arXiv:1711.00697v3 [quant-ph] 5 Apr 2024

Approximating quantum channels by completely positive maps with small Kraus rank

Cécilia Lancien[1] and Andreas Winter[2]

[1]Institut Fourier & CNRS, Université Grenoble Alpes, 38610 Gières, France

[2]Departament de Física: Grup d'Informació Quàntica, Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), Spain & Institució Catalana de Recerca i Estudis Avançats, 08010 Barcelona, Spain

We study the problem of approximating a quantum channel by one with as few Kraus operators as possible (in the sense that, for any input state, the output states of the two channels should be close to one another). Our main result is that any quantum channel mapping states on some input Hilbert space A to states on some output Hilbert space B can be compressed into one with order $d \log d$ Kraus operators, where $d = \max(|\mathrm{A}|, |\mathrm{B}|)$, hence much less than $|\mathrm{A}||\mathrm{B}|$. In the case where the channel's outputs are all very mixed, this can be improved to order d. We discuss the optimality of this result as well as some consequences.

1 Introduction

Quantum channels are the most general framework in which the transformations that a quantum system may undergo are described. These are defined as completely positive and trace preserving (CPTP) maps from the set of bounded operators on some input Hilbert space A to the set of bounded operators on some output Hilbert space B. Indeed, to be a physically valid evolution in the open system setting, a linear map $\mathcal{N}$ has to preserve quantum states (i.e. positive semi-definiteness and unit-trace conditions) even when tensorized with the identity map $\mathcal{I}$ on an auxiliary system.

Let us fix here once and for all some notation that we will use repeatedly in the remainder of the paper: Given a Hilbert space H, we shall denote by $\mathcal{L}(\mathrm{H})$ the set of linear operators on H, and by $\mathcal{D}(\mathrm{H})$ the set of density operators (i.e. positive semi-definite and trace 1 operators) on H. Also, whenever H is finite dimensional (which will be the case of all the Hilbert spaces we will deal with in the sequel) we shall denote by $|\mathrm{H}|$ its dimension.

So assume from now on that the Hilbert spaces A and B are finite dimensional. Then, we know by Choi's representation theorem [6] that a CPTP map $\mathcal{N} : \mathcal{L}(\mathrm{A}) \to \mathcal{L}(\mathrm{B})$ can always be written as

$$\mathcal{N} : X \in \mathcal{L}(\mathrm{A}) \mapsto \sum_{i=1}^{s} K_i X K_i^\dagger \in \mathcal{L}(\mathrm{B}), \tag{1}$$

where the operators $K_i : \mathrm{A} \to \mathrm{B}$, $1 \leqslant i \leqslant s$, are called the Kraus operators of $\mathcal{N}$ and satisfy the normalization relation $\sum_{i=1}^{s} K_i^\dagger K_i = \mathbb{1}_{\mathrm{A}}$. The minimal $s \in \mathbf{N}$ such that $\mathcal{N}$ can be decomposed in the Kraus form (1) is called the Kraus rank of $\mathcal{N}$, which we shall

Cécilia Lancien: cecilia.lancien@univ-grenoble-alpes.fr

Andreas Winter: andreas.winter@uab.cat

denote by $r_K(\mathcal{N})$. By Stinespring's dilatation theorem [11], another alternative way of characterizing a CPTP map $\mathcal{N} : \mathcal{L}(\mathrm{A}) \to \mathcal{L}(\mathrm{B})$ is as follows

$$\mathcal{N} : X \in \mathcal{L}(\mathrm{A}) \mapsto \mathrm{Tr}_{\mathrm{E}} \left(V X V^\dagger \right) \in \mathcal{L}(\mathrm{B}), \tag{2}$$

for some environment Hilbert space E and some isometry $V : \mathrm{A} \hookrightarrow \mathrm{B} \otimes \mathrm{E}$ (i.e. $V^\dagger V = \mathbb{1}_{\mathrm{A}}$). In such picture, $r_K(\mathcal{N})$ is then nothing else than the minimal environment dimension $|\mathrm{E}| \in \mathbf{N}$ such that $\mathcal{N}$ may be expressed in the Stinespring form (2). It may be worth pointing out that there is a lot of freedom in representation (1): two sets of Kraus operators $\{K_i,\ 1 \leqslant i \leqslant s\}$ and $\{L_i,\ 1 \leqslant i \leqslant s\}$ give rise to the same quantum channel as soon as there exists a unitary U on $\mathbf{C}^s$ such that, for all $1 \leqslant i \leqslant s$, $L_i = \sum_{j=1}^s U_{ij} K_j$. On the contrary, representation (2) is essentially unique, up to the (usually irrelevant) transformation $V \mapsto (\mathbb{1} \otimes U)V$, for U a unitary on $\mathbf{C}^s$. That is why we will often prefer working with the latter than with the former.

Yet another way of viewing the Kraus rank of a CPTP map $\mathcal{N} : \mathcal{L}(\mathrm{A}) \to \mathcal{L}(\mathrm{B})$ is as the rank of its associated Choi-Jamiołkowski state. Denoting by ψ the maximally entangled state on $\mathrm{A} \otimes \mathrm{A}$, the latter is defined as the state $\tau(\mathcal{N}) = \mathcal{I} \otimes \mathcal{N}(\psi)$ on $\mathrm{A} \otimes \mathrm{B}$. Consequently, it holds that any quantum channel from A to B has Kraus rank at most $|\mathrm{A}||\mathrm{B}|$. And the extremal such quantum channels are those with Kraus rank less than $|\mathrm{A}|$. In particular, the case $r_K(\mathcal{N}) = 1$ corresponds to $\mathcal{N}$ being a unitary, hence reversible, evolution, whereas whenever $r_K(\mathcal{N}) > 1$, one can view $\mathcal{N}$ as a noisy summary of a unitary evolution on a larger system. The Kraus rank of a quantum channel can thus legitimately be seen as a measure of its "complexity": it quantifies the minimal amount of ancillary resources needed to implement it (or equivalently the amount of degrees of freedom in it that one is ignorant of). A natural question in this context would therefore be: given any quantum channel, is it possible to reduce its complexity while not affecting too much its action, or in other words to find a channel with much smaller Kraus rank which approximately simulates it?

One last definition we shall need concerning CP maps is the following: the conjugate (or dual) of a CP map $\mathcal{N} : \mathcal{L}(\mathrm{A}) \to \mathcal{L}(\mathrm{B})$ is the CP map $\mathcal{N}^* : \mathcal{L}(\mathrm{B}) \to \mathcal{L}(\mathrm{A})$ defined by

$$\forall\ X \in \mathcal{L}(\mathrm{A}),\ \forall\ Y \in \mathcal{L}(\mathrm{B}),\ \mathrm{Tr}(\mathcal{N}(X)Y) = \mathrm{Tr}(X\mathcal{N}^*(Y)).$$

It is characterized as well by saying that $\{K_i,\ 1 \leqslant i \leqslant s\}$ is a set of Kraus operators for $\mathcal{N}$ if and only if $\{L_i = K_i^\dagger,\ 1 \leqslant i \leqslant s\}$ is a set of Kraus operators for $\mathcal{N}^*$. Hence obviously, $\mathcal{N}$ and $\mathcal{N}^*$ have same Kraus rank, while the trace-preservingness condition $\sum_{i=1}^s K_i^\dagger K_i = \mathbb{1}$ for $\mathcal{N}$ is equivalent to the unitality condition $\sum_{i=1}^s L_i L_i^\dagger = \mathbb{1}$ for $\mathcal{N}^*$.

The remainder of this paper is organized as follows. In Section 2 we gather all needed background on quantum channel approximation that we are interested in. This includes precise definitions, previous works in this direction, etc. Our main results are then stated and commented in Section 3, while their proofs are relegated to Section 4. In Section 5 we present several corollaries, which have applications in quantum data hiding and locking, amongst other. We finally discuss some open questions in Section 6.

2 Quantum channel approximation: definitions and already known facts

Before going any further, we need to specify what we mean by "approximating a quantum channel", since indeed, several definitions of approximation may be considered. In our setting, the most natural one is probably that of approximation in (1→1)-norm: given CPTP maps $\mathcal{N}, \widehat{\mathcal{N}} : \mathcal{L}(\mathrm{A}) \to \mathcal{L}(\mathrm{B})$, we will say that $\widehat{\mathcal{N}}$ is an ε-approximation of $\mathcal{N}$ in

(1→1)-norm, where $\varepsilon > 0$ is some fixed parameter, if

$$\forall\ \varrho \in \mathcal{D}(\mathrm{A}),\ \left\|\widehat{\mathcal{N}}(\varrho) - \mathcal{N}(\varrho)\right\|_1 \leqslant \varepsilon. \tag{3}$$

At first sight it might appear that an even more natural error quantification in such a context would be in terms of the completely-bounded (1→1)-norm (aka diamond norm) [1]. That is, in order to call $\widehat{\mathcal{N}}$ an ε-approximation of $\mathcal{N}$, we would require that, for any Hilbert space A',

$$\forall\ \varrho \in \mathcal{D}(\mathrm{A} \otimes \mathrm{A}'),\ \left\|\widehat{\mathcal{N}} \otimes \mathcal{I}(\varrho) - \mathcal{N} \otimes \mathcal{I}(\varrho)\right\|_1 \leqslant \varepsilon. \tag{4}$$

Nevertheless, this notion of approximation is too strong for our purposes. Indeed, if $\mathcal{N}$ and $\widehat{\mathcal{N}}$ satisfy Equation (4), it implies in particular that their associated Choi-Jamiołkowski states have to be ε-close in trace-norm. And this, in general, is possible only if $\mathcal{N}$ and $\widehat{\mathcal{N}}$ have the same, or at least comparable, number of Kraus operators, so that no environment dimensionality reduction can be achieved.

The question of quantum channel compression has already been studied in one specific case, which is the one of the fully randomizing (or depolarizing) channel. Let us recall what is known there. The fully randomizing channel $\mathcal{R} : \mathcal{L}(\mathrm{A}) \to \mathcal{L}(\mathrm{A})$ is the CPTP map with same input and output spaces defined by

$$\mathcal{R} : X \in \mathcal{L}(\mathrm{A}) \mapsto (\mathrm{Tr}\, X)\frac{\mathbb{1}}{|\mathrm{A}|} \in \mathcal{L}(\mathrm{A}),$$

so that, in particular, all input states $\varrho \in \mathcal{D}(\mathrm{A})$ are sent to the maximally mixed state $\mathbb{1}/|\mathrm{A}| \in \mathcal{D}(\mathrm{A})$. $\mathcal{R}$ has maximal Kraus rank $|\mathrm{A}|^2$ (because $\tau(\mathcal{R})$ is simply $\mathbb{1}/|\mathrm{A}|^2$, and hence has rank $|\mathrm{A}|^2$). This was of course to be expected, if adhering to the intuitive idea that the bigger is the Kraus rank of channel, the noisier is the channel. One possible minimal Kraus decomposition for $\mathcal{R}$ is

$$\mathcal{R} : X \in \mathcal{L}(\mathrm{A}) \mapsto \frac{1}{|\mathrm{A}|^2}\sum_{j,k=1}^{|\mathrm{A}|} V_{jk} X V_{jk}^\dagger \in \mathcal{L}(\mathrm{A}),$$

where for each $1 \leqslant j, k \leqslant |\mathrm{A}|$, $V_{jk} = X^j Z^k$ with X and Z the generalized Pauli shift and phase operators on A. It was initially established in [8] and later improved in [2] that there exist almost randomizing channels with drastically smaller Kraus rank. More specifically, the following was proved: for any $0 < \varepsilon < 1$, the CPTP map $\mathcal{R}$ can be ε-approximated in (1→1)-norm by a CPTP map $\widehat{\mathcal{R}}$ with Kraus rank at most $C|\mathrm{A}|/\varepsilon^2$, where $C > 0$ is a constant that is independent of $|\mathrm{A}|$ and ε. Actually, something stronger was established, namely

$$\forall\ \varrho \in \mathcal{D}(\mathrm{A}),\ \left\|\widehat{\mathcal{R}}(\varrho) - \mathcal{R}(\varrho)\right\|_\infty \leqslant \frac{\varepsilon}{|\mathrm{A}|},$$

which obviously implies that, for any $1 \leqslant p \leqslant \infty$, $\widehat{\mathcal{R}}$ is an ε-approximation of $\mathcal{R}$ in (1→p)-norm, in the sense that

$$\forall\ \varrho \in \mathcal{D}(\mathrm{A}),\ \left\|\widehat{\mathcal{R}}(\varrho) - \mathcal{R}(\varrho)\right\|_p \leqslant \frac{\varepsilon}{|\mathrm{A}|^{1-1/p}}. \tag{5}$$

The question we investigate here is whether such kind of statement actually holds true for any channel. Note however that, for a channel which is not the fully randomizing one, imposing an approximation in Schatten-p-norm up to an error $\varepsilon/|\mathrm{B}|^{1-1/p}$, as the

error appearing in Equation (5), is potentially too strong. Indeed, the fully randomizing channel is such that all outputs are the maximally mixed state, and thus have p-norm equal to $1/|\mathrm{B}|^{1-1/p}$. But other channels might have outputs which are much less mixed, and thus have much larger p-norm. It would therefore seem more accurate to quantify closeness in terms of relative error. Hence, given a CPTP map $\mathcal{N} : \mathcal{L}(\mathrm{A}) \to \mathcal{L}(\mathrm{B})$, we would be interested in finding a CPTP map $\widehat{\mathcal{N}} : \mathcal{L}(\mathrm{A}) \to \mathcal{L}(\mathrm{B})$ with Kraus rank as small as possible, and such that

$$\forall\ \varrho \in \mathcal{D}(\mathrm{A}),\ \left\|\widehat{\mathcal{N}}(\varrho) - \mathcal{N}(\varrho)\right\|_p \leqslant \varepsilon \left\|\mathcal{N}(\varrho)\right\|_p . \tag{6}$$

3 Statement of the main results

Our strategy in order to prove an approximation result of the form given by Equation (3) will be to first establish approximation in terms of operator ordering of the outputs for all inputs. Concretely, given a CPTP map $\mathcal{N} : \mathcal{L}(\mathrm{A}) \to \mathcal{L}(\mathrm{B})$, the idea will be to look for a CPTP map $\widehat{\mathcal{N}} : \mathcal{L}(\mathrm{A}) \to \mathcal{L}(\mathrm{B})$ with Kraus rank as small as possible, and such that

$$\forall\ \varrho \in \mathcal{D}(\mathrm{A}),\ (1-\varepsilon)\mathcal{N}(\varrho) \leqslant \widehat{\mathcal{N}}(\varrho) \leqslant (1+\varepsilon)\mathcal{N}(\varrho). \tag{7}$$

The approximation statement that we establish in Theorem 3.1 is not exactly that of Equation (7), but very close to it.

Theorem 3.1. *Fix $0 < \varepsilon < 1$ and let $\mathcal{N} : \mathcal{L}(\mathrm{A}) \to \mathcal{L}(\mathrm{B})$ be a CPTP map with Kraus rank $|\mathrm{E}| \geqslant |\mathrm{A}|, |\mathrm{B}|$. Then, there exists a CP map $\widehat{\mathcal{N}} : \mathcal{L}(\mathrm{A}) \to \mathcal{L}(\mathrm{B})$ with Kraus rank at most $C \max(|\mathrm{A}|, |\mathrm{B}|) \log(|\mathrm{E}|/\varepsilon)/\varepsilon^2$ (where $C > 0$ is a universal constant) and such that*

$$\forall\ \varrho \in \mathcal{D}(\mathrm{A}),\ -\varepsilon\left(\mathcal{N}(\varrho) - \frac{\mathbb{1}}{|\mathrm{B}|}\right) \leqslant \widehat{\mathcal{N}}(\varrho) - \mathcal{N}(\varrho) \leqslant \varepsilon\left(\mathcal{N}(\varrho) + \frac{\mathbb{1}}{|\mathrm{B}|}\right). \tag{8}$$

Remark 3.2. *Note that if $\widehat{\mathcal{N}}$ satisfies Equation (8), then it especially implies that it approximates $\mathcal{N}$ in any Schatten-norm in a sense close to that of Equation (6), namely*

$$\forall\ p \in \mathbf{N},\ \forall\ \varrho \in \mathcal{D}(\mathrm{A}),\ \left\|\widehat{\mathcal{N}}(\varrho) - \mathcal{N}(\varrho)\right\|_p \leqslant \varepsilon\left(\left\|\mathcal{N}(\varrho)\right\|_p + \frac{1}{|\mathrm{B}|^{1-1/p}}\right).$$

In particular, we have the $(1\to1)$-norm approximation of $\mathcal{N}$ by $\widehat{\mathcal{N}}$

$$\forall\ \varrho \in \mathcal{D}(\mathrm{A}),\ \left\|\widehat{\mathcal{N}}(\varrho) - \mathcal{N}(\varrho)\right\|_1 \leqslant 2\varepsilon,$$

in which we can further impose that $\widehat{\mathcal{N}}$ is strictly, and not just up to an error 2ε, trace preserving (see the proof of Theorem 3.1).

The main interest of the approximation statements in Theorem 3.1 and Remark 3.2 is their universality. They indeed establish that any quantum channel can be approximated by one that has Kraus rank of order $\max(|\mathrm{A}|, |\mathrm{B}|)\log(|\mathrm{E}|)$. They of course do not exclude the existence of channels for which a better compression would be possible, but provide an achievable compression in the worst-case scenario.

One important question at this point is the one of optimality in the above results. A first obvious observation to make in order to answer it is the following: if a CP map has Kraus rank s, then it necessarily sends rank 1 inputs to output states of rank at most s. This is of course informative only if s is smaller than the output space dimension. But as

we shall see, having this in mind will be useful to prove that certain channels cannot be compressed further than as guaranteed by Theorem 3.1.

Our constructions will be based on the existence of so-called tight normalized frames. Namely, for any $N, d \in \mathbf{N}$ with $N \geqslant d$, there exist unit vectors $|\psi_1\rangle, \ldots, |\psi_N\rangle$ in $\mathbf{C}^d$ such that

$$\frac{1}{N}\sum_{k=1}^{N} |\psi_k\rangle\langle\psi_k| = \frac{\mathbb{1}}{d}.$$

Denoting by $\{|j\rangle,\ 1 \leqslant j \leqslant d\}$ an orthonormal basis of $\mathbf{C}^d$, a possible way of constructing such vectors is e.g. to make the choice

$$\forall\ 1 \leqslant k \leqslant N,\ |\psi_k\rangle = \frac{1}{\sqrt{d}}\sum_{j=1}^{d} e^{2i\pi jk/N}|j\rangle. \tag{9}$$

Note that if this so, then any basis vector $|j\rangle$, $1 \leqslant j \leqslant d$, is such that, for each $1 \leqslant k \leqslant N$, $|\langle\psi_k|j\rangle|^2 = 1/d$.

Let us now come back to our objective. What we want to exhibit here are CPTP maps $\mathcal{N} : \mathcal{L}(\mathrm{A}) \to \mathcal{L}(\mathrm{B})$ with either one or the other of the following two properties: if a CP map $\widehat{\mathcal{N}} : \mathcal{L}(\mathrm{A}) \to \mathcal{L}(\mathrm{B})$ satisfies

$$\forall\ R \in \mathcal{H}_+(\mathrm{A}),\ (1-\varepsilon)\mathcal{N}(R) - \varepsilon(\operatorname{Tr} R)\frac{\mathbb{1}}{|\mathrm{B}|} \leqslant \widehat{\mathcal{N}}(R) \leqslant (1+\varepsilon)\mathcal{N}(R) + \varepsilon(\operatorname{Tr} R)\frac{\mathbb{1}}{|\mathrm{B}|}, \tag{10}$$

then it necessarily has to be such that either $r_K(\widehat{\mathcal{N}}) \geqslant |\mathrm{A}|$ or $r_K(\widehat{\mathcal{N}}) \geqslant |\mathrm{B}|$. Besides, note that the CP maps $\mathcal{N}, \widehat{\mathcal{N}}$ fulfilling condition (10) above is equivalent to the conjugate CP maps $\mathcal{N}^*, \widehat{\mathcal{N}}^*$ fulfilling condition (11) below

$$\forall\ R \in \mathcal{H}_+(\mathrm{B}),\ (1-\varepsilon)\mathcal{N}^*(R) - \varepsilon(\operatorname{Tr} R)\frac{\mathbb{1}}{|\mathrm{B}|} \leqslant \widehat{\mathcal{N}}^*(R) \leqslant (1+\varepsilon)\mathcal{N}^*(R) + \varepsilon(\operatorname{Tr} R)\frac{\mathbb{1}}{|\mathrm{B}|}. \tag{11}$$

Indeed, we have the following chain of equivalences

$$\begin{aligned}
&\forall\ R \in \mathcal{H}_+(\mathrm{A}),\ \widehat{\mathcal{N}}(R) \leqslant (1+\varepsilon)\mathcal{N}(R) + \varepsilon(\operatorname{Tr} R)\frac{\mathbb{1}}{|\mathrm{B}|} \\
\Leftrightarrow\ &\forall\ R \in \mathcal{H}_+(\mathrm{A}), S \in \mathcal{H}_+(\mathrm{B}),\ \operatorname{Tr}\left(\widehat{\mathcal{N}}(R)S\right) \leqslant (1+\varepsilon)\operatorname{Tr}\left(\mathcal{N}(R)S\right) + \varepsilon(\operatorname{Tr} R)\operatorname{Tr}\left(\frac{\mathbb{1}}{|\mathrm{B}|}S\right) \\
\Leftrightarrow\ &\forall\ R \in \mathcal{H}_+(\mathrm{A}), S \in \mathcal{H}_+(\mathrm{B}),\ \operatorname{Tr}\left(R\widehat{\mathcal{N}}^*(S)\right) \leqslant (1+\varepsilon)\operatorname{Tr}\left(R\mathcal{N}^*(S)\right) + \varepsilon(\operatorname{Tr} S)\operatorname{Tr}\left(\frac{\mathbb{1}}{|\mathrm{B}|}R\right) \\
\Leftrightarrow\ &\forall\ S \in \mathcal{H}_+(\mathrm{B}),\ \widehat{\mathcal{N}}^*(S) \leqslant (1+\varepsilon)\mathcal{N}^*(S) + \varepsilon(\operatorname{Tr} S)\frac{\mathbb{1}}{|\mathrm{B}|},
\end{aligned}$$

where the first and third equivalences are by the characterization of positive semidefinite matrices, and the second equivalence is by definition of conjugate maps. And similarly for the other inequality. Depending on what we want to establish, it will be more convenient to work with either requirement (10) or requirement (11)

Assume first of all that $|\mathrm{B}| \geqslant |\mathrm{A}|$, and consider $\mathcal{M} : \mathcal{L}(\mathrm{A}) \to \mathcal{L}(\mathrm{B})$ a so-called quantum-classical channel (aka measurement). More specifically, define the CPTP map

$$\mathcal{M} : X \in \mathcal{L}(\mathrm{A}) \mapsto \frac{|\mathrm{A}|}{|\mathrm{B}|}\sum_{i=1}^{|\mathrm{B}|}\langle\psi_i|X|\psi_i\rangle|x_i\rangle\langle x_i| \in \mathcal{L}(\mathrm{B}), \tag{12}$$

where $\{|x_i\rangle,\ 1 \leqslant i \leqslant |\mathrm{B}|\}$ is an orthonormal basis of B and $|\psi_1\rangle, \ldots, |\psi_{|\mathrm{B}|}\rangle$ are unit vectors of A, defined in terms of an orthonormal basis $\{|j\rangle,\ 1 \leqslant j \leqslant |\mathrm{A}|\}$ of A as by Equation (9).

Note that this tight normalized frame assumption implies that $\{(|\mathrm{A}|/|\mathrm{B}|)|\psi_i\rangle\langle\psi_i|\}_{1\leqslant i\leqslant|\mathrm{B}|}$ forms a rank-1 POVM on A (hence a posteriori the justification of the denomination for $\mathcal{M}$). Setting, for each $1 \leqslant i \leqslant |\mathrm{B}|$, $K_i = \sqrt{|\mathrm{A}|/|\mathrm{B}|}\,|x_i\rangle\langle\psi_i|$, we can clearly re-write $\mathcal{M} : X \in \mathcal{L}(\mathrm{A}) \mapsto \sum_{i=1}^{|\mathrm{B}|} K_i X K_i^\dagger \in \mathcal{L}(\mathrm{B})$, so $r_K(\mathcal{M}) \leqslant |\mathrm{B}|$. And what we actually want to show is that it is even impossible to approximate $\mathcal{M}$ in the sense of Theorem 3.1 with strictly less than $|\mathrm{B}|$ Kraus operators. Observe that by construction, say, $|1\rangle$ is such that, for each $1 \leqslant i \leqslant |\mathrm{B}|$, $|\langle\psi_i|1\rangle|^2 = 1/|\mathrm{A}|$, so that $\mathcal{M}(|1\rangle\langle 1|) = \mathbb{1}/|\mathrm{B}|$. Yet, assume that $\widehat{\mathcal{M}} : \mathcal{L}(\mathrm{A}) \to \mathcal{L}(\mathrm{B})$ is a CPTP map such that $\mathcal{M}, \widehat{\mathcal{M}}$ fulfill Equation (10) for some $0 < \varepsilon < 1/2$. Then, the l.h.s. of Equation (10) yields in particular, $\widehat{\mathcal{M}}(|1\rangle\langle 1|) \geqslant (1 - 2\varepsilon)\,\mathbb{1}/|\mathrm{B}|$, so that $\widehat{\mathcal{M}}(|1\rangle\langle 1|)$ has to have full rank. And therefore, it cannot be that $r_K(\widehat{\mathcal{M}}) < |\mathrm{B}|$.

Assume now that $|\mathrm{A}| \geqslant |\mathrm{B}|$, and consider $\mathcal{N} : \mathcal{L}(\mathrm{A}) \to \mathcal{L}(\mathrm{B})$ a so-called classical-quantum channel. More specifically, define the CPTP map

$$\mathcal{N} : X \in \mathcal{L}(\mathrm{A}) \mapsto \sum_{i=1}^{|\mathrm{A}|} \langle x_i|X|x_i\rangle |\psi_i\rangle\langle\psi_i| \in \mathcal{L}(\mathrm{B}), \tag{13}$$

where $\{|x_i\rangle,\ 1 \leqslant i \leqslant |\mathrm{A}|\}$ is an orthonormal basis of A and $|\psi_1\rangle, \ldots, |\psi_{|\mathrm{A}|}\rangle$ are unit vectors in B. Setting, for each $1 \leqslant i \leqslant |\mathrm{A}|$, $K_i = |\psi_i\rangle\langle x_i|$, we can clearly re-write $\mathcal{N} : X \in \mathcal{L}(\mathrm{A}) \mapsto \sum_{i=1}^{|\mathrm{A}|} K_i X K_i^\dagger \in \mathcal{L}(\mathrm{B})$, so $r_K(\mathcal{N}) \leqslant |\mathrm{A}|$. Now, we want to show that, at least for certain choices of $|\psi_1\rangle, \ldots, |\psi_{|\mathrm{A}|}\rangle$, it is even impossible to approximate $\mathcal{N}$ in the sense of Theorem 3.1 with strictly less than $|\mathrm{A}|$ Kraus operators. For that, we impose that they are defined in terms of an orthonormal basis $\{|j\rangle,\ 1 \leqslant j \leqslant |\mathrm{B}|\}$ of B as by Equation (9). Since the conjugate of $\mathcal{N}$ is the CP unital map

$$\mathcal{N}^* : X \in \mathcal{L}(\mathrm{B}) \mapsto \sum_{i=1}^{|\mathrm{A}|} \langle\psi_i|X|\psi_i\rangle |x_i\rangle\langle x_i| \in \mathcal{L}(\mathrm{A}),$$

we have in this case that $\mathcal{M} = (|\mathrm{B}|/|\mathrm{A}|)\mathcal{N}^*$ is precisely of the form (12) (with the roles of A and B switched). Hence, as we already showed, if $\widehat{\mathcal{M}} : \mathcal{L}(\mathrm{B}) \to \mathcal{L}(\mathrm{A})$ is a CPTP map such that $\mathcal{M}, \widehat{\mathcal{M}}$ fulfil Equation (10) (with the roles of A and B switched) for some $0 < \varepsilon < 1/2$, then it cannot be that $r_K(\widehat{\mathcal{M}}) < |\mathrm{A}|$. This means equivalently that if $\widehat{\mathcal{N}}^* : \mathcal{L}(\mathrm{B}) \to \mathcal{L}(\mathrm{A})$ is a CP map such that $\mathcal{N}^*, \widehat{\mathcal{N}}^*$ fulfil Equation (11) for some $0 < \varepsilon < 1/2$, then it cannot be that $r_K(\widehat{\mathcal{N}}) = r_K(\widehat{\mathcal{N}}^*) < |\mathrm{A}|$.

Summarizing, we just established that $n \geqslant \max(|\mathrm{A}|, |\mathrm{B}|)$ is for sure necessary in Theorem 3.1. But it is not clear whether or not the $\log|\mathrm{E}|$ factor can be removed. In the case of "well-behaved" channels, whose range is only composed of sufficiently mixed states, we can answer affirmatively, as an immediate implication of Theorem 3.3 below. However, we leave the question open in general.

Theorem 3.3. *Fix $0 < \varepsilon < 1$ and let $\mathcal{N} : \mathcal{L}(\mathrm{A}) \to \mathcal{L}(\mathrm{B})$ be a CPTP map with Kraus rank $|\mathrm{E}| \geqslant |\mathrm{A}|, |\mathrm{B}|$. Then, there exists a CP map $\widehat{\mathcal{N}} : \mathcal{L}(\mathrm{A}) \to \mathcal{L}(\mathrm{B})$ with Kraus rank at most $C\max(|\mathrm{A}|, |\mathrm{B}|)/\varepsilon^2$ (where $C > 0$ is a universal constant) and such that*

$$\sup_{\varrho\in\mathcal{D}(\mathrm{A})} \left\|\widehat{\mathcal{N}}(\varrho) - \mathcal{N}(\varrho)\right\|_\infty \leqslant \varepsilon \sup_{\varrho\in\mathcal{D}(\mathrm{A})} \|\mathcal{N}(\varrho)\|_\infty .$$

As a straightforward consequence of Theorem 3.3 we get the following: If $\mathcal{N} : \mathcal{L}(\mathrm{A}) \to \mathcal{L}(\mathrm{B})$ is such that $\|\mathcal{N}(\varrho)\|_\infty \leqslant c/|\mathrm{B}|$ for all $\varrho \in \mathcal{D}(\mathrm{A})$, then $\|\widehat{\mathcal{N}}(\varrho) - \mathcal{N}(\varrho)\|_\infty \leqslant c\varepsilon/|\mathrm{B}|$ for all $\varrho \in \mathcal{D}(\mathrm{A})$, and hence $\|\widehat{\mathcal{N}}(\varrho) - \mathcal{N}(\varrho)\|_1 \leqslant c\varepsilon$ for all $\varrho \in \mathcal{D}(\mathrm{A})$. This means that, if $\mathcal{N}$ sends any input state to an output state that has small operator norm, then it can be

approximated by a CP map with Kraus rank of order $\max(|\mathrm{A}|, |\mathrm{B}|)$ in $(1\to\infty)$-norm up to error of order $\varepsilon/|\mathrm{B}|$, and thus a fortiori in $(1\to1)$-norm up to error ε.

Before moving on to the full proof of Theorems 3.1 and 3.3 let us briefly explain the main ideas in it. These two existence results of CPTP maps having some desired properties actually stem from proving that suitably constructed random ones have them with high probability. One thus has to show that for the random CPTP map $\widehat{\mathcal{N}}$ the probability is high that, for every input state ϱ, $\widehat{\mathcal{N}}(\varrho)$ is close to $\mathcal{N}(\varrho)$. This is achieved in two steps: establishing first that this holds for a given input state ϱ and second that it in fact holds for all of them simultaneously. The fact that the individual probability of deviating from average is small is a consequence of the concentration of measure phenomenon in high dimensions. Deriving from there that the global deviation probability is also small is done by discretizing the input set and using a union bound. This line of proof is extremely standard in asymptotic geometric analysis (this is for instance how Dvoretzky's theorem is obtained from Levy's lemma) or in large dimensional probability theory (this is for instance how the supremum of an empirical process is upper bounded through generic chaining). In our case though, the first step requires a careful analysis of the sub-exponential behavior of a certain random variable.

4 Proofs of the main results

As a crucial step in establishing Theorems 3.1 and 3.3, we will need a large deviation inequality for sums of independent ψ_1 (aka sub-exponential) random variables. Recall that the ψ_1-norm of a random variable X (which quantifies the exponential decay of the tail) may be defined via the growth of moments

$$\|X\|_{\psi_1} = \sup_{p\in\mathbf{N}} \frac{\left(\mathbf{E}\,|X|^p\right)^{1/p}}{p}.$$

This definition is more practical than the standard definition through the Orlicz function $x \mapsto e^x - 1$, and leads to an equivalent norm (see [5], Corollary 1.1.6). The large deviation inequality for a sum of independent ψ_1 random variables is known as Bernstein's inequality and is quoted below.

Theorem 4.1 (Bernstein's inequality, see e.g. [5], Theorem 1.2.5)**.** *Let $X_1, \ldots, X_n$ be n independent ψ_1 random variables. Set $M = \max_{1\leqslant i\leqslant n} \|X_i\|_{\psi_1}$ and $\sigma^2 = \sum_{1\leqslant i\leqslant n} \|X_i\|_{\psi_1}^2/n$. Then,*

$$\forall\ t > 0,\ \mathbf{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}(X_i - \mathbf{E}\,X_i)\right| > t\right) \leqslant \exp\left(-c_0 n \min\left(\frac{t^2}{\sigma^2}, \frac{t}{M}\right)\right),$$

where $c_0 > 0$ is a universal constant.

Our application of Bernstein's inequality to a suitably chosen sum of independent ψ_1 random variables will yield Proposition 4.2 below. Note that in the latter, as well as in several other places in the remainder of the paper, we shall use the following shorthand notation, whenever no confusion is at risk: given a unit vector ϕ in $\mathbf{C}^n$, we also denote by ϕ the corresponding pure state $|\phi\rangle\langle\phi|$ on $\mathbf{C}^n$.

Proposition 4.2. *Let $\mathcal{N} : \mathcal{L}(\mathrm{A}) \to \mathcal{L}(\mathrm{B})$ be a CPTP map with Kraus rank $|\mathrm{E}|$, defined by*

$$\forall\ \varrho \in \mathcal{D}(\mathrm{A}),\ \mathcal{N}(\varrho) = \mathrm{Tr}_{\mathrm{E}}\left[V\varrho V^\dagger\right], \tag{14}$$

for some isometry $V : \mathrm{A} \hookrightarrow \mathrm{B} \otimes \mathrm{E}$.

For any given unit vector φ *in* E *define next the CP map* $\mathcal{N}_\varphi : \mathcal{L}(\mathrm{A}) \to \mathcal{L}(\mathrm{B})$ *by*

$$\forall\ \varrho \in \mathcal{D}(\mathrm{A}),\ \mathcal{N}_\varphi(\varrho) = |\mathrm{E}|\ \mathrm{Tr}_{\mathrm{E}}\left[(\mathbb{1} \otimes \varphi)\, V \varrho V^\dagger\, (\mathbb{1} \otimes \varphi)\right]. \tag{15}$$

Now, fix unit vectors x *in* A, y *in* B, *and pick random unit vectors* $\varphi_1, \ldots, \varphi_n$ *in* E, *independently and uniformly. Then,*

$$\forall\ 0 < \varepsilon < 1,\ \mathbf{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} \langle y|\mathcal{N}_{\varphi_i}(x)\,|y\rangle - \langle y|\mathcal{N}(x)\,|y\rangle\right| > \varepsilon \langle y|\mathcal{N}(x)\,|y\rangle\right) \leqslant e^{-cn\varepsilon^2},$$

where $c > 0$ *is a universal constant.*

Note that, by construction, the CP map $\mathcal{N}_\varphi : \mathcal{L}(\mathrm{A}) \to \mathcal{L}(\mathrm{B})$ defined in Equation (15) has Kraus rank 1. Indeed, it has $V_\varphi = \sqrt{|\mathrm{E}|}\,(\mathbb{1} \otimes \varphi)\, V : \mathrm{A} \hookrightarrow \mathrm{B} \otimes \mathrm{span}\{\varphi\}$ as Stinespring embedding, which has effective environment dimension 1.

In order to derive this concentration result, we will need first of all an estimate on the ψ_1-norm of a certain random variable appearing in our construction. This is the content of Lemma 4.3 below.

Lemma 4.3. *Fix* $d, s \in \mathbf{N}$. *Let* σ *be a state on* $\mathbf{C}^d \otimes \mathbf{C}^s$ *and* y *be a unit vector in* $\mathbf{C}^d$. *Next, for* φ *a uniformly distributed unit vector in* $\mathbf{C}^s$ *define the random variable*

$$X_\varphi(\sigma, y) = \mathrm{Tr}\left[y \otimes \varphi\, \sigma\right].$$

Then, $X_\varphi(\sigma, y)$ *is a* ψ_1 *random variable with mean and* ψ_1*-norm satisfying*

$$\mathbf{E}\, X_\varphi(\sigma, y) = \frac{1}{s}\,\mathrm{Tr}\left[y \otimes \mathbb{1}\, \sigma\right] \quad \textit{and} \quad \|X_\varphi(\sigma, y)\|_{\psi_1} \leqslant \frac{1}{s}\,\mathrm{Tr}\left[y \otimes \mathbb{1}\, \sigma\right]. \tag{16}$$

Proof. To begin with, recall that, for any $p \in \mathbf{N}$, we have, for φ a uniformly distributed unit vector in $\mathbf{C}^s$,

$$\mathbf{E}\, \varphi^{\otimes p} = \frac{1}{\binom{s+p-1}{p}} P_{\mathrm{Sym}^p(\mathbf{C}^s)},$$

where $P_{\mathrm{Sym}^p(\mathbf{C}^s)}$ denotes the orthogonal projector onto the completely symmetric subspace of $(\mathbf{C}^s)^{\otimes p}$. Indeed, $\mathbf{E}\, \varphi^{\otimes p}$ commutes with all $U^{\otimes p}$, for $U \in \mathcal{U}(\mathbf{C}^s)$, so by Schur's lemma it has to be proportional to $P_{\mathrm{Sym}^p(\mathbf{C}^s)}$, and the normalization is given by observing that $\mathrm{Tr}(\mathbf{E}\, \varphi^{\otimes p}) = 1$, together with the fact that $\dim(\mathrm{Sym}^p(\mathbf{C}^s)) = \binom{s+p-1}{p}$.

Now, setting $\sigma_y = \mathrm{Tr}_{\mathbf{C}^d}\left[y \otimes \mathbb{1}\, \sigma\right]$, positive sub-normalized operator on $\mathbf{C}^s$, we see that $X_\varphi(\sigma, y) = \mathrm{Tr}\left[\varphi\, \sigma_y\right]$. Hence, we clearly have first of all the first statement in Equation (16), namely

$$\mathbf{E}\, X_\varphi(\sigma, y) = \frac{1}{s}\,\mathrm{Tr}\left[\mathbb{1}\, \sigma_y\right] = \frac{1}{s}\,\mathrm{Tr}\left[y \otimes \mathbb{1}\, \sigma\right].$$

What is more, for any $p \in \mathbf{N}$, $|X_\varphi(\sigma, y)|^p = (\mathrm{Tr}\left[\varphi\, \sigma_y\right])^p = \mathrm{Tr}\left[\varphi^{\otimes p}\, \sigma_y^{\otimes p}\right]$. And therefore,

$$\mathbf{E}\, |X_\varphi(\sigma, y)|^p = \frac{1}{\binom{s+p-1}{p}}\,\mathrm{Tr}\left[P_{\mathrm{Sym}^p(\mathbf{C}^s)} \sigma_y^{\otimes p}\right] \leqslant \frac{1}{\binom{s+p-1}{p}}\,\mathrm{Tr}\left[\sigma_y^{\otimes p}\right] \leqslant \left(\frac{p}{s}\,\mathrm{Tr}\left[\sigma_y\right]\right)^p,$$

where the last inequality is simply by the rough bounds $p! \leqslant p^p$ and $(s+p-1)!/(s-1)! \geqslant s^p$.

So in the end, we get as wanted the second statement in Equation (16), namely

$$\|X_\varphi(\sigma, y)\|_{\psi_1} = \sup_{p \in \mathbf{N}} \frac{\left(\mathbf{E}\, |X_\varphi(\sigma, y)|^p\right)^{1/p}}{p} \leqslant \frac{1}{s}\,\mathrm{Tr}\left[y \otimes \mathbb{1}\, \sigma\right].$$

This concludes the proof of Lemma 4.3. □

Proof of Proposition 4.2. Note first of all that we can obviously re-write

$$\langle y|\mathcal{N}(x)|y\rangle = \mathrm{Tr}\left[y \otimes \mathbb{1}\, VxV^\dagger\right] \text{ and } \forall\ \varphi \in S_{\mathrm{E}},\ \langle y|\mathcal{N}_\varphi(x)|y\rangle = |\mathrm{E}|\ \mathrm{Tr}\left[y \otimes \varphi\, VxV^\dagger\right].$$

Next, for each $1 \leqslant i \leqslant n$, define the random variable $Y_i = \langle y|\mathcal{N}_{\varphi_i}(x)|y\rangle$. By Lemma 4.3, combined with the observation just made above, we know that these are independent ψ_1 random variables with mean $\langle y|\mathcal{N}(x)|y\rangle$ and ψ_1-norm upper bounded by $\langle y|\mathcal{N}(x)|y\rangle$. So by Bernstein's inequality, recalled as Theorem 4.1, we get that

$$\forall\, t > 0,\ \mathbf{P}\left(\left|\frac{1}{n}\sum_{i=1}^n Y_i - \langle y|\mathcal{N}(x)|y\rangle\right| > t\right) \leqslant \exp\left(-c_0 n \min\left(\frac{t^2}{\langle y|\mathcal{N}(x)|y\rangle^2}, \frac{t}{\langle y|\mathcal{N}(x)|y\rangle}\right)\right),$$

where $c_0 > 0$ is a universal constant. And hence,

$$\forall\ 0 < \varepsilon < 1,\ \mathbf{P}\left(\left|\frac{1}{n}\sum_{i=1}^n Y_i - \langle y|\mathcal{N}(x)|y\rangle\right| > \varepsilon\langle y|\mathcal{N}(x)|y\rangle\right) \leqslant e^{-c_0 n\varepsilon^2},$$

which is precisely the result announced in Proposition 4.2. □

Having at hand the "fixed x, y" concentration inequality of Proposition 4.2, we can now get its "for all x, y" counterparts by a standard net-argument. This appears as the following Propositions 4.4 and 4.5.

Proposition 4.4. *Let $\mathcal{N} : \mathcal{L}(\mathrm{A}) \to \mathcal{L}(\mathrm{B})$ be a CPTP map, as characterized by Equation (14), and for each unit vector φ in E define the CP map $\mathcal{N}_\varphi : \mathcal{L}(\mathrm{A}) \to \mathcal{L}(\mathrm{B})$ as in Equation (15). Next, for $\varphi_1, \ldots, \varphi_n$ independent uniformly distributed unit vectors in E, set $\mathcal{N}_{\varphi^{(n)}} = \left(\sum_{i=1}^n \mathcal{N}_{\varphi_i}\right)/n$. Then, for any $0 < \varepsilon < 1$,*

$$\mathbf{P}\left(\forall\ x \in S_{\mathrm{A}}, y \in S_{\mathrm{B}},\ \left|\langle y|\mathcal{N}_{\varphi^{(n)}}(x) - \mathcal{N}(x)|y\rangle\right| \leqslant \varepsilon\langle y|\mathcal{N}(x)|y\rangle + \frac{\varepsilon}{|\mathrm{B}|}\right)$$
$$\geqslant 1 - \left(\frac{24|\mathrm{E}||\mathrm{B}|}{\varepsilon}\right)^{2(|\mathrm{A}|+|\mathrm{B}|)} e^{-cn\varepsilon^2},$$

where $c > 0$ is a universal constant.

Note that, by construction, the random CP map $\mathcal{N}_{\varphi^{(n)}} : \mathcal{L}(\mathrm{A}) \to \mathcal{L}(\mathrm{B})$ introduced above has Kraus rank at most n. Indeed, it is a convex combination of n Kraus rank 1 random CP maps $\mathcal{N}_{\varphi_i} : \mathcal{L}(\mathrm{A}) \to \mathcal{L}(\mathrm{B})$, $1 \leqslant i \leqslant n$.

Proof. Fix $0 < \alpha, \beta < 1$ and consider $\mathcal{A}_\alpha, \mathcal{B}_\beta$ minimal α, β-nets within the unit spheres of A, B, so that by a standard volumetric argument $|\mathcal{A}_\alpha| \leqslant (3/\alpha)^{2|\mathrm{A}|}, |\mathcal{B}_\beta| \leqslant (3/\beta)^{2|\mathrm{B}|}$ (see e.g. [10], Chapter 4). Then, by Proposition 4.2 and the union bound, we get that, for any $\varepsilon > 0$,

$$\mathbf{P}\left(\forall\ x \in \mathcal{A}_\alpha, y \in \mathcal{B}_\beta,\ \left|\langle y|\mathcal{N}_{\varphi^{(n)}}(x) - \mathcal{N}(x)|y\rangle\right| \leqslant \varepsilon\langle y|\mathcal{N}\left(x\right)|y\rangle\right)$$
$$\geqslant 1 - \left(\frac{3}{\alpha}\right)^{2|\mathrm{A}|}\left(\frac{3}{\beta}\right)^{2|\mathrm{B}|} e^{-cn\varepsilon^2}. \tag{17}$$

Now, fix $\varepsilon > 0$ and suppose that $\mathcal{E} : \mathcal{L}(\mathrm{A}) \to \mathcal{L}(\mathrm{B})$ is a Hermiticity-preserving map which is such that

$$\forall\ x \in \mathcal{A}_\alpha,\ \forall\ y \in \mathcal{B}_\beta,\ |\langle y|\mathcal{E}(x)|y\rangle| \leqslant \varepsilon\langle y|\mathcal{N}\left(x\right)|y\rangle. \tag{18}$$

Assume that $\mathcal{E}$ additionally satisfies the boundedness property

$$\forall\ x \in S_{\mathrm{A}},\ \forall\ y \in S_{\mathrm{B}},\ |\langle y|\mathcal{E}(x)|y\rangle| \leqslant |\mathrm{E}|. \tag{19}$$

Note that if $\mathcal{E}$ is Hermicity-preserving, then

$$\begin{aligned}
\forall\ y \in S_{\mathrm{B}},\quad \sup_{v,v' \in S_{\mathrm{A}}} |\langle y|\mathcal{E}(|v\rangle\langle v'|)|y\rangle| &= \sup_{v \in S_{\mathrm{A}}} |\langle y|\mathcal{E}(|v\rangle\langle v|)|y\rangle|, \\
\forall\ x \in S_{\mathrm{A}},\quad \sup_{w,w' \in S_{\mathrm{B}}} |\langle w|\mathcal{E}(|x\rangle\langle x|)|w'\rangle| &= \sup_{w \in S_{\mathrm{B}}} |\langle w|\mathcal{E}(|x\rangle\langle x|)|w\rangle|.
\end{aligned}$$

Indeed, this is because for any X, $\mathcal{E}(X^\dagger) = \mathcal{E}(X)^\dagger$. Hence, it will be useful to us later on to keep in mind that assumption (19) is actually equivalent to

$$\forall\ x, x' \in S_{\mathrm{A}},\ \forall\ y, y' \in S_{\mathrm{B}},\ \begin{cases} |\langle y|\mathcal{E}(|x\rangle\langle x'|)|y\rangle| \leqslant |\mathrm{E}| \\ |\langle y|\mathcal{E}(|x\rangle\langle x|)|y'\rangle| \leqslant |\mathrm{E}| \end{cases}.$$

Then, for any unit vectors $x \in S_{\mathrm{A}}$, $y \in S_{\mathrm{B}}$, we know by definition that there exist $\tilde{x} \in \mathcal{A}_\alpha$, $\tilde{y} \in \mathcal{B}_\beta$ such that $\|x - \tilde{x}\| \leqslant \alpha$, $\|y - \tilde{y}\| \leqslant \beta$. Hence, first of all

$$\begin{aligned}
|\langle y|\mathcal{E}(|x\rangle\langle x|)|y\rangle| &\leqslant |\langle \tilde{y}|\mathcal{E}(|x\rangle\langle x|)|\tilde{y}\rangle| + |\langle y - \tilde{y}|\mathcal{E}(|x\rangle\langle x|)|\tilde{y}\rangle| + |\langle y|\mathcal{E}(|x\rangle\langle x|)|y - \tilde{y}\rangle| \\
&\leqslant |\langle \tilde{y}|\mathcal{E}(|x\rangle\langle x|)|\tilde{y}\rangle| + 2\beta|\mathrm{E}|,
\end{aligned}$$

where the second inequality follows from the boundedness property (19) of $\mathcal{E}$, combined with the fact that $\|y - \tilde{y}\| \leqslant \beta$. Then similarly, because $\|x - \tilde{x}\| \leqslant \alpha$,

$$\begin{aligned}
|\langle \tilde{y}|\mathcal{E}(|x\rangle\langle x|)|\tilde{y}\rangle| &\leqslant |\langle \tilde{y}|\mathcal{E}(|\tilde{x}\rangle\langle \tilde{x}|)|\tilde{y}\rangle| + |\langle \tilde{y}|\mathcal{E}(|x - \tilde{x}\rangle\langle \tilde{x}|)|\tilde{y}\rangle| + |\langle \tilde{y}|\mathcal{E}(|x\rangle\langle x - \tilde{x}|)|\tilde{y}\rangle| \\
&\leqslant |\langle \tilde{y}|\mathcal{E}(|\tilde{x}\rangle\langle \tilde{x}|)|\tilde{y}\rangle| + 2\alpha|\mathrm{E}|.
\end{aligned}$$

Putting together the two previous upper bounds, we see that we actually have

$$|\langle y|\mathcal{E}(|x\rangle\langle x|)|y\rangle| \leqslant |\langle \tilde{y}|\mathcal{E}(|\tilde{x}\rangle\langle \tilde{x}|)|\tilde{y}\rangle| + 2|\mathrm{E}|(\alpha + \beta) \leqslant \varepsilon\langle \tilde{y}|\mathcal{N}(|\tilde{x}\rangle\langle \tilde{x}|)|\tilde{y}\rangle + 2|\mathrm{E}|(\alpha + \beta),$$

where the second inequality is by assumption (18) on $\mathcal{E}$. Now, arguing just as before (using this time that $\mathcal{N}$ satisfies the boundedness property $|\langle y|\mathcal{N}(|x\rangle\langle x'|)|y'\rangle| \leqslant 1$ for any $x, x' \in S_{\mathrm{A}}$ and $y, y' \in S_{\mathrm{B}}$), we get

$$\langle \tilde{y}|\mathcal{N}(|\tilde{x}\rangle\langle \tilde{x}|)|\tilde{y}\rangle \leqslant \langle y|\mathcal{N}(|x\rangle\langle x|)|y\rangle + 2(\alpha + \beta).$$

So eventually, what we obtain is

$$|\langle y|\mathcal{E}(x)|y\rangle| \leqslant \varepsilon(\langle y|\mathcal{N}(x)|y\rangle + 2(\alpha + \beta)) + 2|\mathrm{E}|(\alpha + \beta) \leqslant \varepsilon\langle y|\mathcal{N}(x)|y\rangle + 4|\mathrm{E}|(\alpha + \beta).$$

Therefore, choosing $\alpha = \beta = \varepsilon/(8|\mathrm{E}||\mathrm{B}|)$ (and observing that, by the way $\mathcal{N}_{\varphi^{(n)}}$ is constructed, $\mathcal{N}_{\varphi^{(n)}} - \mathcal{N}$ fulfills condition (19)), it follows from Equation (17) that, for any $\varepsilon > 0$,

$$\begin{aligned}
&\mathbf{P}\left(\forall\ x \in S_{\mathrm{A}}, y \in S_{\mathrm{B}},\ \left|\langle y|\mathcal{N}_{\varphi^{(n)}}(x) - \mathcal{N}(x)|y\rangle\right| \leqslant \varepsilon\langle y|\mathcal{N}\left(x\right)|y\rangle + \frac{\varepsilon}{|\mathrm{B}|}\right) \\
&\qquad \geqslant 1 - \left(\frac{24|\mathrm{E}||\mathrm{B}|}{\varepsilon}\right)^{2(|\mathrm{A}|+|\mathrm{B}|)} e^{-cn\varepsilon^2},
\end{aligned}$$

which is exactly what we wanted to show. □

Proposition 4.5. *Let $\mathcal{N} : \mathcal{L}(\mathrm{A}) \to \mathcal{L}(\mathrm{B})$ be a CPTP map, as characterized by Equation (14), and for each unit vector φ in E define the CP map $\mathcal{N}_\varphi : \mathcal{L}(\mathrm{A}) \to \mathcal{L}(\mathrm{B})$ as in Equation (15). Next, for $\varphi_1, \ldots, \varphi_n$ independent uniformly distributed unit vectors in E, set $\mathcal{N}_{\varphi^{(n)}} = \big(\sum_{i=1}^n \mathcal{N}_{\varphi_i}\big)/n$. Then, for any $0 < \varepsilon < 1$,*

$$\mathbf{P}\left(\sup_{x\in S_\mathrm{A}, y\in S_\mathrm{B}} \left|\langle y|\mathcal{N}_{\varphi^{(n)}}(x) - \mathcal{N}(x)|y\rangle\right| \leqslant \varepsilon \sup_{x\in S_\mathrm{A}, y\in S_\mathrm{B}} \langle y|\mathcal{N}(x)|y\rangle\right) \geqslant 1 - 225^{|\mathrm{A}|+|\mathrm{B}|} e^{-cn\varepsilon^2},$$

where $c > 0$ is a universal constant.

Proof. We will argue in a way very similar to what was done in the proof of Proposition 4.4, and hence skip some of the details here. Again, fix $0 < \alpha, \beta < 1/4$ and consider $\mathcal{A}_\alpha, \mathcal{B}_\beta$ minimal α, β-nets within the unit spheres of A, B. Now, fix $\varepsilon > 0$ and suppose that $\mathcal{E} : \mathcal{L}(\mathrm{A}) \to \mathcal{L}(\mathrm{B})$ is a Hermiticity-preserving map which is such that,

$$\forall\ x \in \mathcal{A}_\alpha,\ \forall\ y \in \mathcal{B}_\beta,\ |\langle y|\mathcal{E}(x)|y\rangle| \leqslant \varepsilon\langle y|\mathcal{N}(x)|y\rangle.$$

Then, for any unit vectors $x \in S_\mathrm{A}$, $y \in S_\mathrm{B}$,

$$\begin{aligned}
&|\langle y|\mathcal{E}(|x\rangle\langle x|)|y\rangle| \\
&\quad\leqslant \varepsilon\big(\langle y|\mathcal{N}(|x\rangle\langle x|)|y\rangle + 2\alpha \sup_{v,v'}\langle y|\mathcal{N}(|v\rangle\langle v'|)|y\rangle + 2\beta \sup_{w,w'}\langle w|\mathcal{N}(|\tilde{x}\rangle\langle\tilde{x}|)|w'\rangle\big) \\
&\qquad + 2\alpha \sup_{v,v'}\left|\langle\tilde{y}|\mathcal{E}(|v\rangle\langle v'|)|\tilde{y}\rangle\right| + 2\beta \sup_{w,w'}\left|\langle w|\mathcal{E}(|x\rangle\langle x|)|w'\rangle\right|,
\end{aligned}$$

where $\tilde{x} \in \mathcal{A}_\alpha$, $\tilde{y} \in \mathcal{B}_\beta$ are such that $\|x - \tilde{x}\| \leqslant \alpha$, $\|y - \tilde{y}\| \leqslant \beta$. And consequently, taking supremum over unit vectors $x \in S_\mathrm{A}$, $y \in S_\mathrm{B}$, we get

$$\sup_{x,y}|\langle y|\mathcal{E}(x)|y\rangle| \leqslant \varepsilon(1 + 2(\alpha+\beta))\sup_{x,y}\langle y|\mathcal{N}(x)|y\rangle + 2(\alpha+\beta)\sup_{x,y}|\langle y|\mathcal{E}(x)|y\rangle|,$$

that is equivalently,

$$\sup_{x,y}|\langle y|\mathcal{E}(x)|y\rangle| \leqslant \varepsilon\,\frac{1 + 2(\alpha+\beta)}{1 - 2(\alpha+\beta)}\sup_{x,y}\langle y|\mathcal{N}(x)|y\rangle.$$

Therefore, choosing $\alpha = \beta = 1/5$, so that $(1 + 2(\alpha+\beta))/(1 - 2(\alpha+\beta)) = 9$ and $3/\alpha = 3/\beta = 15$, we eventually obtain that, for any $0 < \varepsilon < 1$,

$$\mathbf{P}\left(\sup_{x\in S_\mathrm{A}, y\in S_\mathrm{B}} \left|\langle y|\mathcal{N}_{\varphi^{(n)}}(x) - \mathcal{N}(x)|y\rangle\right| \leqslant 9\varepsilon \sup_{x\in S_\mathrm{A}, y\in S_\mathrm{B}} \langle y|\mathcal{N}(x)|y\rangle\right) \geqslant 1 - 15^{2(|\mathrm{A}|+|\mathrm{B}|)} e^{-cn\varepsilon^2},$$

which, after relabelling 9ε in ε, implies precisely the result announced in Proposition 4.5. □

Proof of Theorem 3.1. Because operator-ordering is preserved by convex combinations, it follows from Proposition 4.4 that there exists a universal constant $c > 0$ such that, for any $\varepsilon > 0$,

$$\begin{aligned}
\mathbf{P}\left(\forall\ \rho \in \mathcal{D}(\mathrm{A}),\ -\varepsilon\left(\mathcal{N}(\rho) + \frac{\mathbb{1}}{|\mathrm{B}|}\right) \leqslant \mathcal{N}_{\varphi^{(n)}}(\rho) - \mathcal{N}(\rho) \leqslant \varepsilon\left(\mathcal{N}(\rho) + \frac{\mathbb{1}}{|\mathrm{B}|}\right)\right) \\
\geqslant 1 - \left(\frac{24|\mathrm{E}||\mathrm{B}|}{\varepsilon}\right)^{2(|\mathrm{A}|+|\mathrm{B}|)} e^{-cn\varepsilon^2}.
\end{aligned}$$

The r.h.s. of the latter inequality becomes larger than, say, $1/2$ as soon as n is larger than $C\max(|\mathrm{A}|, |\mathrm{B}|)\log(|\mathrm{E}|/\varepsilon)/\varepsilon^2$, for $C > 0$ some universal constant.

Recapitulating, what we have shown so far is that there exists a completely positive map $\mathcal{N}^{(n)}$ with Kraus rank $n \leqslant C \max(|\mathrm{A}|, |\mathrm{B}|) \log(|\mathrm{E}|/\varepsilon)/\varepsilon^2$, for $C > 0$ some universal constant, such that,

$$\forall\ \rho \in \mathcal{D}(\mathrm{A}),\ -\varepsilon \left(\mathcal{N}(\rho) + \frac{\mathbb{1}}{|\mathrm{B}|} \right) \leqslant \mathcal{N}^{(n)}(\rho) - \mathcal{N}(\rho) \leqslant \varepsilon \left(\mathcal{N}(\rho) + \frac{\mathbb{1}}{|\mathrm{B}|} \right). \tag{20}$$

In particular, Equation (20) implies that, for any $\rho \in \mathcal{D}(\mathrm{A})$, $\left| \mathrm{Tr}\left(\mathcal{N}^{(n)}(\rho) \right) - 1 \right| \leqslant 2\varepsilon$, so that $\mathcal{N}^{(n)}$ is almost trace preserving, up to an error 2ε. As a consequence of Equation (20), we also have

$$\forall\ \rho \in \mathcal{D}(\mathrm{A}),\ \left\| \mathcal{N}^{(n)}(\rho) - \mathcal{N}(\rho) \right\|_1 \leqslant 2\varepsilon, \tag{21}$$

and to get only such trace-norm approximation, it is actually possible to impose that $\mathcal{N}^{(n)}$ is strictly trace preserving. Indeed, denote by $\{K_1, \ldots, K_n\}$ a set of Kraus operators for $\mathcal{N}^{(n)}$, and set $S = \sum_{i=1}^n K_i^\dagger K_i$. Equation (21) guarantees that $\|S - \mathbb{1}\|_\infty \leqslant 2\varepsilon$, so that S is in particular invertible, as soon as $\varepsilon < 1/2$. Hence, consider the completely positive map $\widehat{\mathcal{N}}^{(n)}$ having $\{K_1 S^{-1/2}, \ldots, K_n S^{-1/2}\}$ as a set of Kraus operators, which means that $\widehat{\mathcal{N}}^{(n)}(\cdot) = \mathcal{N}^{(n)}(S^{-1/2} \cdot S^{-1/2})$. The latter is trace preserving by construction, and such that

$$\forall\ \rho \in \mathcal{D}(\mathrm{A}),\ \left\| \widehat{\mathcal{N}}^{(n)}(\rho) - \mathcal{N}(\rho) \right\|_1 \leqslant \left\| \widehat{\mathcal{N}}^{(n)}(\rho) - \mathcal{N}^{(n)}(\rho) \right\|_1 + \left\| \mathcal{N}^{(n)}(\rho) - \mathcal{N}(\rho) \right\|_1 \leqslant 8\varepsilon. \tag{22}$$

Indeed, for any $\rho \in \mathcal{D}(\mathrm{A})$, we have the chain of inequalities

$$\|\rho - S^{1/2}\rho S^{1/2}\|_1 \leqslant \left(\|S^{1/2}\|_\infty + \|\mathbb{1}\|_\infty \right) \|\rho\|_1 \|S^{1/2} - \mathbb{1}\|_\infty \leqslant (1 + \varepsilon + 1)\, 2\varepsilon \leqslant 6\varepsilon,$$

where the first inequality follows from the triangle and Hölder inequalities (after simply noticing that, setting $\Delta = \mathbb{1} - S^{1/2}$, we can rewrite $\rho - S^{1/2}\rho S^{1/2}$ as $\Delta\rho\mathbb{1} + S^{1/2}\rho\Delta$), while the second inequality is because, for any $0 < x < 1/2$, $(1 + 2x)^{1/2} \leqslant 1 + x$ and $(1 - 2x)^{1/2} \geqslant 1 - 2x$, so that $\|S^{1/2}\|_\infty \leqslant 1 + \varepsilon$ and $\|\mathbb{1} - S^{1/2}\|_\infty \leqslant 2\varepsilon$. This implies that, for any $\rho \in \mathcal{D}(\mathrm{A})$,

$$\left\| \widehat{\mathcal{N}}^{(n)}(\rho) - \mathcal{N}^{(n)}(\rho) \right\|_1 = \left\| \widehat{\mathcal{N}}^{(n)}\left(\rho - S^{1/2}\rho S^{1/2} \right) \right\|_1 \leqslant \left\| \rho - S^{1/2}\rho S^{1/2} \right\|_1 \leqslant 6\varepsilon, \tag{23}$$

where the first inequality is because $\widehat{\mathcal{N}}^{(n)}$ is a CPTP map, and hence has $(1 \to 1)$-norm equal to 1. Combining (23) and (21), we get the last inequality in (22).

This concludes the proof of Theorem 3.1 and of Remark 3.2 following it. □

Proof of Theorem 3.3. By extremality of pure states amongst all states, it follows from Proposition 4.5 that there exists a universal constant $c > 0$ such that, for any $\varepsilon > 0$,

$$\mathbf{P}\left(\sup_{\rho \in \mathcal{D}(\mathrm{A})} \left\| \mathcal{N}_{\varphi^{(n)}}(\rho) - \mathcal{N}(\rho) \right\|_\infty \leqslant \varepsilon \sup_{\rho \in \mathcal{D}(\mathrm{A})} \|\mathcal{N}(\rho)\|_\infty \right) \geqslant 1 - 225^{|\mathrm{A}|+|\mathrm{B}|} e^{-cn\varepsilon^2}.$$

The r.h.s. of the latter inequality becomes larger than, say, $1/2$ as soon as n is larger than $C \max(|\mathrm{A}|, |\mathrm{B}|)/\varepsilon^2$, for $C > 0$ some universal constant. And the proof of Theorem 3.3 is thus complete. □

Remark 4.6. *The way we construct a random quantum channel approximating with high probability a quantum channel of interest is by starting from its Stinespring representation.*

Indeed, as briefly explained in the introduction, starting from one of its Kraus representations is a priori not as convenient, because of their non uniqueness. However, our random construction can be rephrased in terms of Kraus operators, in a way that does not depend on the chosen representation. The correspondence is as follows: Assume that our CPTP map $\mathcal{N} : \mathcal{L}(\mathrm{A}) \to \mathcal{L}(\mathrm{B})$ *can be written in the Stinespring and Kraus pictures, respectively, as*

$$\mathcal{N}(X) = \mathrm{Tr}_{\mathrm{E}}\left(VXV^\dagger\right) = \sum_{i=1}^{|\mathrm{E}|} K_i X K_i^\dagger.$$

Then, given a unit vector $\varphi \in \mathrm{E}$*, we define in Equation (15) the CP map* $\mathcal{N}_\varphi : \mathcal{L}(\mathrm{A}) \to \mathcal{L}(\mathrm{B})$ *by*

$$\mathcal{N}_\varphi(X) = \mathrm{Tr}_{\mathrm{E}}\left(V_\varphi X V_\varphi^\dagger\right), \text{ where } V_\varphi = \sqrt{|\mathrm{E}|}\,(\mathbb{1}\otimes\varphi)\,V.$$

Now, the latter can be equivalently defined by

$$\mathcal{N}_\varphi(X) = K_\varphi X K_\varphi^\dagger, \text{ where } K_\varphi = \sqrt{|\mathrm{E}|}\sum_{i=1}^{|\mathrm{E}|}\varphi_i K_i.$$

Hence, sampling Kraus operators at random from the set $\{K_1, \ldots, K_{|\mathrm{E}|}\}$ *does not work in general (and it is not even clear how to chose a representation for which it would), but sampling them as random weighted sums from the set* $\{K_1, \ldots, K_{|\mathrm{E}|}\}$ *does work (whatever the chosen representation).*

This is in contrast with the case of the fully randomizing channel, studied in [8] and [2]. For the latter there is a well-identified distribution to sample Kraus operators from (namely Haar-distributed unitaries), and even a distinguished Kraus decomposition to directly subsample Kraus operators from (namely that built from generalized Pauli shift and phase operators).

5 Consequences and applications

5.1 Approximation in terms of output entropies or fidelities

This section gathers some (more or less straightforward) corollaries of Theorem 3.1 concerning approximation of quantum channels in other distance measures than the (1→1)-norm distance mostly studied up to now.

Given a state ϱ on some Hilbert space H, we define, for any $p \in]1, \infty[$, its Rényi entropy of order p as

$$S_p(\varrho) = -\frac{p}{p-1}\log\|\varrho\|_p,$$

and the latter definition is extended by continuity to $p \in \{1, \infty\}$ as

$$S_1(\varrho) = S(\varrho) = -\mathrm{Tr}(\varrho\log\varrho) \text{ and } S_\infty(\varrho) = -\log\lambda_{\max}(\varrho).$$

Rényi p-entropies thus measure the amount of information present in a quantum state, generalizing the case $p = 1$ of the von Neumann entropy. Besides, given states ρ, σ on some Hilbert space H, their fidelity is defined as $F(\rho,\sigma) = \|\sqrt{\rho}\sqrt{\sigma}\|_1$.

Now, given a channel $\mathcal{N}$, from some input Hilbert space A to some output Hilbert space B, it is important to understand quantities such as its minimum output Rényi p-entropy, i.e.

$$S_p^{\min}(\mathcal{N}) = \min_{\rho\in\mathcal{D}(\mathrm{A})} S_p(\mathcal{N}(\rho)),$$

or its maximum output fidelity with a fixed state σ on B, i.e.

$$F^{\max}(\mathcal{N},\sigma) = \max_{\rho\in\mathcal{D}(\mathrm{A})} F(\mathcal{N}(\rho),\sigma).$$

Indeed, in quantum Shannon theory, these are relevant for their own sake in the asymptotic memoryless setting, while smoothed versions of them show up in the one-shot setting. It is thus of interest to have a channel $\widehat{\mathcal{N}}$ which is less complex than $\mathcal{N}$ but nevertheless shares approximately the same $S_p^{\min}$ and $F^{\max}(\cdot,\sigma)$. What is more, as we will see shortly, channel approximation results in terms of output von Neumann entropy (and perhaps also in terms of other entropies or fidelities) could be the key to establishing channel-dependent lower bounds on the achievable Kraus rank reduction.

Proposition 5.1. *Let $\mathcal{N}:\mathcal{L}(\mathrm{A})\to\mathcal{L}(\mathrm{B})$ be a CPTP map, and assume that the CP map $\widehat{\mathcal{N}}:\mathcal{L}(\mathrm{A})\to\mathcal{L}(\mathrm{B})$ satisfies*

$$\forall\ \varrho\in\mathcal{D}(\mathrm{A}),\ (1-\varepsilon)\mathcal{N}(\varrho)-\varepsilon\frac{\mathbb{1}}{|\mathrm{B}|}\leqslant\widehat{\mathcal{N}}(\varrho)\leqslant(1+\varepsilon)\mathcal{N}(\varrho)+\varepsilon\frac{\mathbb{1}}{|\mathrm{B}|},\tag{24}$$

for some $0<\varepsilon<1/2$. Then, for any $p\in]1,\infty]$, $\widehat{\mathcal{N}}$ is close to $\mathcal{N}$ in terms of output p-entropies, in the sense that

$$\forall\ \varrho\in\mathcal{D}(\mathrm{A}),\ \left|S_p(\widehat{\mathcal{N}}(\varrho))-S_p(\mathcal{N}(\varrho))\right|\leqslant\frac{p}{p-1}4\varepsilon.$$

Proof. Setting $\sigma=\mathcal{N}(\varrho)$, $\widehat{\sigma}=\widehat{\mathcal{N}}(\varrho)$ and $\tau=\mathbb{1}/|\mathrm{B}|$, we can re-write Equation (24) as the two inequalities

$$\widehat{\sigma}\leqslant(1+\varepsilon)\sigma+\varepsilon\tau\ \text{ and }\ \sigma\leqslant\frac{1}{1-\varepsilon}\widehat{\sigma}+\frac{\varepsilon}{1-\varepsilon}\tau\leqslant(1+2\varepsilon)\widehat{\sigma}+2\varepsilon\tau.$$

By operator monotonicity and the triangle inequality for $\|\cdot\|_p$, these imply the two estimates

$$\|\widehat{\sigma}\|_p\leqslant(1+\varepsilon)\|\sigma\|_p+\varepsilon\|\tau\|_p\ \text{ and }\ \|\sigma\|_p\leqslant(1+2\varepsilon)\|\widehat{\sigma}\|_p+2\varepsilon\|\tau\|_p.\tag{25}$$

Now, from the first inequality in Equation (25), we get

$$\log\|\widehat{\sigma}\|_p\leqslant\log\|\sigma\|_p+\log(1+\varepsilon)+\frac{\varepsilon}{1+\varepsilon}\frac{\|\tau\|_p}{\|\sigma\|_p}\leqslant\log\|\sigma\|_p+2\varepsilon,$$

where we used first that log is non-decreasing, then twice that $\log(1+x)\leqslant x$, and finally that $\|\sigma\|_p\leqslant\|\tau\|_p$. Similarly, we derive from the second inequality in Equation (25) that

$$\log\|\sigma\|_p\leqslant\log\|\widehat{\sigma}\|_p+4\varepsilon.$$

Multiplying the two previous inequalities by $-p/(p-1)<0$, we eventually obtain

$$S_p(\widehat{\sigma})\geqslant S_p(\sigma)-\frac{p}{p-1}2\varepsilon\ \text{ and }\ S_p(\sigma)\geqslant S_p(\widehat{\sigma})-\frac{p}{p-1}4\varepsilon.$$

The conclusion of Proposition 5.1 then follows. □

Proposition 5.2. *Let $\mathcal{N}:\mathcal{L}(\mathrm{A})\to\mathcal{L}(\mathrm{B})$ be a CPTP map, and assume that the CP map $\widehat{\mathcal{N}}:\mathcal{L}(\mathrm{A})\to\mathcal{L}(\mathrm{B})$ satisfies*

$$\forall\ \varrho\in\mathcal{D}(\mathrm{A}),\ \left\|\widehat{\mathcal{N}}(\varrho)-\mathcal{N}(\varrho)\right\|_1\leqslant\frac{2\varepsilon}{\log|\mathrm{B}|},\tag{26}$$

for some $0<\varepsilon<1/2$. Then, $\widehat{\mathcal{N}}$ is close to $\mathcal{N}$ in terms of output entropies, in the sense that

$$\forall\ \varrho\in\mathcal{D}(\mathrm{A}),\ \left|S(\widehat{\mathcal{N}}(\varrho))-S(\mathcal{N}(\varrho))\right|\leqslant4\sqrt{\varepsilon}.$$

Proof. By Fannes-Audenaert inequality [3] (see also [12] for a streamlined proof), Equation (26) implies that

$$\left|S(\widehat{\mathcal{N}}(\varrho)) - S(\mathcal{N}(\varrho))\right| \leqslant \varepsilon - \frac{\varepsilon}{\log|\mathrm{B}|}\log\left(\frac{\varepsilon}{\log|\mathrm{B}|}\right) - \left(1 - \frac{\varepsilon}{\log|\mathrm{B}|}\right)\log\left(1 - \frac{\varepsilon}{\log|\mathrm{B}|}\right).$$

Now, for any $0 < x < 1/2$, on the one hand $x\log(1/x) \leqslant \sqrt{x}$, and on the other hand $\log(1/(1-x)) \leqslant \log(1+2x) \leqslant 2x$ so that $(1-x)\log(1/(1-x)) \leqslant 2x$. Hence,

$$\left|S(\widehat{\mathcal{N}}(\varrho)) - S(\mathcal{N}(\varrho))\right| \leqslant \varepsilon + \frac{2\varepsilon}{\log|\mathrm{B}|} + \sqrt{\frac{\varepsilon}{\log|\mathrm{B}|}},$$

and the conclusion of Proposition 5.2 follows. □

Theorem 5.3. *Fix $0 < \varepsilon < 1$ and let $\mathcal{N} : \mathcal{L}(\mathrm{A}) \to \mathcal{L}(\mathrm{B})$ be a CPTP map with Kraus rank $|\mathrm{E}| \geqslant |\mathrm{A}|, |\mathrm{B}|$. Then, there exists a CP map $\widehat{\mathcal{N}} : \mathcal{L}(\mathrm{A}) \to \mathcal{L}(\mathrm{B})$ with Kraus rank at most $C\max(|\mathrm{A}|,|\mathrm{B}|)\log(|\mathrm{E}|/\varepsilon)/\varepsilon^2$ (where $C > 0$ is a universal constant) and such that*

$$\forall\ p \in]1,\infty],\ \forall\ \varrho \in \mathcal{D}(\mathrm{A}),\ \left|S_p(\widehat{\mathcal{N}}(\varrho)) - S_p(\mathcal{N}(\varrho))\right| \leqslant \frac{p}{p-1}\varepsilon.$$

Besides, there also exists a CPTP map $\widehat{\mathcal{N}}' : \mathcal{L}(\mathrm{A}) \to \mathcal{L}(\mathrm{B})$ with Kraus rank at most $C\max(|\mathrm{A}|,|\mathrm{B}|)\log^5(|\mathrm{E}|/\varepsilon^2)/\varepsilon^4$ (where $C > 0$ is a universal constant) and such that

$$\forall\ \varrho \in \mathcal{D}(\mathrm{A}),\ \left|S(\widehat{\mathcal{N}}'(\varrho)) - S(\mathcal{N}(\varrho))\right| \leqslant \varepsilon. \tag{27}$$

Proof. This is a direct consequence of Theorem 3.1, combined with Propositions 5.1 and 5.2. □

We already argued about optimality in Theorem 3.1, showing that there exist CPTP maps $\mathcal{N} : \mathcal{L}(\mathrm{A}) \to \mathcal{L}(\mathrm{B})$ for which at least $|\mathrm{A}|$ or $|\mathrm{B}|$ Kraus operators are needed to approximate them in the sense of Equation (8). We will now establish that, even to get the weaker notion of approximation of Equation (27), a Kraus rank of at least $|\mathrm{A}|$ or $|\mathrm{B}|$ might, in some cases, still be necessary.

Let $\mathcal{N} : X \in \mathcal{L}(\mathrm{A}) \mapsto \mathrm{Tr}_{\mathrm{E}}(VXV^\dagger) \in \mathcal{L}(\mathrm{B})$ be a CPTP map with isometry $V : \mathrm{A} \hookrightarrow \mathrm{B} \otimes \mathrm{E}$. Given $\varrho \in \mathcal{D}(\mathrm{A})$, we consider its input entropy $S(\varrho)$, its output entropy $S(\mathcal{N}(\varrho))$, and its entropy exchange $S(\varrho, \mathcal{N})$. The latter quantity is defined as follows: let $\varphi_{A'A}$ be an extension of ϱ_A, $\widetilde{\varphi}_{A'BE} = (\mathbb{1}_{A'} \otimes V_{A\to BE})\,\varphi_{A'A}$, and set

$$S(\varrho_A, \mathcal{N}_{A\to B}) = S(\mathrm{Tr}_{\mathrm{E}}\,\widetilde{\varphi}_{A'BE}) = S(\mathrm{Tr}_{\mathrm{A'B}}\,\widetilde{\varphi}_{A'BE}).$$

By non-negativity of the loss and the noise of a quantum channel, we then have (see [7], Section 4.5)

$$\forall\ \varrho \in \mathcal{D}(\mathrm{A}),\ |S(\varrho) - S(\mathcal{N}(\varrho))| \leqslant S(\varrho, \mathcal{N}).$$

Yet, for any $\varrho \in \mathcal{D}(\mathrm{A})$, obviously $S(\varrho, \mathcal{N}) \leqslant \log|\mathrm{E}|$. And hence as a consequence,

$$\log|\mathrm{E}| \geqslant \max\{|S(\varrho) - S(\mathcal{N}(\varrho))|\ :\ \varrho \in \mathcal{D}(\mathrm{A})\}.$$

In particular, we may derive the two following lower bounds on $|\mathrm{E}|$, for certain CPTP maps $\mathcal{N}$,

$$\exists\ \psi_{\mathrm{A}} :\ \mathcal{N}(\psi_A) = \frac{\mathbb{1}_B}{|\mathrm{B}|} \implies |\mathrm{E}| \geqslant |\mathrm{B}| \text{ and } \exists\ \psi_B :\ \mathcal{N}\left(\frac{\mathbb{1}_A}{|\mathrm{A}|}\right) = \psi_B \implies |\mathrm{E}| \geqslant |\mathrm{A}|\ . \tag{28}$$

And this remains approximately true for an approximation of $\mathcal{N}$. Concretely, let $\widehat{\mathcal{N}} : \mathcal{L}(\mathrm{A}) \to \mathcal{L}(\mathrm{B})$ be a CPTP map such that

$$\forall\ \varrho \in \mathcal{D}(\mathrm{A}),\ \left|S(\widehat{\mathcal{N}}(\varrho)) - S(\mathcal{N}(\varrho))\right| \leqslant \varepsilon.$$

If $\mathcal{N}$ satisfies the first condition in Equation (28), then

$$S(\widehat{\mathcal{N}}\,(\psi_A)\,) \geqslant S(\mathcal{N}\,(\psi_A)\,) - \varepsilon = \log|\mathrm{B}| - \varepsilon,$$

which implies that

$$\log r_K(\widehat{\mathcal{N}}) \geqslant \log|\mathrm{B}| - \varepsilon,\ \text{i.e. } r_K(\widehat{\mathcal{N}}) \geqslant e^{-\varepsilon}|\mathrm{B}|.$$

And if $\mathcal{N}$ satisfies the second condition in Equation (28), then

$$S\left(\widehat{\mathcal{N}}\left(\frac{\mathbb{1}_A}{|\mathrm{A}|}\right)\right) \leqslant S\left(\mathcal{N}\left(\frac{\mathbb{1}_A}{|\mathrm{A}|}\right)\right) + \varepsilon = \varepsilon,$$

which implies that

$$\log r_K(\widehat{\mathcal{N}}) \geqslant \log|\mathrm{A}| - \varepsilon,\ \text{i.e. } r_K(\widehat{\mathcal{N}}) \geqslant e^{-\varepsilon}|\mathrm{A}|.$$

Hence, the conclusion of this study is that, in Theorem 5.3, $r_K(\widehat{\mathcal{N}}) \geqslant (1-\varepsilon)\max(|\mathrm{A}|,|\mathrm{B}|)$ is for sure necessary, in general, to have the entropy approximation (27). It additionally tells us that there is a channel-dependent lower bound on $r_K(\widehat{\mathcal{N}})$ so that the latter holds (and hence even more so that the stronger notion of approximation in (1→1)-norm holds), namely

$$\log r_K(\widehat{\mathcal{N}}) \geqslant (1-\varepsilon)\max\left\{|S(\varrho) - S(\mathcal{N}(\varrho))|\ :\ \varrho \in \mathcal{D}(\mathrm{A})\right\}. \tag{29}$$

Let us point out though that what we established here is optimality of our results only in the sense that there exist some quantum channels for which $\max(|\mathrm{A}|,|\mathrm{B}|)$ is necessary as approximating Kraus rank. It is however a more subtle question to find, for each given quantum channel $\mathcal{N}$, what is the optimal approximating Kraus rank $\hat{r}_K(\mathcal{N})$. We leave this issue open, but in view of Equation (29), a possible conjecture could be that

$$\log \hat{r}_K(\mathcal{N}) \simeq \max\left\{|S(\varrho) - S(\mathcal{N}(\varrho))|\ :\ \varrho \in \mathcal{D}(\mathrm{A})\right\}.$$

We use the occasion to also formulate an information-theoretic version of this question, namely: we wish to ε-approximate the channel $\mathcal{N}^{\otimes n}$ (for concreteness, say in (1→1)-norm) by one of minimum Kraus rank $\hat{r}_K(\mathcal{N}, n, \varepsilon)$, and we would like to determine the value of

$$R(\mathcal{N}) := \sup_{\varepsilon>0} \limsup_{n\to+\infty} \frac{1}{n}\ \log \hat{r}_K(\mathcal{N}, n, \varepsilon).$$

The latter quantity has the natural operational interpretation, as the minimum rate of qubits needed in the environment, per channel realisation, to approximate many copies of the channel. One could thus hope to get information theoretic lower and upper bounds on it. The above reasoning shows, for instance, that

$$R(\mathcal{N}) \geqslant \max\left\{|S(\varrho) - S(\mathcal{N}(\varrho))|\ :\ \varrho \in \mathcal{D}(\mathrm{A})\right\},$$

but we will not develop this notion further in the present paper.

Proposition 5.4. *Let $\mathcal{N} : \mathcal{L}(\mathrm{A}) \to \mathcal{L}(\mathrm{B})$ be a CPTP map, and assume that the CP map $\widehat{\mathcal{N}} : \mathcal{L}(\mathrm{A}) \to \mathcal{L}(\mathrm{B})$ satisfies*

$$\forall\ \varrho \in \mathcal{D}(\mathrm{A}),\ (1-\varepsilon)\mathcal{N}(\varrho) - \varepsilon\frac{\mathbb{1}}{|\mathrm{B}|} \leqslant \widehat{\mathcal{N}}(\varrho) \leqslant (1+\varepsilon)\mathcal{N}(\varrho) + \varepsilon\frac{\mathbb{1}}{|\mathrm{B}|}, \tag{30}$$

for some $0 < \varepsilon < 1/2$. Then, $\widehat{\mathcal{N}}$ is close to $\mathcal{N}$ in terms of output fidelities, in the sense that

$$\forall\ \varrho \in \mathcal{D}(\mathrm{A}), \forall\ \omega \in \mathcal{D}(\mathrm{B}),\ \left|F\big(\widehat{\mathcal{N}}(\varrho),\omega\big) - F\big(\mathcal{N}(\varrho),\omega\big)\right| \leqslant \frac{3}{\sqrt{2}}\sqrt{\varepsilon}.$$

Proof. As noted in the proof of Proposition 5.1, setting $\sigma = \mathcal{N}(\varrho)$, $\widehat{\sigma} = \widehat{\mathcal{N}}(\varrho)$ and $\tau = \mathbb{1}/|\mathrm{B}|$, we can re-write Equation (30) as the two inequalities $\widehat{\sigma} \leqslant (1+\varepsilon)\sigma + \varepsilon\tau$ and $\sigma \leqslant (1+2\varepsilon)\widehat{\sigma} + 2\varepsilon\tau$. By operator monotonicity of $F(\cdot,\omega)$, and the fact that it is upper bounded by 1, these imply the two estimates

$$F(\widehat{\sigma},\omega) \leqslant \sqrt{1+\varepsilon}F(\sigma,\omega) + \sqrt{\varepsilon}F(\tau,\omega) \leqslant F(\sigma,\omega) + \frac{\varepsilon}{2} + \sqrt{\varepsilon},$$

$$F(\sigma,\omega) \leqslant \sqrt{1+2\varepsilon}F(\widehat{\sigma},\omega) + \sqrt{2\varepsilon}F(\tau,\omega) \leqslant F(\widehat{\sigma},\omega) + \varepsilon + \sqrt{2\varepsilon}.$$

Finally, just observing that $\varepsilon \leqslant \sqrt{\varepsilon/2}$ for $0 < \varepsilon < 1/2$, the conclusion of Proposition 5.4 directly follows. □

Theorem 5.5. *Fix $0 < \varepsilon < 1$ and let $\mathcal{N} : \mathcal{L}(\mathrm{A}) \to \mathcal{L}(\mathrm{B})$ be a CPTP map with Kraus rank $|\mathrm{E}| \geqslant |\mathrm{A}|, |\mathrm{B}|$. Then, there exists a CP map $\widehat{\mathcal{N}} : \mathcal{L}(\mathrm{A}) \to \mathcal{L}(\mathrm{B})$ with Kraus rank at most $C\max(|\mathrm{A}|,|\mathrm{B}|)\log(|\mathrm{E}|/\varepsilon)/\varepsilon^4$ (where $C > 0$ is a universal constant) and such that*

$$\forall\ \varrho \in \mathcal{D}(\mathrm{A}), \forall\ \omega \in \mathcal{D}(\mathrm{B}),\ \left|F\big(\widehat{\mathcal{N}}(\varrho),\omega\big) - F\big(\mathcal{N}(\varrho),\omega\big)\right| \leqslant \varepsilon.$$

Proof. This is a direct consequence of Theorem 3.1, combined with Proposition 5.4. □

5.2 Destruction of correlations with few resources and data hiding.

It was observed in [8], Section 3, that an ε-randomizing channel (i.e. a channel which is an ε-approximation of the fully randomizing channel) approximately destroys the correlations between the system it acts on and any system the latter might be coupled to, in the following two senses: First of all, a state which is initially just classically correlated becomes almost uncorrelated (or in other words any separable state is sent close to a product state, in 1-norm distance). And secondly, whatever the initial state, the correlations present in it become almost invisible to local observers (or in other words any state is sent to close to a product state, in one-way-LOCC-norm). Hence, having an ε-randomizing channel with few Kraus operators can be seen as having an efficient way to decouple a system of interest from its environment. Thanks to Theorem 3.1, we can generalize these results to Theorem 5.6 below.

Theorem 5.6. *Let $\mathrm{A}, \mathrm{B}, \mathrm{C}$ be Hilbert spaces, and assume that $d = \max(|\mathrm{A}|,|\mathrm{B}|) < +\infty$. For any $0 < \varepsilon < 1$ and $\sigma_B^* \in \mathcal{D}(\mathrm{B})$, there exists a CPTP map $\widehat{\mathcal{N}} : \mathcal{L}(\mathrm{A}) \to \mathcal{L}(\mathrm{B})$ with Kraus rank at most $Cd\log(d/\varepsilon)/\varepsilon^2$ (where $C > 0$ is a universal constant) and such that*

$$\forall\ \varrho_{AC} \in \mathcal{S}(\mathrm{A}:\mathrm{C}),\ \left\|\widehat{\mathcal{N}} \otimes \mathcal{I}(\varrho_{AC}) - \sigma_B^* \otimes \varrho_C\right\|_1 \leqslant \varepsilon, \tag{31}$$

$$\forall\ \varrho_{AC} \in \mathcal{D}(\mathrm{A}\otimes\mathrm{C}),\ \left\|\widehat{\mathcal{N}} \otimes \mathcal{I}(\varrho_{AC}) - \sigma_B^* \otimes \varrho_C\right\|_{\mathbf{LOCC}^{\to(\mathrm{B}:\mathrm{C})}} \leqslant \varepsilon. \tag{32}$$

Proof. Define the completely forgetful CPTP map $\mathcal{N} : X_A \in \mathcal{L}(\mathrm{A}) \mapsto (\operatorname{Tr} X_A)\sigma_B^* \in \mathcal{L}(\mathrm{B})$ (i.e. $\mathcal{N}$ sends every input state to the output state σ_B^*). By Theorem 3.1, there exists a CPTP map $\widehat{\mathcal{N}} : \mathcal{L}(\mathrm{A}) \to \mathcal{L}(\mathrm{B})$ with Kraus rank at most $Cd\log(d/\varepsilon)/\varepsilon^2$ such that

$$\forall\ \varrho_A \in \mathcal{D}(\mathrm{A}),\ \left\|\widehat{\mathcal{N}}(\varrho_A) - \mathcal{N}(\varrho_A)\right\|_1 \leqslant \varepsilon \text{ i.e. } \left\|\widehat{\mathcal{N}}(\varrho_A) - \sigma_B^*\right\|_1 \leqslant \varepsilon.$$

Now, following the exact same route as in the proofs of Lemmas III.1 and III.2 in [8], we get that this implies precisely Equations (31) and (32), respectively. We will therefore only briefly recall the arguments here.

Concerning Equation (31), let $\varrho_{AC} \in \mathcal{S}(\mathrm{A} : \mathrm{C})$, i.e. $\varrho_{AC} = \sum_x p_x \varrho_A^{(x)} \otimes \varrho_C^{(x)}$. Then,

$$\begin{aligned}\left\|\widehat{\mathcal{N}} \otimes \mathcal{I}(\varrho_{AC}) - \sigma_B^* \otimes \varrho_C\right\|_1 &= \left\|\sum_x p_x \left(\widehat{\mathcal{N}}\left(\varrho_A^{(x)}\right) - \sigma_B^*\right) \otimes \varrho_C^{(x)}\right\|_1 \\ &\leqslant \sum_x p_x \left\|\widehat{\mathcal{N}}\left(\varrho_A^{(x)}\right) - \sigma_B^*\right\|_1 \\ &\leqslant \varepsilon,\end{aligned}$$

where the last inequality is because, by assumption, for each x, $\|\widehat{\mathcal{N}}(\varrho_A^{(x)}) - \sigma_B^*\|_1 \leqslant \varepsilon$, and $\sum_x p_x = 1$.

As for Equation (32), let $\mathrm{M} = \left(M_B^x \otimes M_C^{(x)}\right)_x \in \mathbf{LOCC}^{\to}(\mathrm{B} : \mathrm{C})$, i.e. for each x, $0 \leqslant M_B^{(x)}, M_C^{(x)} \leqslant \mathbb{1}$, and $\sum_x M_B^{(x)} = \mathbb{1}$. Then, for any $\varrho_{AC} \in \mathcal{D}(\mathrm{A} \otimes \mathrm{C})$,

$$\begin{aligned}&\left\|\widehat{\mathcal{N}} \otimes \mathcal{I}(\varrho_{AC}) - \mathcal{N} \otimes \mathcal{I}(\varrho_{AC})\right\|_{\mathrm{M}} \\ &\quad= \sum_x \left|\operatorname{Tr}\left[M_B^{(x)} \otimes M_C^{(x)} \left(\widehat{\mathcal{N}} \otimes \mathcal{I}(\varrho_{AC}) - \mathcal{N} \otimes \mathcal{I}(\varrho_{AC})\right)\right]\right| \\ &\quad= \sum_x \left|\operatorname{Tr}\left[\left(\widehat{\mathcal{N}}^*\left(M_B^{(x)}\right) - \mathcal{N}^*\left(M_B^{(x)}\right)\right) \otimes M_C^{(x)} \varrho_{AC}\right]\right| \\ &\quad\leqslant \sum_x \left\|\widehat{\mathcal{N}}^*\left(M_B^{(x)}\right) - \mathcal{N}^*\left(M_B^{(x)}\right)\right\|_\infty \\ &\quad\leqslant \varepsilon,\end{aligned}$$

where the next-to-last inequality is because, $\|\varrho_{AC}\|_1 \leqslant 1$ and for each x, $\|M_C^{(x)}\|_\infty \leqslant 1$, while the last inequality is because, by assumption, for each x, $\|\widehat{\mathcal{N}}^*(M_B^{(x)}) - \mathcal{N}^*(M_B^{(x)})\|_\infty \leqslant \varepsilon \operatorname{Tr} M_B^{(x)}/|\mathrm{B}|$, and $\sum_x \operatorname{Tr} M_B^{(x)} = |\mathrm{B}|$. □

Remark 5.7. *Note that the completely forgetful CPTP map*

$$\mathcal{N} : X_A \in \mathcal{L}(\mathrm{A}) \mapsto (\operatorname{Tr} X_A)\sigma_B^* \in \mathcal{L}(\mathrm{B})$$

has Kraus rank equal to $|\mathrm{A}| \times \operatorname{rank}(\sigma_B^*)$*. Hence, the content of Theorem 5.6 is interesting only for states* σ_B^* *of sufficiently high rank, namely* $\operatorname{rank}(\sigma_B^*) \geqslant Cd\log d/|\mathrm{A}|$*.*

In a similar vein, Theorem 3.3 implies that any bipartite state which is sufficiently mixed can be used for data hiding with a bipartite state of much lower rank. In brief, two bipartite states are called data hiding if there exists a global measurement that allows to distinguish them very well, while they remain poorly distinguishable by all local measurements and classical communication. Most known examples of good data hiding states (such as Werner states or random states) are of high rank. Theorem 5.8 below shows how to construct, for any 'truly' high rank state, an associated low rank state such that the pair is data hiding.

Theorem 5.8. *Fix $0 < \varepsilon < 1$ and let $\tau \in \mathcal{D}(\mathrm{A} \otimes \mathrm{A})$ be such that $\|\tau\|_\infty \leqslant C/|\mathrm{A}|^2$ (where $C > 0$ is a universal constant). Then, there exists a state $\hat{\tau} \in \mathcal{D}(\mathrm{A} \otimes \mathrm{A})$, with rank at most $C'|\mathrm{A}|/\varepsilon^2$ (where $C' > 0$ is a universal constant), satisfying*

$$\|\tau - \hat{\tau}\|_{\mathbf{LOCC}^{\rightarrow}(\mathrm{A}:\mathrm{A})} \leqslant C\varepsilon \text{ and } \|\tau - \hat{\tau}\|_1 \geqslant 2\left(1 - \frac{C'C}{\varepsilon^2|\mathrm{A}|}\right). \tag{33}$$

Proof. Let $\mathcal{N} : \mathcal{L}(\mathrm{A}) \to \mathcal{L}(\mathrm{A})$ be the quantum channel whose Choi-Jamiołkowski state is τ. It is easy to check that, since $\|\tau\|_\infty \leqslant C/|\mathrm{A}|^2$, $\mathcal{N}$ is such that

$$\forall\ \varrho \in \mathcal{D}(\mathrm{A}),\ \|\mathcal{N}(\varrho)\|_\infty \leqslant \frac{C}{|\mathrm{A}|}.$$

Indeed, by definition of τ, for any $\varrho \in \mathcal{D}(\mathrm{A})$,

$$\begin{aligned}
\|\mathcal{N}(\varrho)\|_\infty &= \max_{\|X\|_1 \leqslant 1} \mathrm{Tr}(\mathcal{N}(\varrho)X) \\
&= |\mathrm{A}| \times \max_{\|X\|_1 \leqslant 1} \mathrm{Tr}(\tau\, X \otimes \varrho^T) \\
&\leqslant |\mathrm{A}| \times \max_{\|Y\|_1 \leqslant 1} \mathrm{Tr}(\tau Y) \\
&= |\mathrm{A}| \times \|\tau\|_\infty \\
&\leqslant \frac{C}{|\mathrm{A}|}.
\end{aligned}$$

Hence, by Theorem 3.3, there exists $\widehat{\mathcal{N}}$ with Kraus rank at most $C'|\mathrm{A}|/\varepsilon^2$ such that

$$\forall\ \varrho \in \mathcal{D}(\mathrm{A}),\ \|\mathcal{N}(\varrho) - \widehat{\mathcal{N}}(\varrho)\|_1 \leqslant C\varepsilon.$$

Let $\hat{\tau} \in \mathcal{D}(\mathrm{A} \otimes \mathrm{A})$ be the Choi-Jamiołkowski state associated to $\widehat{\mathcal{N}}$. It has rank equal to the Kraus rank of $\widehat{\mathcal{N}}$, i.e. at most $C'|\mathrm{A}|/\varepsilon^2$. By the same reasoning as in the proof of Equation (32) in Theorem 5.6, we have, denoting by ψ the maximally entangled state on $\mathrm{A} \otimes \mathrm{A}$,

$$\|\tau - \hat{\tau}\|_{\mathbf{LOCC}^{\rightarrow}(\mathrm{A}:\mathrm{A})} = \left\|\mathcal{N} \otimes \mathcal{I}(\psi) - \widehat{\mathcal{N}} \otimes \mathcal{I}(\psi)\right\|_{\mathbf{LOCC}^{\rightarrow}(\mathrm{A}:\mathrm{A})} \leqslant C\varepsilon.$$

This proves the first inequality in Equation (33). As for the second one, it follows from the fact that τ has largest eigenvalue at most $C/|\mathrm{A}|^2$, while $\hat{\tau}$ has rank at most $C'|\mathrm{A}|/\varepsilon^2$. Indeed, we therefore have

$$\|\tau - \hat{\tau}\|_1 \geqslant \frac{C'|\mathrm{A}|}{\varepsilon^2}\left(\frac{\varepsilon^2}{C'|\mathrm{A}|} - \frac{C}{|\mathrm{A}|^2}\right) + 1 - \frac{C'C}{\varepsilon^2|\mathrm{A}|} = 2\left(1 - \frac{C'C}{\varepsilon^2|\mathrm{A}|}\right),$$

which is exactly the announced result. □

Applying Theorem 5.8 with $\varepsilon = 1/|\mathrm{A}|^\alpha$, for some $0 < \alpha < 1/2$, shows that, for any state τ on $\mathrm{A} \otimes \mathrm{A}$ such that $\|\tau\|_\infty \leqslant C/|\mathrm{A}|^2$, we can construct a state $\hat{\tau}$ on $\mathrm{A} \otimes \mathrm{A}$ with rank at most $C'|\mathrm{A}|^{1+2\alpha}$ satisfying

$$\|\tau - \hat{\tau}\|_{\mathbf{LOCC}^{\rightarrow}(\mathrm{A}:\mathrm{A})} \leqslant \frac{C}{|\mathrm{A}|^\alpha} \text{ and } \|\tau - \hat{\tau}\|_1 \geqslant 2\left(1 - \frac{C'}{|\mathrm{A}|^{1-2\alpha}}\right).$$

5.3 Werner channels.

An interesting case to which Theorem 3.3 applies is that of the so-called Werner channels. These are defined as the family of CPTP maps

$$\mathcal{W}_\lambda : X \in \mathcal{L}(\mathrm{A}) \mapsto \frac{1}{|\mathrm{A}| + 2\lambda - 1} \left[(\operatorname{Tr} X)\mathbb{1} + (2\lambda - 1)X^T \right] \in \mathcal{L}(\mathrm{A}),\ 0 \leqslant \lambda \leqslant 1.$$

Denoting by ς and α the symmetric and anti-symmetric states on $\mathrm{A} \otimes \mathrm{A}$, it is easy to check that, for each $0 \leqslant \lambda \leqslant 1$, the Choi-Jamiołkowski state $\tau(\mathcal{W}_\lambda)$ associated to $\mathcal{W}_\lambda$ is nothing else than the Werner state $\rho_\lambda = \lambda\varsigma + (1-\lambda)\alpha$. Hence, $\mathcal{W}_\lambda$ has Kraus rank $|\mathrm{A}|^2$ whenever $0 < \lambda < 1$, and $|\mathrm{A}|(|\mathrm{A}|+1)/2$, resp. $|\mathrm{A}|(|\mathrm{A}|-1)/2$, when $\lambda = 1$, resp. $\lambda = 0$, i.e. in any case full or almost full Kraus rank. These channels are thus typically of the kind that we would like to compress into more economical ones. What is more, they have the property of having only very mixed output states. Indeed,

$$\max_{\varrho \in \mathcal{D}(\mathrm{A})} \|\mathcal{W}_\lambda(\varrho)\|_\infty = \begin{cases} 2\lambda/(|\mathrm{A}| + 2\lambda - 1) \text{ if } \lambda \geqslant 1/2 \\ 1/(|\mathrm{A}| + 2\lambda - 1) \text{ if } \lambda < 1/2 \end{cases} \leqslant \frac{2}{|\mathrm{A}|}.$$

So by Theorem 3.3, we get that, for each $0 \leqslant \lambda \leqslant 1$, given $0 < \varepsilon < 1$, there exists a CP map $\widehat{\mathcal{W}}_\lambda : \mathcal{L}(\mathrm{A}) \to \mathcal{L}(\mathrm{A})$ with Kraus rank at most $C|\mathrm{A}|/\varepsilon^2$ (where $C > 0$ is a universal constant) such that

$$\forall\ \varrho \in \mathcal{D}(\mathrm{A}),\ \left\| \widehat{\mathcal{W}}_\lambda(\varrho) - \mathcal{W}_\lambda(\varrho) \right\|_\infty \leqslant \frac{\varepsilon}{|\mathrm{A}|}.$$

In words, this means that the Werner CPTP maps can be $(\varepsilon/|\mathrm{A}|)$-approximated in $(1 \to \infty)$-norm distance (hence in particular ε-approximated in $(1 \to 1)$-norm distance) by CP maps having Kraus rank $C|\mathrm{A}|/\varepsilon^2 \ll |\mathrm{A}|^2$.

Remark 5.9. *Note that, as a special case of this approximation result for Werner channels, we recover the well-known approximation result for the fully randomizing channel $\mathcal{R}$, recalled in Section 2. Indeed, $\mathcal{W}_{1/2}$ is nothing else than $\mathcal{R}$.*

6 Discussion

We have generalized in several senses the result established in [8] and [2]. First, we have shown that it holds for all quantum channels and not only for the fully randomizing one: any CPTP map from $\mathcal{L}(\mathrm{A})$ to $\mathcal{L}(\mathrm{B})$ can be ε-approximated in $(1 \to 1)$-norm distance by a CPTP map with Kraus rank of order $d \log(d/\varepsilon)/\varepsilon^2$, where $d = \max(|\mathrm{A}|, |\mathrm{B}|)$. Second, we have established that a stronger notion of approximation can actually be proven, namely an ε-ordering of the two CP maps, which allows to derive approximation results in terms of various output quantities (that are tighter than those induced by the rougher norm distance closeness). In the case where the channel under consideration is, as the fully randomizing channel, very noisy (meaning that all output states are very mixed), the extra $\log(d/\varepsilon)$ factor in our result can be removed. However, we do not know if this is true in general. On a related note, our study of optimality shows that there exist channels which cannot be compressed below order d Kraus operators (even to achieve the weakest notions of approximation). But what about channel-dependent upper and lower bounds? For a given channel, would there be a more clever construction than ours (i.e. a non-universal one) that would enable its compression to a number of Kraus operators whose log would be, for instance, of the order of its maximum input-output entropy difference?

Furthermore, full or partial derandomization of our construction would be desirable in practice. Here again the main difficulty is that most of the techniques which apply to very noisy channels may fail in general. Let us specify a bit what we mean. In [2], two approximation schemes were proposed for the fully randomizing channel $\mathcal{R} : \mathcal{L}(\mathbf{C}^d) \to \mathcal{L}(\mathbf{C}^d)$. They consisted in taking as Kraus operators $\{U_i/\sqrt{n},\ 1 \leqslant i \leqslant n\}$ with $U_1, \ldots, U_n$ sampled either from the Haar measure on $\mathcal{U}(\mathbf{C}^d)$ or from any other isotropic (aka unitary 1-design) measure on $\mathcal{U}(\mathbf{C}^d)$. It was then shown that, in order to approximate $\mathcal{R}$ up to error ε/d in $(1\to\infty)$-norm, n of order d/ε^2 was enough in the Haar-distributed case and n of order $d\log^6 d/\varepsilon^2$ was enough in the, more general, isotropically-distributed case. The advantage of the second result compared to the first one is that there exist isotropic measures which are much simpler than the Haar measure on $\mathcal{U}(\mathbf{C}^d)$, in particular discrete ones (e.g. the uniform measure over any unitary orthogonal basis of $\mathcal{L}(\mathbf{C}^d)$). Hence, from a practical point of view, generating such a measure is arguably more realistic than generating the Haar measure (the reader is e.g. referred to [4] for a more precise formulation of the claim that implementing a Haar distributed unitary is hard and an extensive discussion on how to approximate such a unitary by a more easily implementable one). Now, if $\mathcal{N} : \mathcal{L}(\mathbf{C}^d) \to \mathcal{L}(\mathbf{C}^d)$ is a channel, with environment $\mathbf{C}^s$, such that $\sup_{\rho\in\mathcal{D}(\mathbf{C}^d)} \|\mathcal{N}(\rho)\|_\infty \leqslant C/d$, then arguments of the same type apply to our construction: to approximate $\mathcal{N}$ up to error ε/d in $(1\to\infty)$-norm by sampling unit vectors in $\mathbf{C}^s$, order d/ε^2 of them is enough if they are Haar-distributed (which is the content of Theorem 3.1) and order $d\log^6 d/\varepsilon^2$ of them is enough if they are only assumed to be isotropically-distributed. Here as well, the gain in terms of needed amount of randomness is obvious: there exist isotropic measures which are much simpler to sample from than the Haar-measure on $S_{\mathbf{C}^s}$ (e.g. the uniform measure on any orthonormal basis of $\mathbf{C}^s$). Unfortunately, this whole reasoning (based on Dudley's upper bounding of Bernoulli averages by covering number integrals and on a sharp entropy estimate for the suprema of empirical processes in Banach spaces) fails completely for channels that have some of their outputs which are too pure.

Finally, one could ask whether other parameters than the Kraus rank would be relevant ones to try and minimize. As we have already explained, the question we investigate can be seen as a data compression problem: how can we approximate a given protocol while reducing the amount of resources needed? Taking the Kraus rank, an integer-valued non-smooth quantity, as figure of merit in this channel compression task is somehow putting ourselves in a one-shot scenario. But one could wonder how to define an asymptotic version of this question, with a corresponding regularized version of the Kraus rank.

Acknowledgements

This work owes a lot to Guillaume Aubrun. He is actually the one who first raised the question of compressing a general quantum channel in terms of its Kraus operators, following the papers [8] and [2] which were dealing with the case of the fully randomizing channel. He then followed all the steps of the project, including the most off-putting ones, such as proof-reading technical points of the paper. We would also like to thank an anonymous referee for their incredibly constructive comments, which have helped us improve the paper considerably. In particular, the alternative way of viewing our construction, in terms of Kraus operators rather than Stinespring dilation, that we discuss in Remark 4.6, is due to them.

This research was supported by the European Research Council (grants no. 337603 and no. 648913), the Spanish MINECO (projects FIS2013-40627-P, FIS2016-86681-P and PID2019-107609GB-I00) with the support of FEDER funds, the Generalitat de Catalunya

(CIRIT projects 2014-SGR-966 and 2017-SGR-1127), the John Templeton Foundation (grant no. 48322), the French ANR (projects Stoq 14-CE25-0033 and ANR-11-LABX-0040).

References

[1] Dorit Aharonov, Alexei Kitaev, Noam Nisan. Quantum circuits with mixed states. In *Proc. of the 13th Annual ACM Symposium on Theory of Computing (STOC'98)*, 20–30, 1998. DOI:10.1145/276698.276708.

[2] Guillaume Aubrun. On almost randomizing channels with a short Kraus decomposition. *Commun. Math. Phys.* 288(3):1103–1116, 2009. DOI:10.1007/s00220-008-0695-y.

[3] Koenraad M.R. Audenaert. A sharp Fannes-type inequality for the von Neumann entropy. *J. Phys. A* 40(28):8127–8136, 2007. DOI:10.1088/1751-8113/40/28/S18

[4] Fernando G.S.L. Brandão, Aram W. Harrow, Michał Horodecki. Local random quantum circuits are approximate polynomial-designs. *Commun. Math. Phys.* 346(2):397–434, 2016. DOI:10.1007/s00220-016-2706-8.

[5] Djalil Chafaï, Olivier Guédon, Guillaume Lecué, Alain Pajor. *Interactions between compressed sensing, random matrices and high dimensional geometry*. Panoramas et synthèse 37, Société Mathématique de France, 2012. Pdf available here.

[6] Man-Duen Choi. Completely positive linear maps on complex matrices. *Lin. Alg. Appl.* 10(3):285–290, 1975. DOI:10.1016/0024-3795(75)90075-0.

[7] Vittorio Giovannetti, Alexander S. Holevo. Quantum channels and their entropic characteristics. *Rep. Prog. Phys.* 75(046001), 2012. DOI:10.1088/0034-4885/75/4/046001.

[8] Patrick Hayden, Debbie W. Leung, Peter W. Shor, Andreas Winter. Randomizing quantum states: Constructions and applications. *Commun. Math. Phys.* 250(2):371–391, 2004. DOI:10.1007/s00220-004-1087-6.

[9] Cécilia Lancien, Andreas Winter. Distinguishing multi-partite states by local measurements. *Commun. Math. Phys.* 323(2):555–573, 2013. DOI:10.1007/s00220-013-1779-x.

[10] Gilles Pisier. *The Volume of Convex Bodies and Banach Spaces Geometry*. Cambridge Tracts in Mathematics 94, Cambridge University Press, Cambridge, 1989. DOI:10.1017/CBO9780511662454.

[11] William F. Stinespring. Positive functions on C*-algebras. In *Proc. of the American Mathematical Society* 6(2):211–216, 1955. DOI:10.2307/2032342.

[12] Andreas Winter. Tight uniform continuity bounds for quantum entropies: conditional entropy, relative entropy distance and energy constraints. *Commun. Math. Phys.* 347(1):291–313, 2016. DOI:10.1007/s00220-016-2609-8.