Preprocessing operations and the reverse compression

Matheus Capela\textsuperscript{1}\textsuperscript{*} and Fabio Costa\textsuperscript{2}\textsuperscript{†}

\textsuperscript{1}Institute of Physics, Federal University of Goi\’\'as, POBOX 131, 74001-970, Goi\’\'nia, Brazil
\textsuperscript{2}Centre for Engineered Quantum Systems, School of Mathematics and Physics, The University of Queensland, St Lucia, QLD 4072, Australia

(Dated: January 8, 2021)

The task of compression of data – as stated by the source coding theorem – is one of the cornerstones of information theory. Data compression usually exploits statistical redundancies in the data according to its prior distribution. Motivated by situations where one does not have access to the statistics of data, but has some information about a transformation that is going to be applied to it, we propose a novel method for data compression called reverse compression. It is defined in such a way that works for both classical and quantum information processes, and furthermore relies exclusively on the channel to be used: all input data leading to indistinguishable outputs is compressed to the same state, regardless of their prior distribution. Moreover, this process can be characterized as a higher order operation within the type of preprocessing. We also consider as an example the application of the method to the classical and quantum erasure channel. The examples suggest that noiseless reverse compression can take place only in trivial cases, although meaningful instances of noisy reverse compression can exist.

I. INTRODUCTION

Information theory is ubiquitous feature of modern science. Its cornerstone is Shannon’s mathematical theory of communication\textsuperscript{[1]}, dealing with two main problems: channel coding and source coding. Channel capacity, defined in the former as the amount of information that can be reliably sent through a channel, is shown to be equal to the mutual information between the input and output random variables maximized over all possible input probability distributions. Thus, it relies only on the channel features. In the latter, the concept of Shannon entropy characterizes the limit of how much a random source can be reliably compressed. Since then, many links with other branches of science have been achieved. For instance, it allowed a deep connection with statistical mechanics through the Jaynes’ maximum entropy principle\textsuperscript{[2]}, setting the equilibrium probability distribution associated with a thermodynamical (statistical) system as the one maximizing Shannon entropy respecting the relevant physical constraints.

We address here the problem of compression of a source without relying on its statistical features, but only on the knowledge of properties of the channel in which the messages are to be send through. The practical situation of interest can be thought, for instance, as the task of transmitting a picture through some known processing algorithm or device, without any prior knowledge of the statistical distribution of the pixels composing the picture. For example, the device could be some analysis tool, returning a set of properties of the picture. By knowing in advance that the device has little sensitivity to some aspects of the picture (e.g., the output might not depend much on certain differences in color), we can save communication bandwidth by compressing the image before sending it to the device. As we want the protocol to be independent of the input distribution, Shannon’s compression method does not help towards this task.

We call such a procedure reverse compression of a source with respect to a channel. We formalize this preprocessing operation for both classical and quantum information processes. We also show examples where compression is possible given some finite noise. However, it appears that no noiseless compression can be achieved, except for trivial examples.

The reverse compression protocol introduced here can be understood as a particular case of superchannel operation, which is a problem dealt with by Shannon in the early days of Information Theory. His interest was in the pre-ordering of classical channels. He was able to show that if a channel can simulate another one by preprocessing and postprocessing with shared randomness, and there is a code for the latter, then there is a code at least as efficient for the former\textsuperscript{[3]}. Much development has been achieved in this problem relating the characterization of specific classes of channels\textsuperscript{[4]–[10]}. Nevertheless, it remains as an open area of research. An area of statistical analysis which is a particular case of the one considered by Shannon is the Theory of Comparison\textsuperscript{[5]}. It consists of the simulation of channels with the same input space by postprocessing only. Recent developments and extension to the Quantum Information Theory have been achieved in\textsuperscript{[10]–[14]}. On the other hand, it is also possible to define relevant infor-
mation processing tasks involving exclusively prepro-
cessing, for instance, the reverse compression presented here. The authors in [13] have considered the prepro-
cessing of Positive Operator Value Measures (POVM).

Recent interest has been shown in the quantum the-
ory of superchannels. For instance, Ref. [14] has con-
sidered the problem of quantum channel simulation by superchannels with preprocessing and postprocessing with quantum side information, and the extension of the notion of entropy for channels. With a more physical motivation, the authors in [15] have consid-
ered the problem of quantum channel simulation by so called thermal operations, and derived that in the asymptotic limit the thermodynamical cost of the simu-
lation is given by a unique measure: the difference in thermodynamical capacity of the correspondent chan-
nels. Furthermore, superchannels extend natura-

ally to more general higher order transformations, which are central in the study of quantum causal structures [14- 23].

The paper is organized as follows. In section II we de-

fine the notation to be used along the text. In section III we discuss the Shannon compression method in order to make clear the distinction with the prepro-
cessing task considered here. In section IV we de-

fine the reverse compression of data for memoryless class-

cial channels used without feedback. The exposition starts from the case of single-shot uses of the channel, leading to the asymptotic setting, with arbitrarily many uses. In section V we consider the definition of reverse compression of data for quantum channels. As a general rule we consider as an application the case of the erasure channel. Finally, in section VI we make our final remarks and discussions.

II. NOTATION AND PRELIMINARIES

We denote random variables by English upper case letters, their outcomes by lower case letters, and alphabets – the collection of all possible outcomes, also called the values of a random variable – by calligraphic letters. For instance, \( X \) represents a random variable with out-

comes denoted by \( x \), while \( \mathcal{X} \) represents its alphabet. In what follows, we only consider alphabets with a finite number of outcomes denoted as \(|\mathcal{X}|\). More generally, the symbol \(|\cdot|\) represents the cardinality of a set.

A probability distribution of the random variable \( X \), denoted as \( P(X) \), is an assignment of real numbers to each outcome in its alphabet, written as \( P(x) \) and called probability masses. The probability masses satis-

fy \( 0 \leq P(x) \leq 1 \) for any outcome, and \( \sum_{x \in \mathcal{X}} P(x) = 1 \).

The expected value of a real valued random variable \( R \) is defined as \( E(R) := \sum_{r \in \mathcal{R}} P(r) r \) and its variance as \( \text{var}(R) := E([R - E(R)]^2) \). A sequence of \( k \) ran-

dom variables \( X^k := (X_1, \ldots, X_k) \) taking values in the same alphabet \( \mathcal{X} \) is said to be independent and identically distributed (i.i.d.) if it is drawn according to a probability of the form \( P(X^k) = \prod_{i=1}^k P(X_i) \), where \( P(X_i) = P(X_j) \) for any \( 1 \leq i, j \leq k \). We define the sam-

ple average of a sequence of real-valued random vari-

ables as \( R^k := \frac{1}{k} \sum_{i=1}^k R_i \). Note the sample average is de-


fined here as a random variable assigning the arithmetic average to each outcome \( r^k \). Additionally, we say the se-

quence of real-valued joint random variables \((R^k)_{k \in \mathbb{N}}\), each drawn according to \( P(R^k) \), converges in probability to a real number \( \zeta \) if \( \Pr \left( |R^k - \zeta| \leq \delta \right) \rightarrow 1 \) as \( k \rightarrow \infty \) for arbitrary positive real number \( \delta \), and write \( R^k \xrightarrow{P} \zeta \) in such case.

The fidelty of two probability distributions \( P \) and \( Q \) of the same random variable \( X \) is defined as

\[
F[P(X), Q(X)] = \left( \sum_x \sqrt{P(x)Q(x)} \right)^2.
\]

It is a quantity regarded as a measure of distin-

guishability of probability distributions such that \( 0 \leq F(P, Q) \leq 1 \), and \( F(P, Q) = 1 \) if and only if \( P = Q \).

A communication system consists of a sender en-
coding messages emitted by a source to be transmit-
ted through a channel – the physical medium allowing communication between distant parties – and decoded by a receiver. The channel is described by a probability distribution of the output variable \( Y \) conditional to the input variable \( X \), and denoted by \( P(Y|X) \). Moreover, the channel can be understood as a process that trans-
forms input probability distributions \( P(X) \) into output ones \( P(Y) \) by the relation \( P(Y) = \sum_{x \in \mathcal{X}} P(Y|x)P(x) \).

The quantum analogue of the alphabet of a random variable is the space of pure quantum states. Therefore, to each random variable \( X \) with alphabet \( \mathcal{X} \) we assign a collection of unit orthogonal vectors \((|x\rangle)_{x \in \mathcal{X}}\), and de-

fine the associated Hilbert space by any linear composi-
tion of them denoted by \( \mathcal{H}_X := \text{Span} \{|x\rangle \}_{x \in \mathcal{X}} \cong \mathbb{C}^{|\mathcal{X}|}\). We denote the space of linear operators acting on a vec-

tor space \( \mathcal{H} \) by \( \mathcal{L}(\mathcal{H}) \), and the set of states of the cor-

responding quantum system – positive linear operators with unit trace – as \( \mathcal{D}(\mathcal{H}) \).

A linear map \( M^{X \rightarrow Y} : \mathcal{L}(\mathcal{H}_X) \rightarrow \mathcal{L}(\mathcal{H}_Y) \) is called co-

completely positive (CP) if its extension \( T^R \otimes M^{X \rightarrow Y} \) to any finite dimensional system \( R \) is positive, where \( \mathcal{T}^R : \mathcal{L}(\mathcal{H}_R) \rightarrow \mathcal{L}(\mathcal{H}_R) \) is the identity map. We say \( M^{X \rightarrow Y} \) is trace preserving (TP) if \( \text{Tr}_Y \circ M^{X \rightarrow Y} = \text{Tr}_X \) is a quantum channel is a linear completely positive and trace preserving (CPTP) map. We refer the reader to [24] for a more complete discussion on quantum channels.

The fidelity of the quantum states \( \rho \) and \( \sigma \) is de-

fined here as \( F(\rho, \sigma) := \text{Tr}(\sqrt{\sqrt{\rho} \sigma \sqrt{\rho}})^2 \), following
the original definition in [23]. If the states considered are diagonal in the same basis, assuming the form \( \rho = \sum_x P(x) |x\rangle \langle x| \) and \( \sigma = \sum_x Q(x) |x\rangle \langle x| \), then Equation (11) is recovered.

III. OVERVIEW ON SHANNON COMPRESSION

The compression of a source is defined by the task of minimizing redundancy of the messages to be sent through a channel relying on its probability distribution, and thus maximizing the efficiency of the communication protocol. This section is based on [28, 29].

Shannon’s compression theorem states that there is an encoding procedure that is asymptotically reliable, and the goal of this section is to make mathematically precise such terminology. Shannon’s compression method is based on the law of large numbers stating that a sequence of i.i.d real valued random variables \( R^n \) with finite variance converges in probability to its expected value, \( \overline{R^n} \rightarrow E(R) \). Taking \( R_k = \log P(X_k) \), the law of large numbers implies the convergence \( \log P(X_k) \rightarrow H(X) \), which means that for any positive real number \( \delta \) we have

\[
\Pr \left[ \left| \frac{1}{k} \log P(X_1, \ldots, X_k) - H(X) \right| \leq \delta \right] \rightarrow 1 \tag{2}
\]

as \( k \rightarrow \infty \).

It suggests we can take the set of likely sequences as consisting of the so called \( \delta \)-typical sequences defined by the property \( \left| \log P(X_k) - H(X) \right| \leq \delta \), where \( \delta \) can be made as any arbitrarily small positive real number. The set of all \( \delta \)-typical sequences of length \( k \), for which the probability can be made arbitrarily close to its expected value for sufficiently large \( k \), is denoted here by \( T^{(k)}_\delta \). Moreover, we can encode only the typical sequences and neglect the atypical ones with probability of decoding a sequence wrong arbitrarily close to zero, as it follows from Equation (2) that \( \Pr \left[ X^k \setminus T^{(k)}_\delta \right] \rightarrow 0 \) as \( k \rightarrow \infty \).

It can be shown that the number of typical sequences one needs to encode ranges as \( (1 - \delta)2^{[kH(X) - \delta]} \leq |T^{(k)}_\delta| \leq 2^{[kH(X) + \delta]} \), and for \( \delta \rightarrow 0 \) one needs to encode \( |T^{(k)}_0| = 2^{kH(X)} \) sequences with probability of error arbitrarily close to zero if \( k \) is large enough (\( k \rightarrow \infty \)). Note that \( H(X) \) may not be an integer and therefore \( |T^{(k)}_0| \) should be replaced in the equation above by the smallest integer greater than or equal to it, but we keep this notation for simplicity. It means that we can reliably compress a variable by encoding only \( 2^{kH(X)} \) sequences, rather than encoding all the \( |X|^k \). This is often called as the direct part of the Shannon compression theorem: the proof of existence of a compression-decompression scheme

\[
( f : X^k \rightarrow \{0, 1\}^{kH(X)}, g : \{0, 1\}^{kH(X)} \rightarrow X^k )
\]

with arbitrarily small probability of error defined by \( \Pr \left[ g \circ f(X^k) \neq X^k \right] \) for sufficiently large \( k \). The only case where there is no advantage in this encoding method is when all outcomes are uniformly distributed, and therefore all the sequences are typical. The converse part of the Shannon compression theorem states that for any compression-decompression scheme \( ( f : X^k \rightarrow \{0, 1\}^n, g : \{0, 1\}^n \rightarrow X^k ) \) with a compression rate \( \frac{n}{k} \) less than \( H(X) \) has an arbitrarily large probability of error.

IV. CLASSICAL REVERSE COMPRESSION

A. Single-Shot Settings

We start with a definition of preprocessing for single-shot scenarios, i.e., information processing protocols with a single use of the channel. For single-shot settings the definition of reverse compression takes a simpler and clear form, and furthermore its extension to asymptotic scenarios is an immediate application considered subsequently.

The question addressed here is the possibility of replacement of the input variable of a classical channel by a less redundant one, as it is the case for the Shannon compression method, but without relying on the knowledge of the input variable distribution \( P(X) \) nevertheless. Instead, the idea is to look for redundancies in how the input is mapped to the output: if two input variables produce very similar output distributions, then they can be replaced by a single one. Importantly, here one compresses the input based on information of a transformation that will take place in the future, hence the terminology reverse compression. Moreover, the main difference between the Shannon compression method and ours lies in the task of communication: the goal of Shannon compression is to communicate a message through a noiseless channel with arbitrarily small probability of error, while reverse compression is defined in the following by channel indistinguishability after preprocessing of data.

Two probability distributions \( P \) and \( Q \) are equal if and only if their fidelity is equal to unity, \( F(P, Q) = 1 \). Allowing for slightly distinguishable events, we say that probability distributions \( P \) and \( Q \) are \( \epsilon \)-indistinguishable if \( F(P, Q) \geq 1 - \epsilon \), and denote it as \( P \preceq Q \). The reverse compression of a source with respect to a channel consists of replacing the input ran-
random variable $X$ by a random variable $Z$ with smaller alphabet such that the output variable $Y$ conditional distribution remains approximately the same. This stage is called here compression. The appropriate mapping from $Z$ into $X$ (necessary to feed the variable into the desired channel) is called decompression. Figure 1 shows a diagram representing the process.

Consider a partition of the input variable’s alphabet according to the following criteria. We take sets $A_z$ of outcomes of $X$ such that, for some arbitrarily fixed $\epsilon > 0$, the conditional probabilities $P(Y|X=x)$ and $P(Y|X=x')$ are $\epsilon$-indistinguishable, i.e.,

$$F[P(Y|X=x), P(Y|X=x')] \geq 1 - \epsilon,$$  \hspace{1cm} (3)

for any $x$ and $x'$ in $A_z$.

A collection of non-empty sets $A_z$ is a partition of the input alphabet $X$ if they are pair-wise disjoint and they cover $X$, namely $\cup_z A_z = X$. The mapping $A_z \ni x \mapsto z$ defines a collective random variable $Z$, interpreted as the compressed input. The channel from $Z$ to $Y$ is given by

$$P(Y|Z=z) = \sum_{x \in A_z} P(Y|X=x)P(X=x|Z=z),$$  \hspace{1cm} (4)

where $P(X|Z=z)$ is any distribution with support in $A_z$, that is, $P(X=x|Z=z) = 0$ if $x \notin A_z$.

Given the definition above, we have, for each $x \in A_z$

$$F[P(Y|X=x), P(Y|Z=z)] \geq 1 - \epsilon.$$  \hspace{1cm} (5)

Note that as the collection $P_\epsilon = \{A_z\}_{z \in Z}$ is a partition of the input alphabet, the random variable $Z$ has a smaller (or equal) cardinality than $X$. It is worth to note that a partition $P = \{A_z\}_{z \in Z}$ of $X$ is equivalent to a surjective mapping $\lambda : X \mapsto Z$. Each map $\lambda$ of that type corresponds to a partition with elements defined as $A_z := \lambda^{-1}(z)$, where $\lambda^{-1}(z) := \{x \in X | \lambda(x) = z\}$ is the inverse image of the set $\{z\}$ under the mapping $\lambda$. Moreover, a partition is equivalent to an appropriate mapping on the input alphabet. The partition $P_\epsilon$ is called a reverse compression. The channel $P(X|Z)$ in Eq. 4 is a decompression. For definiteness, we can choose an arbitrary but fixed $z_\epsilon \in A_z$ for each subset and define the decompression as the mapping from $z$ to this $x_\epsilon$: $P(X=x|Z=z) = \delta_{x_\epsilon z}$. See Figure 2. A decompression with such a definition is clearly not unique. It follows the formal definition. Let $0 \leq \epsilon \leq 1$.

**Definition 1 (Single-Shot Reverse Compression).** An $\epsilon$-reverse compression of a channel $P(Y|X)$ is the smallest partition $P_\epsilon$ such that each pair $x$ and $x'$ belonging to the same element of $P_\epsilon$ satisfies $F[P(Y|x), P(Y|x')] \geq 1 - \epsilon$.

The quality of the reverse compression is clearly quantified by the cardinality of the variable $Z$, and for this matter we define the compressibility of a channel as
\[ \Gamma_\epsilon[P(Y|X)] := \frac{|X| - |P_x|}{|X| - 1}. \]  
(6)

The compressibility defines the amount of redundancy in the input, relative to the channel \( P(Y|X) \), and it is clearly a monotonically increasing function of \( \epsilon \). It also satisfies \( 0 \leq \Gamma_\epsilon[P(Y|X)] \leq 1 \), as the worst possible reverse compression consists in assigning a set to each possible input outcome and the best reverse compression is given by collecting all input outcomes into a single set.

Two limit cases can be readily identified. For a channel that outputs a variable independent of the input, \( P(Y|X = x) = P(Y|X = x') \) for all \( x, x' \), maximal compression is possible: we can replace all input variables with a single, arbitrary one without any effect on the output statistics, resulting in \( \Gamma_\epsilon = 1 \) for all \( \epsilon \). On the other end of the spectrum, an identity channel \( P(Y = y|X = x) = \delta_{yx} \) does not tolerate any compression: for any two distinct inputs \( x \neq x' \), the output distributions are perfectly distinguishable, \( F[P(Y|x), P(Y|x')] = 0 \), which means that, for any \( \epsilon > 0 \), the only \( \epsilon \)-reverse compression is the partition that assigns each \( x \in X \) to a distinct set, resulting in \( \Gamma_\epsilon = 0 \).

B. Asymptotic Settings

One can always take the asymptotic regime as a particular case of the single-shot setting mentioned above by considering the channel acting on infinitely many input variables, thus producing as many output ones. The channel \( P(Y|X) \) in the asymptotic regime is a conditional distribution of the type \( P(Y_1, Y_2, \cdots | X_1, X_2, \cdots) \). The reverse compression is identified in this case as a global operation on the joint input random variables, as illustrated in Figure 3.

In order to use the previous definition of reverse compression for an asymptotic scenario we need to define a partition with the same properties as above but defined on the space of input sequences. For doing so, consider \( k \) copies of the memoryless channel used without feedback, described by

\[ P(Y^k|X^k) = \prod_{i=1}^{k} P(Y_i|X_i). \]  
(7)

For a channel with such properties, the fidelity of the output distribution conditional on two different inputs takes the simple product form

\[ F[P(Y^k|x^k), P(Y^k|x')^k] = \prod_{i=1}^{k} F[P(Y_i|x_i), P(Y_i|x'_i)] \]  
(8)

where \( x^k \) and \( x^k \) are two arbitrary input sequences. It is clear that \( F[P(Y^k|x^k), P(Y^k|x')^k] = 1 \) if and only if each term \( F[P(Y_i|x_i), P(Y_i|x'_i)] = 1 \). A proof of Equation (8) can be found in the Appendix. It yields the definition of reverse compression for asymptotic settings.

**Definition 2 (Asymptotic Reverse Compression).** An asymptotic \( \epsilon \)-reverse compression of a memoryless channel used without feedback is a sequence of partitions \( \{P^{(k)}_{\epsilon}\}_{k \in \mathbb{N}} \) such that each \( P^{(k)}_{\epsilon} \) is an \( \epsilon \)-reverse compression of the channel \( P(Y^k|X^k) = \prod_{i=1}^{k} P(Y_i|X_i) \).

The above definition clearly does not define a unique sequence of partitions, but any \( \epsilon \)-reverse compression is equivalent for the purpose of the compression task. In the asymptotic regime, we quantity the quality of the method by the asymptotic compressibility defined as

\[ \Delta_\epsilon[P(Y|X)] := \lim_{k \to \infty} \Gamma^{(k)}_\epsilon[P(Y|X)], \]  
(9)

where \( \Gamma^{(k)}_\epsilon[P(Y|X)] \) is the compressibility of \( k \) independent uses of the channel \( P(Y|X) \). Similarly to the single-shot case, the asymptotic compressibility is a monotonically increasing function of \( \epsilon \). Therefore, the compressibility could have been used in the definition of reverse compression. Alternatively we could have defined it requesting the partition \( P^{(k)}_{\epsilon} \) to have maximum compressibility.

C. Examples

1. The erasure channel

The study case considered here is the erasure channel, outputting the erasure symbol denoted as \( a \) with
probability \( \eta \), and transmitting correctly the input symbol emitted by the source with probability \( 1 - \eta \). If the input space is \( \mathcal{X} = \{1, \cdots, r\} \), then the output space is \( \mathcal{Y} = \{1, \cdots, r, a\} \). The channel is described by the conditional probability distribution

\[
P_\eta(y|x) = (1 - \eta)\delta_{x,y} + \eta\delta_{x,y},
\]

with \( 0 \leq \eta \leq 1 \).

For the erasure, channel we have (see Appendix [2])

\[
F[P_\eta(y^k|x^k), P_\eta(y^k|\tilde{x}^k)] = \eta^{2S(x^k, \tilde{x}^k)},
\]

where \( S(x^k, \tilde{x}^k) \) is the number of different components of the sequences \( x^k \) and \( \tilde{x}^k \). For the problem considered here, we want to define the optimal partition – or appropriate bounds on it – such that, for each pair of input sequences \( x^k_i, \tilde{x}^k_i \) belonging to the same element of the partition, we have \( \eta^{2S(x^k, \tilde{x}^k)} \geq 1 - \epsilon \). Then, for fixed \( 0 < \epsilon, \eta < 1 \) we must have

\[
S(x^k, \tilde{x}^k) \leq \frac{1}{2} \log(1 - \epsilon) - \log \eta.
\]

Consider the case where \( \eta = 1/2 \), i.e., in each use of the channel the input variable \( x \) is perfectly mapped into the variable \( y = x \) with probability 1/2 or it is mapped to an error variable \( y = a \) with probability 1/2. In this case, the fidelity threshold reduces to \( (0.25)^{|S(x^k, \tilde{x}^k)|} \geq 1 - \epsilon \). Just to be clear, the best partition one could do is \( \{X^k\} \) and the worst possible partition would be \( \{\{x^k_i\}_{i \in X^k}\} \). Let us consider the partition \( \{\{x^k_i, \tilde{x}^k_i\}, \{w^k_i\}_{i \in X^\prime \setminus \{x^k, \tilde{x}^k\}}\} \), which has \( |X^k| - 1 \) elements. The best we could do is to take \( x^k \) and \( \tilde{x}^k \) with only one different component such that \( |S(x^k, \tilde{x}^k)| = 1 \), for instance \( x^k = (0,0,\cdots,0) \) and \( \tilde{x}^k = (1,0,\cdots,0) \). In this case, we should allow for an error \( \epsilon \geq 0.85 \) in fidelity. Thus we cannot compress input variables according to the reverse compression method with arbitrarily high fidelity for the erasure channel. The following example considers an example of asymptotic process with positive compressibility.

A possible way to make sure the reverse fidelity of two \( k \)-sequences with respect to the channel \( P(Y^k|X^k) \),

\[
\tilde{F}(x^k, \tilde{x}^k) := F[P(Y^k|x^k), P(Y^k|\tilde{x}^k)],
\]

is arbitrarily large is exploiting the factorization property stated in the Equation [3], i.e., if \( \tilde{F}(x_i, \tilde{x}_i) \geq (1 - \epsilon) \), then \( \tilde{F}(x^k, \tilde{x}^k) \geq 1 - \epsilon \).

Inspired by the erasure channel, consider the case where one can take partitions of the input space – with respect to the reverse compression – such that each sequence belonging to the same element of the partition has no more than \( s \) different components. If \( P_\epsilon^{(k)} \) is a reverse compression of the channel \( P(Y^k|X^k) \), then \( S(x^k, \tilde{x}^k) \leq s \) for each pair \( x^k_i, \tilde{x}^k_i \) belonging to the same element of the partition. We must have then

\[
|P_\epsilon^{(k)}| \leq |X|^k-s.
\]

We have furthermore the positive asymptotic compressibility \( \Delta \geq 1 - |X|^{-s} \). In Appendix [3] we conjecture that the equality is achieved. A similar situation for the single-shot regime is considered in the next example.

2. The generalized erasure channel

We can define a generalization of the erasure channel by considering several output erasure symbols as follows. Let \( P = \{A_i\}_{i=1} \) be a partition of \( \mathcal{X} \). Define \( P(Y|X) \) to be the channel such that \( P(Y = y_i|X = x_i) = (1 - \eta_i) \) and \( P(Y = \tilde{a}|X = x_i) = \eta_i \) for any \( x_i \in A_i \), with \( 0 \leq \eta_i \leq 1 \), and \( i = 1, \cdots, d \). See Figure [4]. If \( \min\eta_i^2 \geq 1 - \epsilon \), then the mapping \( x_i \mapsto x_i^\prime \) is an \( \epsilon \)-reverse compression of the channel, with an arbitrarily fixed \( x_i^\prime \) in \( A_i \). Its compressibility is equal to \( \Gamma_\epsilon = (|\mathcal{X}| - d)/(|\mathcal{X}| - 1) \).

The next section considers the reverse compression method for quantum communication schemes.

V. QUANTUM REVERSE COMPRESSION

In order to extend the reverse compression method to quantum processes it is crucial to properly define the concept of partition for quantum systems, also often called coarse graining. The definition of quantum
coarse graining to be considered here aims to deal with the problem of reverse compression, but it is possible to have alternative definitions depending on the problems addressed as it can be found for instance in [28,31].

We start by considering how the classical coarse graining associated with the reverse compression is represented as a CPTP map, i.e., a quantum channel. We then analyse the erasure channel, and confirm that the results remain valid using this framework. After this, we propose a natural generalization of quantum coarse graining, thus defining the reverse compression of a quantum channel.

A coarse graining should map the state space of a system into a state space of smaller dimension, and furthermore it must be regarded as a quantum channel $\Lambda^{X\rightarrow Z} : \mathcal{D}(\mathcal{H}_X) \rightarrow \mathcal{D}(\mathcal{H}_Z)$ such that $\mathcal{H}_Z$ is a subspace with $\dim(\mathcal{H}_Z) \leq \dim(\mathcal{H}_X)$. Denoting the correspondent classical partition by $\{A_z\}_{z \in \Lambda}$, we define the projector on the subspace spanned by $\{|x_z\rangle\}_{x_z \in A_z}$ as

$$\Pi_z := \sum_{x_z \in A_z} |x_z\rangle \langle x_z|,$$

where $|\psi\rangle$ stands for the pure state $|\psi\rangle \langle \psi|$. The quantum operation corresponding to $\lambda : x_z \mapsto z$ is the measurement of a state $\rho$ in the subspace corresponding to $A_z$ and the preparation of the pure state $|z\rangle$, represented by

$$\Lambda^{X\rightarrow Z}_z(\rho) := |z\rangle \langle z| \text{Tr}(\Pi_z \rho).$$

It is a CP mapping $\mathcal{L}(\mathcal{H}_X) \rightarrow \mathcal{L}(\mathcal{H}_Z)$ with decomposition $\Lambda^{X\rightarrow Z}_z(\rho) = \sum_{x_z \in A_z} K^{(z)}_{x_z} \rho K^{(z)*}_{x_z}$ given by Kraus operators defined as $K^{(z)}_{x_z} = |z\rangle \langle x_z|$. The quantum coarse graining corresponding to a classical partition is defined by summing over all outcomes

$$\Lambda^{X\rightarrow Z} := \sum_{z \in \Lambda} \Lambda^{X\rightarrow Z}_z.$$  

(17)

It is clearly a CPTP map as its Kraus operators satisfy $\sum_{z \in \Lambda} \sum_{x_z \in A_z} K^{(z)*}_{x_z} K^{(z)}_{x_z} = \mathbb{1}_X$, where $\mathbb{1}_X$ is the identity operator acting on $\mathcal{H}_X$.

Partial trace as a coarse graining.—Consider the bipartite Hilbert space $\mathcal{H}_X = \mathcal{H}_Z \otimes \mathcal{H}_W$ and the coarse-graining

$$\Lambda^{X\rightarrow Z}(\rho) = \text{Tr}_W(\rho) = \sum_w \langle w | \langle w| \rho (|w\rangle \langle w|),$$

(18)

where $\{|w\rangle\}_{w \in \mathcal{W}}$ is an orthonormal basis of $\mathcal{H}_W$. It corresponds to a classical partition of $X = Z \times W$ by setting the elements $A_z := \{(z, w) : w \in \mathcal{W}\}$. That is, a uniform partition.

Similarly to its classical version, the quantum erasure channel outputs the input state $\rho$ with probability $1 - \eta$ and outputs a pure erasure state $|a\rangle$ with probability $\eta$. The erasure state is orthogonal to any input one, that is, $\text{Tr}(|a\rangle \langle a| \rho) = 0$. It means the input Hilbert space $\mathcal{H}_X$ must have smaller dimension than the output Hilbert space $\mathcal{H}_Y$, and more precisely $\mathcal{H}_Y \cong \mathbb{C}^{X'_{\eta} + 1}$. This channel is denoted here as

$$\mathcal{M}_\eta^{X\rightarrow Y}(\rho) = (1-\eta)\rho + \eta |a\rangle \langle a|.$$

(19)

Consider the case where the input states are taken to be pure states of the canonical basis. The reverse compression here consists in defining the optimal partition such that each pair of pure states $|x_z\rangle$ and $|\tilde{x}_z\rangle$ satisfies the condition

$$\tilde{F}([x_z], [\tilde{x}_z]) := \mathcal{F}(\mathcal{M}_\eta^{X\rightarrow Y}([x_z]), \mathcal{M}_\eta^{X\rightarrow Y}([\tilde{x}_z])) \geq 1 - \epsilon.$$

(20)

As the reverse fidelity $\tilde{F}$ of distinct pure states $|x\rangle$ and $|\tilde{x}\rangle$ with respect to the erasure channel is equal to $\eta^2$, then the quantum reverse compression goes in the same lines as its classical counterpart: it is either possible to perfectly (reversely) compress every pure state in the canonical basis to the same fixed pure state if $\eta^2 \geq 1 - \epsilon$, or it is not possible to (reversely) compress at all if $\eta^2 < 1 - \epsilon$.

The coarse graining operation defined in Equation [17] maps distinct collections of pure orthogonal states into distinct fixed orthogonal pure states. We can consider more general coarse grainings, without a direct classical analogue, that do not rely on a particular basis choice. In order to define reverse compression for such cases, we introduce the vectorial kernel of a linear map $\Lambda^{X\rightarrow X'} : \mathcal{L}(\mathcal{H}_X) \rightarrow \mathcal{L}(\mathcal{H}_{X'})$, with $\mathcal{H}_X \cong \mathcal{H}_{X'}$, as the set

$$\mathcal{K}(\Lambda^{X\rightarrow X'}) := \{|\psi\rangle \in \mathcal{H}_{X'} : \Lambda^{X\rightarrow X'}(\rho) |\psi\rangle = 0, \forall \rho \in \mathcal{D}(\mathcal{H}_X)\}.$$  

(21)

This is clearly a subspace of the associated channel’s input vector space. More precisely it is a subspace of the output space of the coarse graining operation, which is defined to be isomorphic to the input one. The range of the map $\Lambda^{X\rightarrow X'}$,

$$\mathcal{R}(\Lambda^{X\rightarrow X'}) := \{\Lambda^{X\rightarrow X'}(\rho) : \rho \in \mathcal{D}(\mathcal{H}_X)\}.$$  

(22)

is equal to the set of linear operators acting on the orthogonal complement of the vector kernel of the map,
The quantum erasure channel can be reversely compressed if and only if indistinguishable, and denote it as of a quantum coarse graining plays the role of the classical compressed input space $Z$, in the sense that, after coarse graining, only states from within the range are sent through the channel.

We say channels $R^{X\rightarrow Y}$ and $T^{X\rightarrow Y}$ are $\epsilon$-indistinguishable, and denote it as $R^{X\rightarrow Y} \leq T^{X\rightarrow Y}$, if and only if they satisfy $F[R^{X\rightarrow Y}(\rho), T^{X\rightarrow Y}(\rho)] \geq 1 - \epsilon$ for any state $\rho$ in $H_X$. The reverse compression method for an arbitrary channel in the most general setting is defined as follows.

**Definition 3** (Single-Shot Quantum Reverse Compression). An $\epsilon$-reverse compression of the quantum channel $N^{X\rightarrow Y}$ is a quantum channel $\Lambda^{X\rightarrow X'}$ such that

$$N^{X\rightarrow Y} \circ \Lambda^{X\rightarrow X'} \notin R^{X\rightarrow Y},$$

with maximum quantum compressibility

$$\Gamma_e(M^{X\rightarrow Y}) := \frac{\dim K(\Lambda^{X\rightarrow X'})}{\dim H^X - 1}. \tag{24}$$

A non-trivial compression is identified as a non-full range CPTP map $\Lambda^{X\rightarrow X'} : \mathcal{L}(H_X) \rightarrow \mathcal{L}(H_{X'})$. Considering this most general definition, the quantum reverse compression is clearly a superchannel acting on the channel of interest. A superchannel is a linear transformation of quantum channels into quantum channels, with possibly different dimensions of input and output spaces. A transformation of this type (for the case of single-party) occur in general as an isometric preprocessing and postprocessing with a side identity channel [14]. In what follows, we show that, even allowing for arbitrary quantum coarse grainings, the reverse compression holds the same result for the erasure channel.

**Theorem 1.** The quantum erasure channel can be reversely compressed if and only if $\eta^2 \geq 1 - \epsilon$.

**Proof.** Direct part.—If $\eta^2 \geq 1 - \epsilon$, then the input space can be perfectly compressed to a fixed state. It can be easily checked from the reverse fidelity for arbitrary input states as

$$F[\rho, \Lambda^{X\rightarrow X'}(\rho)] := F[M^{X\rightarrow Y}(\rho), M^{X\rightarrow Y}_{\eta}(\Lambda^{X\rightarrow X'}(\rho))] = \left(1 - \eta^2\right) F[\rho, \Lambda^{X\rightarrow X'}(\rho)] + \eta^2 \geq \eta^2. \tag{25}$$

As a consequence, we can reversely compress all the input space to an arbitrary pure input state according to $\Lambda^{X\rightarrow X'}(\rho) = [1$. This clearly satisfies the threshold $F[\rho, \Lambda^{X\rightarrow X'}(\rho)] \geq 1 - \epsilon$, and achieves the highest possible compressibility $\Gamma_e(M^{X\rightarrow Y}) = 1$.

Converse part.—If $\eta^2 < 1 - \epsilon$, then no reverse compression is possible — meaning that any reverse compression $\Lambda^{X\rightarrow X'}$ would be a full range map. In order to check this, suppose there is a non-trivial vector in the vector kernel of the map, $|\psi\rangle \in K(\Lambda^{X\rightarrow X'})$. In such a case it holds that $F([\psi], \Lambda^{X\rightarrow X'}(|\psi\rangle)) = 0$, and therefore, using the expression (25) for the fidelity of the erasure channel, we find $F[M^{X\rightarrow Y}_{\eta}(\psi), M^{X\rightarrow Y}_{\eta}(\Lambda^{X\rightarrow X'}(|\psi\rangle))] = \eta^2 < 1 - \epsilon$. We conclude that, for $\eta^2 < 1 - \epsilon$, the reverse compression map must have a trivial vector kernel, achieving compressibility $\Gamma_e(M^{X\rightarrow Y}) = 0$.

As it was the case for the classical reverse compression, we can easily extend the quantum reverse compression to the asymptotic regime by considering its action on several copies of the input space as follows.

**Definition 4** (Asymptotic Quantum Reverse Compression). An asymptotic quantum $\epsilon$-reverse compression of a channel $N^{X\rightarrow Y}$ is a sequence of quantum channels $\left(\Lambda_k^{X\rightarrow X'^k} : \mathcal{L}(H_{X^k}) \rightarrow \mathcal{L}(H_{X'^k})\right)_{k \in \mathbb{N}}$, such that each channel $\Lambda_k$ is an $\epsilon$-reverse compression of $k$ independent copies of the channel $N$, with compressibility $\Gamma_e^{(k)}(N) := \Gamma_e(N^k)$. The asymptotic compressibility is defined as the limit $\Gamma_e(N^{X\rightarrow Y}) := \lim_{k \rightarrow \infty} \Gamma_e^{(k)}(N)$.

For the case of separable input states the asymptotic quantum reverse compression works similarly to its classical counterpart. The more general case of entangled input states is left for future studies.

VI. CONCLUSION

We have considered here the task of preprocessing classical and quantum systems according to a method we called reverse compression. It can be understood as complementary to Shannon compression. While the latter is irrelevant for sources with uniform probability distributions, the former is irrelevant for identity channels, the case where distinct data is perfectly distinguishable for the receiver. Additionally, we have considered the erasure channel as a study case. This presents an all vs nothing situation where, depending on the amount of noise tolerated, it is either possible to perfectly compress the input of the channel or it is not possible at all. This happens because the noise in the erasure channel does not depend upon the input
state. A more detailed analysis of asymptotic settings is left for future studies as well. Moreover, the study of different preprocessing operations, and general channel simulation tasks is an open area.

ACKNOWLEDGMENTS

F.C. acknowledges support through an Australian Research Council Discovery Early Career Researcher Award (DE170100712). M.C. acknowledges support from the the funding agency CNPq. M.C. acknowledges the warmly hospitality of the School of Mathematics and Physics at the University of Queensland, and also thanks to Lucas Chibebe Céleri for useful discussions. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. We acknowledge the traditional owners of the land on which the University of Queensland is situated, the Turrbal and Jagera people.

[1] C. E. Shannon, “A mathematical theory of communication,” The Bell system technical journal, vol. 27, no. 3, pp. 379–423, 1948.
[2] E. T. Jaynes, “Information theory and statistical mechanics,” Physical review, vol. 106, no. 4, p. 620, 1957.
[3] C. E. Shannon, “A note on a partial ordering for communication channels,” Information and control, vol. 1, no. 4, pp. 390–397, 1958.
[4] Y. Zhang and C. Tepedelenlioğlu, “Analytical and numerical characterizations of shannon ordering for discrete memoryless channels,” IEEE transactions on information theory, vol. 60, no. 1, pp. 72–83, 2013.
[5] R. Nasser, “Characterizations of two channel orderings: Input-degradedness and the shannon ordering,” IEEE Transactions on Information Theory, vol. 64, no. 10, pp. 6759–6770, 2018.
[6] F. Haddadpour, M. H. Yassaei, S. Beigi, A. Gohari, and M. R. Aref, “Simulation of a channel with another channel,” IEEE Transactions on Information Theory, vol. 63, no. 5, pp. 2659–2677, 2016.
[7] R. G. D’Oliveira and M. Firer, “A distance between channels: the average error of mismatched channels,” Designs, Codes and Cryptography, vol. 87, no. 2-3, pp. 481–493, 2019.
[8] D. Blackwell, “Equivalent comparisons of experiments,” The annals of mathematical statistics, pp. 265–272, 1953.
[9] A. Jenčová, “Comparison of quantum channels and statistical experiments,” in 2016 IEEE International Symposium on Information Theory (ISIT), pp. 2249–2253, IEEE, 2016.
[10] F. Buscemi, “Comparison of noisy channels and reverse data-processing theorems,” in 2017 IEEE Information Theory Workshop (ITW), pp. 489–493, IEEE, 2017.
[11] F. Buscemi, M. Keyl, G. M. D’Ariano, P. Perinotti, and R. F. Werner, “Clean positive operator valued measures,” Journal of mathematical physics, vol. 46, no. 8, p. 082109, 2005.
[12] G. Gour, “Comparison of quantum channels by superchannels,” IEEE Transactions on Information Theory, vol. 65, no. 9, pp. 5980–5994, 2019.
[13] P. Faist, M. Berta, and F. Brandão, “Thermodynamic capacity of quantum processes,” Physical review letters, vol. 122, no. 20, p. 200601, 2019.
[14] G. Chiribella, G. M. D’Ariano, and P. Perinotti, “Transforming quantum operations: Quantum supermaps,” EPL (Europhysics Letters), vol. 83, no. 3, p. 30004, 2008.
1. Reverse fidelity of memoryless channels used without feedback

The fidelity of conditional probabilities distributions \( P(Y^k|x^k) = \prod_{i=1}^{k} P(Y_i|x_i) \) and \( P(Y^k|x^k) = \prod_{i=1}^{k} P(Y_i|x_i) \) can be written as \( F[P(Y^k|x^k), P(Y^k|x^k)] = \prod_{i=1}^{k} F[P(Y_i|x_i), P(Y_i|x_i)] \). The proof follows as

\[
F[P(Y^k|x^k), P(Y^k|x^k)] = \left[ \sum_{y^k} \sqrt{P(y^k|x^k)P(y^k|x^k)} \right]^2 \\
= \left[ \sum_{y^k} \left( \prod_{i=1}^{k} P(y_i|x_i)P(y_i|x_i) \right) \right]^2 \\
= \left[ \prod_{i=1}^{k} \sum_{y_i} \sqrt{P(y_i|x_i)P(y_i|x_i)} \right]^2 \\
= \prod_{i=1}^{k} F[P(y_i|x_i), P(y_i|x_i)]. \tag{26}
\]

A quantum version of Equation \( 26 \) can be achieved by considering the action of \( k \) independent copies of a channel \( N^{X^k \rightarrow Y^k} = N^{X_1 \rightarrow Y_1} \otimes \cdots \otimes N^{X_k \rightarrow Y_k} \) on i.i.d copies of states \( \rho \) and \( \sigma \), that is, \( \rho^{X^k} = \rho^{X_1} \otimes \cdots \otimes \rho^{X_k} \) and \( \sigma^{X^k} = \sigma^{X_1} \otimes \cdots \otimes \sigma^{X_k} \). It follows that

\[
F\left[ N^{X^k \rightarrow Y^k}(\rho^{X^k}), N^{X^k \rightarrow Y^k}(\sigma^{X^k}) \right] = \left[ \text{Tr}_{Y^k} \left( \sqrt{\sqrt{N^{X^k \rightarrow Y^k}(\rho^{X^k})}N^{X^k \rightarrow Y^k}(\sigma^{X^k})\sqrt{N^{X^k \rightarrow Y^k}(\rho^{X^k})} \right)^2 \right] \\
= \prod_{i=1}^{k} \left[ \text{Tr}_{Y_i} \left( \sqrt{\sqrt{N^{X_i \rightarrow Y_i}(\rho^{X_i})}N^{X_i \rightarrow Y_i}(\sigma^{X_i})\sqrt{N^{X_i \rightarrow Y_i}(\rho^{X_i})} \right)^2 \right] \\
= \prod_{i=1}^{k} F\left[ N^{X_i \rightarrow Y_i}(\rho^{X_i}), N^{X_i \rightarrow Y_i}(\sigma^{X_i}) \right]. \tag{27}
\]

2. Reverse fidelity of the erasure channel

We present here a detailed derivation of the reverse fidelity of an erasure channel. Recall that the erasure channel is described by the conditional probability distribution \( P(y|x) = (1 - \eta) \delta_{x,y} + \eta \delta_{a,y} \), with \( x \in X := \{1, \cdots, r\} \), \( y \in Y := \{1, \cdots, r, a\} \), and \( 0 \leq \eta \leq 1 \). It follows then
\[ \tilde{F}(x, \tilde{x}) \triangleq F[P(Y|x), P(Y|\tilde{x})] = \left[ \sum_{y \in Y} \sqrt{[(1-\eta)\delta_{x,y} + \eta \delta_{u,y}][(1-\eta)\delta_{\tilde{x},y} + \eta \delta_{u,y}]} \right]^2 = \left[ \sum_{y \in Y} \sqrt{(1-\eta)^2 \delta_{x,y} \delta_{\tilde{x},y} + (1-\eta)\eta \delta_{x,y} \delta_{u,y} + \eta (1-\eta) \delta_{u,y} \delta_{\tilde{x},y} + (\eta \delta_{u,y})^2} \right]^2 = \begin{cases} 1, & x = \tilde{x} \\ \eta^2, & x \neq \tilde{x} \end{cases} \] (28)

Considering Equation (28) we get the reverse fidelity of a memoryless erasure channel used without feedback
\[
\tilde{F}(x^k, \tilde{x}^k) \triangleq F[P(Y^k|x^k), P(Y^k|\tilde{x}^k)] = \prod_{i=1}^{k} F[P(Y_i|x_i), P(Y_i|\tilde{x}_i)] = \left( \eta^2 \right)^{S(x^k, \tilde{x}^k)} \] (29)

where \( S(x^k, \tilde{x}^k) \) is the number of different components of \( x \) and \( \tilde{x} \).

3. On the conjecture considered in section IV C

We consider here the asymptotic reverse compression of channels, allowing for some error \( e \) in fidelity, such that the elements of the associated partition \( P_e^{(k)} \) cannot have more than \( s \) different components. First we show the validity of the upper bound
\[
|P_e^{(k)}| \leq |X|^{k-s}. \] (30)

It can be proved by building a partition with such properties and with cardinality equal to \( |X|^{k-s} \). As an example, we consider binary sequences \( (X = \{0, 1\}) \) of length \( k = 3 \). All possible sequences can be written in lexicographical order as follows: 000, 001, 010, 011, 100, 101, 110 and 111. If \( s = 0 \) there is nothing to be done as the only possible partition is the trivial \( \{(x^3)\}_{x^3} \). Therefore, in this case we can only have \( |P_e^{(3)}| = 8 = |X|^3 \). For \( s = 1 \), we can merge two sequences together as in the partition \( \{A_1, A_2, A_3, A_4\} \), where \( A_1 = \{000, 001\}, A_2 = \{010, 011\}, A_3 = \{100, 101\} \) and \( A_4 = \{110, 111\} \). This gives us \( |P_e^{(3)}| = 4 = |X|^{3-1} \). For \( s = 2 \), it is possible to define a partition following the same reasoning, \( \{A_1 \cup A_2 = \{000, 001, 010, 011\}, A_3 \cup A_4 = \{100, 101, 110, 111\}\} \), and for \( s = 3 \) all the sequences can be merged together generating the partition \( \{A_1 \cup A_2 \cup A_3 \cup A_4\} \), in both cases saturating the bound (30). This procedure can clearly be extended to arbitrary \( k \)-ary input spaces, and there is a partition \( Q_e^{(k)} \) with the desired properties, and such that \( |Q_e^{(k)}| = |X|^{k-s} \).

For the preceding example it can be shown there is no partition with the defining properties such that it has fewer than \(|X|^{k-s}\) elements. For \( s = 0 \) and 3, the proof is trivial. In order to prove that for \( s = 2 \), define the partition of \( \{0, 1\}^3 \) given by sets \( B_1 = \{000, 111\}, B_2 = \{001, 110\}, B_3 = \{010, 101\} \) and \( B_4 = \{011, 100\} \). Such partition is defined in a way that each element \( B_1 \) consists of sequences with \( s = 3 \) different components. Therefore, the desired assertion follows by the pigeonhole principle for \( s = 2 \) as each \( B_j \) contains two strings. That is, we can at most take a partition with two elements \( A_1 \) and \( A_2 \), such that each \( A_i \) (\( i = 1, 2 \)) contains at most one string in \( B_j \) (\( j = 1, \ldots, 4 \)). For instance, one such partition could be with elements \( A_1 = \{000, 001, 010, 100\} \), and \( A_1 = \{110, 111, 101, 111\} \). For \( s = 1 \), we can prove it by considering the partition with sets \( B_1 = \{000, 011, 110, 101\} \) and \( B_1 = \{111, 100, 000, 010\} \). It is defined in such a way that \( x, y \in B_j (j = 1, 2) \), implies \( S(x, y) \geq 2 \). Therefore, the proof follows similarly by the pigeonhole principle as each \( B_j \) contains 4 strings. We refer the reader to [32] for a clear presentation on the application of the pigeonhole principle to problems in combinatorics. Note that for \( k = 4 \) it is already not so simple to show that, as the same line of reasoning does not hold. By this reason, we conjecture that the equality must hold for any value of \( k \).
4. Asymptotic compressibility of the generalized erasure channel

We show here that the generalized erasure channel defined in [VWC2] has zero asymptotic compressibility. It can be seen by exploiting its definition that \( \tilde{F}(x_i, x_j) = 0 \), for \( x_i \in \mathcal{A}_i \) and \( x_j \in \mathcal{A}_j \), with \( i \neq j \). Therefore, if \( \epsilon < 1 \) and one wants to reversely compress the input sequences of \( \mathcal{X}^k \), one cannot map sequences with components belonging to different sets \( \mathcal{A}_i \) to the same compressed one. Otherwise, it would lead to zero reverse fidelity. Then, the best we can do is to compress the sequences with all the components in the same set \( \mathcal{A}_i \) to the same sequence. That means we are exploiting the partition

\[
\mathcal{X}^k = \left( \bigcup_{i=1}^{d} \mathcal{A}_i^k \right) \cup \mathcal{B},
\]

where \( \mathcal{B} \) contains all the sequences which is not in any of the sets \( \mathcal{A}_i^k \). Precisely, the partition \( \mathcal{P}_\epsilon^{(k)} = \{ \mathcal{A}_1^k, \cdots, \mathcal{A}_d^k, \mathcal{B} \} \) has cardinality \( |\mathcal{P}_\epsilon^{(k)}| = d + |\mathcal{B}| \).

Note that \( \mathcal{X} = \bigcup_{i=1}^{d} \mathcal{A}_i \) and \( |\mathcal{X}| = \sum_{i=1}^{d} |\mathcal{A}_i| \). Therefore, it follows that

\[
|\mathcal{X}^k| = \left( \sum_{i=1}^{d} |\mathcal{A}_i| \right)^k = \sum_{i=1}^{d} |\mathcal{A}_i|^k + |\mathcal{B}|,
\]

and we must have

\[
|\mathcal{B}| = \left( \sum_{i=1}^{d} |\mathcal{A}_i| \right)^k - \sum_{i=1}^{d} |\mathcal{A}_i|^k.
\]

By definition, the compressibility of \( k \) independent uses of the generalized erasure channel is upper bounded by

\[
\Gamma^{(k)}_{\epsilon} \leq \frac{\sum_{i=1}^{d} |\mathcal{A}_i|^k - d}{\left( \sum_{i=1}^{d} |\mathcal{A}_i| \right)^k - 1}.
\]

Moreover, if \( \epsilon < 1 \), then the asymptotic compressibility of the generalized erasure channel is \( \Delta_{\epsilon} = 0 \).