Classical limits and learning gaps for compound bosonic fading channels

Lluis Eriksson

26 October 2026

Abstract

We study classical communication over noisy mixtures of phase-insensitive bosonic Gaussian channels. Both thermal noise and photon budget scale with an energy parameter. An explicit measure–prepare representation converts every quantum block code, including entangled codewords and collective receivers, into a classical code with identical error and an exactly quantified cost increment. The same simulation works for an unknown fixed mixing law. A Gaussian-smoothing estimate and a uniform finite-alphabet reduction give an explicit $O(E^{-1})$ comparison with a classical compound variational value. This value optimizes one input distribution before taking the worst mixing law. We also prove a strict separation between classical optimal modulation and the time-sharing hull of Gaussian coherent modulations for a matched-reset example. Matching bounds determine the high-energy limit of a regularized Holevo functional with Gaussian average marginals. A block-uniform, energy-controlled continuity modulus removes both exact marginal Gaussianity and Gaussianity of the full average state. We prove that almost every product coherent Gaussian random codebook from a fixed finite deterministic schedule has the required approximate marginals, controlling empirical moments separately. Strict cost slack or an explicit common rescaling enforces physical feasibility. This yields an operational converse for reliable sequences in the defined typical class, below an unrestricted non-Gaussian common-code rate at sufficiently large energy. It does not assert a good decoder for every random book or cover every protocol described as Gaussian. The same explicit classical comparison extends to free prior entanglement. For a finite compound family with classical feedback the limiting value instead optimizes after the worst law is fixed. A matched-reset/AWGN pair has a strict learning gap: one feedback bit after a zero-photon pilot eventually outperforms arbitrary prior entanglement without feedback. The Gaussian, coding and continuity tools used in the argument have established antecedents.

1 Model and operational statement

All logarithms are natural, and h_2 denotes binary Shannon entropy.

Write $\phi_v(z) = (\pi v)^{-1} e^{-|z|^2/v}$, so that $\mathbb{E}|G_v|^2 = v$. A thermal state γ_t has mean photon number t. Let $\Phi_{a,t}$ be the phase-insensitive channel sending a coherent input $|z\rangle$ to $D(az)\gamma_t D(az)^*$. Its characteristic function is

$$\chi_{\Phi_{a,t}(\rho)}(\xi) = \chi_\rho(\bar{a}\xi) e^{-[t+(1-|a|^2)/2]|\xi|^2}.$$

Physicality requires $t \geq \max(0, |a|^2 - 1)$.

Fix $K < \infty$, $0 < v_* \leq v^* < \infty$, and the compact parameter space

$$\Theta = \{(a, v) \in \mathbb{C} \times \mathbb{R} : |a|^2 \leq K, \quad v_* \leq v \leq v^*\}.$$

Let Γ be a nonempty weakly closed set of probability laws on Θ. For $\nu \in \Gamma$ define

$$N_{E,\nu} = \int_\Theta \Phi_{a,vE}\, d\nu(a, v), \qquad W_\nu(y|x) = \int_\Theta \phi_v(y - ax)\, d\nu(a, v).$$

The unknown law ν is fixed throughout each code; conditional on it, branches are sampled independently in different uses. Neither party receives the branch label or the unknown law. This is a compound memoryless model, not an arbitrarily varying channel.

A quantum code with M equally likely messages has m-mode states ρ_j, a single decoding POVM $\{D_j\}$, and expected photon cost

$$M^{-1}\sum_j \operatorname{Tr}\rho_j \sum_{i=1}^{m} n_i \le mE.$$

The average message error must tend to zero uniformly over $\nu \in \Gamma$. Write $C_{\mathrm{opt}}(E,\Gamma)$ for its supremum of rates. It is an unassisted *classical* capacity of a quantum channel, not a capacity for transmitting qubits.

For a law μ on $\mathbb{C}$, write $I_\mu(W_\nu) = I(X;Y)$, and set

$$V_\Gamma(P) = \sup_{\mu:\int |x|^2 d\mu \le P} \inf_{\nu\in\Gamma} I_\mu(W_\nu).$$

It is essential that the input optimization precede the worst-law operation. This is generally different from the worst separately optimized capacity.

Theorem 1 (Uniform classical comparison)**.** *For $E > 0$, $E \ge 2K/v_*$, the preceding model obeys*

$$V_\Gamma(1) - \frac{2}{v_* E} \;\le\; C_{\mathrm{opt}}(E,\Gamma) \;\le V_\Gamma(1) + \frac{\log(1+K/v_*) + 4K/v_*}{E}. \tag{1}$$

For $K = 0$ both operational capacities are zero. The constants are independent of Γ; K, v_*, v^* stay fixed.

Coding length tends to infinity at fixed E, before energy tends to infinity. No uniform blocklength, efficient optimizer, fixed-error strong converse or positivity for every family is asserted. In particular, a family containing a completely replacing channel has value zero. If $\Gamma = \Gamma_E$ changes between energy levels, (1) compares to $V_{\Gamma_E}(1)$; it does not assert that this sequence converges.

2 An exact simulation preserving the common decoder

Put $k = |a|^2$. For $t \ge k$, the heterodyne measure–prepare representation is

$$\Phi_{a,t}(\rho) = \int_{\mathbb{C}} Q_\rho(\alpha) D(a\alpha)\gamma_{t-k} D(a\alpha)^*\, d^2\alpha, \qquad Q_\rho(\alpha) = \pi^{-1}\langle\alpha|\rho|\alpha\rangle. \tag{2}$$

This is a specialization of the established Gaussian entanglement-breaking representation [1, 2]. Indeed, the characteristic function of the heterodyne outcome is $\chi_\rho(\eta)e^{-|\eta|^2/2}$; multiplying it by the preparation factor $e^{-(t-k+1/2)|\xi|^2}$, at $\eta = \bar{a}\xi$, gives the characteristic function in Section 1. Thus the noise subtraction in (2) is k, and

$$\int |\alpha|^2 Q_\rho(\alpha)\, d^2\alpha = \operatorname{Tr}\rho n + 1. \tag{3}$$

The identity extends to correlated m-mode Husimi outcomes, with total cost $\operatorname{Tr}\rho\sum_i n_i + m$.

For $t > k$, refine the prepared thermal state by its positive coherent-state representation. For every quantum code this gives the classical Markov chain

$$J \longrightarrow A^m \longrightarrow B^m \longrightarrow \widehat{J}, \quad P(\widehat{J} = j | B^m = \beta^m) = \langle\beta^m|D_j|\beta^m\rangle. \tag{4}$$

Given A^m, the middle transition is a product of mixtures with variance $vE - k$. Both the stochastic Husimi encoder and the final Born decision kernel are independent of ν. Equation (4) reproduces the error of the quantum code for *every* law, with its correlations intact. The hidden B^m is a converse simulation variable, not an observation supplied to the physical receiver.

Normalize $X = A/\sqrt{E}$, $Y = B/\sqrt{E}$. Define

$$W_{E,-,\nu}(y|x) = \int \phi_{v-k/E}(y - ax)\, d\nu, \qquad W_{E,+,\nu}(y|x) = \int \phi_{v+1/E}(y - ax)\, d\nu.$$

The lower channel is realized by coherent input $|\sqrt{E}x\rangle$ and heterodyne reception. The upper simulation has mean cost at most $1 + 1/E$, by (3). Consequently the operational classical compound capacities satisfy the exact sandwich

$$C_{\mathrm{cmp}}(W_{E,+}, \Gamma, 1) \le C_{\mathrm{opt}}(E, \Gamma) \le C_{\mathrm{cmp}}(W_{E,-}, \Gamma, 1 + 1/E). \tag{5}$$

No additivity or Gaussian optimizer of the quantum mixture is used. For continuous ν, the state integrals are trace-norm Bochner integrals; probabilities and nonnegative moments also permit Tonelli.

3 Uniform information estimates

Lemma 2 (Entropy continuity)**.** *Suppose planar densities $f_n, f \le L < \infty$ converge in L^1 and their second moments are at most $M < \infty$. Then $h(f_n) \to h(f)$.*

Proof. On a fixed ball use the uniform continuity of $-u \log u$ on $[0, L]$. Outside radius R, let δ be the mass. The restricted, normalized density has second moment at most M/δ. Gaussian maximum entropy therefore gives

$$-\delta \log L \le -\int_{|y|>R} f \log f \le \delta \log(\pi e M/\delta^2).$$

Here $\delta \le M/R^2$; both bounds tend uniformly to zero as $R \to \infty$, with value zero when $\delta = 0$. This also proves finiteness of the entropies in the assertion. □

Lemma 3 (Noise sensitivity)**.** *Fix any input law of finite second moment and any mixture of bounded gains. If component variances follow $v_\theta = v_0 + \theta d \in [\ell, U]$, $\ell > 0$, with $|d| \le D$, then*

$$|I_\mu(W_{\theta_1}) - I_\mu(W_{\theta_0})| \le \frac{2D}{\ell} |\theta_1 - \theta_0|. \tag{6}$$

The constant is independent of the input peak, its cost and the mixing law.

Proof. First take a compactly supported input. Let $f_{a,v} = (a_\# \mu) * \phi_v$, $f = \int f_{a,v} d\nu$, and $J = \int d \nabla f_{a,v} d\nu$. Complex variance v means real-coordinate variance $v/2$, hence $\partial_v \phi_v = \Delta \phi_v / 4$. Differentiation and integration by parts yield

$$\partial_\theta h(f) = \tfrac{1}{4} \int J \cdot \nabla \log f.$$

The Gaussian score conditional-expectation identity and Jensen give $F(f_{a,v}) \le 4/v \le 4/\ell$, where $F(f) = \int |\nabla f|^2 / f$ is planar Fisher information. Weighted Cauchy–Schwarz gives

$$F(f) \le 4/\ell, \qquad \int |J|^2 / f \le 4D^2/\ell.$$

Thus $|\partial_\theta h(f)| \le D/\ell$. To justify the differentiations, bounded centers and variances give common Gaussian upper tails, a Gaussian lower density bound, and polynomial-times-Gaussian bounds for derivatives; boundary terms vanish. These bounds hold uniformly over atoms or continuous mixtures.

For a general input, replace its tail outside radius R by zero. The coupled outputs converge in L^1, are bounded by $(\pi\ell)^{-1}$, and have second moment at most $K\mathbb{E}|X|^2 + U$. Lemma 2 passes the integrated entropy inequality to the limit. This does not require differentiability for the untruncated heavy-tailed input. For a fixed letter x, the same entropy bound is independent of x. Conditional entropies are integrable, bounded below by a constant and above by $\log[\pi e(K|x|^2+U)]$. Integrate the conditional bound and subtract it from the output bound to obtain (6). □

Revealing the independent branch only for an information upper bound gives

$$0 \le I_\mu(W_\nu) \le \int \log(1 + |a|^2 P/v)\, d\nu \le \log(1 + KP/v_*). \tag{7}$$

No branch is revealed to the actual decoder. The function V_Γ is finite, nondecreasing and concave, and $V_\Gamma(0) = 0$. Indeed input concavity holds simultaneously for every law and survives the infimum. In particular

$$V_\Gamma(1 + u) \le (1 + u)V_\Gamma(1). \tag{8}$$

4 One code for a continuous compound family

We supply the reduction to the established finite-alphabet compound coding mechanism [3, 4]; a finite-output theorem alone does not justify continuous observations or physical costs.

Proposition 4. *For the Gaussian-smoothed family above and $P > 0$,*

$$C_{\mathrm{cmp}}(W, \Gamma, P) = V_\Gamma(P). \tag{9}$$

The same holds for either family in (5) at fixed admissible E.

Proof. For the converse, Fano and data processing applied to a common stochastic encoder give, for each ν,

$$(1 - \epsilon) \log M - h_2(\epsilon) \le I(J; Y_\nu^m) \le \sum_i I_{\mu_i}(W_\nu) \le m I_{\bar\mu}(W_\nu), \qquad \bar\mu = m^{-1} \sum_i \mu_i.$$

The law $\bar\mu$ has cost at most P and is independent of ν. Taking the infimum over ν, then optimizing feasible $\bar\mu$, gives the required order. The block inequality can be proved without subtracting infinite differential entropies: the KL chain rule with product output marginals gives

$$\sum_i I(X_i; Y_i) = I(X^m; Y^m) + D(P_{Y^m} \| \textstyle\prod_i P_{Y_i}).$$

All terms are finite in our bounded-noise regime by (7).

For achievability choose a feasible input near $V_\Gamma(P)$. Mixing it with a small mass at zero makes its cost strictly below P; concavity loses arbitrarily little information uniformly in ν by (7). Truncate its tail to zero, then quantize a bounded input ball. The Gaussian kernel has the global input-gradient bound $\sqrt{2K}/(\pi\sqrt{e}\, v_*^{3/2})$. Thus their density difference is at most this constant times $\mathbb{E}|X_l - X|$. Under this coupling the output densities converge uniformly on every output ball, uniformly in ν, and have common density and second-moment bounds.

The proof of Lemma 2, split into that ball and its complement, gives uniform output entropy convergence. The conditional entropy is jointly continuous on compact input and parameter domains and is bounded by a constant plus $\log(1+|x|^2)$. Its tail integral under the fixed input vanishes. Thus one finite input alphabet with strict cost slack approximates all the information quantities simultaneously.

For this finite input law choose nested finite output partitions: expanding dyadic boxes, their refined cells, and one exterior cell. They generate the Borel sigma algebra. Their mutual informations increase to the continuous-output information by posterior martingale convergence; the posterior entropy is bounded by the finite input entropy. Each partition information is continuous in ν: Gaussian kernels depend continuously on (a, v) in L^1, so every Borel cell probability is continuous, including the unbounded exterior cell. The limit is also continuous: bounded means and compact positive variances give common Gaussian tails and local density convergence. Dini's theorem on compact Γ now supplies one finite partition with uniformly small loss. This partition is fixed before code length.

For completeness, let A, B be these finite alphabets, and let $R < \inf_W I(p, W)$. Choose constant-composition types $p_n \to p$ on its positive support, retaining strict cost slack. Draw $M = \lceil e^{nR} \rceil$ codewords uniformly and independently from their type class. Decode a unique codeword whose empirical mutual information with the output exceeds $R+\eta$, with $\eta > 0$ below the rate slack. This decoder is independent of W. Hoeffding bounds for output counts within each input-symbol class, followed by finite-alphabet entropy continuity, make the transmitted word fail the threshold with probability e^{-cn}, uniformly in W. For any fixed output word a random competing codeword crosses the threshold with probability at most $\mathrm{poly}(n)e^{-n(R+\eta)}$: a conditional type class has size at most $e^{nH(X|Y)}$, whereas the input type class has size at least $(n+1)^{-|A|}e^{nH(p_n)}$. Sum over the polynomial number of joint types and over competitors.

To choose one deterministic codebook, cover the finite stochastic-matrix family in per-row total variation by a net of radius n^{-2}, of cardinality at most $(Cn^2)^{|A||B|}$. Choose its points from the family; all retain the rate slack. The expected *sum* of average errors on this net tends to zero, so a deterministic codebook has uniformly vanishing net error. Product-channel total variation changes every decoder error by at most nn^{-2}. Therefore its error vanishes over the whole family. There is no exchange of an expected supremum with a supremum of expectations.

Each word has exact additive cost $n\sum_x p_n(x)|x|^2 < nP$. Mapping x to $|\sqrt{E}x\rangle$ enforces physical expected photon cost $E\sum_i |x_i|^2$ for every message. Fix the slack and finite approximations first, let $n \to \infty$, then remove those approximations. Equivalently, for each rate strictly below the variational value, choose all finite approximations once before taking the coding limit. If that value is zero, the converse already settles the zero-rate case. The result asserts no uniform coding length or efficient search. □

Proof of Theorem 1. For the plus family use (6) with $D = 1, \ell = v_*$, over an interval of length $1/E$. For the minus family use $D = K, \ell = v_*/2$, valid when $E \geq 2K/v_*$. The bounds survive infima and suprema over the same laws and input constraints. Consequently

$$V_{E,+,\Gamma}(1) \geq V_\Gamma(1) - 2/(v_*E), \qquad V_{E,-,\Gamma}(P) \leq V_\Gamma(P) + 4K/(v_*E).$$

Apply (9) in (5), then (8) and (7). This gives (1). The optimized coherent-modulation heterodyne rate is within $[\log(1+K/v_*)+(4K+2)/v_*]/E$ of the unrestricted quantum rate. □

5 A matched-reset application: a strict Gaussian-hull gap

Fix $k, r > 0$, $0 < \lambda < 1$, $q = 1 - \lambda$, $s = k + r$, and

$$W_\theta(y|x) = \lambda\phi_r(y - \sqrt{k}e^{i\theta}x) + q\phi_s(y).$$

For $\mu_0 = \mathcal{CN}(0,1)$, the average output is exactly ϕ_s. Let F denote the signal/reset flag and $H(x) = H(F|Y, X = x)$. It is radial, bounded, positive at finite x, and tends to zero at infinity. The last assertion follows by separating the two widely separated Gaussian clouds and applying binary Fano. It follows that H is not affine in $|x|^2$.

There exists a compactly supported bounded radial h, normalized by $\|h\|_\infty \le 1$, with

$$\int h\, d\mu_0 = \int |x|^2 h\, d\mu_0 = 0, \qquad a_0 = \int hH\, d\mu_0 > 0. \tag{10}$$

Indeed, if all such moment-annihilating tests annihilated H, two fixed radial moment correctors would imply that H equals an affine function of $|x|^2$ on every bounded radial domain, hence globally almost everywhere. Boundedness and decay would make it zero, contradicting positivity. Reverse the sign and normalize a nonzero test. The moment-corrector matrix is invertible by taking two disjoint annuli with distinct conditional squared radii.

For $\mu_\epsilon = (1 + \epsilon h)\mu_0$, (10) preserves mass and cost. Since the output second moment stays s, Gaussian cross entropy and the binary flag identity give

$$I_{\mu_\epsilon}(W_0) - I_{\mu_0}(W_0) = \epsilon a_0 - D(p_{Y,\epsilon}\|\phi_s) \ge \epsilon a_0 - \epsilon^2 \int h^2\, d\mu_0. \tag{11}$$

The last inequality is KL data processing and $\log(1 + u) \le u$. A fixed sufficiently small $\epsilon > 0$ gives a positive, non-numerical improvement.

Let G_ϑ range over all real two-dimensional Gaussian laws on $\mathbb{C}$, including displaced, anisotropic and singular laws. Write $\vartheta = (m, V)$, $p(\vartheta) = |m|^2 + \operatorname{Tr} V$, $J(\vartheta) = I_{G_\vartheta}(W_0)$, and define

$$H_G = \sup_{\eta : \int p\, d\eta \le 1} \int J\, d\eta. \tag{12}$$

Here the time-sharing setting is known to both parties and does not itself carry a message.

Proposition 5 (Strict Gaussian scheduling gap)**.** *For this matched-reset channel, $H_G < C(1)$, where $C(P) = \sup_{\mu:\mathbb{E}|X|^2 \le P} I_\mu(W_0)$.*

Proof. The symmetry, injectivity and strict-growth strategy has nearby antecedents in fading problems [8, 7]; we check it directly for the present channel and Gaussian scheduling class. The output-law map is injective: subtract $q\phi_s$ and Fourier transform the remaining convolution, whose Gaussian factor is nowhere zero. Thus output entropy, and hence mutual information, is strictly concave in the input law.

If a Gaussian law attained $C(1)$, rotation covariance and strict concavity would force it to be circular $\mathcal{CN}(0, P)$. The finite concave nondecreasing capacity $C(P)$ is unbounded, since revealing the independent flag costs at most $h_2(\lambda)$ and gives

$$C(P) \ge \lambda \log(1 + kP/r) - h_2(\lambda).$$

Such a function is strictly increasing: a flat interval would, by concavity and monotonicity, force all subsequent slopes to vanish. Therefore $P = 1$, contradicted by (11). No Gaussian law attains $C(1)$.

The parameter sublevel sets of p are compact. Laws η with $\int p\, d\eta \le 1$ are tight and weakly closed, hence compact. Gaussian smoothing and Lemma 2 show that J is continuous, including at singular covariance matrices. Furthermore

$$0 \le J(\vartheta) \le \lambda \log(1 + kp(\vartheta)/r).$$

The upper bound is sublinear in p, so its tails over $p > R$ are uniformly integrable under the mean-cost constraint. Compactly supported continuous truncation then makes $\eta \mapsto \int J\, d\eta$ continuous. The supremum (12) is attained.

For an attaining η_*, the barycenter $\bar{\mu} = \int G_\vartheta \, d\eta_*$ has cost at most one. To justify averaging entropies, every output density is bounded below by $q\phi_s$ and above by a common constant. Thus $|\log p_{Y,\bar{\mu}}(y)| \le C + C'|y|^2$, and the average output second moment is finite. Individual output entropies have integrable bounds by constants and by $C+\log(1+p(\vartheta))$; the conditional entropy of $W(.|x)$ is bounded through its binary flag identity. Fubini therefore gives the exact identity

$$I_{\bar{\mu}}(W_0) - H_G = \int D(p_{Y,\vartheta} \| p_{Y,\bar{\mu}}) \, d\eta_* \ge 0.$$

If $H_G = C(1)$, equality forces almost all output laws to coincide. Injectivity makes almost all Gaussian input laws identical to $\bar{\mu}$, which would be a Gaussian maximizer. This contradiction proves the strict gap. Its size and its optimizer are not computed. □

For any correlated multivariate Gaussian modulation on product coherent states with total cost mE, positive thermal preparation gives exactly

$$\chi(X^m : N_{E,\theta}^{\otimes m}) \le I(X^m : Y^m) \le \sum_i I_{\mu_i}(W_\theta) \le m H_G. \tag{13}$$

The first inequality uses the preparation chain $X^m \to Y^m \to |\sqrt{E} Y^m\rangle$, valid for coherent letters without the input heterodyne cost increment. Here $N_{E,\theta} = \lambda \Phi_{\sqrt{k} e^{i\theta}, rE} + q\gamma_{sE}$ Tr; we may take $E \ge 2k/r$, which also ensures physicality and the hypotheses of Theorem 1. The last uses Gaussian marginals and their average cost, allowing individual costs above one. Known, non-message-bearing time sharing of such blocks is also covered. This does *not* bound squeezed or entangled Gaussian quantum signal states, message-bearing covariance selection, or every protocol described as Gaussian.

Choose a finite input law of strict cost below one whose information exceeds H_G by a fixed amount, using Proposition 4's approximation argument. Heterodyne reception loses at most $2/(rE)$ by (6), so it eventually beats the entire benchmark (13). For an unknown fixed phase, $W_\theta(y|x) = W_0(e^{-i\theta} y|x)$ implies $I_\mu(W_\theta) = I_\mu(W_0)$ for every input law. The common code, rather than this equality alone, is supplied by (9). No phase oracle is assumed.

5.1 Ensembles with Gaussian average quantum states

Consider any ensemble $\{p_j, \rho_j\}$ on m modes with average state $\bar{\rho} = \sum_j p_j \rho_j$ Gaussian and $\operatorname{Tr} \bar{\rho} \sum_i n_i \le mE$. The average state may be displaced, squeezed or entangled across uses; the conditional states need not be Gaussian. Continuous ensembles use the relative-entropy definition of Holevo information. Then

$$\frac{1}{m} \chi\big(\{p_j, N_{E,\theta}^{\otimes m}(\rho_j)\}\big) \le H_G + \frac{\lambda \log(1 + k/r) + 4k/r}{E}. \tag{14}$$

The simulation in (4) implies

$$\chi(J : \text{output}) \le I(J; B^m) \le I(A^m; B^m) \le \sum_i I_{\mu_i}(W_{E,-,\theta}).$$

For continuous labels, the first inequality is $D(\sigma_j \| \bar{\sigma}) \le D(P_{B|j} \| P_B)$, integrated in j, without subtracting an entropy of the label. To justify it, approximate the density ratios by conditional expectations on increasing finite partitions of B, preparing the average coherent state in each cell. Finite-partition data processing, L^1 convergence of the ratios, trace-norm convergence of the prepared states, and lower semicontinuity of relative entropy pass to the limit. Tonelli averages the inequalities. The remaining steps are classical data processing and the block KL chain rule. Linearity of the Husimi function makes its average law $Q_{\bar{\rho}}$, which is multivariate

Gaussian for any Gaussian quantum state: heterodyne is a Gaussian measurement [9]. Its normalized marginals μ_i are real planar Gaussian laws, with average cost at most $1 + 1/E$. Correlations in A^m are allowed in the displayed block inequality. The noise-sensitivity bound, uniform in each marginal cost, gives

$$I_{\mu_i}(W_{E,-,\theta}) \le I_{\mu_i}(W_0) + \frac{4k}{rE}.$$

Writing $H_G(P)$ for (12) at cost P, the sum is therefore at most $mH_G(1+1/E)+4mk/(rE)$. The hull is finite, nondecreasing and concave in cost, with $H_G(0) = 0$; mix its parameter probability laws to see concavity. Thus $H_G(1+1/E) \le H_G+H_G/E$. Revealing the independent signal flag and applying Jensen in cost also yields $H_G \le \lambda \log(1+k/r)$, proving (14). All output entropies are finite under the photon constraint and the uniform preparation-noise floor. Known time sharing among such blocks is covered by summing their marginal costs before applying $H_G(P)$.

By Proposition 5, a fixed finite strict-cost law can have $I_\mu(W_0) \ge H_G + \delta$ for some $\delta > 0$. Its coherent-modulation heterodyne rate exceeds the supremum of (14), over all block lengths, by at least

$$\delta - \frac{\lambda \log(1 + k/r) + (4k + 2)/r}{E}, \tag{15}$$

which is positive for sufficiently large E. The common unknown-phase decoder is again supplied by (9).

Equation (14) uses the global Gaussian-average condition. Its distinction from a Gaussian ensemble is standard; see [12, Remark 2], where a thermal average of Fock states gives an example. That work concerns Gaussian measurement channels, not the present non-Gaussian signal/reset mixture. Finite codebooks generally do not have an exactly Gaussian average. Section 6 gives a quantitative bound under approximately Gaussian average marginals and proves that a precisely defined typical Gaussian random-codebook class satisfies it. Message-dependent covariance choices and arbitrary deterministic books are not automatically covered.

5.2 A matching limit for the ensemble benchmark

Define the regularized Gaussian-average Holevo functional

$$R_G(E) = \sup_{m\ge1} \frac{1}{m} \sup_{\substack{\{p_j,\rho_j\}:\ \bar\rho \text{ Gaussian} \\ \operatorname{Tr}\bar\rho \sum_i n_i \le mE}} \chi\big(\{p_j, N_{E,0}^{\otimes m}(\rho_j)\}\big). \tag{16}$$

Corollary 6 (Gaussian-average Holevo limit)**.** *For $E \ge 2k/r$, the functional (16) satisfies*

$$H_G - \frac{2}{rE} \le R_G(E) \le H_G + \frac{\lambda \log(1 + k/r) + 4k/r}{E}. \tag{17}$$

In particular $R_G(E) \to H_G$.

Proof. The upper bound is (14). For the lower, take a parameter probability law nearly optimal in (12). Mixing it with a small mass at the zero-cost point Gaussian creates strict cost slack with arbitrarily small loss of its average J. Replace its high-cost tail by that point. The information loss tends to zero since $J(\vartheta) \le \lambda \log(1 + kp(\vartheta)/r) = o(p(\vartheta))$. On a compact parameter sublevel, continuity of p, J permits finite quantization; rational approximation of the resulting weights retains the cost slack. Thus rational finite time-sharing laws approximate H_G from below with strict mean cost below one.

For weights with denominator m, use m independent coherent Gaussian modulations in their respective parameter settings. Their average quantum input is a product Gaussian state, of photon cost $E \sum_i p(\vartheta_i) < mE$. Heterodyne information adds over these independent modes,

and Holevo information is at least the information of that physical measurement. Lemma 3 gives at least $\sum_i J(\vartheta_i) - 2m/(rE)$. Divide by m and remove the finite approximations at fixed E. This proves the lower bound in (17) without claiming a finite-message code with exactly Gaussian average. □

The quantum channels satisfy $N_{E,\theta} = \mathcal{U}_\theta \circ N_{E,0}$, where $\mathcal{U}_\theta$ rotates the output and leaves the thermal reset invariant. Consequently (16) is phase independent even for arbitrary conditional input states. For the compact family of unknown fixed phases, the classical compound value is $C(1)$, by the output-rotation identity. Equations (1), (17) and Proposition 5 give the exact limiting separation

$$\lim_{E\to\infty} \left[C_{\mathrm{opt}}(E, \text{all fixed phases}) - R_G(E)\right] = C(1) - H_G > 0. \tag{18}$$

The common decoder in the operational term comes from (9), whereas R_G remains the explicitly defined Holevo functional. Its block supremum is not silently replaced by an operational constrained-code capacity. The positive constant in (18) is not numerically evaluated.

6 Robust marginal bounds and typical Gaussian codebooks

The full-block Gaussian condition in (16) can be weakened substantially. Write $D_{\mathrm{tr}}(\rho, \sigma) = \frac{1}{2}\|\rho - \sigma\|_1$. The following classical estimate makes the cost of an approximation explicit, uniformly in block length. Common-mixture and inflated-energy continuity methods have established antecedents [13, 14]; we give a direct proof with the particular channel cost needed here.

Lemma 7 (Information continuity with an explicit cost)**.** *Let W be a memoryless mixture of complex Gaussian channels with $|a|^2 \le K$ and component variance at least $\ell > 0$. If laws Q, G on $\mathbb{C}^m$ both have total second moment at most mP, and $\mathrm{TV}(Q, G) \le d \le 1$, then*

$$|I_Q(W^{\otimes m}) - I_G(W^{\otimes m})| \le h_2(d) + md \log\left(1 + \frac{KP}{\ell d}\right). \tag{19}$$

The product notation refers to independent branches of the same mixture. The correction at $d = 0$ is defined as zero. An upper noise bound is not required for this lemma. The estimate is uniform in the mixing law.

Proof. For any input with coordinate costs c_i, reveal the independent branch vector Z^m only to obtain an information upper bound. Conditioned on a finite-variance branch vector, Gaussian maximum entropy, subadditivity and Jensen give

$$0 \le I(X^m; Y^m) \le I(X^m; Y^m, Z^m) = I(X^m; Y^m | Z^m) \le \sum_i \log(1 + Kc_i/\ell) \le m \log(1 + KP/\ell).$$

These are relative-entropy chain rules. Conditional Gaussian smoothing has bounded density and finite second moment, so the conditional calculations are finite. After integrating the branch bound, mutual information stays finite even when the mixed output entropy is infinite; no subtraction of those mixed differential entropies is used.

For $0 < d < 1$, a common positive submeasure of Q, G of mass $1 - d$ exists: shrink their maximal common submeasure if their actual total variation is less than d. Normalize it to R, and write $Q = (1 - d)R + dS$, $G = (1 - d)R + dT$. Both exceptional laws have cost at most mP/d. A binary mixture label B, followed by the channel, gives exactly

$$I_Q = (1 - d)I_R + dI_S + I_Q(B; Y^m), \qquad I_G = (1 - d)I_R + dI_T + I_G(B; Y^m).$$

The last terms lie in $[0, h_2(d)]$, and I_S, I_T lie in $[0, m\log(1 + KP/(\ell d))]$. Subtract their common parts to prove (19), with one binary entropy. The endpoints follow from equality or the preceding cost bound. This common-submeasure proof is essential: the right side is not globally monotone in d. A uniform bound on differences also bounds differences of the infima over a channel family; it does not interchange an input supremum and a channel infimum. □

For the matched-reset channel, let $\bar{\rho}_i$ denote the average input marginal of mode i. Finite photon energy ensures finite first moments and symmetrized quadrature covariance. The uncertainty inequality $V_i + i\Omega/2 \geq 0$ defines its unique moment-matched Gaussian state σ_i, with exactly the same energy $e_i = \operatorname{Tr} \bar{\rho}_i n$. This comparison allows squeezing. Set

$$d_i = D_{\mathrm{tr}}(\bar{\rho}_i, \sigma_i), \quad d = m^{-1} \sum_i d_i, \quad P_i = (e_i + 1)/E, \quad c_0 = \lambda \log(1 + k/r) + 4k/r.$$

Theorem 8 (Robust marginal-Gaussian converse)**.** *For $E \geq 2k/r$, any m-mode ensemble of average total photon cost at most mE, with arbitrary conditional states and arbitrary correlations in its average, satisfies*

$$\frac{\chi(J : N_{E,\theta}^{\otimes m})}{m} \leq H_G + \frac{c_0}{E} + h_2(d) + d \log\left(1 + \frac{2k(1 + 1/E)}{rd}\right). \tag{20}$$

The last two terms vanish at $d = 0$ and tend to zero as $d \to 0$, uniformly in m. No bound $e_i \leq E$ is imposed on individual modes. The same conclusion holds for continuous ensembles.

Proof. The simulation and continuous-label data processing used for (14) give

$$\chi \leq \sum_i I_{Q_{\bar{\rho}_i, E}}(W_{E,-,\theta}).$$

Here heterodyne contraction bounds the total variation to $Q_{\sigma_i,E}$ by d_i, and both laws have second moment P_i. Lemma 7, with noise floor $r/2$, therefore bounds each replacement by

$$h_2(d_i) + d_i \log(1 + 2kP_i/(rd_i)).$$

The nonmaximal common part in its proof justifies using d_i, without a false monotonicity assertion.

The comparison Husimi marginal is real planar Gaussian. Noise sensitivity costs at most $4k/(rE)$ per mode, uniformly in its cost. The Gaussian hull combines all those costs before concavity is applied, giving $H_G(1 + 1/E) \leq H_G + \lambda \log(1 + k/r)/E$. Finally $(u, p) \mapsto u \log(1 + (2k/r)p/u)$, extended by zero when $u = 0$, is jointly concave: it is the perspective of $\log(1 + (2k/r)p)$. Jensen for this function and for h_2, and $m^{-1} \sum_i P_i \leq 1 + 1/E$, prove (20). All mutual informations are finite by the cost estimate in Lemma 7; output quantum energy is finite in the stated model. Thus no entropy ambiguity is hidden in the chain rules. □

Define a distinct, larger functional $R_{\mathrm{marg}}(E)$ by replacing the global Gaussianity condition in (16) by Gaussianity of every one-mode average marginal. The joint average need not be Gaussian. Theorem 8 at $d = 0$ and the same product-Gaussian lower construction used for (17) imply

$$H_G - \frac{2}{rE} \leq R_{\mathrm{marg}}(E) \leq H_G + \frac{c_0}{E}. \tag{21}$$

This is a Holevo functional, not an asserted operational capacity. For approximately Gaussian marginals with $d \leq \epsilon_m \to 0$, continuity at zero of the correction in (20) suffices for its uniform vanishing; global monotonicity of that correction is unnecessary.

6.1 A precise typical-codebook class

A discrete empirical distribution of amplitudes has total variation one from a continuous Gaussian. We instead control the empirical quantum states and their moments separately. Fix $E > 0$, a finite list of real planar Gaussian laws $G_l = N_2(u_l, V_l)$, $V_l \geq 0$, with $p_l = |u_l|^2 + \operatorname{Tr} V_l < \infty$, and a deterministic known schedule $l_m(i)$, $1 \leq i \leq m$. This schedule carries no message; the list and E are fixed before $m \to \infty$. Let $M_m \geq e^{Rm}$, $R > 0$. Independently over messages and coordinates, draw $X_{ji} \sim G_{l_m(i)}$ and use equiprobable product coherent words $\bigotimes_i |\sqrt{E}X_{ji}\rangle$. Independence between experiments at different lengths is not required. Set $P_m = m^{-1} \sum_i p_{l_m(i)}$.

Proposition 9 (Typical marginal Gaussianity and physical cost)**.** *For the preceding experiment, almost surely*

$$\max_{i \leq m} D_{\mathrm{tr}}(\widehat{\rho}_i, \widehat{\sigma}_i) \longrightarrow 0,$$

where $\widehat{\rho}_i$ is the empirical average marginal and $\widehat{\sigma}_i$ its own moment-matched Gaussian. If $P_m \leq 1 - \eta$ eventually for a fixed $\eta > 0$, the physical average cost is at most mE almost surely eventually. If only $P_m \leq 1$, the common correction below gives exact feasible cost, preserves the displayed convergence, and changes the average error of any given channel/decoder by $o(1)$, uniformly in the channel. No existence of a reliable decoder is asserted by this proposition.

Proof. Let $\tau_l = \int |\sqrt{E}x\rangle\langle\sqrt{E}x| \, dG_l(x)$ and $N_* = E \max_l p_l$. Projection Π_L onto photon numbers $0, \ldots, L$ has rank $L+1$ and $\operatorname{Tr} \tau_l(I - \Pi_L) \leq N_*/(L+1)$. The nonnormalized truncation estimate $\|\rho - \Pi_L \rho \Pi_L\|_1 \leq 2\sqrt{\operatorname{Tr} \rho(I - \Pi_L)}$ and Jensen control both empirical and design tails in expectation. For projected independent rank-one samples, Hilbert–Schmidt variance is at most $1/M$, so their mean trace fluctuation is at most $\sqrt{(L+1)/M}$. Consequently, with $M = M_m$,

$$\mathbb{E}\|\widehat{\rho}_i - \tau_{l_m(i)}\|_1 \leq 4\sqrt{\frac{N_*}{L+1}} + \sqrt{\frac{L+1}{M}}.$$

Choose $L + 1$ of order $\sqrt{M}$. Markov at threshold $M^{-1/8}$, then a union bound over m modes, gives failure probability at most $C_E m M_m^{-1/8}$, summable in m. Borel–Cantelli proves uniform convergence to the design states.

That alone does not control moment matching. Separately put $\widehat{u}_i = M^{-1} \sum_j X_{ji}$, $\widehat{T}_i = M^{-1} \sum_j X_{ji} X_{ji}^{\mathsf{T}}$, and $\widehat{V}_i = \widehat{T}_i - \widehat{u}_i \widehat{u}_i^{\mathsf{T}}$. The finite Gaussian list has uniformly finite fourth moments, hence

$$\mathbb{E}|\widehat{u}_i - u_l|^2 \leq C/M,$$
$$\mathbb{E}\|\widehat{T}_i - (V_l + u_l u_l^{\mathsf{T}})\|_F^2 \leq C/M.$$

Chebyshev and a union bound at threshold $M_m^{-1/8}$ give failure bounds $C m M_m^{-3/4}$, summable in m. Almost surely, means and raw second moments converge uniformly with error $O(M_m^{-1/8})$. Thus covariances converge uniformly, even at a singular design covariance.

The moment-matched quantum Gaussian is the coherent mixture with displacement law $N_2(\widehat{u}_i, \widehat{V}_i)$. For $[q, p] = i$, its mean is $\sqrt{2E}\widehat{u}_i$ and covariance is $I/2 + 2E\widehat{V}_i$. Coupling two Gaussian laws by the same standard real Gaussian vector and using the coherent overlap gives

$$D_{\mathrm{tr}}(\tau(u, V), \tau(u', V')) \leq \sqrt{E}\big(|u - u'|^2 + \|\sqrt{V} - \sqrt{V'}\|_F^2\big)^{1/2}.$$

The positive matrix square root is continuous on positive semidefinite matrices, including the boundary: uniformly approximate the scalar square root by polynomials on a compact spectral interval. The empirical covariances stay eventually in such an interval. The last display and the triangle inequality prove convergence to the empirical moment-matched states, rather than merely to the design states.

The empirical cost is $\widehat{P}_m = (mM_m)^{-1}\sum_{ij}|X_{ji}|^2 = m^{-1}\sum_i \operatorname{Tr}\widehat{T}_i$. Its difference from P_m is $O(M_m^{-1/8})$ almost surely. Fixed slack therefore proves eventual feasibility. For $P_m \le 1$ without slack put

$$t_m = \min(1, \widehat{P}_m^{-1/2}), \qquad X'_{ji} = t_m X_{ji},$$

with $t_m = 1$ at zero cost. This is a common correction fixed with the codebook, not a message-bearing operation. Its cost is exactly feasible and $1 - t_m = O(M_m^{-1/8})$ almost surely. The rescaled empirical means/covariances and their Gaussian states converge to the same design values. Moreover coherent overlap and Jensen bound the average whole-block trace perturbation by

$$M_m^{-1}\sum_j D_{\mathrm{tr}}(\rho'_j, \rho_j) \le |1 - t_m|\sqrt{mE\widehat{P}_m} = O(\sqrt{mE}M_m^{-1/8}) = o(1).$$

Trace-distance contraction controls the change in average decoding error for any channel and any fixed decoder, uniformly over channels. This transfers reliability if it was already proved; it does not create it. The corrected random book is no longer iid, and is described explicitly as the iid design followed by this common adjustment. □

Corollary 10 (Operational consequence)**.** *At fixed $E \ge 2k/r$, any reliable equiprobable-message sequence with cost mE and average marginal distances $d_m \to 0$ obeys*

$$\limsup_{m\to\infty} \frac{\log M_m}{m} \le H_G + \frac{c_0}{E}. \tag{22}$$

This applies almost surely to the feasible typical books in Proposition 9, if they have vanishing average error. It includes collective decoders and the unknown-fixed-phase problem. For sufficiently large E, an unrestricted non-Gaussian common code from (9) has a rate strictly above the right side of (22).

Proof. For error ϵ_m, Holevo and Fano give

$$(1 - \epsilon_m)\log M_m/m - h_2(\epsilon_m)/m \le \chi_m/m.$$

Use (20) and then take the block limit at fixed E. For uniform error over phases, fix any phase for the converse. The non-Gaussian strict-cost rate used in (15) is at least $H_G + \delta - 2/(rE)$; it exceeds (22) when $E > (c_0 + 2/r)/\delta$, with the earlier admissibility conditions. Its common decoder is provided by (9), not by the random-book theorem. □

6.2 Two necessary cautions

Trace closeness alone does not control the moment-matched Gaussian, even with uniformly bounded energy. Indeed $\rho_L = (1-1/L)|0\rangle\langle 0| + (1/L)|L\rangle\langle L|$ has energy one and trace distance $1/L$ to vacuum, while its Gaussian with the same moments is always γ_1. Thus $D_{\mathrm{tr}}(\rho_L, \gamma_1) \to 1/2$, not zero. This is why the empirical moments above were controlled separately. Also, for a saturated circular design $\mathcal{CN}(0,1)$, the actual cost is the average of mM_m independent unit-mean exponentials, so $\Pr(\widehat{P}_m \le 1) \to 1/2$ by the central limit theorem. Design cost in expectation does not imply typical feasibility without slack or correction.

No theorem here covers every deterministic realization named Gaussian, arbitrary post-selection, arbitrary message weights, message-bearing covariance schedules, or design lists growing with block length. No trace approximation of the whole average to a product Gaussian is claimed or needed. Equation (21) remains a functional statement; (22) is a separate converse for a defined operational code class, not a proof of its exact capacity or random-code achievability.

7 Entanglement assistance and learning feedback

The common-input comparison also admits a resource extension. Free prior entanglement does not remove its limiting minimax restriction; feedback can instead change which input law is selected after learning the channel. The minimax and training mechanisms have established antecedents [15, 16, 17]. We prove the energy-controlled bosonic extension directly rather than importing a finite-dimensional theorem.

An assisted code starts with a message-independent, law-independent shared state of arbitrary finite dimension per block. Its dimension may grow without a rate bound. The receiver may retain bosonic quantum memory and perform joint decoding. Feedback, when allowed, is noiseless classical communication only from receiver to sender; there is no free forward side channel. For an adaptive protocol the expected photon budget, averaged over messages and all protocol randomness, is at most mE separately under every actual ν. Denote these capacities by $C_{\mathrm{EA}}(E,\Gamma)$ and $C_{\mathrm{EA,FB}}(E,\Gamma)$, and the unassisted feedback capacity by $C_{\mathrm{FB}}(E,\Gamma)$. All are vanishing-error classical communication capacities. Put

$$A = \log(1 + K/v_*) + 4K/v_*, \qquad B = 2/v_*, \qquad C_\nu(P) = \sup_{\mu:\mathbb{E}|X|^2\le P} I_\mu(W_\nu).$$

Lemma 11 (A reference-preserving classical bottleneck)**.** *The exact minus-channel representation in Section 2 remains valid with an arbitrary retained reference. Let M, R, X, Y be its joint state, where X, Y are classical and Y is produced by $W_{E,-,\nu}$ independently of MR conditional on X. At finite input cost,*

$$I(M;RY) - I(M;R) = I(M;Y|R) = I(MR;Y) - I(R;Y) \le I(X;Y). \tag{23}$$

These are quantum mutual informations, defined by relative entropy; M is finite. Preparing the actual output from Y preserves the same upper bound for the receiver's information increment.

Proof. Heterodyne the channel input, put $X = \alpha/\sqrt{E}$, apply $W_{E,-,\nu}$, and prepare $|\sqrt{E}Y\rangle$. The operator-valued conditional measure of MR given X is retained. Its nonselective marginal is unchanged, although its conditional states generally are steered. The measure–prepare identity is an equality of normal channels on all inputs, hence a complete identity on references. The positive thermal preparation variance is $vE - |a|^2$; the condition $E \ge 2K/v_*$ gives minus noise at least $v_*/2$.

Applying the preparation channel $x \mapsto \rho^x_{MR}$ to the classical joint law of X, Y and using relative-entropy data processing gives $I(MR;Y) \le I(X;Y)$. Revealing the independent mixture branch for an upper bound yields $I(X;Y) \le \log(1 + 2KP/v_*)$ for $\mathbb{E}|X|^2 \le P < \infty$. Thus $I(MR;Y)$ and $I(R;Y)$ are finite, while message information is at most $H(M)$. The relative-entropy chain rule therefore justifies (23) without subtracting infinite marginal entropies, even for bosonic receiver memory. Finally prepare the quantum output and use data processing while retaining R. This is not an inequality for increments of optimized accessible information. □

Theorem 12 (Uniform assisted comparisons)**.** *For $E > 0$, $E \ge 2K/v_*$, the no-feedback model obeys*

$$V_\Gamma(1) - B/E \le C_{\mathrm{opt}}(E,\Gamma) \le C_{\mathrm{EA}}(E,\Gamma) \le V_\Gamma(1) + A/E. \tag{24}$$

If Γ is finite, then

$$\inf_{\nu\in\Gamma} C_\nu(1) - B/E \le C_{\mathrm{FB}}(E,\Gamma) \le C_{\mathrm{EA,FB}}(E,\Gamma) \le \inf_{\nu\in\Gamma} C_\nu(1) + A/E. \tag{25}$$

The upper feedback bound holds without the finiteness restriction. In particular, prior entanglement can improve the no-feedback rate by at most $(A+B)/E$ in the stated scaling. This is not a claim of zero entanglement advantage at finite energy or at fixed signal energy.

Proof. Without feedback, trace-preserving encoding on the sender's half leaves the message independent of the receiver's initial reference. Joint heterodyne on the encoded block has one law independent of ν, including any preshared randomness. Lemma 11 gives $I(M;B^mR) \le I(X^m;Y^m)$. The exact memoryless identity

$$\sum_i I(X_i;Y_i) = I(X^m;Y^m) + D(P_{Y^m} \| \bigotimes_i P_{Y_i})$$

and input concavity bound this by $mI_{\bar\mu}(W_{E,-,\nu})$, where $\bar\mu = m^{-1}\sum_i \mu_i$ is common to all laws and has cost at most $1 + 1/E$. The extra cost is exactly one heterodyne photon per mode. Fano with the worst-law error implies $C_{\mathrm{EA}} \le V_{E,-,\Gamma}(1 + 1/E)$. The noise sensitivity and cost concavity used in Theorem 1 give (24); its unassisted lower bound is still an allowed assisted code.

For feedback fix the actual ν. Before forward use i, retain all receiver memory and transcript copies in R_i. Sender-local trace-preserving operations, including message- and feedback-controlled operations, do not change its joint marginal with M. Apply the complete representation to this actual next input only. Equation (23) bounds the actual quantum mutual information increment by $I(X_i;Y_i)$. The receiver's local instrument, retaining its outcome, is a channel on its previous memory; copying and sending that classical outcome back cannot increase its information about M. Hence

$$I(M;R_{\mathrm{final}}) \le \sum_i I(X_i;Y_i) \le \sum_i C_{E,-,\nu}(P_i(\nu)) \le mC_{E,-,\nu}(1 + 1/E).$$

Here $P_i(\nu) = \mathbb{E}_\nu |X_i|^2$ may depend on the actual law through feedback. The per-law physical budget gives $\sum_i P_i(\nu) \le m(1+1/E)$; no per-use or per-history bound is assumed. Nonnegative costs make every term finite. Apply Fano and the same uniform noise/cost estimates at this fixed law, and then take its infimum. This proves the upper part of (25).

For the finite-family direct part fix E first. Group identical coherent/heterodyne kernels. Each pair of distinct kernels is separated by some finite-cost input probe and a Borel output bin. Finitely many pairs need only finitely many such experiments. Use positive fractions of t_m independent, message-independent probes, with $t_m \to \infty$, $t_m/m \to 0$. Their bounded empirical bin frequencies identify the finite kernel index with vanishing worst-index error. Feed the estimate back once. For each index choose a known-law coherent/heterodyne payload code, with the same message count and rate strictly below $\min_\nu C_{E,+,\nu}(1-\delta)$, using fixed slack $\delta > 0$. Proposition 4 supplies the required finite-alphabet physical codes. Every selected book obeys that budget, including when the estimate is wrong. Selection is independent of the message. If the finite pilot's mean normalized cost is Q, then $t_mQ + (m - t_m)(1-\delta) \le m$ eventually. Estimation error plus the matching payload error tends uniformly to zero. After the block limit, remove δ by continuity of each finite concave capacity at one, and apply the plus-noise bound. This gives the lower part of (25). No uniform coding length in E, efficient identification, or direct theorem for an infinite feedback family is asserted. □

Proposition 13 (A learning gap that prior entanglement cannot close)**.** *There is a two-law family in the present optical model such that, for all sufficiently large E, unassisted classical feedback achieves a strictly higher classical rate than arbitrary prior entanglement without feedback. One feedback bit after a zero-photon pilot suffices for the achievability assertion.*

Proof. Use the matched-reset kernel of Section 5 at phase zero, with $k, r > 0$, $0 < \lambda < 1$, $q = 1 - \lambda$, $s = k + r$, and let its unit-cost capacity be C_1. It is positive and finite, and (11) shows $G_1 = I_{\mathcal{CN}(0,1)}(W_1) < C_1$. Define

$$b^2 = r(e^{C_1} - 1), \quad W_2(y|x) = \phi_r(y - bx), \qquad \Gamma_2 = \{\lambda\delta_{(\sqrt{k},r)} + q\delta_{(0,s)}, \delta_{(b,r)}\}.$$

The flag upper bound gives $0 < b^2 \le r[(1+k/r)^\lambda - 1] < k$. Thus both laws belong to one rectangle with $K = k, v_* = r, v^* = s$. The AWGN kernel has capacity $\log(1+b^2/r) = C_1$, uniquely attained by $\mathcal{CN}(0,1)$: equality in the output entropy bound requires the centered circular Gaussian output with full available variance, and deconvolution by its nonvanishing Gaussian characteristic function uniquely determines the input.

We justify strictness rather than infer it from noncoincident formal optimizers. Input laws of second moment at most one form a weakly compact set, by tightness and lower semicontinuity of the moment. Under weak convergence their output densities converge pointwise, since each fixed Gaussian kernel is bounded continuous in the input. Scheffe's lemma then gives total-variation convergence. Output densities have one common upper bound and second-moment bound. Lemma 2, including its uniform tail estimate, makes their differential entropies continuous. For W_1, conditional entropy is a fixed weighted Gaussian entropy plus $h_2(\lambda) - H(F|Y, X = x)$, a bounded continuous function of x; for W_2 it is constant. Both mutual informations are therefore continuous on this weakly compact cost set. In particular C_1 and the common-input maximum are attained. Equality of that maximum with C_1 would force its input to maximize W_2, hence to be $\mathcal{CN}(0,1)$, which gives only $G_1 < C_1$ on W_1. Consequently

$$\Delta = C_1 - V_{\Gamma_2}(1) > 0. \tag{26}$$

For a concrete zero-cost pilot send vacuum repeatedly and heterodyne. Put $u_E = r + 1/E$, $w_E = s + 1/E$. The bit $Z = \mathbf{1}\{|Y|^2 \le u_E\}$ has probabilities

$$p_2 = 1 - e^{-1}, \qquad p_1 = \lambda(1-e^{-1}) + q(1-e^{-u_E/w_E}), \qquad d_E = p_2 - p_1 = q(e^{-u_E/w_E} - e^{-1}) > 0.$$

The midpoint test on t empirical bits has each conditional error at most $\exp(-td_E^2/2)$, by the one-sided Hoeffding inequality. Its transmitted photon cost is zero. For fixed E, take $t = \lceil a \log m \rceil$ with any fixed $a > 0$, and send the resulting index bit back. The payload construction in the preceding proof now applies. With $A_0 = \log(1+k/r) + 4k/r$, $B_0 = 2/r$, Theorem 12 yields

$$C_{\mathrm{FB}}(E, \Gamma_2) - C_{\mathrm{EA}}(E, \Gamma_2) \ge \Delta - (A_0 + B_0)/E. \tag{27}$$

It is positive when $E \ge 2k/r$ and $E > (A_0 + B_0)/\Delta$. Neither C_1, Δ, nor this existential threshold is numerically evaluated. This compares different resources under a fixed unknown law, not entanglement and feedback for a known Gaussian channel. □

8 Prior work, limits and verification context

Gaussian measure–prepare structure, the single Gaussian classical limit, data processing, Fano, heat-flow score methods, and type coding are prior tools [1, 2, 3, 4]. Recent general Gaussian-encoding optimality results for single-mode Gaussian channels are given by Adesso, Lami and Giovannetti [11]. Their Gaussian-channel hypothesis excludes the generic mixtures studied here; a mixture of Gaussian components is not in general a Gaussian channel. Their fixed-error energy-rescaling formula is not imported into our compound average-error problem. Uniform quantization using compactness and Dini-type arguments is also explicit in earlier compound-channel work [5, 6]. The Mitran–Devroye–Tarokh operational model assumes that the channel parameter is known to the receiver; its theorem is therefore not a substitute for our decoder without that side information. Our use of its quantization mechanism is separately justified above. Classical on/off fading already motivates non-Gaussian input optimization [7]. Non-Gaussian advantages in entanglement-assisted quantum fading are also prior research [8]. Near-capacity coherent modulation for non-Gaussian attenuator environments and additive classical noise is also established [10]. Those bounds depend on the environment and are independent of input energy at fixed channel. Our joint scaling and unknown-law comparison

must be distinguished from that earlier setting; coherent near-optimality itself is not new. Our example therefore does not discover the broad non-Gaussian phenomenon. Common-mixture continuity methods and exceptional-energy penalties have antecedents in Winter [13] and Shirokov [14]. Winter's energy-constrained bounds use Gibbs hypotheses; Shirokov's ensemble distances do not simply become distances between average marginals. Neither attribution replaces our direct proofs of (19)–(22). The contribution here is the quantitative common simulation, Gaussian-hull separation, and robust marginal converse for the stated joint scaling, including the defined typical books and their physical cost correction. These tools and that consequence are distinguished from a general Gaussian-protocol impossibility theorem or an exact Gaussian-code capacity.

The results do not cover within-block changes of the unknown law, channel memory, unbounded gains, a vanishing normalized noise floor, or arbitrary nonclassical reset states. The feedback direct theorem requires a finite family; the assistance convention is arbitrary finite-dimensional prior entanglement per block, with unbounded growth allowed. No separate theorem for a differently defined infinite-resource assistance model is claimed. The compound minimax change and training mechanisms are prior results [15, 16, 17]. Constrained entanglement-assisted capacity is longstanding theory [18]; its general formula is not substituted for the direct energy-controlled argument. Single Gaussian-channel entanglement-assisted formulas already give related simultaneous high-energy limits [19]. They do not by themselves prove a uniform compound-mixture comparison. Finite-energy entanglement advantages, including noisy optical experiments [20], are consistent with the stated limit. The additional resource result is quantitative and specific; standalone significance and worldwide priority are not established. Expected photon cost is not a deterministic photon-number bound. A directed primary-source search is not an exhaustive literature audit. The proof does not identify the general classical fading optimizer.

AI and responsibility disclosure. The named author directed substantial OpenAI Codex assistance in derivation and writing. Technical audits requested GPT-6 Astra xhigh, with shared filesystem and reused context declared. Runtime identity, effort, isolation and statistical independence are not independently attested. No independent human review, Lean formalization, certified worldwide originality or predicted assessment score is claimed. Assessment reports must be read with their exact artifact hashes, declared participation and limitations. The unscored technical audit716 examined the marginal and typical-book arguments. Unscored audit745 examined the resource extension and matched-reset/AWGN construction, including the explicit vacuum pilot. Its quantum-mutual-information correction, reference-memory and finiteness requirements, per-law cost and limiting-order conditions, and antecedent/priority reservations are incorporated here. Both are authoring assistance with shared files and declared context, not independent votes or manuscript assessments. The exact combined PDF of this revision has not yet received a complete assessment. Published v1 and v2 and their own exact-artifact reports remain separate and unchanged; their scores are not assigned to this revision. No target grade is used.

References

[1] A. S. Holevo, Entanglement-breaking channels in infinite dimensions, arXiv:0802.0235v1 (2008).

[2] A. S. Holevo and V. Giovannetti, Quantum channels and their entropic characteristics, arXiv:1202.6480v1 (2012), especially Sections 3.4–3.5 and 4.3.

[3] I. Csiszár, The method of types, IEEE Transactions on Information Theory 44, 2505–2523 (1998), doi:10.1109/18.720546.

[4] I. Bjelaković and H. Boche, Classical capacities of averaged and compound quantum channels, arXiv:0710.3027v2 (2009), Section II for the classical finite-alphabet formulation.

[5] P. Mitran, N. Devroye and V. Tarokh, On compound channels with side-information at the transmitter, author draft, 20 December 2005, Lemmas 3–5. Published in IEEE Transactions on Information Theory 52, 1745–1755 (2006), doi:10.1109/TIT.2006.871044. The cited draft is not identified as the final published byte version.

[6] R. F. Schaefer and S. Loyka, The secrecy capacity of compound Gaussian MIMO wiretap channels, arXiv:1509.02182v1 (2015), Section III for common quantization; its secrecy model and Gaussian optimizer are not imported here.

[7] Tu, Li and Dhillon, An information-theoretic framework for vision-aided communication over memoryless on-off fading channels, npj Wireless Technology 2, 80 (2026), doi:10.1038/s44459-026-00074-7.

[8] G. Catalano, M. Fanizza, F. A. Mele, G. De Palma and V. Giovannetti, Convex combinations of bosonic pure-loss channels, arXiv:2604.26874v2 (2026), selected discussion of entanglement-assisted fading and non-Gaussian inputs. Version 2, revised 30 September 2026.

[9] C. Weedbrook, S. Pirandola, R. García-Patrón, N. J. Cerf, T. C. Ralph, J. H. Shapiro and S. Lloyd, Gaussian quantum information, arXiv:1110.3234v1 (2011), Sections II.E and II.E.2 in the PDF (II.5 and II.5.2 in numeric HTML) for Gaussian measurement outcomes and heterodyne.

[10] S. Huber and R. König, Coherent state coding approaches the capacity of non-Gaussian bosonic channels, arXiv:1710.06734v2 (2018), Theorem 1. Journal of Physics A 51, 184001, doi:10.1088/1751-8121/aab7ff.

[11] G. Adesso, L. Lami and V. Giovannetti, Classical capacity of single-mode bosonic Gaussian channels, arXiv:2610.22308v1 (15 October 2026), Section IV and Theorem 2 for the Gaussian-channel scope.

[12] A. Holevo, On the classical capacity of general quantum Gaussian measurement, Entropy 23, 377 (2021), doi:10.3390/e23030377, Theorem 1 and Remark 2 for the distinction between Gaussian average and Gaussian ensemble; its Gaussian-measurement optimization is not imported here.

[13] A. Winter, Tight uniform continuity bounds for quantum entropies: conditional entropy, relative entropy distance and energy constraints, arXiv:1507.07775v6 (12 January 2016), Lemma 2 and the energy-constrained discussion, Proposition 14 and Lemma 15.

[14] M. E. Shirokov, Uniform continuity bounds for information characteristics of quantum channels depending on input dimension and on input energy, arXiv:1610.08870v5 (3 November 2018), Section 6.2, Propositions 7–8 for ensemble-distance continuity; those hypotheses are not identified with marginal closeness.

[15] M. Berta, H. Gharibyan and M. Walter, Entanglement-assisted capacities of compound quantum channels, arXiv:1603.02282v1 (7 March 2016), Section II Theorems 1, 3–4 and Section VI for finite-dimensional compound and feedback antecedents.

[16] H. Boche, G. Janßen and S. Kaltenstadler, Entanglement-assisted classical capacities of compound and arbitrarily varying quantum channels, arXiv:1606.09314v3 (25 January 2017), Theorem 3.1 and Section 4.1 Remark 2 for the common-input converse and growing finite assistance dimensions.

[17] B. Shrader and H. Permuter, Feedback capacity of the compound channel, arXiv:0711.0705v1 (5 November 2007), Section VII for classical training and inf–sup.

[18] A. S. Holevo, Entanglement-assisted capacity of constrained quantum channel, arXiv:quant-ph/0211170v2 (4 January 2003), Proposition 4 under its entropy hypotheses.

[19] H. Shi, Z. Zhang and Q. Zhuang, Practical route to entanglement-assisted communication over noisy bosonic channels, arXiv:1909.11112v4 (12 March 2020), Section III equations (2)–(4). Phys. Rev. Applied 13, 034029 (2020).

[20] S. Hao, H. Shi, W. Li, Q. Zhuang and Z. Zhang, Entanglement-assisted communication surpassing the ultimate classical capacity, arXiv:2101.07482v1 (19 January 2021), optical experimental context.