

Finite-copy optimal cloning of mixed qudit states:\\ exact optima, a closed-form cloner, and the structure of covariant sector channels

Hikaru Wakaura

hikaruwakaura@gmail.com

QuantScape Inc

Research Article

Keywords: no-cloning theorem , quantum cloner.

Posted Date: October 26th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11480035/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

Finite-copy optimal cloning of mixed qudit states: exact optima, a closed-form cloner, and the structure of covariant sector channels

Hikaru Wakaura^{1,*}

¹*QIRI (Quantum Integrated Research Institute Inc.),
1-16-3 Akasaka, Minato-ku, Tokyo 107-0061, Japan*

(Dated: October 23, 2026)

Copying is the most basic operation on information, and the no-cloning theorem is the sharpest statement of how quantum information differs from classical. The best approximate copier of pure states has been known for decades, and the asymptotic theory of cloning mixed states at fixed gain $\gamma = m/n$ was recently completed by Schur–Weyl methods, which give closed-form limiting fidelities and an asymptotically optimal “Schur–Cartan” cloner. Whether those constructions are also good at the finite copy numbers available in experiments, and what the finite-copy optimum is, was unknown. We show that the finite-copy optimum is a concave program over Young-label transition rules and extremal covariant maps between irreducible sectors, solve it exactly for qubits up to $n = 40$ and for qutrits and ququarts at small n , and show that the asymptotically optimal cloners lose up to 0.11 in root fidelity at $n \leq 20$, while purify–clone–trace fails at the degenerate spectrum. The loss is entirely in the label rule: the Cartan sector channel is finite-copy optimal for qubits, and in higher dimension it coincides with the lowest-Casimir (minimal Parthasarathy–Ranga Rao–Varadarajan) component, which is within 10^{-3} of optimal. A closed-form cloner, a Knothe–Rosenblatt coupling of the input and target label laws followed by lowest-Casimir channels, is asymptotically optimal, improves the Schur–Cartan cloner by up to 0.115, and comes with Karush–Kuhn–Tucker and Frank–Wolfe certificates bracketing the true optimum to 10^{-3} without semidefinite programming. The analysis isolates a single representation-theoretic question, the pairwise optimality of the PRV channel at intermediate spectral mixedness, settled here in the cold and hot limits and for single-copy inputs, which governs how much coherent cloning advantage survives at the copy numbers of near-term quantum networks.

I. INTRODUCTION

Every classical information technology rests on the ability to copy data at will, and quantum theory forbids exactly that operation for unknown states [1, 2]. The no-cloning theorem is therefore the departure point for quantum cryptography, quantum money, and the design of quantum networks, where one state must often be shared among several parties. In all of these settings the relevant question is quantitative rather than absolute: how well can n copies of an unknown state be turned into $m > n$ approximate copies, and by which physical process? For pure states this question was answered in the 1990s [3–6]: the optimal universal cloner projects the input onto the symmetric subspace of the output, its fidelity has a closed form, and for single-clone fidelity the limit of many clones reduces to state estimation [7], although for the joint fidelity used here optimal measure-and-prepare outputs need not be independent copies of an estimate [8].

Mixed states resisted a comparable treatment for two decades [9, 10]. The difficulty is structural: n copies of a mixed state ρ decompose under Schur–Weyl duality into a classical Young label, which carries the spectrum, and an irreducible sector, which carries the eigenbasis, and a cloner must act on both. Recent work resolved the asymptotic problem. Fei and Wang [11] determined

the optimal global fidelity at fixed gain $\gamma > 1$ for simple full-rank spectra, as a product of one-mode thermal affinities, and attained it with a Schur–Cartan cloner that processes the label classically and applies a covariant Cartan-sector channel to the irreducible state. Fanizza *et al.* [12] and Jeon *et al.* [13] showed, independently, that purify–clone–trace (PCT), the Werner cloner applied to a random purification [14], has the optimal sample complexity $n = \Theta((m-n)rd/\epsilon)$ for rank- r states, a worst-case high-fidelity statement whose order of limits differs from the fixed-gain one [11]. Together these results fix the limiting constants and the scaling of mixed-state cloning.

However, the limiting constants say nothing about the regime that matters for implementations, where n is a few tens at most, the output is not divided among asymptotically many clones, and the spectrum may be close to degenerate; moreover, the asymptotically optimal constructions are not designed to be optimal at any finite n , so neither the size of their finite-copy loss nor the shape of the true finite-copy optimum was known. In this study, we determine the exact finite-copy optimum over covariant, permutation-symmetric cloners by reducing it to a concave program whose variables are a transition rule on Young labels and a mixture of extremal covariant maps between irreducible sectors. We found that the Schur–Cartan cloner of Ref. [11] is far from optimal at finite n , by up to 0.11 in root fidelity for qubits at $n = 20$ and by up to 0.115 for ququarts at $n = 3$, and that the whole loss sits in its label rule, because the output label is sampled independently of the input label and incom-

* h.wakaura@qiri.co.jp

patible pairs fall back to a maximally mixed output. The result shows that the sector channel is already optimal: for qubits the Cartan coupling is used with weight exactly one by the optimum at every size we solved, and in higher dimension it is the component of lowest quadratic Casimir in $\mathcal{H}_\mu^* \otimes \mathcal{H}_\nu$, the minimal Parthasarathy–Ranga Rao–Varadarajan (PRV) component [15, 16], which remains optimal to within 10^{-3} . This identification yields a closed-form cloner that is asymptotically optimal and near-optimal at every finite size, together with certificates that bound its distance to the true optimum without solving a semidefinite program.

II. SETTING AND REDUCTION

A. Global fidelity at fixed gain

Let $\rho_p = \text{diag}(p_1, \dots, p_d)$ with $p_1 > \dots > p_d > 0$ and $\rho_{p,g} = g\rho_p g^\dagger$ for $g \in U(d)$. A cloner is a channel $\mathcal{M}$ from $(\mathbb{C}^d)^{\otimes n}$ to $(\mathbb{C}^d)^{\otimes m}$, $m > n$, and its figure of merit is the root Uhlmann fidelity [17] of the entire output with the product target,

$$F(\mathcal{M}) = \inf_g \left\| \sqrt{\mathcal{M}(\rho_{p,g}^{\otimes n})} \sqrt{\rho_{p,g}^{\otimes m}} \right\|_1, \quad (1)$$

exactly the global criterion of Ref. [11]. Schur–Weyl duality gives $(\mathbb{C}^d)^{\otimes N} = \bigoplus_{\lambda \vdash N} \mathcal{H}_\lambda \otimes \mathcal{S}_\lambda$, and

$$\rho_p^{\otimes N} = \bigoplus_{\lambda} P_{N,p}(\lambda) \rho_{p,\lambda} \otimes \frac{I_{\mathcal{S}_\lambda}}{\dim \mathcal{S}_\lambda}, \quad (2)$$

with $P_{N,p}(\lambda) = \dim \mathcal{S}_\lambda s_\lambda(p)$, where s_λ is the Schur polynomial and $\rho_{p,\lambda} = \pi_\lambda(\rho_p)/s_\lambda(p)$ is diagonal in a weight basis of the irreducible $U(d)$ -module $\mathcal{H}_\lambda$. The Fei–Wang cloners [11] measure the input label μ , draw an output label ν from a transition rule $K(\nu|\mu)$, and apply the Cartan-sector channel

$$C_{\mu,\nu}(X) = \frac{\dim \mathcal{H}_\mu}{\dim \mathcal{H}_\nu} V_{\mu,\omega}^\dagger (X \otimes I_{\mathcal{H}_\omega}) V_{\mu,\omega}, \quad \omega = \nu - \mu, \quad (3)$$

when ω is a partition (the pair is *compatible*), with $V_{\mu,\omega}$ the isometry onto the highest-weight component of $\mathcal{H}_\mu \otimes \mathcal{H}_\omega$; otherwise they output $\text{Tr}(X)I/\dim \mathcal{H}_\nu$. The known-spectrum rule is $K(\nu|\mu) = P_{m,p}(\nu)$; the spectrum-independent rule rounds $\gamma\mu$ randomly. Their fidelities converge to $F_{\text{orb}}(\gamma, p) = \prod_{i < j} f_\gamma(p_j/p_i)$ and $F_{\text{univ}} = F_{\text{cl}} F_{\text{orb}}$ respectively, with $f_\gamma(q) = (\sqrt{\gamma} + \sqrt{q(\gamma - 1 + q)})/(\gamma + q)$ and $F_{\text{cl}} = (2\sqrt{\gamma}/(1 + \gamma))^{(d-1)/2}$ [11]. PCT applies Werner’s cloner [5] in dimension dr to a random purification and traces the purifying registers; its limit F_{PCT} lies strictly below F_{univ} for simple full-rank spectra, while for pure states PCT and the Schur–Cartan cloner coincide [11].

B. Exact finite-copy optimum as a concave program

By covariantization and joint concavity of the fidelity, the infimum over g in Eq. (1) is attained by $U(d)$ -covariant channels, and twirling the output over S_m (which fixes the target) and the input over S_n costs nothing (Appendix B, where every step of this subsection is derived). The Choi operator of such a channel therefore lies in the commutant of $U(d) \times S_n \times S_m$ and decomposes as

$$J = \bigoplus_{\mu \vdash n, \nu \vdash m} J^{\mu\nu} \otimes I_{\mathcal{S}_\mu^* \otimes \mathcal{S}_\nu}, \quad J^{\mu\nu} \in \text{End}(\mathcal{H}_\mu^* \otimes \mathcal{H}_\nu)^{U(d)}, \quad (4)$$

with trace preservation $\sum_\nu \text{Tr}_\nu J^{\mu\nu} = I_\mu$. Each $J^{\mu\nu}$ is a positive element of the commutant, that is, a positive-weighted sum over the irreducible components $\lambda \subset \mathcal{H}_\mu^* \otimes \mathcal{H}_\nu$ of projectors (density matrices on the multiplicity spaces when a component repeats). Writing $\Xi_\nu = \sum_\mu P_{n,p}(\mu) \Phi^{\mu\nu}(\rho_{p,\mu})$ for the unnormalized output block and $T_\nu = \dim \mathcal{S}_\nu \pi_\nu(\rho_p)$ for the target block, the fidelity is

$$F = \sum_\nu \sum_w \sqrt{T_{\nu,w}} \text{Tr} \sqrt{\Xi_{\nu,w}}, \quad (5)$$

where w runs over the weight spaces of $\mathcal{H}_\nu$ (T_ν is scalar on each, and Ξ_ν is weight-block-diagonal by torus covariance). Equation (5) is concave in the sub-channels, so the finite-copy optimum is a concave program: for $d = 2$ every weight space is one-dimensional, the extremal covariant maps between spin j and spin J are indexed by an intermediate spin $L \in \{|J - j|, \dots, J + j\}$ with explicit Clebsch–Gordan kernels, and the program has $\sim 10^4$ scalar variables at $n = 20$; for $d \geq 3$ we solve the semidefinite program over the $J^{\mu\nu}$ with $U(d)$ -covariance imposed through the Lie algebra, blocked by relative weight. We validated the $d = 2$ program against a direct Choi-matrix SDP (agreement 2×10^{-5}), the two $d \geq 3$ engines against each other (agreement 10^{-9}), and all engines against Werner’s pure-state value $\sqrt{(n+1)/(m+1)}$. Irreducible representations are built from Gelfand–Tsetlin matrix elements [18, 19], so no tensor-power embedding is required.

III. FINITE-COPY RESULTS FOR QUBITS

Figure 1 compares, for $d = 2$, the exact optimum with the Fei–Wang cloners and PCT for $n \leq 20$ and $\gamma \in \{1.5, 2, 3\}$. Three features are common to all panels. First, the optimum approaches F_{orb} from above and slowly: for $p = (0.6, 0.4)$, $\gamma = 2$ it is 0.982 at $n = 20$ and 0.970 at $n = 40$ against the limit 0.926, so the asymptotic constant is not predictive for mildly mixed states at experimentally relevant sizes. Second, the known-spectrum cloner C_n^p converges from below and is strongly suboptimal: the worst case in our grid is $p = (0.9, 0.1)$, $\gamma = 1.5$,

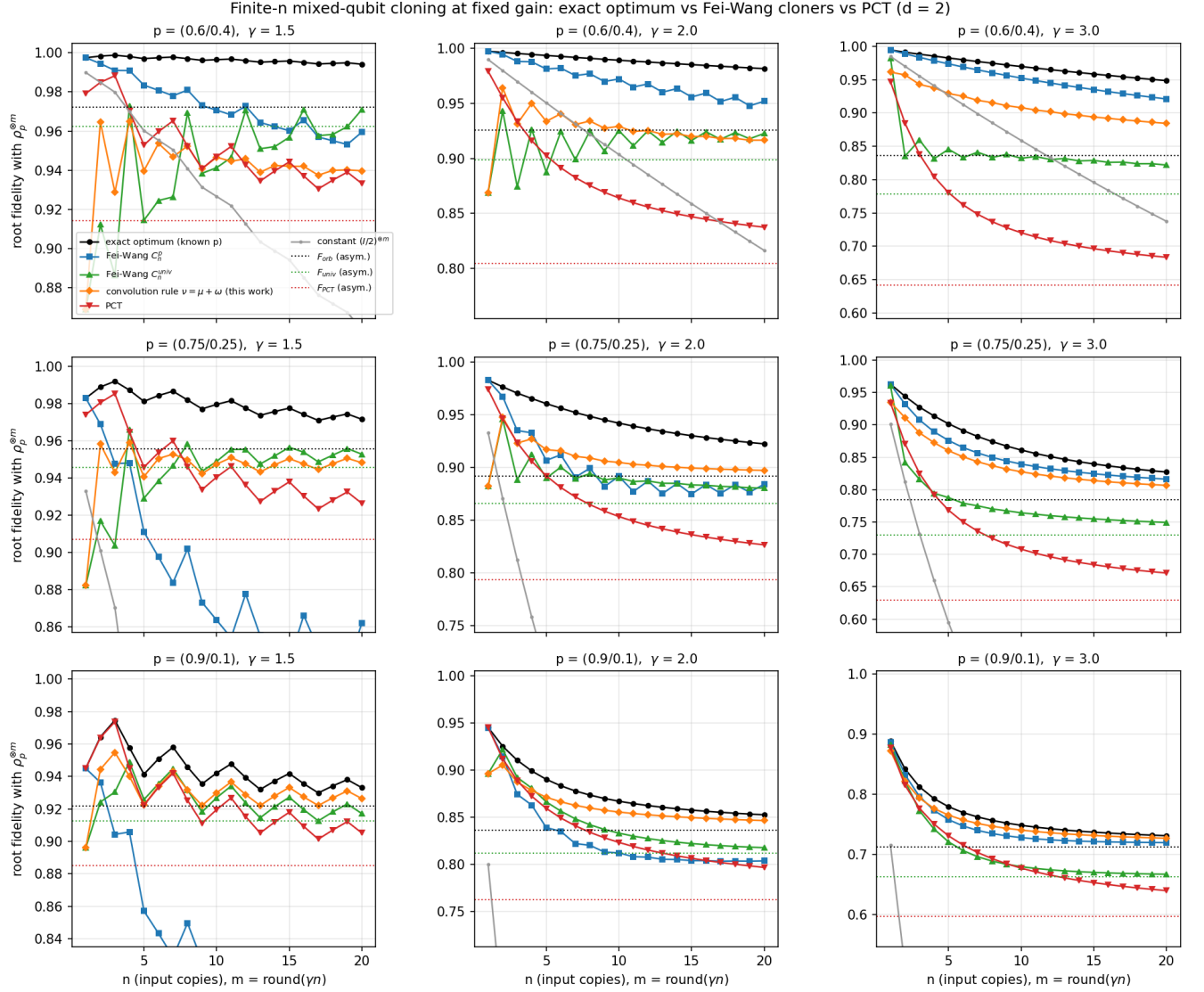


FIG. 1. **The finite-copy optimum lies well above the asymptotic constant while the Schur–Cartan cloner lies well below it.** Root fidelity versus input copies n at gain $\gamma = m/n$ for three qubit spectra (rows) and three gains (columns): exact optimum (black), Fei–Wang known-spectrum cloner C_n^p (blue), spectrum-independent cloner C_n^{univ} (green), purify–clone–trace (red), the constant channel $(I/2)^{\otimes m}$ (grey), and the limits F_{orb} , F_{univ} , F_{PCT} (dotted). At $\gamma = 1.5$, $p = (0.9, 0.1)$ the known-spectrum cloner is beaten by its own spectrum-independent variant and by PCT.

$n = 20$, with 0.824 against the optimum 0.933. Decomposing its output shows why: the rule $K(\nu|\mu) = P_{m,p}(\nu)$ ignores the input label, so the fraction of probability mass on incompatible pairs, which are sent to the maximally mixed state, is 0.19 at $n = 10$ and still 0.09 at $n = 100$ for $p = (0.6, 0.4)$, $\gamma = 2$, and on compatible pairs the fluctuation of the ratio $(\mu_1 - \mu_2)/(\nu_1 - \nu_2)$ around $1/\gamma$ lowers the mean sector fidelity to 0.913 at $n = 100$, against 0.943 for a monotone coupling and the limit 0.926. Third, PCT is the weakest of the quantum cloners for $n \geq 3$, with the single exception of $p = (0.9, 0.1)$, $\gamma = 1.5$ where it beats C_n^p , and converges to F_{PCT} from above.

Figure 2 quantifies the non-commutation of the degen-

erate and large- n limits anticipated in Ref. [11]. At fixed n the optimum tends to one as $q \rightarrow 1$, because the state becomes known, while $F_{\text{orb}}(\gamma, q) \rightarrow 2\sqrt{\gamma}/(1 + \gamma) < 1$; at $n = 20$ the optimum is already 0.994 at $q = 0.8$ against the limit 0.934. The known-spectrum cloner is exact at $q = 1$ (the Cartan channel maps I/dim to I/dim) and agrees with the optimum to four digits for $q \geq 0.98$. PCT does not become trivial at the degenerate point: the random purification of $I/2$ is maximally entangled and its Werner clone is lossy, giving 0.84 at $n = 20$.

Figure 3 addresses the unknown-spectrum problem. The finite-copy minimax over a spectral set K is the same concave program with a max–min objective. For $n \leq 20$

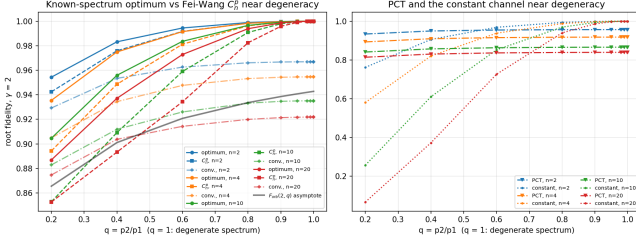


FIG. 2. Near a spectral degeneracy the finite-copy optimum tends to one while the asymptotic constant does not. Left: exact optimum (solid), C_n^p (dashed) and the convolution rule (dash-dotted) for $\gamma = 2$ as functions of $q = p_2/p_1$ at $n = 2, 4, 10, 20$, against $F_{\text{orb}}(2, q) \rightarrow 2\sqrt{2}/3$. Right: PCT (dash-dotted) saturates at 0.84 for $n = 20$ even at $q = 1$, where the constant channel (dotted) is exact.

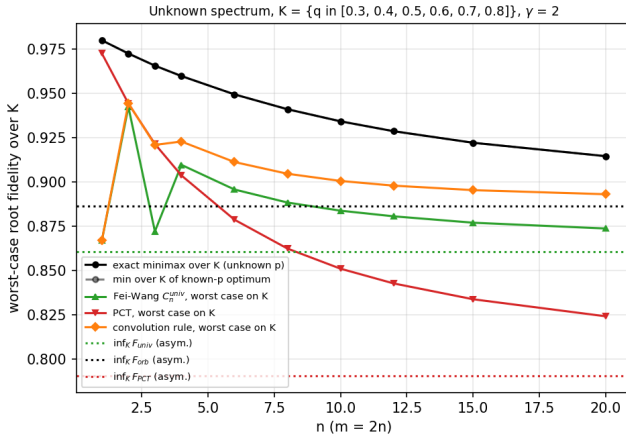


FIG. 3. The cost of not knowing the spectrum is absent at $n \leq 20$. Worst-case root fidelity over the spectral set $K = \{q \in [0.3, 0.8]\}$ at $\gamma = 2$: the exact minimax over K (black) coincides with the minimum over K of the known-spectrum optimum (grey, hidden) to five digits, whereas the asymptotic values $\inf_K F_{\text{univ}} = 0.861$ and $\inf_K F_{\text{orb}} = 0.886$ (dotted) differ by the classical factor $F_{\text{cl}} = 0.971$.

it equals the worst-case known-spectrum optimum to five digits, for a wide set and for sets of width $\propto n^{-1/2}$ alike: the channel designed for the hardest spectrum does at least as well on the easier ones. The separation that the asymptotic theory predicts, $F_{\text{cl}} = 0.971$, begins to appear only at $n = 30\text{--}40$ (1×10^{-4} and 4×10^{-4}), so the classical factor is a genuinely large- n effect.

IV. STRUCTURE OF THE OPTIMUM

A. The sector channel is the minimal PRV component

Extracting the optimal sub-channels for $d = 2$ in eight instances with $6 \leq n \leq 20$ and $\gamma \in \{1.5, 2, 3\}$ shows that the optimum uses the Cartan coupling $L = J - j$

with weight at least 0.994, and typically 1.000; the whole finite-copy gap is in the label rule. For $d = 3, 4$ the same statement holds up to a small, well-characterized admixture. The following theorem explains which component the Cartan channel is.

Theorem 1 (Cartan = lowest Casimir = minimal PRV). *For any labels $\mu \vdash n, \nu \vdash m$, the irreducible component of $\mathcal{H}_\mu^* \otimes \mathcal{H}_\nu$ with the smallest quadratic Casimir is the minimal PRV component, whose highest weight is the dominant representative $\text{dom}(\nu - \mu)$ of the Weyl orbit of $\nu - \mu$, and it has multiplicity one. When $\nu - \mu$ is a partition it is the component of the Cartan channel (3).*

Proof sketch. The lowest weight of $\mathcal{H}_\mu^*$ is $-\mu$, so by the theorem of Parthasarathy, Ranga Rao and Varadarajan [15, 16] $\mathcal{H}_\mu^* \otimes \mathcal{H}_\nu$ contains the module of highest weight $\text{dom}(\nu - \mu)$ with multiplicity one. Every component has highest weight $\theta \succeq \text{dom}(\nu - \mu)$ in dominance order: this follows from the Lidskii–Wielandt eigenvalue inequality once the Littlewood–Richardson coefficient is realized by a Hermitian triple (Klyachko, Knutson–Tao). The Casimir $C(\theta) = \langle \theta, \theta + 2\rho \rangle$ is strictly increasing along dominance order. The full proof, with the Casimir computation and the Hermitian-triple argument, is in Appendix C; for compatible pairs $\text{dom}(\nu - \mu) = \omega$ and the component is the one of the intertwiner $V_{\mu, \omega}$ (Appendix B3). $\square$

This suggests a single rule for every pair, compatible or not: use the extremal covariant channel supported on the lowest-Casimir component (we call it the *lowest-Casimir channel*). On incompatible pairs the Fei–Wang fallback to the maximally mixed state costs up to 0.03 in $d = 3$, because the output blocks are no longer one-dimensional, while the lowest-Casimir channel is the optimal choice: restricting incompatible pairs to it and optimizing the rule reproduces, to six digits in all 28 instances we solved ($d = 3, n \leq 3, m \leq 6$), the optimum in which incompatible pairs carry arbitrary covariant maps and compatible pairs the Cartan channel. The remaining difference from the fully unrestricted optimum, at most 9×10^{-4} , is the compatible-pair admixture analysed next.

B. An exact criterion for non-Cartan admixture

Let x be the optimum of the program restricted to lowest-Casimir channels, with rule K , and define the marginal gain of routing label μ into block ν through component λ with multiplicity density M ,

$$g_{\mu\nu\lambda}(M; x) = P_{n,p}(\mu) \sum_w \sqrt{T_{\nu,w}} \frac{1}{2} \text{Tr}(\Xi_{\nu,w}(x)^{-1/2} \Phi_{\lambda,M}(\rho_{p,\mu})|_w), \quad (6)$$

and $g_{\mu\nu\lambda}^* = \max_M g_{\mu\nu\lambda}(M; x)$, the top eigenvalue of a $k \times k$ matrix for a component of multiplicity k .

Theorem 2 (KKT criterion). *Suppose $\Xi_{\nu,w}(x) \succ 0$ on every block. (i) For each μ the gains $g_{\mu\nu\lambda_{\min}}(x)$ take a common value η_μ on all ν used by K and are $\leq \eta_\mu$ elsewhere. (ii) x is the unrestricted optimum if and only if $g_{\mu\nu\lambda}^*(x) \leq \eta_\mu$ for all (μ, ν, λ) . (iii) In general $\text{OPT} \leq F(x) + \sum_\mu (\max_{\nu\lambda} g_{\mu\nu\lambda}^*(x) - \eta_\mu)_+$.*

Proof sketch. $\text{Tr} \sqrt{\cdot}$ is differentiable at positive operators with $\frac{d}{dt} \text{Tr} \sqrt{X + tA}|_0 = \frac{1}{2} \text{Tr}(X^{-1/2}A)$, so the directional derivative of F at x towards any feasible x' is $\sum_\mu \sum_{\nu\lambda M} w'_\mu(\nu, \lambda, M) g_{\mu\nu\lambda}(M; x)$. (i) is the Karush–Kuhn–Tucker condition for the simplex constraint $\sum_\nu K(\nu|\mu) = 1$. For (iii), concavity gives $F(x') \leq F(x) + DF(x)[x' - x]$; since $\text{Tr} \sqrt{\cdot}$ is positively homogeneous of degree $\frac{1}{2}$, $DF(x)[x] = F(x)/2 = \sum_\mu \eta_\mu$ by (i), while $DF(x)[x'] \leq \sum_\mu \max_{\nu\lambda} g_{\mu\nu\lambda}^*$. (ii) follows from (iii) and from the fact that a violated inequality gives a feasible direction of strictly positive derivative. Each step is written out in Appendix D. $\square$

Theorem 2 converts the structural question into a computable certificate. On 14 instances in $d = 3, 4$ the set of violations predicted by (ii) coincided exactly with the admixture observed in the unrestricted SDP (a fifteenth instance was settled by the certificate alone), and the violating component was always the one that embeds $\mathcal{H}_\nu$ into $\mathcal{H}_\mu \otimes \mathcal{H}_{\lambda'}$ through the *row-shaped* ancilla $\lambda' = (|\omega|, 0, \dots, 0)$; no violation ever involved an incompatible pair or a higher component. The mechanism is the concavity of Eq. (5) across weight blocks: the admixed component places weight where the Cartan output is thin. When a block ν is fed by a second input label μ' with $\nu - \mu'$ row-shaped, that label already supplies the missing pattern and no admixture occurs; the admixture therefore appears only for blocks fed by a single compatible label and with weight multiplicities above one, in the examples $(1, 0, 0) \rightarrow (2, 1, 0)$ at $(1, 3)$, $(2, 1, 0) \rightarrow (3, 2, 0)$ at $(3, 5)$ and $(2, 1, 1) \rightarrow (3, 2, 1)$ at $(4, 6)$, with lowest-Casimir fractions 0.88, 0.94 and 0.91, and total fidelity gains below 9×10^{-4} . The criterion is intrinsically spectrum-dependent: the pair $(2, 1, 1) \rightarrow (3, 2, 1)$ at $(n, m) = (4, 6)$ mixes for $p = (0.6, 0.3, 0.1)$ and not for $p = (0.7, 0.2, 0.1)$ with identical combinatorial data, so no spectrum-free combinatorial rule can be complete. A quantitative form of the single-feeder mechanism follows from operator monotonicity of $x \mapsto x^{-1/2}$ (Proposition 1 in Appendix D): if some μ' feeds ν with mass κ and its lowest-Casimir output Y dominates the admixture pattern, $\Phi_\lambda(\rho_\mu) \leq cY$ blockwise, then $g_{\mu\nu\lambda}^* \leq \frac{1}{2} P_{n,p}(\mu) \sqrt{c/(P_{n,p}(\mu')\kappa)} \sum_w \sqrt{T_{\nu,w}} \text{Tr} \sqrt{\Phi_\lambda(\rho_\mu)}|_w$.

V. A CLOSED-FORM CLONER AND ITS CERTIFICATES

A. Any coupling of the label laws is asymptotically optimal

Theorem 3. *Let p have a simple full-rank spectrum, $m_n/n \rightarrow \gamma > 1$, and let J_n be any coupling of $P_{n,p}$ and $P_{m_n,p}$. The lifted cloner with rule $K_n(\nu|\mu) = J_n(\mu, \nu)/P_{n,p}(\mu)$ and Cartan channels on compatible pairs has fidelity converging to $F_{\text{orb}}(\gamma, p)$, uniformly for p in compact subsets of the simple chamber and for all eigenbases.*

The proof uses two facts about the lifted cloners that we prove here and two asymptotic inputs that we quote. Write $T_N(p) = \{\lambda \vdash N : \|\lambda/N - p\|_\infty \leq N^{-1/3}\}$ for the typical labels.

Lemma 1 (block form of the fidelity). *For any rule K and any covariant sub-channels $E_{\mu\nu}$, the lifted cloner $Q[K]$ satisfies, for every g ,*

$$F(Q[K](\rho_{p,g}^{\otimes n}), \rho_{p,g}^{\otimes m}) = \sum_\nu \sqrt{P_{m,p}(\nu)} F(\Xi_\nu, \rho_{p,\nu}),$$

with $\Xi_\nu = \sum_\mu P_{n,p}(\mu) K(\nu|\mu) E_{\mu\nu}(\rho_{p,\mu})$ and $\text{Tr} \Xi_\nu = \sum_\mu P_{n,p}(\mu) K(\nu|\mu)$.

Proof. Covariance of $Q[K]$ and of the target under $g^{\otimes m}$ reduces to $g = I$. By construction $Q[K](\rho_p^{\otimes n}) = \bigoplus_\nu \Xi_\nu \otimes I_{\mathcal{S}_\nu} / \dim \mathcal{S}_\nu$, while Eq. (2) gives $\rho_p^{\otimes m} = \bigoplus_\nu P_{m,p}(\nu) \rho_{p,\nu} \otimes I_{\mathcal{S}_\nu} / \dim \mathcal{S}_\nu$. The root fidelity is additive over orthogonal blocks, multiplicative over tensor factors, equal to one on identical normalized factors, and homogeneous of degree $\frac{1}{2}$ in each argument. $\square$

Lemma 2 (label concentration). $P_{N,p}(T_N(p)^c) \leq (N+1)^{d(d+1)/2} e^{-2N^{1/3}}$ for every N and every spectrum p .

Proof. By the large-deviation bound for the Young-label law [20], $P_{N,p}(\lambda) \leq (N+1)^{d(d-1)/2} e^{-ND(\lambda/N||p)}$ with D the relative entropy. Pinsker's inequality [21, Lemma 11.6.1] gives $D(\lambda/N||p) \geq 2\|\lambda/N - p\|_1^2 \geq 2\|\lambda/N - p\|_\infty^2$, and there are at most $(N+1)^d$ labels. $\square$

The two quoted inputs are the uniform Cartan-sector limit and the converse of Ref. [11]: (Q1) for every compact K in the simple chamber there are $\varepsilon_n \rightarrow 0$ such that, for all sufficiently large n , every $p \in K$, $\mu \in T_n(p)$ and $\nu \in T_{m_n}(p)$ form a compatible pair with $F(C_{\mu,\nu}(\rho_{p,\mu}), \rho_{p,\nu}) \geq F_{\text{orb}}(\gamma, p) - \varepsilon_n$ (Proposition 2.3 of Ref. [11]); (Q2) for every sequence of channels $\mathcal{M}_n$, $\limsup_n \inf_g F(\mathcal{M}_n(\rho_{p,g}^{\otimes n}), \rho_{p,g}^{\otimes m_n}) \leq F_{\text{orb}}(\gamma, p)$ (Theorem 1.1 of Ref. [11]). Both are statements about the Cartan channel and about arbitrary channels respectively; neither involves the label rule.

Proof of Theorem 3. Fix $g = I$ by covariance. By Lemma 1, $F_n = \sum_\nu \sqrt{P_{m,p}(\nu)} F(\Xi_\nu, \rho_{p,\nu})$ with $\Xi_\nu =$

$\sum_{\mu} J_n(\mu, \nu) E_{\mu\nu}(\rho_{p,\mu})$. Restrict the sum to typical inputs:

$$\Xi_{\nu}^{\text{typ}} = \sum_{\mu \in T_n(p)} J_n(\mu, \nu) E_{\mu\nu}(\rho_{p,\mu}) \leq \Xi_{\nu},$$

$$Q^{\text{typ}}(\nu) = \text{Tr} \Xi_{\nu}^{\text{typ}} \leq P_{m,p}(\nu).$$

Since $A \mapsto F(A, C) = \text{Tr} \sqrt{\sqrt{C} A \sqrt{C}}$ is operator monotone and concave, and homogeneous of degree $\frac{1}{2}$,

$$F_n \geq \sum_{\nu \in T_m(p)} \sqrt{P_{m,p}(\nu) Q^{\text{typ}}(\nu)} \times \sum_{\mu \in T_n(p)} \frac{J_n(\mu, \nu)}{Q^{\text{typ}}(\nu)} F(E_{\mu\nu}(\rho_{p,\mu}), \rho_{p,\nu}).$$

For n large, (Q1) makes every pair in the double sum compatible, so $E_{\mu\nu} = C_{\mu,\nu}$ and each inner fidelity is at least $F_{\text{orb}}(\gamma, p) - \varepsilon_n$. Using $\sqrt{P_{m,p} Q^{\text{typ}}} \geq Q^{\text{typ}}$,

$$F_n \geq (F_{\text{orb}}(\gamma, p) - \varepsilon_n) J_n(T_n(p) \times T_m(p))$$

$$\geq (F_{\text{orb}}(\gamma, p) - \varepsilon_n) (1 - P_{n,p}(T_n^c) - P_{m,p}(T_m^c)),$$

where the last step uses only the marginals of the coupling, and the two tail probabilities vanish uniformly in p by Lemma 2. Hence $\liminf_n F_n \geq F_{\text{orb}}(\gamma, p)$ uniformly on K ; (Q2) gives the matching $\limsup$. $\square$

The independent coupling of Ref. [11] is one instance; the finite-copy advantage of other couplings is therefore a question of incompatible mass and of the variance of the sector ratio. The natural multi-dimensional monotone coupling is the Knothe–Rosenblatt (KR) coupling, which matches the first rows of μ and ν by quantiles and then each subsequent row conditionally. It has exact marginals, hence is asymptotically optimal, and its incompatible mass decays quickly (below 10^{-3} at $n = 40$ in $d = 3$ and 2×10^{-3} in $d = 4$), whereas the independent coupling leaves 21–33% ($d = 3$) and 42–56% ($d = 4$) of the mass incompatible at $n = 40$. We call *KR+min* the cloner with the KR rule and lowest-Casimir channels on every pair.

Tables I and II summarize $d = 3, 4$. *KR+min* improves the Fei–Wang rule by up to 0.107 ($d = 3$) and 0.115 ($d = 4$) and is within 1.1×10^{-2} of the exact or lowest-Casimir optimum everywhere, with the gap growing slowly with n (1×10^{-3} at (4, 6), 7×10^{-3} at (5, 10), 9×10^{-3} at (6, 12)). When the spectrum is unknown, the spectrum-independent rounding rule of Ref. [11] remains the construction with a proven guarantee.

B. Certificates without semidefinite programming

Two consequences of Theorem 2 make the closed-form cloner auditable. First, if the restricted optimum (a program with one scalar per label pair) shows no violation,

TABLE I. Root fidelities in $d = 3$, $p = (0.6, 0.3, 0.1)$ unless stated: Fei–Wang known-spectrum cloner, *KR+min*, optimum restricted to lowest-Casimir channels (rule optimized), and unrestricted optimum. For (5, 10) and (6, 12) the unrestricted semidefinite program is out of reach and the last column is the primal–dual bracket of Sec. VB, which contains the optimum.

n	m	FW	KR+min	OPT _{min}	OPT
2	3 ^b	0.9105	0.9721	0.9727	0.9727
2	4	0.9267	0.9486	0.9493	0.9493
3	4 ^b	0.8718	0.9787	0.9799	0.9799
3	5	0.9117	0.9599	0.9607	0.9608
4	6	0.8850	0.9674	0.9684	0.9684
4	8	0.8509	0.9177	0.9228	0.9228
5	10	0.8103	0.9049	0.9123	[0.9117, 0.9130]
5	10 ^b	0.7808	0.8802	0.8932	[0.8922, 0.8940]
6	12	0.7886	0.8944	0.9029	[0.9022, 0.9035]

^b $p = (0.7, 0.2, 0.1)$.

TABLE II. Root fidelities in $d = 4$ for $p = (0.4, 0.3, 0.2, 0.1)$ (upper block) and $p = (0.55, 0.25, 0.15, 0.05)$ (lower block). The last column is the weight of the lowest-Casimir component in the unrestricted optimum.

n	m	FW	KR+min	OPT _{min}	OPT	lowest
1	3	0.9598	0.9606	0.9606	0.9617	0.863
2	3	0.9570	0.9889	0.9889	0.9889	1.000
2	4	0.9538	0.9704	0.9715	0.9715	0.996
3	6	0.9226	0.9600	0.9615	0.9615 ^c	—
2	3	0.8944	0.9685	0.9685	0.9685	1.000
2	4	0.8882	0.9228	0.9241	0.9241	0.977
3	4	0.8601	0.9746	0.9758	0.9758	1.000

^c Certified equal to OPT_{min} by Theorem 2(ii) without solving the unrestricted SDP.

then OPT equals it exactly; this certified the $d = 4$ instance (3, 6) in Table II, whose unrestricted SDP was out of reach, and reproduced a 65-minute SDP value at (4, 8), $d = 3$, in nine minutes. Second, the variational formula $\text{Tr} \sqrt{X} = \min_{Y \succ 0} \frac{1}{2} [\text{Tr}(XY^{-1}) + \text{Tr} Y]$ (Cauchy–Schwarz and the arithmetic–geometric mean inequality, with equality at $Y = \sqrt{X}$; Appendix A) gives, for every block-diagonal $Y \succ 0$, the weak-duality bound

$$\text{OPT} \leq \frac{1}{2} \sum_w \sqrt{T_w} \text{Tr} Y_w$$

$$+ \sum_{\mu} \max_{\nu \lambda M} \frac{1}{2} P_{n,p}(\mu) \sum_w \sqrt{T_w} \text{Tr} (\Phi_{\lambda, M}(\rho_{\mu})|_w Y_w^{-1}), \quad (7)$$

which coincides with the Frank–Wolfe linearization at $Y = \sqrt{\Xi(x)}$ and tightens along Frank–Wolfe iterations [22] that use only eigendecompositions and a per-label maximization (Appendix F). Starting from *KR+min*, 80 iterations bracket the optimum to 4×10^{-5} at (1, 3) and

to 8×10^{-4} at $(2, 4)$, $(3, 5)$ in $d = 3$ and $(2, 4)$ in $d = 4$, in seconds, against one-shot linearization gaps of 0.02–0.08 on the same instances. At $(5, 10)$ in $d = 3$, where the unrestricted program has 70 label pairs with sector dimensions up to 2520, 150 iterations take two minutes and bracket the optimum to 1.3×10^{-3} and 1.8×10^{-3} for the two spectra of Table I, and at $(6, 12)$, with 133 pairs and sector dimensions up to 5670, twelve minutes give a bracket of width 1.3×10^{-3} ; all three brackets contain the lowest-Casimir optimum, so the closed-form restriction loses at most 10^{-3} there as well.

VI. PAIRWISE OPTIMALITY OF THE PRV CHANNEL

The results above reduce the design of finite-copy cloners to one representation-theoretic question: for a single pair (μ, ν) , which irreducible component λ of $\mathcal{H}_\mu^* \otimes \mathcal{H}_\nu$ maximizes the pair fidelity $F_\lambda = F(\Phi_\lambda(\rho_{p,\mu}), \rho_{p,\nu})$?

Lemma 3 (compression representation). *For every component λ with isometric intertwiner $V_\lambda : \mathcal{H}_\nu \rightarrow \mathcal{H}_\mu \otimes \mathcal{H}_\lambda$, $\pi_\nu(\rho_p) = V_\lambda^\dagger(\pi_\mu(\rho_p) \otimes \pi_\lambda(\rho_p))V_\lambda$. Hence, with $\mathcal{C}_\lambda(\sigma) = V_\lambda^\dagger(\rho_{p,\mu} \otimes \sigma)V_\lambda$, the channel output and the target are the same compression of the input with a uniform and with a thermal ancilla, $\Phi_\lambda(\rho_{p,\mu}) \propto \mathcal{C}_\lambda(I/\dim \lambda)$ and $\rho_{p,\nu} \propto \mathcal{C}_\lambda(\pi_\lambda(\rho_p)/s_\lambda(p))$.*

The lemma holds because the intertwining relation is polynomial in the group element and therefore extends from $U(d)$ to $GL(d, \mathbb{C}) \ni \rho_p$ (Appendix E); the limiting value $F_{\text{orb}} = F(\tau^{(\gamma)}, \tau)$ of Ref. [11] is its asymptotic form. It yields the two limits, whose proofs are in Appendix E.

Theorem 4 (cold limit). *For a compatible pair and $\rho_p \propto e^{-\beta h}$ with $\beta \rightarrow \infty$, $F_\lambda^2 \rightarrow (\dim \mathcal{H}_\mu / \dim \mathcal{H}_\nu) \|(\langle hw_\mu | \otimes I)V_\lambda | hw_\nu \rangle\|^2$, which equals $\dim \mathcal{H}_\mu / \dim \mathcal{H}_\nu$ for the PRV component and is strictly smaller for every other component.*

Proof sketch. The blocks converge to highest-weight projectors and $F(X, |\psi\rangle\langle\psi|)^2 = \langle\psi|X|\psi\rangle$. Equality $\|(\langle hw_\mu | \otimes I)V_\lambda | hw_\nu \rangle\| = 1$ means $V_\lambda | hw_\nu \rangle = | hw_\mu \rangle \otimes y$ with $\|y\| = 1$, and applying the raising operators $E \otimes 1 + 1 \otimes E$, which annihilate $V_\lambda | hw_\nu \rangle$, forces y to be a highest-weight vector of weight $\nu - \mu$; hence $\lambda = \nu - \mu$ (Appendix E). $\square$

Theorem 5 (hot limit). *For $p = (1 + \epsilon h)/d$ with $\sum_i h_i = 0$ and $H_\lambda = \pi_\lambda(\text{diag } h)$,*

$$1 - F_\lambda = \frac{\epsilon^2}{8 \dim \mathcal{H}_\nu} \|V_\lambda^\dagger(I \otimes H_\lambda)V_\lambda\|_{\text{HS}}^2 + O(\epsilon^3), \quad (8)$$

and $(c'_\lambda)^2 \|H_\nu\|^2 \leq \|V_\lambda^\dagger(I \otimes H_\lambda)V_\lambda\|^2 \leq \dim \mathcal{H}_\nu C_0(\lambda) \|h\|^2 / (d^2 - 1)$ with $c'_\lambda = (C_0(\nu) + C_0(\lambda) - C_0(\mu)) / (2C_0(\nu))$, where $C_0(\theta) = C(\theta) - |\theta|^2/d$ is the traceless ($\mathfrak{su}(d)$) Casimir and $\|h\|^2 = \sum_i h_i^2$. If the PRV component is leakage-free, $\Phi_{\text{PRV}}(H_\mu) \propto H_\nu$

for all h , which holds whenever the adjoint representation occurs once in $\text{End}(\mathcal{H}_\nu)$ and in particular for all $d = 2$, then the PRV component minimizes the quadratic coefficient among all components provided $C_0(\lambda) + C_0(\text{PRV}) \geq 2(C_0(\mu) - C_0(\nu))$, which is automatic for $C_0(\nu) \geq C_0(\mu)$.

The identity uses $H_\nu = V_\lambda^\dagger(H_\mu \otimes I + I \otimes H_\lambda)V_\lambda$ and the second-order expansion of the fidelity around the maximally mixed state; the lower bound is the projection onto the Lie-algebra copy of the adjoint representation, whose coefficient follows from Schur's lemma and the Casimir identity $\sum_a (X_a \otimes 1)(1 \otimes X_a) = \frac{1}{2}(C_0^{\text{tot}} - C_0^{(\mu)} - C_0^{(\lambda)})$, and the upper bound from the invariance of the induced bilinear form; Appendix E derives each step. For $d = 2$ the coefficient is the Landé form $\gamma_L = (J(J+1) + j(j+1) - L(L+1)) / (2J(J+1))$ of the projection theorem [23], and

$$1 - F_L = \frac{1}{6}\epsilon^2 J(J+1)(1 - \gamma_L)^2 + O(\epsilon^3) \quad (9)$$

for $p = ((1 + \epsilon)/2, (1 - \epsilon)/2)$: for $J \geq j$ the Cartan coupling $L = J - j$ is optimal, while for $J < j$ it is not. This predicts, and numerics confirm, that the naive conjecture “the PRV component maximizes F_λ for every pair and spectrum” is false: for the depolarizing pair $\mu = (5, 0) \rightarrow \nu = (5, 2)$ at $q = 0.6$ the component $2L = 4$ beats the PRV component (0.9968 against 0.9956), whereas at $q \leq 0.3$ the PRV component wins again. All six failures among 240 qubit cases have $J < j$ and high temperature; in $d = 3$ the PRV component was maximal in all 240 pairs tested, including 132 incompatible ones (in $d = 4$ the certificates of Theorem 2 never showed a violation on an incompatible pair), and the hot-limit minimality held as a matrix inequality over all spectral directions in 151 of 151 cases, with two exact ties. We therefore state:

Conjecture 1. For compatible pairs the PRV component maximizes the pair fidelity at every spectrum; for all pairs it does so at sufficiently low temperature; for depolarizing incompatible pairs at high temperature the maximizer is determined by the transfer coefficient of Theorem 5.

The failing pairs have $J < j$, an output label less polarized than the input, and so lie outside the typical sets of Proposition 2.3 of Ref. [11]; no certificate ever showed a violation on an incompatible pair, so the design rule “lowest-Casimir channel on every pair” stands; its justification is pairwise optimality on typical pairs together with Theorem 2.

For qubits the conjecture is a classical statement. By Lemma 3 the output weight law q_L and the target t are the same Clebsch–Gordan kernel applied to a uniform and to a thermal ancilla, and pairwise optimality for all monotone divergences follows if there is a stochastic map S with $Sq_L = q_L^*$ and $St = t$, that is, if (q_L^*, t) is a relative majorization [24–26] of (q_L, t) (the existence of S is equivalent to the ordering of the Lorenz curves of

the two pairs [27, Ch. 14]). Such an S exists in all 3240 cases with $J \geq j$ that we solved by linear programming, whereas for $J < j$ it fails in 121 of 336 cases, consistent with the region of the counterexamples; it is in general neither covariant nor tridiagonal, and the inequality it encodes splits into stochastic dominance $q_L \geq_{\text{st}} q_{L^*}$, which follows from a spectrum-free partial-sum inequality of Clebsch–Gordan squares that we verified in 288 of 288 cases, and a likelihood-ratio condition that fails for $j \geq 2$. For a single input copy the whole statement is elementary.

Theorem 6 (single-copy qubit input). *For $j = \frac{1}{2}$, all J and all $q \in (0, 1)$, the Cartan coupling $L^* = J - \frac{1}{2}$ is a relative majorization of the only other component $L = J + \frac{1}{2}$ with respect to the thermal target; hence it maximizes every monotone divergence to the target, in particular the pair fidelity.*

Proof. With $N = 2J$, $a = J - M$ and $p = 1/(1 + q)$, the Clebsch–Gordan squares are linear in M and give $q_{L^*}(a) \propto pN - (2p - 1)a$, $q_L(a) \propto (2p - 1)a + p + (1 - p)(N + 1)$, $t(a) \propto q^a$. Both likelihood ratios $q(a)/t(a)$ are increasing in a ($\ell_{L^*}(a + 1)/\ell_{L^*}(a) \geq 1$ reduces to $a \leq p(N - 1)/(2p - 1)$, true for $a \leq N - 1$), and $q_L - q_{L^*}$ is an increasing affine function of a with zero sum, so all tail sums of q_L dominate those of q_{L^*} . With identical orderings the Lorenz curves of the two pairs have the same breakpoints and the dominance of tail sums is the Lorenz ordering; the Blackwell–Strassen theorem [24, 25] supplies S . $\square$

VII. CONCLUSION

We have determined the exact finite-copy optimum of mixed-state cloning at fixed gain as a concave program over Young-label rules and extremal covariant sector maps, solved it for qubits up to $n = 40$ and for qutrits and ququarts at small n , and compared it with the asymptotically optimal Schur–Cartan cloner of Ref. [11] and with purify–clone–trace. This is the first exact solution of the finite-copy problem beyond pure states, and the first identification of the optimal sector channel as the minimal PRV component. The trends are:

- the optimum approaches the asymptotic constant from above and slowly, so that at $n = 20$ the constant underestimates the achievable fidelity by 0.05 for $p = (0.6, 0.4)$, $\gamma = 2$;
- the Schur–Cartan cloner approaches it from below and loses up to 0.11 ($d = 2$), 0.107 ($d = 3$) and 0.115 ($d = 4$), entirely through its label rule, while PCT loses up to 0.26 and never reaches the trivial optimum at a spectral degeneracy;
- the closed-form cloner KR+min, asymptotically optimal by the coupling theorem, closes this gap to at most 1.1×10^{-2} and is certifiable to 10^{-3}

by Frank–Wolfe dual bounds without semidefinite programming.

In the broader development of cloning theory these results connect the estimation-theoretic asymptotics of mixed states [11, 20, 28] with the finite-copy optimality results for pure states [5, 6, 29], and they replace the asymptotic “Cartan” prescription by the finite-copy prescription “lowest-Casimir channel plus optimally coupled labels”. If finite-copy cloners are instead built from the asymptotic constructions, the coherent advantage of quantum cloning over purify–clone–trace at $n \sim 10$ –20, of order 0.1 in fidelity, will be forfeited; conversely, a security analysis that models an adversary by the asymptotic fidelity underestimates the adversary by up to 0.06 at these copy numbers. The remaining question is the intermediate-temperature pairwise optimality of the PRV channel, now reduced to a convex-order inequality for Hahn-polynomial-tilted hypergeometric means, and the hot-limit coefficient for outputs with adjoint multiplicity above one. The design rule does not depend on its answer: Theorem 2(ii) certifies the lowest-Casimir restriction instance by instance, with or without Conjecture 1, and Theorem 3 secures the asymptotic optimality of the KR rule independently of the sector question.

DATA AVAILABILITY

The data that support the findings of this article, namely the reference outputs behind Tables I and II and Figs. 1–3, and the source code that generated them, the Python package `finite_n_cloning_repro` version 1.0.0 with the scripts that regenerate every table and figure and regression tests against every number quoted here, are publicly available at Zenodo [30]. The same record archives the manuscript source.

Appendix A: Fidelity: definitions and the facts used

For positive operators A, C on a finite-dimensional space, the root fidelity is

$$F(A, C) = \|\sqrt{A}\sqrt{C}\|_1 = \text{Tr} \sqrt{\sqrt{C} A \sqrt{C}}, \quad (\text{A1})$$

the two expressions agreeing because $\|X\|_1 = \text{Tr} \sqrt{X^\dagger X}$ with $X = \sqrt{A}\sqrt{C}$ [17, 31]. We use the following properties.

(F1) *Block additivity.* If $A = \bigoplus_i A_i$ and $C = \bigoplus_i C_i$ on the same orthogonal decomposition, then $\sqrt{C} A \sqrt{C} = \bigoplus_i \sqrt{C_i} A_i \sqrt{C_i}$, the square root of a block-diagonal operator is block-diagonal, and the trace adds: $F(A, C) = \sum_i F(A_i, C_i)$.

(F2) *Tensor multiplicativity.* $\sqrt{C_1 \otimes C_2} (A_1 \otimes A_2) \sqrt{C_1 \otimes C_2} = (\sqrt{C_1} A_1 \sqrt{C_1}) \otimes (\sqrt{C_2} A_2 \sqrt{C_2})$, the

square root of a tensor product is the tensor product of the square roots, and the trace multiplies: $F(A_1 \otimes A_2, C_1 \otimes C_2) = F(A_1, C_1)F(A_2, C_2)$.

(F3) *Homogeneity.* $F(cA, C) = F(A, cC) = \sqrt{c} F(A, C)$ for $c \geq 0$, and $F(A, cI) = \sqrt{c} \text{Tr} \sqrt{A}$. In particular $F(I_k/k, I_k/k) = 1$ and $F(I_k, I_k/k) = \sqrt{k}$.

(F4) *Monotonicity.* If $0 \leq A \leq B$ then $F(A, C) \leq F(B, C)$: indeed $\sqrt{CA}\sqrt{C} \leq \sqrt{CB}\sqrt{C}$, the square root is operator monotone [32, Prop. V.1.8], and the trace preserves the order.

(F5) *Concavity.* $A \mapsto F(A, C)$ is concave: $A \mapsto \sqrt{CA}\sqrt{C}$ is linear and $X \mapsto \text{Tr} \sqrt{X}$ is concave on positive operators, because the square root is operator concave [32, Thm. V.2.5]. The fidelity is moreover jointly concave in (A, C) [31, Thm. 3.26], which is what the covariantization argument of Appendix B uses.

(F6) *Derivative.* For $X \succ 0$ and Hermitian A ,

$$\frac{d}{dt} \text{Tr} \sqrt{X + tA} \Big|_{t=0} = \frac{1}{2} \text{Tr}(X^{-1/2}A). \quad (\text{A2})$$

This is the general rule $\frac{d}{dt} \text{Tr} f(X + tA)|_0 = \text{Tr} f'(X)A$ for a function f differentiable on the spectrum [32, Thm. V.3.3]; in the trace case it reduces to first-order perturbation theory, $\frac{d}{dt} \sum_i f(\lambda_i(t)) = \sum_i f'(\lambda_i) \langle i|A|i \rangle$ with $\{|i\rangle\}$ an eigenbasis of X , continuity handling degenerate eigenvalues.

(F7) *Second order at the maximally mixed state.* Let D be the dimension and A, B Hermitian and traceless. Then

$$F\left(\frac{I + \epsilon A}{D}, \frac{I + \epsilon B}{D}\right) = 1 - \frac{\epsilon^2}{8D} \text{Tr}(A - B)^2 + O(\epsilon^3). \quad (\text{A3})$$

Derivation. With $(I + \epsilon B)^{1/2} = I + \frac{\epsilon}{2}B - \frac{\epsilon^2}{8}B^2 + O(\epsilon^3)$, the operator under the square root in Eq. (A1) is $D^{-2}[I + \epsilon(A + B) + \frac{\epsilon^2}{2}(AB + BA) + O(\epsilon^3)]$. Using $\sqrt{I + \epsilon X + \epsilon^2 Y} = I + \frac{\epsilon}{2}X + \epsilon^2(\frac{1}{2}Y - \frac{1}{8}X^2) + O(\epsilon^3)$ with $X = A + B$, $Y = \frac{1}{2}(AB + BA)$, and $\text{Tr} A = \text{Tr} B = 0$, the trace is $D^{-1}[D + \epsilon^2(\frac{1}{2} \text{Tr} AB - \frac{1}{8} \text{Tr}(A + B)^2)] = 1 - \frac{\epsilon^2}{8D} \text{Tr}(A - B)^2$.

(F8) *Variational form of $\text{Tr} \sqrt{X}$.* For $X \succ 0$,

$$\text{Tr} \sqrt{X} = \min_{Y \succ 0} \frac{1}{2} [\text{Tr}(XY^{-1}) + \text{Tr} Y], \quad (\text{A4})$$

with the minimum at $Y = \sqrt{X}$. *Derivation.* For any $Y \succ 0$, the Cauchy–Schwarz inequality for the Hilbert–Schmidt inner product gives $\text{Tr} \sqrt{X} = \text{Tr}[(Y^{-1/2}X^{1/2})^\dagger Y^{1/2}] \leq \|Y^{-1/2}X^{1/2}\|_2 \|Y^{1/2}\|_2 = \sqrt{\text{Tr}(XY^{-1}) \text{Tr} Y} \leq \frac{1}{2}[\text{Tr}(XY^{-1}) + \text{Tr} Y]$, the last step being the arithmetic–geometric mean inequality; at $Y = \sqrt{X}$ both sides equal $\text{Tr} \sqrt{X}$.

Appendix B: Schur–Weyl reduction and covariant channels

1. The block form of $\rho_p^{\otimes N}$

Schur–Weyl duality decomposes $(\mathbb{C}^d)^{\otimes N} = \bigoplus_{\lambda \vdash N, \ell(\lambda) \leq d} \mathcal{H}_\lambda \otimes \mathcal{S}_\lambda$ as a $U(d) \times S_N$ -module, where $\mathcal{H}_\lambda$ is the irreducible polynomial $GL(d)$ -module of highest weight λ and $\mathcal{S}_\lambda$ the Specht module [33, §6.1]. For $g \in GL(d)$ the operator $g^{\otimes N}$ acts as $\bigoplus_\lambda \pi_\lambda(g) \otimes I_{\mathcal{S}_\lambda}$, so with $g = \rho_p$,

$$\rho_p^{\otimes N} = \bigoplus_\lambda \pi_\lambda(\rho_p) \otimes I_{\mathcal{S}_\lambda}. \quad (\text{B1})$$

The character of $\mathcal{H}_\lambda$ is the Schur polynomial, $\text{Tr} \pi_\lambda(\rho_p) = s_\lambda(p_1, \dots, p_d)$ [33, §6.1][34, I.3], so normalizing each block gives Eq. (2) with $P_{N,p}(\lambda) = \dim \mathcal{S}_\lambda s_\lambda(p)$, the Young-label law of Keyl and Werner [20]. Since $\rho_p = \exp(\sum_i \log p_i E_{ii})$ lies in the diagonal torus, $\pi_\lambda(\rho_p)$ is diagonal in any weight basis, with entry $p^w = \prod_i p_i^{w_i}$ on a vector of weight w ; the Gelfand–Tsetlin basis is such a basis [18, 19].

2. Covariant permutation-symmetric channels

Let Φ be a channel from $\mathcal{H}_{\text{in}} = (\mathbb{C}^d)^{\otimes n}$ to $\mathcal{H}_{\text{out}} = (\mathbb{C}^d)^{\otimes m}$ with Choi operator $J = \sum_{a,b} |a\rangle\langle b| \otimes \Phi(|a\rangle\langle b|)$, viewed as an operator on $\mathcal{H}_{\text{in}}^* \otimes \mathcal{H}_{\text{out}}$, where $\mathcal{H}_{\text{in}}^*$ carries the conjugate representation [31, Ch. 2]. Covariance, $\Phi(u^{\otimes n} X u^{\dagger \otimes n}) = u^{\otimes m} \Phi(X) u^{\dagger \otimes m}$ for all $u \in U(d)$, and permutation symmetry on the input and on the output are equivalent to J commuting with the representation of $U(d) \times S_n \times S_m$ on $\mathcal{H}_{\text{in}}^* \otimes \mathcal{H}_{\text{out}} = \bigoplus_{\mu \vdash n, \nu \vdash m} (\mathcal{H}_\mu^* \otimes \mathcal{H}_\nu) \otimes (\mathcal{S}_\mu^* \otimes \mathcal{S}_\nu)$. The $S_n \times S_m$ -isotypic components are indexed by the pairs (μ, ν) , and inequivalent pairs are inequivalent irreducible modules, so by Schur’s lemma J has no cross terms between different pairs and acts as the identity on each $\mathcal{S}_\mu^* \otimes \mathcal{S}_\nu$:

$$J = \bigoplus_{\mu, \nu} J^{\mu\nu} \otimes I_{\mathcal{S}_\mu^* \otimes \mathcal{S}_\nu}, \quad J^{\mu\nu} \in \text{End}(\mathcal{H}_\mu^* \otimes \mathcal{H}_\nu)^{U(d)}. \quad (\text{B2})$$

Complete positivity is $J \succeq 0$, that is, every $J^{\mu\nu} \succeq 0$. Trace preservation, $\text{Tr}_{\text{out}} J = I_{\text{in}}$, reads $\sum_\nu \dim \mathcal{S}_\nu \text{Tr}_{\mathcal{H}_\nu} J^{\mu\nu} = I_{\mathcal{H}_\mu}$ for every μ . Define $\Phi^{\mu\nu}(X) = \dim \mathcal{S}_\nu \text{Tr}_{\mathcal{H}_\mu^*}[J^{\mu\nu}(X^T \otimes I)]$, a completely positive covariant map with $\sum_\nu \text{Tr} \Phi^{\mu\nu}(X) = \text{Tr} X$; by covariance $\text{Tr} \Phi^{\mu\nu}(X) = K(\nu|\mu) \text{Tr} X$ with $K(\nu|\mu) \geq 0$, $\sum_\nu K(\nu|\mu) = 1$, so $\Phi^{\mu\nu} = K(\nu|\mu) E_{\mu\nu}$ with $E_{\mu\nu}$ a covariant channel. This is the lifted form used in Lemma 1.

Applying Φ to the input (2), the output in the sector

ν is

$$\sum_{\mu} P_{n,p}(\mu) \operatorname{Tr}_{\mathcal{H}_{\mu}^* \otimes \mathcal{S}_{\mu}^*} \left[(J^{\mu\nu} \otimes I) (\rho_{p,\mu}^T \otimes \frac{I_{\mathcal{S}_{\mu}^*}}{\dim \mathcal{S}_{\mu}} \otimes I) \right] \\ = \frac{\Xi_{\nu}}{\dim \mathcal{S}_{\nu}} \otimes I_{\mathcal{S}_{\nu}}, \quad (\text{B3})$$

with $\Xi_{\nu} = \sum_{\mu} P_{n,p}(\mu) \Phi^{\mu\nu}(\rho_{p,\mu})$, while the target in the same sector is $T_{\nu} \otimes I_{\mathcal{S}_{\nu}} / \dim \mathcal{S}_{\nu}$ with $T_{\nu} = \dim \mathcal{S}_{\nu} \pi_{\nu}(\rho_p) = P_{m,p}(\nu) \rho_{p,\nu}$. By (F1)–(F3),

$$F = \sum_{\nu} F\left(\frac{\Xi_{\nu}}{\dim \mathcal{S}_{\nu}}, T_{\nu}\right) F\left(I_{\mathcal{S}_{\nu}}, \frac{I_{\mathcal{S}_{\nu}}}{\dim \mathcal{S}_{\nu}}\right) = \sum_{\nu} F(\Xi_{\nu}, T_{\nu}), \quad (\text{B4})$$

using $F(I_{\mathcal{S}}, I_{\mathcal{S}} / \dim \mathcal{S}) = \sqrt{\dim \mathcal{S}}$ and (F3). Both Ξ_{ν} and T_{ν} commute with the diagonal torus, hence are block-diagonal over the weight spaces w of $\mathcal{H}_{\nu}$, and T_{ν} is the scalar $T_{\nu,w} = \dim \mathcal{S}_{\nu} p^w$ on each; (F1) and (F3) then give Eq. (5). Equation (B4) with $T_{\nu} = P_{m,p}(\nu) \rho_{p,\nu}$ is Lemma 1.

Covariantization. For an arbitrary channel $\mathcal{M}$ and its twirl $\overline{\mathcal{M}} = \int du \mathcal{U}_m^{\dagger} \circ \mathcal{M} \circ \mathcal{U}_n$ (and similarly over $\mathcal{S}_n, \mathcal{S}_m$), joint concavity (F5) and invariance of the target give $\inf_g F(\overline{\mathcal{M}}(\rho_{p,g}^{\otimes n}), \rho_{p,g}^{\otimes m}) \geq \inf_g F(\mathcal{M}(\rho_{p,g}^{\otimes n}), \rho_{p,g}^{\otimes m})$, so the optimum in Eq. (1) is attained on covariant permutation-symmetric channels, for which the infimum over g is attained at $g = I$.

3. Extremal covariant maps and their normalization

By Schur's lemma the commutant $\operatorname{End}(\mathcal{H}_{\mu}^* \otimes \mathcal{H}_{\nu})^{U(d)}$ is $\bigoplus_{\lambda} \operatorname{End}(\mathbb{C}^{k_{\lambda}})$, where λ runs over the irreducible components and k_{λ} is the multiplicity: choosing isometries B_a^{λ} ($a = 1, \dots, k_{\lambda}$) onto the k_{λ} copies, generated from orthonormal highest-weight vectors by one fixed sequence of lowering operators so that $\Pi_{ab}^{\lambda} = B_a^{\lambda} B_b^{\lambda\dagger}$ are intertwiners, every element is $\sum_{\lambda} \sum_{ab} M_{ab}^{\lambda} \Pi_{ab}^{\lambda}$ and positivity is $M^{\lambda} \succeq 0$ for all λ . The partial trace $\operatorname{Tr}_{\mathcal{H}_{\nu}} \Pi_{ab}^{\lambda}$ is an intertwiner of the irreducible $\mathcal{H}_{\mu}^*$, hence a scalar, and its trace is $\operatorname{Tr} \Pi_{ab}^{\lambda} = \delta_{ab} \dim \mathcal{H}_{\lambda}$; thus $\operatorname{Tr}_{\mathcal{H}_{\nu}} \Pi_{ab}^{\lambda} = \delta_{ab} (\dim \mathcal{H}_{\lambda} / \dim \mathcal{H}_{\mu}) I$. A trace-preserving covariant map supported on the component λ therefore has Choi operator

$$J_{\lambda,M} = \frac{\dim \mathcal{H}_{\mu}}{\dim \mathcal{H}_{\lambda}} \sum_{ab} M_{ab} \Pi_{ab}^{\lambda}, \quad M \succeq 0, \operatorname{Tr} M = 1, \quad (\text{B5})$$

and the extreme points of the set of covariant channels are the rank-one M . For a component of multiplicity one the covariant channel supported on it is unique. The Cartan channel (3) is covariant, trace preserving (Schur's lemma applied to $\operatorname{Tr}_{\mathcal{H}_{\omega}} V_{\mu,\omega} V_{\mu,\omega}^{\dagger}$ [11]), and supported on the component of highest weight ω , which has multiplicity one (Appendix C); by uniqueness it coincides with Φ_{ω} . In the same way, the map

$X \mapsto (\dim \mathcal{H}_{\mu} / \dim \mathcal{H}_{\nu}) V_{\lambda}^{\dagger} (X \otimes I_{\lambda}) V_{\lambda}$ built from an isometric intertwiner $V_{\lambda} : \mathcal{H}_{\nu} \rightarrow \mathcal{H}_{\mu} \otimes \mathcal{H}_{\lambda}$ is the covariant channel of the component λ whenever that component has multiplicity one, and in general it is $\Phi_{\lambda,M}$ for the M determined by V_{λ} .

4. Qubits: Clebsch–Gordan kernels and the Landé coefficient

For $d = 2$ write $\mathcal{H}_{\mu} \cong \mathcal{H}_j$ with $2j = \mu_1 - \mu_2$ and $\mathcal{H}_{\nu} \cong \mathcal{H}_J$, $2J = \nu_1 - \nu_2$. The components of $\mathcal{H}_j^* \otimes \mathcal{H}_J \cong \mathcal{H}_j \otimes \mathcal{H}_J$ are the spins $L \in \{|J - j|, \dots, J + j\}$, each with multiplicity one, and $V_L |J, M\rangle = \sum_m \langle j, m; L, M - m | J, M \rangle |j, m\rangle |L, M - m\rangle$. For a weight-diagonal input $X = \sum_m x_m |j, m\rangle \langle j, m|$,

$$\langle J, M | \Phi_L(X) | J, M' \rangle \\ = \frac{2j+1}{2J+1} \sum_m \langle J, M | (|j, m\rangle \langle j, m| \otimes I) | J, M' \rangle x_m \\ = \delta_{MM'} \frac{2j+1}{2J+1} \sum_m |\langle j, m; L, M - m | J, M \rangle|^2 x_m, \quad (\text{B6})$$

because the ancilla states $|L, M - m\rangle$ and $|L, M' - m\rangle$ are orthogonal unless $M = M'$. Trace preservation follows from the orthogonality relation $\sum_M |\langle j, m; L, M - m | J, M \rangle|^2 = (2J+1)/(2j+1)$ [23, Eq. (3.5.14)]. In the lowering-number indexing $a = J - M$, $a_1 = j - m$ this is the kernel quoted in Appendix G.

The Landé coefficient is the projection theorem for vector operators [23, §5.4]: within the multiplet J embedded in $\mathcal{H}_j \otimes \mathcal{H}_L$, the components of the ancilla spin $\mathbf{L}$ satisfy $V_L^{\dagger} (I \otimes L_z) V_L = a_L J_z$ with $a_L = \langle \mathbf{L} \cdot \mathbf{J} \rangle / J(J+1)$. From $\mathbf{j} = \mathbf{J} - \mathbf{L}$, $j(j+1) = J(J+1) + L(L+1) - 2\langle \mathbf{L} \cdot \mathbf{J} \rangle$, so

$$a_L = \frac{J(J+1) + L(L+1) - j(j+1)}{2J(J+1)}, \quad \gamma_L = 1 - a_L. \quad (\text{B7})$$

Appendix C: Proof of Theorem 1

1. The quadratic Casimir and dominance order

Let E_{ij} be the matrix units of $\mathfrak{gl}(d)$ and $C = \sum_{i,j} E_{ij} E_{ji}$ the quadratic Casimir; it is central and acts on the irreducible module of highest weight θ by a scalar $C(\theta)$. On a highest-weight vector v , $E_{ij} v = 0$ for $i < j$ and $E_{ii} v = \theta_i v$; using $E_{ij} E_{ji} = E_{ji} E_{ij} + E_{ii} - E_{jj}$ for $i < j$,

$$Cv = \left[\sum_i \theta_i^2 + \sum_{i < j} (\theta_i - \theta_j) \right] v \\ = \sum_i \theta_i (\theta_i + d + 1 - 2i) v = \langle \theta, \theta + 2\rho \rangle v, \quad (\text{C1})$$

with $\rho = \frac{1}{2}(d-1, d-3, \dots, 1-d)$ and $\langle \cdot, \cdot \rangle$ the Euclidean product. The traceless part $C_0(\theta) = C(\theta) - |\theta|^2/d$, $|\theta| = \sum_i \theta_i$, is the Casimir of $\mathfrak{su}(d)$ on the same module.

For integral weights θ, θ' with $|\theta| = |\theta'|$, dominance order $\theta' \succeq \theta$ means $\sum_{i \leq k} \theta'_i \geq \sum_{i \leq k} \theta_i$ for all k , equivalently $\theta' - \theta = \sum_{i < d} c_i(e_i - e_{i+1})$ with $c_i = \sum_{l \leq i} (\theta'_l - \theta_l) \geq 0$ [35, §13.2].

Lemma 4. *If θ, θ' are dominant, $|\theta| = |\theta'|$ and $\theta' \succeq \theta$, then $C(\theta') \geq C(\theta) + 2 \sum_i c_i$, with equality in $C(\theta') = C(\theta)$ only for $\theta' = \theta$.*

Proof. From (C1), $C(\theta') - C(\theta) = \langle \theta' - \theta, \theta' + \theta + 2\rho \rangle = \sum_i c_i \langle e_i - e_{i+1}, \theta' + \theta + 2\rho \rangle = \sum_i c_i [(\theta'_i - \theta'_{i+1}) + (\theta_i - \theta_{i+1}) + 2] \geq 2 \sum_i c_i$, since dominance of θ, θ' makes the two differences nonnegative. $\square$

2. The PRV component

The lowest weight of $\mathcal{H}_\mu^*$ is $-\mu$, and its highest weight is $\mu^* = (-\mu_d, \dots, -\mu_1) = -w_0\mu$. Parthasarathy, Ranga Rao and Varadarajan proved that for dominant weights α, β the tensor product $\mathcal{H}_\alpha \otimes \mathcal{H}_\beta$ contains the irreducible module of highest weight $\text{dom}(\alpha + w_0\beta)$ with multiplicity exactly one [15]; Kumar proved the conjecture of the same paper that $\text{dom}(\alpha + w\beta)$ occurs for every w [16]. With $\alpha = \nu$, $\beta = \mu^*$ and $w_0\mu^* = -\mu$, the component of highest weight $\text{dom}(\nu - \mu)$ occurs in $\mathcal{H}_\mu^* \otimes \mathcal{H}_\nu$ with multiplicity one. For a compatible pair, $\nu - \mu$ is already dominant and the vector $V_{\mu, \omega}$ of Eq. (3) identifies this component with the Cartan channel (Appendix B3).

3. Every component dominates the PRV component

Lemma 5. *Every irreducible component of $\mathcal{H}_\mu^* \otimes \mathcal{H}_\nu$ has highest weight $\theta \succeq \text{dom}(\nu - \mu)$.*

Proof. All components have $|\theta| = m - n$. By duality, $\text{Hom}(\mathcal{H}_\theta, \mathcal{H}_\mu^* \otimes \mathcal{H}_\nu) \cong \text{Hom}(\mathcal{H}_\mu \otimes \mathcal{H}_\theta, \mathcal{H}_\nu)$, and tensoring with $\det^t$, $t \geq \max(0, -\theta_d)$, turns θ into a partition $\theta' = \theta + t\mathbf{1}$ and ν into $\nu' = \nu + t\mathbf{1}$ without changing the multiplicity. Thus the Littlewood–Richardson coefficient $c_{\mu\theta'}^{\nu'}$ is positive [33, §A.1]. By Klyachko’s theorem and the saturation theorem of Knutson and Tao, positivity of $c_{\mu\theta'}^{\nu'}$ is equivalent to the existence of $d \times d$ Hermitian matrices A, B with spectra μ, θ' and $A + B$ with spectrum ν' [36, 37][38, Thm. 11]. For such matrices the Lidskii–Wielandt inequality [32, Thm. III.4.1] states that for every k and every k -subset $S \subset \{1, \dots, d\}$,

$$\sum_{i \in S} \nu'_i \leq \sum_{i \in S} \mu_i + \sum_{j \leq k} \theta'_j. \quad (\text{C2})$$

Taking the maximum over S , $\sum_{j \leq k} \theta'_j \geq \max_{|S|=k} \sum_{i \in S} (\nu'_i - \mu_i) = \sum_{i \leq k} \text{dom}(\nu' - \mu)_i$, which is $\theta' \succeq \text{dom}(\nu - \mu) + t\mathbf{1}$, i.e. $\theta \succeq \text{dom}(\nu - \mu)$. $\square$

For compatible pairs an elementary argument suffices: $c_{\mu\theta'}^{\nu'} > 0$ means that the skew shape ν/μ admits a Littlewood–Richardson filling of content θ' , in which every entry in row r is at most r [34, I.9]; hence the rows $1, \dots, k$ of ν/μ , which contain $\sum_{i \leq k} (\nu_i - \mu_i)$ boxes, contain only entries $\leq k$, of which there are $\sum_{j \leq k} \theta'_j$ in total, and $\theta' \succeq \nu - \mu$ follows.

4. Conclusion of the proof

By Lemma 5 every component $\theta \neq \text{dom}(\nu - \mu)$ satisfies $\theta \succeq \text{dom}(\nu - \mu)$ with some $c_i > 0$, so by Lemma 4 $C(\theta) > C(\text{dom}(\nu - \mu))$. Hence the component of smallest Casimir is the PRV component, it is unique, and it has multiplicity one. Numerically, in all 1095 components of the 240 pairs of $d = 3$ and the $d = 4$ pairs with $n \leq 3$, $m \leq 6$ the lowest Casimir eigenspace had highest weight $\text{dom}(\nu - \mu)$ and dimension $\dim \mathcal{H}_{\text{dom}(\nu - \mu)}$, as the theorem requires. $\square$

Appendix D: Proof of Theorem 2 and the single-feeder bound

1. The program and its derivative

A point x of the program assigns to each input label μ a probability distribution w_μ over triples (ν, λ, M) with M a density matrix on the multiplicity space of $\lambda \subset \mathcal{H}_\mu^* \otimes \mathcal{H}_\nu$; the output blocks are $\Xi_\nu(x) = \sum_\mu P_{n,p}(\mu) \sum_{\lambda, M} w_\mu(\nu, \lambda, M) \Phi_{\lambda, M}(\rho_{p, \mu})$, affine in x , and $F(x) = \sum_{\nu, w} \sqrt{T_{\nu, w}} \text{Tr} \sqrt{\Xi_{\nu, w}(x)}$. This covers every covariant permutation-symmetric channel by Appendix B3; the restriction “lowest-Casimir channels” is w_μ supported on $\lambda = \lambda_{\min}(\mu, \nu)$, $M = 1$, in which case w_μ is the rule $K(\cdot|\mu)$.

If $\Xi_{\nu, w}(x) \succ 0$ on every block, (F6) gives the directional derivative of F at x towards any feasible x' ,

$$DF(x)[x'] = \sum_\mu \sum_{\nu, \lambda, M} w'_\mu(\nu, \lambda, M) g_{\mu\nu\lambda}(M; x), \quad (\text{D1})$$

with g as defined before Theorem 2; g is linear in M , so $g_{\mu\nu\lambda}^* = \max_M g = \lambda_{\max}(\Gamma)$ for the $k \times k$ matrix $\Gamma_{ab} = g_{\mu\nu\lambda}(\frac{1}{2}(|a\rangle\langle b| + |b\rangle\langle a|); x)$.

2. Proof of (i)

At the optimum x of the restricted program, F is a concave function of the rules $K(\cdot|\mu)$, each ranging over a simplex. The Karush–Kuhn–Tucker conditions for maximizing a differentiable concave function over a product of simplices [39, §5.5] state that for each μ the partial derivatives $\partial F / \partial K(\nu|\mu) = g_{\mu\nu\lambda_{\min}}(x)$ take a common value η_μ on the support of $K(\cdot|\mu)$ and are at most η_μ off the support.

3. Proof of (iii)

By concavity, $F(x') \leq F(x) + DF(x)[x' - x]$ for every feasible x' . The map $x \mapsto \Xi(x)$ is positively homogeneous of degree one and $\text{Tr} \sqrt{\cdot}$ of degree one half, so $F(tx) = \sqrt{t}F(x)$ and, differentiating at $t = 1$, $DF(x)[x] = F(x)/2$; by (i) and (D1), $DF(x)[x] = \sum_{\mu} \sum_{\nu} K(\nu|\mu)\eta_{\mu} = \sum_{\mu} \eta_{\mu}$. Also by (D1), $DF(x)[x'] \leq \sum_{\mu} \max_{\nu, \lambda} g_{\mu\nu\lambda}^*$, since w'_{μ} is a probability distribution. Hence $F(x') \leq F(x) + \sum_{\mu} (\max_{\nu, \lambda} g_{\mu\nu\lambda}^* - \eta_{\mu})$; the maximum is at least η_{μ} because the lowest-Casimir triples on the support of K attain η_{μ} , so the positive part changes nothing.

4. Proof of (ii)

If $g_{\mu\nu\lambda}^* \leq \eta_{\mu}$ for all triples, (iii) gives $\text{OPT} \leq F(x)$, so x is optimal. Conversely, if $g_{\mu\nu\lambda}(M; x) > \eta_{\mu}$ for some triple, let x' replace w_{μ} by the point mass at (ν, λ, M) and leave the other labels unchanged; then $DF(x)[x' - x] = g_{\mu\nu\lambda}(M; x) - \sum_{\nu'} K(\nu'|\mu)\eta_{\mu} = g - \eta_{\mu} > 0$, so F increases along the feasible segment from x to x' and x is not optimal. $\square$

5. The single-feeder bound

Proposition 1. *Suppose a label μ' feeds the block ν with rule weight $\kappa = K(\nu|\mu') > 0$ in x , with lowest-Casimir output $Y = \Phi_{\lambda_{\min}}(\rho_{p, \mu'})$, and that for some component λ of a label μ and some $c > 0$ the admixture pattern satisfies $\Phi_{\lambda}(\rho_{p, \mu})|_w \leq cY|_w$ on every weight block. Then*

$$g_{\mu\nu\lambda}^*(x) \leq \frac{1}{2}P_{n,p}(\mu) \sqrt{\frac{c}{P_{n,p}(\mu')\kappa}} \times \sum_w \sqrt{T_{\nu,w}} \text{Tr} \sqrt{\Phi_{\lambda}(\rho_{p, \mu})|_w}. \quad (\text{D2})$$

Proof. Write $A_w = \Phi_{\lambda}(\rho_{p, \mu})|_w$. Since $\Xi_{\nu,w}(x) \geq P_{n,p}(\mu')\kappa Y|_w \geq sA_w$ with $s = P_{n,p}(\mu')\kappa/c$, and $t \mapsto t^{-1/2}$ is operator monotone decreasing on positive definite operators [32, Ch. V], we have $\Xi_{\nu,w}^{-1/2} \leq s^{-1/2}A_w^{-1/2}$ on the support of A_w (approximating a singular A_w by $A_w + \delta I$ and letting $\delta \rightarrow 0$). Hence $\text{Tr}(\Xi_{\nu,w}^{-1/2}A_w) = \text{Tr}(A_w^{1/2}\Xi_{\nu,w}^{-1/2}A_w^{1/2}) \leq s^{-1/2}\text{Tr}\sqrt{A_w}$, and the claim follows from the definition of g (the multiplicity density only changes the pattern A_w , to which the same argument applies). $\square$

Appendix E: Proofs for Section VI

1. Lemma 3

The intertwiner V_{λ} satisfies $\pi_{\nu}(u) = V_{\lambda}^{\dagger}(\pi_{\mu}(u) \otimes \pi_{\lambda}(u))V_{\lambda}$ for all $u \in U(d)$. Both sides are polyno-

mial functions of the matrix entries of u (the modules are polynomial representations of $GL(d)$), and $U(d)$ is Zariski dense in $GL(d, \mathbb{C})$, so the identity extends to all $g \in GL(d, \mathbb{C})$ [33, §6.1], in particular to $g = \rho_p$. Dividing by $s_{\nu}(p) = s_{\mu}(p)s_{\lambda}(p) \cdot [s_{\nu}(p)/(s_{\mu}(p)s_{\lambda}(p))]$ gives $\rho_{p, \nu} = \frac{s_{\mu}(p)s_{\lambda}(p)}{s_{\nu}(p)}V_{\lambda}^{\dagger}(\rho_{p, \mu} \otimes \rho_{p, \lambda})V_{\lambda}$, while $\Phi_{\lambda}(\rho_{p, \mu}) = \frac{\dim \mathcal{H}_{\mu} \dim \mathcal{H}_{\lambda}}{\dim \mathcal{H}_{\nu}}V_{\lambda}^{\dagger}(\rho_{p, \mu} \otimes \frac{I}{\dim \mathcal{H}_{\lambda}})V_{\lambda}$. Both are the compression $\mathcal{C}_{\lambda}$ of the input with an ancilla in the thermal state $\rho_{p, \lambda}$ and in the maximally mixed state, respectively. $\square$

2. Theorem 4

As $\beta \rightarrow \infty$ with $\rho_p \propto e^{-\beta h}$, $h_1 < \dots < h_d$, the normalized blocks converge to the highest-weight projectors, $\rho_{p, \mu} \rightarrow |hw_{\mu}\rangle\langle hw_{\mu}|$ and $\rho_{p, \nu} \rightarrow |hw_{\nu}\rangle\langle hw_{\nu}|$, because $p^w/p^{\theta} \rightarrow 0$ for every weight $w \neq \theta$ of a module with highest weight θ . The fidelity is continuous, and for pure targets $F(X, |\psi\rangle\langle\psi|) = \sqrt{\langle\psi|X|\psi\rangle}$, so

$$F_{\lambda}^2 \rightarrow \langle hw_{\nu}|\Phi_{\lambda}(|hw_{\mu}\rangle\langle hw_{\mu}|)|hw_{\nu}\rangle = \frac{\dim \mathcal{H}_{\mu}}{\dim \mathcal{H}_{\nu}} \|(\langle hw_{\mu}| \otimes I)V_{\lambda}|hw_{\nu}\rangle\|^2. \quad (\text{E1})$$

The norm is at most one. It equals one iff $V_{\lambda}|hw_{\nu}\rangle = |hw_{\mu}\rangle \otimes y$ with $\|y\| = 1$. Applying a raising operator $E = E_{i, i+1}$ in the diagonal action, $0 = (E \otimes 1 + 1 \otimes E)V_{\lambda}|hw_{\nu}\rangle = |hw_{\mu}\rangle \otimes Ey$, so y is a highest-weight vector of $\mathcal{H}_{\lambda}$, of weight $\nu - \mu$ by weight counting; hence $\lambda = \nu - \mu$ and the component is the PRV (Cartan) one. For every other component the norm is strictly smaller than one. $\square$

3. Theorem 5

Expansion of the blocks. With $p = (1 + \epsilon h)/d$, $\sum_i h_i = 0$, and $H_{\theta} = \pi_{\theta}(\text{diag } h)$ (diagonal in the Gelfand-Tsetlin basis, $\text{Tr } H_{\theta} = 0$ by Weyl-group symmetry of the weights), $\pi_{\theta}(\rho_p) = d^{-|\theta|}\pi_{\theta}(I + \epsilon \text{diag } h) = d^{-|\theta|}(I + \epsilon H_{\theta} + O(\epsilon^2))$, so $\rho_{p, \theta} = (I + \epsilon H_{\theta})/\dim \mathcal{H}_{\theta} + O(\epsilon^2)$. The channel is covariant and $\Phi_{\lambda}(I_{\mu}) = (\dim \mathcal{H}_{\mu}/\dim \mathcal{H}_{\nu})I_{\nu}$ (Schur's lemma), and the intertwining property of V_{λ} for the Lie algebra reads $H_{\nu} = V_{\lambda}^{\dagger}(H_{\mu} \otimes I + I \otimes H_{\lambda})V_{\lambda}$. Hence

$$\Phi_{\lambda}(\rho_{p, \mu}) = \frac{1}{\dim \mathcal{H}_{\nu}} [I + \epsilon(H_{\nu} - T_{\lambda}(h))] + O(\epsilon^2), \quad (\text{E2})$$

$$T_{\lambda}(h) = V_{\lambda}^{\dagger}(I \otimes H_{\lambda})V_{\lambda}, \quad (\text{E3})$$

and (F7) with $A = H_{\nu} - T_{\lambda}(h)$, $B = H_{\nu}$ gives the displayed formula $1 - F_{\lambda} = \frac{\epsilon^2}{8 \dim \mathcal{H}_{\nu}} \|T_{\lambda}(h)\|_{\text{HS}}^2 + O(\epsilon^3)$.

Lower bound. The map $h \mapsto T_{\lambda}(h)$ from the traceless diagonal matrices extends to a linear map $T_{\lambda} : \mathfrak{su}(d) \rightarrow \text{End}(\mathcal{H}_{\nu})$, $X \mapsto V_{\lambda}^{\dagger}(I \otimes \pi_{\lambda}(X))V_{\lambda}$, which is equivariant for the adjoint action of $U(d)$ on $\mathfrak{su}(d)$ and conjugation on $\text{End}(\mathcal{H}_{\nu})$. The subspace $\mathcal{A} = \{\pi_{\nu}(X) : X \in \mathfrak{su}(d)\} \subset \text{End}(\mathcal{H}_{\nu})$ is an irreducible submodule isomorphic to the

adjoint representation, and the orthogonal projection P_A for the Hilbert–Schmidt product is equivariant; so $P_A \circ T_\lambda$ is an equivariant map from the adjoint representation to itself and by Schur’s lemma equals $c'_\lambda \pi_\nu$ for a scalar c'_λ . Then $\|T_\lambda(h)\|^2 \geq \|P_A T_\lambda(h)\|^2 = (c'_\lambda)^2 \|\pi_\nu\|^2$. To compute c'_λ , take an orthonormal basis $\{X_a\}$ of $\mathfrak{su}(d)$ for the trace form and sum: $\sum_a \text{Tr}(\pi_\nu(X_a) T_\lambda(X_a)) = c'_\lambda \sum_a \text{Tr} \pi_\nu(X_a)^2 = c'_\lambda \dim \mathcal{H}_\nu C_0(\nu)$. On the left, with $P = V_\lambda V_\lambda^\dagger$ the projector onto $\mathcal{H}_\nu \subset \mathcal{H}_\mu \otimes \mathcal{H}_\lambda$ and $\pi_\nu(X) = V_\lambda^\dagger (X \otimes 1 + 1 \otimes X) V_\lambda$,

$$\begin{aligned} & \sum_a \text{Tr}[P(X_a \otimes 1 + 1 \otimes X_a)(1 \otimes X_a)] \\ &= \text{Tr}\left[P\left(\sum_a (X_a \otimes 1)(1 \otimes X_a) + 1 \otimes C_0^{(\lambda)}\right)\right], \quad (\text{E4}) \end{aligned}$$

and the Casimir identity $\sum_a (X_a \otimes 1)(1 \otimes X_a) = \frac{1}{2}(C_0^{\text{tot}} - C_0^{(\mu)} \otimes 1 - 1 \otimes C_0^{(\lambda)})$ (expand $C_0^{\text{tot}} = \sum_a (X_a \otimes 1 + 1 \otimes X_a)^2$) evaluates on the range of P to $\dim \mathcal{H}_\nu [\frac{1}{2}(C_0(\nu) - C_0(\mu) - C_0(\lambda)) + C_0(\lambda)]$. Therefore

$$c'_\lambda = \frac{C_0(\nu) + C_0(\lambda) - C_0(\mu)}{2C_0(\nu)}. \quad (\text{E5})$$

Here the traceless Casimir C_0 must be used; with C in its place the coefficient is wrong whenever $|\lambda| \neq 0$ (we checked both against direct evaluation of $\text{Tr}(H_\nu T_\lambda(h))/\|\pi_\nu\|^2$).

Upper bound. $\|T_\lambda(h)\|^2 = \text{Tr}[V_\lambda^\dagger (I \otimes H_\lambda) P (I \otimes H_\lambda) V_\lambda] \leq \text{Tr}[P(I \otimes H_\lambda^2)]$ since $P \leq I$. The bilinear form $(X, Y) \mapsto \text{Tr}[P(I \otimes \pi_\lambda(X) \pi_\lambda(Y))]$ on $\mathfrak{su}(d)$ is invariant under the adjoint action (because P commutes with the diagonal action), hence a multiple $m \text{Tr}(XY)$ of the trace form, and summing over an orthonormal basis gives $m(d^2 - 1) = \text{Tr}[P(I \otimes C_0^{(\lambda)})] = \dim \mathcal{H}_\nu C_0(\lambda)$. With $\|h\|^2 = \text{Tr}(\text{diag } h)^2$ this is the stated bound $\|T_\lambda(h)\|^2 \leq \dim \mathcal{H}_\nu C_0(\lambda) \|h\|^2 / (d^2 - 1)$.

Leakage-free PRV component. If $\Phi_{\text{PRV}}(H_\mu) \propto H_\nu$ for all h , then $T_{\text{PRV}}(h) = H_\nu - (\dim \mathcal{H}_\nu / \dim \mathcal{H}_\mu) \Phi_{\text{PRV}}(H_\mu) \in \mathcal{A}$, so the lower bound is attained: $\|T_{\text{PRV}}(h)\|^2 = (c'_{\text{PRV}})^2 \|\pi_\nu\|^2$. For any other component, $\|T_\lambda(h)\|^2 \geq (c'_\lambda)^2 \|\pi_\nu\|^2$, and $(c'_\lambda)^2 \geq (c'_{\text{PRV}})^2$ holds whenever $c'_\lambda - c'_{\text{PRV}} = (C_0(\lambda) - C_0(\text{PRV})) / 2C_0(\nu) \geq 0$, which is Theorem 1 (the traceless and full Casimirs differ by the same constant on all components of a fixed pair), and $c'_\lambda + c'_{\text{PRV}} \geq 0$, which is $C_0(\lambda) + C_0(\text{PRV}) \geq 2(C_0(\mu) - C_0(\nu))$. This condition is automatic when $C_0(\nu) \geq C_0(\mu)$. The adjoint representation occurs once in $\text{End}(\mathcal{H}_\nu)$ whenever ν is a one-row or one-column shape (up to determinant factors) and for every ν when $d = 2$; in those cases $\Phi_{\text{PRV}}(H_\mu)$, which is an equivariant image of the adjoint representation, must lie in $\mathcal{A}$, so the PRV component is automatically leakage-free. $\square$

Qubits. For $d = 2$, $h = \text{diag}(1, -1) = 2j_z$, $H_\lambda = 2L_z$ and, by (B7), $T_L(h) = 2a_L J_z$ with $\|J_z\|^2 = \text{Tr } J_z^2 =$

$\frac{1}{3}J(J+1)(2J+1)$; the formula gives

$$\begin{aligned} 1 - F_L &= \frac{\epsilon^2}{8(2J+1)} \cdot 4a_L^2 \cdot \frac{J(J+1)(2J+1)}{3} + O(\epsilon^3) \\ &= \frac{\epsilon^2}{6} J(J+1)(1 - \gamma_L)^2 + O(\epsilon^3), \quad (\text{E6}) \end{aligned}$$

which we verified against the exact pair fidelity at $\epsilon = 10^{-2}$ to four digits for several (j, L, J) . Since γ_L decreases strictly in L , for $J \geq j$ one has $\gamma_{J-j} - 1 = -(J-j)/J \leq 0$ and $|\gamma_L - 1|$ is smallest at $L = J-j$; for $J < j$, $\gamma_{j-J} - 1 = (j-J)/(J+1) > 0$ and γ_L crosses one as L grows, so the minimizer is the L with $L(L+1)$ closest to $j(j+1) - J(J+1)$, which is not the PRV component in general.

Appendix F: Weak duality and the Frank–Wolfe certificate

By (F8) applied blockwise, for any x ,

$$F(x) = \min_{Y \succ 0} \frac{1}{2} \sum_{\nu, w} \sqrt{T_{\nu, w}} [\text{Tr}(\Xi_{\nu, w}(x) Y_{\nu, w}^{-1}) + \text{Tr } Y_{\nu, w}], \quad (\text{F1})$$

the minimum over weight-block-diagonal Y . Since $\max_x \min_Y \leq \min_Y \max_x$, for every fixed Y ,

$$\begin{aligned} \text{OPT} &\leq \frac{1}{2} \sum_{\nu, w} \sqrt{T_{\nu, w}} \text{Tr } Y_{\nu, w} \\ &\quad + \max_x \frac{1}{2} \sum_{\nu, w} \sqrt{T_{\nu, w}} \text{Tr}(\Xi_{\nu, w}(x) Y_{\nu, w}^{-1}). \quad (\text{F2}) \end{aligned}$$

The second term is linear in x , and x ranges over a product of simplices (one per μ), so the maximum is attained at a vertex: for each μ the best triple (ν, λ, M) , with M the top eigenvector of the $k \times k$ matrix obtained by inserting $\Phi_{\lambda, M}(\rho_{p, \mu})$; this gives the bound displayed in Sec. VB. At $Y = \sqrt{\Xi(x)}$ the first term is $F(x)/2$ and the second is $\sum_\mu \max_{\nu, \lambda} g_{\mu\nu\lambda}^*(x)$ by (A2), so the bound coincides with the Frank–Wolfe linearization $F(x) + DF(x)[x' - x]$ maximized over x' (Appendix D). The Frank–Wolfe iteration [22] moves x towards the maximizing vertex by a line search, which requires only the objective (eigendecompositions of the weight blocks); by concavity $F(x_k) \leq \text{OPT} \leq$ the bound at $Y = \sqrt{\Xi(x_k)}$ at every step, and the primal and dual sequences converge to OPT at rate $O(1/k)$, standard for Frank–Wolfe on a compact convex domain. Strong duality in Eq. (A4) holds because the objective is a minimum over Y of functions linear in x and the domain of x is compact and convex.

Appendix G: Numerical methods

Qubits. For $d = 2$, Eq. (5) with weight-diagonal blocks reads $F = \sum_\nu \sum_a \sqrt{\Xi_\nu[a] T_\nu[a]}$. The extremal covariant

maps from spin j to spin J act on weight-diagonal inputs as $\Phi^{(L)}(X)[a] = \frac{2j+1}{2J+1} \sum_{a_1} |\langle j, a_1; L, a - a_1 | J, a \rangle|^2 X[a_1]$ (lowering-number indexing; derived in Appendix B 4); the squared Clebsch–Gordan coefficients are obtained by diagonalizing J^2 on each total-weight subspace. The optimum over rules and intermediate spins is a second-order cone program in $\sim 10^4$ variables (Clarabel [40], cross-checked with SCS to 10^{-6}); PCT is evaluated in closed form through the Cauchy decomposition of $\text{Sym}^m(\mathbb{C}^2 \otimes \mathbb{C}^2)$ in exact rational arithmetic and validated against a direct 4^m -dimensional computation.

Qudits. Irreducible representations are generated from the orthonormal Gelfand–Tsetlin matrix elements [18, 19], checked against the $\mathfrak{gl}(d)$ relations and Casimir values. The lowest-Casimir channel of a pair is the normalized projector onto the lowest eigenspace of the quadratic Casimir on $\mathcal{H}_\mu^* \otimes \mathcal{H}_\nu$; for compatible pairs it coincides with the Cartan channel built from an explicit tensor-power realization (agreement 10^{-6} in the Choi matrix). The unrestricted optimum imposes covariance of

each $J^{\mu\nu}$ through the $\mathfrak{u}(d)$ generators, blocks the variable by relative weight, and represents $\text{Tr} \sqrt{\cdot}$ through the SDP $\max\{\text{Tr} Y : \begin{pmatrix} X & Y \\ Y^\top & I \end{pmatrix} \succeq 0\}$. Isotypic components with their intertwiners are obtained from the joint weight decomposition of the highest-weight space.

Appendix H: Reproducibility

All engines, regression tests against the numbers quoted here, the scripts that regenerate every table and figure, and the reference outputs are collected in the package `finite_n_cloning_repro` (Python; `numpy`, `scipy`, `cvxpy` with Clarabel), archived together with this manuscript at Zenodo [30], [doi:10.5281/zenodo.23588754](https://doi.org/10.5281/zenodo.23588754). The quick verification runs in under a minute; the full reproduction takes several hours, dominated by the $d = 3$ (4, 8) unrestricted SDP (65 minutes) and the $d = 3$ (5, 10) restricted program (80 minutes); the certificates of Sec. VB at (5, 10) and (6, 12) take two and twelve minutes.

-
- [1] W. K. Wootters and W. H. Zurek, A single quantum cannot be cloned, *Nature* **299**, 802 (1982).
 - [2] D. Dieks, Communication by EPR devices, *Phys. Lett. A* **92**, 271 (1982).
 - [3] V. Bužek and M. Hillery, Quantum copying: Beyond the no-cloning theorem, *Phys. Rev. A* **54**, 1844 (1996).
 - [4] N. Gisin and S. Massar, Optimal quantum cloning machines, *Phys. Rev. Lett.* **79**, 2153 (1997).
 - [5] R. F. Werner, Optimal cloning of pure states, *Phys. Rev. A* **58**, 1827 (1998).
 - [6] M. Keyl and R. F. Werner, Optimal cloning of pure states, testing single clones, *J. Math. Phys.* **40**, 3283 (1999).
 - [7] J. Bae and A. Acín, Asymptotic quantum cloning is state estimation, *Phys. Rev. Lett.* **97**, 030402 (2006).
 - [8] G. Chiribella and Y. Yang, Optimal asymptotic cloning machines, *New J. Phys.* **16**, 063005 (2014).
 - [9] V. Scarani, S. Iblisdir, N. Gisin, and A. Acín, Quantum cloning, *Rev. Mod. Phys.* **77**, 1225 (2005).
 - [10] H. Barnum, C. M. Caves, C. A. Fuchs, R. Jozsa, and B. Schumacher, Noncommuting mixed states cannot be broadcast, *Phys. Rev. Lett.* **76**, 2818 (1996).
 - [11] J. Fei and J. Wang, Asymptotically optimal mixed-state cloning, *arXiv:2609.35986* (2026).
 - [12] M. Fanizza, D. Grinko, T. Scharnhorst, and J. Spilecki, Optimal cloning of mixed states, *arXiv:2608.27298* (2026).
 - [13] S. Jeon, V. Sohn, and C. Oh, Optimal copy complexity of quantum state cloning, *arXiv:2608.24484* (2026).
 - [14] E. Tang, J. Wright, and M. Zhandry, Conjugate queries can help, *arXiv:2510.07622* (2025).
 - [15] K. R. Parthasarathy, R. Ranga Rao, and V. S. Varadarajan, Representations of complex semi-simple Lie groups and Lie algebras, *Ann. Math.* **85**, 383 (1967).
 - [16] S. Kumar, Proof of the Parthasarathy–Ranga Rao–Varadarajan conjecture, *Invent. Math.* **93**, 117 (1988).
 - [17] A. Uhlmann, The “transition probability” in the state space of a $\ast$ -algebra, *Rep. Math. Phys.* **9**, 273 (1976).
 - [18] I. M. Gelfand and M. L. Tsetlin, Finite-dimensional representations of the group of unimodular matrices, *Dokl. Akad. Nauk SSSR* **71**, 825 (1950).
 - [19] A. Alex, M. Kalus, A. Huckleberry, and J. von Delft, A numerical algorithm for the explicit calculation of $\text{SU}(n)$ and $\text{SL}(n, \mathbb{C})$ Clebsch–Gordan coefficients, *J. Math. Phys.* **52**, 023507 (2011).
 - [20] M. Keyl and R. F. Werner, Estimating the spectrum of a density operator, *Phys. Rev. A* **64**, 052311 (2001).
 - [21] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. (Wiley, Hoboken, 2006).
 - [22] M. Frank and P. Wolfe, An algorithm for quadratic programming, *Naval Res. Logist. Quart.* **3**, 95 (1956).
 - [23] A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton University Press, Princeton, 1957).
 - [24] D. Blackwell, Equivalent comparisons of experiments, *Ann. Math. Statist.* **24**, 265 (1953).
 - [25] V. Strassen, The existence of probability measures with given marginals, *Ann. Math. Statist.* **36**, 423 (1965).
 - [26] E. Ruch, R. Schraner, and T. H. Seligman, The mixing distance, *J. Chem. Phys.* **69**, 386 (1978).
 - [27] A. W. Marshall, I. Olkin, and B. C. Arnold, *Inequalities: Theory of Majorization and Its Applications*, 2nd ed. (Springer, New York, 2011).
 - [28] M. Guță and K. Matsumoto, Optimal cloning of mixed Gaussian states, *Phys. Rev. A* **74**, 032305 (2006).
 - [29] V. Brzić, D. Grinko, M. Studziński, and M. T. Quintino, Optimal physical approximations of pure-state cloning and transposition are complementary channels, *arXiv:2603.23628* (2026).
 - [30] H. Wakaura, *Finite-copy optimal cloning of mixed qudit states: manuscript and reproducibility package finite_n_cloning_repro*, Zenodo, version 1.0.0 (2026), <https://doi.org/10.5281/zenodo.23588754>.

- [31] J. Watrous, *The Theory of Quantum Information* (Cambridge University Press, 2018).
- [32] R. Bhatia, *Matrix Analysis*, Graduate Texts in Mathematics, Vol. 169 (Springer, New York, 1997).
- [33] W. Fulton and J. Harris, *Representation Theory: A First Course*, Graduate Texts in Mathematics, Vol. 129 (Springer, New York, 1991).
- [34] I. G. Macdonald, *Symmetric Functions and Hall Polynomials*, 2nd ed. (Oxford University Press, Oxford, 1995).
- [35] J. E. Humphreys, *Introduction to Lie Algebras and Representation Theory*, Graduate Texts in Mathematics, Vol. 9 (Springer, New York, 1972).
- [36] A. A. Klyachko, Stable bundles, representation theory and Hermitian operators, *Selecta Math. (N.S.)* **4**, 419 (1998).
- [37] A. Knutson and T. Tao, The honeycomb model of $GL_n(\mathbb{C})$ tensor products I: Proof of the saturation conjecture, *J. Amer. Math. Soc.* **12**, 1055 (1999).
- [38] W. Fulton, Eigenvalues, invariant factors, highest weights, and Schubert calculus, *Bull. Amer. Math. Soc.* **37**, 209 (2000).
- [39] S. Boyd and L. Vandenberghe, *Convex Optimization* (Cambridge University Press, Cambridge, 2004).
- [40] P. J. Goulart and Y. Chen, Clarabel: An interior-point solver for conic programs with quadratic objectives, *arXiv:2405.12762* (2024).