Quantum Channel Stein's Lemma with an Exponential Strong Converse

Minbo Gao*[1,2], Zhengfeng Ji†[3], and Chenghua Liu‡[1,2]

[1]Institute of Software, Chinese Academy of Sciences, Beijing, China
[2]University of Chinese Academy of Sciences, Beijing, China
[3]Department of Computer Science and Technology, Tsinghua University, Beijing, China

Abstract

We establish the quantum channel Stein lemma at every fixed type-I error tolerance in $(0,1)$ and prove an exponential strong converse under arbitrary adaptive strategies. For any finite-dimensional channel pair, the optimal type-II exponent equals the regularized channel relative entropy and is attained by parallel strategies. When this divergence is finite, every strictly larger type-II rate forces the probability of accepting the null hypothesis to decay exponentially, uniformly over strategies and quantum memories. The proof rests on a uniform Stinespring approximation: tensor powers of the null dilation are approximated by tensor powers of the alternative dilation after applying an auxiliary linear map to the environment. The operator-norm error is exponentially small under a squared-norm bound at any rate above the regularized relative entropy. We construct the approximation from a weak testing bound by iterative rate reduction and a fixed-block tensor expansion with an exact residual. It implies right continuity at order one of the regularized sandwiched Rényi channel divergence, which yields the adaptive converse through the channel chain rule. We also obtain a sharp hockey-stick threshold and a subchannel smoothing asymptotic equipartition property for fixed and subexponentially vanishing diamond-norm errors.

1 Introduction

Stein's lemma identifies relative entropy as the optimal type-II error exponent in asymmetric hypothesis testing [1, Sec. 11.8]. It is a basic result in the classical theory of error exponents [2]. Hiai and Petz established the quantum state version [3], and Ogawa and Nagaoka proved the strong converse [4], giving the same exponent at every fixed type-I error tolerance in $(0,1)$. Later work determined the strong-converse exponent [5] and the second-order asymptotics [6, 7]. These results concern independent copies of a fixed pair of states, with optimization over the final measurement.

Channel discrimination also optimizes how the states to be measured are produced. Even for a memoryless unknown channel, the experiment may use inputs entangled across all channel uses and adaptive processing with a quantum memory. The resulting output states need not be tensor powers of any fixed pair. A converse must therefore control entire sequences of experiments, rather than the repetition of a single state discrimination problem.

For channels $\mathcal{N}$ and $\mathcal{M}$, the natural candidate for the Stein exponent is the regularized channel relative entropy

$$d = D^{\infty}(\mathcal{N}\|\mathcal{M}) = \sup_{k\geq 1}\frac{1}{k}D(\mathcal{N}^{\otimes k}\|\mathcal{M}^{\otimes k}), \tag{1.1}$$

*gmb17@tsinghua.org.cn
†jizhengfeng@tsinghua.edu.cn
‡liuch.russell@gmail.com

where $D(\mathcal{N}\|\mathcal{M})$ is the largest output relative entropy obtained from a common input and an arbitrary reference. Every rate below d is achievable: choose a suitable input block, repeat it independently, and apply the state Stein lemma. The difficulty is the fixed-error converse. If $\beta_n^*(\varepsilon)$ is the optimal parallel type-II error at type-I error at most ε, binary data processing gives only

$$-\frac{1}{n}\log\beta_n^*(\varepsilon) \leq \frac{d+1/n}{1-\varepsilon}. \tag{1.2}$$

For $0 < d < \infty$, this leaves a gap at every fixed $\varepsilon > 0$. Adaptive strategies add a further difficulty: the input at a given use can already differ under the two hypotheses.

We prove that d is the fixed-error Stein exponent for arbitrary finite-dimensional channel pairs, under both parallel and general adaptive strategies. For finite d, we also establish an exponential strong converse: for every $R > d$, any adaptive n-use test with type-II error at most 2^{-nR} accepts the null hypothesis with probability at most 2^{-nc_R}, where $c_R > 0$ is independent of the strategy and its memory dimensions. Thus neither adaptive control nor the choice of a fixed error tolerance changes the first-order exponent. The precise statement is given in Theorem 2.1.

Related work. Hayashi studied classical channel discrimination with adaptive input selection [8]. For quantum channels, Cooney, Mosonyi, and Wilde determined strong-converse exponents when the alternative is a replacer channel [9]. Wilde, Berta, Hirche, and Kaur developed amortized channel divergences for adaptive converses and proved a strong Stein lemma for classical–quantum channels [10]. Wang and Wilde formulated the resource theory of asymmetric channel distinguishability and established its weak Stein characterization [11].

For general channel pairs, Fang, Fawzi, Renner, and Sutter proved a relative-entropy chain rule that identifies amortization with regularization and determines the weak adaptive Stein exponent [12]. Bergh, Datta, Salzmann, and Wilde gave quantitative comparisons between adaptive and parallel discrimination strategies for finitely many channel uses [13]. Fawzi and Fawzi established the corresponding sandwiched Rényi chain rule and characterized the optimal adaptive strong-converse exponent [14]. The remaining question was whether that exponent is strictly positive at every rate above d. As explained in [14, Remark 5.6], locating this threshold requires right continuity at order one of the regularized sandwiched Rényi channel divergence. Fang, Gour, and Wang further developed the connections among weak Stein results, exponential strong converses, continuity, and asymptotic equipartition properties formulated through output-state smoothing [15].

Smoothing provides another formulation of the threshold problem. For states, smooth entropies and max-relative entropy have asymptotic limits governed by von Neumann and relative entropies [16–18]. For channels, the normalization of the approximating map matters. Gour related subchannel smoothing to the parallel Stein strong converse and sharp testing thresholds, while showing that smoothing over trace-preserving channels can have a strictly larger asymptotic rate [19]. That work also develops uniform completely positive approximations from testing information. Our smoothing result concerns subchannels and does not impose exact trace preservation on the approximating maps.

Our approach. The central step is a uniform comparison of Stinespring dilations. We approximate tensor powers of the null dilation by an auxiliary linear map on the alternative environment, with exponentially small error and a squared-norm rate arbitrarily close to d from above. The same map works for every input and reference system. This is the additional control needed to pass from a weak testing estimate to the fixed-error threshold: it bounds acceptance probabilities directly, proves the Rényi continuity required by the adaptive chain rule, and yields dominated subchannel approximations. Section 2.3 states and explains the approximation theorem, and Section 2.4 outlines its proof.

The results and proof overview appear in Section 2. We derive the Stein theorem from the approximation in Section 3, then prove the approximation itself in Section 4. Section 5 treats testing thresholds and subchannel smoothing, and Section 6 discusses further questions.

2 Main Results and Proof Overview

2.1 Notation and discrimination model

All Hilbert spaces are finite dimensional, and n denotes a positive integer. We write $\mathcal{L}(A, B)$ for linear operators from A to B, $\mathcal{L}(A) = \mathcal{L}(A, A)$, and $\mathcal{S}(A)$ for density operators on A. A channel $\mathcal{N} : A \to B$ is a completely positive trace-preserving (CPTP) map from $\mathcal{L}(A)$ to $\mathcal{L}(B)$. For completely positive (CP) maps, $\Phi \leq_{\mathrm{CP}} \Psi$ means that $\Psi - \Phi$ is CP. We write $A^n = A^{\otimes n}$, identify tensor products under the canonical permutations of their factors, and omit identity subscripts only when the system is unambiguous.

The identity operator and identity channel on A are I_A and id_A. The adjoint, transpose in a fixed basis, and Moore–Penrose inverse are $X^\dagger$, X^T, and X^+, respectively. For a subspace K, Π_K is its orthogonal projection. For Hermitian X, X_+ denotes its positive part. The operator and trace norms are $\|X\|_\infty$ and $\|X\|_1$; for vectors we use the Euclidean norm $\|x\|$. Logarithms are base two unless denoted by ln, and $0 \log 0 = 0$. The binary entropy is $h_2(x) = -x \log x - (1 - x) \log(1 - x)$ for $x \in [0, 1]$. We refer to [20] for standard facts about quantum channels and matrix norms.

For states ρ, σ, their relative entropy is $D(\rho\|\sigma) = \operatorname{Tr} \rho(\log \rho - \log \sigma)$ when $\operatorname{supp} \rho \subseteq \operatorname{supp} \sigma$, and is $+\infty$ otherwise. An effect $0 \leq T \leq I$ represents acceptance of the null hypothesis. Its acceptance probabilities under ρ and σ are $p = \operatorname{Tr} T\rho$ and $q = \operatorname{Tr} T\sigma$, respectively; the type-I and type-II errors are $1 - p$ and q. For $0 < \varepsilon < 1$, define

$$D_H^\varepsilon(\rho\|\sigma) = -\log \inf_{0 \leq T \leq I} \{\operatorname{Tr} T\sigma : \operatorname{Tr} T\rho \geq 1 - \varepsilon\}.$$

A parallel n-use strategy may be taken to prepare a pure state $|\psi\rangle_{RA^n}$ and measure RB^n after the unknown channel has acted on all inputs. The two output states are

$$\rho_n^\psi = (\mathrm{id}_R \otimes \mathcal{N}^{\otimes n})(|\psi\rangle\langle\psi|), \qquad \sigma_n^\psi = (\mathrm{id}_R \otimes \mathcal{M}^{\otimes n})(|\psi\rangle\langle\psi|). \tag{2.1}$$

Purifying an input cannot reduce the divergences used here or worsen a test, since the purifying system may be ignored at the final measurement. A pure input has Schmidt rank at most $\dim A^n$ across the reference–input cut, so $R \simeq A^n$ suffices up to a reference isometry. We use this reduction in all parallel optimizations below, without restricting entanglement across channel uses.

We use the same divergence symbol for states and channels; the arguments distinguish them. For a fixed reference $R \simeq A$, optimize over unit vectors $|\psi\rangle_{RA} \in R \otimes A$:

$$D(\mathcal{N}\|\mathcal{M}) = \sup_{|\psi\rangle_{RA}} D\big((\mathrm{id}_R \otimes \mathcal{N})(|\psi\rangle\langle\psi|)\|(\mathrm{id}_R \otimes \mathcal{M})(|\psi\rangle\langle\psi|)\big). \tag{2.2}$$

Set $a_n = D(\mathcal{N}^{\otimes n}\|\mathcal{M}^{\otimes n})$. Product inputs and additivity of state relative entropy give $a_{n+m} \geq a_n + a_m$; hence Fekete's lemma yields

$$d = D^\infty(\mathcal{N}\|\mathcal{M}) = \lim_{n\to\infty} \frac{a_n}{n} = \sup_{n\geq 1} \frac{a_n}{n}, \tag{2.3}$$

where $+\infty$ is allowed. The optimal parallel type-II error is

$$\beta_n^*(\varepsilon) = \inf_{|\psi\rangle,\, 0\le T\le I} \{\operatorname{Tr} T\sigma_n^\psi : \operatorname{Tr} T\rho_n^\psi \ge 1-\varepsilon\}. \tag{2.4}$$

Here $R \simeq A^n$ is fixed, $|\psi\rangle \in R\otimes A^n$ is a unit vector, and T acts on $R\otimes B^n$. The feasible set is compact, so the infimum is attained.

An adaptive strategy alternates the unknown channel with arbitrary CPTP maps on the preceding output and a quantum memory. Its initial state and intermediate maps are the same under both hypotheses. Intermediate measurements and classical feedback are included by storing outcomes in the memory. Each memory is finite dimensional, but its dimension is unrestricted and may depend on n. We write $\beta_{n,\mathrm{ad}}^*(\varepsilon)$ for the corresponding infimum of the type-II error under the same type-I constraint. This is the standard adaptive model used in [10, 11]. For either strategy class, p and q denote the probabilities of accepting the null hypothesis when the channel is $\mathcal{N}$ and $\mathcal{M}$, respectively.

2.2 The Stein threshold

The first theorem determines the fixed-error exponent and gives an exponential bound above it. The bound applies to every adaptive strategy, including those with memory dimensions that grow with the blocklength.

Theorem 2.1 (Channel Stein lemma and exponential strong converse)**.** *Let $\mathcal{N},\mathcal{M} : A\to B$ be channels and set $d = D^\infty(\mathcal{N}\|\mathcal{M})$. For every $0<\varepsilon<1$,*

$$\lim_{n\to\infty} -\frac{1}{n}\log \beta_{n,\mathrm{ad}}^*(\varepsilon) = \lim_{n\to\infty} -\frac{1}{n}\log\beta_n^*(\varepsilon) = d. \tag{2.5}$$

If $d<\infty$, then for every $R>d$ there is $c_R>0$ such that all adaptive n-use tests satisfy

$$q \le 2^{-nR} \quad\Longrightarrow\quad p\le 2^{-nc_R} \qquad (n\ge 1). \tag{2.6}$$

In (2.6), c_R depends only on the channels and R. Every nonnegative rate below d is attained by repeating a suitable fixed input block, with type-I error tending to zero. Above d, the probability of accepting the null hypothesis tends to zero exponentially. The fixed-error threshold is consequently independent of both the chosen tolerance and access to adaptive control.

The limits in (2.5) are understood in the extended real sense. When $d=+\infty$, parallel tests have zero type-II error at all sufficiently large blocklengths; Proposition 3.6 gives an explicit construction. For replacer channels $\mathcal{N}(X) = \operatorname{Tr}(X)\rho$ and $\mathcal{M}(X)=\operatorname{Tr}(X)\sigma$, the threshold is $D(\rho\|\sigma)$. If $\mathcal{N}=\mathcal{M}$, then $p=q$ and $\beta_n^*(\varepsilon)=1-\varepsilon$.

2.3 Uniform Stinespring approximation

Our main technical result concerns the dilations of the two channels. A Stinespring isometry $V_N : A\to B\otimes \mathsf{E}_N$ represents $\mathcal{N}$ as $\mathcal{N}(X) = \operatorname{Tr}_{\mathsf{E}_N} V_N X V_N^\dagger$ [21]; choose an analogous dilation V_M of $\mathcal{M}$. Rather than compare the output states separately for each input, we approximate the dilation itself. Specifically, we seek an auxiliary linear map C_n between the two environment spaces such that $(I_{B^n}\otimes C_n)V_M^{\otimes n}$ is close to $V_N^{\otimes n}$. The following theorem controls both the approximation error and the norm of this map.

Theorem 2.2 (Exponential Stinespring approximation)**.** *Let $\mathcal{N}, \mathcal{M} : A \to B$ be channels with $d = D^{\infty}(\mathcal{N}\|\mathcal{M}) < \infty$, and fix Stinespring isometries $V_N : A \to B \otimes \mathsf{E}_N$ and $V_M : A \to B \otimes \mathsf{E}_M$. For every $S > d$, there are $K, \gamma > 0$ and auxiliary linear maps $C_n : \mathsf{E}_M^{\otimes n} \to \mathsf{E}_N^{\otimes n}$, for all sufficiently large n, such that $\|C_n\|_\infty \le 2^{nS/2}$ and*

$$\left\| V_N^{\otimes n} - (I_{B^n} \otimes C_n) V_M^{\otimes n} \right\|_\infty \le K e^{-\gamma n}. \tag{2.7}$$

Write $\widehat{V}_n = (I_{B^n} \otimes C_n) V_M^{\otimes n}$ and $e_n = \left\| V_N^{\otimes n} - \widehat{V}_n \right\|_\infty$. For any unit vector on $R \otimes A^n$, the two operators give output vectors at Euclidean distance at most e_n, because tensoring with I_R preserves the operator norm. The same auxiliary map therefore works uniformly over all inputs and reference dimensions.

The norm constraint determines how this approximation can be used in a test. Applying C_n can increase an acceptance amplitude by a factor of at most $\|C_n\|_\infty$, so its squared norm is the corresponding probability factor. This explains the normalization $\|C_n\|_\infty^2 \le 2^{nS}$. Stinespring continuity estimates [22] provide related comparisons between dilation and channel distance. Here we use the direct uniform output bound above, because C_n is an auxiliary map whose norm may grow exponentially with n.

Finite divergence already gives an exact comparison $V_N = (I_B \otimes C_*) V_M$ with $\|C_*\|_\infty \le 2^{L/2}$ at some finite domination rate $L \ge d$, where $L = \log c$ for any finite $c \ge 1$ with $\mathcal{N} \le_{\mathrm{CP}} c\mathcal{M}$; see Lemma 3.1 and Corollary 3.3. The gain in Theorem 2.2 is that allowing an exponentially small error reduces the required rate to any $S > d$, including $S < L$. Moreover, C_n and $C_*^{\otimes n}$ act between the same prescribed environments. Their difference is therefore an exact auxiliary map for the residual $V_N^{\otimes n} - \widehat{V}_n$. This fact is used both in the tensor construction and in the Rényi estimates.

The approximation also controls the channel after the environment is discarded. Dividing $\widehat{V}_n$ by $1 + e_n$ makes it a contraction, which defines a completely positive, trace-nonincreasing map, or subchannel. This map is exponentially close to $\mathcal{N}^{\otimes n}$ in diamond norm and is dominated, in completely positive order, by $2^{nS}\mathcal{M}^{\otimes n}$. Thus approximation and domination can be imposed simultaneously at any rate above d. We prove this consequence in Section 5.2 and use it to determine the subchannel smoothing rate. We discuss possible quantitative refinements of this approximation in Section 6.

2.4 Overview of the proof

We first derive the statistical results from the approximation, then explain its construction. The construction uses only a weak testing bound.

From approximation to testing. For a parallel input and a binary test, the approximation in Theorem 2.2 gives

$$\sqrt{p} \le \|C_n\|_\infty \sqrt{q} + e_n.$$

Indeed, the test acts on the output and reference, whereas C_n acts on the environment. If $d < S < R$ and $q \le 2^{-nR}$, this bounds $\sqrt{p}$ by $2^{-n(R-S)/2} + Ke^{-\gamma n}$. The parallel strong converse follows, uniformly over all inputs and tests.

For adaptive strategies, we use the approximation to prove Rényi continuity. Each null output purification is the sum of an approximate part with cost rate S and a small residual with a finite cost rate L. A weighted Schatten-norm estimate then bounds the regularized Rényi divergence by S for orders sufficiently close to one. Since $S > d$ is arbitrary, this proves continuity at order one. The

Fawzi–Fawzi chain rule [14] then bounds the divergence increase at each adaptive use. Choosing the order sufficiently close to one gives a positive converse exponent at every $R > d$.

Constructing the approximation. For any $r > d$, relative-entropy data processing bounds the supremum of $p - 2^{nr}q$ over parallel inputs and tests by $d/r + o(1)$. Semidefinite duality converts this into a uniform approximation using an auxiliary map of norm at most $2^{nr/2}$ and with error at most a constant $\delta < 1$. We also have the exact auxiliary map at rate L. The task is to obtain the accuracy of the exact representation at a rate close to r.

Directly repeating the low-rate approximation can accumulate error. Instead, let X be such an approximation and Y an accurate one at a higher rate. Write $Y = X + H$ and expand $(X + H)^{\otimes m}$. Keeping only terms with at most a suitable fraction of H factors lowers the cost rate. Since $\|H\|_\infty < 1$, choosing this fraction sufficiently close to one makes the omitted tail exponentially small in m. The internal blocklength is also allowed to grow, which controls the error in replacing $Y^{\otimes m}$ by the target. A fixed number of iterations gives a stretched-exponential error at every rate above r; this is Lemma 4.4.

We then choose a sufficiently accurate block X of length k and keep k fixed. Its exact residual $H = V_N^{\otimes k} - X$ satisfies $(X + H)^{\otimes m} = V_N^{\otimes km}$. Because $\|H\|_\infty$ can be made arbitrarily small, using only a small fraction of residual factors achieves exponential accuracy with an arbitrarily small increase in rate. There is now no error from repeating an approximate target. Lemma 4.5 formalizes this second construction.

3 Proof of the Channel Stein Theorem

We use Theorem 2.2 to prove the parallel converse, Rényi continuity, and the adaptive converse, in that order. We first record the algebraic facts needed in these deductions. The approximation theorem is proved independently in Section 4, using only a weak testing bound.

3.1 Support and exact environmental comparison

For a linear map $\Phi : A \to B$, its unnormalized Choi operator is $J_\Phi = (\mathrm{id}_A \otimes \Phi)(|\Omega\rangle\langle\Omega|)$, where $|\Omega\rangle = \sum_{i=1}^{\dim A} |i\rangle|i\rangle$ and the first copy of A is a reference. The Choi criterion states that Φ is CP exactly when $J_\Phi \geq 0$; a channel also satisfies $\mathrm{Tr}_B\, J_\Phi = I_A$ [23]. This turns finite divergence into an operator-order condition.

Lemma 3.1 (Finite domination and regularization). *For channels $\mathcal{N}, \mathcal{M} : A \to B$, finiteness of $D^\infty(\mathcal{N}\|\mathcal{M})$ is equivalent to $\operatorname{supp} J_\mathcal{N} \subseteq \operatorname{supp} J_\mathcal{M}$, and to $\mathcal{N} \leq_{\mathrm{CP}} c\mathcal{M}$ for some finite $c \geq 1$.*

Under these equivalent conditions, the proof gives the explicit choice $c = \left\|(J_\mathcal{M}^+)^{1/2} J_\mathcal{N} (J_\mathcal{M}^+)^{1/2}\right\|_\infty$. We write $L = \log c$ for this domination rate; it satisfies $0 \leq a_n \leq nd \leq nL$. If support inclusion fails, then $a_n = +\infty$ for every n.

Proof. The Choi criterion and finite-dimensional matrix order identify support inclusion with domination by a finite multiple. Under support inclusion, the stated value of c gives $J_\mathcal{N} \leq cJ_\mathcal{M}$; taking traces shows $c \geq 1$. To see why this domination persists under tensor powers, let $\Phi \leq_{\mathrm{CP}} \Psi$ be CP maps. The identity

$$\Psi^{\otimes(n+1)} - \Phi^{\otimes(n+1)} = (\Psi^{\otimes n} - \Phi^{\otimes n}) \otimes \Psi + \Phi^{\otimes n} \otimes (\Psi - \Phi)$$

shows the induction step: if the difference at n is CP, then both terms on the right are CP, since tensor products of CP maps are CP. Starting with $\Psi - \Phi$ therefore proves $\Phi^{\otimes n} \leq_{\mathrm{CP}} \Psi^{\otimes n}$ for every n. Applying this with $\Phi = \mathcal{N}$ and $\Psi = c\mathcal{M}$ gives $\mathcal{N}^{\otimes n} \leq_{\mathrm{CP}} c^n \mathcal{M}^{\otimes n}$.

Complete positivity allows us to tensor this difference with id_R and apply it to any input $|\psi\rangle\langle\psi|$. Thus every output pair satisfies $\rho_n^\psi \leq c^n \sigma_n^\psi$, which also implies $\operatorname{supp}\rho_n^\psi \subseteq \operatorname{supp}\sigma_n^\psi$. Work on $\operatorname{supp}\sigma_n^\psi$ and let I denote its identity operator. Adding ηI for $\eta > 0$ makes both states strictly positive on this subspace. Since $c \geq 1$, we have

$$\rho_n^\psi + \eta I \leq c^n \sigma_n^\psi + \eta I \leq c^n(\sigma_n^\psi + \eta I).$$

Operator monotonicity of the logarithm and taking the trace against ρ_n^ψ give

$$\operatorname{Tr}\rho_n^\psi[\log(\rho_n^\psi + \eta I) - \log(\sigma_n^\psi + \eta I)] \leq n \log c.$$

Letting $\eta \downarrow 0$ yields $D(\rho_n^\psi \| \sigma_n^\psi) \leq n \log c$ for every input $|\psi\rangle$. Taking the supremum over inputs gives $a_n \leq n \log c = nL$.

As noted in (2.3), superadditivity implies $d = \sup_n a_n/n$; nonnegativity now gives all stated bounds. Conversely, a normalized maximally entangled input produces $J_{\mathcal{N}}/\dim A$ and $J_{\mathcal{M}}/\dim A$. If support inclusion fails, these states have infinite relative entropy, as do their tensor powers. Hence $a_n = +\infty$ for every n, excluding $d < \infty$. □

To lift Choi domination to fixed dilations, we use the following finite-dimensional form of Douglas factorization [24].

Lemma 3.2 (Matrix factorization)**.** *Let $F : \mathsf{E} \to \mathsf{H}$ and $G : \mathsf{F} \to \mathsf{H}$ be linear maps. If $FF^\dagger \leq tGG^\dagger$ for $t \geq 0$, then $F = GD$ for a linear map $D : \mathsf{E} \to \mathsf{F}$ with $\|D\|_\infty \leq \sqrt{t}$.*

Proof. The order assumption implies $\ker G^\dagger \subseteq \ker F^\dagger$, hence $\operatorname{ran} F \subseteq \operatorname{ran} G$. Set $D = G^+F$, so $GD = F$. Congruence of the assumed inequality by G^+ yields $DD^\dagger \leq tG^+GG^\dagger(G^+)^\dagger = t\Pi_{\operatorname{ran} G^\dagger} \leq tI_{\mathsf{F}}$. Taking the operator norm proves the claim, including $t = 0$. □

Fix Stinespring isometries $V_N : A \to B \otimes \mathsf{E}_N$ and $V_M : A \to B \otimes \mathsf{E}_M$ for the channel pair. Choose orthonormal environment bases, write $V_N|x\rangle = \sum_{j=1}^u N_j|x\rangle \otimes |j\rangle$ with $u = \dim \mathsf{E}_N$, and similarly use $v = \dim \mathsf{E}_M$ Kraus operators for V_M. For $K : A \to B$, define $\operatorname{vec}(K) = \sum_i |i\rangle \otimes K|i\rangle$ and let $F_N = [\operatorname{vec}(N_1) \ \cdots \ \operatorname{vec}(N_u)]$, with F_M defined in the same way. Then $F_N F_N^\dagger = J_{\mathcal{N}}$ and $F_M F_M^\dagger = J_{\mathcal{M}}$.

If $F_0 = F_M D$, its corresponding dilation operator is $(I_B \otimes D^\mathsf{T})V_M$: the transpose is required by this column convention. Transposition preserves the operator norm even for rectangular matrices. More generally, reshaping $V|x\rangle = \sum_j K_j|x\rangle \otimes |j\rangle$ into the column matrix $F = [\operatorname{vec}(K_1) \ \cdots]$ gives

$$\operatorname{Tr}_B FF^\dagger = (V^\dagger V)^\mathsf{T}, \qquad \|V\|_\infty^2 = \left\| \operatorname{Tr}_B FF^\dagger \right\|_\infty. \tag{3.1}$$

Indeed, the (i, j) entry of the partial trace is $\sum_a \langle K_a j, K_a i\rangle$, which is the (j, i) entry of $\sum_a K_a^\dagger K_a$. The identity does not require V to be an isometry.

The factorization lemma now provides an exact comparison in the prescribed environments. This representation will also control the residuals of approximate comparisons.

Corollary 3.3 (Exact auxiliary map)**.** *Fix Stinespring isometries $V_N : A \to B \otimes \mathsf{E}_N$ and $V_M : A \to B \otimes \mathsf{E}_M$ for channels $\mathcal{N}, \mathcal{M}$. If $\mathcal{N} \leq_{\mathrm{CP}} c\mathcal{M}$ for $c \geq 1$, there is an auxiliary linear map $C_* : \mathsf{E}_M \to \mathsf{E}_N$ with $\|C_*\|_\infty \leq \sqrt{c}$ and $V_N = (I_B \otimes C_*)V_M$.*

Proof. Apply Lemma 3.2 to $F_N F_N^\dagger \leq cF_M F_M^\dagger$. It gives $F_N = F_M D$ with $D : \mathsf{E}_N \to \mathsf{E}_M$ and $\|D\|_\infty \leq \sqrt{c}$. The column convention above gives the claimed identity with $C_* = D^\mathsf{T} : \mathsf{E}_M \to \mathsf{E}_N$. □

3.2 The parallel converse and achievability

An operator-norm approximation controls every input and test through the following acceptance-amplitude bound.

Lemma 3.4 (Acceptance amplitudes). *Fix Stinespring isometries $V_N : A \to B \otimes \mathsf{E}_N$ and $V_M : A \to B \otimes \mathsf{E}_M$ for $\mathcal{N}, \mathcal{M}$. For $C : \mathsf{E}_M \to \mathsf{E}_N$, set $e = \|V_N - (I_B \otimes C)V_M\|_\infty$. For any reference-assisted input and binary test, let p, q be its acceptance probabilities under $\mathcal{N}, \mathcal{M}$, respectively. Then*

$$\sqrt{p} \leq \|C\|_\infty \sqrt{q} + e.$$

Proof. After purifying the input and enlarging R, it suffices to consider a unit vector $|\psi\rangle_{RA}$ and an effect $0 \leq T \leq I_{RB}$. Set $x = (I_R \otimes V_N)|\psi\rangle$ and $y = (I_R \otimes V_M)|\psi\rangle$. Tensoring the error with I_R gives $\|x - (I_{RB} \otimes C)y\| \leq e$. As $\sqrt{T}$ is a contraction and acts on systems disjoint from C,

$$\begin{aligned}\sqrt{p} &= \|(\sqrt{T} \otimes I_{\mathsf{E}_N})x\| \\ &\leq \|(I_{RB} \otimes C)(\sqrt{T} \otimes I_{\mathsf{E}_M})y\| + e \leq \|C\|_\infty \sqrt{q} + e.\end{aligned}$$

The argument is independent of the size of the reference. □

Applying this lemma to tensor-power channels gives the parallel converse. The matching lower bound requires only a fixed input block and the state Stein lemma.

Proposition 3.5 (Parallel Stein lemma and exponential converse). *Let $\mathcal{N}, \mathcal{M} : A \to B$ be channels with $d = D^\infty(\mathcal{N}\|\mathcal{M}) < \infty$. For every $0 < \varepsilon < 1$, $\lim_{n\to\infty} -n^{-1} \log \beta_n^*(\varepsilon) = d$. For every $R > d$, there are $K_R, \gamma_R > 0$ such that all parallel n-use tests satisfy*

$$q \leq 2^{-nR} \quad \Longrightarrow \quad p \leq K_R e^{-\gamma_R n} \tag{3.2}$$

for all sufficiently large n.

Proof. Fix $d < S < R$, and take C_n, e_n from Theorem 2.2. Lemma 3.4 applies to the fixed tensor-power dilations. When $q \leq 2^{-nR}$, it gives

$$p \leq \left(2^{-n(R-S)/2} + Ke^{-\gamma n}\right)^2 \leq (1+K)^2 e^{-2\min\{(R-S)\ln 2/2, \gamma\} n}.$$

This proves (3.2). For fixed ε, its right side is eventually smaller than $1-\varepsilon$, so every feasible test in (2.4) has $q > 2^{-nR}$. Thus $\beta_n^*(\varepsilon) \geq 2^{-nR}$ eventually. Since $R > d$ is arbitrary, the limit superior of the normalized negative logarithm is at most d.

For the lower bound, fix a blocklength k and a pure input for k uses, with output states ρ, σ. Finite domination ensures support inclusion. The state Stein lemma gives

$$\lim_{m\to\infty} \frac{1}{m} D_H^\varepsilon(\rho^{\otimes m}\|\sigma^{\otimes m}) = D(\rho\|\sigma).$$

Its direct part also attains each strictly smaller rate with null acceptance tending to one [3,4]. Write $n = mk + \ell$ with $0 \leq \ell < k$. Repeat the input block m times, use any fixed state on the remaining inputs, and ignore the remaining outputs. This is an admissible parallel strategy, so

$$-\log \beta_n^*(\varepsilon) \geq D_H^\varepsilon(\rho^{\otimes m}\|\sigma^{\otimes m}). \tag{3.3}$$

Since $m/n \to 1/k$, the limit inferior is at least $D(\rho\|\sigma)/k$. Taking the supremum over the input and then over k gives $\sup_k a_k/k = d$. This proves the limit and also justifies the below-threshold achievability assertion following Theorem 2.1. □

If Choi support inclusion fails, a zero-probability event under the alternative gives a direct test.

Proposition 3.6 (Failure of Choi support inclusion)**.** *Let $\mathcal{N}, \mathcal{M} : A \to B$ be channels with $\operatorname{supp} J_{\mathcal{N}} \not\subseteq \operatorname{supp} J_{\mathcal{M}}$. There is $\lambda \in (0, 1]$ such that, for every $n \geq 1$, a parallel n-use test has $q_n = 0$ and $p_n = 1 - (1 - \lambda)^n$.*

Consequently, $\beta_n^*(\varepsilon) = 0$ for all sufficiently large n at every fixed $0 < \varepsilon < 1$.

Proof. A normalized maximally entangled input gives the states $\rho = J_{\mathcal{N}}/\dim A$ and $\sigma = J_{\mathcal{M}}/\dim A$. Let $Q = \Pi_{\ker \sigma}$. Support failure implies $\lambda = \operatorname{Tr} Q\rho > 0$, while $\operatorname{Tr} Q\sigma = 0$. Repeat the input independently, measure $\{Q, I - Q\}$ on each output and reference, and accept the null if at least one outcome is Q. The stated probabilities follow by independence. Its type-I error $(1 - \lambda)^n$ eventually lies below every fixed positive tolerance. $\square$

3.3 Rényi continuity

To handle adaptive strategies, we need a divergence with a chain rule for different input states under the two hypotheses. We use the sandwiched Rényi divergence, introduced in [25, 26]. For $\alpha > 1$, a state σ, and a positive operator τ supported on $\operatorname{supp} \sigma$, define

$$\widetilde{Q}_\alpha(\tau\|\sigma) = \operatorname{Tr}\left(\sigma^{-\frac{\alpha-1}{2\alpha}} \tau \sigma^{-\frac{\alpha-1}{2\alpha}}\right)^\alpha. \tag{3.4}$$

Here inverse powers are taken on $\operatorname{supp} \sigma$, and the trace is evaluated on that subspace. For a state ρ, set $\widetilde{D}_\alpha(\rho\|\sigma) = (\alpha - 1)^{-1} \log \widetilde{Q}_\alpha(\rho\|\sigma)$, with value $+\infty$ without support inclusion. We use data processing [27, 28] and $\widetilde{D}_\alpha \geq D$ [25]. For a rectangular matrix F, its Schatten norm is $\|F\|_p = (\operatorname{Tr}(F^\dagger F)^{p/2})^{1/p}$, for $p \geq 1$.

Define $\widetilde{D}_\alpha(\mathcal{N}\|\mathcal{M})$ by the same reference-assisted optimization as (2.2), with D replaced by $\widetilde{D}_\alpha$, and let

$$d_\alpha := \widetilde{D}_\alpha^\infty(\mathcal{N}\|\mathcal{M}) = \lim_{n\to\infty} \frac{1}{n} \widetilde{D}_\alpha(\mathcal{N}^{\otimes n}\|\mathcal{M}^{\otimes n}).$$

For $d < \infty$ and $1 < \alpha \leq 2$, this limit equals the supremum of the normalized block values and lies in $[d, L]$, where L is the domination rate of Lemma 3.1. Indeed, superadditivity follows from product inputs, and the upper bound follows from the small-trace estimate below applied to $\rho_n^\psi \leq 2^{nL} \sigma_n^\psi$.

The adaptive chain rule gives the threshold d once d_α converges to d as $\alpha \downarrow 1$. The next theorem establishes this convergence.

Theorem 3.7 (Right continuity of the regularized divergence)**.** *For channels $\mathcal{N}, \mathcal{M} : A \to B$ with $d = D^\infty(\mathcal{N}\|\mathcal{M}) < \infty$,*

$$\lim_{\alpha\downarrow 1} \widetilde{D}_\alpha^\infty(\mathcal{N}\|\mathcal{M}) = d. \tag{3.5}$$

In fact, let L be the domination rate from Lemma 3.1. For $d < S < L$, take an approximation with error at most $Ke^{-\gamma n}$ from Theorem 2.2. We will prove that, for $1 < \alpha \leq 2$,

$$\widetilde{D}_\alpha^\infty(\mathcal{N}\|\mathcal{M}) \leq \max\left\{S,\ L - \frac{2\gamma}{(\alpha - 1)\ln 2}\right\}. \tag{3.6}$$

To prove this bound, we must control the residual after weighting by a negative power of σ. Small trace alone is insufficient; the following lemma also uses domination by σ, without requiring a lower bound on its nonzero eigenvalues.

Lemma 3.8 (Small-trace dominated terms)**.** *Let $\sigma \in \mathcal{S}(\mathsf{H})$, $t \geq 0$, and $1 < \alpha \leq 2$. Every positive operator $\tau \leq t\sigma$ satisfies $\widetilde{Q}_\alpha(\tau\|\sigma) \leq t^{\alpha-1} \operatorname{Tr} \tau$. Consequently, if $F_0, F_1 : \mathsf{E} \to \mathsf{H}$ are linear maps, $F = F_0 + F_1$, and $F_j F_j^\dagger \leq t_j \sigma$ with $t_j \geq 0$, then*

$$\widetilde{Q}_\alpha(FF^\dagger\|\sigma)^{1/2\alpha} \leq \sum_{j=0}^{1} t_j^{(\alpha-1)/(2\alpha)} \|F_j\|_2^{1/\alpha}. \tag{3.7}$$

Proof. The case $t = 0$ has $\tau = 0$. Otherwise work on $\operatorname{supp} \sigma$ and put $B = \sigma^{-(\alpha-1)/2\alpha} \tau \sigma^{-(\alpha-1)/2\alpha}$. Domination gives $B \leq t\sigma^{1/\alpha}$. The function $x \mapsto x^{\alpha-1}$ is operator monotone for $1 < \alpha \leq 2$, so $B^{\alpha-1} \leq t^{\alpha-1}\sigma^{(\alpha-1)/\alpha}$. Multiplying by B inside the trace preserves this inequality:

$$\operatorname{Tr} B^\alpha \leq t^{\alpha-1} \operatorname{Tr} B\sigma^{(\alpha-1)/\alpha} = t^{\alpha-1} \operatorname{Tr} \tau. \tag{3.8}$$

For the second assertion, $\widetilde{Q}_\alpha(FF^\dagger\|\sigma)^{1/2\alpha} = \|\sigma^{-(\alpha-1)/2\alpha} F\|_{2\alpha}$. Apply the Schatten-norm triangle inequality to $F_0 + F_1$ and use (3.8) with $\tau = F_j F_j^\dagger$. As $\operatorname{Tr} F_j F_j^\dagger = \|F_j\|_2^2$, this gives (3.7), including zero terms. $\square$

Proof of Theorem 3.7. If $d = L$, the bounds $d \leq d_\alpha \leq L$ already prove continuity. Otherwise fix $d < S < L$, the dilations, and auxiliary maps C_n from Theorem 2.2. Corollary 3.3 provides C_* with $\|C_*\|_\infty \leq 2^{L/2}$ in the same environments.

Fix any pure input with a reference for n uses, and let ρ, σ be its output states. Reshape the null output purification, its approximation, and their difference into matrices F, F_0, F_1 whose columns are indexed by the null environment. Then $F = F_0 + F_1$, $FF^\dagger = \rho$, $\|F_0\|_2 \leq 1 + e_n$, and $\|F_1\|_2 \leq e_n$, where $e_n \leq Ke^{-\gamma n}$. If G is the alternative purification's column matrix, then $GG^\dagger = \sigma$, $F_0 = GC_n^\mathsf{T}$, and $F_1 = G(C_*^{\otimes n} - C_n)^\mathsf{T}$. The last identity follows from the exact environmental representation. Consequently,

$$F_0 F_0^\dagger \leq 2^{nS}\sigma, \qquad F_1 F_1^\dagger \leq 4\,2^{nL}\sigma,$$

since $\|C_*^{\otimes n} - C_n\|_\infty \leq 2 \cdot 2^{nL/2}$. Here $F = F_0 + F_1$ is a decomposition of purification matrices, not a positive-operator decomposition of ρ; the cross terms are handled by the Schatten-norm triangle inequality.

Lemma 3.8 gives, uniformly over the input,

$$\begin{aligned}\widetilde{Q}_\alpha(\rho\|\sigma)^{1/2\alpha} \leq{} & (1 + e_n)^{1/\alpha} 2^{nS(\alpha-1)/2\alpha} \\ & + e_n^{1/\alpha} (4\,2^{nL})^{(\alpha-1)/2\alpha}.\end{aligned}$$

For fixed S and α, the two summands are bounded by constant multiples of 2^{nu_0} and 2^{nu_1}, where $u_0 = (\alpha - 1)S/2\alpha$ and $u_1 = (\alpha - 1)L/2\alpha - \gamma/(\alpha \ln 2)$. The constants are independent of the input and n. Hence the logarithm of their sum is at most $n \max\{u_0, u_1\} + O(1)$, uniformly over all inputs. Multiplying by $2\alpha/(\alpha - 1)$, optimizing over the input, and dividing by n proves (3.6) upon taking the limit.

For each fixed $S \in (d, L)$, the second term of the maximum in (3.6) is below S whenever $\alpha > 1$ is sufficiently close to one. Therefore $d \leq \liminf_{\alpha\downarrow 1} d_\alpha \leq \limsup_{\alpha\downarrow 1} d_\alpha \leq S$. Letting $S \downarrow d$ completes the proof. $\square$

The exponential error is needed here because it yields the term $-2\gamma/((\alpha - 1)\ln 2)$ in (3.6). A subexponential error bound would give no such reduction after division by n.

3.4 The adaptive converse

The Fawzi–Fawzi chain rule [14, Corollary 5.2] applies to different input states. For $\alpha > 1$, channels $\mathcal{N}, \mathcal{M} : A \to B$, and states ρ, σ on a common reference and A, it states

$$\begin{aligned} &\widetilde{D}_\alpha\big((\mathrm{id} \otimes \mathcal{N})(\rho) \| (\mathrm{id} \otimes \mathcal{M})(\sigma)\big) \\ &\qquad \le \widetilde{D}_\alpha(\rho\|\sigma) + \widetilde{D}_\alpha^\infty(\mathcal{N}\|\mathcal{M}). \end{aligned} \tag{3.9}$$

Combined with continuity, this bounds every use in an adaptive protocol at a rate arbitrarily close to d.

Proof of Theorem 2.1. Assume $d < \infty$ and fix $R > d$. By Theorem 3.7, choose $1 < \alpha \le 2$ such that $d_\alpha = \widetilde{D}_\alpha^\infty(\mathcal{N}\|\mathcal{M}) < R$. The two hypotheses start from the same state, with divergence zero. Inductively, (3.9) adds at most d_α at each use, while data processing prevents any increase under the intervening CPTP maps. The final-state divergence is therefore at most nd_α. Each step permits an arbitrary common reference, so the bound is independent of the memory dimensions.

Consider a final test with $q \le 2^{-nR}$. If $p = 0$, the desired bound is immediate. If $p > 0$, finite final-state divergence implies $q > 0$, and $R > d \ge 0$ gives $q < 1$. Data processing under the binary measurement, keeping only the term $p^\alpha q^{1-\alpha}$ inside the logarithm, gives

$$\begin{aligned} nd_\alpha &\ge \frac{1}{\alpha - 1} \log\big[p^\alpha q^{1-\alpha} + (1-p)^\alpha (1-q)^{1-\alpha}\big] \\ &\ge \frac{\alpha}{\alpha - 1} \log p - \log q. \end{aligned}$$

Rearranging and using $-\log q \ge nR$ proves

$$p \le 2^{-n\frac{\alpha-1}{\alpha}(R-d_\alpha)}.$$

Thus (2.6) holds for every $n \ge 1$ with $c_R = (\alpha - 1)(R - d_\alpha)/\alpha > 0$.

For fixed ε, the requirement $p \ge 1 - \varepsilon$ is incompatible with this bound for sufficiently large n. Hence the adaptive limit superior in (2.5) is at most d. The parallel block strategies in Proposition 3.5 are adaptive strategies as well and give the matching lower bound. If $d = +\infty$, Lemma 3.1 and Proposition 3.6 provide parallel tests with zero type-II error at all sufficiently large blocklengths. Both fixed-error limits are then $+\infty$. □

This proves positivity of the strong-converse exponent characterized in [14, Theorem 5.5] at every $R > d$.

4 Uniform Exponential Approximation

We now prove Theorem 2.2. The one-shot comparison below applies to arbitrary channel pairs. Finite divergence is needed only afterwards, to obtain both a constant-error approximation from relative entropy and an exact representation at finite cost rate. Two tensor constructions then improve the approximation.

4.1 From testing to a uniform approximation

We use the state hockey-stick quantity $E_t(\rho\|\sigma) = \operatorname{Tr}(\rho - t\sigma)_+$ for $t \geq 0$, extending the definition for $t \geq 1$ in [30]. For channels $\mathcal{N}, \mathcal{M} : A \to B$, define its channel extension by

$$E_t(\mathcal{N}\|\mathcal{M}) = \sup_{|\psi\rangle} \operatorname{Tr}(\rho_1^\psi - t\sigma_1^\psi)_+ = \sup_{|\psi\rangle,\, 0\leq T\leq I_{RB}} (p - tq), \tag{4.1}$$

where $\rho_1^\psi, \sigma_1^\psi$ are the output states in (2.1), and $p = \operatorname{Tr} T\rho_1^\psi$ and $q = \operatorname{Tr} T\sigma_1^\psi$. The second equality follows from the variational formula for the positive part. Pure inputs with $R \simeq A$ suffice, and $0 \leq E_t \leq 1$. For $t \geq 1$, this is the unrestricted-measurement case of [31, Def. 3]. For $0 \leq t < 1$, that reference subtracts $(1-t)_+$.

The next semidefinite representation converts a bound on all inputs and tests into a single positive operator. It is an instance of the semidefinite approach to completely bounded norms [29]; we give the primal and dual explicitly to fix the normalization.

Lemma 4.1 (Channel testing duality)**.** *Let $\mathcal{N}, \mathcal{M} : A \to B$ be channels, let $t \geq 0$, and set $\Delta = J_\mathcal{N} - tJ_\mathcal{M}$. Then*

$$E_t(\mathcal{N}\|\mathcal{M}) = \max_{\substack{\omega\in\mathcal{S}(A)\\ 0\leq Q\leq \omega\otimes I_B}} \operatorname{Tr}\Delta Q \tag{4.2}$$

$$= \min_{Y\geq 0,\ Y\geq\Delta} \|\operatorname{Tr}_B Y\|_\infty . \tag{4.3}$$

Here Q, Y act on $A \otimes B$. Both extrema are attained.

Proof. Identify the reference with A. Up to an isometry on the reference, a pure input has the form $(\sqrt{\omega} \otimes I_A)|\Omega\rangle$ for some $\omega \in \mathcal{S}(A)$. Pulling the output effect T through $\sqrt{\omega}$ gives $Q = (\sqrt{\omega} \otimes I_B)T(\sqrt{\omega} \otimes I_B)$, which satisfies $0 \leq Q \leq \omega \otimes I_B$ and $p - tq = \operatorname{Tr}\Delta Q$. Conversely, this order constraint forces Q to be supported on $\operatorname{supp}\omega \otimes B$. Congruence by the inverse of $\sqrt{\omega}$ on its support gives an effect on that subspace; extending it by zero gives a valid T. This proves (4.2).

Introduce a positive multiplier Y for $Q \leq \omega \otimes I_B$ and a real multiplier λ for $\operatorname{Tr}\omega = 1$. On $Q, \omega \geq 0$, the Lagrangian is

$$\lambda + \operatorname{Tr}[(\Delta - Y)Q] + \operatorname{Tr}[(\operatorname{Tr}_B Y - \lambda I_A)\omega].$$

Its supremum is finite exactly when $Y \geq \Delta$ and $\operatorname{Tr}_B Y \leq \lambda I_A$. For positive Y, the least such λ is $\|\operatorname{Tr}_B Y\|_\infty$, giving the dual in (4.3). Strict primal feasibility is witnessed by $\omega = I_A/\dim A$ and $Q = \eta I_{AB}$ with $0 < \eta < 1/\dim A$. Strict dual feasibility follows by taking $Y = yI_{AB}$ with $y > \max\{0, \|\Delta\|_\infty\}$ and then $\lambda > y \dim B$. Semidefinite strong duality applies. The primal feasible set is compact. On a dual sublevel set, $Y \geq 0$ and $\operatorname{Tr} Y \leq (\dim A)\,\|\operatorname{Tr}_B Y\|_\infty$ bound the norm of Y; the set is also closed. This proves attainment on both sides. $\square$

Fix dilations $V_N : A \to B \otimes \mathsf{E}_N$ and $V_M : A \to B \otimes \mathsf{E}_M$. For $t \geq 0$, let

$$s_t = \min_{\substack{C:\mathsf{E}_M\to\mathsf{E}_N\\ \|C\|_\infty\leq\sqrt{t}}} \|V_N - (I_B \otimes C)V_M\|_\infty . \tag{4.4}$$

The following comparison turns testing control into an approximation in the prescribed environment. This fixed codomain will make it possible to subtract and tensor auxiliary maps constructed at different rates.

Proposition 4.2. *Let $\mathcal{N}, \mathcal{M} : A \to B$ be channels. Fix Stinespring isometries $V_N : A \to B \otimes \mathsf{E}_N$ and $V_M : A \to B \otimes \mathsf{E}_M$ for $\mathcal{N}$ and $\mathcal{M}$, respectively, and define s_t by (4.4) using these isometries. Then*

$$s_t^2 \le E_t(\mathcal{N}\|\mathcal{M}) \le 2s_t - s_t^2 \qquad (t \ge 0).$$

Proof. The set of maps $C : \mathsf{E}_M \to \mathsf{E}_N$ satisfying $\|C\|_\infty \le \sqrt{t}$ is compact in finite dimensions. Since $C \mapsto \|V_N - (I_B \otimes C)V_M\|_\infty$ is continuous, the minimum defining s_t is attained. The choice $C = 0$ gives $s_t \le 1$. Take a dual optimizer Y from Lemma 4.1, and put $A_0 = tJ_\mathcal{M}$ and $H = A_0 + Y$. The dual constraint $Y \ge J_\mathcal{N} - tJ_\mathcal{M}$ implies $J_\mathcal{N} \le tJ_\mathcal{M} + Y = H$. Write the fixed dilations as $V_N|x\rangle = \sum_j N_j|x\rangle \otimes |j\rangle$ and $V_M|x\rangle = \sum_k M_k|x\rangle \otimes |k\rangle$ in orthonormal environment bases. As in Section 3.1, collect the vectorized Kraus operators into the column matrices $F_N = [\mathrm{vec}(N_1) \ \cdots \ \mathrm{vec}(N_u)]$ and $F_M = [\mathrm{vec}(M_1) \ \cdots \ \mathrm{vec}(M_v)]$, where $\mathrm{vec}(K) = \sum_i |i\rangle \otimes K|i\rangle$. These satisfy $F_N F_N^\dagger = J_\mathcal{N}$ and $F_M F_M^\dagger = J_\mathcal{M}$. Define $F_0 = A_0 H^+ F_N$ and $F_1 = Y H^+ F_N$. Here H^+ is the Moore–Penrose pseudoinverse of H. Since $\operatorname{ran} F_N \subseteq \operatorname{supp} H$, we have $F_0 + F_1 = HH^+F_N = F_N$. Moreover,

$$F_0 F_0^\dagger \le A_0 H^+ A_0 \le A_0, \qquad F_1 F_1^\dagger \le Y H^+ Y \le Y. \tag{4.5}$$

The first inequality in each chain uses $J_\mathcal{N} \le H$. For the second, let $0 \le Z \le H$ and put $K = (H^+)^{1/2} Z^{1/2}$. Then $KK^\dagger \le I$, hence $K^\dagger K \le I$. Congruence of $Z^{1/2} H^+ Z^{1/2} \le I$ by $Z^{1/2}$ gives $ZH^+Z \le Z$. Apply this with $Z = A_0$ and $Z = Y$.

Since $F_0 F_0^\dagger \le A_0 = tF_M F_M^\dagger$, Lemma 3.2 gives $F_0 = F_M D$ for a map $D : \mathsf{E}_N \to \mathsf{E}_M$ with $\|D\|_\infty \le \sqrt{t}$. Set $C = D^\mathsf{T} : \mathsf{E}_M \to \mathsf{E}_N$. Transposition preserves the operator norm, so C is feasible in (4.4). Under the column convention of Section 3.1, applying $I_B \otimes C$ to V_M changes its Kraus-column matrix to $F_M C^\mathsf{T} = F_M D = F_0$. More explicitly, write $C|k\rangle = \sum_j C_{jk}|j\rangle$. For any input vector $|x\rangle$, the residual acts as

$$[V_N - (I_B \otimes C)V_M]|x\rangle = \sum_j \left(N_j - \sum_k C_{jk} M_k \right) |x\rangle \otimes |j\rangle.$$

Its jth column is therefore $\mathrm{vec}(N_j - \sum_k C_{jk} M_k) = \mathrm{vec}(N_j) - \sum_k C_{jk}\,\mathrm{vec}(M_k)$. This is precisely the jth column of $F_N - F_0$, so the residual has column matrix $F_N - F_0 = F_1$.

The reshaping identity (3.1) expresses the squared operator norm of this residual as $\left\|\mathrm{Tr}_B F_1 F_1^\dagger\right\|_\infty$. By (4.5), $F_1 F_1^\dagger \le Y$; taking the partial trace preserves this order, as does taking the operator norm of positive operators. Since Y is a dual optimizer, we obtain

$$\|V_N - (I_B \otimes C)V_M\|_\infty^2 = \left\|\mathrm{Tr}_B F_1 F_1^\dagger\right\|_\infty \le \|\mathrm{Tr}_B Y\|_\infty = E_t(\mathcal{N}\|\mathcal{M}).$$

Minimizing over feasible C gives $s_t^2 \le E_t(\mathcal{N}\|\mathcal{M})$. The argument also applies at $t = 0$, when the factorization forces $D = 0$ and hence $C = 0$.

For the upper bound, choose an auxiliary map C attaining the minimum in (4.4). Its approximation error is s_t and $\|C\|_\infty \le \sqrt{t}$, so Lemma 3.4 gives

$$\sqrt{p} \le \|C\|_\infty \sqrt{q} + s_t \le \sqrt{tq} + s_t$$

for every input and effect, with acceptance probabilities p, q. If the testing objective $p - tq$ is positive, then $z = \sqrt{p} - \sqrt{tq} > 0$, and the preceding inequality gives $z \le s_t \le 1$. Factoring the difference of squares yields

$$p - tq = z(\sqrt{p} + \sqrt{tq}) = z(2\sqrt{p} - z) \le 2z - z^2 \le 2s_t - s_t^2.$$

Here the first inequality uses $\sqrt{p} \le 1$, and the second uses that $z \mapsto 2z - z^2$ is increasing on $[0,1]$. If $p - tq \le 0$, the same bound holds because $2s_t - s_t^2 \ge 0$. Thus every input and effect satisfies $p - tq \le 2s_t - s_t^2$. Taking the supremum of $p - tq$ over all inputs and effects gives $E_t(\mathcal{N}\|\mathcal{M})$ by (4.1), proving the upper bound. □

Uniform CP approximations from testing information also appear in [19]. Here the approximation is expressed by one linear auxiliary map with a fixed codomain, independently of both the input and the test. This is the form required by the tensor constructions. For a channel pair with $d = D^\infty(\mathcal{N}\|\mathcal{M}) < \infty$, abbreviate $E_n(r) = E_{2^{nr}}(\mathcal{N}^{\otimes n}\|\mathcal{M}^{\otimes n})$. The next lemma supplies the only information-theoretic estimate needed to begin amplification.

Lemma 4.3. *Let $\mathcal{N}, \mathcal{M} : A \to B$ be channels with $d = D^\infty(\mathcal{N}\|\mathcal{M}) < \infty$. For every $r > 0$ and $n \ge 1$,*

$$E_n(r) \le \min\left\{1, \frac{nd+1}{nr}\right\}. \tag{4.6}$$

Proof. Fix a pure input $|\psi\rangle_{RA^n}$ for n parallel channel uses, and let $\rho_n^\psi, \sigma_n^\psi$ be the output states defined in (2.1). Choose an effect $0 \le T \le I_{RB^n}$ corresponding to acceptance of the null hypothesis. The acceptance probabilities are $p = \operatorname{Tr} T\rho_n^\psi$ under $\mathcal{N}^{\otimes n}$ and $q = \operatorname{Tr} T\sigma_n^\psi$ under $\mathcal{M}^{\otimes n}$. By (4.1), $E_n(r)$ is the supremum of $p - 2^{nr}q$ over these inputs and effects. Since the claimed upper bound is nonnegative, every choice with $p - 2^{nr}q \le 0$ already satisfies it. It therefore suffices to consider choices with $p - 2^{nr}q > 0$. For such a choice, $p > 0$ and $q < 2^{-nr}$, since $p \le 1$. By Lemma 3.1, there is a finite $c \ge 1$ such that $\rho_n^\psi \le c^n \sigma_n^\psi$. Taking the trace against the positive effect T gives $p = \operatorname{Tr} T\rho_n^\psi \le c^n \operatorname{Tr} T\sigma_n^\psi = c^n q$. Thus $q = 0$ would force $p = 0$, contradicting $p > 0$. By the definition of the channel relative entropy and (2.3), this input satisfies $D(\rho_n^\psi\|\sigma_n^\psi) \le a_n \le nd$. The binary measurement $\{T, I - T\}$ maps the two output states to the probability distributions $(p, 1-p)$ and $(q, 1-q)$. Relative entropy cannot increase under this measurement, so

$$\begin{aligned} nd \ge D(\rho_n^\psi\|\sigma_n^\psi) &\ge p\log\frac{p}{q} + (1-p)\log\frac{1-p}{1-q} \\ &= p\log(1/q) + (1-p)\log(1/(1-q)) - h_2(p) \\ &\ge p\log(1/q) - h_2(p). \end{aligned}$$

The equality uses $h_2(p) = -p\log p - (1-p)\log(1-p)$, and the last inequality drops the nonnegative term $(1-p)\log(1/(1-q))$. Since $0 < q < 2^{-nr}$, we have $\log(1/q) > nr$; also $h_2(p) \le 1$. Consequently, $nd \ge pnr - 1$, or $p \le (nd+1)/(nr)$. Since $p - 2^{nr}q \le p$, every positive objective value is at most $(nd+1)/(nr)$. The zero effect and $E_n(r) \le 1$ prove (4.6).

□

We now convert this testing bound into the approximation needed for the tensor argument. Fix Stinespring isometries V_N, V_M for the two channels, and choose a rate $r > d$. Since $d/r < 1$, we can choose a constant δ with $\sqrt{d/r} < \delta < 1$. By Lemma 4.3,

$$E_n(r) \le \frac{nd+1}{nr} = \frac{d}{r} + \frac{1}{nr} \le \delta^2$$

for all sufficiently large n; the last inequality holds because $\delta^2 - d/r > 0$ is fixed.

Apply Proposition 4.2 to the channels $\mathcal{N}^{\otimes n}, \mathcal{M}^{\otimes n}$, using the fixed tensor-power dilations $V_N^{\otimes n}, V_M^{\otimes n}$ and $t = 2^{nr}$. The bound $s_t^2 \le E_t$ says that the minimum squared approximation error under the norm constraint $\|C\|_\infty \le \sqrt{t}$ is at most $E_n(r)$. Since this minimum is attained, there is a map $C_n^{\mathrm{w}} : \mathsf{E}_M^{\otimes n} \to \mathsf{E}_N^{\otimes n}$ such that

$$\|C_n^{\mathrm{w}}\|_\infty^2 \le 2^{nr}, \qquad \left\|V_N^{\otimes n} - (I_{B^n} \otimes C_n^{\mathrm{w}})V_M^{\otimes n}\right\|_\infty \le \sqrt{E_n(r)} \le \delta.$$

Thus the weak testing bound supplies an auxiliary map at any rate $r > d$, with an operator-norm error uniformly bounded by a constant strictly below one. The same auxiliary map works for every input and reference system. This estimate alone does not show that the error tends to zero as n grows.

When $r < L$, the next lemma combines these auxiliary maps with the exact auxiliary map C_* from Corollary 3.3, whose tensor powers have zero error at the finite rate L. The weak auxiliary maps supply the lower rate r, while the exact auxiliary map supplies an accurate starting point and a finite bound on the cost of corrections. The tensor argument uses the strict inequality $\delta < 1$ to control terms containing many corrections, and obtains vanishing error at every rate above r. If $r \ge L$, the exact auxiliary maps already meet the rate bound.

4.2 Two-rate tensor amplification

We improve the approximation in two stages. In this subsection, we combine a low-rate approximation of constant error with an accurate representation at a higher rate, obtaining vanishing error at every rate above the low rate. In Section 4.3, we then fix one sufficiently accurate block and expand it together with its exact residual to obtain exponentially small error, with an arbitrarily small further increase in rate.

The argument is algebraic and does not use channel divergences. For isometries $V : A \to B \otimes \mathsf{E}$ and $W : A \to B \otimes \mathsf{F}$, write $\mathcal{T}_n(C) = (I_{B^n} \otimes C)W^{\otimes n}$ for $C : \mathsf{F}^{\otimes n} \to \mathsf{E}^{\otimes n}$. Canonical permutations group the output and environment factors. The maps $\mathcal{T}_n$ are linear and satisfy

$$\mathcal{T}_{n+m}(C \otimes D) = \mathcal{T}_n(C) \otimes \mathcal{T}_m(D). \tag{4.7}$$

Throughout this and the next subsection, an auxiliary map at blocklength n is a linear map from $\mathsf{F}^{\otimes n}$ to $\mathsf{E}^{\otimes n}$. Its cost rate is at most r when its squared norm is at most 2^{nr}. The next lemma combines a low-rate constant-error approximation with a finite-rate exact representation.

Lemma 4.4 (Two-rate tensor amplification)**.** *Suppose $V = \mathcal{T}_1(C_*)$ with $\|C_*\|_\infty \le 2^{L/2}$ for some $L \ge 0$. Assume that, for some $0 \le r < L$ and $0 \le \delta < 1$, for every sufficiently large n, there exists an auxiliary map C_n^{w} with $\|C_n^{\mathrm{w}}\|_\infty \le 2^{nr/2}$ and $\|\mathcal{T}_n(C_n^{\mathrm{w}}) - V^{\otimes n}\|_\infty \le \delta$.*

Then for every $R > r$, there exist $K, \gamma > 0$ and $a \in (0, 1]$ such that, for all sufficiently large N, there are auxiliary maps C_N with

$$\|C_N\|_\infty \le 2^{NR/2}, \qquad \left\|\mathcal{T}_N(C_N) - V^{\otimes N}\right\|_\infty \le Ke^{-\gamma N^a}. \tag{4.8}$$

Proof. For a rate S, let $\mathsf{P}(S)$ denote the following property: there exist $K, \gamma > 0$ and $a \in (0, 1]$ such that, for every sufficiently large n, there exists an auxiliary map C_n^{s} with $\|C_n^{\mathrm{s}}\|_\infty \le 2^{nS/2}$ and error $e_n = \|\mathcal{T}_n(C_n^{\mathrm{s}}) - V^{\otimes n}\|_\infty \le Ke^{-\gamma n^a}$. All constants and the initial blocklength may depend on S, but not on n. The exact auxiliary maps $C_*^{\otimes n}$ give $\mathsf{P}(L)$ with zero error and $a = 1$.

Step 1: Fix the truncation fraction. Set $A_0 = 1 + \delta$ and $B_0 = (1 + \delta)/2 < 1$. At $x = 1$, the continuous function $(\ln 2)h_2(x) + (1 - x)\ln A_0 + x \ln B_0$ equals $\ln B_0 < 0$. There are therefore $\beta \in (0, 1)$ and $\kappa > 0$ such that

$$(\ln 2)h_2(x) + (1 - x)\ln A_0 + x \ln B_0 \le -\kappa \quad \text{for } \beta \le x \le 1. \tag{4.9}$$

Keep these constants fixed throughout the iteration. Choosing β close to one ensures that the decay of B_0^j outweighs the number of omitted terms in the tensor expansion.

Step 2: Lower the rate of an accurate approximation. We show that

$$\mathsf{P}(S) \Longrightarrow \mathsf{P}(S') \quad \text{if } S > r \text{ and } S' > (1-\beta)r + \beta S. \tag{4.10}$$

Assume $\mathsf{P}(S)$. At a sufficiently large blocklength n, put $X = \mathcal{T}_n(C_n^{\mathrm{w}})$, $Y = \mathcal{T}_n(C_n^{\mathrm{s}})$, and $H = Y - X$. Then $\|X\|_\infty \le A_0$. Since $e_n \to 0$, we may also assume $e_n \le (1-\delta)/2$, giving $\|H\|_\infty \le \delta + e_n \le B_0$. Linearity expresses H by the auxiliary map $C_n^{\mathrm{s}} - C_n^{\mathrm{w}}$, whose norm is at most $2 \cdot 2^{nS/2}$ because $S > r$.

Expand $Y^{\otimes m} = (X+H)^{\otimes m}$, and keep only terms with at most $\lfloor \beta m \rfloor$ factors H. In detail, for $J \subseteq \{1, \ldots, m\}$ let $G_{J,j} = H$ if $j \in J$ and $G_{J,j} = X$ otherwise, and set

$$Z_{n,m} = \sum_{\substack{J \subseteq \{1,\ldots,m\} \\ |J| \le \lfloor \beta m \rfloor}} \bigotimes_{j=1}^{m} G_{J,j}. \tag{4.11}$$

All summands have the same domain and codomain. The expansion selects X or H independently on each tensor factor. Using $\binom{m}{j} \le 2^{m h_2(j/m)}$ and (4.9), the omitted terms obey

$$\begin{aligned} \|Z_{n,m} - Y^{\otimes m}\|_\infty &\le \sum_{j > \beta m} \binom{m}{j} A_0^{m-j} B_0^j \\ &\le (m+1) e^{-\kappa m}. \end{aligned}$$

To represent $Z_{n,m}$ by an auxiliary map, recall that $X = \mathcal{T}_n(C_n^{\mathrm{w}})$ and $H = \mathcal{T}_n(C_n^{\mathrm{s}} - C_n^{\mathrm{w}})$. For each subset $J \subseteq \{1, \ldots, m\}$, define

$$A_{J,j} = \begin{cases} C_n^{\mathrm{s}} - C_n^{\mathrm{w}}, & j \in J, \\ C_n^{\mathrm{w}}, & j \notin J, \end{cases} \qquad D_{n,m} = \sum_{\substack{J \subseteq \{1,\ldots,m\} \\ |J| \le \lfloor \beta m \rfloor}} \bigotimes_{j=1}^{m} A_{J,j}.$$

Each tensor product is a map from $\mathsf{F}^{\otimes nm}$ to $\mathsf{E}^{\otimes nm}$. By (4.7), applying $\mathcal{T}_{nm}$ to this tensor product gives $\bigotimes_{j=1}^m G_{J,j}$, the corresponding summand in (4.11). Linearity then gives $\mathcal{T}_{nm}(D_{n,m}) = Z_{n,m}$. A subset of size j contributes $m - j$ factors with norm at most $2^{nr/2}$ and j factors with norm at most $2 \cdot 2^{nS/2}$. Multiplying these bounds gives $2^j 2^{n[(m-j)r + jS]/2}$ for each tensor product. There are $\binom{m}{j}$ subsets of size j, so the triangle inequality gives

$$\begin{aligned} \|D_{n,m}\|_\infty &\le \sum_{j=0}^{\lfloor \beta m \rfloor} \binom{m}{j} 2^j 2^{n[(m-j)r + jS]/2} \\ &\le 2^{nm[(1-\beta)r + \beta S]/2} \sum_{j=0}^{\lfloor \beta m \rfloor} \binom{m}{j} 2^j \\ &\le 3^m 2^{nm[(1-\beta)r + \beta S]/2}. \end{aligned} \tag{4.12}$$

The second inequality uses $S > r$ and $j \le \beta m$ to bound $(m-j)r + jS \le m[(1-\beta)r + \beta S]$. The last inequality extends the sum to $j = m$ and uses $\sum_{j=0}^m \binom{m}{j} 2^j = 3^m$.

The truncation estimate above bounds $\|Z_{n,m} - Y^{\otimes m}\|_\infty$. To bound the error relative to the target $V^{\otimes nm}$, we must also control $\|Y^{\otimes m} - (V^{\otimes n})^{\otimes m}\|_\infty$, since Y only approximates $V^{\otimes n}$ with error

e_n. The telescoping expansion of this difference has m terms, each with one factor $Y - V^{\otimes n}$ and all other factors of norm at most $1 + e_n$. Hence the total error is

$$\|\mathcal{T}_{nm}(D_{n,m}) - V^{\otimes nm}\|_\infty \le (m+1)e^{-\kappa m} + me_n(1+e_n)^{m-1}. \tag{4.13}$$

Step 3: Cover every sufficiently large blocklength. For a large integer N, choose $n = \lfloor\sqrt{N}\rfloor$, $m = \lfloor N/n\rfloor$, and $\ell = N - nm$, so $0 \le \ell < n$. Pad with the exact auxiliary map: $C_N = D_{n,m} \otimes C_*^{\otimes \ell}$. The associated operator is $Z_{n,m} \otimes V^{\otimes \ell}$; since V is an isometry, padding does not increase the error. The empty tensor power is interpreted as the scalar identity. Writing $\bar{S} = (1-\beta)r + \beta S$, (4.12) gives

$$\|C_N\|_\infty^2 \le 3^{2m} 2^{nm\bar{S} + \ell L}, \qquad \frac{nm\bar{S} + \ell L + 2m\log 3}{N} = \bar{S} + o(1).$$

The last identity uses $nm/N \to 1$, $\ell/N \to 0$, and $m/N \to 0$. Thus $\|C_N\|_\infty \le 2^{NS'/2}$ for every fixed $S' > \bar{S}$ and all sufficiently large N.

We bound the two error terms in (4.13) separately. Since $n = \lfloor\sqrt{N}\rfloor$ and $m = \lfloor N/n\rfloor$, both are bounded above and below by positive constant multiples of $\sqrt{N}$ for all large N. In particular, $n^a \ge bN^{a/2}$ for some $b > 0$, and

$$me_n \le mKe^{-\gamma n^a} \le K_1\sqrt{N}\, e^{-\gamma b N^{a/2}} \longrightarrow 0.$$

Thus $(1+e_n)^{m-1} \le e^{me_n}$ is eventually at most 2. The second error term therefore satisfies

$$me_n(1+e_n)^{m-1} \le 2K_1\sqrt{N}\, e^{-\gamma b N^{a/2}} \le K_2 e^{-(\gamma b/2)N^{a/2}}$$

for all large N. The last inequality uses $\log N = o(N^{a/2})$, so the factor $\sqrt{N}$ can be absorbed by halving the decay constant.

For the first error term, m is of order $\sqrt{N}$, so $(m+1)e^{-\kappa m} \le K_3 e^{-\kappa_1\sqrt{N}}$ for suitable $K_3, \kappa_1 > 0$, by the same argument for the polynomial prefactor. Since $a \le 1$, we have $\sqrt{N} \ge N^{a/2}$, giving the bound $K_3 e^{-\kappa_1 N^{a/2}}$. Adding the two bounds and recalling that exact padding preserves the error yields

$$\left\|\mathcal{T}_N(C_N) - V^{\otimes N}\right\|_\infty \le K' e^{-\gamma' N^{a/2}}$$

for suitable $K', \gamma' > 0$. Hence the new approximation satisfies $\mathsf{P}(S')$ with exponent $a/2$, proving (4.10).

Step 4: Iterate a fixed number of times. Let $\theta = (1+\beta)/2 \in (\beta, 1)$ and $S_j = r + \theta^j(L-r)$ for $j = 0, 1, \ldots$. These rates start at $S_0 = L$ and decrease to r. Moreover,

$$S_{j+1} - \left[(1-\beta)r + \beta S_j\right] = (\theta - \beta)\theta^j(L-r) > 0,$$

so each successive pair meets the strict rate condition in (4.10). Starting from the exact representation, which gives $\mathsf{P}(L)$ with exponent $a = 1$, we may therefore apply that implication repeatedly. Each application halves the error exponent, so after j iterations we have $\mathsf{P}(S_j)$ with $a = 2^{-j}$.

Now fix a target rate $r < R < L$. Since $\theta^j(L-r) \to 0$, a finite j satisfies $\theta^j(L-r) < R - r$, or equivalently $S_j < R$. The auxiliary maps supplied by $\mathsf{P}(S_j)$ then satisfy $\|C_N\|_\infty^2 \le 2^{NS_j} \le 2^{NR}$ and have error at most $Ke^{-\gamma N^{2^{-j}}}$ for all sufficiently large N. We choose this j before letting N grow. Thus $2^{-j} > 0$ is a fixed exponent, and the constants and initial blocklength may depend on R but not on N. If $R \ge L$, the exact auxiliary maps already satisfy the rate bound with zero error. This proves (4.8). $\square$

4.3 Exponential accuracy from a fixed block

We continue with the isometries V, W and environment spaces of Section 4.2. The preceding lemma gives approximations whose error tends to zero at every rate above the low rate. We now use this vanishing error to obtain exponential decay, allowing an arbitrarily small further increase in the cost rate.

Choose one sufficiently accurate approximation X at blocklength k and keep k fixed as the total blocklength grows. The exact auxiliary map provides a representation of the residual $H = V^{\otimes k} - X$, so that

$$(X + H)^{\otimes m} = V^{\otimes km}.$$

Thus the full expansion equals the target exactly. Truncating it introduces only the omitted-tail error; there is no additional error from repeating an approximate target, as in (4.13). Since $\|H\|_\infty$ can be made arbitrarily small by choosing k large enough, we can retain terms with only a small fraction of residual factors while keeping the tail exponentially small in m. This limits the increase in cost rate. With k fixed, exponential decay in m also gives exponential decay in the total blocklength.

Lemma 4.5 (Fixed-block exponential amplification)**.** *Suppose $V = \mathcal{T}_1(C_*)$ with $\|C_*\|_\infty \le 2^{L/2}$ for some $L \ge 0$. At a rate $S \ge 0$, assume that there exist auxiliary maps A_n for all sufficiently large n, with $\|A_n\|_\infty \le 2^{nS/2}$ and $\varepsilon_n = \|\mathcal{T}_n(A_n) - V^{\otimes n}\|_\infty \to 0$.*

Then for every $R > S$, there exist $K, \gamma > 0$ such that, for all sufficiently large N, there are auxiliary maps C_N satisfying

$$\|C_N\|_\infty \le 2^{NR/2}, \qquad \left\|\mathcal{T}_N(C_N) - V^{\otimes N}\right\|_\infty \le K e^{-\gamma N}. \tag{4.14}$$

Proof. If $R \ge L$, the exact auxiliary maps suffice, so assume $S < R < L$. Choose $\eta \in (0,1)$ with $\bar{R} = S + \eta(L - S) < R$, and set $z = 4^{1/\eta}$. Since $\varepsilon_n \to 0$, there is a blocklength k large enough that $\varepsilon_k \le 1/(1+z)$ and $v := \bar{R} + 2\log 3/k < R$. Keep k fixed as $N \to \infty$.

Put $X = \mathcal{T}_k(A_k)$ and $H = V^{\otimes k} - X = \mathcal{T}_k(C_*^{\otimes k} - A_k)$. Then $\|X\|_\infty \le 1 + \varepsilon_k$, $\|H\|_\infty \le \varepsilon_k$, and $(X+H)^{\otimes m} = V^{\otimes km}$ exactly. As in (4.11), let Z_m contain the terms containing at most $\lfloor \eta m \rfloor$ factors H. The difference from $V^{\otimes km}$ is the omitted sum of terms with $j > \eta m$ residual factors. There are $\binom{m}{j}$ terms with exactly j such factors, each of norm at most $(1+\varepsilon_k)^{m-j}\varepsilon_k^j$. Hence

$$\begin{aligned}
\left\|Z_m - V^{\otimes km}\right\|_\infty &\le \sum_{j > \eta m} \binom{m}{j} (1+\varepsilon_k)^{m-j} \varepsilon_k^j \\
&\le z^{-\eta m} \sum_{j=0}^{m} \binom{m}{j} (1+\varepsilon_k)^{m-j} (z\varepsilon_k)^j \\
&= z^{-\eta m} (1 + \varepsilon_k + z\varepsilon_k)^m \le 2^{-m}.
\end{aligned}$$

For the second inequality, $z > 1$ and $j > \eta m$ imply $\varepsilon_k^j \le z^{-\eta m}(z\varepsilon_k)^j$. We then extend the sum to all $0 \le j \le m$, adding only nonnegative terms, and apply the binomial theorem. Finally, our choices $z = 4^{1/\eta}$ and $\varepsilon_k \le 1/(1+z)$ give $z^{-\eta} = 1/4$ and $1 + (1+z)\varepsilon_k \le 2$. The resulting bound is therefore $(\frac{1}{4} \cdot 2)^m = 2^{-m}$.

Replace X by A_k and H by $C_*^{\otimes k} - A_k$ in the same truncated sum to obtain an auxiliary map D_m with $\mathcal{T}_{km}(D_m) = Z_m$. Since $S < L$, the auxiliary map $C_*^{\otimes k} - A_k$ representing the correction has norm at most $2 \cdot 2^{kL/2}$. Therefore

$$\begin{aligned}\|D_m\|_\infty &\le \sum_{j \le \eta m} \binom{m}{j} 2^j 2^{k[(m-j)S + jL]/2} \\ &\le 3^m 2^{km\bar{R}/2} = 2^{kmv/2}.\end{aligned}$$

For $N = km + \ell$ with $0 \le \ell < k$, set $C_N = D_m \otimes C_*^{\otimes \ell}$. Exact padding preserves the error, and $\|C_N\|_\infty^2 \le 2^{kmv+\ell L} = 2^{Nv+\ell(L-v)} \le 2^{NR}$ for all sufficiently large N: the bounded remainder $\ell(L-v)$ is absorbed by $N(R-v) > 0$. Finally, $m = \lfloor N/k \rfloor$ implies $2^{-m} \le 2e^{-(\ln 2)N/k}$. This proves (4.14) with $K = 2$ and $\gamma = (\ln 2)/k$. □

We can now complete the approximation theorem. The rate margins are chosen before the blocklength, so each application of amplification uses constants independent of the total number of uses.

Proof of Theorem 2.2. Let L be as in Lemma 3.1. Corollary 3.3 provides C_* with $\|C_*\|_\infty \le 2^{L/2}$. For a target rate $S \ge L$, the auxiliary maps $C_*^{\otimes n}$ are exact and have the required norm. Otherwise $d < S < L$. Fix rates $d < r < S_0 < S$ and choose $\delta \in (\sqrt{d/r}, 1)$. Combining Lemma 4.3 with Proposition 4.2 gives the required constant-error approximation at rate r. Apply Lemma 4.4 with $V = V_N$, $W = V_M$, and target rate S_0 to obtain vanishing error at that rate. Lemma 4.5, with input rate S_0 and target rate S, upgrades the error to an exponential. Both lemmas cover every sufficiently large blocklength and preserve the prescribed tensor-power environments. The exact case covers $d = L$ as well, including $L = 0$. □

5 Testing and Smoothing Consequences

The approximation theorem gives more than the fixed-error testing limit. It also identifies the threshold of the optimized testing tail and the asymptotic cost of approximating the null channel by a subchannel dominated by the alternative. Throughout this section, $\mathcal{N}, \mathcal{M} : A \to B$ are channels with $d = D^\infty(\mathcal{N}\|\mathcal{M}) < \infty$.

5.1 A sharp testing threshold

Below d, block repetition produces tests whose null acceptance tends to one. Above d, the parallel converse forces that acceptance to vanish whenever the testing objective is positive. These two facts locate the threshold of the hockey-stick divergence.

Corollary 5.1 (Hockey-stick threshold)**.** *Let $\mathcal{N}, \mathcal{M} : A \to B$ be channels with $d = D^\infty(\mathcal{N}\|\mathcal{M}) < \infty$, and set $E_n(r) = E_{2^{nr}}(\mathcal{N}^{\otimes n}\|\mathcal{M}^{\otimes n})$ for $r \ge 0$, where E_t is defined in (4.1). Then*

$$\lim_{n\to\infty} E_n(r) = \begin{cases} 1, & 0 \le r < d, \\ 0, & r > d. \end{cases}$$

For each fixed $r > d$, convergence to zero is exponential.

Proof. A positive objective value $p-2^{nr}q$ implies $q<2^{-nr}$ and is at most p. For $r>d$, Proposition 3.5 therefore bounds every such value by $K_r e^{-\gamma_r n}$ for all large n. Taking the supremum proves the exponential upper bound.

For $0\le r<d$, choose a blocklength k and a pure input whose output states ρ,σ satisfy $D(\rho\|\sigma)/k>r$. Fix u with $kr<u<D(\rho\|\sigma)$. The direct part of the state Stein lemma gives tests on m copies with $p_m\to 1$ and $q_m\le 2^{-mu}$ for all sufficiently large m. At total blocklength $n=mk+\ell$, $0\le\ell<k$, use the padding construction from (3.3). It follows that $E_n(r)\ge p_m-2^{nr-mu}\to 1$, because $mu/n\to u/k>r$. Together with $E_n(r)\le 1$, this proves the lower-rate limit. □

No conclusion is asserted at the critical rate $r=d$.

5.2 Uniform approximation by dominated subchannels

A subchannel is a completely positive trace-nonincreasing map. For a Hermiticity-preserving map $\Phi:\mathcal{L}(A)\to\mathcal{L}(B)$, its diamond norm can be evaluated as

$$\|\Phi\|_\diamond=\sup_{\substack{|\psi\rangle\in R\otimes A\\ \||\psi\rangle\|=1}}\|(\mathrm{id}_R\otimes\Phi)(|\psi\rangle\langle\psi|)\|_1\,,$$

where $R\simeq A$ suffices [20, 29]. We use the unhalved norm throughout. Thus this distance controls all inputs and references simultaneously.

The approximating dilation need not be a contraction. A rescaling makes it one while preserving the domination rate and changing the error by only a constant factor.

Corollary 5.2 (Dominated subchannels)**.** *Let $\mathcal{N},\mathcal{M}:A\to B$ be channels with $d=D^\infty(\mathcal{N}\|\mathcal{M})<\infty$. For every $S>d$, there are $K,\gamma>0$ and subchannels $\mathcal{L}_n:A^n\to B^n$, for all sufficiently large n, such that*

$$\mathcal{L}_n\le_{\mathrm{CP}}2^{nS}\mathcal{M}^{\otimes n},\qquad \|\mathcal{L}_n-\mathcal{N}^{\otimes n}\|_\diamond\le Ke^{-\gamma n}.\tag{5.1}$$

Proof. Fix dilations and take C_n,e_n from Theorem 2.2 at rate S. Put $\widehat{V}_n=(I_{B^n}\otimes C_n)V_M^{\otimes n}$ and $\widetilde{V}_n=\widehat{V}_n/(1+e_n)$. Since $\left\|\widehat{V}_n\right\|_\infty\le 1+e_n$, the operator $\widetilde{V}_n$ is a contraction. It defines a CP map $\mathcal{L}_n(X)=\mathrm{Tr}_{\mathsf{E}_N^{\otimes n}}\widetilde{V}_nX\widetilde{V}_n^\dagger$. For $X\ge 0$, $\mathrm{Tr}\,\mathcal{L}_n(X)=\mathrm{Tr}\,X\widetilde{V}_n^\dagger\widetilde{V}_n\le\mathrm{Tr}\,X$, so this map is a subchannel.

Let $F_{M,n}$ be the Kraus-column matrix of $V_M^{\otimes n}$ and $D_n=(C_n/(1+e_n))^{\mathsf{T}}$. Then

$$J_{\mathcal{L}_n}=F_{M,n}D_nD_n^\dagger F_{M,n}^\dagger\le 2^{nS}F_{M,n}F_{M,n}^\dagger=2^{nS}J_{\mathcal{M}^{\otimes n}}.$$

This proves the CP domination. For the error, the identity $(1+e_n)V_N^{\otimes n}-\widehat{V}_n=e_nV_N^{\otimes n}+(V_N^{\otimes n}-\widehat{V}_n)$ gives $\left\|V_N^{\otimes n}-\widetilde{V}_n\right\|_\infty\le 2e_n/(1+e_n)\le 2e_n$. For a pure input with any reference, let x,y be the output vectors of these two operators. They have norm at most one and $\|x-y\|\le 2e_n$. Decomposing their rank-one difference yields

$$\||x\rangle\langle x|-|y\rangle\langle y|\|_1\le(\|x\|+\|y\|)\|x-y\|\le 4e_n.$$

The partial trace is trace-norm contractive on Hermitian operators. Taking that trace and optimizing over inputs therefore gives $\|\mathcal{L}_n-\mathcal{N}^{\otimes n}\|_\diamond\le 4e_n$. Absorbing the factor four into the prefactor proves (5.1). □

We express the resulting rate through max-relative entropy. Following the order-based definition for states [16] and its channel extension [11], for a nonzero CP map Φ and a channel Ψ on the same spaces set $D_{\max}(\Phi\|\Psi) = \log\inf\{t > 0 : \Phi \le_{\mathrm{CP}} t\Psi\}$, with value $+\infty$ if no finite t is feasible. For $0 < \varepsilon < 1$, define

$$D^{\varepsilon,\le}_{\max}(\mathcal{N}\|\mathcal{M}) = \inf_{\substack{\mathcal{L} \text{ a subchannel} \\ \|\mathcal{L}-\mathcal{N}\|_\diamond \le \varepsilon}} D_{\max}(\mathcal{L}\|\mathcal{M}). \tag{5.2}$$

The superscript $\le$ records that trace loss is allowed. Since $\|\mathcal{N}\|_\diamond = 1$, the condition $\varepsilon < 1$ excludes the zero map; finite values of (5.2) can nevertheless be negative. This is a channel analogue of the smooth-entropy asymptotics for states [17,18], with the normalization constraint on the approximant treated separately [19].

The dominated subchannels give the upper bound on the smoothing rate. For the lower bound, closeness to the null channel forces every feasible approximant to preserve the acceptance probability of a good test, up to the smoothing error. Together these observations give the following asymptotic identity.

Corollary 5.3 (Subchannel asymptotic equipartition)**.** *Let $\mathcal{N},\mathcal{M} : A \to B$ be channels with $d = D^\infty(\mathcal{N}\|\mathcal{M}) < \infty$. For every $0 < \varepsilon < 1$,*

$$\lim_{n\to\infty} \frac{1}{n} D^{\varepsilon,\le}_{\max}(\mathcal{N}^{\otimes n}\|\mathcal{M}^{\otimes n}) = d. \tag{5.3}$$

The converse follows from a one-shot estimate: for every $n \ge 1$ and $0 < \eta < 1-\varepsilon$, the quantity $D^{\varepsilon,\le}_{\max}(\mathcal{N}^{\otimes n}\|\mathcal{M}^{\otimes n})$ is at least $-\log\beta^*_n(\eta) + \log(1-\eta-\varepsilon)$. We establish this bound as part of the proof.

Proof. Let $\mathcal{L}$ be a feasible subchannel in (5.2), and take any $t > 0$ with $\mathcal{L} \le_{\mathrm{CP}} t\mathcal{M}^{\otimes n}$. For any pure input $|\psi\rangle$ and effect T, let p, q be the acceptance probabilities under $\mathcal{N}^{\otimes n}, \mathcal{M}^{\otimes n}$. Diamond-norm closeness and CP domination imply

$$p - \varepsilon \le \operatorname{Tr} T(\mathrm{id}_R \otimes \mathcal{L})(|\psi\rangle\langle\psi|) \le tq.$$

If $p \ge 1-\eta$, then $q \ge (1-\eta-\varepsilon)/t$. Taking the infimum over all such inputs and tests gives $\beta^*_n(\eta) \ge (1-\eta-\varepsilon)/t$, and therefore $\log t \ge -\log\beta^*_n(\eta) + \log(1-\eta-\varepsilon)$. Taking infima over feasible t and then over $\mathcal{L}$, with the convention $\inf\varnothing = +\infty$, proves the one-shot estimate. Fixing $\eta = (1-\varepsilon)/2$ and applying Proposition 3.5 gives the asymptotic lower bound d.

For the upper bound, take any $S > d$. Corollary 5.2 gives, at all sufficiently large n, a feasible subchannel with max-relative entropy at most nS. Thus the limit superior is at most S. Let $S \downarrow d$ to obtain (5.3). □

The same argument allows a varying smoothing radius. If its logarithmic inverse is sublinear in n, the exponentially accurate approximants are eventually feasible at that radius.

Corollary 5.4 (Subexponential smoothing errors)**.** *Let $\mathcal{N},\mathcal{M} : A \to B$ be channels with $d = D^\infty(\mathcal{N}\|\mathcal{M}) < \infty$. Suppose $0 < \varepsilon_n \le \varepsilon_* < 1$ for a fixed ε_*, and $\log(1/\varepsilon_n) = o(n)$. Then*

$$\lim_{n\to\infty} \frac{1}{n} D^{\varepsilon_n,\le}_{\max}(\mathcal{N}^{\otimes n}\|\mathcal{M}^{\otimes n}) = d. \tag{5.4}$$

Proof. Since $\varepsilon_n \le \varepsilon_*$, the feasible set at radius ε_n is contained in that at radius ε_*. Corollary 5.3 therefore gives the lower bound. For any $S > d$, Corollary 5.2 supplies subchannels with domination 2^{nS} and error $Ke^{-\gamma n}$. The assumption on ε_n implies $\ln(1/\varepsilon_n) = o(n)$; hence $Ke^{-\gamma n} \le \varepsilon_n$ for all sufficiently large n. These subchannels are feasible at radius ε_n, giving a limit superior at most S. Letting $S \downarrow d$ proves the claim. □

The sequence of radii need not be monotone or converge to zero. Trace loss, however, is essential to the class of approximants used here. Rescaling gives a contraction, and completing the associated subchannel to a trace-preserving map need not preserve the domination rate. The asymptotic equipartition property with trace-preserving channel smoothing in fact fails in general [19].

6 Discussion

We have established $D^{\infty}(\mathcal{N}\|\mathcal{M})$ as the fixed-error Stein threshold under arbitrary adaptive control. The proof also gives a way to pass from a weak testing bound to exponential accuracy without first identifying the optimizing input states. Once the weak bound is converted into a uniform environmental approximation, finite CP domination supplies an exact representation of the residual. The tensor constructions can then improve the error at an arbitrarily small increase in cost rate. This explains how the same weak estimate leads to the parallel strong converse, Rényi continuity, and the adaptive converse.

The main quantitative question is how the approximation error depends on the rate gap above d. In Lemma 4.5, the error exponent is $\gamma = (\ln 2)/k$, where the blocklength k is chosen after fixing the rate gap. This proves positivity for every fixed gap, but the construction does not optimize the required blocklength. Sharper control of this dependence would give quantitative continuity bounds through (3.6), and hence explicit lower bounds on the strong-converse exponent near the threshold. It would also give bounds on the domination rate needed for a prescribed exponential smoothing error, extending the regime covered by Corollary 5.4.

Acknowledgment

OpenAI Codex was used in proof development, mathematical checking, literature work, and preparation of the source manuscript, and in expanding, reorganizing, and typesetting the present exposition.

References

[1] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ, USA: Wiley, 2006, doi: 10.1002/047174882X.

[2] R. E. Blahut, “Hypothesis testing and information theory,” *IEEE Trans. Inf. Theory*, vol. 20, no. 4, pp. 405–417, 1974, doi: 10.1109/TIT.1974.1055254.

[3] F. Hiai and D. Petz, “The proper formula for relative entropy and its asymptotics in quantum probability,” *Commun. Math. Phys.*, vol. 143, pp. 99–114, 1991, doi: 10.1007/BF02100287.

[4] T. Ogawa and H. Nagaoka, “Strong converse and Stein’s lemma in quantum hypothesis testing,” *IEEE Trans. Inf. Theory*, vol. 46, no. 7, pp. 2428–2433, 2000, doi: 10.1109/18.887855.

[5] M. Mosonyi and T. Ogawa, "Quantum hypothesis testing and the operational interpretation of the quantum Rényi relative entropies," *Commun. Math. Phys.*, vol. 334, no. 3, pp. 1617–1648, 2015, doi: 10.1007/s00220-014-2248-x.

[6] K. Li, "Second-order asymptotics for quantum hypothesis testing," *Ann. Statist.*, vol. 42, no. 1, pp. 171–189, 2014, doi: 10.1214/13-AOS1185.

[7] M. Tomamichel and M. Hayashi, "A hierarchy of information quantities for finite block length analysis of quantum tasks," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7693–7710, 2013, doi: 10.1109/TIT.2013.2276628.

[8] M. Hayashi, "Discrimination of two channels by adaptive methods and its application to quantum system," *IEEE Trans. Inf. Theory*, vol. 55, no. 8, pp. 3807–3820, 2009, doi: 10.1109/TIT.2009.2023726.

[9] T. Cooney, M. Mosonyi, and M. M. Wilde, "Strong converse exponents for a quantum channel discrimination problem and quantum-feedback-assisted communication," *Commun. Math. Phys.*, vol. 344, pp. 797–829, 2016, doi: 10.1007/s00220-016-2645-4.

[10] M. M. Wilde, M. Berta, C. Hirche, and E. Kaur, "Amortized channel divergence for asymptotic quantum channel discrimination," *Lett. Math. Phys.*, vol. 110, pp. 2277–2336, 2020, doi: 10.1007/s11005-020-01297-7.

[11] X. Wang and M. M. Wilde, "Resource theory of asymmetric distinguishability for quantum channels," *Phys. Rev. Research*, vol. 1, Art. no. 033169, 2019, doi: 10.1103/PhysRevResearch.1.033169.

[12] K. Fang, O. Fawzi, R. Renner, and D. Sutter, "Chain rule for the quantum relative entropy," *Phys. Rev. Lett.*, vol. 124, Art. no. 100501, 2020, doi: 10.1103/PhysRevLett.124.100501.

[13] B. Bergh, N. Datta, R. Salzmann, and M. M. Wilde, "Parallelization of adaptive quantum channel discrimination in the non-asymptotic regime," *IEEE Trans. Inf. Theory*, vol. 70, no. 4, pp. 2617–2636, 2024, doi: 10.1109/TIT.2024.3355929.

[14] H. Fawzi and O. Fawzi, "Defining quantum divergences via convex optimization," *Quantum*, vol. 5, Art. no. 387, 2021, doi: 10.22331/q-2021-01-26-387.

[15] K. Fang, G. Gour, and X. Wang, "Towards the ultimate limits of quantum channel discrimination and quantum communication," *Sci. China Inf. Sci.*, vol. 68, Art. no. 180509, 2025, doi: 10.1007/s11432-024-4488-0.

[16] N. Datta, "Min- and max-relative entropies and a new entanglement monotone," *IEEE Trans. Inf. Theory*, vol. 55, no. 6, pp. 2816–2826, 2009, doi: 10.1109/TIT.2009.2018325.

[17] M. Tomamichel, R. Colbeck, and R. Renner, "A fully quantum asymptotic equipartition property," *IEEE Trans. Inf. Theory*, vol. 55, no. 12, pp. 5840–5847, 2009, doi: 10.1109/TIT.2009.2032797.

[18] N. Datta, M. Mosonyi, M.-H. Hsieh, and F. G. S. L. Brandão, "A smooth entropy approach to quantum hypothesis testing and the classical capacity of quantum channels," *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 8014–8026, 2013, doi: 10.1109/TIT.2013.2282160.

[19] G. Gour, "Failure of the asymptotic equipartition property for quantum channels," 2026, arXiv:2609.15004v1. [Online]. Available: https://arxiv.org/abs/2609.15004v1

[20] J. Watrous, *The Theory of Quantum Information*, Cambridge, U.K.: Cambridge University Press, 2018. [Online]. Available: https://cs.uwaterloo.ca/~watrous/TQI/

[21] W. F. Stinespring, "Positive functions on C^*-algebras," *Proc. Amer. Math. Soc.*, vol. 6, no. 2, pp. 211–216, 1955, doi: 10.1090/S0002-9939-1955-0069403-4.

[22] D. Kretschmann, D. Schlingemann, and R. F. Werner, "The information-disturbance tradeoff and the continuity of Stinespring's representation," *IEEE Trans. Inf. Theory*, vol. 54, no. 4, pp. 1708–1717, 2008, doi: 10.1109/TIT.2008.917696.

[23] M.-D. Choi, "Completely positive linear maps on complex matrices," *Linear Algebra Appl.*, vol. 10, no. 3, pp. 285–290, 1975, doi: 10.1016/0024-3795(75)90075-0.

[24] R. G. Douglas, “On majorization, factorization, and range inclusion of operators on Hilbert space,” *Proc. Amer. Math. Soc.*, vol. 17, pp. 413–415, 1966, doi: 10.1090/S0002-9939-1966-0203464-1.

[25] M. Müller-Lennert, F. Dupuis, O. Szehr, S. Fehr, and M. Tomamichel, “On quantum Rényi entropies: A new generalization and some properties,” *J. Math. Phys.*, vol. 54, Art. no. 122203, 2013, doi: 10.1063/1.4838856.

[26] M. M. Wilde, A. Winter, and D. Yang, “Strong converse for the classical capacity of entanglement-breaking and Hadamard channels via a sandwiched Rényi relative entropy,” *Commun. Math. Phys.*, vol. 331, no. 2, pp. 593–622, 2014, doi: 10.1007/s00220-014-2122-x.

[27] S. Beigi, “Sandwiched Rényi divergence satisfies data processing inequality,” *J. Math. Phys.*, vol. 54, Art. no. 122202, 2013, doi: 10.1063/1.4838855.

[28] R. L. Frank and E. H. Lieb, “Monotonicity of a relative Rényi entropy,” *J. Math. Phys.*, vol. 54, Art. no. 122201, 2013, doi: 10.1063/1.4838835.

[29] J. Watrous, “Semidefinite programs for completely bounded norms,” *Theory Comput.*, vol. 5, pp. 217–238, 2009, doi: 10.4086/toc.2009.v005a011.

[30] N. Sharma and N. A. Warsi, “On the strong converses for the quantum channel capacity theorems,” arXiv:1205.1712, 2012. [Online]. Available: https://arxiv.org/abs/1205.1712

[31] T. Nuradha, V. Singh, and M. M. Wilde, “Measured hockey-stick divergence and its applications to quantum pufferfish privacy,” arXiv:2501.12359, 2025. [Online]. Available: https://arxiv.org/abs/2501.12359