{1/2} \). Trace distance and fidelity are related by the following inequalities

\[
1 - \sqrt{F(\rho_1, \rho_2)} \leq D(\rho_1, \rho_2) \leq \sqrt{1 - F(\rho_1, \rho_2)} ,
\]

therefore states which have high values of fidelity are also close in trace distance, and vice-versa.
Appendix B: Quasi-Local Algebras

Quasi-local algebras are the proper mathematical tools to describe infinitely extended quantum lattice systems (Bratteli and Robinson 1979). For the sake of simplicity let us consider a chain of infinitely many qubits (spin) placed in a one dimensional lattice $\mathbb{Z}$. Then, to each lattice site $j \in \mathbb{Z}$ attach a Hilbert space $\mathcal{H}_j \simeq \mathbb{C}^2$ and consider the associated algebra of bounded operators $\mathcal{A}_j = B(\mathcal{H}_j)$. If one restricts to a finite part of the chain, say a set $\Lambda \subset \mathbb{Z}$, it is possible to define the following tensor products

$$\mathcal{H}_\Lambda = \otimes_{j \in \Lambda} \mathcal{H}_j, \quad \mathcal{A}_\Lambda = B(\mathcal{H}_\Lambda) = \otimes_{j \in \Lambda} \mathcal{A}_j. \quad (B1)$$

Operators $a \in \mathcal{A}_\Lambda$ are called local operators as they are operators ‘localized’ in $\Lambda$. Clear examples of local observables are the Pauli operators $\sigma_x, \sigma_y, \sigma_z$, attached to the site $j \in \mathbb{Z}$. However, if the region $\Lambda$ is infinite the situation becomes tricky. In fact, although it is still possible to attach an Hilbert space $\mathcal{H}$ to the whole chain such that local operators $\sigma_j$ act on $\mathcal{H}$, this cannot be done in a unique way. It is then preferable to proceed by considering the algebraic properties of local operators not requiring a global Hilbert space $\mathcal{H}$. To this end, let us first notice that having two finite regions $\Lambda, \Lambda'$ such that $\Lambda \subset \Lambda' \subset \mathbb{Z}$, the operators

$$a \in \mathcal{A}_\Lambda \quad \text{and} \quad a \otimes \mathbbm{1}_{\Lambda' \setminus \Lambda} \in \mathcal{A}_{\Lambda'}, \quad (B2)$$

describe the same physical object. Therefore, one can identify $\mathcal{A}_\Lambda$ with the sub-algebra $\mathcal{A}_\Lambda \otimes \mathbbm{1}_{\Lambda' \setminus \Lambda}$ of $\mathcal{A}_{\Lambda'}$ through the map

$$\mathcal{A}_\Lambda \ni a \mapsto a \otimes \mathbbm{1}_{\Lambda' \setminus \Lambda} \in \mathcal{A}_{\Lambda'}. \quad (B3)$$

Doing so one gets a system of matrix algebras $\mathcal{A}_\Lambda$ which are ordered by inclusion

$$\Lambda \subset \Lambda' \Rightarrow \mathcal{A}_\Lambda \subset \mathcal{A}_{\Lambda'}, \quad \forall \text{ finite } \Lambda, \Lambda' \subset \mathbb{Z}. \quad (B4)$$

This construction leads to the possibility of defining, for two arbitrary local operators $a_1, a_2$,

i) their linear combination $\mu_1 a_1 + \mu_2 a_2$ with $\mu_1, \mu_2 \in \mathbb{C}$;

ii) their product $a_1 a_2$;

iii) their adjoints $a_1^\dagger$ (resp. $a_2^\dagger$).

To this end one only needs to find a region $\Lambda$ such that the matrix algebra $\mathcal{A}_\Lambda$ contains both operators $a_1, a_2$.

More precisely one can introduce the space of all local operators by the union

$$\mathcal{A}^{loc} := \bigcup_{\text{finite } \Lambda \subset \mathbb{Z}} \mathcal{A}_\Lambda, \quad (B5)$$

and then equip this space with a vector space structure, a product (associative and bilinear), a $1$-operation and a unit element $\mathbbm{1}$. In such a way $\mathcal{A}^{loc}$ becomes the algebra of local observables.

Going further on, one can associate to each $a \in \mathcal{A}^{loc}$ its norm $\|a\|$ by finding the region $\Lambda$ such that $a \in \mathcal{A}_\Lambda$ and then using the standard operator norm. A problem is given by the fact that $\mathcal{A}^{loc}$ is not complete with respect to such a norm, i.e., not all Cauchy sequences converge. This problem can be overcome by taking the norm-closure of $\mathcal{A}^{loc}$ and get the algebra of quasi-local observables

$$\mathcal{A}^Z = \overline{\mathcal{A}^{loc}}^{\|\|}. \quad (B6)$$

Here quasi-local stands for the fact that $\mathcal{A}^Z$ besides all local observables, also contains non-local observables which can be approximated in norm by local ones. $\mathcal{A}^Z$ results a $C^*$-algebra and its elements can be regarded in many respects as bounded operators (Bratteli and Robinson 1979).

It is also useful to consider methods to transform abstract elements $a \in \mathcal{A}^Z$ into operators $\pi(a)$ acting on a Hilbert space $\mathcal{H}$, i.e., representations of quasi-local algebra. A representation $\pi$ on Hilbert space $\mathcal{H}$ is a homomorphism $\pi : \mathcal{A}^Z \to B(\mathcal{H})$, i.e., a linear map satisfying $\pi(ab) = \pi(a)\pi(b)$ and $\pi(a^\dagger) = \pi(a)^\dagger$. Unfortunately for spin chains there is not a unique representation that can be used for all purposes, but one has to choose the representation which is most appropriate to the given physical context. This ambiguity also reflects on the definition of states. In fact one might be tempted to use density operators $\rho$ on the Hilbert space $\mathcal{H}$ which carry the representation $\pi$. However since different representations correspond to different physical contexts one should use all possible representations (in fact each density operator in any representation can describe a state). Clearly, it would be much better to describe states in a way independent from the representation. Thus a state of $\mathcal{A}^Z$ is defined as a linear functional $\psi : \mathcal{A}^Z \to \mathbb{C}$ which is positive ($\psi(a^\dagger a) \geq 0, \forall a \in \mathcal{A}$) and normalized ($\psi(\mathbbm{1}) = 1$). This means that given a representation $\pi$ and a density operator $\rho$ on $\mathcal{H}$, the corresponding state is the functional $\psi_\rho(a) = tr(\pi(a)\rho).$

The possibility to find for each state $\psi$ a Hilbert space $\mathcal{H}$ carrying a representation $\pi$ and a density operator $\rho$ such that $\psi = \psi_\rho$ is guaranteed by the so called Gelfand-Naimark-Segal theorem (Bratteli and Robinson 1979). It states that each state $\psi$ can be represented by a state vector $|\psi\rangle$ on a suitable Hilbert space. In other words, it is like to say that it is always possible to provide a ‘purification’ of the state $\psi$.

One can introduce into the quasi-local algebra $\mathcal{A}^Z$ a shift operation $T : \mathcal{A}^Z \to \mathcal{A}^Z$ by the following action:

$$\mathcal{A}^Z \ni a \simeq a \otimes \mathbbm{1}_{\mathcal{A}} \mapsto T(a) := \mathbbm{1}_{\mathcal{A}} \otimes a \simeq a \in \mathcal{A}^{A+1}, \quad (B7)$$

where $A \otimes \mathbbm{1}_{\mathcal{A}}$ (resp. $\mathbbm{1}_{\mathcal{A}} \otimes A$) stands for the tensor product between $A$ belonging to $\mathcal{A}^Z$ and identity of $\mathcal{A}$ on the site to the right of $\Lambda$ (resp. between identity of $\mathcal{A}$ on the site to the left of $\Lambda$ and $a$ belonging to $\mathcal{A}^\Lambda$).
Moving from the action of the shift $T$, it is possible to introduce the notion of stationary state $\psi$ on $A^\mathbb{Z}$ when $\psi \circ T = \psi$ holds true. The set of stationary states on $A^\mathbb{Z}$ turns out to be convex. Then a state $\psi$ on $A^\mathbb{Z}$ is called ergodic (with respect to the shift) if it is extremal on this set.

Appendix C: Decomposition for non-anticipatory quantum channels

This section provides an explicit derivation of the decomposition of non-anticipatory channels in Eq. (80) based on a generalization of analysis presented in [Beckman et al., 2001; Eggeling, Schlingemann and Werner, 2001; Kretschmann and Werner, 2005; Piani et al., 2006].

Let then $\{\Phi^{(n)}; n = 1, 2, \ldots \}$ be a family of CPTP maps which describes a non-antipatory quantum channel. Adopting the unitary representation in Eq. (7), for each $n$ one can define a unitary transformation $W_{q_0, q_{n-1}, \ldots, q_1, M}^{(n)}$ coupling the first $n$ carriers to a common environment $M$ which allows one to write

$$
\Phi^{(n)}(\rho_Q^{(n)}) = \text{Tr}_M[W^{(n)}_{q_0, q_{n-1}, \ldots, q_1, M}]_{q_0, q_{n-1}, \ldots, q_1, M} \cdot (\rho_Q^{(n)} \otimes \omega_M^{(n)}) W^{(n)}_{q_0, q_{n-1}, \ldots, q_1, M} \cdot \omega_M^{(n)}.
$$

(C1)

The environmental state $\omega_M^{(n)}$ is in general a function of $n$ but is independent on the input state $\rho_Q^{(n)}$. Without loss of generality here it is assumed to be a pure state, $\omega_M^{(n)} = |\omega^{(n)}\rangle_M \langle \omega^{(n)}|$. In general the unitary couplings $W^{(n)}_{q_0, q_{n-1}, \ldots, q_1, M}$ can have a complicated dependence upon $n$; however since the channel is non-anticipatory they must obey the following rule

$$
\text{Tr}_{M, q_{n-1}, \ldots, q_1, M} [W^{(n)}_{q_0, q_{n-1}, \ldots, q_1, M} \rho_Q^{(n)} \otimes \omega_M^{(n)} W^{(n)\dagger}_{q_0, q_{n-1}, \ldots, q_1, M}] = 
\text{Tr}_M [W^{(n-1)}_{q_{n-1}, \ldots, q_1, M} \rho_Q^{(n-1)} \otimes \omega_M^{(n-1)} W^{(n-1)\dagger}_{q_{n-1}, \ldots, q_1, M}],
$$

where $\rho_{Q_{(n-1)}} = \text{Tr}_{q_{n-1}, \ldots, q_1, M} [\rho_Q^{(n)}]$ is the reduced density operator of $\rho_Q^{(n)}$ associated with the first $n-1$ carriers.

Applying this relation to a pure input state $\rho_Q^{(n)}$ of the form $|\psi\rangle_{q_0} \otimes |\phi\rangle_{q_{n-1}, \ldots, q_1, M}$ one notices that the vectors $W^{(n)}_{q_0, q_{n-1}, \ldots, q_1, M} |\psi\rangle_{q_0} \otimes |\phi\rangle_{q_{n-1}, \ldots, q_1, M} \otimes |\omega^{(n)}\rangle_M$ and $W^{(n-1)}_{q_{n-1}, \ldots, q_1, M} |\phi\rangle_{q_{n-1}, \ldots, q_1} \otimes |\omega^{(n-1)}\rangle_M$ are both purifications of the state $\Phi^{(n-1)}(\rho_Q^{(n-1)})$. Therefore there must exist a unitary transformation $U_{q_{n-1}, M}$ acting on the latter which satisfies the identity [Nielsen and Chuang, 2000]

$$
W^{(n)}_{q_0, q_{n-1}, \ldots, q_1, M} |\omega^{(n)}\rangle_M = U_{q_0, M} W^{(n-1)}_{q_{n-1}, \ldots, q_1, M} |\omega^{(n-1)}\rangle_M.
$$

(C2)

Iterating this $n$ times yields

$$
W^{(n)}_{q_0, q_{n-1}, \ldots, q_1, M} |\omega^{(0)}\rangle_M = U_{q_0, M} U_{q_{n-1}, M} \cdots U_{q_1, M} |\omega^{(0)}\rangle_M.
$$

(C3)

which replaced into Eq. (C1) implies Eq. (80).

Appendix D: Explicit derivation of capacity upper bounds

Here we present an explicit derivation of the upper bounds [158] and [159] for the classical and quantum capacities defined in Eq. (144) of a (non necessarily memoryless) quantum channel $\Phi^{(n)}$.

1. Upper bound for the classical capacity

The derivation of Eq. (158) follows by merging the Holevo Bound [Holevo, 1973 a,b] with the classical Fano inequality [Cover and Thomas, 1991]. For this purpose one reminds that given two random variables $X$ and $Y$ connected by conditional probability distribution $p(y|x)$ and a correspondence rule which assigns values of $Y$ to each of the values of $X$, the Fano inequality allows one to lower bound the mutual information $I(X : Y)$ as

$$
I(X : Y) \geq H(X) - h(p_e) - P_e \log_2(|X| - 1), \quad \text{(D1)}
$$

where $|X|$ is the number of elements of the variable $X$, $h(p)$ is the binary Shannon entropy (66), and finally $P_e$ is the average error probability that the correspondence rule is violated by the conditional probability $p(y|x)$. Specifically, assuming for simplicity that the $X$ and $Y$ span the same alphabet of symbols and that correspondence rule assign to $Y$ the same symbol on $X$, we have $P_e = 1 - \sum_x p(x) p(y = x|x)$. Identify then $X$ with the messages $m$ that Alice is mapping from $M$ to $N$ carriers via the coding channel $\Phi^{(k\rightarrow n)}_E$, and with and $Y$ the elements of $M$ which Bob is retrieving via his decoding mapping $\Phi^{(n\rightarrow k)}_D$. Under the assumption that the probability that Alice is selecting the messages from $M$ with uniform probability, and reminding that $M$ contains $2^k$ elements, Eq. (D1) yields

$$
I(X : Y) \geq k - h(p_e) - P_e \log_2(2^k - 1)
\geq k - h(e) - ek,
$$

(D2)

where we used the fact that $P_e \leq 1 - F_{\text{min}} < \epsilon$, with $F_{\text{min}}$ being the minimum fidelity achieved by the selected code. By reorganizing the various terms and by dividing by $n$ we then get

$$
\frac{k}{n} \leq \frac{I(X : Y)}{(1 - \epsilon)n} + \frac{h(e)}{(1 - \epsilon)n}.
$$

(D3)

Remind next that $I(X : Y)$ is the mutual information associated to the ensemble of codewords $\mathcal{E} = \{p_m = 2^{-k}; \Phi^{(k\rightarrow n)}_E(m)\}$ generated by Alice and received by Bob through the channel $\Phi^{(n)}$. The Holevo Bound (72) implies then

$$
\frac{k}{n} \leq \frac{\chi(\mathcal{E}; \Phi^{(n)})}{(1 - \epsilon)n} + \frac{h(e)}{(1 - \epsilon)n}
\leq \frac{\max_{\mathcal{E}} \chi(\mathcal{E}; \Phi^{(n)})}{(1 - \epsilon)n} + \frac{h(e)}{(1 - \epsilon)n}.
$$

(D4)
Since this inequality holds for all encoding/decoding strategies entering in the capacity definition \( (144) \), by taking the limit on \( k \to \infty \) and \( \epsilon \to 0 \) we finally get the inequality \( (158) \).

The same derivation detailed above can be used to prove the converse part of the HSW theorem for the one-shot classical capacity \( C_1(\Phi) \) of a memoryless channel \( \Phi \), i.e.,

\[
C_1(\Phi) \leq \max_{\varepsilon} \chi(\mathcal{E}; \Phi) . \tag{D5}
\]

Indeed, exploiting the fact that the channel is memoryless and the coding procedure uses only separable codewords \( \Phi^{(k-n)}(m) \) one can apply the sub-additivity of the classical mutual information \( \text{Cover and Thomas} \ 1991 \) Gallager \ 1968 \) to replace Eq. \( (122) \) as

\[
\sum_{i=1}^{n} I(X_i ; Y_i) \geq k - h(P_e) - P_e \log_2(2^k - 1) \\
\geq k - h(\epsilon) - ek , \tag{D6}
\]

where, for \( i = 1, \ldots, n, I(X_i ; Y_i) \) is the mutual information associated with the classical input of the \( i \)-th channel use. Applying the Holevo Bound to all these terms we get

\[
\frac{k}{n} \leq \max_{\varepsilon} \chi(\mathcal{E}; \Phi) + \frac{h(\epsilon)}{(1 - \epsilon)n} , \tag{D7}
\]

that finally leads to Eq. \( (D5) \) when taking \( \epsilon \to 0 \).

In conclusion, it is worth remarking that Eq. \( (158) \) can be used to prove an inequality which in some cases happens to be useful for deriving explicit expression for \( C \) – see e.g., Sec. VI.A.1 This is obtained by noticing that

\[
\max_{\varepsilon} \chi(\mathcal{E}; \Phi^{(n)}) \leq \max_{\rho} S(\Phi^{(n)}(\rho)) - \min_{\rho} S(\Phi^{(n)}(\rho)) \\
\leq n \log_2 d - S_{\min}(\Phi^{(n)}) . \tag{D8}
\]

with \( S_{\min}(\Phi^{(n)}) = \min_{\rho} S(\Phi^{(n)}(\rho)) \) being the minimum entropy one can reach at the output of the channel. In the above derivation the first inequality follows directly from the definition of the Holevo information, while the second from the fact that the entropy of a state of \( \mathcal{H}_Q^{(n)} \) is not larger than \( \log_2 d^n \) (\( d \) being the dimension of \( \mathcal{H}_Q \)). Replacing this into Eq. \( (158) \) we then get

\[
C \leq \log_2 d - \lim_{n \to \infty} \frac{S_{\min}(\Phi^{(n)})}{n} . \tag{D9}
\]

2. Upper bound for the quantum capacity

The derivation which follows is an adaptation to the memory channel scenario of the proof of \( (\text{Barunum} \ 1998) \) Nielsen, and Schumacher \ 1998 \). The starting point in this case is the the quantum Fano inequality presented in Eq. \( (65) \), the data-processing inequality \( (70) \) and the bounds \( (63) \). From them one can easily verify that given a generic density matrix \( \tau \) of the reference set \( \mathcal{M} \) the following relation holds

\[
S(\tau) \\
\leq \frac{J(\tau; \Phi_D^{(n-k)} \circ \Phi_E^{(k-n)})}{n} + \frac{2}{n} h(F_\epsilon(\tau; \Phi_D^{(n-k)} \circ \Phi_E^{(k-n)})) \\
+ \frac{4}{n} \log_2(4^k - 1) \leq \frac{\max_{\rho} J(\rho; \Phi^{(n)})}{n} + \frac{2}{n} h(1 - 3\epsilon/2) + \frac{6}{n} \epsilon k , \tag{D10}
\]

where the second inequality has been obtained by using again the data-processing inequality \( (70) \) and maximizing \( J \) over all possible inputs states \( \rho \) of the \( n \) channels uses, and by using the fact that the minimum fidelity of the code is lower bounded by \( 1 - \epsilon \) and the fact that this implies that the corresponding entanglement fidelity fulfills \( F_\epsilon(\tau; \Phi_D^{(n-k)} \circ \Phi_E^{(k-n)})) > 1 - 3\epsilon/2 \), – see Eq. \( (63) \). Specifying Eq. \( (D10) \) for the totally mixed state state of \( \mathcal{M} \), we then get

\[
\frac{k}{n} \leq \frac{\max_{\rho} J(\rho; \Phi^{(n)})}{n} + \frac{2}{n} h(1 - 3\epsilon/2) + \frac{6}{n} \epsilon k ,
\]

which in the limit of \( k \to \infty \) and \( \epsilon \to 0 \) yields finally \( (159) \).

References

Abeyesinghe, A., and Hayden, P., 2003, Phys. Rev. A 68, 062319.
Abeyesinghe, A., Devetak, I., Hayden, P., and Winter, A., 2009, Proceedings of the Royal Society A 465, 2537.
Adami, C., and Cerf, N. J., 1997, Phys. Rev. A 56, 3470.
Alicki, R., and Lendi, K., 1987, \textit{Semigroups and Applications}, Lecture Notes in Physics, Springer-Verlag, Berlin.
Ahlswede, R., 1968, Z. Wahrscheinlichkeitstheor. Verwandte Geb. 11, 61.
Alicki, R., and Lendi, K., 1987, \textit{Quantum Dynamical Semigroups and Applications}, Lecture Notes in Physics (Springer-Verlag, Berlin).
Amosov, G. G., 2007, J. Math. Phys. 48, 2104.
Amosov, G. G., and Mancini, S., 2009, Quant. Inf. & Comp. 7, 594.
Amosov, G. G., and Holevo, A. S., 2000, Probl. Inform. Transm. 36, 25.
Arikat, E., 2009, IEEE Trans. Inf. Theory 55, 3051.
Arimoto, S., 1973, IEEE Trans. Inf. Theory 19, 357.
Arshed, N., Toor, A. H., and Lidar, D. A., 2010, Phys. Rev. A 81, 062353.
Ashikhmin, A., and Knill, E., 2001, IEEE Trans. Inf. Theory 47, 3065.
Aubrun, G., Szarek, S., and Werner, E., 2010, Comm. Math. Phys. 305, 85.
Ball, J., Dragan, A., and Banaszek, K., 2004, Phys. Rev. A 69, 042324.
Ball, J., and Banaszek, K., 2005, Open Sys. & Inf. Dynamics 12, 121.
Quantum information processing with Clifford quantum cellular automata, PhD Thesis, Leibniz University, Hannover, http://www.pik-potsdam.de/members/johannes/dissertation-jg.pdf.

Gütschow, J., Uphoff, S., Werner, R. F., and Zimborás, Z., 2009, Comm. Math. Phys. 289, 76.

Hayashi, M., 2007, Phys. Rev. A 5, 1753.

Hayden, P., and Winter, A., 2008, Comm. Math. Phys. 284, 263-280.

Hayden, P., Horodecki, M., Winter, A., and Yard, J., 2008, Open System & Information Dynamics 15, 7.

Hausladen, P., Schumacher, B., Westmoreland, M., and Wootters, W. K., 1995, Annals of the New York Academy of Sciences 755, 698.

Hausladen, P., Jozsa, R., Schumacher, B., Westmoreland, M., and Wootters, W. K., 1996, Phys. Rev. A 54, 1869.

Helstrom, C. W., 1976, Quantum Detection and Estimation Theory, (London, Academic Press).

Hill, R., 1985, A First Course in Coding Theory, (Clarendon Press, Oxford).

Hiroshima, T., 2006, Phys. Rev. A 73, 022330.

Holevo, A. S., 1997, Probabilistic and Statistical Aspects of Quantum Theory, (Clarendon Press).

Holevo, A. S., 2009, arXiv:quant-ph/0904.2557.

Grace, M. D., Dominy, J., Kosut, R. L., Brif, C., and Rabitz, H., 2010, New J. Phys. 12, 015001.

Grassl, M., Beth, T., and Pellizzari, T., 1997, Phys. Rev. A 56, 33-38.

Grassl, M., and Roetteler, M., 2006, IEEE International Symposium on Information Theory, 1109-1113.

Gray, R. M., 2006, IEEE Trans. Inf. Theory 52, 192.

Gray, R. M., 1972, IEEE Trans. Inf. Theory 18, 725.

Grassl, M., Beth, T., and Pellizzari, T., 1997, Phys. Rev. A 56, 33-38.

Gütschow, J., Uphoff, S., Werner, R. F., and Zimborás, Z., 2009, Comm. Math. Phys. 289, 76.

Hayashi, M., 2007, Phys. Rev. A 5, 1753.

Hastings, M. B., 2009, Nat. Phys. 43, 1087.

Hamada, M., 2002, J. Math. Phys. 43, 4382.

Hastings, M. B., 2009, Nat. Phys. 5, 255.

Hayashi, M., 2007, Phys. Rev. A 76, 062301.

Hayashi, M., 2008, Eprint arXiv:0805.4092.

Hayashi, M., 2009, Comm. Math. Phys. 289, 1087.

Hayashi, M., and Nagaoka, H., 2003, IEEE Trans. Inf. Theory 49, 1753.

Jacobs, K., 1962, Matrix Analysis, (Cambridge University Press).

Holevo, A. S., Sohma, M., and Hirota, O., 1999, Phys. Rev. A 59, 1820.

Holevo, A. S., and Werner, R. F., 2001, Phys. Rev. A 63, 032312.

Holevo, A. S., and Giovannetti, V., 2012, Rep. Prog. Phys. 85, 126501.

Holevo, A. S., 1973, Probl. Inf. Transm. 9, 63.

Holevo, A. S., 1974, Probl. Inf. Transm. 10, 1753.

Holevo, A. S., 1975, Probl. Peredachi Inf. 11, 38.

Holevo, A. S., 1976, The Theory of Operator Algebras Vol. II, (American Mathematical Society).

Holevo, A. S., and Giovannetti, V., 2012, Rep. Prog. Phys. 85, 126501.

Holevo, A. S., and Werner, R. F., 2001, Phys. Rev. A 63, 032312.

Holevo, A. S., and Giovannetti, V., 2012, Rep. Prog. Phys. 85, 126501.

Holevo, A. S., 1982, Stabilizer Codes and Quantum Error Correction, PhD Thesis California, Institute of Technology, Pasadena Ca.

Holevo, A. S., 1997, Quantum Systems, Channels, Information, (De Gruyter).

Holevo, A. S., Shom, M., and Hirot, O., 1999, Phys. Rev. A 59, 1820.

Holevo, A. S., and Werner, R. F., 2001, Phys. Rev. A 63, 032312.
Wilde, M. M., and Brun, T. A., 2010, Quantum Inf. Processing 9, 509.
Wilde, M. M., Houshmand, M., and Hosseini-Khayat, S., 2011, Proceedings of the International Symposium on Information Theory, pp. 450-454.
Wilde, M. M., and Guha, S. 2011, arXiv:1109.2591.
Winter, A., and Yang, D., 2013, arXiv:1306.1586.
Winter, A., 1999, PhD Thesis, Universität Bielefeld.
Winter, A., 1999, IEEE Trans. Info. Theory 45, 2481.
Wolf, M. M., and Pérez-Garcia, D., 2007, Phys. Rev. A 75, 012303.
Wolf, M. M., Ortiz, G., Verstraete, F., and Cirac, J. I., 2006, Phys. Rev. Lett. 97, 110403.
Wolfowitz, J., 1964, Coding Theorem and Information Theory, (Springer, New York).
Yin, J., 2013, arXiv:1306.0672.
Zanardi, P., and Rasetti, M., 1997, Phys. Rev. Lett. 79, 3306.
Ziman, M., and Bužek, V., 2005, Phys. Rev. A 72, 022110.
Ziman, M., Štehachovič, P., Bužek, V., Scarani, V., and Gisin, N., 2002, Phys. Rev. A 65, 042105.
Zwanzig, R., 1960, J. Chem. Phys. 33, 1338.