Causal postselection and the Stein exponent for quantum superchannels

Yuxuan Zhang

13 November 2026 Version 1.0

Abstract

We study binary discrimination of finite-dimensional quantum superchannels when the two components of different uses may be interleaved in any causally valid order and their internal memories remain inaccessible. We prove that the general, nested-adaptive and fully parallel Stein exponents coincide with the regularized superchannel relative entropy. Above this rate, null acceptance decays exponentially at every fixed type-I tolerance below one. The proof combines polynomial postselection for causal two-stage combs with a common-amplitude approximation measured in a strategy norm. Canonical covariant dilations ensure that the postselection atoms are physical two-combs. A positive-slack minimax argument supplies one approximation valid for all parallel testers; a tensor amplification argument then makes its error exponentially small. The result extends methods for ordinary channel discrimination while preserving the causal constraints of a superchannel. This is an internally reviewed complete proof candidate; independent expert review and comprehensive priority assessment are pending.

1 Operational question and main theorem

A superchannel maps channels $\mathcal{N} : A \to B$ to channels $C \to D$ and has a physical realization

$$\Theta(\mathcal{N}) = \mathcal{D} \circ (\mathcal{N} \otimes \mathrm{id}_S) \circ \mathcal{E}, \quad \mathcal{E} : C \to AS, \quad \mathcal{D} : BS \to D. \tag{1}$$

The memory S is internal. A tester may supply the external inputs, retain its own memory and choose the inserted channel, but cannot act on S. In a fully parallel strategy, all first components precede a joint inserted channel and all second components follow it. Nested strategies recursively insert complete uses into one another. General strategies may interleave the components of different uses, provided each first component precedes its matched second component [1].

Let α_n be the probability of rejecting the null and β_n the probability of accepting it under the alternative. For a strategy class S, define

$$\zeta_{\mathsf{S}} = \lim_{\varepsilon \downarrow 0} \liminf_{n \to \infty} -\frac{1}{n} \log_2 \inf_{P \in \mathsf{S}_n : \alpha_n(P) \le \varepsilon} \beta_n(P). \tag{2}$$

All rate logarithms below are base two. The slot dimensions are fixed and finite; a tester's finite ancillary dimensions are unrestricted.

Write $D_{\mathrm{sc}}(\Theta_1 \| \Theta_2)$ for the supremum of the output-state relative entropy over common one-use testers, allowing side wires. Define

$$d_0 = D_{\mathrm{sc}}^{\infty}(\Theta_1 \| \Theta_2) := \sup_{k \ge 1} \frac{1}{k} D_{\mathrm{sc}}(\Theta_1^{\otimes k} \| \Theta_2^{\otimes k}). \tag{3}$$

The output states may be obtained from any joint input and joint inserted channel. The parallel coding characterization of this rate is established in [1, Sec. 5.3].

Theorem 1. *For every finite-dimensional pair of deterministic superchannels with the operational access in (1),*

$$\zeta_{\mathrm{general}} = \zeta_{\mathrm{nested}} = \zeta_{\mathrm{parallel}} = d_0. \tag{4}$$

If $d_0 < \infty$, then for every $r' > d_0$ there are constants $K, \gamma > 0$ such that every general n-use test with $\beta_n \le 2^{-nr'}$ satisfies $1 - \alpha_n \le Ke^{-\gamma n}$ for all sufficiently large n.

This includes classical choices of later causal orders based on earlier outcomes. It concerns causally valid interleavings of the specified components, not an assumption that the tester accesses the internal memory.

For ordinary memoryless channels, Gao, Ji and Liu obtain an exponential adaptive converse [2], and Zhu and Wang treat general process-matrix testers [3]. Their approximation, positive-slack and postselection methods supply the starting point. Here an arbitrary ordinary channel on the combined slots cannot replace a superchannel: it might signal from the second input to the first output. The proof preserves the two-stage causal constraint inside every postselection atom and controls the approximation in the corresponding strategy norm.

2 Polynomial postselection preserving two-stage causality

We use unnormalized Choi operators and vectorizations. A deterministic joint n-block two-comb is a positive operator R_n with

$$\mathrm{Tr}_{D^n} R_n = I_{B^n} \otimes E_n, \qquad \mathrm{Tr}_{A^n} E_n = I_{C^n}. \tag{5}$$

The trace characterization and its realization by channels are standard [5, Theorem 6].

We first record the isometry form of fixed-marginal postselection from [4, Theorem 1 and Appendix A]. The normalization matters when it is applied to the two dilations.

Lemma 2. *Let $q \le p$ and let $V_n : (\mathbb{C}^q)^{\otimes n} \to (\mathbb{C}^p)^{\otimes n}$ be an isometry satisfying $V_n P_\pi^{(q)} = P_\pi^{(p)} V_n$ for every permutation. With Haar measure on one-copy isometries $V : \mathbb{C}^q \to \mathbb{C}^p$,*

$$|V_n\rangle\langle V_n| \le c_n(p,q) \int (|V\rangle\langle V|)^{\otimes n}\, dV, \qquad c_n(p,q) = \binom{n+pq-1}{pq-1}. \tag{6}$$

Proof. Schur–Weyl duality decomposes input and output into $Q_\lambda^{(q)} \otimes W_\lambda$ and $Q_\lambda^{(p)} \otimes W_\lambda$. Write their representation dimensions as $q_\lambda, p_\lambda, f_\lambda$. Covariance gives $V_n = \bigoplus_\lambda A_\lambda \otimes I_{W_\lambda}$, with $A_\lambda^* A_\lambda = I$. If $J : \mathbb{C}^q \to \mathbb{C}^p$ is a fixed inclusion, the Haar isometry is UJ, and Schur orthogonality gives the corresponding block of the average Ω_n as

$$\frac{I_{p_\lambda}}{p_\lambda} \otimes I_{q_\lambda} \otimes |I_{W_\lambda}\rangle\langle I_{W_\lambda}|. \tag{7}$$

Its eigenvalue on the last support is f_λ / p_λ; the target component has squared norm $q_\lambda f_\lambda$. Hence

$$\langle V_n | \Omega_n^+ | V_n \rangle = \sum_\lambda p_\lambda q_\lambda = \dim \mathrm{Sym}^n(\mathbb{C}^p \otimes \mathbb{C}^q) = c_n(p,q).$$

The Cauchy decomposition gives the middle identity. Finally, $|v\rangle\langle v| \le \langle v, \Omega^+ v\rangle \Omega$ for v in the range of a positive Ω. This proves the result. In particular the Specht dimensions cancel; no minimum-weight bound is substituted. □

Proposition 3. *Let R_n satisfy (5) and be invariant under simultaneous permutations of complete external-slot tuples. Put $a = d_A, b = d_B, c = d_C, d = d_D$, $s = ca$ and $t = dbs$. There is a fixed probability measure on physical one-copy two-combs $R(V, W)$ such that*

$$R_n \le P(n) \int R(V,W)^{\otimes n}\, dV\, dW, \tag{8}$$

$$P(n) = \binom{n + (ca)^2 - 1}{(ca)^2 - 1} \binom{n + (cabd)^2 - 1}{(cabd)^2 - 1}. \tag{9}$$

The measure uses independent Haar isometries $V : C \to AS$ and $W : BS \to DT$, with $\dim S = s$ and $\dim T = t$.

Proof. Take the canonical square-root Choi purification of the channel E_n. It gives a Stinespring isometry $V_n : C^n \to A^n S^n$, with $S = CA$ per copy, including unused directions when the Choi rank is smaller. Since E_n commutes with complete-factor permutations, its positive square root does too. Permutation matrices are real, so vectorization gives

$$V_n P_\pi^C = (P_\pi^A \otimes P_\pi^S) V_n. \tag{10}$$

The strategy realization theorem gives a second channel $\mathcal{D}_n : B^n S^n \to D^n$ realizing R_n after this particular V_n. To see why the canonical dilation is admissible, first use a minimal Stinespring dilation of E_n. Every other dilation contains its memory by an isometry. Transport the second channel through that isometry and extend it by an arbitrary CPTP action on the unused memory complement.

Average $\mathcal{D}_n$ over matching input permutations of $B^n S^n$ and output permutations of D^n. Equation (10) and invariance of R_n imply that each conjugated channel still realizes R_n after V_n. The averaged channel is covariant. Its canonical Stinespring isometry $W_n : B^n S^n \to D^n T^n$, with $T = DBS$ per copy, is covariant by the same square-root argument.

Apply Lemma 2 with $(p, q) = (as, c)$ and (dt, bs). Tensor the two positive inequalities, contract the S^n legs with the fixed maximally entangled link vector and trace T^n. This contraction is a fixed CP map on the tensor product of the pure Choi operators. Each iid pair produces the physical comb $R(V, W)^{\otimes n}$, and the product of the two constants is exactly (9). □

The proposition is a finite-n domination statement. It does not assume that R_n extends symmetrically to arbitrarily many copies. Infinite exchangeability results for causal processes [6] and marginal or almost-iid channel approximations [7] have different quantifiers.

3 A strategy norm for common amplitudes

Let $\mathcal{S}_n$ be the deterministic joint two-combs in (5), and $\mathcal{T}_n$ their compact convex set of deterministic co-testers. Thus $\operatorname{Tr} TR = 1$ for every $T \in \mathcal{T}_n$ and $R \in \mathcal{S}_n$. Both sets contain full-rank elements. For $Y \ge 0$, positive-strategy duality [5, Theorem 9] gives

$$g_n(Y) := \max_{T \in \mathcal{T}_n} \operatorname{Tr} TY = \min\{\lambda \ge 0 : Y \le \lambda R,\ R \in \mathcal{S}_n\}. \tag{11}$$

The extension to any positive Y follows by positive scaling, since a sufficient multiple of a full-rank deterministic comb dominates Y. Compactness supplies the minimum, with $Y = 0$ handled separately.

For a rectangular matrix Z whose rows belong to the external-slot space, define

$$\|Z\|_n := \sqrt{g_n(ZZ^*)} = \max_{T\in\mathcal{T}_n} \|\sqrt{T}Z\|_{\mathrm{F}}. \tag{12}$$

The expression is a norm, and the full-rank tester makes it nondegenerate. It satisfies

$$\|ZC\|_n \le \|Z\|_n \|C\|_{\mathrm{op}}, \qquad \|Z\otimes Z'\|_{n+m} \le \|Z\|_n \|Z'\|_m. \tag{13}$$

For the tensor bound, dominate ZZ^* and $Z'Z'^*$ using (11); the product of their deterministic combs is again a deterministic joint comb. The norm is invariant under complete-slot permutations, and equals one if ZZ^* is deterministic.

Let R_1, R_2 be the one-copy Choi operators of the unknown superchannels, and put $A = \sqrt{R_1}, B = \sqrt{R_2}$ and $A_n = A^{\otimes n}, B_n = B^{\otimes n}$. If $\operatorname{supp} R_1 \subseteq \operatorname{supp} R_2$, choose $C_* = B^+ A$ and finite $L \ge 0$ with

$$A = BC_*, \qquad \|C_*\|_{\mathrm{op}} \le 2^{L/2}. \tag{14}$$

In particular $\|A_n\|_n = \|B_n\|_n = 1$.

4 One approximation for every parallel tester

Assume for now that $d_0 < \infty$ and that the support inclusion above holds. For $r > d_0$ define the parallel hockey-stick score

$$h_n(r) = \max_{T\in\mathcal{T}_n,\, 0\le Q\le T} \operatorname{Tr} Q(R_1^{\otimes n} - 2^{nr} R_2^{\otimes n}). \tag{15}$$

For the binary output probabilities p, q, data processing gives $D_{\mathrm{bin}}(p\|q) \le nd_0$. A nonnegative score implies $q \le p2^{-nr}$ and hence

$$D_{\mathrm{bin}}(p\|q) \ge pnr - 1 \ge (p - 2^{nr}q)nr - 1.$$

Thus

$$0 \le h_n(r) \le \frac{nd_0 + 1}{nr}. \tag{16}$$

The common positive slack is

$$h_n(r) = \min_{Y\ge 0,\, Y\ge R_1^{\otimes n} - 2^{nr} R_2^{\otimes n}} g_n(Y). \tag{17}$$

For a fixed T, SDP strong duality gives

$$\max_{0\le Q\le T} \operatorname{Tr} QX = \inf_{Y\ge 0,\, Y\ge X} \operatorname{Tr} TY. \tag{18}$$

A large multiple of identity is strictly dual feasible. The right side need not attain its infimum for singular T. Minimax nevertheless interchanges the compact convex tester maximum and the convex slack infimum. Afterwards, a full-rank $T_0 \in \mathcal{T}_n$ gives $g_n(Y) \ge \lambda_{\min}(T_0) \operatorname{Tr} Y$ for positive Y. A global minimizing sequence is bounded, and closedness gives the minimum in (17).

For a minimizing Y, the supported Douglas factorization of $A_n A_n^* \le 2^{nr} B_n B_n^* + Y$ supplies a contraction $K = (K_1; K_2)$ with

$$A_n = 2^{nr/2} B_n K_1 + \sqrt{Y} K_2.$$

Taking $C_n^{\mathrm{w}} = 2^{nr/2} K_1$ gives, for all sufficiently large n,

$$\|C_n^{\mathrm{w}}\|_{\mathrm{op}} \le 2^{nr/2}, \qquad \|A_n - B_n C_n^{\mathrm{w}}\|_n \le \sqrt{h_n(r)} \le \delta < 1. \tag{19}$$

This is a single matrix valid for all parallel testers.

5 Exponential amplification in the strategy norm

We adapt the tensor amplification of [2, Lemmas 4.4–4.5]. Its algebra uses the triangle inequality, tensor submultiplicativity, $\|A_n\|_n = 1$, the compatible linear maps $C \mapsto B_n C$, and the finite exact cost (14). These properties hold by (13); the cost norm remains the ordinary operator norm.

Lemma 4. *For every $R > d_0$ there are $K, \gamma > 0$ and matrices C_n such that, for all sufficiently large n,*

$$\|C_n\|_{\mathrm{op}} \le 2^{nR/2}, \qquad \|A_n - B_n C_n\|_n \le K e^{-\gamma n}. \tag{20}$$

Proof. Fix $d_0 < r < R$ and the weak approximation (19). If $r \ge L$, use the exact map. Otherwise put $a = 1 + \delta$ and $b = (1 + \delta)/2 < 1$. Choose $\theta \in (0, 1)$ sufficiently near one and $k_0 > 0$ such that, for $x \in [\theta, 1]$,

$$(\ln 2) h_2(x) + (1 - x) \ln a + x \ln b \le -k_0, \tag{21}$$

where h_2 is binary entropy in bits.

Suppose an accurate cost-S sequence $Y_n = B_n C_n^{\mathrm{s}}$ has error $e_n \le K e^{-\gamma n^u}$, $u > 0$, while $X_n = B_n C_n^{\mathrm{w}}$ is the weak cost-r sequence. Initially take $S = L$ and exact error zero. For large n, $H_n = Y_n - X_n$ obeys $\|H_n\|_n \le \delta + e_n \le b$ and $\|X_n\|_n \le a$. Expand $(X_n + H_n)^{\otimes m}$ and retain terms with at most $\lfloor \theta m \rfloor$ factors H_n. Equation (21) bounds the omitted norm by $(m + 1) e^{-k_0 m}$. Every retained term has a preimage under $C \mapsto B_{nm} C$. Summing its cost gives

$$\|C_{nm}\|_{\mathrm{op}} \le 3^m 2^{nm((1-\theta)r + \theta S)/2}. \tag{22}$$

The error from $Y_n^{\otimes m}$ to A_{nm} is at most $m e_n (1 + e_n)^{m-1}$, by telescoping.

For general large N, take $n = \lfloor \sqrt{N} \rfloor$, $m = \lfloor N/n \rfloor$ and pad the fewer than n remaining factors with the exact map. The rate overhead tends to zero and the error stays stretched exponential, with power at least $u/2$. Thus an accurate cost S implies every accurate cost $S' > (1 - \theta) r + \theta S$. A finite number of iterations, fixed before N grows, reaches any prescribed cost above r.

To obtain exponential error, choose $r < S < R$ and one sufficiently accurate fixed block of length k at cost S. Let

$$X = B_k C_k, \qquad H = A_k - X = B_k (C_*^{\otimes k} - C_k), \qquad \|H\|_k \le \epsilon_k.$$

If $R \ge L$, use the exact map; otherwise choose $\eta > 0$ with $S + \eta(L - S) < R$. Retain terms with at most $\lfloor \eta m \rfloor$ residuals in $(X + H)^{\otimes m}$. For $z = 4^{1/\eta}$ and $\epsilon_k \le (1 + z)^{-1}$, the omitted norm is at most

$$z^{-\eta m} (1 + \epsilon_k + z \epsilon_k)^m \le 2^{-m}. \tag{23}$$

The cost is at most $3^m 2^{km(S + \eta(L-S))/2}$. Taking k larger ensures $S + \eta(L - S) + 2 \log_2(3)/k < R$. Exact padding of fewer than k factors adds bounded cost, absorbed by a strict rate margin for all large N. The error is exponential in N, proving (20). □

6 From parallel control to general interleavings

Average C_n over matching environment-factor permutations. The square roots A_n, B_n intertwine the corresponding external permutations, and the strategy norm is invariant and convex. Both bounds in (20) survive, while the residual $H_n = A_n - B_n C_n$ has permutation-invariant $Y_n = H_n H_n^*$. By (11) choose a deterministic joint comb S_n with $Y_n \le e_n^2 S_n$, where $e_n = K e^{-\gamma n}$. Symmetrizing S_n preserves the inequality. Proposition 3 gives

$$Y_n \le e_n^2 P(n) \int R(V, W)^{\otimes n} \, dV \, dW. \tag{24}$$

For any general tester let T_{g} be its positive deterministic operator and $0 \le Q_{\mathrm{g}} \le T_{\mathrm{g}}$ its null-acceptance effect. Positivity follows by expressing the circuit links as Bell contractions of Choi operators and tester Kraus vectors, summing over branches. For every physical one-use comb R,

$$\operatorname{Tr} T_{\mathrm{g}} R^{\otimes n} = 1. \tag{25}$$

Each matched first component precedes its second component. Replacing the unknown comb in every use by any such R therefore leaves a deterministic acyclic circuit. Classical decisions about later orders are included by summing their positive branch operators.

The tester need not normalize an arbitrary correlated joint comb in $\mathcal{S}_n$. The iid physical factors in (24) are what makes (25) sufficient. Set $p = \operatorname{Tr} Q_{\mathrm{g}} R_1^{\otimes n}$ and $q = \operatorname{Tr} Q_{\mathrm{g}} R_2^{\otimes n}$. Then

$$\begin{aligned} \sqrt{p} = \|\sqrt{Q_{\mathrm{g}}} A_n\|_{\mathrm{F}} &\le \|\sqrt{Q_{\mathrm{g}}} B_n\|_{\mathrm{F}} \|C_n\|_{\mathrm{op}} + \|\sqrt{Q_{\mathrm{g}}} H_n\|_{\mathrm{F}} \\ &\le 2^{nR/2}\sqrt{q} + \sqrt{P(n)} K e^{-\gamma n}. \end{aligned} \tag{26}$$

For $r' > d_0$, choose $d_0 < R < r'$. If $q \le 2^{-nr'}$, both terms decrease exponentially: P has fixed degree because the slot dimensions are fixed. Squaring gives the uniform exponential converse in Theorem 1.

7 Support, achievability and the nested corollary

A legal parallel tester prepares maximally entangled pairs on C, C' and B_0, B'. It sends C into the first component, stores the returned A, sends B_0 into the second, and retains C', A, B', D. Its output is the normalized comb Choi state $R_i/(d_C d_B)$. No internal memory is accessed.

If $\operatorname{supp} R_1 \not\subseteq \operatorname{supp} R_2$, a measurement on the unsupported part, repeated on independent readouts, gives zero type-II error and vanishing type-I error. The parallel rate is infinite, and so are the larger strategy classes. Otherwise $R_1 \le cR_2$ for some finite c, which bounds every common n-block output relative entropy by $n \log_2 c$. Thus d_0 is finite and the exact factorization (14) applies.

For achievability, fix any block length and common tester approaching the corresponding relative-entropy supremum in (3). Repeat that block and apply state Stein testing. Fewer than one block of unused copies may be discarded. Optimizing over blocks attains d_0. Since parallel strategies are nested and nested strategies are general, the converse proves (4). In particular:

Corollary 5. *Nested-adaptive discrimination never improves the vanishing-type-I Stein exponent over fully parallel discrimination for finite-dimensional deterministic superchannels with inaccessible internal memory.*

An alternative proof of this nested corollary, included in the source package, combines the CP trace-nonincreasing approximation of [2] with a heralded completion and the finite-state AEP of [8]. That route is a source-dependent synthesis and is not a second independent contribution beyond Theorem 1.

8 Attribution and review

The fixed-marginal isometry bound, strategy duality and tensor amplification are inherited tools. The added interfaces proposed here are physical two-stage postselection with canonical covariant memories, the strategy norm for a common approximation, and its use under general interleaving testers. The positive-slack argument is closely related to the ordinary-channel work of [3].

The finite postselection statement differs from infinite exchangeable causal-process representations [6]. The ordinary-channel theorems of [7] describe reduced channels and almost-iid approximations and do not directly preserve the two-stage constraints needed here. The source package records these comparisons and the exact original questions. A bounded literature check does not establish comprehensive priority.

Sequential internal review checked the canonical-memory realization, Schur normalization, common-slack minimax, norm transfer and general-tester normalization. It corrected an earlier assertion that the fixed-tester dual in (18) always attains its optimum. The present proof uses an infimum there and proves global attainment separately. An exact physical counterexample to the earlier assertion is in the package.

New rational calculations construct Haar moments through a permutation commutant Gram pseudoinverse and reproduce the relevant normalization at two and three qubit copies. Incorrect constants are detected. These are finite controls, not a certification of the all-dimensional theorem. There has been no independent expert or proof-assistant review. The author directed the research and workflow; AI systems assisted with derivation, source comparison, computation and drafting.

References

[1] C. Hirche, *Quantum Network Discrimination*, arXiv:2103.02404v2 (2021).

[2] M. Gao, Z. Ji and C. Liu, *Quantum Channel Stein's Lemma with an Exponential Strong Converse*, arXiv:2609.27196v1 (2026).

[3] C. Zhu and X. Wang, *Quantum Channel Stein Theorem beyond Definite Causal Order*, arXiv:2609.30268v1 (2026).

[4] S. Nahar, D. Tupkary, Y. Zhao, N. Lütkenhaus and E. Y.-Z. Tan, *Postselection technique for optical Quantum Key Distribution with improved de Finetti reductions*, arXiv:2403.11851v3 (2024).

[5] G. Gutoski and J. Watrous, *Toward a General Theory of Quantum Games*, arXiv:quant-ph/0611234v2 (2007).

[6] F. Costa, J. Barrett and S. Shrapnel, *A de Finetti theorem for quantum causal structures*, Quantum **9**, 1628 (2025), arXiv:2403.10316v3.

[7] L. Ye, B. Bergh and N. Datta, *Quantum de Finetti theorems for states and channels in any distance measure*, arXiv:2609.40343v1 (2026).

[8] K. Fang, O. Fawzi, R. Renner and D. Sutter, *A chain rule for the quantum relative entropy*, arXiv:1909.05826 (2019).