Electronic Colloquium on Computational Complexity, Report No. 140 (2026)

# Binary code rate bounds via classical–quantum channels

Omar Alrabiah* Venkatesan Guruswami†

**Abstract**

We derive the four principal asymptotic rate-distance tradeoffs for binary codes—Plotkin, Elias–Bassalygo, and the two McEliece–Rodemich–Rumsey–Welch (MRRW) bounds—from one theorem, the "pretty good criterion." If the bit error rate under the pretty good measurement (PGM)—the quantum analog of posterior sampling—of a binary-input output-symmetric classical–quantum (cq) channel lies below $\delta$, then every length-$n$ binary code, linear or nonlinear, of relative distance $\delta$ has rate at most the channel's capacity, up to an $O(n^{-1/2})$ correction. Rate–distance bounds thereby reduce to a channel design problem, wherein the task is to minimize channel capacity subject to the posterior bit error rate constraint. Via the pretty good criterion, the binary erasure channel (BEC) yields Plotkin, the binary symmetric channel (BSC) yields Elias–Bassalygo, the pure-state channel (PSC) yields the first MRRW bound, and a masked PSC yields the second MRRW bound exactly.

This framework is then instantiated with new channels to improve upon the MRRW bounds. Specifically, the mixed-qubit channel (MQC), a mixed-state version of PSC, strictly improves the first MRRW bound at every $0 < \delta < \frac{1}{2}$, while the masked mixed-qubit channel (2MQC) strictly improves the second MRRW bound throughout the same interval.

*Department of Electrical Engineering and Computer Sciences, UC Berkeley, Berkeley, CA 94709, USA. Email: oalrabiah@berkeley.edu. Research supported in part by a Saudi Arabian Cultural Mission (SACM) Scholarship, a UC Noyce initiative award, ONR grant N00014-24-1-2491, and V. Guruswami's Simons Investigator Award.

†Department of Electrical Engineering and Computer Sciences, Department of Mathematics, and the Simons Institute for the Theory of Computing, UC Berkeley, Berkeley, CA 94709, USA. Email: venkatg@berkeley.edu. Research supported in part by a Simons Investigator Award, ONR grant N00014-24-1-2491, and NSF grant 2211972.

ISSN 1433-8092

# Contents

# 1 Introduction

The theory of error-correcting codes developed around two error models. In Hamming's model [Ham50], errors are adversarial, and a code of minimum distance $d$ guarantees recovery from every pattern of fewer than $d/2$ substitutions, wherever they occur. In Shannon's model [Sha48], errors are stochastic, produced by a fixed channel that corrupts the transmitted symbols at random, and reliability asks that the block error probability vanish as the blocklength grows. The two models measure a code against different standards—worst-case correction radius versus typical-case reliability—and they are not interchangeable, for rare error patterns of a given weight may be undecodable even when random errors of that weight are usually harmless. Each model carries its own extremal theory—the tradeoff between rate and minimum distance on the Hamming side, capacity on the Shannon side.

On the Shannon side, the extremal theory was settled at its birth. The best information transmission rate across a memoryless channel is dictated by one corresponding number, namely its capacity. At every rate below the capacity, there are codes whose block decoding error vanishes [Sha48], and at every rate above it, the block decoding error of every code tends to one [Wol57, Str64, Ari73].

On the Hamming side, the tradeoff is far from being completely understood. Let $A_2(n, d)$ be the largest cardinality of a (possibly nonlinear) binary length-$n$ code with minimum distance at least $d$, and define the asymptotic rate function

$$R_2(\delta) \coloneqq \limsup_{n\to\infty} \frac{1}{n} \log A_2(n, \lceil \delta n \rceil). \tag{1}$$

From below, the classic Gilbert–Varshamov greedy argument [Gil52, Var57] gives $R_2(\delta) \geqslant 1 - h(\delta)$, where $h$ is the binary entropy function. This remains the best asymptotic lower bound on the rate of binary codes.[1] From above, McEliece, Rodemich, Rumsey, and Welch (MRRW) [MRRW77] proved two upper bounds. The first is $R_2(\delta) \leqslant R_{\mathrm{MRRW}}(\delta) \coloneqq h\left(\frac{1}{2} - \sqrt{\delta(1-\delta)}\right)$, and the second, which is never weaker, was the benchmark for nearly five decades. Since then, no stronger general bound for binary codes has been shown, even for linear codes.

The rates in the Shannon model far outpace their Hamming-model counterparts. For example, for the binary symmetric channel (BSC), which flips each transmitted bit independently with probability $p$, its capacity is $1 - h(p)$. Correcting the same fraction $p$ of substitutions in Hamming's model, wherever they occur, requires a relative minimum distance $2p$, and the achievable rate is then at most $R_2(2p)$. Since $R_2(\delta) = 0$ at $\delta \geqslant \frac{1}{2}$, by the Plotkin bound [Plo60], no positive rate corrects an adversarial error fraction of $\frac{1}{4}$, whereas the capacity remains positive for every $p < \frac{1}{2}$.

List decoding bridges the two models. Introduced by Elias [Eli57] and Wozencraft [Woz58], it keeps the adversarial error model but weakens what the decoder must return, allowing a short list of candidates in place of a single codeword. For every $R < 1 - h(p)$ there exist binary codes of rate $R$ that are list decodable up to radius $p$ with bounded list size [ZP81, Eli91], whereas no rate above $1 - h(p)$ is possible with subexponential list size [Gur04]. The list decoding model thereby provides an avenue to approach Shannon capacity even when the errors are adversarial.

[1] There have, however, been lower-order improvements. In particular, Jiang and Vardy [JV04] sharpened the Gilbert–Varshamov bound by a factor of order $n$, while Gaborit and Zémor [GZ08] obtained an analogous refinement for certain infinite families of rate-$\frac{1}{2}$ binary linear codes. By contrast, for square alphabet sizes $q \geqslant 49$, the Tsfasman–Vlăduţ–Zink algebraic-geometric construction [TVZ82] improves upon the $q$-ary Gilbert–Varshamov exponent.

### 1.1 Main results

This work crosses between the two error models on a different bridge, drawing on a feature built into the Hamming model from the outset—*obliviousness* to the channel. A minimum-distance guarantee is not tied to a specific noise model. In particular, it is not necessarily tailored to the binary symmetric channel (BSC), the binary erasure channel (BEC), or any other binary memoryless symmetric channel, for that matter. Indeed, previous works have shown quantitative aspects of this, such as [TZ00, HSS21, PSW24, KYH26].

Exploiting this connection in reverse, we can leverage the model-agnostic error-resilience offered by minimum distance to turn bounds for Shannon-model channels into Hamming-model bounds. Whenever codes of relative distance $\delta$ remain reliably decodable over a fixed channel, the channel-coding converse caps their rate by its capacity.

Our result below is a formal manifestation of this intuition and serves as the principal workhorse behind our approach. The strength of our approach draws from an expansion of the scope of this connection to include channels with quantum outputs.

Fix a finite-dimensional binary-input output-symmetric classical–quantum (cq) channel with output density matrices $\sigma_0, \sigma_1 \in \mathsf{D}(\mathbb{C}^d)$. Two quantities of interest govern our approach: (i) the *uniform-prior Holevo information* $\chi(\sigma_0, \sigma_1)$, which equals the channel capacity by output-symmetry (Proposition 6), and (ii) the *posterior bit error rate* $p_e(\sigma_0, \sigma_1)$, the error probability of posterior sampling a single bit. For non-commuting outputs, this is defined as the bit error rate of the pretty good measurement (PGM), the quantum extension of posterior sampling. Both objects are defined formally in Sections 2.1 and 2.2. With this backdrop, we now formally introduce our main theorem.

**Theorem 1** (The pretty good criterion)**.** *Fix $\delta \in (0, \frac{1}{2})$ and a binary code $C \subseteq \{0,1\}^n$ with distance $d_{\min}(C) \geqslant \delta n$ and rate $R(C) := n^{-1} \log |C|$. Fix a finite-dimensional binary-input output-symmetric[2] classical–quantum (cq) channel with output states $\sigma_0, \sigma_1 \in \mathsf{D}(\mathbb{C}^d)$. If $p_e(\sigma_0, \sigma_1) < \delta$, then*

$$R(C) \leqslant \chi(\sigma_0, \sigma_1) + O(n^{-1/2}), \tag{2}$$

*where the implicit constant depends only on $\delta$ and the cq channel $\Phi_\sigma$. In particular, maximizing over all binary codes with minimum distance at least $\delta n$ and letting $n \to \infty$ gives*

$$R_2(\delta) \leqslant \chi(\sigma_0, \sigma_1). \tag{3}$$

The criterion is already informative for classical channels, where the outputs $\sigma_0$ and $\sigma_1$ commute. Applied to the BEC, it recovers the Plotkin bound [Plo60], and applied to the BSC, it recovers the Elias–Bassalygo bound [Eli55, Bas65] (Sections 4.1 and 4.2). A correspondence between these two bounds and their respective channels has been previously noted in a different context [Bla77, Dal15].

For the state-of-the-art MRRW bounds, however, it was not previously known whether they can be captured as successful decoding over an appropriate channel. As it now turns out, there are barriers to such a pursuit: the Elias–Bassalygo bound is the strongest that Theorem 1 can yield when restricted to binary-input *classical* channels.[3] Thus, if such a channel were to exist, it ought to be a genuinely quantum classical–quantum (cq) channel with non-commuting output states.

[2] The output-symmetry hypothesis is a mild constraint. Indeed, consider $\widehat{\sigma}_0 := \frac{1}{2}\sigma_0 \oplus \frac{1}{2}\sigma_1$ and $\widehat{\sigma}_1 := \frac{1}{2}\sigma_1 \oplus \frac{1}{2}\sigma_0$. For any $\sigma_0$ and $\sigma_1$, the pair $(\widehat{\sigma}_0, \widehat{\sigma}_1)$ is output-symmetric. Moreover, $p_e(\widehat{\sigma}_0, \widehat{\sigma}_1) = p_e(\sigma_0, \sigma_1)$ and $\chi(\widehat{\sigma}_0, \widehat{\sigma}_1) = \chi(\sigma_0, \sigma_1)$. This lift can double the output dimension, but it leaves the leading asymptotic rate bound unchanged.

[3] *Proof sketch.* For $q_y := \mathbf{Pr}[\mathbf{b} = 1 \mid \mathbf{y} = y]$, one has $p_e = \mathbf{E}[2q_{\mathbf{y}}(1 - q_{\mathbf{y}})]$ and $\chi = 1 - \mathbf{E}[h(q_{\mathbf{y}})]$. Since $h(q) = f(2q(1-q))$ for the increasing concave function $f(t) := h((1 - \sqrt{1-2t})/2)$, Jensen's inequality gives $\chi \geqslant 1 - f(p_e)$, with the equality case being BSC. Letting $p_e \to \delta^-$ yields the Elias–Bassalygo bound.

A driving question in this work, therefore, is determining whether the MRRW bound can be realized as successful decoding over an appropriate cq channel with genuinely quantum outputs. This leads us to the pure-state channel (PSC), whose connection to the first MRRW bound is laid out in Sections 1.4 and 1.5. Concretely, the PSC is the BSC with a coherent output. Writing $|x\rangle$ for the standard basis vector indexed by $x \in \{0,1\}^n$, define the Bernoulli superposition state $|\mathrm{Ber}(p)^n\rangle := |\mathrm{Ber}(p)\rangle^{\otimes n}$ whose amplitudes are the square roots of the BSC noise probabilities, where $|\mathrm{Ber}(p)\rangle := \sqrt{1-p}\,|0\rangle + \sqrt{p}\,|1\rangle$. The blocklength-$n$ PSC then maps each codeword to the corresponding coherent noisy codeword—the superposition over all possible noisy codewords. More formally, the PSC is defined by the map

$$c \longmapsto |\sigma_c\rangle := X^c\,|\mathrm{Ber}(p)^n\rangle = \sum_{e\in\{0,1\}^n} \sqrt{p^{\mathrm{wt}(e)}(1-p)^{n-\mathrm{wt}(e)}}\,|c\oplus e\rangle,$$

where $X^c := X^{c_1}\otimes\cdots\otimes X^{c_n}$ and $X$ is the X-Pauli matrix defined as $X|0\rangle = |1\rangle$ and $X|1\rangle = |0\rangle$. Note that each coordinate of the output state $\sigma_c$ carries one of two nonorthogonal states $X^{c_i}|\mathrm{Ber}(p)\rangle$. Moreover, measuring $\sigma_c$ in the computational basis yields the outcome $c \oplus e$ where $e \sim \mathrm{Ber}(p)^n$. Thus, decoding over $\mathrm{PSC}(p)$ is no harder than decoding over $\mathrm{BSC}(p)$. The PSC additionally retains coherence among the possible error patterns, which the computational-basis measurement discards.

The pretty good measurement (PGM) provides a way to exploit this coherence, and for binary linear codes it does so *optimally*, thanks to previous work on discriminating geometrically uniform quantum states [BKMH97, UTHH99, EF01, EMV04, ZCCL25, RP21, PR25]. In classical terms, for the PSC, posterior sampling and maximum-a-posteriori (MAP) decoding coincide for linear codes. In fact, this optimality persists down at the bit level (Fact 28). In particular, no explicit decoder for one codeword bit can beat the corresponding bit of the PGM.

Now, fix a binary linear code $C \leqslant \mathbb{F}_2^n$ with $d_{\min}(C) \geqslant \delta n$, let $\mathbf{c}$ be a uniformly random codeword from $C$, and let $\hat{\mathbf{c}}$ be the outcome of the PGM applied to the output state $|\sigma_{\mathbf{c}}\rangle$. Since measuring the $i$'th qubit in the computational basis recovers the $i$'th codeword bit $\mathbf{c}_i$ with error $p$, bitwise optimality therefore tells us that $\mathbf{Pr}[\hat{\mathbf{c}}_i \neq \mathbf{c}_i] \leqslant p$ for every $i \in \{1,\ldots,n\}$. Thus,

$$\delta n \cdot \mathbf{Pr}[\hat{\mathbf{c}} \neq \mathbf{c}] \leqslant \mathbf{E}[d(\hat{\mathbf{c}},\mathbf{c})] = \sum_{i=1}^{n} \mathbf{Pr}[\hat{\mathbf{c}}_i \neq \mathbf{c}_i] \leqslant pn,$$

where the first inequality holds because $\mathbf{c}$ and $\hat{\mathbf{c}}$ are distinct codewords on the event $\{\hat{\mathbf{c}} \neq \mathbf{c}\}$. Hence $\mathbf{Pr}[\hat{\mathbf{c}} = \mathbf{c}] \geqslant 1 - p/\delta$, meaning that for every $p < \delta$, a single measurement of the PSC output returns the whole transmitted codeword with constant probability (Corollary 18). This is exactly the operational conclusion that a constant-size list would have supplied, and it is the sense in which the first MRRW bound is the list-decoding threshold of the PSC, as speculated in Section 1.4. Indeed, when coupled with the well-established strong converse for cq channel coding [Win99, ON99, MO17], we immediately conclude that the code's rate cannot exceed the channel capacity. More concretely, we arrive at the rate bound $R(C) \leqslant h(\frac{1}{2} - \sqrt{p(1-p)}) + O(n^{-1/2})$ for every $p < \delta$. Thus, letting $p \to \delta$ gives the first MRRW bound via our framework.

In the preceding argument, it turns out that linearity was really only a convenience rather than a necessity. Indeed, it only entered through bitwise optimality, which is what allowed the block PGM to be compared against a single-qubit measurement. For an arbitrary code, the same bound $\mathbf{Pr}[\hat{\mathbf{c}}_i \neq \mathbf{c}_i] \leqslant p$ follows instead from three general properties of the PGM: (1) closure under coarse-graining (Lemma 9), which is a key feature of posterior sampling in the classical case, (2)

monotonicity of its error under discarding output systems (Lemma 10), and (3) the worst-case optimality of the uniform input prior (Lemma 11). Nor was anything tied to pure-state outputs, as those three properties and the strong converse hold for every binary-input output-symmetric cq channel. The derivation therefore extends seamlessly to general cq channels, which is exactly the pretty good criterion (Theorem 1).

Now, the pretty good criterion (Theorem 1) provides us with the freedom to develop other rate upper bounds, specifically in the choice of the pair of output-symmetric density matrices $\sigma_0, \sigma_1$. As we will later see, one can choose $\sigma_0, \sigma_1$ to recover the second MRRW *exactly*, albeit with four-dimensional states instead of qubit states. But for now, still focusing on channels with qubit outputs, there is the following natural question: is the first MRRW bound the strongest bound attainable by a qubit-output channel? As it so happens, it is not! By considering appropriate mixed qubits $\sigma_0$ and $\sigma_1$ and their corresponding mixed-qubit channel (MQC) (Section 5.1), we obtain a bound $R_{\mathrm{MQC}}(\delta)$ that strictly improves upon the first MRRW bound throughout the entire interval $\delta \in (0, \frac{1}{2})$.

**Theorem 2** (MQC improvement over the first MRRW bound)**.** *Define*

$$R_{\mathrm{MQC}}(\delta) \coloneqq \min_{0 \leqslant r \leqslant \delta} \left[ s\left((1-2r)^2\right) - s\left(2(\delta - r)(1-2r)\right) \right],$$

*where* $s(D) \coloneqq h((1-\sqrt{1-D})/2)$. *For every* $\delta \in (0, \frac{1}{2})$, *we have*

$$R_2(\delta) \leqslant R_{\mathrm{MQC}}(\delta) < R_{\mathrm{MRRW}}(\delta).$$

Unlike the first MRRW bound, the MQC bound is additionally never weaker than the Elias–Bassalygo bound. Indeed, when $r = 0$, it recovers the bound obtained by $\mathrm{BSC}(J(\delta))$ where $J(\delta) \coloneqq (1-\sqrt{1-2\delta})/2$ is the binary Johnson radius [Joh62]. Moreover, when $r = \delta$, it recovers the bound obtained by $\mathrm{PSC}(\delta)$. In fact, by Remark 5.2, the MQC rate bound is never worse than *any* rate bound obtained by any other binary-input output-symmetric cq channel with qubit outputs.

The MQC does not by itself outperform the second MRRW bound in the low-distance regime (where it is better than the first MRRW bound). The second MRRW bound instead enters our framework through a channel family with higher-dimensional outputs. This is achieved by *channel masking* (Section 2.5), an operation that enables one to consider non-uniform prior output states. Briefly, it is defined as follows: given an input bit $b$, the channel draws a hidden bit $m \sim \mathrm{Ber}(\alpha)$ for a fixed parameter $\alpha \in [0, 1]$ and reveals the "masked" bit $b \oplus m$, along with an accompanying quantum state $\tau_m$ as a "quantum hint" for the hidden bit $m$. Decoding $b$ is therefore equivalent to decoding the "mask" $m$ from a generally non-uniform state ensemble. Applying this operation to the PSC gives the masked PSC and recovers the second MRRW bound *exactly* (Corollary 22).

Applying the same masking operation to the MQC gives the masked mixed-qubit channel (2MQC). This family contains the masked PSC and, when the masking bias $\alpha = 1/2$, recovers the MQC (along with an input-independent, and thus inconsequential, classical bit). These inclusions make the 2MQC a natural refinement, but the strict improvement is not automatic; it follows from the optimization and analysis underlying the next theorem.

**Theorem 3** (2MQC improvement over the second MRRW bound)**.** *For every* $\delta \in (0, \frac{1}{2})$, *we have*[4]

$$R_2(\delta) \leqslant R_{\mathrm{2MQC}}(\delta) < R^{(2)}_{\mathrm{MRRW}}(\delta).$$

[4]See Table 2 for numerical estimates of the improvement.

*Here, $R_{\mathrm{2MQC}}(\delta)$ is the variational bound defined in (46), and $R^{(2)}_{\mathrm{MRRW}}(\delta)$ is the second MRRW bound in (31).*

It is worth noting that the four channels corresponding to the previously known rate bounds—Plotkin, Elias–Bassalygo, and the two MRRW bounds—have a common structural feature. Namely, they are all orthogonal mixtures of pure states. Thus, their mixedness can be perfectly distinguished by examining their block label. On the other hand, generic MQC outputs have no common one-dimensional orthogonal decomposition, so their "mixedness" cannot be entirely distinguished by a readable label. In this sense, the MQC introduces genuine quantum mixtures beyond the orthogonal pure-state mixtures underlying the four previous rate bounds.

**Rate bounds for LDPC codes.** So far, the discussion has only focused on arbitrary binary codes with good distance. A natural question is whether this framework is amenable to accounting for further structural properties. Indeed, as was shown in [IS15, SY26], one can exploit the dual covering radius methods developed for the first MRRW bound to show sharper rate-distance tradeoffs for LDPC codes. Given the connection between the PSC and dual covering methods (Proposition 5), it becomes natural to wonder if one can obtain similar state-of-the-art rate bounds for LDPC codes [SY26].

Fix a positive integer $w \geqslant 3$ denoting the locality of the LDPC code. For integers $n, d \geqslant 1$, let $A_{2,w}(n, d)$ be the largest cardinality of a binary linear code $C \leqslant \mathbb{F}_2^n$ with $d_{\min}(C) \geqslant d$ whose dual is generated by parity checks of weight at most $w$, and put

$$R_w(\delta) \coloneqq \limsup_{n\to\infty} \frac{1}{n} \log A_{2,w}(n, \lceil \delta n \rceil).$$

For simplicity, we focus here on $w = 3$, although the setup extends to general $w$. We denote the one-check PSC curve by $U_3^{\mathrm{PSC}}$ and the Shangguan–Yang curve by $B_3^{\mathrm{SY}}$; both are defined formally in (58) and (56), respectively. Adapting the pretty good criterion (Theorem 1) and the bitwise optimality of PGM for binary linear codes (Fact 28) using one sparse parity check yields the following theorem, proven in Section 8.

**Theorem 4** (PSC-locality improvement for LDPC codes)**.** *For every $\delta \in (0, \frac{1}{2})$,*

$$R_3(\delta) \leqslant U_3^{\mathrm{PSC}}(\delta) < R_{\mathrm{MRRW}}(\delta). \tag{4}$$

*Moreover, for every $\delta \in [\frac{1}{6}, \frac{1}{2})$, we have*

$$U_3^{\mathrm{PSC}}(\delta) < B_3^{\mathrm{SY}}(\delta). \tag{5}$$

We close this subsection with two open problems.

**Optimizing the cq channel.** By the pretty good criterion (Theorem 1), the search for sharper upper bounds on $R_2(\delta)$ for $\delta \in (0, \frac{1}{2})$ now reduces to minimizing the Holevo information subject to keeping the posterior bit error rate below $\delta$. For a fixed output dimension $d \geqslant 2$, the closed formulas in Sections 2.3 and 2.4 make this the problem of minimizing, over output-symmetric density matrices $\sigma_0, \sigma_1 \in \mathsf{D}(\mathbb{C}^d)$,

$$\chi(\sigma_0, \sigma_1) = S\left(\frac{\sigma_0 + \sigma_1}{2}\right) - \frac{1}{2}\left[S(\sigma_0) + S(\sigma_1)\right] \tag{6}$$

subject to the constraint that

$$p_{\mathrm{e}}(\sigma_0, \sigma_1) = \mathrm{Tr}\left(\sigma_0(\sigma_0 + \sigma_1)^{-1/2}\sigma_1(\sigma_0 + \sigma_1)^{-1/2}\right) \leqslant \delta, \tag{7}$$

where the inverse is taken over $\mathrm{supp}(\sigma_0 + \sigma_1)$. Note that by Footnote 2, assuming that $\sigma_0$ and $\sigma_1$ are output-symmetric is only a mild restriction to this optimization problem. The MQC and 2MQC show that noncommuting mixed outputs can improve upon the standard choices, but they do not determine the optimal channel in Theorem 1. Pure outputs can attain at best the first MRRW bound, while the exact optimum across all finite dimensions remains open. As a simple sanity check, Fano's inequality [Fan61] and the Holevo bound [Hol73a] show that $\chi(\sigma_0, \sigma_1) \geqslant 1 - h(\delta)$ under the constraint $p_e(\sigma_0, \sigma_1) \leqslant \delta$, so the method cannot cross the Gilbert–Varshamov bound. Nonetheless, the limitations of this optimization, specifically in relation to previous barriers [Sam01, NS05, Sam25], remain to be understood.

**BSC decoding up to the minimum distance.** The most conspicuous unresolved question that we wish to flag is in fact entirely classical. For every fixed $0 < \varepsilon < \delta < \frac{1}{2}$, does maximum-likelihood decoding on $\mathrm{BSC}(\delta - \varepsilon)$ succeed with high probability, uniformly over all binary linear codes $C \leqslant \mathbb{F}_2^n$ with $d_{\min}(C) \geqslant \delta n$? The best general guarantee is decoding up to the Johnson radius $J(\delta) = (1 - \sqrt{1 - 2\delta})/2$ via the Johnson bound [Joh62] and the sharp-threshold theorem of Tillich and Zémor [TZ00][5]. Kougang-Yombi and Hązła [KYH26] make progress on this problem by proving that the Johnson barrier can be surpassed for linear codes over larger alphabets. Note that an affirmative answer to this problem, when combined with the BSC strong converse and the Gilbert–Varshamov bound, would prove that the linear analog of $R_2(\delta)$ precisely equals $1 - h(\delta)$.

### 1.2 Why posterior sampling?

In the classical setting, the go-to decoder is maximum-a-posteriori (MAP) decoding, which returns a codeword of largest posterior probability.[6] In place of MAP, the decoder used in this work is posterior sampling, whose bit error rate is what dictates the pretty good criterion (Theorem 1), *not* MAP. Now, posterior sampling is not optimal by definition, since it merely samples the posterior that MAP maximizes. Nonetheless, it is never far behind: it errs at most twice as often as MAP.[7] In fact, Barnum and Knill [BK02] prove an analogous result for the pretty good measurement (PGM)—the quantum extension of posterior sampling—and the optimal quantum measurement. Since the channel coding strong converse [Wol57, Str64, Ari73, Win99, ON99] only requires a constant chance of successful block decoding to kick in, this loss is immaterial.

The main motive to prefer posterior sampling over MAP is its closure under coarse-graining. In particular, blockwise posterior sampling *simultaneously* samples from the *bitwise* posteriors as well. Indeed, let $\mathbf{c}$ be uniform on a binary code $C \subseteq \{0,1\}^n$. Let $\mathbf{y}$ be its output when passed through a classical binary memoryless symmetric channel. Given $\mathbf{y} = y$, draw $\hat{\mathbf{c}}$ from the posterior law of $\mathbf{c}$. That is, sample $\hat{\mathbf{c}}$ from $C$ with chance proportional to $\mathbf{Pr}[\mathbf{c} = \hat{\mathbf{c}} \mid \mathbf{y} = y]$. The output is a codeword $\hat{\mathbf{c}} \in C$ by construction, and, conditioned on $\mathbf{y} = y$, the transmitted word $\mathbf{c}$ and the decoded word $\hat{\mathbf{c}}$ are independent draws from the same posterior, so sampling commutes with taking coordinates.

[5] This sharp threshold result was more recently generalized to $q$-ary symmetric channels by [PSW24].

[6] MAP decoding is the Bayesian form of maximum-likelihood decoding (MLD): it maximizes the posterior probability $\mathbf{Pr}[\mathbf{c} = c \mid \mathbf{y} = y]$ rather than the likelihood $\mathbf{Pr}[\mathbf{y} = y \mid \mathbf{c} = c]$, and so it incorporates the prior on $\mathbf{c}$. For the uniform codeword prior used throughout this work, the two rules coincide.

[7] *Proof sketch.* Conditional on an output $y$, set $q_x := \mathbf{Pr}[\mathbf{c} = x \mid \mathbf{y} = y]$ for $x \in C$. Posterior sampling succeeds with probability $\sum_{x \in C} q_x^2$, whereas MLD succeeds with probability $\max_{x \in C} q_x$. Averaging the pointwise upper bound $\sum_{x \in C} q_x^2 \leqslant \max_{x \in C} q_x$ and combining $\sum_{x \in C} q_x^2 \geqslant (\max_{x \in C} q_x)^2$ with Jensen's inequality gives $e_{\mathrm{MLD}} \leqslant p_e \leqslant 2e_{\mathrm{MLD}} - e_{\mathrm{MLD}}^2 \leqslant 2e_{\mathrm{MLD}}$.

Indeed, for every coordinate $i \in \{1, \ldots, n\}$, one has

$$\mathbf{Pr}[\hat{\mathbf{c}}_i = b \mid \mathbf{y} = y] = \sum_{c \in C : c_i = b} \mathbf{Pr}[\mathbf{c} = c \mid \mathbf{y} = y] = \mathbf{Pr}[\mathbf{c}_i = b \mid \mathbf{y} = y].$$

Thus, the $i$'th coordinate of the sampled codeword is itself a posterior sample of the transmitted bit. On the other hand, block MAP decoding outputs a codeword, but its coordinates escape single-letter control, since the bits of the most likely codeword need not be the most likely.[8]

While this contrast between MAP and posterior sampling might appear trifling, these fine distinctions are dramatically amplified in the quantum setting. Indeed, let us revisit the pure-state channel (PSC), defined by the map

$$c \longmapsto |\sigma_c\rangle = \sum_{e \in \{0,1\}^n} \sqrt{p^{\mathrm{wt}(e)}(1-p)^{n-\mathrm{wt}(e)}}\, |c \oplus e\rangle.$$

As mentioned previously, one must exploit the coherence in $\sigma_{\mathbf{c}}$ to benefit from the PSC. This coherence is a scarce resource: classically, one may read an output without disturbing it, whereas a quantum output generally affords only one measurement. That measurement should therefore extract as much information about the transmitted codeword $\mathbf{c}$ as possible. The coarse-graining property of the PGM, inherited from classical posterior sampling (Lemma 8), is indispensable for this purpose. MAP decoding has no analogous property in the quantum setting.

As discussed in Section 1.4, successful MAP decoding in the classical setting often relies on combinatorial arguments, most of which falter in the quantum setting. The quantum formulation instead makes the pretty good criterion (Theorem 1)—an unadorned criterion that nonetheless underpins all known rate upper bounds—transparent, while also furnishing a streamlined treatment of the classical case.

### 1.3 How does mixing help?

At first glance, the pretty good criterion (Theorem 1) might seem to favor either fully classical states—leading to the Elias–Bassalygo bound—or pure quantum states—leading to the first MRRW bound. Its recovery of the second MRRW bound suggests that mixed states can do better, as the MQC and 2MQC constructions demonstrate.

To explain conceptually why moving away from pure states can lower the bound in Fig. 2, we first describe how the MQC is realized.

The MQC output states can be realized as passing the PSC output states through the bit-flip channel. Formally, let $\mathcal{B}_\eta$ be the X-Pauli extension of BSC($\eta$), given by $\mathcal{B}_\eta(\varrho) \coloneqq (1-\eta)\varrho + \eta X \varrho X$. For $b \in \{0,1\}$ and $r \in (0, \frac{1}{2})$, let $\sigma_b^{\mathrm{PSC}}(r)$ be the output states of PSC($r$). Then, for a BSC noise level $\eta \in (0, \frac{1}{2})$, the output states of MQC are defined as[9]

$$\sigma_b^{\mathrm{MQC}}(r, \eta) \coloneqq \mathcal{B}_\eta\left(\sigma_b^{\mathrm{PSC}}(r)\right) = (1-\eta)\sigma_b^{\mathrm{PSC}}(r) + \eta \sigma_{b\oplus 1}^{\mathrm{PSC}}(r).$$

[8] For example, suppose a uniform codeword from $C = \{000, 110, 101\}$ is transmitted over BSC(3/7) and that $y = 000$ was received. The posterior probabilities of the three codewords are $8/17, 9/34, 9/34$, so blockwise MAP returns 000. The posterior probabilities that the three coordinates equal 1 are instead $9/17, 9/34, 9/34$, so bitwise MAP returns $100 \notin C$.

[9] Using the fact that the bit-flip channel commutes with the PSC (cf. Sections 5.1.1 and 5.2.1), one can recast the MQC at the blocklength level as the map $c \mapsto c \oplus \mathbf{e} \mapsto X^{\mathbf{c}\oplus\mathbf{e}} |\mathrm{Ber}(r)^n\rangle$, where $\mathbf{e} \sim \mathrm{Ber}(\eta)^n$. Put compactly, we have the identity $\mathrm{MQC}(r, \eta) = \mathrm{PSC}(r) \circ \mathrm{BSC}(\eta)$. The fact that $\mathbf{e}$ is not retained is precisely what produces "mixedness" in the MQC. Note that this perspective is supplemental to the current discussion but is implicitly exploited in the proof of Proposition 13 and is closely related to Remark 5.1.

On one hand, by data processing—the monotonicity of quantum relative entropy under channels [Lin75, Uhl77]—adding BSC noise naturally decreases the Holevo information of this new channel. On the other hand, also by data processing (Lemma 10), the PGM bit error rate goes up (cf. Remark 5.4). Thus, the success of this approach comes down to numerically inspecting whether the Holevo information can decay faster than the PGM bit error rate. Remarkably, for a tiny value of $\eta$, the drop in Holevo information outpaces the increase in the PGM bit error rate, leading to an improvement over the bound attained by the PSC, i.e., the first MRRW bound (see Fig. 2). This is the content of Proposition 25.

### 1.4 Unique decoding, list decoding, and PGM decoding

There is a useful analogy behind our work that we expound on in this subsection. If one insists on including the rate bounds derived from unique-decoding arguments in our narrative, then Hamming [Ham50] and Singleton [Sin64] bounds should correspond to the BSC and BEC, respectively. By resorting to more standard notions than the ones in Theorem 1, one can make the case that the stronger Elias–Bassalygo and Plotkin bounds should instead be interpreted as the 'list-decoding thresholds' for the BSC and BEC, respectively. This alludes to a preliminary theory of unique-decoding and list-decoding thresholds for all channels. Upon initial inspection, the Tietäväinen [Tie90] bound and the first MRRW bound corroborate this pattern, the former being the unique decoding threshold and the latter being the list-decoding threshold of some appropriate channel. When presented with this pattern, one could surmise the existence of a channel whose unique-decoding and list-decoding thresholds are prescribed by the Tietäväinen and the first MRRW bounds, respectively. As the preceding discourse has inadvertently revealed, the pure-state channel (PSC) precisely fulfills that role, but that comes at the expense of reassessing what list-decoding is ultimately offering us, as we will now explain. This analogy is summarized in the table below.

| **Channel** | **Unique decoding** | **List/PGM decoding** |
|---|---|---|
| BSC($p$) | Hamming: $p < \delta/2$ gives $R_2(\delta) \leqslant 1 - h(\delta/2)$ | Elias–Bassalygo: $p < J(\delta)$ gives $R_2(\delta) \leqslant 1 - h(J(\delta))$[10] |
| BEC($p$) | Singleton: $p < \delta$ gives $R_2(\delta) \leqslant 1 - \delta$ | Plotkin: $p < 2\delta$ gives $R_2(\delta) \leqslant 1 - 2\delta$ |
| PSC($p$) | Tietäväinen: computational-basis decoding at $p < \delta/2$ gives $R_2(\delta) \leqslant h\left(\frac{1}{2} - \sqrt{\frac{\delta}{2}\left(1 - \frac{\delta}{2}\right)}\right)$ | first MRRW: PGM bit error rate $p < \delta$ gives $R_2(\delta) \leqslant h\left(\frac{1}{2} - \sqrt{\delta(1-\delta)}\right)$ |

The fact that the Tietäväinen bound naturally corresponds to the unique-decoding analog over PSC is not difficult to observe. Indeed, measuring the output of PSC($p$) in the computational basis yields an output over the classical channel BSC($p$). Thus, if $p < \delta/2$, then the concentration of the error fraction below $\delta/2$ and the minimum distance assumption lead to the block error tending to zero as $n \to \infty$. Invoking the Holevo bound [Hol73a] and letting $p \to (\delta/2)^-$ therefore yields the rate bound

$$R_2(\delta) \leqslant h\left(\frac{1}{2} - \sqrt{\frac{\delta}{2}\left(1 - \frac{\delta}{2}\right)}\right),$$

[10] Here, $J(\delta) = (1 - \sqrt{1 - 2\delta})/2$ is the binary Johnson radius [Joh62].

which is the Tietäväinen bound [Tie90].[11] Note that this bound did not leverage the coherence of the PSC output at all.

Enticed by this observation, one is therefore led to speculate that the first MRRW bound might be attainable over PSC for $p < \delta$ via list-decoding. As the preceding discussion reveals, this roadway runs into a revealing obstacle. Indeed, the coherence of the output that one needs to crucially leverage is not amenable to the combinatorial nature of list-decoding, eventually leading us to consider the alternative (and more concrete) approach of PGM decoding.

Nonetheless, the two approaches have a key common feature worth spelling out. Observe that the constant-size list already gives the operational conclusion needed by the rate argument. Namely, if a decoder returns a list of at most $L = O(1)$ codewords containing $\mathbf{c}$ with probability $1 - o(1)$, then choosing uniformly from that list returns $\mathbf{c}$ with probability at least $(1 - o(1))/L$, yielding a constant chance of successful block decoding. This is exactly what one derives from applying PGM over PSC (Corollary 18) and what is ultimately needed. Whereas the PGM approach engulfs both classical and quantum channels, it appears that list decoding does not, as far as we understand.

### 1.5 Syndrome duality

We now motivate our shift to quantum in reproving the first MRRW bound by presenting a connection between decodability over PSC and the well-established (and elegant) dual covering arguments of [FT05, NS09]. For a binary linear code $C \leqslant \mathbb{F}_2^n$ whose distance is $d$, their arguments prove the first MRRW bound by showing that translates of the Hamming ball of radius $\frac{n}{2} - \sqrt{d(n-d)} + o(n)$ by the words of $C^\perp$ cover at least a $1/n$ fraction of $\mathbb{F}_2^n$. Since $|C| \cdot |C^\perp| = 2^n$, such a covering forces $|C|$ to be at most $n$ times the volume of that ball. Letting $d/n \to \delta$, we arrive at the first MRRW bound.

Two vital features of this route stand out: (i) it is carried out entirely on the dual code $C^\perp$, and (ii) the relative radius $\frac{1}{2} - \sqrt{\delta(1-\delta)}$ that it requires is exactly the Bernoulli parameter that will govern the dual side of the PSC. By observing the Holevo information identity for PSC below, the connection between our PSC approach for the first MRRW bound and previous approaches becomes well-motivated. Note that this proposition has been observed in previous works [Ren18, RP21, CT24, BCT26]. We mention it here as it substantiates our channel-based approach with previous approaches to the first MRRW bound.

**Proposition 5** (Syndrome duality — [BCT26, Proposition 4];[RP21, Lemma 16])**.** *Let $\mathbf{c}$ be a uniform codeword of a binary linear code $C \leqslant \mathbb{F}_2^n$. Let $\sigma_{\mathbf{c}}$ be the output state of the input codeword $\mathbf{c}$ over* PSC$(p)$*. Let $\mathbf{z} \sim \mathrm{Ber}(\frac{1}{2} - \sqrt{p(1-p)})^n$. Then*

$$I(\mathbf{c}; \sigma_{\mathbf{c}}) = H\left(\mathbf{z} \bmod C^\perp\right), \tag{8}$$

*where the entropy on the right is that of the random coset of $\mathbf{z}$ in $\mathbb{F}_2^n / C^\perp$. Note that by (13), the left-hand side is the block-level Holevo information.*[12]

[11] The main focus of [Tie90] was on proving a much stronger claim. Namely, it showed that the relative *exact* dual covering radius of a binary linear code of relative distance $\delta$ is at most $\frac{1}{2} - \sqrt{\frac{\delta}{2}\left(1 - \frac{\delta}{2}\right)}$.

[12] As an interesting consequence of syndrome duality, if one measures $\sigma_{\mathbf{c}}$ in the computational basis, the quantum data processing inequality thus yields the inequality $H(\mathbf{z} \bmod C^\perp) = I(\mathbf{c}; \sigma_{\mathbf{c}}) \geqslant I(\mathbf{c}; \mathbf{c} + \mathbf{e})$ where $\mathbf{e} \sim \mathrm{Ber}(p)^n$. This inequality by itself recovers a statistical analog of Tietäväinen [Tie90] but not the first MRRW bound. Nonetheless, this elegant inequality readily follows from Proposition 5, albeit using a non-trivial but simple quantum tool, and, to the best of our knowledge, seems not to have been previously noted in the literature.

Indeed, one can reinterpret the works of [FT05, NS09] as constructing a distribution $\mathcal{D}$ over $\mathbb{F}_2^n$ satisfying the following conditions for a sample $\mathbf{v} \sim \mathcal{D}$: (i) permuting coordinates of $\mathbf{v}$ yields the same distribution $\mathcal{D}$, (ii) the event $\mathrm{wt}(\mathbf{v}) \leqslant \frac{n}{2} - \sqrt{d(n-d)} + o(n)$ occurs with probability 1, and (iii) the quantity $\mathbf{v} \bmod C^\perp$ attains at least $1/n$ of the elements in $\mathbb{F}_2^n/C^\perp$. Given conditions (i) and (ii), it seems natural to consider beforehand whether one can just instantiate a similar approach but with $\mathrm{Ber}(\frac{1}{2} - \sqrt{\delta(1-\delta)})^n$ in place of $\mathcal{D}$. By Proposition 5, this is indeed what decoding over the PSC is achieving under the hood.

Next, we sketch the proof of Proposition 5 by setting up the PSC as a pure bipartite system. Indeed, write $|C\rangle \coloneqq |C|^{-1/2} \sum_{c\in C} |c\rangle$ for the uniform coherent superposition over the code. Let $A$ and $B$ be two $n$-qubit registers, and let $\mathrm{CNOT}_{A\to B} \coloneqq \sum_{x\in\{0,1\}^n} |x\rangle\langle x|_A \otimes X_B^x$ be the transversal CNOT with control $A$ and target $B$. Applying it to the code register and the coherent Bernoulli state, and using $X^c |\mathrm{Ber}(p)^n\rangle = |\sigma_c\rangle$, gives

$$|\Psi_C(p)\rangle \coloneqq \mathrm{CNOT}_{A\to B}\left(|C\rangle_A \otimes |\mathrm{Ber}(p)^n\rangle_B\right) = \frac{1}{\sqrt{|C|}} \sum_{c\in C} |c\rangle_A \otimes |\sigma_c\rangle_B. \tag{9}$$

Measuring $A$ in the computational basis leaves $B$ in the PSC output of a uniform codeword, so this pure state purifies the ensemble of Proposition 5.

Now apply the Hadamard gate $\mathsf{H}$ to each of the $2n$ qubits. On the code register it exchanges the code for its dual, $\mathsf{H}^{\otimes n} |C\rangle = |C^\perp\rangle$. On the noise register, it moves the Bernoulli parameter to $q \coloneqq \frac{1}{2} - \sqrt{p(1-p)}$. That is, $\mathsf{H}^{\otimes n} |\mathrm{Ber}(p)^n\rangle = |\mathrm{Ber}(q)^n\rangle$. It additionally reverses the direction of every CNOT, leading us to the identity

$$\begin{aligned} \mathsf{H}^{\otimes 2n} |\Psi_C(p)\rangle &= \mathrm{CNOT}_{B\to A}\left(|C^\perp\rangle_A \otimes |\mathrm{Ber}(q)^n\rangle_B\right) \\ &= \frac{1}{\sqrt{|C^\perp|}} \sum_{d\in C^\perp} \sum_{z\in\{0,1\}^n} \sqrt{q^{\mathrm{wt}(z)}(1-q)^{n-\mathrm{wt}(z)}}\, |d\oplus z\rangle_A \otimes |z\rangle_B. \end{aligned} \tag{10}$$

The two sides reflect two different but dual communication problems. Measuring (9) in the computational basis returns a uniform codeword $\mathbf{c}$ in $A$ and $\mathbf{c}\oplus\mathbf{e}$ in $B$ with $\mathbf{e} \sim \mathrm{Ber}(p)^n$, which is the ordinary problem of decoding a codeword of $C$ against Bernoulli-$p$ noise. Measuring (10) the same way returns $\mathbf{z} \sim \mathrm{Ber}(q)^n$ in $B$ and $\mathbf{z} + \mathbf{c}^\perp$ in $A$, where $\mathbf{c}^\perp$ is a uniformly random codeword from $C^\perp$. What has to be recovered on that side is the Bernoulli string, and what obstructs recovering it is the dual code—in a phrase, decoding Bernoulli-$q$ against the noise $C^\perp$. In this dual decoding problem, only $\mathbf{z}$ modulo $C^\perp$ can be recovered, which is exactly the coset appearing in (8).[13]

### 1.6 Related works

We now position our study within several relevant research topics.

[13] Our nomenclature behind "syndrome duality" comes from attempting to ground a classical intuition initially suggested by Section 1.4. Indeed, first observe that the coset $\mathbf{z} + C^\perp$ corresponds to the syndrome of $\mathbf{z}$ with respect to the code $C^\perp$. More precisely, if $G \in \mathbb{F}_2^{k\times n}$ is a parity-check matrix for $C^\perp$ (equivalently, a generator matrix of $C$) then $x \mapsto Gx$ has kernel exactly $C^\perp$ and therefore identifies the quotient $\mathbb{F}_2^n/C^\perp$ with the syndrome space of $C^\perp$. Ideally, one would like to say that the syndrome distribution $\mathrm{Ber}(q)^n \pmod{C^\perp} = (\mathrm{Ber}(p)^n)^\perp \pmod{C^\perp}$ is linked to the dual syndrome distribution $\mathrm{Ber}(p)^n \pmod{C}$ in some form. Ultimately, it turns out that $\mathrm{Ber}(q)^n \pmod{C^\perp}$ is instead linked to "$C \pmod{\mathrm{Ber}(p)^n}$," and it is this latter notion that seems to be best captured in the quantum identity linking (9) and (10).

**The linear-programming route to the MRRW bounds.** The original proofs of both MRRW bounds run through Delsarte's linear program [Del72, Del73]. By linear programming (LP) duality, one can thus obtain rate upper bounds by finding dual LP solutions, and that involves analyzing polynomials with nonnegative Krawtchouk expansions that are nonpositive for all evaluations from the distance $d$ to $n$. McEliece, Rodemich, Rumsey, and Welch [MRRW77] constructed those certifying the first and second bounds, the second via the Bassalygo–Elias reduction to constant-weight codes. The optimal choices of such polynomials were later systematized by Levenshtein [Lev95].

Another pathway to the Delsarte LP is through the Lovász theta function of the Hamming graph [Lov79, MRR78, Sch79]. This approach has been strengthened in several directions: semidefinite constraints from the Terwilliger algebra [Sch05], complete hierarchies for linear codes [CJJ22, LL23b, CJJ+26], and spectrally constructed dual solutions in general association schemes [CDA25]. These relaxations tighten finite-length bounds, yet no asymptotic improvement over [MRRW77] has been extracted from them. One can view our information-theoretic object—a cq channel over which minimum distance forces decodability—as a proxy for the dual certificates of the Delsarte LP in a different optimization problem. The resulting MQC and 2MQC bounds (Theorems 2 and 3) lead to asymptotic improvements over the first and second MRRW bounds.

Considering the quantum component of this work, we mention the extension of this line of work to quantum codes. Following the LP-style approach, one can place the parameters of a quantum code into a Delsarte-style linear program [AL99] that Rains sharpened through shadow enumerators [Rai99]. These apply in particular to the CSS codes [CS96, Ste96], where the quantum code is assembled from a nested pair of classical codes, and more generally to stabilizer codes through their additive GF(4) description [CRSS98], with the analysis again running through Krawtchouk polynomials.

**Alternative routes to the first MRRW bound.** The first MRRW bound has since been reproved through approaches different from the LP/SDP bounds, some of which are directly relevant to our PSC derivation (see Section 1.5). Friedman and Tillich [FT05] and Navon and Samorodnitsky [NS09] reproved the first MRRW bound through a neat dual covering argument, which involved reinterpreting the distance of $C$ as the expansion of the Cayley graph $\mathrm{Cay}(\mathbb{F}_2^n/C^\perp, \{e_1, \ldots, e_n\})$ and arguing about the covering radius of the dual code $C^\perp$. Through syndrome duality (Proposition 5), this approach directly links to our PSC derivation, though the pretty good criterion (Theorem 1) needs no dual code and applies to nonlinear codes. The same identity also places the PSC beside code smoothing [BLVW19, YZ21, DADRT23, PB23, DAR25, PB25, AGN+26], which asks when noise added to a random codeword flattens toward uniform. Indeed, the information the PSC retains is precisely what smoothing has not yet erased. Finally, we mention other similar Fourier–analytic proofs by Samorodnitsky [Sam23] and Linial and Loyfer [LL23a] and additionally barriers to pushing the Fourier–covering route toward the second MRRW bound [Sam25].

**Rate-distance tradeoffs for LDPC codes.** For Gallager's LDPC codes [Gal62], the strongest known rate–distance bounds also come from the dual-covering side. Iceland and Samorodnitsky [IS15] sharpened the coset-ball bound of [NS09] for sparse duals, and Shangguan and Yang [SY26] sharpened it further through a combination of the coset-weight generating function and a shortening argument from [BHL06]. In light of syndrome duality (Proposition 5), these approaches are naturally linked to the PSC-locality refinement in Theorem 4.

**Classical coding bounds and cq channels.** Perhaps the work closest in spirit to ours, in the sense of bounding a quantity in classical zero-error information theory using cq channels, is Dalai's

celebrated work [Dal13]. Dalai gave a quantum-information-theoretic interpretation of Lovász's bound on the Shannon capacity of a graph. He associated orthonormal representations of the confusability graph with pure-state classical–quantum channels and showed that the zero-error bound arising as the infinite-parameter limit of the cq sphere-packing bound recovers Lovász's theta function. Thus, classical Shannon capacity can be upper-bounded through a cq-channel converse framework. Dalai and Winter [DW17] proved a sphere-packing bound for constant-composition cq codes and obtained an Elias-type bound by optimizing over a family of test channels. Their use of this optimization is perhaps the closest precedent for our channel-design viewpoint.

**PSC decoding and channel duality.** The PSC has been studied extensively as a communication channel in its own right. Indeed, it is one instance of a large body of work on quantum state discrimination [Hel69, Hol73b, YKL75, HW94, BK02]. In the case of binary linear codes, decoding over PSC with a uniform prior is equivalent to state discrimination of *geometrically uniform* ensembles, for which we know that PGM achieves the optimal block (and bit) decoding error [BKMH97, UTHH99, EF01, EMV04, ZCCL25, PR25].

The PSC has also been studied through a channel coding lens. Indeed, Burnashev and Holevo [BH98] initiated the reliability-function analysis of PSC; Wilde and Guha [WG13] constructed capacity-achieving polar codes for binary-input cq channels; Renes [Ren17] introduced belief propagation with quantum messages; Piveteau and Renes [PR25] gave efficient optimal decoders for codes with efficient trellis representations; and Mandal and Pfister [MP26] developed quantum message passing for group-covariant generalizations. Renes [Ren18] placed the PSC within a general duality between a code on a cq channel and its dual code on the dual channel, whose BSC–PSC specialization, made explicit by Rengaswamy and Pfister [RP21], is the source of the syndrome duality used here (Proposition 5). The decoders of Piveteau and Renes [PR25] proceed by quantum belief propagation, and their optimality holds bitwise as well as blockwise: grouping the outcomes of the block PGM according to a single codeword bit again yields the Helstrom-optimal measurement for that bit. That bitwise form is what our LDPC refinement rests on (Fact 28). Various decoding criteria have also been studied: code-aware unambiguous discrimination [MSPJ26], its code-agnostic linear-programming counterpart [BC25], and zero-error list decoding for pure-state channels [DGL26] certify figures of merit different from the posterior bit error rate driving our converse.

**Regev's reduction and decoded quantum interferometry.** A recent algorithmic line makes coherent PSC decoding a computational primitive. Regev's quantum reduction for lattices [Reg09] was adapted to random linear codes by Debris-Alazard, Remaud, and Tillich [DART24]; Chailloux and Tillich [CT24] formulated the resulting quantum-decoding problem, whose Bernoulli case is exactly decoding over the binary PSC, and Blanvillain, Chailloux, and Tillich [BCT26] extended the analysis to general memoryless noise and the induced PGM threshold. Decoded quantum interferometry [JSW+25] runs the same pipeline to reduce optimization to decoding: a coherent decoder uncomputes the codeword from a superposition of its noisy translates so that a Fourier transform can sample the dual code. Chailloux and Tillich [CT25] prove a generic reduction from syndrome decoding to coset sampling that accommodates imperfect decoders, Marwaha et al. [MFGH25] show that DQI constructively realizes a coding bound based on the MacWilliams identity, and Anschuetz, Gamarnik, and Lu [AGL25] prove overlap-gap obstructions for typical MAX-$k$-XOR-SAT instances. For all of these, the value of the PSC lies in decoding it efficiently and coherently. Our focus, in contrast, on PSC is purely information-theoretic. As such, it enables us to

consider the PGM, which seems in general difficult to efficiently implement. A general-purpose quantum algorithm for the PGM is known [GLM+22], while efficient decoders over the PSC have so far been confined to structured families, such as codes with small trellises [PR25] or those reachable by quantum message passing [Ren17, MP26].

**Quantum proofs of classical coding statements.** There have been some notable successes in the use of quantum methods in establishing purely classical statements. We already mentioned Dalai's work connecting the Lovász theta function to cq channels [Dal13]. Kerenidis and de Wolf [KdW04] showed an exponential lower bound on the blocklength of any 2-query locally decodable code (2-LDC) by constructing a quantum random-access code (QRAC) from the 2-LDC and invoking Nayak's lower bound [Nay99]. Our approach happens to be related to that of [KdW04]. Indeed, in the high-noise regime $p = \frac{1}{2} - \Omega(n^{-1/2})$, one can show that the output of a random codeword from any fixed 2-LDC is in fact a QRAC. The connection to our approach becomes more pronounced when one replaces Nayak's bound with the PGM in the proof of [KdW04], as was done in [LdW25]. For more on the quantum method, see the survey of Drucker and de Wolf [DdW11].

Finally, Regev's reduction and the quantum method suggest a natural synthesis: the former has primarily been developed as an algorithmic tool, while the latter has largely served as a proof technique. Syndrome duality (Proposition 5) reframes the Fourier transform at the heart of Regev's reduction as an information-theoretic identity rather than a sampling procedure. From this perspective, the present work represents a first step toward unifying these viewpoints, and we believe this direction has considerable further potential.

### 1.7 Concurrent work

Concurrent work by OpenAI [Ope26] also improves the first and second MRRW bounds, albeit using a different approach. We stress that all our results reported here were obtained prior to the announcement of [Ope26]. The announcement [Ope26] did, however, influence our choice to share our work at this moment, while we are pursuing a more comprehensive understanding of our framework.

A comparative check with GPT-5.6 Sol Pro suggested that the quantitative improvements to the first MRRW bound in our work and [Ope26] match, and that the improvement to the second MRRW bound in [Ope26] is captured by our framework through an appropriate four-dimensional binary-input output-symmetric cq channel, though we have not verified this.

A defining aspect of our work is its foundation in the general modular framework of cq channels. This resulting viewpoint reduces the derivation of new bounds to finding suitable channels and makes the framework readily applicable to $q$-ary codes and LDPC codes.[14]

### 1.8 Acknowledgments

We are grateful to Thiago Bergamaschi, Anthony (Chi-Fang) Chen, Louis Golowich, Dominik Hangleiter, Nathan Ju, Zeph Landau, Tony Metger, Chris Pattison, Angelos Pelecanos, Ewin Tang, Francisca Vasconcelos, Ági Villányi, and John Wright for helpful conversations surrounding

[14] Given the close connection between binary codes and high-dimensional sphere packing [DGS77, KL78, CE03], we suspect that our cq channel coding framework could shed new light on spherical codes as well. We leave this as an interesting direction for future work.

quantum information theory at the early stages of this project. The first author especially thanks Jack Spilecki for numerous insightful discussions surrounding PGM and the Petz recovery map.

### 1.9 AI use statement

For a couple of years, the authors toyed with the possibility of reconciling the two approaches to obtaining the exponential lower bounds for 3-query locally correctable codes, via the Kikuchi method [KM24a, Yan24, KM24b] and the dual covering radius approach [IS20, AG24], in a manner that also yields the first MRRW bound in the spirit of [NS09].

The recent activity of instantiating Regev's reduction [Reg09] for binary linear codes [CLZ22, DART24, CT24, JSW+25, MFGH25, CT25, PR25, BCT26, KOW26] provided inspiration in this regard. It rekindled the old results of [Tie90], from which the authors conceived of the list-decoding table in Section 1.4. Coupled with the simple realization that the Hadamard transform of $|\mathrm{Ber}(p)^n\rangle$ is just $|\mathrm{Ber}(1/2 - \sqrt{p(1-p)})^n\rangle$, which was eerily similar to the formulas appearing in [Tie90, MRRW77], it led the authors to investigate a new quantum proof of the first MRRW bound, ideally one without complex calculations. This investigation was also partially inspired by the success of the quantum method [KdW04] within the scope of locally decodable codes.

After a long-term investigation involving several interactions with GPT-5.2, GPT-5.4, and GPT-5.5, the authors made progress on the problem and arrived at Corollary 18, along with its potential to prove the first MRRW bound when applied to the PSC. Following the release of GPT-5.6, further interactions led to a PSC sharp-threshold result analogous to [TZ00, PSW24].[15] Combined with Corollary 18, this yielded a new quantum-based proof of the first MRRW bound for binary linear codes.

Then, when the authors presented this new quantum-based proof of the first MRRW bound to GPT-5.6 Sol Pro and asked whether the same framework could recover, or possibly improve upon, the second MRRW bound, GPT-5.6 Sol Pro heuristically proposed the MQC as a candidate channel. Later interactions led the authors to realize that the strong converse for cq channels suffices to complete the proofs of the desired rate upper bound claimed in Theorem 1, leading to a new proof and improvement of the first MRRW bound. Subsequent interactions focused on the second MRRW bound led to the proposals of the masked PSC and the 2MQC.

Codex and Claude Code were employed when drafting the paper. The exposition and mathematical setup of this work were orchestrated by the authors, and the authors take full responsibility for the contents of this paper.

### 1.10 Organization

In Section 2, we introduce the quantum information theory tools needed to prove Theorem 1, specifically the PGM error data processing inequality (Lemma 10) and the cq channel coding strong converse (Theorem 7). In the same section, we include closed formulas needed to derive rate bounds, and additionally the masking operation for cq channels to derive the second MRRW and 2MQC rate bounds. In Section 3, we prove the pretty good criterion (Theorem 1). In Section 4, we rederive the standard bounds of Plotkin, Elias–Bassalygo, and both MRRW bounds from the BEC, BSC, PSC,

[15] This appears to be provable by using syndrome duality (Proposition 5) to recast the sharp-threshold problem as a cutoff phenomenon on a suitable Cayley graph, after which the methods of [Sal24] can be adapted to obtain an exponential cutoff phenomenon.

and masked PSC, respectively. In Section 5, we introduce the MQC and 2MQC and derive their corresponding bounds. In Section 6, we show that MQC and 2MQC yield strict improvements over the first and second MRRW bounds over all $\delta \in (0, \frac{1}{2})$. In Section 7, we lay down a blueprint for how one can extend the pretty good criterion to $q$-ary codes. Finally, Section 8 proves the LDPC refinement (Theorem 4), while Section A presents all the deferred proofs from Section 2.

## 2 Preliminaries

This section fixes the notation, recalls standard quantum-information tools [NC10, Wil21, Ren22], and packages the calculations reused by the channel constructions. For completeness, the proofs of the stated results in this section are presented together in Appendix A.

We first fix the classical coding notation. For $u, v \in \{0,1\}^n$, let $d(u,v)$ denote their Hamming distance. For ease of exposition, we will assume throughout this work that $|C| \geqslant 2$ for every binary code $C \subseteq \{0,1\}^n$ considered. The rate of a code $C$ is defined to be $R(C) \coloneqq n^{-1} \log |C|$. Its minimum distance is defined to be $d_{\min}(C) \coloneqq \min\{d(c,c') : c, c' \in C,\ c \neq c'\}$. Throughout this work, $\mathbf{c}$ denotes a uniformly random codeword from $C$.

Next, we fix the quantum notation. A ket $|y\rangle$ is a unit column vector, whereas $\langle y|$ is its conjugate transpose. The expression $|\varphi\rangle\langle\psi|$ denotes the outer product of $|\varphi\rangle$ and $\langle\psi|$, while $\langle\varphi|\psi\rangle$ denotes the inner product between $|\varphi\rangle$ and $|\psi\rangle$. In particular, $|\varphi\rangle\langle\varphi|$ is the rank-one orthogonal projector onto the span of $|\varphi\rangle$.

Next, we introduce the quantum extension of a classical probability distribution.

**Definition 2.1** (Density matrices)**.** *A* ***density matrix*** *is a finite-dimensional positive semidefinite matrix $\varrho$ satisfying the normalization* $\mathrm{Tr}(\varrho) = 1$*. We write* $\mathsf{D}(\mathbb{C}^d)$ *for the set of $d \times d$ density matrices,* $\mathrm{supp}(\varrho)$ *for the range of $\varrho$, and $\Pi_\varrho$ for the orthogonal projector onto that range.*

For a density matrix $\varrho$, its von Neumann entropy is defined to be $S(\varrho) \coloneqq -\mathrm{Tr}(\varrho \log \varrho)$. For $t \in [0,1]$, the binary entropy function is defined to be $h(t) \coloneqq -t \log t - (1-t)\log(1-t)$. Note that all logarithms throughout this work are base 2.

In order for a receiver to extract a classical outcome from a quantum state, they would need to measure it. This leads us to the formalism of a (general) measurement.

**Definition 2.2** (Positive operator-valued measurement (POVM))**.** *A* ***positive operator-valued measurement (POVM)*** *with outcome set $\mathcal{X}$ is a family of $d \times d$ positive semidefinite matrices $\{M_{\hat{x}} \in \mathbb{C}^{d\times d} : \hat{x} \in \mathcal{X}\}$ satisfying the identity $\sum_{\hat{x}\in\mathcal{X}} M_{\hat{x}} = I_d$. Measuring a density matrix $\varrho \in \mathsf{D}(\mathbb{C}^d)$ with respect to this POVM returns $\hat{x}$ with probability* $\mathrm{Tr}(\varrho M_{\hat{x}})$*.*

Throughout this work, Hermitian matrices will be called **observables**, the quantum extension of classical random variables. Consequently, we will call the POVM matrices $M_{\hat{x}}$ observables. Note that projective (von Neumann) measurements are POVMs in which the observables are orthogonal projection matrices ($M_{\hat{x}}^2 = M_{\hat{x}}$ for all $\hat{x} \in \mathcal{X}$ and $M_{\hat{x}_1} M_{\hat{x}_2} = 0$ for all $\hat{x}_1, \hat{x}_2 \in \mathcal{X}$ with $\hat{x}_1 \neq \hat{x}_2$). While projective (von Neumann) measurements are the prototypical notion of measurements, POVMs can be instantiated as projective measurements using ancilla qubits [NC10, Section 2.2.8].

## 2.1 Binary-input cq channels

In this subsection, we recall quantum channels and then specialize to binary-input output-symmetric cq channels and their memoryless extensions.

**Definition 2.3** (Quantum channel)**.** *Fix positive integers $d_A$ and $d_B$. A **quantum channel** is a linear map $\mathcal{N} : \mathbb{C}^{d_A \times d_A} \to \mathbb{C}^{d_B \times d_B}$ satisfying the following two properties:*

(i) (Trace preservation) *For every $X \in \mathbb{C}^{d_A \times d_A}$, one has $\mathrm{Tr}(\mathcal{N}(X)) = \mathrm{Tr}(X)$.*

(ii) (Complete positivity) *For every positive integer $r$, the map $\mathcal{N} \otimes \mathrm{id}_r$ takes positive semidefinite matrices in $\mathbb{C}^{d_A r \times d_A r}$ to positive semidefinite matrices in $\mathbb{C}^{d_B r \times d_B r}$, where $\mathrm{id}_r$ is the identity map on $r \times r$ matrices.*

*We say $\mathcal{N}$ is a **classical–quantum (cq) channel** if it has the following measure-and-prepare form. Fix a finite input alphabet $\mathcal{X}$, set $d_A = |\mathcal{X}|$, and choose one output state $\sigma_x \in \mathsf{D}(\mathbb{C}^{d_B})$ for each $x \in \mathcal{X}$. Relative to the orthonormal basis $\{|x\rangle\}_{x \in \mathcal{X}}$, the corresponding quantum channel acts as*

$$\mathcal{N}(\varrho) := \sum_{x \in \mathcal{X}} \langle x| \varrho |x\rangle \, \sigma_x.$$

*In this work, we abbreviate channels with the simpler notation $\mathcal{N} : x \mapsto \sigma_x$.*

We use two standard quantum channels. For an $n$-partite system of $d$-dimensional states $\varrho_1 \otimes \cdots \otimes \varrho_n$ and an index $i \in \{1, \ldots, n\}$, the **partial trace channel to the $i$'th qudit** $\mathcal{T}_i : \mathbb{C}^{d^n \times d^n} \to \mathbb{C}^{d \times d}$ is defined on product states by

$$\mathcal{T}_i : \varrho_1 \otimes \cdots \otimes \varrho_n \longmapsto \varrho_i. \tag{11}$$

By linearity, this definition extends to non-product states. Operationally, $\mathcal{T}_i$ discards every qudit except the $i$'th one. We use it to reduce blockwise error to bitwise error in the proof of Theorem 17.

The other quantum channel we consider in this work is the **X-Pauli channel** (or bit-flip channel). For $\eta \in [0, 1]$, it is the random-unitary qubit channel

$$\mathcal{B}_\eta(\varrho) := (1 - \eta)\varrho + \eta X \varrho X. \tag{12}$$

Here, $X$ is the Pauli observable that exchanges the computational-basis vectors. Thus, $\mathcal{B}_\eta$ applies $X$ with probability $\eta$ and otherwise does nothing. In Bloch coordinates, it leaves the $X$ component unchanged and multiplies the $Z$ component by $1 - 2\eta$ (cf. (32)). In this work, the channel $\mathcal{B}_\eta$ is used as post-processing to construct the MQC in Sections 1.3 and 5.1.

The binary-input cq channels employed in this work form a particularly simple class of cq channels, defined as follows.

**Definition 2.4** (Binary-input cq channel)**.** *For a positive integer $d$, a **binary-input classical–quantum (cq) channel** is a map $\Phi_\sigma : \{0, 1\} \to \mathsf{D}(\mathbb{C}^d)$ specified by $\Phi_\sigma(b) = \sigma_b$. It is **output-symmetric** if there exists a $d \times d$ unitary matrix $U$ that exchanges the outputs, meaning that $U\sigma_0 U^\dagger = \sigma_1$ and $U\sigma_1 U^\dagger = \sigma_0$. The memoryless extension of $\Phi_\sigma$ sends a codeword $c = (c_1, \ldots, c_n)$ to the product state $\sigma_c := \Phi_\sigma^{\otimes n}(c) = \sigma_{c_1} \otimes \cdots \otimes \sigma_{c_n}$.*

Next, we define the first key quantity, the Holevo information [Hol73a].

**Definition 2.5** (Holevo information)**.** *The **Holevo information** of a finite ensemble of density matrices $\mathcal{E} \coloneqq \{(\pi_x, \sigma_x) : x \in \mathcal{X}\}$ with average state $\bar{\sigma} \coloneqq \sum_{x\in\mathcal{X}} \pi_x \sigma_x$ is defined to be*

$$\chi(\mathcal{E}) \coloneqq I(\mathbf{x}; \sigma_{\mathbf{x}}) = S(\bar{\sigma}) - \sum_{x\in\mathcal{X}} \pi_x S(\sigma_x). \tag{13}$$

*For the uniform binary ensemble $\mathcal{E} = \{(\frac{1}{2}, \sigma_0), (\frac{1}{2}, \sigma_1)\}$, we write $\chi(\sigma_0, \sigma_1) \coloneqq \chi(\mathcal{E})$.*[16]

The operational meaning of Holevo information is as follows. For a random label $\mathbf{x} \sim \pi$, the Holevo information is precisely the von Neumann mutual information $I(\mathbf{x}; \sigma_{\mathbf{x}})$.[17] For commuting outputs $\{\sigma_x\}_{x\in\mathcal{X}}$, it reduces to ordinary mutual information.

Next, we define the quantum extension of channel capacity for cq channels. For comparison, the capacity of a classical discrete memoryless channel $W : x \mapsto W(\cdot \mid x)$ is $C(W) = \max_\pi I(\mathbf{x}; \mathbf{y})$, where $\mathbf{x} \sim \pi$ and $\mathbf{y}$ is the channel output. The cq definition is formally identical, with Holevo information in place of classical mutual information.

**Definition 2.6** (Classical (HSW) capacity of cq channels)**.** *For any cq channel $\Phi : \mathcal{X} \to \mathsf{D}(\mathbb{C}^d)$ with $\Phi(x) = \sigma_x$, the **classical capacity**, normalized per channel use, is defined to be*

$$C(\Phi) \coloneqq \limsup_{n\to\infty} \frac{1}{n} \max_{\pi^{(n)}} \chi\left(\{(\pi^{(n)}(x^n), \sigma_{x_1} \otimes \cdots \otimes \sigma_{x_n}) : x^n \in \mathcal{X}^n\}\right),$$

*where the maximum is over all probability distributions $\pi^{(n)}$ on $\mathcal{X}^n$.*

The above capacity has a single-letter characterization as the *Holevo capacity* of the cq channel, defined as

$$\max_\pi \chi\left(\{(\pi(x), \sigma_x) : x \in \mathcal{X}\}\right),$$

where the maximum is over all probability distributions $\pi$ on $\mathcal{X}$. Further, when the cq channel is output-symmetric, the Holevo capacity equals the Holevo information for the uniform prior, as stated in the following proposition. Consequently, this justifies the reference to the Holevo information as the channel capacity throughout this work. The proof appears in Section A.3.

**Proposition 6** (Capacity of a binary-input output-symmetric cq channel)**.** *Let $\Phi_\sigma : b \mapsto \sigma_b$ be a binary-input output-symmetric cq channel. Its capacity satisfies*

$$C(\Phi_\sigma) = \chi(\sigma_0, \sigma_1).$$

As a slight digression, the Holevo–Schumacher–Westmoreland (HSW) coding theorem [Hol98, SW97, HJS+96] identifies cq channel capacity with the maximum Holevo information over input priors, much like Shannon's formula in the classical case. Its positive side states that, at every rate below $C(\Phi)$, there are length-$n$ codes and decoder POVMs whose average block error tends to zero.

Now, we turn to the more relevant side of the cq channel coding theorem, namely the previously established cq channel coding strong converse [Win99, ON99, MO17] that turns nonvanishing block-decoding success into a rate bound. Specifically, [Win99, Theorem 13] shows that every code

[16] When the states $\{\sigma_x\}_{x\in\mathcal{X}} \subseteq \mathsf{D}(\mathbb{C}^d)$ are specified without an associated prior distribution, we take the prior to be uniform when defining the Holevo information of the ensemble. This convention is assumed throughout this work.

[17] To be precise, this mutual information is relative to the bipartite state $\sum_{x\in\mathcal{X}} \pi_x |x\rangle\langle x| \otimes \sigma_x$.

with decoding error at most $\lambda < 1$ has rate at most $C(\Phi) + O_\lambda(1/\sqrt{n})$, while [ON99, Theorem 1] lower-bounds the average error probability of an arbitrary code by an exponential expression that forces the success probability to decay exponentially at every rate above the capacity.[18]

For our purposes, the version below is tailored specifically to binary-input output-symmetric cq channels. Thus, by Proposition 6, we can state it with the Holevo information $\chi(\sigma_0, \sigma_1)$ in place of the Holevo capacity. For completeness, its proof appears in Section A.4.

**Theorem 7** (cq channel coding strong converse — [Win99, Theorem 13] and [ON99, Theorem 1])**.** *Fix a finite-dimensional binary-input output-symmetric cq channel $\Phi_\sigma$. There is a constant $L_\sigma \geqslant 1$, depending only on its two output states $\sigma_0$ and $\sigma_1$, such that the following holds. Let $C \subseteq \{0,1\}^n$ be any binary code, and let $\mathbf{c}$ be a uniformly random codeword from $C$. For any $\varepsilon > 0$ and decoder POVM $\{D_c : c \in C\}$, if $\hat{\mathbf{c}}$ is obtained by measuring $\sigma_{\mathbf{c}}$ with this POVM, then*

$$\mathbf{Pr}[\hat{\mathbf{c}} = \mathbf{c}] \leqslant 2^{-n[R(C)-\chi(\sigma_0,\sigma_1)-2\varepsilon]} + 4\exp\left(-\varepsilon^2 n / L_\sigma^2\right). \tag{14}$$

*Consequently, any fixed positive rate gap above $\chi(\sigma_0, \sigma_1)$ forces the average decoding success probability to decay exponentially in $n$.*

*Remark* 2.1 (Qualitative decay suffices). For the asymptotic conclusion of Theorem 1, it suffices if the successful block decoding probability tends to zero, at the expense of having weaker finite blocklength estimates. The exponential decay in Theorem 7 is used in the proof of Theorem 1 to obtain the precise $O(n^{-1/2})$ lower-order term in Theorem 1, which is more than enough for the asymptotic version. ◁

## 2.2 The pretty good measurement (PGM)

Throughout this subsection, we fix a finite ensemble of density matrices $\mathcal{E} := \{(\pi_x, \sigma_x) : x \in \mathcal{X}\}$, where $\pi_x \geqslant 0$ and $\sum_{x\in\mathcal{X}} \pi_x = 1$, and write $\bar{\sigma} := \sum_{x\in\mathcal{X}} \pi_x \sigma_x = \mathbf{E}_{\mathbf{x}\sim\pi}[\sigma_{\mathbf{x}}]$ for its average state. This subsection defines the pretty good measurement (PGM), also called the square-root measurement [Bel75, Hol79, HW94, HJS+96], records its exact compatibility with coarse-graining, and states the data-processing consequence used in Section 3. See Renes [Ren22, Section 11.5] for a textbook treatment of PGM and [Wri24, Lectures 6 and 7] for an instructive gentle exposition of it.

**Definition 2.7** (Pretty good measurement (PGM))**.** *The **pretty good measurement (PGM)** of the ensemble $\mathcal{E}$ is the POVM over* $\mathrm{supp}(\bar{\sigma})$ *with observables $M_{\hat{x}} := \pi_{\hat{x}} \bar{\sigma}^{-1/2} \sigma_{\hat{x}} \bar{\sigma}^{-1/2}$, where inverses are over* $\mathrm{supp}(\bar{\sigma})$.[19]

When $\mathcal{E}$ is a classical ensemble (i.e., the states $\{\sigma_x\}_{x\in\mathcal{X}}$ are simultaneously diagonalizable), the PGM reduces to classical posterior sampling. Indeed, assume without loss of generality that all $\sigma_x$ are diagonal $d \times d$ density matrices. Then the average state $\bar{\sigma}$ and the operators $\{M_{\hat{x}}\}_{\hat{x}\in\mathcal{X}}$ are also diagonal. Hence, conditioned on observing $\ell \in \{1, \ldots, d\}$, the PGM returns $\hat{x} \in \mathcal{X}$ with probability

$$(M_{\hat{x}})_{\ell,\ell} = (\pi_{\hat{x}} \bar{\sigma}^{-1/2} \sigma_{\hat{x}} \bar{\sigma}^{-1/2})_{\ell,\ell} = \frac{\pi_{\hat{x}} \mathbf{Pr}[\boldsymbol{\ell} = \ell \mid \mathbf{x} = \hat{x}]}{\mathbf{Pr}[\boldsymbol{\ell} = \ell]} = \mathbf{Pr}[\mathbf{x} = \hat{x} \mid \boldsymbol{\ell} = \ell],$$

which is precisely posterior sampling.

Next, we define our second key quantity, the PGM error, which we interchangeably call posterior error in this work.

[18] The exact strong-converse exponent was later determined in [MO17].

[19] Note that these operators sum to the support projector $\Pi_{\bar{\sigma}}$.

**Definition 2.8** (PGM error)**.** *Let $\hat{\mathbf{x}}$ be the outcome of the PGM for the ensemble $\mathcal{E}$. The **PGM error** of $\mathcal{E}$ is defined to be the quantity*

$$p_{\mathrm{e}}(\mathcal{E}) \coloneqq \mathbf{Pr}[\hat{\mathbf{x}} \neq \mathbf{x}] = \sum_{\substack{\hat{x},x\in\mathcal{X}:\\ \hat{x}\neq x}} \pi_x \operatorname{Tr}(\sigma_x M_{\hat{x}}).$$

*For binary ensembles $\mathcal{E}_\alpha = \{(1-\alpha,\sigma_0),(\alpha,\sigma_1)\}$, we write $p_{\mathrm{e},\alpha}(\sigma_0,\sigma_1) \coloneqq p_{\mathrm{e}}(\mathcal{E}_\alpha)$. In the more special uniform-prior case when $\alpha = \frac{1}{2}$, we write $p_{\mathrm{e}}(\sigma_0,\sigma_1) \coloneqq p_{\mathrm{e},\frac{1}{2}}(\sigma_0,\sigma_1)$ and call it the **PGM bit error rate** of the output state pair $(\sigma_0,\sigma_1)$.*

We next record the following central property of PGM, which was previously exploited in [KT08, BJL23, LdW25]. Its proof appears in Section A.5.

**Lemma 8** (PGM is closed under coarse-graining)**.** *Fix any classical map $f : \mathcal{X} \to \mathcal{Y}$. Let $\mathcal{F} \coloneqq \{(\nu_y,\omega_y) : y \in \mathcal{Y}\}$ be the pushforward ensemble of $\mathcal{E}$ under $f$, where $\nu_y \coloneqq \mathbf{Pr}[f(\mathbf{x}) = y]$ and $\omega_y \coloneqq \mathbf{E}[\sigma_{\mathbf{x}} \mid f(\mathbf{x}) = y]$. For a sample $\mathbf{x}$ from $\mathcal{E}$, if $\hat{\mathbf{x}} \in \mathcal{X}$ is the outcome of the PGM of $\mathcal{E}$ when applied to $\sigma_{\mathbf{x}}$, then $f(\hat{\mathbf{x}})$ is the outcome of the PGM of $\mathcal{F}$ when applied to $\omega_{f(\mathbf{x})}$. As a result, we have the identity*

$$\mathbf{Pr}[f(\hat{\mathbf{x}}) \neq f(\mathbf{x})] = p_{\mathrm{e}}(\mathcal{F}).$$

In Section 3, the ensemble of interest will be $\mathcal{E} = \{(|C|^{-1},\sigma_c) : c \in C\}$, which we will benignly conflate with $C$. For such ensembles, we will crucially rely on the following specialization of Lemma 8 applied to the projection map $f_i(c) = c_i$. This identity also explains why the PGM is preferable here to a hard maximum-a-posteriori (MAP) decoder.

**Lemma 9** (Blockwise PGM is simultaneously bitwise PGM)**.** *For a uniform sample $\mathbf{c}$ from $C$, let $\hat{\mathbf{c}} \in C$ be the outcome of the PGM applied to $\sigma_{\mathbf{c}}$. For any $i \in \{1,\dots,n\}$ and $b \in \{0,1\}$, let $\pi_i(b) \coloneqq \mathbf{Pr}[\mathbf{c}_i = b]$, and let $\sigma_i(b) \coloneqq \mathbf{E}[\sigma_{\mathbf{c}} \mid \mathbf{c}_i = b]$. Then $\hat{\mathbf{c}}_i$ is the outcome of the PGM of the induced binary ensemble $\mathcal{E}_i \coloneqq \{(\pi_i(0),\sigma_i(0)),(\pi_i(1),\sigma_i(1))\}$ when applied to $\sigma_i(\mathbf{c}_i)$. Consequently,*

$$\mathbf{Pr}[\hat{\mathbf{c}}_i \neq \mathbf{c}_i] = p_{\mathrm{e}}(\mathcal{E}_i).$$

We next record the monotonicity of the PGM error under post-processing the quantum outputs through any quantum channel. Note that this is a crucial step in establishing Theorem 17. Its proof appears in Section A.2.

**Lemma 10** (PGM error is nondecreasing under quantum channels — [Ren22, Section 11.5])**.** *Let $\mathcal{N}$ be a completely positive trace-preserving map. Applying $\mathcal{N}$ to every state in $\mathcal{E}$ cannot lower the PGM error. That is,*

$$p_{\mathrm{e}}(\mathcal{E}) \leqslant p_{\mathrm{e}}\left(\{(\pi_x,\mathcal{N}(\sigma_x)) : x \in \mathcal{X}\}\right).$$

In Theorem 17, we apply the preceding lemma to the product ensemble $\mathcal{E} = \{(|C|^{-1},\sigma_c) : c \in C\}$. The partial trace $\mathcal{T}_i$ then discards every output except the $i$'th one.

Next, to handle the bit errors in the proof of Theorem 1 after reducing to them, we will need the following prior-independent bound on each bit. Its proof appears in Section A.6.

**Lemma 11** (Uniform prior maximizes the PGM bit error rate)**.** *Let $\sigma_0,\sigma_1 \in \mathsf{D}(\mathbb{C}^d)$ be an output-symmetric pair. For any $\gamma \in [0,1]$, let $p_{\mathrm{e},\gamma}(\sigma_0,\sigma_1)$ denote the PGM bit error rate for the binary ensemble $\mathcal{E}_\gamma = \{(1-\gamma,\sigma_0),(\gamma,\sigma_1)\}$. Then*

$$p_{\mathrm{e},\gamma}(\sigma_0,\sigma_1) \leqslant p_{\mathrm{e},1/2}(\sigma_0,\sigma_1) = p_{\mathrm{e}}(\sigma_0,\sigma_1).$$

### 2.3 Closed formulas for the PGM bit error rate computations

In this subsection, we record closed formulas needed for our PGM bit error rate computations. For output states $\sigma_0, \sigma_1 \in \mathsf{D}(\mathbb{C}^d)$ and bias parameter $\alpha \in [0,1]$, define the binary ensemble $\mathcal{E}_\alpha := \{(1-\alpha, \sigma_0), (\alpha, \sigma_1)\}$. We adopt the shorthand $\bar{\sigma}_\alpha := (1-\alpha)\sigma_0 + \alpha\sigma_1$ for their average state throughout this work. In the uniform-prior case when $\alpha = \frac{1}{2}$, we simply write $\bar{\sigma} := \bar{\sigma}_{\frac{1}{2}}$. First, we record the following PGM trace identity. Its proof appears in Section A.7.

**Proposition 12** (Binary PGM error formulas)**.** *The PGM error of the binary ensemble $\mathcal{E}_\alpha$ satisfies the identity*

$$p_{\mathrm{e},\alpha}(\sigma_0, \sigma_1) = 2\alpha(1-\alpha)\operatorname{Tr}\left(\sigma_0 \bar{\sigma}_\alpha^{-1/2} \sigma_1 \bar{\sigma}_\alpha^{-1/2}\right). \tag{15}$$

*Moreover, if both output states are pure, i.e., $\sigma_b = |\psi_b\rangle\langle\psi_b|$ for unit vectors $|\psi_0\rangle, |\psi_1\rangle \in \mathbb{C}^d$ and $f := \left|\langle\psi_0|\psi_1\rangle\right|^2$, then*

$$p_{\mathrm{e},\alpha}(\sigma_0, \sigma_1) = \frac{2\alpha(1-\alpha)f}{1 + 2\sqrt{\alpha(1-\alpha)(1-f)}}. \tag{16}$$

*In particular, in the uniform-prior case when $\alpha = \frac{1}{2}$, this becomes $p_{\mathrm{e}}(\sigma_0, \sigma_1) = \frac{1}{2}(1 - \sqrt{1-f})$.*

Finally, we record how the PGM behaves when a classical binary channel precedes a cq preparation. This composition rule will be used when computing PGM bit error rates of MQC and 2MQC below. We present the proof in Section A.7.

**Proposition 13** (Binary PGM error under classical pre-processing)**.** *Fix $\eta \in [0,1]$ and $\mathbf{e} \sim \mathrm{Ber}(\eta)$. Consider the binary-input cq channel*

$$b \longmapsto b \oplus \mathbf{e} \longmapsto \sigma_{b\oplus\mathbf{e}}. \tag{17}$$

*Fix $\alpha \in [0,1]$, and let $\widehat{\mathbf{b}}$ be the outcome of the PGM for the resulting cq channel in (17) when given a non-uniform bit $\mathbf{b} \sim \mathrm{Ber}(\alpha)$ as input. Set $\mathbf{y} := \mathbf{b} \oplus \mathbf{e}$ and $\theta := \mathbf{Pr}[\mathbf{y} = 1] = \alpha + \eta - 2\alpha\eta$. Let $\widehat{\mathbf{y}}$ be the outcome of the PGM for the intermediate ensemble $\{(1-\theta, \sigma_0), (\theta, \sigma_1)\}$. Then the PGM outcome $\widehat{\mathbf{b}}$ can be generated by first producing $\widehat{\mathbf{y}}$ and then drawing $\widehat{\mathbf{b}}$ from the posterior law*

$$\mathbf{Pr}[\widehat{\mathbf{b}} = b \mid \widehat{\mathbf{y}} = y] = \mathbf{Pr}[\mathbf{b} = b \mid \mathbf{y} = y].$$

*Moreover, if we let $\tau_b := (1-\eta)\sigma_b + \eta\sigma_{b\oplus 1}$ for $b \in \{0,1\}$ be the output states of the composed cq channel in (17), then the binary PGM error for the composed channel satisfies*

$$p_{\mathrm{e},\alpha}(\tau_0, \tau_1) = 2\alpha(1-\alpha) - \left[\frac{\alpha(1-\alpha)(1-2\eta)}{\theta(1-\theta)}\right]^2 \cdot \left[2\theta(1-\theta) - p_{\mathrm{e},\theta}(\sigma_0, \sigma_1)\right]. \tag{18}$$

*In particular, for the uniform prior $\alpha = \frac{1}{2}$, the PGM error is*[20]

$$p_{\mathrm{e}}(\tau_0, \tau_1) = \frac{1}{2}\left[1 - (1-2\eta)^2(1 - 2p_{\mathrm{e}}(\sigma_0, \sigma_1))\right]. \tag{19}$$

[20] A more intuitive way to think about the uniform-prior formula is the following: the map $\mathbf{b} \mapsto \widehat{\mathbf{b}}$ can be decomposed as the composition of the maps $\mathbf{b} \mapsto \mathbf{b} \oplus \mathbf{e} = \mathbf{y}$ followed by $\mathbf{y} \mapsto \sigma_{\mathbf{y}} \mapsto \widehat{\mathbf{y}}$, then followed by $\widehat{\mathbf{y}} \mapsto \widehat{\mathbf{b}}$. Since $\mathbf{b}$ has a uniform prior, these three maps can therefore be recast as passing through the independent bit channels $\mathrm{BSC}(\eta)$, $\mathrm{BSC}(p_{\mathrm{e}})$, $\mathrm{BSC}(\eta)$, respectively.

## 2.4 Closed formulas for the Holevo information computations

In this subsection, we record the closed entropy formula used in the Holevo-information calculations of Sections 4 and 5. Every qubit appearing there is real in the computational basis and therefore lies in the plane spanned by the Pauli matrices

$$X := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad Z := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{20}$$

The following proposition reads the von Neumann entropy directly from these Pauli coefficients. Its proof appears in Section A.8.

**Proposition 14** (Qubit entropy in Pauli coordinates)**.** *For $D \in [0,1]$, define*

$$s(D) := h\left(\frac{1-\sqrt{1-D}}{2}\right). \tag{21}$$

*If $x, z \in \mathbb{R}$ satisfy $x^2 + z^2 \leqslant 1$, then $\varrho := \frac{1}{2}(I + xX + zZ)$ is a density matrix with entropy*

$$S\left(\frac{1}{2}(I + xX + zZ)\right) = h\left(\frac{1-\sqrt{x^2+z^2}}{2}\right) = s(1-x^2-z^2).$$

*Note that* $4\det(\varrho) = 1 - x^2 - z^2$.

## 2.5 Channel masking

This subsection introduces the masking operation used in defining the channels in Sections 4.4 and 5.2, which converts a non-uniform binary ensemble $\mathcal{G}_\alpha := \{(1-\alpha, \tau_0), (\alpha, \tau_1)\}$ into a binary-input output-symmetric cq channel. Similar to the previous subsection, we adopt the shorthand $\bar{\tau}_\alpha := (1-\alpha)\tau_0 + \alpha\tau_1$ for their average state.

The idea is as follows. When transmitting an input bit $b \in \{0,1\}$, the channel draws a hidden mask $m \sim \mathrm{Ber}(\alpha)$ and sends the receiver the pair $(b \oplus m, \tau_m)$. In this situation, the quantum output $\tau_m$ serves as a "quantum hint" for the mask $m$. Moreover, decoding $b$ is equivalent to decoding $m$. Thus, decoding becomes equivalent to the task of distinguishing between $\tau_0$ and $\tau_1$ under the non-uniform prior $(1-\alpha, \alpha)$.

We now proceed with its formal definition.

**Definition 2.9** (Masking)**.** *Consider density matrices $\tau_0, \tau_1 \in \mathsf{D}(\mathbb{C}^d)$ and parameter $\alpha \in [0,1]$. Their $\alpha$-**masking** is the binary output pair*

$$\Sigma_0 := (1-\alpha)\tau_0 \oplus \alpha\tau_1,$$
$$\Sigma_1 := \alpha\tau_1 \oplus (1-\alpha)\tau_0.$$

*Throughout, we denote this operation by* $\mathrm{Mask}_\alpha(\tau_0, \tau_1) := (\Sigma_0, \Sigma_1)$.

Next, we state two propositions for deriving the rate upper bounds of the masked PSC (mPSC) in Section 4.4 and the masked mixed-qubit channel (2MQC) in Section 5.2. The first shows that the PGM bit error rate of the masked pair equals that of the binary $\{(1-\alpha, \tau_0), (\alpha, \tau_1)\}$ ensemble.

Intuitively, at the bitwise level, the PGM of the masked pair reads the block label $b \oplus m$, applies the PGM of this non-uniform binary ensemble to the "quantum hint" $\tau_m$ to produce a mask estimate $\widehat{m}$, and returns $\widehat{b} \coloneqq (b \oplus m) \oplus \widehat{m}$. It therefore errs exactly when $\widehat{m} \neq m$, which is why the two error rates coincide. Its proof appears in Section A.9.

**Proposition 15** (PGM bit error rate of a masked pair)**.** *Let* $(\Sigma_0, \Sigma_1) \coloneqq \mathrm{Mask}_\alpha(\tau_0, \tau_1)$. *Then we have the PGM error identity*

$$p_{\mathrm{e}}(\Sigma_0, \Sigma_1) = p_{\mathrm{e}}\left(\mathcal{G}_\alpha\right) = 2\alpha(1-\alpha)\,\mathrm{Tr}\left(\tau_0 \bar{\tau}_\alpha^{-1/2} \tau_1 \bar{\tau}_\alpha^{-1/2}\right). \tag{22}$$

Next, we record the Holevo information of a masked pair, which readily follows from considering the description given in this subsection's second paragraph and the observation that $\chi(\Sigma_0, \Sigma_1) = I(b; (b \oplus m, \tau_m)) = I(b; b \oplus m) + I(b; \tau_m \mid b \oplus m)$. Its proof appears in Section A.9.

**Proposition 16** (Holevo information of a masked pair)**.** *Let* $(\Sigma_0, \Sigma_1) \coloneqq \mathrm{Mask}_\alpha(\tau_0, \tau_1)$. *Then we have the Holevo information identity*

$$\chi(\Sigma_0, \Sigma_1) = 1 - h(\alpha) + S(\bar{\tau}_\alpha) - (1-\alpha)S(\tau_0) - \alpha S(\tau_1). \tag{23}$$

We close with the equal-weight specialization that will relate the MQC and 2MQC rate bounds.

*Remark* 2.2 (Masking with a uniform prior). Suppose $(\tau_0, \tau_1)$ is output-symmetric, and let $U$ be a unitary that exchanges its two states. If $(\Sigma_0, \Sigma_1) \coloneqq \mathrm{Mask}_{\frac{1}{2}}(\tau_0, \tau_1)$ and $W \coloneqq I \oplus U$, then

$$W\Sigma_b W^\dagger = \frac{1}{2}\tau_b \oplus \frac{1}{2}\tau_b \cong \tau_b \otimes \frac{I}{2}, \qquad b \in \{0,1\}.$$

Thus, a uniform-prior mask recovers the original channel together with an input-independent uniform block label. In particular, it preserves both the PGM bit error rate and the uniform-input Holevo information, so optimizing over a masked channel family cannot give a weaker rate upper bound than optimizing over the original family. ◁

## 3 Proof of the main theorem

Throughout this section, $\sigma_0$ and $\sigma_1$ denote the outputs of a finite-dimensional binary-input output-symmetric cq channel $\Phi_\sigma$ (Definition 2.4), and we write $\bar{\sigma} \coloneqq (\sigma_0 + \sigma_1)/2$ for their average state. We will also fix a binary code $C \subseteq \{0,1\}^n$ throughout this section. For ease of exposition, we will conflate $C$ with the uniform ensemble $\{(|C|^{-1}, \sigma_c) : c \in C\}$.

We prove the pretty good criterion (Theorem 1) in two stages. First, in Theorem 17, we upper bound the expected Hamming distance between the transmitted codeword $\mathbf{c}$ and the PGM outcome $\hat{\mathbf{c}}$ by the channel's PGM bit error rate $p_{\mathrm{e}}(\sigma_0, \sigma_1)$, from which then the constraint $p_{\mathrm{e}}(\sigma_0, \sigma_1) < \delta$ converts that expectation bound into a constant block decoding success (Corollary 18).[21] By coupling this with the cq channel coding strong converse (Theorem 7) and setting parameters appropriately, we conclude Theorem 1.

We begin with the expected Hamming distance upper bound.

[21] Note that the code's distance assumption is invoked only once in the argument, which is at Corollary 18.

**Theorem 17** (PGM expected Hamming distance)**.** *Consider any binary-input output-symmetric cq channel $\Phi_\sigma$. Suppose $\mathbf{c}$ is a uniform sample from $C$. Let $\sigma_{\mathbf{c}} = \Phi_\sigma^{\otimes n}(\mathbf{c})$ be its output state, and let $\hat{\mathbf{c}}$ be the outcome of the PGM of $C$ applied to $\sigma_{\mathbf{c}}$. Then*

$$\mathbf{E}\left[d(\hat{\mathbf{c}}, \mathbf{c})\right] \leqslant n p_{\mathrm{e}}(\sigma_0, \sigma_1).$$

*Proof.* First, observe that

$$\mathbf{E}[d(\hat{\mathbf{c}}, \mathbf{c})] = \mathbf{Pr}[\hat{\mathbf{c}}_1 \neq \mathbf{c}_1] + \cdots + \mathbf{Pr}[\hat{\mathbf{c}}_n \neq \mathbf{c}_n].$$

Thus, it suffices to show that

$$\mathbf{Pr}[\hat{\mathbf{c}}_i \neq \mathbf{c}_i] \leqslant p_{\mathrm{e}}(\sigma_0, \sigma_1)$$

for each $i \in \{1, \ldots, n\}$. Indeed, let $\pi_i(b) \coloneqq \mathbf{Pr}[\mathbf{c}_i = b]$ for $b \in \{0, 1\}$, and define the binary ensemble $C_i \coloneqq \{(\pi_i(0), \mathbf{E}[\sigma_{\mathbf{c}} \mid \mathbf{c}_i = 0]), (\pi_i(1), \mathbf{E}[\sigma_{\mathbf{c}} \mid \mathbf{c}_i = 1])\}$. By the blockwise-to-bitwise PGM lemma (Lemma 9), we see that

$$\mathbf{Pr}[\hat{\mathbf{c}}_i \neq \mathbf{c}_i] = p_{\mathrm{e}}(C_i).$$

For any $b \in \{0, 1\}$, linearity of the $i$'th partial trace channel $\mathcal{T}_i$ and (11) give

$$\mathcal{T}_i\left(\mathbf{E}[\sigma_{\mathbf{c}} \mid \mathbf{c}_i = b]\right) = \mathbf{E}[\mathcal{T}_i(\sigma_{\mathbf{c}}) \mid \mathbf{c}_i = b] = \mathbf{E}[\sigma_{\mathbf{c}_i} \mid \mathbf{c}_i = b] = \sigma_b.$$

Applying PGM data processing (Lemma 10) to this partial trace maps $C_i$ to $\{(\pi_i(0), \sigma_0), (\pi_i(1), \sigma_1)\}$. The uniform-prior bound (Lemma 11) then gives

$$p_{\mathrm{e}}(C_i) \leqslant p_{\mathrm{e},\pi_i(1)}(\sigma_0, \sigma_1) \leqslant p_{\mathrm{e}}(\sigma_0, \sigma_1).$$

Combining the bitwise bounds therefore concludes the theorem. □

By combining this Hamming distance upper bound with the code's minimum distance, we therefore obtain a constant chance of successfully decoding over $\Phi_\sigma^{\otimes n}$.

**Corollary 18** (Constant successful block decoding)**.** *Under the hypotheses of Theorem 17, suppose in addition that $d_{\min}(C) \geqslant \delta n$ for some $\delta > 0$. Then*

$$\mathbf{Pr}[\hat{\mathbf{c}} = \mathbf{c}] \geqslant 1 - \frac{p_{\mathrm{e}}(\sigma_0, \sigma_1)}{\delta}.$$

*In particular, if $p_{\mathrm{e}}(\sigma_0, \sigma_1) < \delta$, then $\mathbf{Pr}[\hat{\mathbf{c}} = \mathbf{c}] \geqslant \Omega(1)$.*

*Proof.* On the event $\{\hat{\mathbf{c}} \neq \mathbf{c}\}$, the two random codewords are distinct, so $d(\hat{\mathbf{c}}, \mathbf{c}) \geqslant d_{\min}(C)$. Thus, $d_{\min}(C)\mathbf{Pr}[\hat{\mathbf{c}} \neq \mathbf{c}] \leqslant \mathbf{E}[d(\hat{\mathbf{c}}, \mathbf{c})]$. Combining this inequality with Theorem 17 proves the claim. □

Now, armed with Corollary 18, we can proceed to the proof of our main result.

*Proof of Theorem 1.* Set $\eta \coloneqq 1 - p_{\mathrm{e}}(\sigma_0, \sigma_1)/\delta > 0$, and let $L_\sigma$ be the channel-dependent constant in Theorem 7. Observe that Corollary 18 gives us $\mathbf{Pr}[\hat{\mathbf{c}} = \mathbf{c}] \geqslant \eta$. Now, apply the cq channel coding strong converse (Theorem 7) with

$$\varepsilon \coloneqq L_\sigma \sqrt{\frac{\log(8/\eta)}{n \log \mathrm{e}}}.$$

This choice of $\varepsilon$ is designed so that the second term on the right-hand side of (14) is then $\eta/2$. This leads us to the inequality

$$\eta \leqslant \mathbf{Pr}[\hat{\mathbf{c}} = \mathbf{c}] \leqslant 2^{-n[R(C)-\chi(\sigma_0,\sigma_1)-2\varepsilon]} + \frac{\eta}{2}.$$

Taking logarithms and rearranging gives

$$\begin{aligned} R(C) &\leqslant \chi(\sigma_0,\sigma_1) + 2\varepsilon + \frac{1}{n}\log\frac{2}{\eta} \\ &= \chi(\sigma_0,\sigma_1) + 2L_\sigma\sqrt{\frac{\log(8/\eta)}{n\log \mathrm{e}}} + \frac{1}{n}\log\frac{2}{\eta} \\ &= \chi(\sigma_0,\sigma_1) + O(n^{-1/2}). \end{aligned}$$

Maximizing over all choices of $C$ and letting $n \to \infty$, we conclude $R_2(\delta) \leqslant \chi(\sigma_0,\sigma_1)$. □

# 4 Rederiving known rate bounds

Recall from (1) that $R_2(\delta) \coloneqq \limsup_{n\to\infty} n^{-1}\log A_2(n, \lceil \delta n \rceil)$ is the asymptotic rate–distance function for binary codes. This section applies Theorem 1 to rederive, in order, the Plotkin [Plo60], Elias–Bassalygo [Eli55, Bas65], first MRRW, and second MRRW bounds [MRRW77]. To help bridge the transition from the classical to the quantum setup for the reader, we will present the calculations of the classical channels in Sections 4.1 and 4.2 in the quantum formalism. When deriving the Holevo information for each channel (excluding BEC), we will recast our qubits in terms of Pauli matrices $X$ and $Z$, defined at (20).

The four subsections follow the same template. We first give the two output states $\sigma_0$ and $\sigma_1$, their average state $\bar{\sigma} \coloneqq (\sigma_0 + \sigma_1)/2$, and the unitary $U$ that witnesses the output-symmetry of their corresponding binary-input cq channel $\Phi_\sigma$. We then compute the PGM bit error rate $p_{\mathrm{e}}(\sigma_0,\sigma_1)$ using the tools presented in Section 2.3, followed by the computation of the Holevo information $\chi(\sigma_0,\sigma_1)$ using Eq. (13) and the tools in Section 2.4. We then apply the pretty good criterion (Theorem 1) to conclude the rate–distance tradeoff obtained from $\Phi_\sigma$.

## 4.1 Binary Erasure Channel (BEC)

We begin with the Binary Erasure Channel (BEC), whose diagonal output states rederive the Plotkin bound (Corollary 19). Recall that BEC transmits the input bit $|b\rangle$ with probability $1-p$ and otherwise outputs the erasure symbol $|?\rangle$ that is orthogonal to the output bits $\{|0\rangle, |1\rangle\}$. On the span of $|0\rangle, |1\rangle, |?\rangle$, its output states are, for $b \in \{0,1\}$,

$$\sigma_b^{\mathrm{BEC}} \coloneqq (1-p)\,|b\rangle\langle b| + p\,|?\rangle\langle ?|\,.$$

The unitary $U^{\mathrm{BEC}} \coloneqq |0\rangle\langle 1| + |1\rangle\langle 0| + |?\rangle\langle ?|$ exchanges the two output states, establishing output-symmetry. Moreover, the resulting average state is

$$\bar{\sigma}^{\mathrm{BEC}} = \frac{1}{2}(1-p)(|0\rangle\langle 0| + |1\rangle\langle 1|) + p\,|?\rangle\langle ?|\,.$$

### 4.1.1 PGM bit error rate

Next, we compute the PGM bit error rate, which we abbreviate as $p_{\mathrm{e}}^{\mathrm{BEC}}(p) \coloneqq p_{\mathrm{e}}(\sigma_0^{\mathrm{BEC}}(p), \sigma_1^{\mathrm{BEC}}(p))$. We specialize the binary PGM formula (Proposition 12) to the uniform prior to derive the PGM bit error rate. To compute the trace in Eq. (15), we first multiply each output state by the inverse square root of the average state. This gives us

$$\sigma_0^{\mathrm{BEC}}(\bar{\sigma}^{\mathrm{BEC}})^{-1/2} = \sqrt{2(1-p)}\,|0\rangle\langle 0| + \sqrt{p}\,|?\rangle\langle ?|\,,$$
$$\sigma_1^{\mathrm{BEC}}(\bar{\sigma}^{\mathrm{BEC}})^{-1/2} = \sqrt{2(1-p)}\,|1\rangle\langle 1| + \sqrt{p}\,|?\rangle\langle ?|\,.$$

Plugging these two quantities into (15), we find that

$$p_{\mathrm{e}}^{\mathrm{BEC}}(p) = \frac{1}{2}\operatorname{Tr}(\sigma_0^{\mathrm{BEC}}(\bar{\sigma}^{\mathrm{BEC}})^{-1/2}\sigma_1^{\mathrm{BEC}}(\bar{\sigma}^{\mathrm{BEC}})^{-1/2}) = \frac{1}{2}\operatorname{Tr}(p\,|?\rangle\langle ?|) = \frac{p}{2}.$$

Thus, $p_{\mathrm{e}}^{\mathrm{BEC}}(p) = p/2$.

*Remark* 4.1. The formula $p_{\mathrm{e}}^{\mathrm{BEC}}(p) = p/2$ can, of course, be derived directly from classical considerations. We invoke Proposition 12 here primarily to illustrate its application and provide a useful perspective for the reader. ◁

### 4.1.2 Holevo information

Next, we compute the Holevo information, which we abbreviate as $\chi_{\mathrm{BEC}}(p) \coloneqq \chi(\sigma_0^{\mathrm{BEC}}, \sigma_1^{\mathrm{BEC}})$. Since each output state $\sigma_b^{\mathrm{BEC}}$ has spectrum $\{1-p, p, 0\}$, we see that

$$S(\sigma_0^{\mathrm{BEC}}) = S(\sigma_1^{\mathrm{BEC}}) = h(p).$$

In addition, the average state $\bar{\sigma}^{\mathrm{BEC}}$ has spectrum $\{(1-p)/2, (1-p)/2, p\}$. Its entropy is

$$S(\bar{\sigma}^{\mathrm{BEC}}) = -p\log p - 2\left(\frac{1-p}{2}\right)\log\left(\frac{1-p}{2}\right) = h(p) + 1 - p.$$

Applying the Holevo identity (Eq. (13)) now gives

$$\chi_{\mathrm{BEC}}(p) = S(\bar{\sigma}^{\mathrm{BEC}}) - \frac{1}{2}\left[S(\sigma_0^{\mathrm{BEC}}) + S(\sigma_1^{\mathrm{BEC}})\right] = (h(p) + 1 - p) - \frac{1}{2}\left[h(p) + h(p)\right] = 1 - p.$$

Thus, the Holevo information is exactly the probability that the input survives the erasure.

### 4.1.3 Rate upper bound

Since $p_{\mathrm{e}}^{\mathrm{BEC}}(p) = p/2$, the pretty good criterion (Theorem 1) applies when $p < 2\delta$. Letting $p \to (2\delta)^-$ therefore leads to the Plotkin bound.

**Corollary 19** (Plotkin bound — [Plo60, pp. 445–450])**.** *For every $\delta \in (0, \frac{1}{2})$, applying the pretty good criterion (Theorem 1) to the BEC yields*

$$R_2(\delta) \leqslant 1 - 2\delta.$$

## 4.2 Binary Symmetric Channel (BSC)

We next consider the Binary Symmetric Channel (BSC), whose diagonal output states rederive the Elias–Bassalygo bound (Corollary 20). Recall that the BSC flips the input bit $b$ with probability $p$. In the Pauli notation introduced at (20), we can cast the output states, for $b \in \{0, 1\}$, as

$$\sigma_b^{\mathrm{BSC}} := \frac{1}{2}\left(I + (-1)^b(1-2p)Z\right). \tag{24}$$

Note that both states are diagonal in the computational basis, hence classical. Since $XZX = -Z$, the unitary $U^{\mathrm{BSC}} := X$ exchanges the outputs, establishing output-symmetry. Moreover, the average state is

$$\bar{\sigma}^{\mathrm{BSC}} = \frac{1}{2}\left(\sigma_0^{\mathrm{BSC}} + \sigma_1^{\mathrm{BSC}}\right) = \frac{I}{2}.$$

Note that this Pauli presentation will be reused for the mixed-qubit channel (MQC) in Section 5.1.

### 4.2.1 PGM bit error rate

Next, we compute the PGM bit error rate, which we abbreviate as $p_{\mathrm{e}}^{\mathrm{BSC}}(p) := p_{\mathrm{e}}(\sigma_0^{\mathrm{BSC}}, \sigma_1^{\mathrm{BSC}})$. Since $\bar{\sigma}^{\mathrm{BSC}} = I/2$, we see that $\left(\bar{\sigma}^{\mathrm{BSC}}\right)^{-1/2} = \sqrt{2}I$. First, note the simple identities $\mathrm{Tr}(Z) = 0$ and $Z^2 = I$; by (15) from Proposition 12 and (24), we find that

$$\begin{aligned} p_{\mathrm{e}}^{\mathrm{BSC}}(p) &= \mathrm{Tr}\left(\sigma_0^{\mathrm{BSC}}\sigma_1^{\mathrm{BSC}}\right) = \frac{1}{4}\mathrm{Tr}\left[\left(I + (1-2p)Z\right)\left(I - (1-2p)Z\right)\right] \\ &= \frac{1}{4}\mathrm{Tr}\left[\left(1 - (1-2p)^2\right)I\right] = \frac{1-(1-2p)^2}{2} = 2p(1-p). \end{aligned}$$

*Remark* 4.2 (BSC posterior sampling). A more elegant way to derive the formula $p_{\mathrm{e}}^{\mathrm{BSC}}(p) = 2p(1-p)$ is to observe that posterior sampling—the classical specialization of the PGM—for $\mathrm{BSC}(p)$ is again $\mathrm{BSC}(p)$. Since $\mathbf{Pr}[\mathbf{e}_1 \oplus \mathbf{e}_2 = 1] = 2p(1-p)$ for independent $\mathbf{e}_1, \mathbf{e}_2 \sim \mathrm{Ber}(p)$, the PGM bit error rate formula follows. ◁

### 4.2.2 Holevo information

Next, we compute the Holevo information, which we abbreviate as $\chi_{\mathrm{BSC}}(p) := \chi(\sigma_0^{\mathrm{BSC}}, \sigma_1^{\mathrm{BSC}})$. Each output state has spectrum $\{1-p, p\}$, so

$$S(\sigma_0^{\mathrm{BSC}}) = S(\sigma_1^{\mathrm{BSC}}) = h(p).$$

As for the average state, the spectrum is $\{\frac{1}{2}, \frac{1}{2}\}$, and hence

$$S(\bar{\sigma}^{\mathrm{BSC}}) = 1.$$

Applying the Holevo identity (Eq. (13)) now gives

$$\begin{aligned} \chi_{\mathrm{BSC}}(p) &= S(\bar{\sigma}^{\mathrm{BSC}}) - \frac{1}{2}\left[S(\sigma_0^{\mathrm{BSC}}) + S(\sigma_1^{\mathrm{BSC}})\right] \\ &= 1 - \frac{1}{2}\left[h(p) + h(p)\right] \\ &= 1 - h(p). \end{aligned}$$

Thus, the Holevo information is the usual mutual information of the BSC.

### 4.2.3 Rate upper bound

Since $p_e^{\mathrm{BSC}}(p) = 2p(1-p)$, the pretty good criterion (Theorem 1) applies when $2p(1-p) < \delta$. This condition is met when $p < J(\delta) := (1-\sqrt{1-2\delta})/2$ (i.e., below the Johnson radius). Letting $p \to J(\delta)^-$ leads to the Elias–Bassalygo bound.

**Corollary 20** (Elias–Bassalygo bound for binary codes — [Eli55, Appendix A, Equations (A.15)–(A.18)];[Bas65, Equations (5) and (8)]). *For every $\delta \in (0, \frac{1}{2})$, applying the pretty good criterion (Theorem 1) to the BSC yields*

$$R_2(\delta) \leqslant R_{\mathrm{EB}}(\delta) := 1 - h\left(\frac{1-\sqrt{1-2\delta}}{2}\right).$$

## 4.3 Pure-State Channel (PSC)

We next consider the Pure-State Channel (PSC), whose pure qubit outputs rederive the first MRRW bound (Corollary 21). For a parameter $p \in [0, \frac{1}{2}]$, the PSC($p$) maps a bit $b \in \{0,1\}$ to one of two coherent Bernoulli states. Here, we define $|\mathrm{Ber}(p)\rangle := \sqrt{1-p}\,|0\rangle + \sqrt{p}\,|1\rangle$ to be the coherent Bernoulli state with parameter $p$. The output states of PSC($p$) are, for $b \in \{0,1\}$,

$$\sigma_b^{\mathrm{PSC}} := X^b\,|\mathrm{Ber}(p)\rangle\langle\mathrm{Ber}(p)|\,X^b.$$

Since $X\,|\mathrm{Ber}(p)\rangle = |\mathrm{Ber}(1-p)\rangle$, the two outputs are $|\mathrm{Ber}(p)\rangle\langle\mathrm{Ber}(p)|$ and $|\mathrm{Ber}(1-p)\rangle\langle\mathrm{Ber}(1-p)|$, which is the promised amplitude swap. In the Pauli notation fixed in Section 2.4, the output states can be recast as

$$\sigma_b^{\mathrm{PSC}} = \frac{1}{2}\left(I + 2\sqrt{p(1-p)}X + (-1)^b(1-2p)Z\right). \tag{25}$$

As a sanity check, a state $\varrho$ is pure if and only if $\mathrm{Tr}(\varrho^2) = 1$. Since $\mathrm{Tr}(X) = \mathrm{Tr}(Z) = 0$ and $\frac{1}{2}(1 + (2\sqrt{p(1-p)})^2 + (1-2p)^2) = 1$, this therefore confirms the purity of the output states. Since $XZX = -Z$, the unitary $U^{\mathrm{PSC}} := X$ exchanges the outputs, establishing output-symmetry. Moreover, their average state is

$$\bar{\sigma}^{\mathrm{PSC}} = \frac{1}{2}\left(I + 2\sqrt{p(1-p)}X\right). \tag{26}$$

### 4.3.1 PGM bit error rate

Next, we compute the PGM bit error rate, which we abbreviate as $p_e^{\mathrm{PSC}}(p) := p_e(\sigma_0^{\mathrm{PSC}}, \sigma_1^{\mathrm{PSC}})$. First, observe that the output states $\sigma_b^{\mathrm{PSC}}$ for $b \in \{0,1\}$ are pure states with fidelity

$$\left|\langle\mathrm{Ber}(p)|\,\mathrm{Ber}(1-p)\rangle\right|^2 = 4p(1-p).$$

Applying Proposition 12, specifically (16), with the uniform prior $\alpha = \frac{1}{2}$ gives us

$$p_e^{\mathrm{PSC}}(p) = \frac{2p(1-p)}{1+\sqrt{1-4p(1-p)}} = p, \tag{27}$$

where the last equality uses $\sqrt{1-4p(1-p)} = 1-2p$ for $p \in [0, \frac{1}{2}]$.

*Remark* 4.3 (Computational-basis PGM for the PSC). For $p \in [0, \frac{1}{2})$, the bit-level PGM for the PSC is the computational-basis measurement $\{|0\rangle\langle 0|, |1\rangle\langle 1|\}$. This measurement is additionally optimal (see Fact 28), so the value $p_e^{\mathrm{PSC}}(p) = p$ computed above is the minimum possible bit error rate. ◁

### 4.3.2 Holevo information

Next, we compute the Holevo information, which we abbreviate as $\chi_{\mathrm{PSC}}(p) \coloneqq \chi(\sigma_0^{\mathrm{PSC}}, \sigma_1^{\mathrm{PSC}})$. Since the output states $\sigma_b^{\mathrm{PSC}}$ are pure, we immediately see that[22]

$$S(\sigma_0^{\mathrm{PSC}}) = S(\sigma_1^{\mathrm{PSC}}) = 0.$$

As for the average state $\bar{\sigma}^{\mathrm{PSC}}$, applying Proposition 14 to the Pauli form in (26) gives

$$S(\bar{\sigma}^{\mathrm{PSC}}) = s\left(1 - 4p(1-p)\right) = s\left((1-2p)^2\right) = h\left(\frac{1}{2} - \sqrt{p(1-p)}\right).$$

Applying the Holevo identity (Eq. (13)) now gives

$$\begin{aligned}\chi_{\mathrm{PSC}}(p) &= S(\bar{\sigma}^{\mathrm{PSC}}) - \frac{1}{2}\left[S(\sigma_0^{\mathrm{PSC}}) + S(\sigma_1^{\mathrm{PSC}})\right] \\ &= h\left(\frac{1}{2} - \sqrt{p(1-p)}\right).\end{aligned}$$

### 4.3.3 Rate upper bound

Since $p_{\mathrm{e}}^{\mathrm{PSC}}(p) = p$, the pretty good criterion (Theorem 1) applies whenever $p < \delta$. Applying Theorem 1, we get $R_2(\delta) \leqslant \chi_{\mathrm{PSC}}(p) = h\left(\frac{1}{2} - \sqrt{p(1-p)}\right)$ for all $p < \delta$. Letting $p \to \delta^-$ leads to the first MRRW bound.

**Corollary 21** (First MRRW bound for binary codes — [MRRW77, Equation (1.5)])**.** *For every $\delta \in (0, \frac{1}{2})$, applying the pretty good criterion (Theorem 1) to the PSC gives*

$$R_2(\delta) \leqslant R_{\mathrm{MRRW}}(\delta) = h\left(\frac{1}{2} - \sqrt{\delta(1-\delta)}\right).$$

## 4.4 Masked PSC (mPSC)

We conclude this section with the masked PSC, whose four-dimensional outputs rederive the second MRRW bound (Corollary 22). Fix $r, \alpha \in [0, \frac{1}{2}]$. Recalling the coherent Bernoulli state $|\mathrm{Ber}(r)\rangle = \sqrt{1-r}\,|0\rangle + \sqrt{r}\,|1\rangle$ from Section 4.3, write $\tau_m(r) \coloneqq X^m\,|\mathrm{Ber}(r)\rangle\langle\mathrm{Ber}(r)|\,X^m$ for $m \in \{0,1\}$, and let $\bar{\tau}_\alpha(r) \coloneqq (1-\alpha)\tau_0(r) + \alpha\tau_1(r)$ be their average state. Following the $\alpha$-masking definition (Definition 2.9), the two output states are

$$\begin{aligned}\sigma_0^{\mathrm{mPSC}}(r,\alpha) &\coloneqq (1-\alpha)\tau_0(r) \oplus \alpha\tau_1(r), \\ \sigma_1^{\mathrm{mPSC}}(r,\alpha) &\coloneqq \alpha\tau_1(r) \oplus (1-\alpha)\tau_0(r).\end{aligned}$$

The block-swap unitary $U^{\mathrm{mPSC}}(\xi \oplus \zeta) \coloneqq \zeta \oplus \xi$ exchanges the outputs, establishing output-symmetry. Moreover, their average state is

$$\bar{\sigma}^{\mathrm{mPSC}} = \frac{1}{2}\left(\sigma_0^{\mathrm{mPSC}} + \sigma_1^{\mathrm{mPSC}}\right) = \frac{1}{2}\left(\bar{\tau}_\alpha(r) \oplus \bar{\tau}_\alpha(r)\right).$$

[22] This can also be seen by applying Proposition 14 along with the Pauli form given in Eq. (25).

*Remark* 4.4 (mPSC includes both PSC and BSC). Since the second MRRW bound was initially conceived as interpolating between the first MRRW bound and the Elias–Bassalygo bound, one might wonder if that can be seen directly from the masked PSC construction. Indeed, when $\alpha = \frac{1}{2}$, since masking with a uniform prior restores the original channel (Remark 2.2), we see that $\mathrm{mPSC}(r, \frac{1}{2})$ is equivalent to $\mathrm{PSC}(r)$ and therefore recovers the first MRRW bound (Corollary 21). When $r = \frac{1}{2}$, one has $\tau_m(\frac{1}{2}) = \tau_{m\oplus 1}(\frac{1}{2})$, so the only information about the mask $m$ is its prior $\mathrm{Ber}(\alpha)$. Thus, $\mathrm{mPSC}(\frac{1}{2}, \alpha)$ is equivalent to $\mathrm{BSC}(\alpha)$ and recovers the Elias–Bassalygo bound (Corollary 20). ◁

### 4.4.1 PGM bit error rate

Next, we compute the PGM bit error rate, which we abbreviate in this subsection as $p_{\mathrm{e}}^{\mathrm{mPSC}}(r, \alpha) \coloneqq p_{\mathrm{e}}(\sigma_0^{\mathrm{mPSC}}(r, \alpha), \sigma_1^{\mathrm{mPSC}}(r, \alpha))$. By the PGM error masking identity (Proposition 15), the mPSC error is the PGM error of the non-uniform pure-state ensemble $\{(1-\alpha, \tau_0(r)), (\alpha, \tau_1(r))\}$. Its squared overlap is

$$|\langle \mathrm{Ber}(r)| \mathrm{Ber}(1-r)\rangle|^2 = 4r(1-r),$$

and because $r \in [0, \frac{1}{2}]$, we have that $\sqrt{1-4r(1-r)} = 1-2r$. Substituting these expressions into the non-uniform pure-state formula (16), as recorded in Proposition 12, gives

$$p_{\mathrm{e}}^{\mathrm{mPSC}}(r, \alpha) \coloneqq p_{\mathrm{e}}\left(\sigma_0^{\mathrm{mPSC}}, \sigma_1^{\mathrm{mPSC}}\right) = \frac{8\alpha(1-\alpha)r(1-r)}{1 + 2(1-2r)\sqrt{\alpha(1-\alpha)}}. \tag{28}$$

### 4.4.2 Holevo information

Next, we compute the Holevo information, which we abbreviate as $\chi_{\mathrm{mPSC}}(r, \alpha) \coloneqq \chi(\sigma_0^{\mathrm{mPSC}}, \sigma_1^{\mathrm{mPSC}})$. Both states in the underlying PSC pair are pure, so

$$S(\tau_0(r)) = S(\tau_1(r)) = 0.$$

Their average state has the Pauli form

$$\bar{\tau}_\alpha(r) = \frac{1}{2}\left(I + 2\sqrt{r(1-r)}X + (1-2\alpha)(1-2r)Z\right).$$

By Proposition 14,

$$\begin{aligned} S(\bar{\tau}_\alpha(r)) &= s\left(1 - 4r(1-r) - (1-2\alpha)^2(1-2r)^2\right) \\ &= s\left(4\alpha(1-\alpha)(1-2r)^2\right). \end{aligned}$$

The masked Holevo identity (Proposition 16) therefore gives

$$\begin{aligned} \chi_{\mathrm{mPSC}}(r, \alpha) &\coloneqq \chi\left(\sigma_0^{\mathrm{mPSC}}, \sigma_1^{\mathrm{mPSC}}\right) \\ &= 1 - h(\alpha) + s\left(4\alpha(1-\alpha)(1-2r)^2\right) \\ &= 1 + s\left(4\alpha(1-\alpha)(1-2r)^2\right) - s\left(4\alpha(1-\alpha)\right). \end{aligned} \tag{29}$$

Here the last step uses $h(\alpha) = s\left(4\alpha(1-\alpha)\right)$.

### 4.4.3 Rate upper bound

To see how the resulting rate bound from mPSC attains exactly the second MRRW in its original form [MRRW77, Equation (1.4)],[23] we introduce the shorthands

$$u \coloneqq 2(1-2r)\sqrt{\alpha(1-\alpha)}, \qquad p \coloneqq \frac{8\alpha(1-\alpha)r(1-r)}{1+2(1-2r)\sqrt{\alpha(1-\alpha)}}.$$

Then (29) can be rewritten as

$$\chi_{\mathrm{mPSC}}(r,\alpha) = 1 + s(u^2) - s\left(u^2 + 2pu + 2p\right),$$

which is close in form to the bound presented in [MRRW77, Equation (1.4)]. Moreover, by (28), we see that in fact

$$p_{\mathrm{e}}^{\mathrm{mPSC}}(r,\alpha) = p. \tag{30}$$

Now, via our shorthands, we can solve for $\alpha$ and $r$ in terms of $u$ and $p$. This gives us, picking the solutions in the range $[0, 1/2]$,

$$(r,\alpha) = \left(\frac{1}{2} - \frac{u}{2\sqrt{u^2+2p(1+u)}}, \frac{1-\sqrt{(1+u)(1-2p-u)}}{2}\right),$$

assuming $u \in [0, 1-2p]$ (and $p \geqslant 0$). Now, by (30), the pretty good criterion (Theorem 1) applies whenever $p < \delta$. Therefore, letting $p \longrightarrow \delta^-$ and minimizing over $u$ yields the second MRRW bound.

**Corollary 22** (Second MRRW bound for binary codes — [MRRW77, Equation (1.4)])**.** *For every $\delta \in (0, \frac{1}{2})$, applying the pretty good criterion (Theorem 1) to the masked PSC gives*

$$R_2(\delta) \leqslant R^{(2)}_{\mathrm{MRRW}}(\delta) \coloneqq \min_{0 \leqslant u \leqslant 1-2\delta} \left[1 + s(u^2) - s\left(u^2 + 2\delta u + 2\delta\right)\right]. \tag{31}$$

The endpoints of this family recover the BSC and PSC bounds already encountered.

*Remark* 4.5 (Qubit cq channels). Naturally, one might wonder if the second MRRW bound could instead be captured by a cq channel with qubit outputs as opposed to a 4-dimensional one. As it turns out, by Remark 5.2, the mixed-qubit channel (MQC) is the binary-input output-symmetric cq channel with qubit outputs that yields the best rate bound among all such cq channels. Since it does not outperform the *second* MRRW bound throughout the entire interval, any cq channel corresponding to the second MRRW bound necessarily has to go to higher dimensional qudits. ◁

## 5 New classical–quantum (cq) channel constructions

We now turn to channel designs introduced in this work. The first one is the mixed-qubit channel (MQC), which we will later show has a resulting rate upper bound that lies strictly below the first MRRW bound (Theorem 2). The second construction, the masked MQC (2MQC), follows a similar approach to the derivation of the second MRRW bound but with MQC in place of PSC. As a result, its ultimate rate upper bound will lie strictly below the second MRRW bound (Theorem 3).

[23] Note that the function $s$ in this work coincides with the function $g$ in the original work [MRRW77]. In this work, the function $s$ plays a similar role to von Neumann entropy $S$ as the binary entropy function $h$ does to Shannon entropy $H$.

### 5.1 Mixed-qubit channel (MQC)

The mixed-qubit channel (MQC) is defined to be the channel that is obtained by passing the output of PSC($r$) through the X-Pauli channel $\mathcal{B}_\eta$ from (12). Fix $r, \eta \in [0, \frac{1}{2}]$. Since $X\sigma_b^{\mathrm{PSC}}(r)X = \sigma_{b\oplus 1}^{\mathrm{PSC}}(r)$, following the description given in Section 1.3, its output states are, for $b \in \{0,1\}$,

$$\sigma_b^{\mathrm{MQC}}(r,\eta) \coloneqq \mathcal{B}_\eta\left(\sigma_b^{\mathrm{PSC}}(r)\right) = \frac{1}{2}\left(I + 2\sqrt{r(1-r)}X + (-1)^b(1-2r)(1-2\eta)Z\right). \tag{32}$$

Put differently, each output is the mixture $(1-\eta)\sigma_b^{\mathrm{PSC}}(r) + \eta\sigma_{b\oplus 1}^{\mathrm{PSC}}(r)$. The unitary $U^{\mathrm{PSC}} \coloneqq X$ exchanges the outputs, establishing output-symmetry. Moreover, their average state is

$$\bar{\sigma}^{\mathrm{MQC}}(r,\eta) = \frac{1}{2}I + \sqrt{r(1-r)}X. \tag{33}$$

*Remark* 5.1 (MQC includes both PSC and BSC). Almost by construction, the BSC and PSC are special cases of the MQC. Indeed, at $\eta = 0$, the $X$-Pauli noise is no longer there, and so MQC$(r,0)$ is exactly PSC$(r)$. Moreover, at $r = 0$, the underlying PSC outputs are the orthogonal computational-basis states, so MQC$(0,\eta)$ is exactly BSC$(\eta)$ (cf. Footnote 9). ◁

*Remark* 5.2 (The MQC encompasses all output-symmetric qubit states). One can in fact observe that, up to symmetry, the MQC encompasses all the possible output-symmetric qubit states. Indeed, write the Bloch vectors of the two output states as $v_0, v_1 \in \mathbb{R}^3$. From this Bloch sphere lens, we see that output-symmetry is equivalent to $\|v_0\|_2 = \|v_1\|_2$. This is moreover equivalent to the orthogonality between their midpoint $m \coloneqq (v_0+v_1)/2$ and half-difference $d \coloneqq (v_0 - v_1)/2$. Since every rotation in SO(3) on the Bloch sphere is realized by a unitary $U \in \mathrm{SU}(2)$ applied to the qubit ($\varrho \mapsto U\varrho U^\dagger$), by orthogonally mapping $m \mapsto (\|m\|_2, 0, 0)$ and $d \mapsto (0, 0, \|d\|_2)$, we can therefore rotate the two output states to $v_b = (x, 0, (-1)^b z)$ with $x, z \geqslant 0$ and $x^2 + z^2 \leqslant 1$. Set $r \coloneqq (1 - \sqrt{1-x^2})/2$ and $\eta \coloneqq (1 - z/\sqrt{1-x^2})/2$. Then $2\sqrt{r(1-r)} = x$ and $(1-2r)(1-2\eta) = z$, which is exactly the MQC pair in (32). ◁

#### 5.1.1 PGM bit error rate

Next, we compute the PGM bit error rate, which we abbreviate in this subsection as $p_{\mathrm{e}}^{\mathrm{MQC}}(r,\eta) \coloneqq p_{\mathrm{e}}(\sigma_0^{\mathrm{MQC}}(r,\eta), \sigma_1^{\mathrm{MQC}}(r,\eta))$. From the definition of MQC in (32), observe that

$$\Phi^{\mathrm{MQC}} = \mathcal{B}_\eta \circ \Phi^{\mathrm{PSC}} = \Phi^{\mathrm{PSC}} \circ \Phi^{\mathrm{BSC}}.$$

Indeed, the commutativity follows from observing that both $\mathcal{B}_\eta$ and $\Phi^{\mathrm{PSC}}$ are just applications of the Pauli $X$ (cf. Footnote 9). Thus, applying (19) from Proposition 13 gives us that

$$p_{\mathrm{e}}^{\mathrm{MQC}}(r,\eta) = \frac{1}{2}\left[1 - (1-2r)(1-2\eta)^2\right]. \tag{34}$$

Equivalently, the MQC PGM first applies the PSC PGM and then adds an independent posterior Ber($\eta$) flip (Remark 4.2).[24] Combined with the prior Ber($\eta$) flip, it gives the PGM bit error rate above.

[24] This is perhaps best observed via the Petz contravariant functoriality (Fact 32(ii)), concretely packaged in Proposition 13.

### 5.1.2 Holevo information

Next, we compute the Holevo information, which we abbreviate in this subsection as $\chi_{\mathrm{MQC}}(r,\eta) \coloneqq \chi(\sigma_0^{\mathrm{MQC}}(r,\eta), \sigma_1^{\mathrm{MQC}}(r,\eta))$. Applying Proposition 14 to the output-state Pauli forms in (32) gives

$$\begin{aligned} S\left(\sigma_b^{\mathrm{MQC}}(r,\eta)\right) &= s\left(1 - 4r(1-r) - (1-2r)^2(1-2\eta)^2\right) \\ &= s\left(4\eta(1-\eta)(1-2r)^2\right) \end{aligned} \tag{35}$$

Similarly, applying Proposition 14 to the average-state Pauli form in (33) gives

$$S\left(\bar{\sigma}^{\mathrm{MQC}}(r,\eta)\right) = s\left(1 - 4r(1-r)\right) = s\left((1-2r)^2\right).$$

Plugging the two values into the Holevo identity (Eq. (13)) therefore gives

$$\chi_{\mathrm{MQC}}(r,\eta) = s\left((1-2r)^2\right) - s\left(4\eta(1-\eta)(1-2r)^2\right). \tag{36}$$

*Remark* 5.3 (Alternative derivation via mutual information). Let us give an alternative and more conceptual derivation of Eq. (36). Let $\mathbf{b} \sim \mathrm{Ber}(\frac{1}{2})$ and $\mathbf{e} \sim \mathrm{Ber}(\eta)$ be independent random bits. Set $\mathbf{y} \coloneqq \mathbf{b} \oplus \mathbf{e} \sim \mathrm{Ber}(\frac{1}{2})$. Since $\mathbf{b} \longmapsto \mathbf{b} \oplus \mathbf{e} \longmapsto \sigma_{\mathbf{b}\oplus\mathbf{e}}^{\mathrm{PSC}}$ is a Markov chain, using the mutual information chain rule, the identity $\sigma_{\mathbf{b}\oplus\mathbf{e}}^{\mathrm{PSC}} = X^{\mathbf{b}}\sigma_{\mathbf{e}}^{\mathrm{PSC}}X^{\mathbf{b}}$, the observation in Footnote 9, and the Holevo identity (Eq. (13)), notice that

$$\begin{aligned} I(\mathbf{y};\sigma_{\mathbf{y}}^{\mathrm{PSC}}) = I(\mathbf{b},\mathbf{e};\sigma_{\mathbf{b}\oplus\mathbf{e}}^{\mathrm{PSC}}) &= I(\mathbf{b};\sigma_{\mathbf{b}\oplus\mathbf{e}}^{\mathrm{PSC}}) + I(\mathbf{e};\sigma_{\mathbf{b}\oplus\mathbf{e}}^{\mathrm{PSC}} \mid \mathbf{b}) \\ &= \chi_{\mathrm{MQC}}(r,\eta) + I(\mathbf{e};\sigma_{\mathbf{e}}^{\mathrm{PSC}}). \end{aligned}$$

Thus, we find that

$$\chi_{\mathrm{MQC}}(r,\eta) = I(\mathbf{y};\sigma_{\mathbf{y}}^{\mathrm{PSC}}) - I(\mathbf{e};\sigma_{\mathbf{e}}^{\mathrm{PSC}}).$$

Because the bit $\mathbf{y}$ remains uniform for every $\eta$, the first term is always $I(\mathbf{y};\sigma_{\mathbf{y}}^{\mathrm{PSC}}) = s((1-2r)^2)$. Thus, the dependence on $\eta$ lies in the second term, which equals $I(\mathbf{e};\sigma_{\mathbf{e}}^{\mathrm{PSC}}) = s(4\eta(1-\eta)(1-2r)^2)$. Thus, mixing starts with the information carried by the underlying PSC and subtracts exactly the information that its output state $\sigma_{\mathbf{y}}$ retains about the discarded flip $\mathbf{e}$ once $\mathbf{b}$ is known. ◁

### 5.1.3 Rate upper bound

Based on Eq. (34), define the function

$$e(r,\eta) \coloneqq \frac{1}{2}\left[1 - (1-2r)(1-2\eta)^2\right]. \tag{37}$$

Then the pretty good criterion (Theorem 1) applies whenever $e(r,\eta) < \delta$. Now, fix any value $r \in [0,\delta]$. Notice that increasing $\eta$ along the interval $[0,\frac{1}{2}]$ decreases the Holevo information, which only sharpens the resulting rate upper bound. On the other hand, it also increases the PGM bit error rate, which is $e(r,\eta)$. Since $\delta$ is the supremum value of $e(r,\eta)$ under which the pretty good criterion (Theorem 1) still applies, then setting $\eta = \eta_\delta(r)$, where $\eta_\delta(r)$ is the unique solution to the equation $e(r,\eta_\delta(r)) = \delta$ that lies in the interval $[0,\frac{1}{2}]$, would achieve the best possible rate upper bound for MQC. Now, solving for $\eta_\delta(r)$ in the equation $e(r,\eta_\delta(r)) = \delta$, we find that

$$\eta_\delta(r) = \frac{1}{2} - \frac{1}{2}\cdot\sqrt{\frac{1-2\delta}{1-2r}}.$$

Thus, by Eq. (36), for any fixed $r \in [0, \delta]$, the lowest Holevo information for which the pretty good criterion (Theorem 1) still applies at this specified value $r$ is

$$\begin{aligned}\chi_{\mathrm{MQC}}(r, \eta_\delta(r)) &= s\left((1-2r)^2\right) - s\left(4\eta_\delta(r)(1-\eta_\delta(r))(1-2r)^2\right) \\ &= s\left((1-2r)^2\right) - s\left(2(\delta - r)(1-2r)\right).\end{aligned}$$

Minimizing over all values $r \in [0, \delta]$ gives us the following rate bound.

**Corollary 23** (MQC rate upper bound)**.** *For every $\delta \in (0, \frac{1}{2})$, applying the pretty good criterion (Theorem 1) to the MQC gives us the rate bound*

$$\begin{aligned}R_2(\delta) \leqslant R_{\mathrm{MQC}}(\delta) &:= \min_{0 \leqslant r \leqslant \delta} \left[s\left((1-2r)^2\right) - s\left(2(\delta - r)(1-2r)\right)\right] \\ &= \min_{1-2\delta \leqslant u \leqslant 1} \left[s(u^2) - s\left(u(u - (1-2\delta))\right)\right],\end{aligned} \tag{38}$$

*where the second line is the change of variables $u := 1 - 2r$.*

This is the rate upper bound stated in Theorem 2. Its strict improvement over the first MRRW bound is proved later in Proposition 25.

*Remark* 5.4 (Holevo information and PGM error data processing). As was noted in Section 1.3, as $\eta$ increases along the interval $[0, \frac{1}{2}]$, we expect the Holevo information to decay while the PGM error to grow. Indeed, by inspecting Eqs. (34) and (36), one can numerically confirm that this is indeed the case here. ◁

## 5.2 Masked mixed-qubit channel (2MQC)

We next apply the masking operation of Section 2.5 to the MQC pair. Fix $r, \eta, \alpha \in [0, \frac{1}{2}]$. Throughout, we will recurrently use the convenient shorthand

$$\theta := \alpha + \eta - 2\alpha\eta. \tag{39}$$

For $m \in \{0, 1\}$, write $\tau_m(r, \eta) := \sigma_m^{\mathrm{MQC}}(r, \eta)$, and let $\bar{\tau}_\alpha(r, \eta) := (1-\alpha)\tau_0(r, \eta) + \alpha\tau_1(r, \eta)$. Here, $r$ is the noise parameter of the underlying PSC, $\eta$ is the probability of the discarded inner $X$-flip, and $\alpha$ is the prior probability of the mask. The 2MQC output states are

$$\begin{aligned}\sigma_0^{\mathrm{2MQC}}(r, \eta, \alpha) &:= (1-\alpha)\tau_0(r, \eta) \oplus \alpha\tau_1(r, \eta), \\ \sigma_1^{\mathrm{2MQC}}(r, \eta, \alpha) &:= \alpha\tau_1(r, \eta) \oplus (1-\alpha)\tau_0(r, \eta).\end{aligned}$$

The block-swap unitary $U^{\mathrm{2MQC}}(\varrho_0 \oplus \varrho_1) := \varrho_1 \oplus \varrho_0$ exchanges the outputs, establishing output-symmetry. Moreover, their average state is

$$\bar{\sigma}^{\mathrm{2MQC}} = \frac{1}{2}\left(\sigma_0^{\mathrm{2MQC}} + \sigma_1^{\mathrm{2MQC}}\right) = \frac{1}{2}\left(\bar{\tau}_\alpha(r, \eta) \oplus \bar{\tau}_\alpha(r, \eta)\right).$$

*Remark* 5.5 (2MQC includes both mPSC and MQC). Almost by construction, the mPSC and MQC are both special cases of the 2MQC. Indeed, at $\eta = 0$, the $X$-Pauli noise is no longer there, and so $\mathrm{2MQC}(r, 0, \alpha)$ is exactly $\mathrm{mPSC}(r, \alpha)$. Moreover, when $\alpha = \frac{1}{2}$, since masking with a uniform prior restores the original channel (Remark 2.2), we see that $\mathrm{2MQC}(r, \eta, \frac{1}{2})$ is equivalent to $\mathrm{MQC}(r, \eta)$. ◁

### 5.2.1 PGM bit error rate

Next, we compute the PGM bit error rate, which we abbreviate in this subsection as $p_{\mathrm{e}}^{\mathrm{2MQC}}(r,\eta,\alpha) \coloneqq p_{\mathrm{e}}(\sigma_0^{\mathrm{2MQC}}(r,\eta,\alpha), \sigma_1^{\mathrm{2MQC}}(r,\eta,\alpha))$. Following the same initial reasoning as in Section 5.1.1, from the definition of MQC in (32), observe that

$$\Phi^{\mathrm{MQC}} = \mathcal{B}_\eta \circ \Phi^{\mathrm{PSC}} = \Phi^{\mathrm{PSC}} \circ \Phi^{\mathrm{BSC}}.$$

Indeed, the commutativity follows from observing that both $\mathcal{B}_\eta$ and $\Phi^{\mathrm{PSC}}$ are just applications of the Pauli $X$ (cf. Footnote 9). Thus, by the PGM error masking identity (Proposition 15) and Eq. (18) from Proposition 13, we find that

$$p_{\mathrm{e}}^{\mathrm{2MQC}}(r,\eta,\alpha) = p_{\mathrm{e},\alpha}(\tau_0(r,\eta),\tau_1(r,\eta)) = 2\alpha(1-\alpha) - \left[\frac{\alpha(1-\alpha)(1-2\eta)}{\theta(1-\theta)}\right]^2 \cdot \left[2\theta(1-\theta) - p_{\mathrm{e},\theta}(\sigma_0,\sigma_1)\right].$$

On a more conceptual level, this is the PGM error of a bit $\mathbf{b} \sim \mathrm{Ber}(\alpha)$ sent through the composition $\mathbf{b} \mapsto \mathbf{y} \coloneqq \mathbf{b} \oplus \mathbf{e} \mapsto \sigma_{\mathbf{y}}^{\mathrm{PSC}}(r)$, where $\mathbf{e} \sim \mathrm{Ber}(\eta)$ and hence $\mathbf{y} \sim \mathrm{Ber}(\theta)$. Now, by applying the non-uniform pure-state formula (16) to the intermediate PSC ensemble, we see that

$$p_{\mathrm{e},\theta}\left(\sigma_0^{\mathrm{PSC}}(r), \sigma_1^{\mathrm{PSC}}(r)\right) = \frac{8\theta(1-\theta)r(1-r)}{1+2(1-2r)\sqrt{\theta(1-\theta)}}.$$

Substituting this PGM error directly into the composition identity (18) of Proposition 13 gives

$$\begin{aligned}
p_{\mathrm{e}}^{\mathrm{2MQC}}(r,\eta,\alpha) &= 2\alpha(1-\alpha) - \left[\frac{\alpha(1-\alpha)(1-2\eta)}{\theta(1-\theta)}\right]^2 \cdot \left[2\theta(1-\theta) - \frac{8\theta(1-\theta)r(1-r)}{1+2(1-2r)\sqrt{\theta(1-\theta)}}\right] \\
&= 2\alpha(1-\alpha)\left[1 - \frac{\alpha(1-\alpha)(1-2\eta)^2}{\theta(1-\theta)} \cdot \left[1 - \frac{4r(1-r)}{1+2(1-2r)\sqrt{\theta(1-\theta)}}\right]\right] \\
&= 2\alpha(1-\alpha)\left[1 - \frac{\alpha(1-\alpha)(1-2\eta)^2}{\theta(1-\theta)} + \frac{4\alpha(1-\alpha)r(1-r)(1-2\eta)^2}{\theta(1-\theta)\left[1+2(1-2r)\sqrt{\theta(1-\theta)}\right]}\right] \\
&= 2\alpha(1-\alpha)\left[\frac{\eta(1-\eta)}{\theta(1-\theta)} + \frac{4\alpha(1-\alpha)r(1-r)(1-2\eta)^2}{\theta(1-\theta)\left[1+2(1-2r)\sqrt{\theta(1-\theta)}\right]}\right].
\end{aligned} \tag{40}$$

Here, we used the identity $4\theta(1-\theta) = 1-(2\alpha-1)^2(1-2\eta)^2$ at the final step, which holds by definition of $\theta$ in (39). Note that when $\alpha = \frac{1}{2}$, we have $\theta = \frac{1}{2}$, and so this recovers the formula given in Eq. (34), which is consistent with Remark 2.2.

### 5.2.2 Holevo information

Next, we compute the Holevo information, which we abbreviate in this subsection as $\chi_{\mathrm{2MQC}}(r,\eta,\alpha) \coloneqq \chi(\sigma_0^{\mathrm{2MQC}}(r,\eta,\alpha), \sigma_1^{\mathrm{2MQC}}(r,\eta,\alpha))$. Recall from Section 5.1.2, specifically (35), that

$$S(\tau_0(r,\eta)) = S(\tau_1(r,\eta)) = s\left(4\eta(1-\eta)(1-2r)^2\right).$$

Now, the average state admits the Pauli form

$$\bar{\tau}_\alpha(r,\eta) = \frac{1}{2}\left(I + 2\sqrt{r(1-r)}X + (1-2\theta)(1-2r)Z\right),$$

so Proposition 14 gives us

$$\begin{aligned} S(\bar{\tau}_\alpha(r,\eta)) &= s\left(1 - 4r(1-r) - (1-2\theta)^2(1-2r)^2\right) \\ &= s\left(4\theta(1-\theta)(1-2r)^2\right). \end{aligned}$$

Thus, by the masked Holevo information identity (Proposition 16), we now find that

$$\chi_{\text{2MQC}}(r,\eta,\alpha) = 1 - h(\alpha) + s\left(4\theta(1-\theta)(1-2r)^2\right) - s\left(4\eta(1-\eta)(1-2r)^2\right). \tag{41}$$

### 5.2.3 Rate upper bound

Based on Eq. (40), define the function

$$E(r,\eta,\alpha) \coloneqq 2\alpha(1-\alpha)\left[\frac{\eta(1-\eta)}{\theta(1-\theta)} + \frac{4\alpha(1-\alpha)r(1-r)(1-2\eta)^2}{\theta(1-\theta)\left[1 + 2(1-2r)\sqrt{\theta(1-\theta)}\right]}\right]. \tag{42}$$

Building on the shorthands for the mPSC in Section 4.4.3 and the MQC in Section 5.1.3, define

$$\begin{aligned} u &\coloneqq 2(1-2r)\sqrt{\theta(1-\theta)}, \\ v &\coloneqq 2(1-2r)\sqrt{\eta(1-\eta)}, \\ A &\coloneqq 4\alpha(1-\alpha). \end{aligned} \tag{43}$$

Here $u^2$ and $v^2$ are exactly the two arguments of $s$ in (41), while $h(\alpha) = s(A)$. Thus, Eqs. (40) and (41) become

$$E(u,v,A) = \frac{A}{2}\left[\frac{v^2}{u^2} + \frac{4Ar(1-r)(1-2r)^2(1-2\eta)^2}{u^2(1+u)}\right] = \frac{A\left[u^2(1-v^2) + v^2(u+v^2)\right] - (u^2-v^2)^2}{2u^2(1+u)},$$

and

$$\chi_{\text{2MQC}}(u,v,A) = 1 + s(u^2) - s(A) - s(v^2). \tag{44}$$

The pretty good criterion (Theorem 1) applies whenever $E(u,v,A) < \delta$. Fix a pair $(u,v)$ with $0 < u \leqslant 1$ and $0 \leqslant v < u$. At this fixed pair, the displayed formula for $E(u,v,A)$ is increasing and affine in $A$, whereas (44) decreases with $A$. As in the MQC calculation (Section 5.1.3), the lowest admissible Holevo information is therefore obtained at $A = \lambda_\delta(u,v)$, where $\lambda_\delta(u,v)$ is the unique solution to the equation $E(u,v,\lambda_\delta(u,v)) = \delta$. Solving for $\lambda_\delta(u,v)$ in this equation, we find that

$$\lambda_\delta(u,v) = \frac{(u^2-v^2)^2 + 2\delta u^2(1+u)}{u^2(1-v^2) + v^2(u+v^2)}. \tag{45}$$

Since we must have $A \in [0,1]$, we therefore see that the optimal value for $A$ must be

$$A_\delta(u,v) \coloneqq \min\left\{1, \frac{(u^2-v^2)^2 + 2\delta u^2(1+u)}{u^2(1-v^2) + v^2(u+v^2)}\right\}.$$

It remains to identify the admissible values of the triple $(u, v, A)$. We claim that a triple is admissible if and only if it satisfies the inequalities

$$0 \leqslant v \leqslant u \leqslant 1 \qquad \text{and} \qquad \frac{u^2 - v^2}{1 - v^2} \leqslant A \leqslant 1.$$

Indeed, both inequalities can be deduced by observing the identity $u^2 - v^2 = A(1-2r)^2(1-2\eta)^2$. As for the converse, it follows by setting $r, \eta, \alpha$ equal to

$$\begin{aligned} \alpha &= \frac{1 - \sqrt{1 - A}}{2}, \\ (1-2r)^2 &= v^2 + \frac{u^2 - v^2}{A}, \\ (1-2\eta)^2 &= 1 - \frac{v^2}{(1-2r)^2}. \end{aligned}$$

Finally, we inspect the values of $(u, v)$ for which there exists an $A$ satisfying $E(u, v, A) < \delta$. Now, we know that $A = (u^2 - v^2)/(1 - v^2)$ is the smallest possible value of $A$. In such a case, the PGM bit error rate equals

$$E\left(u, v, \frac{u^2 - v^2}{1 - v^2}\right) = \frac{v^2(u^2 - v^2)}{2u^2(1 - v^2)}.$$

Because $E(u, v, A)$ increases with $A$ (it is affine in $A$), the range over which there exists an $A$ satisfying $E(u, v, A) < \delta$ is specified precisely by the inequality

$$v^2(u^2 - v^2) < 2\delta u^2(1 - v^2).$$

Now, we take stock of everything and conclude our rate bound. By applying the pretty good criterion (Theorem 1) and a simple continuity argument, we find that

$$R_2(\delta) \leqslant 1 + s(u^2) - s\left(A_\delta(u, v)\right) - s(v^2).$$

Taking the infimum over all feasible pairs $(u, v)$ proves the following corollary.

**Corollary 24** (2MQC rate upper bound)**.** *For every $\delta \in (0, \frac{1}{2})$, applying the pretty good criterion (Theorem 1) to the 2MQC gives*

$$R_2(\delta) \leqslant R_{\text{2MQC}}(\delta) \coloneqq \inf_{\substack{0 \leqslant v \leqslant u \leqslant 1, \\ v^2(u^2 - v^2) < 2\delta u^2(1 - v^2)}} \left[1 + s(u^2) - s\left(A_\delta(u, v)\right) - s(v^2)\right], \tag{46}$$

*where we define* $A_\delta(u, v) \coloneqq \min\left\{1, \frac{(u^2 - v^2)^2 + 2\delta u^2(1+u)}{u^2(1-v^2) + v^2(u+v^2)}\right\}$.

## 6 Comparing the rate bounds

In Section 4, we recovered all four previously known rate upper bounds—Plotkin, Elias–Bassalygo, the first and second MRRW bounds. In Section 5, we produced two new rate upper bounds from the newly introduced cq channels MQC and 2MQC, which are stated in Corollaries 23 and 24. In this section, we analytically and numerically compare the rate bounds of MQC and 2MQC with the first and second MRRW bounds, respectively, and establish Theorems 2 and 3. For the reader's convenience, we recap the parameters and formulas of all six channels in Table 1 below.

Table 1: PGM bit error rates, Holevo information, and optimized rate bounds for the six channel families from Sections 4 and 5. Here, $h$ is the binary entropy function, $J(\delta) \coloneqq (1-\sqrt{1-2\delta})/2$ is the Johnson radius, while $s(D) \coloneqq h((1-\sqrt{1-D})/2)$ is the von Neumann entropy of a qubit with determinant $D/4$. The variables $u, v, A$ are functions of $r, \eta, \alpha$ and are specified at Sections 4.4.3, 5.1.3 and 5.2.3. Here, the functions $e$ and $E$ are defined at Eqs. (37) and (42), respectively.

| **Channel** | **PGM bit error rate** | **Uniform-input $\chi$** | **Upper bound on $R_2(\delta)$** |
|---|---|---|---|
| BEC($p$) | $p/2$ | $1-p$ | $1-2\delta$ |
| BSC($p$) | $2p(1-p)$ | $1-h(p)$ | $1-h(J(\delta))$ |
| PSC($p$) | $p$ | $h(\frac{1}{2}-\sqrt{p(1-p)})$ | $R_{\mathrm{MRRW}}(\delta)$ |
| mPSC($r,\alpha$) | $p$ | $1+s(u^2)-s(u^2+2pu+2p)$ | $R^{(2)}_{\mathrm{MRRW}}(\delta)$ |
| MQC($r,\eta$) | $e(r,\eta)$ | $s\left(u^2\right)-s\left(4\eta(1-\eta)u^2\right)$ | $R_{\mathrm{MQC}}(\delta)$ |
| 2MQC($r,\eta,\alpha$) | $E(u,v,A)$ | $1+s(u^2)-s(A)-s(v^2)$ | $R_{\mathrm{2MQC}}(\delta)$ |

### 6.1 MQC improvement over the first MRRW bound

In this subsection, we prove that slightly deviating away from PSC by considering a small $X$-Pauli noise in the MQC will make the MQC rate bound in Corollary 23 strictly smaller than the first MRRW bound. This is stated in Proposition 25. For ease of exposition, we adopt the shorthand

$$F_\delta(u) \coloneqq s(u^2) - s\left(u(u-(1-2\delta))\right), \tag{47}$$

for the second objective in (38), where $u$ ranges over $[1-2\delta, 1]$. Note that by Remark 5.1, we know that

$$R_{\mathrm{MQC}}(\delta) \leqslant \min\{R_{\mathrm{MRRW}}(\delta), R_{\mathrm{EB}}(\delta)\}.$$

Now, we proceed to show that this is indeed strict for the first MRRW bound.[25] For the reader's convenience, Fig. 1 provides a visual illustration of the argument used in the proof of the next proposition.

**Proposition 25** (Strict improvement over the first MRRW expression)**.** *For every $\delta \in (0, \frac{1}{2})$, we have the strict inequality*

$$R_{\mathrm{MQC}}(\delta) < R_{\mathrm{MRRW}}(\delta).$$

*Proof.* By Remark 5.1 and Eqs. (38) and (47), notice that $F_\delta(1-2\delta) = R_{\mathrm{MRRW}}(\delta)$. Now, fix $\delta$ and choose a perturbation $\zeta \to 0^+$ of $1-2\delta$. Write the difference

$$F_\delta(1-2\delta+\zeta) - F_\delta(1-2\delta) = \left[s((1-2\delta+\zeta)^2) - s((1-2\delta)^2)\right] - s\left((1-2\delta+\zeta)\zeta\right).$$

Now, for $D \to 0^+$, the definition of $s$ given at (21) gives

$$s(D) = \frac{D}{4}\log\frac{1}{D} + O(D).$$

Since $(1-2\delta)^2 \in (0,1)$, the mean value theorem thus gives us

$$s((1-2\delta+\zeta)^2) - s((1-2\delta)^2) = O(\zeta).$$

[25] One can similarly show that for the Elias–Bassalygo bound, the MQC bound strictly improves upon it, but only when $\delta \gtrsim 0.07441947$.

For the second term, which is where the primary gain in improved rate bounds stems from, its argument tends to zero, so the preceding expansion gives

$$\begin{aligned} s((1-2\delta+\zeta)\zeta) &= \frac{(1-2\delta+\zeta)\zeta}{4}\log\frac{1}{(1-2\delta+\zeta)\zeta} + O(\zeta) \\ &= \frac{1-2\delta}{4}\zeta\log\frac{1}{\zeta} + O(\zeta). \end{aligned}$$

Therefore,

$$\frac{F_\delta(1-2\delta+\zeta) - F_\delta(1-2\delta)}{\zeta} = -\frac{1-2\delta}{4}\log\frac{1}{\zeta} + O(1),$$

which tends to $-\infty$ as $\zeta \to 0^+$, as $1-2\delta > 0$. Hence $F_\delta(1-2\delta+\zeta) < F_\delta(1-2\delta)$ for all sufficiently small $\zeta > 0$. Thus, we conclude that

$$R_{\mathrm{MQC}}(\delta) \leqslant F_\delta(1-2\delta+\zeta) < F_\delta(1-2\delta) = R_{\mathrm{MRRW}}(\delta). \qquad \square$$

Now, we have established everything we need to prove Theorem 2, which we now show.

*Proof of Theorem 2.* The MQC rate upper bound in Corollary 23 gives $R_2(\delta) \leqslant R_{\mathrm{MQC}}(\delta)$, while Proposition 25 gives $R_{\mathrm{MQC}}(\delta) < R_{\mathrm{MRRW}}(\delta)$ for every $\delta \in (0, \frac{1}{2})$. Combining the two inequalities proves the theorem. $\square$

**Comparing the curves numerically.** Fig. 1 plots the rate bounds resulting from BEC (Corollary 19), BSC (Corollary 20), PSC (Corollary 21), mPSC (Corollary 22), and the new rate bound from MQC (Corollary 23). We omit the 2MQC rate bound (Corollary 24) from the plot, since the MQC bound alone already outperforms all previously known rate bounds for $\delta > 0.1951$. The inset provides a magnified view of the region around $\delta = \frac{1}{4}$, making the improvement more readily visible.

In Fig. 2, we give numerical substantiation of the proof of Proposition 25 for $\delta = 1/4$. In particular, Fig. 2 shows a numerical evaluation of the full function $u \mapsto F_{1/4}(u)$ on its admissible interval. At the point $u = 1-2\delta = 1/2$, the function $F_{1/4}(u)$ equals the first MRRW bound, as predicted by Remark 5.1. The curve initially falls, reaches the marked minimum, and then rises.

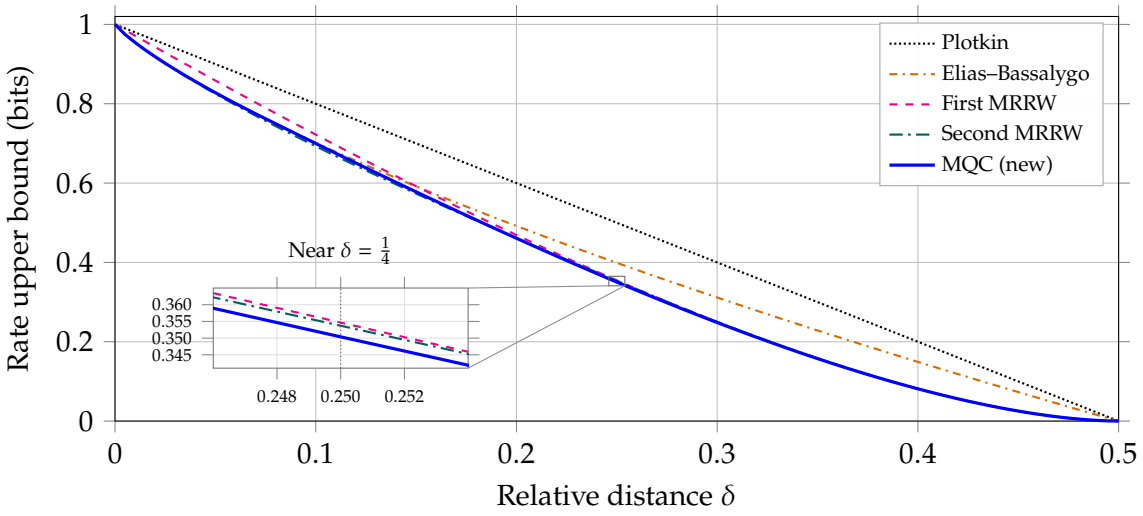


Figure 1: Five benchmark rate curves, shown for $0 \leqslant \delta \leqslant \frac{1}{2}$. The Plotkin [Plo60], Elias–Bassalygo [Eli55, Bas65], and first MRRW [MRRW77] curves are closed form, while the second MRRW [MRRW77] and MQC curves are sampled one-dimensional minima. The two-variable 2MQC infimum from (46) is not plotted; analytically, it lies strictly below the second MRRW throughout $0 < \delta < \frac{1}{2}$ by Proposition 26. The inset magnifies the marked window around $\delta = \frac{1}{4}$, where the sampled MQC value, approximately 0.3503791, lies below the sampled second MRRW bound 0.3537105 and the closed-form first MRRW bound 0.3545789. The MQC improvement over the first MRRW curve is analytic throughout the open interval (Proposition 25); its relation to the second MRRW curve at this point is numerical. All displayed curves use the common endpoint conventions rate 1 at $\delta = 0$ and rate 0 at $\delta = \frac{1}{2}$.

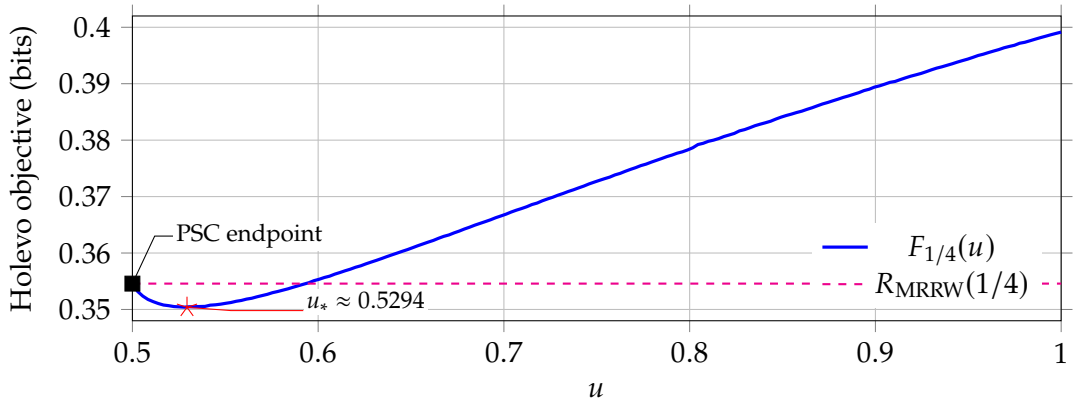


Figure 2: A plot of the MQC Holevo information function $F_{1/4}(u) = s(u^2) - s(u(u - \frac{1}{2}))$ from Eq. (47) over the range $u \in [\frac{1}{2}, 1]$, where $s(D) := h((1 - \sqrt{1 - D})/2)$ as in Proposition 14. The black square marks the PSC endpoint when $u = 1 - 2\delta$ (cf. Remark 5.1), where the MQC Holevo information equals the first MRRW bound [MRRW77], which is approximately $R_{\mathrm{MRRW}}(1/4) \approx 0.3545789$ in this case. The red star marks the numerically obtained minimizer $u_* \approx 0.5293949$ of $F_{1/4}$, whose value is $F_{1/4}(u_*) \approx 0.3503791$.

### 6.2 Low-rate regime

In this subsection, we examine how rapidly the MQC gain over the first MRRW bound decays in the low-rate regime. Define the gap between the two bounds as

$$\Delta(\delta) := R_{\mathrm{MRRW}}(\delta) - R_{\mathrm{MQC}}(\delta).$$

Near the endpoint, this gap is difficult to resolve in Fig. 1. We provide instead Fig. 3, which displays the gap on a logarithmic scale for greater clarity. The formal expansion below characterizes the rate at which the gap decays and identifies the leading order at which the two bounds differ.

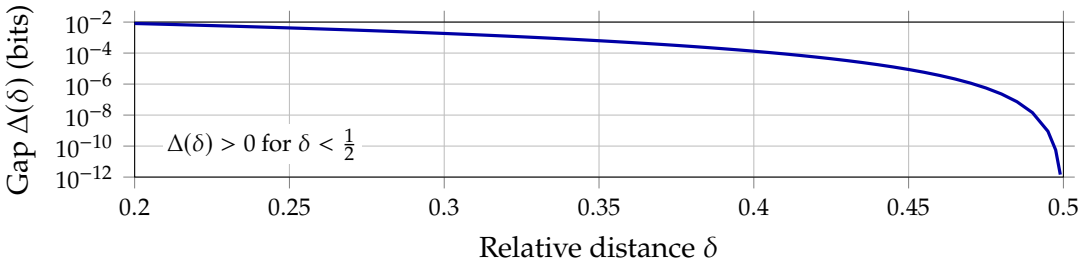


Figure 3: Numerically minimized values of $\Delta(\delta) = R_{\mathrm{MRRW}}(\delta) - R_{\mathrm{MQC}}(\delta)$ for $0.2 \leqslant \delta \leqslant 0.499$, shown on a logarithmic vertical scale. The strict-improvement proposition (Proposition 25) proves $\Delta(\delta) > 0$ for every $\delta \in (0, \frac{1}{2})$. The formal calculation below suggests that $\Delta(\frac{1}{2} - \varepsilon) = \varepsilon^4 \log \mathrm{e} + O(\varepsilon^6 \log(1/\varepsilon))$.

**Formal expansion.** We now heuristically examine the scale visible in Fig. 3. Put $\delta = \frac{1}{2} - \varepsilon$ with $\varepsilon \to 0^+$. The identity $R_{\mathrm{MRRW}}(\frac{1}{2} - \varepsilon) = s(4\varepsilon^2)$ reduces the MRRW expansion to the following small-$D$ expansion of $s$, where $D \to 0^+$:

$$s(D) = \frac{D}{4}\left[\log\frac{4}{D} + \log \mathrm{e}\right] + \frac{D^2}{16}\left[\log\frac{4}{D} - \frac{\log \mathrm{e}}{2}\right] + O\left(D^3 \log\frac{1}{D}\right).$$

Now, note that the admissible interval for $F_{\frac{1}{2}-\varepsilon}$ starts at $u = 2\varepsilon$. Numerical minimizers suggest a displacement of order $\varepsilon^3$ from this endpoint, so we set $u = 2\varepsilon + 8\xi\varepsilon^3$ for a fixed value $\xi > 0$. Substitution into the expansion of $s$ then gives

$$F_{\frac{1}{2}-\varepsilon}(2\varepsilon + 8\xi\varepsilon^3) - s(4\varepsilon^2) = 4\xi\varepsilon^4\left[\log(4\xi) - \log \mathrm{e}\right] + O\left(\varepsilon^6 \log\frac{1}{\varepsilon}\right).$$

Now, to minimize the leading $\varepsilon^4$ coefficient, differentiate with respect to $\xi$. Because all logarithms are base 2, $\frac{d}{d\xi}\log(4\xi) = (\log \mathrm{e})/\xi$. The $\log \mathrm{e}$ terms cancel, so the derivative of $4\xi[\log(4\xi) - \log \mathrm{e}]$ is $4\log(4\xi)$. Thus, the displayed coefficient is minimized at $\xi = 1/4$, corresponding to the candidate value $u = 2\varepsilon + 2\varepsilon^3$. Plugging this value for both the first MRRW bound and the MQC rate and expanding gives us

$$R_{\mathrm{MRRW}}\left(\frac{1}{2} - \varepsilon\right) = \varepsilon^2\left[2\log\frac{1}{\varepsilon} + \log \mathrm{e}\right] + \varepsilon^4\left[2\log\frac{1}{\varepsilon} - \frac{\log \mathrm{e}}{2}\right] + O\left(\varepsilon^6 \log\frac{1}{\varepsilon}\right),$$
$$F_{\frac{1}{2}-\varepsilon}(2\varepsilon + 2\varepsilon^3) = \varepsilon^2\left[2\log\frac{1}{\varepsilon} + \log \mathrm{e}\right] + \varepsilon^4\left[2\log\frac{1}{\varepsilon} - \frac{3\log \mathrm{e}}{2}\right] + O\left(\varepsilon^6 \log\frac{1}{\varepsilon}\right).$$

In particular, taking their difference thus gives us

$$\Delta\left(\frac{1}{2}-\varepsilon\right) \geqslant R_{\mathrm{MRRW}}\left(\frac{1}{2}-\varepsilon\right) - F_{\frac{1}{2}-\varepsilon}(2\varepsilon+2\varepsilon^3) = \varepsilon^4 \log \mathrm{e} + O\left(\varepsilon^6 \log \frac{1}{\varepsilon}\right).$$

First, observe that the positive coefficient is consistent with the strict-improvement proposition (Proposition 25). Moreover, the two bounds agree at the leading $\varepsilon^2 \log(1/\varepsilon)$ scale, the $\varepsilon^2$ term, and in the logarithmic fourth-order term. Their first difference occurs at the term $\varepsilon^4 \log \mathrm{e}$. We remark that this analysis is primarily based on what numerical computations seem to suggest.

### 6.3 Comparing with the second MRRW bound

In this subsection, we prove the strict 2MQC improvement over the second MRRW bound (Proposition 26) and complete the proof of Theorem 3. For ease of exposition, we adopt the shorthand

$$H_\delta(u) \coloneqq 1 + s(u^2) - s(u^2 + 2\delta u + 2\delta), \tag{48}$$

for the objective in (31), where $u$ ranges over $[0, 1-2\delta]$. By Remark 5.5, the 2MQC contains the MQC and mPSC as special cases. Thus, we see that

$$R_{\mathrm{2MQC}}(\delta) \leqslant \min\left\{R_{\mathrm{MQC}}(\delta), R^{(2)}_{\mathrm{MRRW}}(\delta)\right\}.$$

We now show that this inequality for the second MRRW bound is always strict.

**Proposition 26** (Strict improvement over the second MRRW bound)**.** *For every* $\delta \in (0, \frac{1}{2})$,

$$R_{\mathrm{2MQC}}(\delta) < R^{(2)}_{\mathrm{MRRW}}(\delta).$$

*Proof.* Suppose first that $R^{(2)}_{\mathrm{MRRW}}(\delta) = R_{\mathrm{MRRW}}(\delta)$. By Remark 5.5 and Proposition 25, we have that

$$R_{\mathrm{2MQC}}(\delta) \leqslant R_{\mathrm{MQC}}(\delta) < R_{\mathrm{MRRW}}(\delta) = R^{(2)}_{\mathrm{MRRW}}(\delta).$$

For the remainder of the proof, assume $R^{(2)}_{\mathrm{MRRW}}(\delta) < R_{\mathrm{MRRW}}(\delta)$. By (48) and the variational formula of the second MRRW bound (31), we have

$$R^{(2)}_{\mathrm{MRRW}}(\delta) = \min_{0 \leqslant u \leqslant 1-2\delta} H_\delta(u).$$

The continuous function $H_\delta$ attains its minimum on $[0, 1-2\delta]$. The right endpoint is not a minimizer because

$$H_\delta(1-2\delta) = R_{\mathrm{MRRW}}(\delta) > R^{(2)}_{\mathrm{MRRW}}(\delta).$$

The left endpoint is not a minimizer either. Indeed, $s(u^2) = O(u^2 \log(1/u))$ has vanishing one-sided derivative at $u = 0$, and hence

$$H'_\delta(0^+) = -2\delta s'(2\delta) < 0.$$

Thus, we can fix a minimizer $u_* \in (0, 1-2\delta)$.

Now, observe that the mPSC endpoint is when $v = 0$. Indeed, by (43), we have $v = 0$ if and only if $\eta = 0$, which recovers mPSC by Remark 5.5. Now, define the shorthand

$$G_\delta(u, v) \coloneqq 1 + s(u^2) - s\left(\lambda_\delta(u, v)\right) - s(v^2), \tag{49}$$

where $\lambda_\delta(u, v)$ is as defined in Eq. (45). Note that when $v = 0$, we have that $\lambda_\delta(u_*, 0) = u_*^2 + 2\delta u_* + 2\delta$. Thus

$$G_\delta(u_*, 0) = H_\delta(u_*) = R^{(2)}_{\mathrm{MRRW}}(\delta), \tag{50}$$

as expected by Remark 5.5.

We will now perturb $v \to 0^+$ slightly away from 0 on the positive side while keeping the PGM bit error rate fixed. Now, from Eq. (45), we see that $\lambda_\delta(u_*, v) = \lambda_0 + O(v^2)$ for $\lambda_0 \coloneqq \lambda_\delta(u_*, 0) \in (0, 1)$. Thus, for all sufficiently small $v > 0$, we have that $A_\delta(u_*, v) = \lambda_\delta(u_*, v)$. Consequently, by Eq. (46), we have

$$R_{\mathrm{2MQC}}(\delta) \leqslant G_\delta(u_*, v) \tag{51}$$

for $v \to 0^+$. Now, write the difference

$$G_\delta(u_*, v) - G_\delta(u_*, 0) = -\left[s\left(\lambda_\delta(u_*, v)\right) - s(\lambda_0)\right] - s(v^2).$$

By the mean value theorem, we have that

$$s\left(\lambda_\delta(u_*, v)\right) - s(\lambda_0) = O(v^2).$$

For the second term, the expansion used in the proof of Proposition 25 gives

$$s(v^2) = \frac{v^2}{2} \log \frac{1}{v} + O(v^2).$$

Therefore,

$$\frac{G_\delta(u_*, v) - G_\delta(u_*, 0)}{v^2} = -\frac{1}{2} \log \frac{1}{v} + O(1),$$

which tends to $-\infty$ as $v \to 0^+$. Hence

$$G_\delta(u_*, v) < G_\delta(u_*, 0) \tag{52}$$

for all sufficiently small $v > 0$. Combining Eqs. (50) to (52), we therefore conclude that

$$R_{\mathrm{2MQC}}(\delta) \leqslant G_\delta(u_*, v) < G_\delta(u_*, 0) = R^{(2)}_{\mathrm{MRRW}}(\delta). \qquad \square$$

With this proposition at hand, we can now establish Theorem 3.

*Proof of Theorem 3.* The theorem follows by combining the 2MQC rate upper bound $R_2(\delta) \leqslant R_{\mathrm{2MQC}}(\delta)$ (Corollary 24) together with the strict inequality $R_{\mathrm{2MQC}}(\delta) < R^{(2)}_{\mathrm{MRRW}}(\delta)$ (Proposition 26). $\square$

We conclude this section with numerical estimates of our rate bound improvement obtained from the 2MQC (Corollary 24) relative to the previous state-of-the-art second MRRW bound (Corollary 22). This is presented in Table 2. The second MRRW bound column minimizes the one-dimensional objective in (31). The 2MQC value column records reproducible floating-point estimates of the minimal value of the optimization problem in Eq. (46). For the first four rows, the 2MQC column evaluates the exact-error objective $G_\delta(u, v)$ from (49) at the displayed interior coordinates, as $\lambda_\delta(u, v) < 1$ in that regime. For the remaining rows, $A_\delta(u, v) = 1$ instead of $\lambda_\delta(u, v)$.

Table 2: The second MRRW bound compared numerically with floating-point estimates of $R_{\mathrm{2MQC}}(\delta)$, based on estimating the minimal value of the optimization problem in Eq. (46). The displayed gap in the last column is $R^{(2)}_{\mathrm{MRRW}}(\delta)$ minus an estimate of the 2MQC rate bound $R_{\mathrm{2MQC}}(\delta)$. The displayed coordinates $u, v$ are floating-point estimates witnessing the numerical upper bound on $R_{\mathrm{2MQC}}(\delta)$ displayed in the table. Note that all entries are floating-point evaluations, not certified global optima.

| $\delta$ | $u$ | $v$ | $R^{(2)}_{\mathrm{MRRW}}(\delta)$ | 2MQC value | Gap |
|---|---|---|---|---|---|
| 0.05 | 0.0544815 | 0.00135647 | 0.825136808 | 0.825136151 | 0.000000657 |
| 0.10 | 0.1193920 | 0.00618142 | 0.692740743 | 0.692727379 | 0.000013364 |
| 0.15 | 0.1997723 | 0.01681725 | 0.573450044 | 0.573356705 | 0.000093339 |
| 0.20 | 0.3160463 | 0.04188496 | 0.461359604 | 0.460883553 | 0.000476051 |
| 0.25 | 0.5293949 | 0.12474570 | 0.353710543 | 0.350379084 | 0.003331459 |
| 0.30 | 0.4151219 | 0.07923030 | 0.250224912 | 0.248376065 | 0.001848847 |
| 0.35 | 0.3064440 | 0.04443793 | 0.158132937 | 0.157505576 | 0.000627361 |
| 0.40 | 0.2019380 | 0.01978282 | 0.081468915 | 0.081336612 | 0.000132303 |
| 0.45 | 0.1002467 | 0.00497293 | 0.025266128 | 0.025257384 | 0.000008744 |

# 7 Extending to $q$-ary codes

In this section, we sketch how one might extend our framework to $q$-ary codes. Fix an integer $q \geqslant 2$, write $[q] \coloneqq \{0, \dots, q-1\}$, and let $A_q(n, d)$ be the largest cardinality of a code in $[q]^n$ with minimum distance at least $d$. We normalize rates in base $q$ and write

$$R_q(\delta) \coloneqq \limsup_{n\to\infty} \frac{1}{n} \log_q A_q(n, \lceil \delta n \rceil).$$

For $\delta \geqslant 1 - \frac{1}{q}$, the $q$-ary Plotkin bound shows that $R_q(\delta) = 0$. Thus, the nontrivial interval in the $q$-ary setting is $\delta \in (0, 1 - \frac{1}{q})$. Since the pretty good criterion (Theorem 1) is stated for binary codes, we first extend its output-symmetry hypothesis to the cyclic $q$-ary alphabet.

**Definition 7.1** ($q$-ary output-symmetry)**.** *For a positive integer $d$, a $q$-ary-input cq channel $\Phi_\sigma : [q] \to \mathsf{D}(\mathbb{C}^d)$, given by $\Phi_\sigma(x) = \sigma_x$, is **$q$-ary output-symmetric** if there is a unitary $U$ such that*

$$U \sigma_x U^\dagger = \sigma_{x+1}$$

*for every $x \in [q]$, where subscripts and addition are taken modulo $q$. Iterating this identity gives $U^a \sigma_x (U^\dagger)^a = \sigma_{x+a}$ for all $a, x \in [q]$.*

With this terminology in place, the following proposition records the precise $q$-ary extension.

**Proposition 27** ($q$-ary pretty good criterion)**.** *Let $\Phi_\sigma : x \mapsto \sigma_x$ be a $q$-ary output-symmetric cq channel. If the uniform-input PGM symbol error rate $p_{\mathrm{e}}(\{(1/q, \sigma_x) : x \in [q]\}) < \delta$, then every code $C \subseteq [q]^n$ with minimum distance at least $\delta n$ satisfies*

$$\frac{1}{n} \log_q |C| \leqslant \frac{\chi(\{(1/q, \sigma_x)\}_{x\in[q]})}{\log q} + O_{q,\sigma,\delta}\left(\frac{1}{\sqrt{n}}\right). \tag{53}$$

*Consequently,*

$$R_q(\delta) \leqslant \frac{\chi(\{(1/q, \sigma_x)\}_{x\in[q]})}{\log q}.$$

*Proof.* Let $U$ be the unitary witnessing output-symmetry. Write

$$p_{\mathrm{e}} \coloneqq p_{\mathrm{e}}\left(\{(1/q, \sigma_x) : x \in [q]\}\right), \qquad \chi_u \coloneqq \chi\left(\{(1/q, \sigma_x) : x \in [q]\}\right).$$

The PGM coarse-graining and data-processing statements, Lemmas 8 and 10, are already formulated for arbitrary finite ensembles, so they apply here without modification. Since the reduction of blockwise to bitwise PGM in Lemma 9 was specialized to the binary case, we will instead directly invoke Lemma 8 for its $q$-ary analog. The remaining ingredient in the proof of the $q$-ary analog of Theorem 17 is a $q$-ary analog of Lemma 11. We first prove its $q$-ary replacement: for every distribution $\pi$ on $[q]$, the PGM error $p_{\mathrm{e},\pi}$ of $\{(\pi_x, \sigma_x) : x \in [q]\}$ satisfies $p_{\mathrm{e},\pi} \leqslant p_{\mathrm{e}}$.

The proof of the $q$-ary analog of Lemma 11 closely follows the one presented in Section A.6. Indeed, define

$$\bar{\sigma}_\pi \coloneqq \sum_{x\in[q]} \pi_x \sigma_x, \qquad M^\pi_{\hat{x}} \coloneqq \pi_{\hat{x}} \bar{\sigma}_\pi^{-1/2} \sigma_{\hat{x}} \bar{\sigma}_\pi^{-1/2},$$

and introduce a classical system with basis $\{|a\rangle : a \in [q]\}$. For each $x \in [q]$, define

$$\widehat{\sigma}_x \coloneqq \sum_{a\in[q]} \pi_{x-a} |a\rangle\langle a| \otimes \sigma_x.$$

The translated probabilities sum to one, so each $\widehat{\sigma}_x$ is a bona fide density state. Endow the lifted states $\widehat{\sigma}_x$ with the uniform prior. Their average state and PGM operators are, respectively,

$$\overline{\widehat{\sigma}} = \frac{1}{q} \sum_{a\in[q]} |a\rangle\langle a| \otimes U^a \bar{\sigma}_\pi (U^\dagger)^a,$$
$$\widehat{M_{\hat{x}}} = \sum_{a\in[q]} |a\rangle\langle a| \otimes U^a M^\pi_{\hat{x}-a} (U^\dagger)^a.$$

Indeed, conditioned on block $a$, the translated symbol $x - a$ has law $\pi$ and $\sigma_x = U^a \sigma_{x-a} (U^\dagger)^a$. The lifted PGM therefore has success probability

$$\frac{1}{q} \sum_{x\in[q]} \mathrm{Tr}(\widehat{\sigma}_x \widehat{M_x}) = \frac{1}{q} \sum_{a,x\in[q]} \pi_{x-a} \mathrm{Tr}(\sigma_{x-a} M^\pi_{x-a}) = \sum_{x\in[q]} \pi_x \mathrm{Tr}(\sigma_x M^\pi_x).$$

Thus, the lifted uniform ensemble has PGM error exactly $p_{\mathrm{e},\pi}$.

Tracing out the classical system sends $\widehat{\sigma}_x$ to $\sigma_x$. Moreover, the PGM data-processing lemma (Lemma 10) says that this trace-out channel cannot decrease PGM error, and hence

$$p_{\mathrm{e},\pi} \leqslant p_{\mathrm{e}}. \tag{54}$$

This proves the $q$-ary counterpart of Lemma 11.

We now follow the proof of Theorem 17. Draw $\mathbf{c}$ uniformly from $C$, apply the block PGM to $\sigma_{\mathbf{c}} \coloneqq \sigma_{\mathbf{c}_1} \otimes \cdots \otimes \sigma_{\mathbf{c}_n}$, and denote its codeword-valued outcome by $\hat{\mathbf{c}}$. For each coordinate $i$, let $\pi_i(x) \coloneqq \mathbf{Pr}[\mathbf{c}_i = x]$. By Lemma 8, which applies as-is with the coordinate map $c \mapsto c_i$, we have

$$\mathbf{Pr}[\hat{\mathbf{c}}_i \neq \mathbf{c}_i] = p_{\mathrm{e}}\left(\{(\pi_i(x), \mathbf{E}[\sigma_{\mathbf{c}} \mid \mathbf{c}_i = x]) : x \in [q]\}\right).$$

The partial trace channel similarly satisfies $\mathcal{T}_i(\mathbf{E}[\sigma_{\mathbf{c}} \mid \mathbf{c}_i = x]) = \sigma_x$. Applying Lemma 10 as-is and then the $q$-ary uniform-prior bound (54) gives

$$\mathbf{Pr}[\hat{\mathbf{c}}_i \neq \mathbf{c}_i] \leqslant p_{\mathrm{e}}\left(\{(\pi_i(x), \sigma_x) : x \in [q]\}\right) \leqslant p_{\mathrm{e}}.$$

Summing over the coordinates yields the $q$-ary counterpart of Theorem 17:

$$\mathbf{E}[d(\hat{\mathbf{c}}, \mathbf{c})] = \sum_{i=1}^{n} \mathbf{Pr}[\hat{\mathbf{c}}_i \neq \mathbf{c}_i] \leqslant n p_{\mathrm{e}}.$$

The argument of Corollary 18 uses only Hamming distance and therefore extends unchanged. Since distinct codewords are separated by at least $\delta n$, it gives

$$\mathbf{Pr}[\hat{\mathbf{c}} = \mathbf{c}] \geqslant 1 - \frac{p_{\mathrm{e}}}{\delta} =: \eta > 0.$$

It remains to reproduce the final converse step in the proof of Theorem 1. Here we use the $q$-ary output-symmetric counterparts of Proposition 6 and Theorem 7. The same averaging over input shifts and entropy-subadditivity proof as in Proposition 6 gives

$$C(\Phi_\sigma) = \chi_u.$$

Likewise, the proof of Theorem 7 applies using the unitary $U$ from Definition 7.1 and setting $V_c := U^{c_1} \otimes \cdots \otimes U^{c_n}$. Indeed, $V_c \sigma_0^{\otimes n} V_c^\dagger = \sigma_c$, while every power $U^a$ fixes the uniform average state. The decoder-averaging step and both spectral truncations therefore remain unchanged. Thus, there is a channel-dependent constant $L_\sigma \geqslant 1$ such that, for every $\varepsilon > 0$,

$$\mathbf{Pr}[\hat{\mathbf{c}} = \mathbf{c}] \leqslant 2^{-n[n^{-1}\log|C| - \chi_u - 2\varepsilon]} + 4\exp\left(-\varepsilon^2 n / L_\sigma^2\right).$$

Exactly as in the proof of Theorem 1, choose

$$\varepsilon := L_\sigma \sqrt{\frac{\log(8/\eta)}{n \log \mathrm{e}}}.$$

The exponential term is then $\eta/2$, and hence

$$\eta \leqslant \mathbf{Pr}[\hat{\mathbf{c}} = \mathbf{c}] \leqslant 2^{-n[n^{-1}\log|C| - \chi_u - 2\varepsilon]} + \frac{\eta}{2}.$$

Taking logarithms and rearranging gives

$$\begin{aligned}
\frac{1}{n}\log|C| &\leqslant \chi_u + 2\varepsilon + \frac{1}{n}\log\frac{2}{\eta} \\
&= \chi_u + 2L_\sigma\sqrt{\frac{\log(8/\eta)}{n\log\mathrm{e}}} + \frac{1}{n}\log\frac{2}{\eta} \\
&= \chi_u + O_{q,\sigma,\delta}\left(\frac{1}{\sqrt{n}}\right).
\end{aligned}$$

Dividing by $\log q$ proves (53). Maximizing over $C$ and taking $n \to \infty$ gives

$$R_q(\delta) \leqslant \frac{\chi_u}{\log q}.$$

□

Thus, as in the binary setting, the $q$-ary pretty good criterion reduces to channel selection. Indeed, choose a $q$-ary output-symmetric cq channel, read off its uniform-input PGM symbol error and Holevo information, and substitute them into Proposition 27. To state the resulting bounds compactly, define

$$h_q(x) := \frac{h(x) + x\log(q-1)}{\log q},$$
$$J_q(x) := \frac{q-1}{q}\left(1 - \sqrt{1 - \frac{qx}{q-1}}\right),$$
$$\gamma_q(x) := \frac{q - 1 - (q-2)x - 2\sqrt{(q-1)x(1-x)}}{q}.$$

We now specify only the channels. Their PGM symbol errors, Holevo information, and induced rate bounds are collected in Table 3.

**$q$-ary erasure channel ($q$-EC).** For $0 \leqslant p \leqslant 1$, let $|?\rangle$ be orthogonal to $\{|x\rangle\}_{x\in[q]}$, and define

$$\sigma_x^{q\mathrm{EC}}(p) := (1-p)\,|x\rangle\langle x| + p\,|?\rangle\langle ?|\,.$$

**$q$-ary symmetric channel ($q$-SC).** For $0 \leqslant p \leqslant \frac{q-1}{q}$, define the diagonal output states

$$\sigma_x^{q\mathrm{SC}}(p) := (1-p)\,|x\rangle\langle x| + \frac{p}{q-1}\sum_{y\in[q]\setminus\{x\}} |y\rangle\langle y|.$$

**$q$-ary pure-state channel ($q$-PSC).** For $0 \leqslant p \leqslant \frac{q-1}{q}$, define

$$\sigma_x^{q\mathrm{PSC}}(p) := |\psi_x^{(q)}(p)\rangle\langle\psi_x^{(q)}(p)|\,, \qquad |\psi_x^{(q)}(p)\rangle := \sqrt{1-p}\,|x\rangle + \sqrt{\frac{p}{q-1}}\sum_{y\in[q]\setminus\{x\}} |y\rangle.$$

**$q$-ary mixed channel ($q$-MC).** Following the binary MQC construction in Sections 1.3 and 5.1, we place a $q$-ary symmetric channel before a $q$-PSC. Let $\mathcal{T}_\eta^{(q)}$ send $x$ to itself with probability $1-\eta$ and to each $y \neq x$ with probability $\eta/(q-1)$, and let $\mathcal{W}_r^{(q)}$ be the preparation $y \mapsto \sigma_y^{q\mathrm{PSC}}(r)$. Their composition $\mathcal{W}_r^{(q)} \circ \mathcal{T}_\eta^{(q)}$ is the $q$-MC: it prepares the pure state indexed by the perturbed symbol and discards the classical error.

For $0 \leqslant a \leqslant \frac{q-1}{q}$, define $\nu_q(a) := 1 - qa/(q-1)$ as the centered correlation of the $q$-SC with crossover probability $a$. Also denote its inverse by $r_q(v) := (q-1)(1-v)/q$. Fix a desired PGM symbol error $0 \leqslant p < \frac{q-1}{q}$ and $\nu_q(p) \leqslant v \leqslant 1$, and set

$$t_q(p,v) := \sqrt{\frac{\nu_q(p)}{v}}, \qquad \eta_q(p,v) := r_q\left(t_q(p,v)\right).$$

Here $r_q(v)$ is the parameter of the underlying $q$-PSC, while $\eta_q(p,v)$ is the crossover probability of the discarded $q$-SC. Abbreviate

$$\tau_x(v) := \sigma_x^{q\mathrm{PSC}}\left(r_q(v)\right), \qquad \bar{\tau}_v := \frac{1}{q}\sum_{y\in[q]} \tau_y(v).$$

The composition therefore has output

$$\begin{aligned}\sigma_x^{q\mathrm{MC}}(p,v) &\coloneqq \left(\mathcal{W}^{(q)}_{r_q(v)} \circ \mathcal{T}^{(q)}_{\eta_q(p,v)}\right)(x) \\ &= \left(1-\eta_q(p,v)\right)\tau_x(v) + \frac{\eta_q(p,v)}{q-1}\sum_{y\in[q]\setminus\{x\}}\tau_y(v) \\ &= t_q(p,v)\tau_x(v) + \left(1-t_q(p,v)\right)\bar{\tau}_v. \end{aligned} \tag{55}$$

The last equality uses $\sum_{y\in[q]\setminus\{x\}}\tau_y(v) = q\bar{\tau}_v - \tau_x(v)$ and $t_q(p,v) = \nu_q(\eta_q(p,v))$.

For the uniform $q$-PSC ensemble, the PGM is the computational-basis measurement. As in the binary MQC, the PGM reverses the classical pre-processing by applying the same $q$-SC once more. As correlations multiply under composition, the final symbol error $p_{\mathrm{e}}$ satisfies $\nu_q(p_{\mathrm{e}}) = t_q(p,v)^2 v = \nu_q(p)$ and hence equals $p$. At $v = \nu_q(p)$, the pre-processing is noiseless, and the channel is exactly $q$-PSC($p$); at $v = 1$, the pure states are orthogonal and the channel is exactly $q$-SC($J_q(p)$). For alphabet size 2, (55) becomes

$$\mathrm{MQC}\left(\frac{1-v}{2}, \frac{1-t_2(p,v)}{2}\right),$$

with the first argument the underlying binary PSC parameter and the second the discarded BSC crossover probability.

To state its Holevo information compactly, define

$$\begin{aligned} g_q(v) &\coloneqq \gamma_q\left(r_q(v)\right), \\ \ell_q(p,v) &\coloneqq \frac{\left(1-t_q(p,v)\right)g_q(v)}{q-1}, \\ \lambda_{\pm}^{(q)}(p,v) &\coloneqq \frac{1}{2}\Bigg(1-(q-2)\ell_q(p,v) \pm \Bigg[\left(1-(q-2)\ell_q(p,v)\right)^2 \\ &\qquad - \frac{4\left(1-t_q(p,v)\right)\left(1+(q-1)t_q(p,v)\right)}{q-1} g_q(v)\left(1-g_q(v)\right)\Bigg]^{1/2}\Bigg), \\ \Xi_q(p,v) &\coloneqq h_q\left(g_q(v)\right) + \frac{1}{\log q}\Big[\lambda_+^{(q)}(p,v)\log\lambda_+^{(q)}(p,v) + \lambda_-^{(q)}(p,v)\log\lambda_-^{(q)}(p,v) \\ &\qquad + (q-2)\ell_q(p,v)\log\ell_q(p,v)\Big]. \end{aligned}$$

Each $q$-MC output has spectrum $\lambda_+^{(q)}(p,v)$, $\lambda_-^{(q)}(p,v)$, and $\ell_q(p,v)$ with multiplicity $q-2$, while the average state has spectrum $1-g_q(v)$ and $g_q(v)/(q-1)$ with multiplicity $q-1$. These spectra give the normalized Holevo information $\Xi_q(p,v)$, and the uniform-input PGM symbol error is exactly $p$.

The unitary that sends $|x\rangle$ to $|x+1\rangle$ (and fixes $|?\rangle$ for the $q$-EC) shows that all four channels are $q$-ary output-symmetric. For the open channel-design problem, write

$$R^{\star}_{q,\mathrm{cq}}(\delta) \coloneqq \inf_{\substack{\{\sigma_x\}_{x\in[q]}\ q\text{-ary output-symmetric} \\ p_{\mathrm{e}}(\{(1/q,\sigma_x)\}_{x\in[q]})<\delta}} \frac{\chi(\{(1/q,\sigma_x)\}_{x\in[q]})}{\log q}.$$

The final column below follows by choosing a strictly feasible channel in Proposition 27 and then approaching the displayed error threshold by continuity.

Table 3: $q$-ary channel blueprint. All rates and Holevo quantities are normalized by $\log q$. The first four rows give explicit channels used in Proposition 27; the last row records the open optimization over all $q$-ary output-symmetric cq channels.

| **Channel** | **PGM symbol error** | **Uniform-input** $\chi/\log q$ | **Upper bound on** $R_q(\delta)$ |
|---|---|---|---|
| $q$-EC$(p)$ | $\frac{q-1}{q}p$ | $1-p$ | $1-\frac{q}{q-1}\delta$ |
| $q$-SC$(p)$ | $2p-\frac{q}{q-1}p^2$ | $1-h_q(p)$ | $1-h_q\left(J_q(\delta)\right)$ |
| $q$-PSC$(p)$ | $p$ | $h_q\left(\gamma_q(p)\right)$ | $h_q\left(\gamma_q(\delta)\right)$ |
| $q$-MC$(p,v)$ | $p$ | $\Xi_q(p,v)$ | $\min_{\nu_q(\delta)\leqslant v\leqslant 1}\Xi_q(\delta,v)$ |
| Open output-symmetric channel search | $p_{\mathrm{e}}<\delta$ | $\chi(\{(1/q,\sigma_x)\}_x)/\log q$ | $R^{\star}_{q,\mathrm{cq}}(\delta)$ |

The first three rows recover the $q$-ary Plotkin bound [Plo60], the $q$-ary Elias–Bassalygo bound [Eli55, Bas65], and the first Delsarte–Levenshtein linear-programming bound [Del73, Lev95], respectively. The general $q$-ary linear-programming framework originates with Delsarte [Del72, Del73], was developed systematically by Levenshtein [Lev95] and Delsarte and Levenshtein [DL98], and admits more recent explicit dual constructions [CDA25]. Kaipa's anticode-based hybrid, using the optimal-anticode theorem of Ahlswede and Khachatrian, improves both the $q$-ary Elias and Plotkin bounds [AK98, Kai18]. At $q=2$, the $q$-PSC row is the first MRRW bound [MRRW77], while the $q$-MC row becomes the binary MQC bound.

The $q$-MC supplies the desired $q$-ary output-symmetric fixed-error path from $q$-SC$(J_q(p))$ to $q$-PSC$(p)$. We do not claim here that it improves the best known $q$-ary bounds when $q>2$. Its optimized curve should first be compared with Aaltonen's nonbinary linear-programming bound [Aal90], the shortening and straight-line bound of Laihonen and Litsyn [LL98], and the product-association-scheme bounds of Ben-Haim and Litsyn [BHL07]. Searching beyond the $q$-MC family is represented by the last row of the table.

## 8 PSC rate bounds for LDPC codes

This section proves Theorem 4. It is primarily intended as a proof of concept showing how the pretty good criterion (Theorem 1) can incorporate structure specific to a code family and thereby improve a specialized rate–distance bound. For simplicity, we specialize throughout to parity checks of weight at most 3.

We begin with the benchmark bound of Shangguan and Yang [SY26] in the notation of this paper. Set $r_\delta := \frac{1}{2}-\sqrt{\delta(1-\delta)}$. The $w=3$ bound in their Theorem 1.5, with their base-four entropy expanded into base-two logarithms, is

$$B_3^{\mathrm{SY}}(\delta) := \begin{cases} \frac{2}{3}, & r_\delta \geqslant \frac{1}{4}, \\ -r_\delta \log r_\delta - \frac{1-3r_\delta}{3}\log(1-3r_\delta), & 0<r_\delta<\frac{1}{4}, \end{cases} \tag{56}$$

extended continuously by $B_3^{\mathrm{SY}}(\frac{1}{2}) := 0$. Throughout $\frac{2}{5}\leqslant\delta<\frac{1}{2}$, this is the $w=3$ bound obtained in [SY26]. In proving Theorem 4, we will compare the rate bound obtained from the PSC with this

benchmark for $\delta \in [\frac{1}{6}, \frac{1}{2})$.

Recall the two exact optimality statements for PGM decoding over PSC from Section 1.1. Namely, for binary linear codes, the PGM achieves optimal error at both the block and bit levels. In this section, we will rely on PGM's bitwise optimality, stated below.

**Fact 28** (Bitwise PGM optimality for linear codes — [PR25, Proposition 5.1])**.** *Let $C \leqslant \mathbb{F}_2^n$ be linear and $p \in (0, \frac{1}{2})$. For the uniform ensemble $\{(|C|^{-1}, \sigma_c^{\mathrm{PSC}(p)}) : c \in C\}$, let $\hat{\mathbf{c}}$ be the outcome of the PGM. For any coordinate $i \in \{1, \dots, n\}$, the $i$'th outcome bit $\hat{\mathbf{c}}_i$ is the Helstrom-optimal measurement [Hel69] for decoding $\mathbf{c}_i$. In particular, the error of any explicit decoder for $\mathbf{c}_i$ upper-bounds the bitwise error $p_{\mathrm{e},i}$ of the block PGM.*

The explicit decoder we use will apply the equal-prior Helstrom measurement directly to the PSC outputs indexed by one weight-3 parity check containing the desired bit. Define the weight-three error function

$$e_3^{\mathrm{LDPC}}(p) \coloneqq \frac{1 + 4p(1-p) - (1-2p)\sqrt{1 + 12p(1-p)}}{4}. \tag{57}$$

We next bound this Helstrom error and then invoke the bitwise optimality of the code PGM to bound $p_{\mathrm{e},i}$ from above.

**Lemma 29** (One-check PSC decoder)**.** *Let $p \in (0, \frac{1}{2})$. Let $C \leqslant \mathbb{F}_2^n$ have minimum distance greater than one and let $C^\perp$ be generated by checks of weight at most 3. If a uniformly random $\mathbf{c} \in C$ is encoded into the product state $\sigma_{\mathbf{c}}^{\mathrm{PSC}(p)}$, then for every $i \in \{1, \dots, n\}$, the bitwise coarse-graining of the block PGM has error*

$$p_{\mathrm{e},i} \leqslant e_3^{\mathrm{LDPC}}(p) < p.$$

*Proof.* Fix $i \in \{1, \dots, n\}$. Because $C^\perp$ is generated by checks of weight at most 3, some such generating check contains $i$; otherwise, the weight-one vector supported at $i$ would lie in $C$, contrary to the minimum-distance assumption. Choose a minimum-weight $h \in C^\perp$ with $h_i = 1$. If $\mathrm{wt}(h) = 1$, the check fixes $\mathbf{c}_i = 0$, so the bitwise error is zero. Suppose henceforth that $\mathrm{wt}(h) \in \{2, 3\}$. The minimality of $h$ implies that the weight-one vector supported at $i$ is not in $C^\perp$; hence $c \mapsto c_i$ is a nonzero linear functional on $C$, and $\mathbf{c}_i$ is uniform.

By Fact 28, the error of any decoder for $\mathbf{c}_i$ upper-bounds $p_{\mathrm{e},i}$. Suppose first that $h$ has support $\{i, j, k\}$. Minimality implies that the projection of $C$ onto these coordinates is the full even-parity code $\{000, 011, 101, 110\}$. Indeed, any additional linear relation on the projection, or its sum with $h$, would give a check of weight at most two containing $i$. Since $\mathbf{c}$ is uniform on $C$, its projection is uniform as well. Thus, conditioned on $\mathbf{c}_i = b$, the three outputs have states

$$\omega_0 \coloneqq \frac{\sigma_{000}^{\mathrm{PSC}(p)} + \sigma_{011}^{\mathrm{PSC}(p)}}{2}, \qquad \omega_1 \coloneqq \frac{\sigma_{101}^{\mathrm{PSC}(p)} + \sigma_{110}^{\mathrm{PSC}(p)}}{2}.$$

We compute the error of their equal-prior Helstrom measurement. Let $D \coloneqq 4p(1-p)$ and let $V$ be the matrix whose columns are the four pure vectors underlying the states above, ordered as $000, 011, 101, 110$. Any two distinct words in this list have Hamming distance two, so

$$V^\dagger V = (1 - D) I_4 + D \mathbf{1}\mathbf{1}^\top.$$

Writing $\Lambda := \mathrm{diag}(1, 1, -1, -1)$ gives $\omega_0 - \omega_1 = \frac{1}{2} V \Lambda V^\dagger$. The nonzero eigenvalues of this difference are the eigenvalues of $\frac{1}{2}\Lambda V^\dagger V$; its characteristic polynomial is

$$\left(\lambda^2 - \frac{(1-D)^2}{4}\right)\left(\lambda^2 - \frac{(1-D)(1+3D)}{4}\right).$$

Consequently,

$$\|\omega_0 - \omega_1\|_1 = 1 - D + \sqrt{(1-D)(1+3D)},$$

and the Helstrom formula [Wri24, Theorem 3.2] gives

$$\frac{1}{2} - \frac{1}{4}\|\omega_0 - \omega_1\|_1 = e_3^{\mathrm{LDPC}}(p).$$

If $h$ has support $\{i, j\}$, its local projection is $\{00, 11\}$ and the corresponding Helstrom measurement distinguishes the two pure states $\sigma_0^{\mathrm{PSC}(p)} \otimes \sigma_0^{\mathrm{PSC}(p)}$ and $\sigma_1^{\mathrm{PSC}(p)} \otimes \sigma_1^{\mathrm{PSC}(p)}$. Its error is

$$\frac{1}{2}\left(1 - \sqrt{1 - 16p^2(1-p)^2}\right) \leqslant e_3^{\mathrm{LDPC}}(p),$$

where the inequality is equivalent, after two squarings, to

$$2\sqrt{1+D} \geqslant \sqrt{1-D} + \sqrt{1+3D}.$$

The weight-one case has zero error. Therefore, in every case,

$$p_{\mathrm{e},i} \leqslant e_3^{\mathrm{LDPC}}(p).$$

Finally, $e_3^{\mathrm{LDPC}}(p) < p$ is equivalent to

$$1 - 2p + \sqrt{1 + 12p(1-p)} > 2,$$

whose square reduces to $8p(1-2p) > 0$. □

*Remark* 8.1 (Beyond one check). The decoder in Lemma 29 is optimal among measurements restricted to the outputs in the chosen check. A joint measurement on the outputs covered by several checks containing $i$ may nevertheless decrease the bit error further. ◁

The associated PSC rate curve has a closed form. Indeed, in the coordinate $D = 4p(1-p)$, the function in (57) becomes

$$\frac{1 + D - \sqrt{(1-D)(1+3D)}}{4},$$

whose derivative with respect to $D$ is

$$\frac{1}{4}\left(1 - \frac{1-3D}{\sqrt{(1-D)(1+3D)}}\right) > 0.$$

Thus, it increases strictly from zero to $\frac{1}{2}$ as $D$ increases from zero to one. Solving $e_3^{\mathrm{LDPC}}(p) = \delta$ gives

$$4p(1-p) = \delta + \sqrt{\delta(2-3\delta)}.$$

Therefore,

$$U_3^{\mathrm{PSC}}(\delta) \coloneqq \inf_{\substack{0<p<\frac{1}{2} \\ e_3^{\mathrm{LDPC}}(p)<\delta}} h\left(\frac{1}{2}-\sqrt{p(1-p)}\right) = h\left(\frac{1-\sqrt{\delta+\sqrt{\delta(2-3\delta)}}}{2}\right). \tag{58}$$

Because $e_3^{\mathrm{LDPC}}(p) < p$, this constraint admits parameters $p$ beyond $\delta$, and the PSC Holevo information there is strictly smaller than its value at $p = \delta$; Theorem 4 records the resulting strict improvement over the first MRRW expression. Before proving the rate bound, we compare this curve analytically with (56). The estimates below are not fully optimized. Instead, they are chosen for ease of exposition. Fig. 4 displays $U_3^{\mathrm{PSC}}$ against both benchmarks.

**Lemma 30** (Shangguan–Yang interval separation)**.** *The pointwise inequality*

$$U_3^{\mathrm{PSC}}(u) < B_3^{\mathrm{SY}}(u)$$

*holds for* $\frac{1}{6} \leqslant u < \frac{1}{2}$.

*Proof.* For $u \in [\frac{1}{6}, \frac{1}{2})$, set

$$q \coloneqq \frac{1-\sqrt{u+\sqrt{u(2-3u)}}}{2}, \qquad r \coloneqq \frac{1}{2}-\sqrt{u(1-u)}.$$

By (58), $U_3^{\mathrm{PSC}}(u) = h(q)$, while $r$ is the parameter $r_\delta$ in (56). We compare both curves through $h(r)$.

The key estimate is

$$r - q > r^2. \tag{59}$$

To prove it, put $z \coloneqq 1-2u \in (0, \frac{2}{3}]$. Then $r(1-r) = \frac{z^2}{4}$, so (59) is equivalent to $q < \frac{z^2}{4}$. After substituting $u = \frac{1-z}{2}$ and squaring twice, this inequality reduces to

$$1+2z-3z^2 > \left(1+z-2z^2+\frac{z^4}{2}\right)^2.$$

Both squarings preserve the inequality because all quantities involved are positive. Finally,

$$\begin{aligned} &1+2z-3z^2-\left(1+z-2z^2+\frac{z^4}{2}\right)^2 \\ &\qquad = z^3\left(4-5z-z^2+2z^3-\frac{z^5}{4}\right). \end{aligned}$$

For $0 < z \leqslant \frac{2}{3}$, the factor in parentheses satisfies

$$4-5z-z^2+2z^3-\frac{z^5}{4} \geqslant 4-\frac{10}{3}-\frac{4}{9}-\frac{8}{243} = \frac{46}{243} > 0.$$

This proves (59).

The assumption $u \geqslant \frac{1}{6}$ gives

$$r \leqslant \frac{3-\sqrt{5}}{6} < \frac{1}{7}.$$

In particular, $B_3^{\mathrm{SY}}(u)$ lies on the entropy branch of (56). Since $h'(t) = \log \frac{1-t}{t}$ decreases on $(0, \frac{1}{2})$, (59) yields

$$h(r) - h(q) = \int_q^r h'(t)\,dt \geqslant (r-q)\log\frac{1-r}{r} > r^2 \log 6. \tag{60}$$

On the other hand,

$$h(r) - B_3^{\mathrm{SY}}(u) = \int_0^r \log\frac{1-t}{1-3t}\,dt \leqslant (\log \mathrm{e}) \int_0^r \frac{2t}{1-3t}\,dt < \frac{7}{4}(\log \mathrm{e})\, r^2. \tag{61}$$

Here we used $\log(1+x) \leqslant (\log \mathrm{e})x$ and $1 - 3t > \frac{4}{7}$ on the integration interval. Since $\log 6 > \frac{7}{4}\log \mathrm{e}$, subtracting (61) from (60) gives $U_3^{\mathrm{PSC}}(u) < B_3^{\mathrm{SY}}(u)$. □

We now proceed to prove Theorem 4.

*Proof of Theorem 4.* Fix a strictly feasible $p$ in (58) and a code $C$ counted by $A_{2,3}(n, \lceil \delta n \rceil)$, with $n$ sufficiently large that $\lceil \delta n \rceil > 1$. By Lemma 29, every coordinate error of the block PGM satisfies $p_{\mathrm{e},i} \leqslant e_3^{\mathrm{LDPC}}(p)$, and summing these $n$ coordinate bounds gives

$$\mathbf{E}[d(\hat{\mathbf{c}}, \mathbf{c})] \leqslant n e_3^{\mathrm{LDPC}}(p).$$

Since $e_3^{\mathrm{LDPC}}(p) < \delta$, the block-success step (Corollary 18) and the fixed-channel strong converse (Theorem 7), combined exactly as in the proof of Theorem 1, give

$$\frac{1}{n}\log|C| \leqslant h\left(\frac{1}{2} - \sqrt{p(1-p)}\right) + O\left(\frac{1}{\sqrt{n}}\right),$$

where the implicit constant depends on the fixed $p, \delta$ but not on $n$ or $C$. Maximizing over $C$, taking $n \to \infty$, and then taking the infimum over strictly feasible $p$ proves $R_3(\delta) \leqslant U_3^{\mathrm{PSC}}(\delta)$.

By Lemma 29, $e_3^{\mathrm{LDPC}}(\delta) < \delta$. Continuity makes some $p > \delta$ feasible in (58). The PSC Holevo information decreases strictly on $(0, \frac{1}{2})$, so

$$U_3^{\mathrm{PSC}}(\delta) < h\left(\frac{1}{2} - \sqrt{\delta(1-\delta)}\right) = R_{\mathrm{MRRW}}(\delta).$$

The comparison with $B_3^{\mathrm{SY}}$ on $\frac{1}{6} \leqslant \delta < \frac{1}{2}$ is contained in Lemma 30. □

The comparison with the explicit Shangguan–Yang curve therefore holds throughout $\frac{1}{6} \leqslant \delta < \frac{1}{2}$. Since this curve is their $w = 3$ bound throughout $\frac{2}{5} \leqslant \delta < \frac{1}{2}$, the theorem yields an improvement on the latter interval.

## References


[Aal90] M. Aaltonen. A new upper bound on nonbinary block codes. *Discrete Mathematics*, 83(2–3):139–160, 1990. 48

[AG24] O. Alrabiah and V. Guruswami. Near-tight bounds for 3-query locally correctable binary linear codes via rainbow cycles. In *65th IEEE Annual Symposium on Foundations of Computer Science*, pages 1874–1882. IEEE, 2024. doi:10.1109/FOCS61266.2024.00112. 14

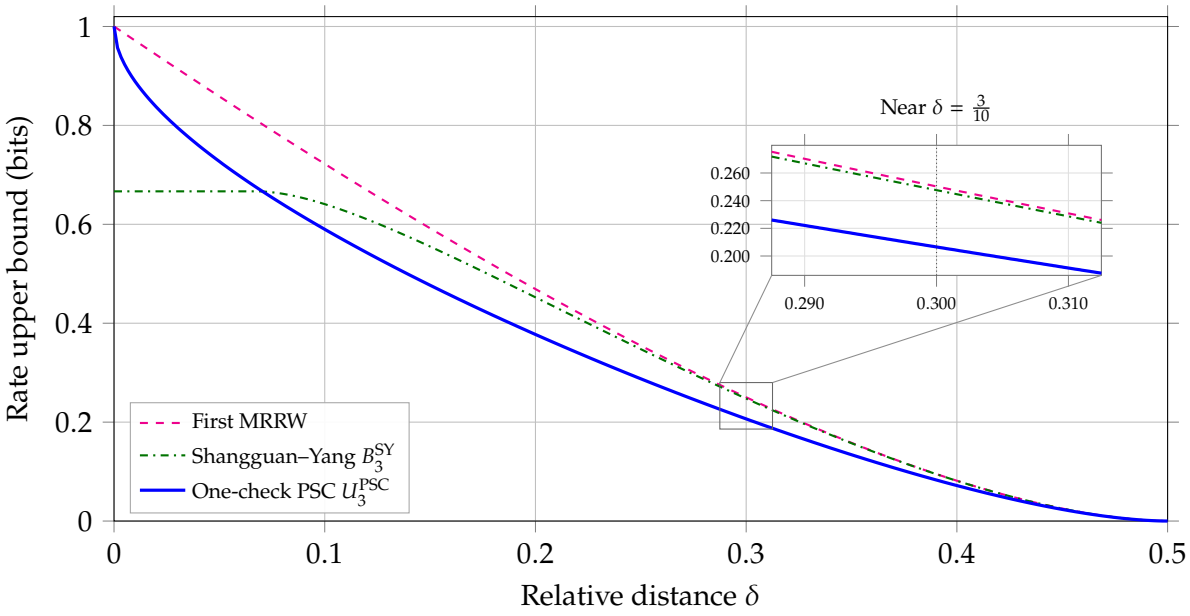


Figure 4: Rate upper bounds for binary linear codes whose duals are generated by checks of weight at most 3. The first MRRW curve [MRRW77] applies to every binary code, whereas $B_3^{SY}$ and $U_3^{PSC}$ apply to this sparse-check class. By Theorem 4 and Lemma 30, $U_3^{PSC}$ lies strictly below the first MRRW bound for $0 < \delta < \frac{1}{2}$ and below $B_3^{SY}$ for $\frac{1}{6} \leqslant \delta < \frac{1}{2}$; numerically, the latter separation begins near $\delta \approx 0.0705$. At $\delta = \frac{3}{10}$, we have $U_3^{PSC} \approx 0.2065$, $B_3^{SY} \approx 0.2476$, and first MRRW $\approx 0.2502$.


[AGL25] E. R. Anschuetz, D. Gamarnik, and J. Z. Lu. Decoded quantum interferometry requires structure. arXiv:2509.14509, 2025. 12

[AGN+26] D. Aggarwal, R. Gupta, H. H. Nguyen, K. Z. Tan, and P. N. Vasudevan. Towards worst-case hardness for low-noise LPN. arXiv:2606.05834, 2026. 11

[AK98] R. Ahlswede and L. H. Khachatrian. The diametric theorem in Hamming spaces—optimal anticodes. *Advances in Applied Mathematics*, 20(4):429–449, 1998. 48

[AL99] A. Ashikhmin and S. Litsyn. Upper bounds on the size of quantum codes. *IEEE Transactions on Information Theory*, 45(4):1206–1215, 1999. doi:10.1109/18.761270. 11

[Ari73] S. Arimoto. On the converse to the coding theorem for discrete memoryless channels. *IEEE Transactions on Information Theory*, 19(3):357–359, 1973. doi:10.1109/TIT.1973.1055007. 1, 6, 68

[Bas65] L. A. Bassalygo. New upper bounds for error-correcting codes. *Problemy Peredachi Informatsii*, 1(4):41–44, 1965. English translation in *Problems of Information Transmission* 1 (1965), no. 4, 32–35. 2, 24, 27, 39, 48

[BC25] Q. Buzet and A. Chailloux. Fine-grained unambiguous measurements. arXiv:2510.07298, 2025. 12

[BCT26] A. Blanvillain, A. Chailloux, and J.-P. Tillich. The quantum decoding problem: Tight achievability bounds and application to Regev's reduction. *IEEE Transactions on Information Theory*, 72(7):4980–4999, 2026. 9, 12, 14

[BDL16] S. Beigi, N. Datta, and F. Leditzky. Decoding quantum information via the Petz recovery map. *Journal of Mathematical Physics*, 57(8):082203, 2016. 64

[Bei13] S. Beigi. Sandwiched Rényi divergence satisfies data processing inequality. *Journal of Mathematical Physics*, 54(12):122202, 2013. 67

[Bel75] V. P. Belavkin. Optimal multiple quantum statistical hypothesis testing. *Stochastics*, 1:315–345, 1975. 18

[BH98] M. V. Burnashev and A. S. Holevo. On the reliability function for a quantum communication channel. *Problems of Information Transmission*, 34(2):97–107, 1998. Correction in "Letter to the Editor," *Problems of Information Transmission* 35 (1999), no. 1, 93–94. 12

[Bha97] Rajendra Bhatia. *Matrix Analysis*, volume 169 of *Graduate Texts in Mathematics*. Springer, New York, 1997. 63

[BHL06] Y. Ben-Haim and S. Litsyn. Upper bounds on the rate of LDPC codes as a function of minimum distance. *IEEE Transactions on Information Theory*, 52(5):2092–2100, 2006. 11

[BHL07] Y. Ben-Haim and S. Litsyn. A new upper bound on the rate of non-binary codes. *Advances in Mathematics of Communications*, 1(1):83–92, 2007. doi:10.3934/amc.2007.1.83. 48

[BJL23] Naresh Goud Boddu, Rahul Jain, and Han-Hsuan Lin. On relating one-way classical and quantum communication complexities. *Quantum*, 7:1010, 2023. 19

[BK02] H. Barnum and E. Knill. Reversing quantum dynamics with near-optimal quantum and classical fidelity. *Journal of Mathematical Physics*, 43(5):2097–2106, 2002. Corrected version available as arXiv:quant-ph/0004088v2. 6, 12, 64

[BKMH97] M. Ban, K. Kurokawa, R. Momose, and O. Hirota. Optimum measurements for discrimination among symmetric quantum states and parameter estimation. *International Journal of Theoretical Physics*, 36(6):1269–1288, 1997. doi:10.1007/BF02435921. 3, 12

[Bla77] R. E. Blahut. Composition bounds for channel block codes. *IEEE Transactions on Information Theory*, 23(6):656–674, 1977. doi:10.1109/TIT.1977.1055792. 2

[BLVW19] Z. Brakerski, V. Lyubashevsky, V. Vaikuntanathan, and D. Wichs. Worst-case hardness for LPN and cryptographic hashing via code smoothing. In *Advances in Cryptology—EUROCRYPT 2019, Part III*, volume 11478 of *Lecture Notes in Computer Science*, pages 619–635. Springer, 2019. doi:10.1007/978-3-030-17659-4_21. 11

[CDA25] A. Chailloux and T. Debris-Alazard. New solutions to Delsarte's dual linear programs. *IEEE Transactions on Information Theory*, 71(1):297–316, 2025. 11, 48

[CE03] H. Cohn and N. Elkies. New upper bounds on sphere packings I. *Annals of Mathematics*, 157(2):689–714, 2003. 13

[CJJ22] L. N. Coregliano, F. G. Jeronimo, and C. Jones. A complete linear programming hierarchy for linear codes. In *13th Innovations in Theoretical Computer Science Conference (ITCS 2022)*, volume 215 of *LIPIcs*, pages 51:1–51:22, 2022. 11

[CJJ+26] L. N. Coregliano, F. G. Jeronimo, C. Jones, N. Linial, and E. Loyfer. Higher-order Delsarte dual LPs: Lifting, constructions and completeness. In *17th Innovations in Theoretical Computer Science Conference (ITCS 2026)*, volume 362 of *LIPIcs*, pages 44:1–44:22, 2026. 11

[CLZ22] Yilei Chen, Qipeng Liu, and Mark Zhandry. Quantum algorithms for variants of average-case lattice problems via filtering. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 372–401. Springer, 2022. 14

[CPS20] C.-F. Chen, G. Penington, and G. Salton. Entanglement wedge reconstruction using the Petz map. *Journal of High Energy Physics*, 2020(1):168, 2020. 64

[CRSS98] A. R. Calderbank, E. M. Rains, P. W. Shor, and N. J. A. Sloane. Quantum error correction via codes over GF(4). *IEEE Transactions on Information Theory*, 44(4):1369–1387, 1998. doi:10.1109/18.681315. 11

[CS96] A. R. Calderbank and P. W. Shor. Good quantum error-correcting codes exist. *Physical Review A*, 54(2):1098–1105, 1996. doi:10.1103/PhysRevA.54.1098. 11

[CT24] A. Chailloux and J.-P. Tillich. The quantum decoding problem. In *19th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2024)*, volume 310 of *LIPIcs*, pages 6:1–6:14, 2024. Full version: arXiv:2310.20651. 9, 12, 14

[CT25] A. Chailloux and J.-P. Tillich. Quantum advantage from soft decoders. In *Proceedings of the 57th Annual ACM Symposium on Theory of Computing*, pages 738–749, 2025. doi:10.1145/3717823.3718319. 12, 14

[DADRT23] T. Debris-Alazard, L. Ducas, N. Resch, and J.-P. Tillich. Smoothing codes and lattices: Systematic study and new bounds. *IEEE Transactions on Information Theory*, 69(9):6006–6027, 2023. doi:10.1109/TIT.2023.3276921. 11

[Dal13] M. Dalai. Lower bounds on the probability of error for classical and classical–quantum channels. *IEEE Transactions on Information Theory*, 59(12):8027–8056, 2013. 12, 13

[Dal15] M. Dalai. Elias bound for general distances and stable sets in edge-weighted graphs. *IEEE Transactions on Information Theory*, 61(5):2335–2350, 2015. doi:10.1109/TIT.2015.2410782; arXiv:1406.1308. 2

[DAR25] T. Debris-Alazard and N. Resch. Worst and average case hardness of decoding via smoothing bounds. In *Public-Key Cryptography—PKC 2025, Part II*, volume 15675 of *Lecture Notes in Computer Science*, pages 363–392. Springer, 2025. doi:10.1007/978-3-031-91823-0_12; IACR ePrint 2022/1744. 11

[DART24] T. Debris-Alazard, M. Remaud, and J.-P. Tillich. Quantum reduction of finding short code vectors to the decoding problem. *IEEE Transactions on Information Theory*, 70(7):5323–5342, 2024. 12, 14

[DdW11] A. Drucker and R. de Wolf. Quantum proofs for classical theorems. *Theory of Computing Library Graduate Surveys*, 2:1–54, 2011. 13

[Del72] P. Delsarte. Bounds for unrestricted codes, by linear programming. *Philips Research Reports*, 27(3):272–289, 1972. 11, 48

[Del73] P. Delsarte. *An Algebraic Approach to the Association Schemes of Coding Theory*. Number 10 in Philips Research Reports Supplements. Philips Research Laboratories, 1973. 11, 48

[DGL26] M. Dalai, F. Girardi, and L. Lami. Zero-error list decoding for classical–quantum channels. arXiv:2601.09786, 2026. 12

[DGS77] P. Delsarte, J.-M. Goethals, and J. J. Seidel. Spherical codes and designs. *Geometriae Dedicata*, 6(3):363–388, 1977. 13

[DL98] P. Delsarte and V. I. Levenshtein. Association schemes and coding theory. *IEEE Transactions on Information Theory*, 44(6):2477–2504, 1998. doi:10.1109/18.720545. 48

[DW17] M. Dalai and A. Winter. Constant compositions in the sphere packing bound for classical–quantum channels. *IEEE Transactions on Information Theory*, 63(9):5603–5617, 2017. 12

[EF01] Y. C. Eldar and G. D. Forney, Jr. On quantum detection and the square-root measurement. *IEEE Transactions on Information Theory*, 47(3):858–872, 2001. 3, 12

[Eli55] P. Elias. Coding for noisy channels. *IRE Convention Record*, 3(part 4):37–46, 1955. 2, 24, 27, 39, 48

[Eli57] P. Elias. List decoding for noisy channels. Technical Report 335, Research Laboratory of Electronics, Massachusetts Institute of Technology, 1957. 1

[Eli91] P. Elias. Error-correcting codes for list decoding. *IEEE Transactions on Information Theory*, 37(1):5–12, 1991. 1

[EMV04] Y. C. Eldar, A. Megretski, and G. C. Verghese. Optimal detection of symmetric mixed quantum states. *IEEE Transactions on Information Theory*, 50(6):1198–1207, 2004. doi:10.1109/TIT.2004.828070; arXiv:quant-ph/0211111. 3, 12

[Fan61] R. M. Fano. *Transmission of Information: A Statistical Theory of Communication*. MIT Press, Cambridge, MA, 1961. 6

[FL13] R. L. Frank and E. H. Lieb. Monotonicity of a relative Rényi entropy. *Journal of Mathematical Physics*, 54(12):122201, 2013. 67

[FT05] J. Friedman and J.-P. Tillich. Generalized Alon–Boppana theorems and error-correcting codes. *SIAM Journal on Discrete Mathematics*, 19(3):700–718, 2005. 9, 10, 11

[Gal62] R. G. Gallager. Low-density parity-check codes. *IRE Transactions on Information Theory*, 8(1):21–28, 1962. doi:10.1109/TIT.1962.1057683. 11

[Gil52] E. N. Gilbert. A comparison of signalling alphabets. *Bell System Technical Journal*, 31(3):504–522, 1952. 1

[GLM+22] A. Gilyén, S. Lloyd, I. Marvian, Y. Quek, and M. M. Wilde. Quantum algorithm for Petz recovery channels and pretty good measurements. *Physical Review Letters*, 128:220502, 2022. 13, 64

[Gur04] V. Guruswami. *List Decoding of Error-Correcting Codes*, volume 3282 of *Lecture Notes in Computer Science*. Springer, 2004. 1

[GZ08] P. Gaborit and G. Zémor. Asymptotic improvement of the Gilbert–Varshamov bound for linear codes. *IEEE Transactions on Information Theory*, 54(9):3865–3872, 2008. doi:10.1109/TIT.2008.928288; arXiv:0708.4164. 1

[Ham50] R. W. Hamming. Error detecting and error correcting codes. *Bell System Technical Journal*, 29(2):147–160, 1950. 1, 8

[Hel69] C. W. Helstrom. Quantum detection and estimation theory. *Journal of Statistical Physics*, 1:231–252, 1969. 12, 49

[HJS+96] P. Hausladen, R. Jozsa, B. Schumacher, M. Westmoreland, and W. K. Wootters. Classical information capacity of a quantum channel. *Physical Review A*, 54:1869–1876, 1996. 17, 18

[Hoe63] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963. 70

[Hol73a] A. S. Holevo. Bounds for the quantity of information transmitted by a quantum communication channel. *Problems of Information Transmission*, 9(3):177–183, 1973. Russian original in *Problemy Peredachi Informatsii* 9 (1973), no. 3, 3–11. 6, 8, 16

[Hol73b] A. S. Holevo. Statistical decision theory for quantum systems. *Journal of Multivariate Analysis*, 3(4):337–394, 1973. doi:10.1016/0047-259X(73)90028-6. 12

[Hol79] Alexander S Holevo. On asymptotically optimal hypothesis testing in quantum statistics. *Theory of Probability & Its Applications*, 23(2):411–415, 1979. 18

[Hol98] A. S. Holevo. The capacity of the quantum channel with general signal states. *IEEE Transactions on Information Theory*, 44(1):269–273, 1998. 17

[HSS21] J. Hązła, A. Samorodnitsky, and O. Sberlo. On codes decoding a constant fraction of errors on the BSC. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 1479–1488. ACM, 2021. doi:10.1145/3406325.3451015. 2

[HW94] P. Hausladen and W. K. Wootters. A "pretty good" measurement for distinguishing quantum states. *Journal of Modern Optics*, 41(12):2385–2390, 1994. 12, 18

[IS15] E. Iceland and A. Samorodnitsky. On coset leader graphs of LDPC codes. *IEEE Transactions on Information Theory*, 61(8):4158–4163, 2015. 5, 11

[IS20] Eran Iceland and Alex Samorodnitsky. On coset leader graphs of structured linear codes. *Discrete & Computational Geometry*, 63(3):560–576, 2020. 14

[Joh62] S. M. Johnson. A new upper bound for error-correcting codes. *IRE Transactions on Information Theory*, 8(3):203–207, 1962. 4, 6, 8

[JRS+18] M. Junge, R. Renner, D. Sutter, M. M. Wilde, and A. Winter. Universal recovery maps and approximate sufficiency of quantum relative entropy. *Annales Henri Poincaré*, 19(10):2955–2978, 2018. 64

[JSW+25] S. P. Jordan, N. Shutty, M. Wootters, A. Zalcman, A. Schmidhuber, R. King, S. V. Isakov, T. Khattar, and R. Babbush. Optimization by decoded quantum interferometry. *Nature*, 646:831–836, 2025. 12, 14

[JV04] T. Jiang and A. Vardy. Asymptotic improvement of the Gilbert–Varshamov bound on the size of binary codes. *IEEE Transactions on Information Theory*, 50(8):1655–1664, 2004. 1

[Kai18] K. Kaipa. An improvement of the asymptotic Elias bound for non-binary codes. *IEEE Transactions on Information Theory*, 64(7):5170–5178, 2018. 48

[KdW04] I. Kerenidis and R. de Wolf. Exponential lower bound for 2-query locally decodable codes via a quantum argument. *Journal of Computer and System Sciences*, 69(3):395–420, 2004. 13, 14

[KL78] G. A. Kabatiansky and V. I. Levenshtein. On bounds for packings on a sphere and in space. *Problems of Information Transmission*, 14(1):1–17, 1978. 13

[KM24a] P. K. Kothari and P. Manohar. An exponential lower bound for linear 3-query locally correctable codes. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pages 776–787. ACM, 2024. 14

[KM24b] Pravesh K Kothari and Peter Manohar. Exponential lower bounds for smooth 3-LCCs and sharp bounds for designs. In *2024 IEEE 65th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1802–1845. IEEE, 2024. 14

[KOW26] Robin Kothari, Ryan O'Donnell, and Kewen Wu. No exponential quantum speedup for $\mathrm{SIS}^\infty$ anymore. In *Proceedings of the 58th Annual ACM Symposium on Theory of Computing*, pages 101–105, 2026. 14

[KT08] R. König and B. M. Terhal. The bounded-storage model in the presence of a quantum adversary. *IEEE Transactions on Information Theory*, 54(2):749–762, 2008. 19

[KYH26] D. Kougang-Yombi and J. Hązła. Weight distribution bounds to relate minimum distance, list decoding, and symmetric channel performance. arXiv:2604.02994, 2026. 2, 6

[LdW25] H.-H. Lin and R. de Wolf. Getting almost all the bits from a quantum random access code. arXiv:2506.01903, 2025. 13, 19

[Lev95] V. I. Levenshtein. Krawtchouk polynomials and universal bounds for codes and designs in Hamming spaces. *IEEE Transactions on Information Theory*, 41(5):1303–1321, 1995. 11, 48

[Lin75] G. Lindblad. Completely positive maps and entropy inequalities. *Communications in Mathematical Physics*, 40(2):147–151, 1975. 8

[LL98] T. Laihonen and S. Litsyn. On upper bounds for minimum distance and covering radius of non-binary codes. *Designs, Codes and Cryptography*, 14(1):71–80, 1998. doi:10.1023/A:1008260505585. 48

[LL23a] N. Linial and E. Loyfer. An elementary proof of the first LP bound on the rate of binary codes. arXiv:2303.16619, 2023. 11

[LL23b] E. Loyfer and N. Linial. New LP-based upper bounds in the rate-vs.-distance problem for binary linear codes. *IEEE Transactions on Information Theory*, 69(5):2886–2899, 2023. 11

[Lov79] L. Lovász. On the Shannon capacity of a graph. *IEEE Transactions on Information Theory*, 25(1):1–7, 1979. 11

[MFGH25] K. Marwaha, B. Fefferman, A. Gheorghiu, and V. Havlicek. On the complexity of decoded quantum interferometry. arXiv:2509.14443, 2025. 12, 14

[MLDS+13] M. Müller-Lennert, F. Dupuis, O. Szehr, S. Fehr, and M. Tomamichel. On quantum Rényi entropies: A new generalization and some properties. *Journal of Mathematical Physics*, 54(12):122203, 2013. 67

[MO17] M. Mosonyi and T. Ogawa. Strong converse exponent for classical–quantum channel coding. *Communications in Mathematical Physics*, 355:373–426, 2017. 3, 17, 18

[MP26] A. Mandal and H. D. Pfister. Quantum message passing for factor graphs over finite abelian groups. arXiv:2604.12186, 2026. 12, 13

[MRR78] R. J. McEliece, E. R. Rodemich, and H. Rumsey, Jr. The Lovász bound and some generalizations. *Journal of Combinatorics, Information & System Sciences*, 3(3):134–152, 1978. 11

[MRRW77] R. J. McEliece, E. R. Rodemich, H. Rumsey, Jr., and L. R. Welch. New upper bounds on the rate of a code via the Delsarte–MacWilliams inequalities. *IEEE Transactions on Information Theory*, 23(2):157–166, 1977. 1, 11, 14, 24, 28, 30, 39, 48, 53

[MSPJ26] A. Mandal, N. Shutty, H. D. Pfister, and S. P. Jordan. Affine filtering measurements and their applications to quantum decoding. arXiv:2606.07852, 2026. 12

[Nay99] A. Nayak. Optimal lower bounds for quantum automata and random access codes. In *40th Annual Symposium on Foundations of Computer Science (FOCS 1999)*, pages 369–376, 1999. 13

[NC10] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 10th anniversary edition, 2010. 15, 66

[NS05] M. Navon and A. Samorodnitsky. On Delsarte's linear programming bounds for binary codes. In *46th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2005)*, pages 327–336, 2005. doi:10.1109/SFCS.2005.55. 6

[NS09] M. Navon and A. Samorodnitsky. Linear programming bounds for codes via a covering argument. *Discrete & Computational Geometry*, 41(2):199–207, 2009. 9, 10, 11, 14

[ON99] T. Ogawa and H. Nagaoka. Strong converse to the quantum channel coding theorem. *IEEE Transactions on Information Theory*, 45:2486–2489, 1999. 3, 6, 17, 18

[Ope26] OpenAI. Improved Bounds for Binary and Spherical Codes. Chapter 2 of *Ten Advances in Mathematics and Theoretical Computer Science*, 2026. Available at `https://cdn.openai.com/pdf/ten-proofs-oai.pdf`. 13

[PB23] M. Pathegama and A. Barg. Smoothing of binary codes, uniform distributions, and applications. *Entropy*, 25(11):1515, 2023. doi:10.3390/e25111515. 11

[PB25] M. Pathegama and A. Barg. Limitations of the decoding-to-LPN reduction via code smoothing. *Designs, Codes and Cryptography*, 93(7):2761–2778, 2025. doi:10.1007/s10623-025-01617-9; arXiv:2408.03742. 11

[Pet84] D. Petz. A dual in von Neumann algebras with weights. *Quarterly Journal of Mathematics*, 35(4):475–483, 1984. 64

[Pet86] D. Petz. Sufficient subalgebras and the relative entropy of states of a von Neumann algebra. *Communications in Mathematical Physics*, 105(1):123–131, 1986. 64

[Pet88] D. Petz. Sufficiency of channels over von Neumann algebras. *Quarterly Journal of Mathematics*, 39(1):97–108, 1988. 64

[Plo60] M. Plotkin. Binary codes with specified minimum distance. *IRE Transactions on Information Theory*, 6(4):445–450, 1960. 1, 2, 24, 25, 39, 48

[PR25] C. Piveteau and J. M. Renes. Efficient and optimal quantum state discrimination via quantum belief propagation. arXiv:2509.19441, 2025. 3, 12, 13, 14, 49

[PSW24] F. Pernice, O. Sprumont, and M. Wootters. List-decoding capacity implies capacity on the $q$-ary symmetric channel. arXiv:2410.20020, 2024. 2, 6, 14

[Rai99] E. M. Rains. Quantum shadow enumerators. *IEEE Transactions on Information Theory*, 45(7):2361–2366, 1999. doi:10.1109/18.796376. 11

[Reg09] O. Regev. On lattices, learning with errors, random linear codes, and cryptography. *Journal of the ACM*, 56(6):34:1–34:40, 2009. Article 34. 12, 14

[Ren17] J. M. Renes. Belief propagation decoding of quantum channels by passing quantum messages. *New Journal of Physics*, 19:072001, 2017. 12, 13

[Ren18] J. M. Renes. Duality of channels and codes. *IEEE Transactions on Information Theory*, 64(1):577–592, 2018. 9, 12

[Ren22] J. M. Renes. *Quantum Information Theory: Concepts and Methods*. De Gruyter Oldenbourg, 2022. 15, 18, 19, 64

[RP21] N. Rengaswamy and H. D. Pfister. On the duality between the BSC and quantum PSC. In *2021 IEEE International Symposium on Information Theory (ISIT)*, pages 2232–2237, 2021. 3, 9, 12

[Sal24] J. Salez. Cutoff for non-negatively curved Markov chains. *Journal of the European Mathematical Society*, 26(11):4375–4392, 2024. 14

[Sam01] A. Samorodnitsky. On the optimum of Delsarte's linear program. *Journal of Combinatorial Theory, Series A*, 96(2):261–287, 2001. 6

[Sam23] A. Samorodnitsky. One more proof of the first linear programming bound for binary codes and two conjectures. *Israel Journal of Mathematics*, 2023. Also available as arXiv:2104.14587. 11

[Sam25] A. Samorodnitsky. On the difficulty to beat the first linear programming bound for binary codes. *IEEE Transactions on Information Theory*, 71(4):2383–2388, 2025. 6, 11

[Sch79] A. Schrijver. A comparison of the Delsarte and Lovász bounds. *IEEE Transactions on Information Theory*, 25(4):425–429, 1979. 11

[Sch05] A. Schrijver. New code upper bounds from the Terwilliger algebra and semidefinite programming. *IEEE Transactions on Information Theory*, 51(8):2859–2866, 2005. 11

[Sha48] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 623–656, 1948. 1

[Sin64] R. C. Singleton. Maximum distance $q$-nary codes. *IEEE Transactions on Information Theory*, 10(2):116–118, 1964. 8

[Ste96] A. M. Steane. Multiple-particle interference and quantum error correction. *Proceedings of the Royal Society of London. Series A*, 452(1954):2551–2577, 1996. doi:10.1098/rspa.1996.0136. 11

[STH16] D. Sutter, M. Tomamichel, and A. W. Harrow. Strengthened monotonicity of relative entropy via pinched Petz recovery map. *IEEE Transactions on Information Theory*, 62(5):2907–2913, 2016. 64

[Str64] Volker Strassen. Asymptotische Abschätzungen in Shannons Informationstheorie. In *Transactions of the Third Prague Conference on Information Theory, Statistical Decision Functions, Random Processes*, pages 689–723, Prague, 1964. Publishing House of the Czechoslovak Academy of Sciences. Conference held in Liblice, Czechoslovakia, in 1962. 1, 6, 68

[SW97] B. Schumacher and M. D. Westmoreland. Sending classical information via noisy quantum channels. *Physical Review A*, 56:131–138, 1997. 17

[SY26] C. Shangguan and Y. Yang. Improved rate-versus-distance upper bounds for LDPC codes. arXiv:2605.01213, 2026. 5, 11, 48

[Tie90] A. A. Tietäväinen. An upper bound on the covering radius as a function of the dual distance. *IEEE Transactions on Information Theory*, 36(6):1472–1474, 1990. 8, 9, 14

[TVZ82] M. A. Tsfasman, S. G. Vlăduţ, and Th. Zink. Modular curves, Shimura curves, and Goppa codes, better than Varshamov–Gilbert bound. *Mathematische Nachrichten*, 109(1):21–28, 1982. doi:10.1002/mana.19821090103. 1

[TZ00] J.-P. Tillich and G. Zémor. Discrete isoperimetric inequalities and the probability of a decoding error. *Combinatorics, Probability and Computing*, 9(5):465–479, 2000. 2, 6, 14

[Uhl77] A. Uhlmann. Relative entropy and the Wigner–Yanase–Dyson–Lieb concavity in an interpolation theory. *Communications in Mathematical Physics*, 54(1):21–32, 1977. 8

[UTHH99] T. S. Usuda, I. Takumi, M. Hata, and O. Hirota. Minimum error detection of classical linear code sending through a quantum channel. *Physics Letters A*, 256(2–3):104–108, 1999. doi:10.1016/S0375-9601(99)00232-7. 3, 12

[Var57] R. R. Varshamov. Estimate of the number of signals in error correcting codes. *Doklady Akademii Nauk SSSR*, 117:739–741, 1957. 1

[WG13] M. M. Wilde and S. Guha. Polar codes for classical–quantum channels. *IEEE Transactions on Information Theory*, 59(2):1175–1187, 2013. 12

[Wil21] M. M. Wilde. From Classical to Quantum Shannon Theory. Second-edition prepublication draft, December 2021, https://www.markwilde.com/qit-notes.pdf, 2021. An earlier version is available as arXiv:1106.1445v8, 2019. 15, 63

[Win99] A. Winter. Coding theorem and strong converse for quantum channels. *IEEE Transactions on Information Theory*, 45:2481–2485, 1999. Also available as arXiv:1409.2536. 3, 6, 17, 18

[Wol57] J. Wolfowitz. The coding of messages subject to chance errors. *Illinois Journal of Mathematics*, 1(4):591–606, 1957. 1, 6, 68

[Wol12] M. M. Wolf. Quantum Channels & Operations: Guided Tour. Lecture notes, https://mediatum.ub.tum.de/doc/1701036/1701036.pdf, 2012. 66

[Woz58] J. M. Wozencraft. List decoding. Quarterly Progress Report 48, Research Laboratory of Electronics, Massachusetts Institute of Technology, 1958. 1

[Wri24] John Wright. Quantum Learning Theory. Course notes for CS 294: Quantum Learning Theory, University of California, Berkeley, https://people.eecs.berkeley.edu/~jswright/quantumlearningtheory24/, 2024. 18, 50

[WWY14] M. M. Wilde, A. Winter, and D. Yang. Strong converse for the classical capacity of entanglement-breaking and Hadamard channels via a sandwiched Rényi relative entropy. *Communications in Mathematical Physics*, 331(2):593–622, 2014. 67

[Yan24] Tal Yankovitz. A stronger bound for linear 3-LCC. In *2024 IEEE 65th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1786–1801. IEEE, 2024. 14

[YKL75] H. P. Yuen, R. S. Kennedy, and M. Lax. Optimum testing of multiple hypotheses in quantum detection theory. *IEEE Transactions on Information Theory*, 21(2):125–134, 1975. doi:10.1109/TIT.1975.1055351. 12

[YZ21] Y. Yu and J. Zhang. Smoothing out binary linear codes and worst-case sub-exponential hardness for LPN. In *Advances in Cryptology—CRYPTO 2021, Part III*, volume 12827 of *Lecture Notes in Computer Science*, pages 473–501. Springer, 2021. doi:10.1007/978-3-030-84252-9_16. 11

[ZCCL25] J. Zhou, S. Chessa, E. Chitambar, and F. Leditzky. On the distinguishability of geometrically uniform quantum states. *Journal of Physics A: Mathematical and Theoretical*, 58:415303, 2025. doi:10.1088/1751-8121/ae0a95; arXiv:2501.12376. 3, 12

[ZP81] V. V. Zyablov and M. S. Pinsker. List cascade decoding. *Problemy Peredachi Informatsii*, 17(4):29–34, 1981. 1

# A Proofs for the preliminaries

In this appendix, for the sake of completeness, we present the non-immediate proofs deferred from Section 2, particularly the PGM error data processing inequality (Lemma 10) and the cq channel coding strong converse (Theorem 7), as those are the two crucial inequalities needed in the proof of the pretty good criterion (Theorem 1).

To this end, we recall some necessary quantum notions and tools. Building on the notion of a density matrix, a $d \times d$ matrix is said to be a *subnormalized state* if it is a positive semidefinite matrix whose trace is at most 1. For any finite-dimensional Hilbert space $V$, let $\mathcal{M}(V)$ denote the vector space of all Hermitian matrices over $V$.

Next, we recall some standard matrix analysis tools needed in the proof of the cq channel coding strong converse in Section A.4. We denote the trace norm of a $d \times d$ complex-valued matrix $X$ by $\|X\|_1 \coloneqq \operatorname{Tr}\sqrt{X^\dagger X}$, the Frobenius norm by $\|X\|_2 \coloneqq \sqrt{\operatorname{Tr}(X^\dagger X)}$, and the spectral norm by $\|X\|_\infty \coloneqq \sqrt{\lambda_{\max}(X^\dagger X)}$. Throughout, matrix inequalities $A \preceq B$ between Hermitian matrices mean that $B - A$ is positive semidefinite. Note that for diagonal density matrices, i.e., classical probability distributions, the quantity $\frac{1}{2}\|\tau - \tau'\|_1$ is the total-variation distance between the probability distributions $\tau$ and $\tau'$.

We now recall a standard inequality that is needed for the proof of the cq channel coding strong converse (Section A.4), which is the gentle measurement lemma, stated in [Wil21, Lemma 9.4.2 and Exercise 9.4.1]. The gentle measurement lemma controls how much a test that passes with high probability disturbs such a state. Classically, conditioning on an event of probability at least $1 - \varepsilon$ moves a distribution by $O(\varepsilon)$ in total variation. In the quantum case, that estimate becomes $O(\sqrt{\varepsilon})$.

**Lemma 31** (Gentle measurement lemma)**.** *If $A$ is a subnormalized $d \times d$ state and $P$ is a projection $d \times d$ matrix, then*

$$\|A - PAP\|_1 \leqslant 2\sqrt{\operatorname{Tr}[A(I-P)]}.$$

*Proof.* Observe that $I - P$ and $P$ are both projection matrices. Combining this observation with the triangle inequality, the matrix Cauchy–Schwarz inequality [Bha97, Corollary IV.2.6], and the inequalities $\operatorname{Tr}(A) \leqslant 1$ and $\max\{\operatorname{Tr}(PA), \operatorname{Tr}((I-P)A)\} \leqslant \operatorname{Tr}(A)$ (as $A$ is a subnormalized state), we find that

$$\begin{aligned}
\|A - PAP\|_1 &= \|(I-P)A + PA(I-P)\|_1 \\
&\leqslant \|(I-P)A\|_1 + \|PA(I-P)\|_1 \\
&\leqslant \|(I-P)A^{1/2}\|_2 \cdot \|A^{1/2}\|_2 + \|PA^{1/2}\|_2 \cdot \|A^{1/2}(I-P)\|_2 \\
&= \sqrt{\operatorname{Tr}[A(I-P)]} \cdot \sqrt{\operatorname{Tr}(A)} + \sqrt{\operatorname{Tr}(PA)} \cdot \sqrt{\operatorname{Tr}[A(I-P)]} \\
&\leqslant \sqrt{\operatorname{Tr}[A(I-P)]} \cdot 1 + 1 \cdot \sqrt{\operatorname{Tr}[A(I-P)]}
\end{aligned}$$

$$= 2\sqrt{\mathrm{Tr}[A(I-P)]}. \qquad \square$$

## A.1 Petz recovery and its structural properties

In this subsection, we develop the Petz recovery map [Pet84, Pet86, Pet88, Ren22], a generalization of the PGM, and prove its structural properties (Fact 32), en route to proving the PGM data-processing lemma (Lemma 10) in Section A.2. A Petz map reverses an arbitrary quantum channel relative to a chosen reference state. When the channel is a cq preparation channel and the reference is its classical prior, the Petz map reduces to the PGM [BK02]. This identification explains the connection with the PGM.

As a brief digression, recent work uses Petz maps and their variants to quantify approximate recovery in strengthened data-processing inequalities [STH16, JRS+18], to construct explicit decoders for quantum communication [BDL16], to reconstruct holographic entanglement wedges [CPS20], and to design quantum algorithms for Petz channels and PGMs [GLM+22].

Returning to the subject matter, we now formally introduce the Petz map.

**Definition A.1** (Petz recovery map)**.** *Let $\Phi : \mathbb{C}^{d_A \times d_A} \to \mathbb{C}^{d_B \times d_B}$ be a quantum channel, and let $\pi \in \mathsf{D}(\mathbb{C}^{d_A})$ be a reference state. Let $\nu := \Phi(\pi) \in \mathsf{D}(\mathbb{C}^{d_B})$. The **Petz recovery map** of $\Phi$ relative to $\pi$ is the linear map $\Phi_\pi^\sharp : \mathbb{C}^{d_B \times d_B} \to \mathbb{C}^{d_A \times d_A}$ defined as*

$$\Phi_\pi^\sharp(Y) := \pi^{1/2} \Phi^* \left( \nu^{-1/2} Y \nu^{-1/2} \right) \pi^{1/2},$$

*where $\Phi^*$ is determined by the identity $\mathrm{Tr}[X^\dagger \Phi^*(Y)] = \mathrm{Tr}[\Phi(X)^\dagger Y]$, and inverse powers act on the indicated supports. Moreover, for $A_1, A_2 \in \mathcal{M}(\mathrm{supp}(\pi))$, define the weighted inner product*

$$\langle A_1, A_2 \rangle_\pi := \mathrm{Tr}\left( A_1^\dagger \pi^{-1/2} A_2 \pi^{-1/2} \right).$$

*Since $\langle A, A \rangle_\pi = \| \pi^{-1/4} A \pi^{-1/4} \|_2^2$, this weighted form is therefore positive definite over $\mathcal{M}(\mathrm{supp}(\pi))$.*

*Remark* A.1 (Petz in the classical case). Let $W(y \mid x)$ be any classical channel with input and© output alphabets $\mathcal{X}$ and $\mathcal{Y}$, let $\pi(x)$ be an input distribution, and let $\nu(y) := \sum_{x \in \mathcal{X}} \pi(x) W(y \mid x)$ be the distribution of $\pi$ when passed through $W$. For every output $y$ with $\nu(y) > 0$, the Petz recovery map is the Bayes reverse channel

$$W_\pi^\sharp(x \mid y) := \frac{\pi(x) W(y \mid x)}{\nu(y)} = \mathbf{Pr}[\mathbf{x} = x \mid \mathbf{y} = y].$$

Thus, Petz recovery is the quantum form of reversing a channel by Bayes' rule relative to a fixed prior. $\triangleleft$

The proof of PGM data processing in Section A.2 uses the following three structural properties of this recovery map.

**Fact 32** (Petz adjointness, functoriality, and contraction)**.** *Following the notation given in Definition A.1, when restricted to $\mathcal{M}(\mathrm{supp}(\nu))$, the Petz recovery map $\Phi_\pi^\sharp$ is a quantum channel with output density matrices residing in $\mathcal{M}(\mathrm{supp}(\pi))$. Moreover, the following three properties hold:*

(i) (Weighted adjointness). *For all $A \in \mathcal{M}(\mathrm{supp}(\pi))$ and $B \in \mathcal{M}(\mathrm{supp}(\nu))$, we have the identity*

$$\langle \Phi(A), B \rangle_\nu = \left\langle A, \Phi_\pi^\sharp(B) \right\rangle_\pi.$$

(ii) (Contravariant functoriality). *For any quantum channel* $\Psi : \mathbb{C}^{d_B \times d_B} \to \mathbb{C}^{d_C \times d_C}$, *we have the composition identity*

$$(\Psi \circ \Phi)_\pi^\sharp = \Phi_\pi^\sharp \circ \Psi_\nu^\sharp.$$

(iii) (Petz–Gram contraction). *For all* $A_1, A_2 \in \mathcal{M}(\mathrm{supp}(\pi))$, *the quantum channel* $K \coloneqq \Phi_\pi^\sharp \circ \Phi$ *satisfies*

$$\langle A_1, K(A_2)\rangle_\pi = \langle \Phi(A_1), \Phi(A_2)\rangle_\nu .$$

*In particular, for any* $A \in \mathbb{C}^{d_A \times d_A}$, *we have*

$$0 \leqslant \langle A, K(A)\rangle_\pi \leqslant \langle A, A\rangle_\pi . \tag{62}$$

*Proof of Fact 32.* For every $B \in \mathcal{M}(\mathrm{supp}(\nu))$, from Definition 2.1, we have that $B\Pi_\nu = B$. Combining this observation with the cyclicity of the trace and the definition of $\Phi^*$ in Definition A.1, we therefore find that

$$\begin{aligned}
\mathrm{Tr}(\Phi_\pi^\sharp(B)) &= \mathrm{Tr}\left[\Phi^*\left(\nu^{-1/2} B \nu^{-1/2}\right)\pi\right] \\
&= \mathrm{Tr}\left[\nu^{-1/2} B \nu^{-1/2} \Phi(\pi)\right] \\
&= \mathrm{Tr}\left[B \nu^{-1/2} \Phi(\pi) \nu^{-1/2}\right] \\
&= \mathrm{Tr}(B\Pi_\nu) \\
&= \mathrm{Tr}(B).
\end{aligned}$$

Hence, $\Phi_\pi^\sharp$ is trace-preserving on $\mathcal{M}(\mathrm{supp}(\nu))$. Since the maps $B \mapsto \nu^{-1/2} B \nu^{-1/2}$, $A \mapsto \pi^{1/2} A \pi^{1/2}$, and $\Phi^*$ are all completely positive,[26] the closure of completely positive maps under composition therefore shows that $\Phi_\pi^\sharp$ is completely positive. Thus, $\Phi_\pi^\sharp$ is a quantum channel when restricted to $\mathcal{M}(\mathrm{supp}(\nu))$. The fact that its output states lie in $\mathcal{M}(\mathrm{supp}(\pi))$ is easily seen from the definition of $\Phi_\pi^\sharp$.

For Fact 32(i), by the defining identity for $\Phi^*$ given in Definition A.1 and the cyclicity property of the trace map, we have that

$$\begin{aligned}
\left\langle A, \Phi_\pi^\sharp(B)\right\rangle_\pi &= \mathrm{Tr}\left[A^\dagger \pi^{-1/2} \pi^{1/2} \Phi^*\left(\nu^{-1/2} B \nu^{-1/2}\right) \pi^{1/2} \pi^{-1/2}\right] \\
&= \mathrm{Tr}\left[\Pi_\pi A^\dagger \Pi_\pi \Phi^*\left(\nu^{-1/2} B \nu^{-1/2}\right)\right] \\
&= \mathrm{Tr}\left[\Phi(A)^\dagger \nu^{-1/2} B \nu^{-1/2}\right] \\
&= \langle \Phi(A), B\rangle_\nu .
\end{aligned}$$

For Fact 32(ii), define $\mu \coloneqq \Psi(\nu)$. For $C \in \mathsf{D}(\mathrm{supp}(\mu))$, applying the preceding adjointness identity twice gives us

$$\langle (\Psi \circ \Phi)(A), C\rangle_\mu = \left\langle \Phi(A), \Psi_\nu^\sharp(C)\right\rangle_\nu = \left\langle A, \Phi_\pi^\sharp \Psi_\nu^\sharp(C)\right\rangle_\pi ,$$

which proves the functoriality formula by uniqueness of the adjoint. Note that the adjoint map, in our case, extends from $\mathcal{M}(\mathrm{supp}(\mu))$ to $\mathcal{M}(\mathbb{C}^{d_C})$ by zero.

[26] To see the complete positivity of $\Phi^*$, adjoint the Kraus representation of $\Phi$.

It remains to prove Fact 32(iii). To this end, we first show that $K$ is positive semidefinite with respect to the $\pi$-weighted inner product $\langle \cdot, \cdot \rangle_\pi$ from Definition A.1, which enables us to invoke the (finite-dimensional) spectral theorem. Indeed, weighted adjointness (i) tells us that

$$\langle A_1, K(A_2) \rangle_\pi = \langle \Phi(A_1), \Phi(A_2) \rangle_\nu .$$

In particular, $\langle A, K(A) \rangle_\pi = \langle \Phi(A), \Phi(A) \rangle_\nu \geqslant 0$. Moreover, conjugate symmetry gives us

$$\langle K(A), B \rangle_\pi = \overline{\langle B, K(A) \rangle_\pi} = \overline{\langle \Phi(B), \Phi(A) \rangle_\nu} = \langle \Phi(A), \Phi(B) \rangle_\nu = \langle A, K(B) \rangle_\pi .$$

Thus, $K$ is positive semidefinite with respect to the $\pi$-weighted inner product. By the spectral theorem with respect to $\langle \cdot, \cdot \rangle_\pi$, we can find a weighted-eigenbasis $\{E_j\}_{j=1}^m$ satisfying $\langle E_j, E_j \rangle_\pi = 1$ and $K(E_j) = \lambda_j E_j$. Positivity gives $\lambda_j = \langle E_j, K(E_j) \rangle_\pi \geqslant 0$.

We next show that every eigenvalue of $K$ lies in $[0, 1]$. Indeed, as a composition of support-restricted quantum channels, $K$ is a quantum channel from $\mathcal{M}(\mathrm{supp}(\pi))$ to itself. Let $K^*$ denote its adjoint with respect to the trace inner product on $\mathcal{M}(\mathrm{supp}(\pi))$, defined by the identity $\mathrm{Tr}[A_1^\dagger K^*(A_2)] = \mathrm{Tr}[K(A_1)^\dagger A_2]$. Now, notice that the trace preservation of $K$ implies that (in fact, equivalent) $K^*$ is unital, i.e., $K^*(\Pi_\pi) = \Pi_\pi$. Indeed, for every $A \in \mathcal{M}(\mathrm{supp}(\pi))$,

$$\mathrm{Tr}\left[A^\dagger K^*(\Pi_\pi)\right] = \mathrm{Tr}\left[K(A)^\dagger \Pi_\pi\right] = \overline{\mathrm{Tr}(K(A))} = \overline{\mathrm{Tr}(A)} = \mathrm{Tr}(A^\dagger \Pi_\pi).$$

The nondegeneracy of the trace inner product over $\mathcal{M}(\mathrm{supp}(\pi))$ thus implies that $K^*(\Pi_\pi) = \Pi_\pi$. Since $K$ is a quantum channel over $\mathcal{M}(\mathrm{supp}(\pi))$, by the Stinespring representation [NC10, Chapter 8], there exist a positive integer $k$ and an isometry $V : \mathrm{supp}(\pi) \to \mathrm{supp}(\pi) \otimes \mathbb{C}^k$ such that we can write $K(A) = \mathrm{Tr}_E(VAV^\dagger)$, where $\mathrm{Tr}_E : \mathcal{M}(\mathrm{supp}(\pi)) \otimes \mathcal{M}(\mathbb{C}^k) \to \mathcal{M}(\mathrm{supp}(\pi))$ is the partial trace map, i.e., the map $A \otimes M \mapsto A \,\mathrm{Tr}(M)$. By uniqueness of adjoints, we can therefore write $K^*(X) = V^\dagger(X \otimes I_k)V$ for all $X \in \mathcal{M}(\mathrm{supp}(\pi))$. Since $VV^\dagger \preceq \Pi_\pi \otimes I_k$, we have

$$\begin{aligned}
K^*(A)^\dagger K^*(A) &= V^\dagger (A^\dagger \otimes I_k) VV^\dagger (A \otimes I_k) V \\
&\preceq V^\dagger (A^\dagger A \otimes I_k) V \\
&= K^*(A^\dagger A) \\
&\preceq K^*(\|A\|_\infty^2 \Pi_\pi) \\
&\preceq \|A\|_\infty^2 K^*(\Pi_\pi) \\
&= \|A\|_\infty^2 \Pi_\pi ,
\end{aligned}$$

where the last two inequalities use the positivity and unitality of $K^*$.[27] Taking spectral norms thus gives us $\|K^*(A)\|_\infty \leqslant \|A\|_\infty$. Hence, if $K^*(A) = \mu A$ with $A \neq 0$, then $|\mu| \leqslant 1$. Since the eigenvalues of $K$ are simply conjugates of those of $K^*$, the preceding spectral bound then gives us $\lambda_j \leqslant 1$. Thus, every eigenvalue of $K$ lies in $[0, 1]$.

Now, we can conclude our proof. Writing $A = \sum_{j=1}^m a_j E_j$, we thus see that

$$0 \leqslant \langle A, K(A) \rangle_\pi = \sum_{j=1}^m \lambda_j \left|a_j\right|^2 \leqslant \sum_{j=1}^m \left|a_j\right|^2 = \langle A, A \rangle_\pi . \qquad \square$$

[27] This argument closely follows the one presented in [Wol12, Section 5.2, Equations (5.2)–(5.3)].

## A.2 Lemma 10: PGM data processing

In this subsection, using the tools we developed in Section A.1, we prove Lemma 10. Throughout this subsection, for an ensemble $\mathcal{E} = \{(\pi_x, \sigma_x) : x \in \mathcal{X}\}$ with PGM observables $\{M_x\}_{x\in\mathcal{X}}$, we adopt the shorthand $p_{\mathrm{s}}(\mathcal{E}) \coloneqq \sum_{x\in\mathcal{X}} \pi_x \operatorname{Tr}(\sigma_x M_x) = 1 - p_{\mathrm{e}}(\mathcal{E})$ for its PGM success probability.

*Proof of Lemma 10.* Define $A_x \coloneqq \pi_x\sigma_x$. Expanding the PGM success formula using the definition of PGM (Definition 2.7), observe that

$$p_{\mathrm{s}}(\mathcal{E}) = \sum_{x\in\mathcal{X}} \pi_x^2 \operatorname{Tr}\left(\sigma_x \bar{\sigma}^{-1/2}\sigma_x\bar{\sigma}^{-1/2}\right) = \sum_{x\in\mathcal{X}} \langle A_x, A_x\rangle_{\bar{\sigma}}.$$

For $x \in \mathcal{X}$, let $\tau_x \coloneqq \mathcal{N}(\sigma_x)$. By linearity of $\mathcal{N}$, we see that the average post-processed state satisfies $\bar{\tau} \coloneqq \sum_{x\in\mathcal{X}} \pi_x\tau_x = \sum_{x\in\mathcal{X}} \pi_x\mathcal{N}(\sigma_x) = \mathcal{N}(\bar{\sigma})$. Now, for every $x \in \mathcal{X}$, by applying the weighted adjointness (i) and contraction (iii) of the Petz map from Fact 32, we see that

$$\langle \mathcal{N}(A_x), \mathcal{N}(A_x)\rangle_{\bar{\tau}} = \left\langle A_x, (\mathcal{N}^{\sharp}_{\bar{\sigma}} \circ \mathcal{N})(A_x)\right\rangle_{\bar{\sigma}} \leqslant \langle A_x, A_x\rangle_{\bar{\sigma}}\,.$$

Summing this inequality and applying the same PGM expansion to the processed ensemble thus shows that

$$p_{\mathrm{s}}(\mathcal{E}) = \sum_{x\in\mathcal{X}} \langle A_x, A_x\rangle_{\bar{\sigma}} \geqslant \sum_{x\in\mathcal{X}} \langle \mathcal{N}(A_x), \mathcal{N}(A_x)\rangle_{\bar{\tau}} = p_{\mathrm{s}}\left(\{(\pi_x, \mathcal{N}(\sigma_x)) : x \in \mathcal{X}\}\right).$$

Since $p_{\mathrm{e}} = 1 - p_{\mathrm{s}}$, it therefore follows that

$$p_{\mathrm{e}}(\mathcal{E}) \leqslant p_{\mathrm{e}}\left(\{(\pi_x, \mathcal{N}(\sigma_x)) : x \in \mathcal{X}\}\right). \qquad \square$$

*Remark* A.2 (Alternative proof via Rényi divergence). In the classical case, Lemma 10 admits a more natural proof by recasting $p_{\mathrm{s}}(\mathcal{E})$ as a conditional collision entropy and invoking Jensen's inequality. Such a proof does extend to the quantum setting via the data processing inequality for the sandwiched order-2 quantum Rényi divergence [MLDS+13, WWY14], established independently by Frank and Lieb [FL13] and by Beigi [Bei13]. For completeness and greater self-containment, we presented this alternative proof via the Petz map. In this formulation, the Petz–Gram inequality in Fact 32(iii) serves as the noncommutative analog of the conditional-Jensen step. ◁

## A.3 Proposition 6: Capacity of a binary-input output-symmetric cq channel

In this subsection, we prove Proposition 6.

*Proof of Proposition 6.* Fix any positive integer $n$, and let $\pi^{(n)}$ be any probability distribution on $\{0,1\}^n$. Write $\chi(\pi^{(n)}; \Phi_\sigma^{\otimes n})$ for the Holevo information of the ensemble that assigns probability $\pi^{(n)}(x^n)$ to $\sigma_{x_1} \otimes \cdots \otimes \sigma_{x_n}$. Define the $n$-partite state

$$\omega_{B^n} \coloneqq \sum_{x^n\in\{0,1\}^n} \pi^{(n)}(x^n)\sigma_{x_1} \otimes \cdots \otimes \sigma_{x_n},$$

and let $\omega_{B_i} \coloneqq \pi_i(0)\sigma_0 + \pi_i(1)\sigma_1$ be its $i$'th subsystem, where $\pi_i$ is the $i$'th-coordinate marginal of $\pi^{(n)}$. By the subadditivity of entropy, we see that

$$\chi(\pi^{(n)}; \Phi_\sigma^{\otimes n}) = S(\omega_{B^n}) - \sum_{i=1}^{n} \left(\pi_i(0)S(\sigma_0) + \pi_i(1)S(\sigma_1)\right)$$

$$\leqslant \sum_{i=1}^{n} \left( S(\omega_{B_i}) - \pi_i(0) S(\sigma_0) - \pi_i(1) S(\sigma_1) \right)$$
$$= \sum_{i=1}^{n} \chi(\pi_i; \Phi_\sigma). \tag{63}$$

Now, define $f(t) \coloneqq \chi(\{(t, \sigma_0), (1-t, \sigma_1)\})$ for $t \in [0,1]$. Notice that the concavity of von Neumann entropy implies that $f$ is concave, while output-symmetry of $\Phi_\sigma$ implies that $f(t) = f(1-t)$. Hence

$$\chi(\{(t, \sigma_0), (1-t, \sigma_1)\}) = f(t) = \frac{1}{2} f(t) + \frac{1}{2} f(1-t) \leqslant f\left(\frac{1}{2}\right) = \chi(\sigma_0, \sigma_1).$$

Therefore, for every $i \in \{1, \ldots, n\}$, we have that

$$\chi(\pi_i; \Phi_\sigma) \leqslant \chi(\sigma_0, \sigma_1). \tag{64}$$

Combining (63) and (64) thus gives us $\chi(\pi^{(n)}; \Phi_\sigma^{\otimes n}) \leqslant n \chi(\sigma_0, \sigma_1)$. Since the uniform product input attains equality by additivity of entropy, we thus see that

$$\frac{1}{n} \max_{\pi^{(n)}} \chi(\pi^{(n)}; \Phi_\sigma^{\otimes n}) = \chi(\sigma_0, \sigma_1).$$

As $n \to \infty$, the left-hand side becomes $C(\Phi_\sigma)$ by Definition 2.6. This completes the proof. □

### A.4 Theorem 7: The cq channel coding strong converse

In this subsection, we prove the cq channel coding strong converse.[28] To facilitate the exposition, we recall for the reader the proof for the classical BSC case. Fix $p \in (0, \frac{1}{2})$, and suppose that $\{\mathcal{D}_c : c \in C\}$ are the decoding regions of a deterministic decoder for BSC($p$). If $\mathbf{c}$ is sent and the noise is $\mathbf{z} \in \{0,1\}^n$, then the decoder succeeds exactly when $\mathbf{c} \oplus \mathbf{z} \in \mathcal{D}_c$. Now, define the function

$$t_n(z) \coloneqq \frac{1}{|C|} \left| \{c \in C : c \oplus z \in \mathcal{D}_c\} \right|.$$

If $\mathbf{z} \sim \mathrm{Ber}(p)^n$ is the BSC noise and $\mathbf{y}$ is uniform on $\{0,1\}^n$, then

$$\mathbf{E}[t_n(\mathbf{z})] = \mathbf{Pr}[\hat{\mathbf{c}} = \mathbf{c}],$$
$$\mathbf{E}[t_n(\mathbf{y})] = \frac{1}{|C|}.$$

The second identity holds because translation preserves the uniform distribution and the decoding regions partition $\{0,1\}^n$. Now, define the one-sided typical set

$$Q_\varepsilon \coloneqq \left\{ z \in \{0,1\}^n : -\log \mathbf{Pr}[\mathbf{e} = z] \geqslant n[h(p) - \varepsilon] \right\},$$

where $\mathbf{e} \sim \mathrm{Ber}(p)^n$. For an independent $\mathbf{z} \sim \mathrm{Ber}(p)^n$, notice that $\mathbf{E}_{\mathbf{z}} \left[ -\log \mathbf{Pr}_{\mathbf{e}}[\mathbf{e} = \mathbf{z}] \right] = n h(p)$. Since $-\log \mathbf{Pr}_{\mathbf{e}}[\mathbf{e} = \mathbf{z}] = \sum_{i=1}^n -\log \mathbf{Pr}_{\mathbf{e}_i}[\mathbf{e}_i = \mathbf{z}_i]$, each of which is independent, applying Hoeffding's inequality on $\mathbf{z}$ therefore gives

$$\mathbf{Pr}[\mathbf{z} \notin Q_\varepsilon] \leqslant \exp\left( -\frac{2\varepsilon^2 n}{[\log((1-p)/p)]^2} \right).$$

[28] For context, the classical strong converse dates back to Wolfowitz [Wol57], with exponential versions following later refinements [Str64, Ari73].

For every $z \in Q_\varepsilon$, the definition of the set and the identity $\mathbf{Pr}[\mathbf{y} = z] = 2^{-n}$ give us the likelihood ratio bound

$$\mathbf{Pr}[\mathbf{z} = z] \leqslant 2^{-n[h(p)-\varepsilon]} = 2^{n[1-h(p)+\varepsilon]}\mathbf{Pr}[\mathbf{y} = z].$$

Since $0 \leqslant t_n(z) \leqslant 1$, we can split the success probability according to whether the noise is typical or not and apply this comparison term by term. That is,

$$\begin{aligned}
\mathbf{Pr}[\hat{\mathbf{c}} = \mathbf{c}] &= \sum_{z\in\{0,1\}^n} \mathbf{Pr}[\mathbf{z} = z]t_n(z) \\
&\leqslant \sum_{z\in Q_\varepsilon} \mathbf{Pr}[\mathbf{z} = z]t_n(z) + \mathbf{Pr}[\mathbf{z} \notin Q_\varepsilon] \\
&\leqslant 2^{n[1-h(p)+\varepsilon]} \sum_{z\in Q_\varepsilon} \mathbf{Pr}[\mathbf{y} = z]t_n(z) + \mathbf{Pr}[\mathbf{z} \notin Q_\varepsilon] \\
&\leqslant 2^{n[1-h(p)+\varepsilon]}\mathbf{E}[t_n(\mathbf{y})] + \mathbf{Pr}[\mathbf{z} \notin Q_\varepsilon] \\
&\leqslant 2^{-n[R(C)-(1-h(p))-\varepsilon]} + \exp\left(-\frac{2\varepsilon^2 n}{[\log((1-p)/p)]^2}\right).
\end{aligned}$$

Thus, whenever $R(C) > 1 - h(p)$ by a fixed amount, choosing $\varepsilon$ smaller than that gap makes both terms exponentially small. This finishes the BSC strong converse proof.

The cq channel proof repeats these three steps. Output symmetry plays a similar role as XOR in the BSC case, producing a matrix $T_n$ that plays the role of $t_n$. Its expectations under $\sigma_0^{\otimes n}$ and $\bar{\sigma}^{\otimes n}$ reproduce the two identities above. Finally, two spectral cutoffs replace the BSC typical set—one cutoff to say that the probability of $\mathbf{z}$ (resp. $\sigma_0^{\otimes n}$) is not too large, and another to say that the probability of $\mathbf{y}$ (resp. $\bar{\sigma}^{\otimes n}$) is not too small—and provide the required matrix version of the likelihood ratio bound.

*Proof of Theorem 7.* Recall that each decoder matrix $D_c$ is designed to recognize its corresponding codeword state $\sigma_c$. Let $U$ be the unitary exchanging $\sigma_0$ and $\sigma_1$, and, for $c \in C$, set $V_c \coloneqq U^{c_1} \otimes \cdots \otimes U^{c_n}$. Then we see that $V_c \sigma_0^{\otimes n} V_c^\dagger = \sigma_c$ and $V_c \bar{\sigma}^{\otimes n} V_c^\dagger = \bar{\sigma}^{\otimes n}$. Following the BSC definition of $t_n$, define the matrix

$$T_n \coloneqq \frac{1}{|C|} \sum_{c\in C} V_c^\dagger D_c V_c.$$

Since every $D_c$ satisfies $0 \preceq D_c \preceq I$, their average satisfies $0 \preceq T_n \preceq I$. The analogs of the two BSC expectations are thus

$$\begin{aligned}
\mathrm{Tr}(\sigma_0^{\otimes n} T_n) &= \frac{1}{|C|} \sum_{c\in C} \mathrm{Tr}(\sigma_0^{\otimes n} V_c^\dagger D_c V_c) = \frac{1}{|C|} \sum_{c\in C} \mathrm{Tr}(\sigma_c D_c) = \mathbf{Pr}[\hat{\mathbf{c}} = \mathbf{c}], \\
\mathrm{Tr}(\bar{\sigma}^{\otimes n} T_n) &= \frac{1}{|C|} \sum_{c\in C} \mathrm{Tr}(\bar{\sigma}^{\otimes n} V_c^\dagger D_c V_c) = \frac{1}{|C|} \sum_{c\in C} \mathrm{Tr}(\bar{\sigma}^{\otimes n} D_c) = \frac{1}{|C|}.
\end{aligned} \tag{65}$$

For the second identity, we used the invariance of $\bar{\sigma}^{\otimes n}$ under every $V_c$ and the POVM identity $\sum_{c\in C} D_c = I$. Thus, $T_n$ is exactly the matrix counterpart of $t_n$.

Let $\lambda_{\min}$ be the smallest positive eigenvalue occurring in either of $\sigma_0$ or $\bar{\sigma}$. The channel-dependent constant in the theorem statement is defined to be

$$L_\sigma \coloneqq \max\{1, \log(1/\lambda_{\min})\}.$$

The first cutoff removes eigenvalues of $\sigma_0^{\otimes n}$ that are atypically large, so as to make the bulk of its spectrum "spread out." To expose the classical random variable underneath this step, choose an eigenbasis of $\sigma_0$ with positive eigenvalues $\lambda_j$. Measuring $\sigma_0$ in this basis returns $j$ with probability $\lambda_j$, so the "surprisal" $-\log \lambda_j$ lies in $[0, L_\sigma]$ and has mean $S(\sigma_0)$. Let $Q$ be the spectral projector of $\sigma_0^{\otimes n}$ onto the positive eigenvalues $\lambda$ satisfying $-\log \lambda \geqslant n[S(\sigma_0) - \varepsilon]$. Hoeffding's inequality [Hoe63], applied to the sum of $n$ independent terms, gives

$$\operatorname{Tr}(\sigma_0^{\otimes n} Q) \geqslant 1 - \exp(-2\varepsilon^2 n / L_\sigma^2). \tag{66}$$

The second cutoff removes eigenspaces on which $\bar{\sigma}^{\otimes n}$ is atypically small. Note that the BSC did not need to perform this, since the distribution corresponding to $\bar{\sigma}^{\otimes n}$ was the uniform distribution over $\{0,1\}^n$, which evidently satisfies this property. Now, to see the relevant probability distribution, choose an eigenbasis $\{|j\rangle\}$ of $\bar{\sigma}$, write $\mu_j > 0$ for the eigenvalue associated with $|j\rangle$, and measure $\sigma_0$ in this basis. The outcome $j$ has probability $\langle j| \sigma_0 |j\rangle$, so $-\log \mu_j$ lies in $[0, L_\sigma]$ and has mean $-\operatorname{Tr}(\sigma_0 \log \bar{\sigma}) = S(\bar{\sigma})$. Let $P$ be the spectral projector of $\bar{\sigma}^{\otimes n}$ onto the eigenvalues $\mu$ satisfying $-\log \mu \leqslant n[S(\bar{\sigma}) + \varepsilon]$. For $n$ copies, the product-basis measurement gives $n$ independent copies of this random variable, so a second application of Hoeffding's inequality yields

$$\operatorname{Tr}(\sigma_0^{\otimes n} P) \geqslant 1 - \exp(-2\varepsilon^2 n / L_\sigma^2). \tag{67}$$

We now show that imposing both cutoffs changes $\sigma_0^{\otimes n}$ only slightly. Set

$$\widetilde{\sigma}_n \coloneqq PQ\sigma_0^{\otimes n} QP.$$

Since $Q$ commutes with $\sigma_0^{\otimes n}$, we have $0 \leq Q\sigma_0^{\otimes n} Q \leq \sigma_0^{\otimes n}$. From (66) and (67), we see that

$$\begin{aligned} \|\sigma_0^{\otimes n} - Q\sigma_0^{\otimes n} Q\|_1 &= \operatorname{Tr}\left(\sigma_0^{\otimes n}(I - Q)\right) \leqslant \exp(-2\varepsilon^2 n / L_\sigma^2), \\ \operatorname{Tr}\left(Q\sigma_0^{\otimes n} Q(I - P)\right) &\leqslant \operatorname{Tr}\left(\sigma_0^{\otimes n}(I - P)\right) \leqslant \exp(-2\varepsilon^2 n / L_\sigma^2). \end{aligned}$$

Using these two inequalities, the triangle inequality, and applying the gentle measurement lemma (Lemma 31) to the subnormalized matrix $Q\sigma_0^{\otimes n} Q$, we thus see that

$$\begin{aligned} \|\sigma_0^{\otimes n} - \widetilde{\sigma}_n\|_1 &\leqslant \|\sigma_0^{\otimes n} - Q\sigma_0^{\otimes n} Q\|_1 + \|Q\sigma_0^{\otimes n} Q - PQ\sigma_0^{\otimes n} QP\|_1 \\ &\leqslant \exp(-2\varepsilon^2 n / L_\sigma^2) + 2\sqrt{\exp(-2\varepsilon^2 n / L_\sigma^2)} \\ &\leqslant 4 \exp\left(-\varepsilon^2 n / L_\sigma^2\right). \end{aligned} \tag{68}$$

Next, we establish the matrix analog of the likelihood ratio bound. By the definitions of $Q$ and $P$, notice that $Q\sigma_0^{\otimes n} Q \leq 2^{-n[S(\sigma_0) - \varepsilon]} Q$ and $P \leq 2^{n[S(\bar{\sigma}) + \varepsilon]} \bar{\sigma}^{\otimes n}$. By combining these two inequalities along with $PQP \leq P$, we see that

$$\begin{aligned} \widetilde{\sigma}_n &= PQ\sigma_0^{\otimes n} QP \\ &\leq 2^{-n[S(\sigma_0) - \varepsilon]} PQP \\ &\leq 2^{-n[S(\sigma_0) - \varepsilon]} P \\ &\leq 2^{n[S(\bar{\sigma}) - S(\sigma_0) + 2\varepsilon]} \bar{\sigma}^{\otimes n} \\ &= 2^{n[\chi(\sigma_0, \sigma_1) + 2\varepsilon]} \bar{\sigma}^{\otimes n}. \end{aligned} \tag{69}$$

This is the quantum replacement for the classical assertion that the likelihood ratio is at most $2^{n(\chi+2\varepsilon)}$ on the retained typical set.

Now, we have all the ingredients to conclude this theorem. Indeed, by the inequality $0 \preceq T_n \preceq I$ and Eqs. (65), (68) and (69), we see that

$$\begin{aligned}\mathbf{Pr}[\hat{\mathbf{c}} = \mathbf{c}] &= \operatorname{Tr}(\sigma_0^{\otimes n} T_n) \\ &= \operatorname{Tr}(\widetilde{\sigma}_n T_n) + \operatorname{Tr}((\sigma_0^{\otimes n} - \widetilde{\sigma}_n) T_n) \\ &\leqslant 2^{n[\chi(\sigma_0,\sigma_1)+2\varepsilon]} \operatorname{Tr}(\bar{\sigma}^{\otimes n} T_n) + \|\sigma_0^{\otimes n} - \widetilde{\sigma}_n\|_1 \\ &= 2^{-n[R(C)-\chi(\sigma_0,\sigma_1)-2\varepsilon]} + 4\exp\left(-\varepsilon^2 n / L_\sigma^2\right).\end{aligned}$$

For the final assertion, consider any family of codes for which $R(C) - \chi(\sigma_0, \sigma_1) \geqslant \Delta$ for some fixed $\Delta > 0$. Choosing $\varepsilon = \Delta/4$ in the preceding bound gives the exponentially decaying estimate

$$\mathbf{Pr}[\hat{\mathbf{c}} = \mathbf{c}] \leqslant 2^{-n\Delta/2} + 4\exp\left(-\Delta^2 n/(16 L_\sigma^2)\right). \qquad \square$$

### A.5 Lemma 8: PGM is closed under coarse-graining

In this subsection, we prove Lemma 8. The proof proceeds by exploiting the linearity of PGMs.

*Proof.* First, observe that

$$\bar{\omega} \coloneqq \sum_{y \in \mathcal{Y}} \nu_y \omega_y = \sum_{y \in \mathcal{Y}} \mathbf{Pr}[f(\mathbf{x}) = y] \mathbf{E}[\sigma_{\mathbf{x}} \mid f(\mathbf{x}) = y] = \mathbf{E}[\sigma_{\mathbf{x}}] = \bar{\sigma}.$$

Let $\{M_{\hat{x}}\}_{\hat{x} \in \mathcal{X}}$ and $\{N_{\hat{y}}\}_{\hat{y} \in \mathcal{Y}}$ be the PGMs of $\mathcal{E}$ and $\mathcal{F}$, respectively. The PGM formula (Definition 2.7) then gives

$$N_{\hat{y}} = \nu_{\hat{y}} \bar{\omega}^{-1/2} \omega_{\hat{y}} \bar{\omega}^{-1/2} = \bar{\sigma}^{-1/2} \left( \sum_{\substack{\hat{x} \in \mathcal{X}: \\ f(\hat{x}) = \hat{y}}} \pi_{\hat{x}} \sigma_{\hat{x}} \right) \bar{\sigma}^{-1/2} = \sum_{\substack{\hat{x} \in \mathcal{X}: \\ f(\hat{x}) = \hat{y}}} M_{\hat{x}}.$$

Thus, the PGM $\{N_{\hat{y}}\}_{\hat{y} \in \mathcal{Y}}$ is the classical post-processing of $\{M_{\hat{x}}\}_{\hat{x} \in \mathcal{X}}$ under $f$. In particular, an outcome $\hat{x}$ of the first PGM becomes $f(\hat{x})$ under the second.

For every $y, \hat{y} \in \mathcal{Y}$, the PGM error formula (Definition 2.8) gives us

$$\mathbf{Pr}[f(\mathbf{x}) = y,\ f(\hat{\mathbf{x}}) = \hat{y}] = \sum_{\substack{x \in \mathcal{X} \\ f(x) = y}} \sum_{\substack{\hat{x} \in \mathcal{X} \\ f(\hat{x}) = \hat{y}}} \pi_x \operatorname{Tr}(\sigma_x M_{\hat{x}}) = \sum_{\substack{x \in \mathcal{X} \\ f(x) = y}} \pi_x \operatorname{Tr}(\sigma_x N_{\hat{y}}) = \nu_y \operatorname{Tr}(\omega_y N_{\hat{y}}).$$

By the PGM error formula (Definition 2.8) once again and the definition of $\omega_y$, we conclude that

$$\mathbf{Pr}[f(\hat{\mathbf{x}}) \neq f(\mathbf{x})] = \sum_{\substack{y, \hat{y} \in \mathcal{Y}: \\ y \neq \hat{y}}} \mathbf{Pr}[f(\mathbf{x}) = y,\ f(\hat{\mathbf{x}}) = \hat{y}] = \sum_{\substack{y, \hat{y} \in \mathcal{Y}: \\ y \neq \hat{y}}} \nu_y \operatorname{Tr}(\omega_y N_{\hat{y}}) = p_{\mathrm{e}}(\mathcal{F}). \qquad \square$$

### A.6 Lemma 11: Worst-case uniform prior

*Proof of Lemma 11.* Consider the binary ensemble that asks to decode a uniformly random bit $\mathbf{b}$ when given as input $(\sigma_{\mathbf{b}}, \mathbf{b} \oplus \mathbf{t})$ for $\mathbf{t} \sim \mathrm{Ber}(\alpha)$. More formally, consider the binary ensemble $\mathcal{E}'_\alpha := \{(\frac{1}{2}, \widehat{\sigma}_0), (\frac{1}{2}, \widehat{\sigma}_1)\}$ where

$$\widehat{\sigma}_0 := \sigma_0 \otimes (1-\alpha)\,|0\rangle\langle 0|_T + \sigma_0 \otimes \alpha\,|1\rangle\langle 1|_T\,, \qquad \widehat{\sigma}_1 := \sigma_1 \otimes \alpha\,|0\rangle\langle 0|_T + \sigma_1 \otimes (1-\alpha)\,|1\rangle\langle 1|_T\,.$$

Here, $T$ is a two-dimensional classical system. We claim that

$$p_{\mathrm{e}}(\widehat{\sigma}_0, \widehat{\sigma}_1) = p_{\mathrm{e},\alpha}(\sigma_0, \sigma_1).$$

Indeed, one can check that the PGM of $\mathcal{E}'_\alpha$ performs the following procedure: it first reads $T$. Then, if $T = 0$, it performs the PGM for the ensemble $\{(1-\alpha, \sigma_0), (\alpha, \sigma_1)\}$. Otherwise, if $T = 1$, it performs the PGM for the ensemble $\{(\alpha, \sigma_0), (1-\alpha, \sigma_1)\}$. Since $p_{\mathrm{e},\alpha}(\sigma_0, \sigma_1) = p_{\mathrm{e},1-\alpha}(\sigma_0, \sigma_1)$ by output-symmetry, we thus see that

$$p_{\mathrm{e}}(\widehat{\sigma}_0, \widehat{\sigma}_1) = \frac{1}{2} p_{\mathrm{e},\alpha}(\sigma_0, \sigma_1) + \frac{1}{2} p_{\mathrm{e},\alpha}(\sigma_0, \sigma_1) = p_{\mathrm{e},\alpha}(\sigma_0, \sigma_1).$$

Now, observe that the quantum channel that traces $T$ out sends $\widehat{\sigma}_b$ to $\sigma_b$ for $b \in \{0, 1\}$. The PGM data-processing lemma (Lemma 10) therefore gives

$$p_{\mathrm{e},\alpha}(\sigma_0, \sigma_1) = p_{\mathrm{e}}(\widehat{\sigma}_0, \widehat{\sigma}_1) \leqslant p_{\mathrm{e},1/2}(\sigma_0, \sigma_1). \qquad \square$$

### A.7 Propositions 12 and 13: Binary PGM formulas

In this subsection, we prove Propositions 12 and 13.

*Proof of Proposition 12.* By Definition 2.7, the PGM matrices for $\mathcal{E}_\alpha$ are $M_0 = (1-\alpha)\bar{\sigma}_\alpha^{-1/2}\sigma_0\bar{\sigma}_\alpha^{-1/2}$ and $M_1 = \alpha\bar{\sigma}_\alpha^{-1/2}\sigma_1\bar{\sigma}_\alpha^{-1/2}$. From Definition 2.8 and the cyclicity of the trace, we thus find that

$$\begin{aligned} p_{\mathrm{e}}(\mathcal{E}_\alpha) &= (1-\alpha)\,\mathrm{Tr}(\sigma_0 M_1) + \alpha\,\mathrm{Tr}(\sigma_1 M_0) \\ &= \alpha(1-\alpha)\,\mathrm{Tr}\left(\sigma_0\bar{\sigma}_\alpha^{-1/2}\sigma_1\bar{\sigma}_\alpha^{-1/2}\right) + \alpha(1-\alpha)\,\mathrm{Tr}\left(\sigma_1\bar{\sigma}_\alpha^{-1/2}\sigma_0\bar{\sigma}_\alpha^{-1/2}\right) \\ &= 2\alpha(1-\alpha)\,\mathrm{Tr}\left(\sigma_0\bar{\sigma}_\alpha^{-1/2}\sigma_1\bar{\sigma}_\alpha^{-1/2}\right). \end{aligned}$$

This proves (15).

As for (16), observe that the average $\bar{\sigma}_\alpha$ admits the Gram factorization $\bar{\sigma}_\alpha = VV^\dagger$, where $V$ is the matrix consisting of the two columns $\sqrt{1-\alpha}\,|\psi_0\rangle$ and $\sqrt{\alpha}\,|\psi_1\rangle$. Now, define the matrix

$$G := V^\dagger V = \begin{pmatrix} 1-\alpha & \sqrt{\alpha(1-\alpha)}\,\langle\psi_0|\psi_1\rangle \\ \sqrt{\alpha(1-\alpha)}\,\langle\psi_1|\psi_0\rangle & \alpha \end{pmatrix}.$$

Observe that $\sqrt{G} = V^\dagger(VV^\dagger)^{-1/2}V = V^\dagger\bar{\sigma}_\alpha^{-1/2}V$. Combining this with the cyclicity of the trace and (15), notice that

$$p_{\mathrm{e}}(\mathcal{E}_\alpha) = 2\alpha(1-\alpha)\,\mathrm{Tr}\left(\sigma_0\bar{\sigma}_\alpha^{-1/2}\sigma_1\bar{\sigma}_\alpha^{-1/2}\right)$$

$$= 2\alpha(1-\alpha)\operatorname{Tr}\left(|\psi_0\rangle\langle\psi_0|\,\bar{\sigma}_\alpha^{-1/2}\,|\psi_1\rangle\langle\psi_1|\,\bar{\sigma}_\alpha^{-1/2}\right)$$
$$= 2\left(\sqrt{\alpha(1-\alpha)}\,\langle\psi_0|\,\bar{\sigma}_\alpha^{-1/2}\,|\psi_1\rangle\right)^2$$
$$= 2\left|\sqrt{G}_{01}\right|^2. \tag{70}$$

Now, since $G$ is a $2\times 2$ matrix, the span of $I$ and $G$ includes all the projections onto the eigenspaces of $G$. Thus, we can express $\sqrt{G}$ as

$$\sqrt{G} = aG + bI$$

for some $a, b \in \mathbb{R}$. If $\lambda_1, \lambda_2 \in \mathbb{R}$ denote the eigenvalues of $G$, then this leads to a simple linear system $a\lambda_i + b = \sqrt{\lambda_i}$ for $i = 1, 2$. Solving for just $a$, we get

$$a = \frac{\sqrt{\lambda_1}-\sqrt{\lambda_2}}{\lambda_1-\lambda_2} = \frac{1}{\sqrt{\lambda_1}+\sqrt{\lambda_2}} = \frac{1}{\sqrt{\lambda_1+\lambda_2+2\sqrt{\lambda_1\lambda_2}}} = \frac{1}{\sqrt{\operatorname{Tr}(G)+2\sqrt{\det(G)}}}.$$

From the definition of $G$, we see that $\operatorname{Tr}(G) = 1$ and $\det(G) = \alpha(1-\alpha)(1-f)$. Thus

$$(\sqrt{G})_{01} = a\cdot(G)_{01} + b\cdot 0 = \frac{1}{\sqrt{\operatorname{Tr}(G)+2\sqrt{\det(G)}}}\cdot\sqrt{\alpha(1-\alpha)}\,\langle\psi_0|\psi_1\rangle = \frac{\sqrt{\alpha(1-\alpha)}\,\langle\psi_0|\psi_1\rangle}{\sqrt{1+2\sqrt{\alpha(1-\alpha)(1-f)}}}. \tag{71}$$

Therefore, combining (70) and (71), we conclude that

$$p_{\mathrm{e}}(\mathcal{E}_\alpha) = \frac{2\alpha(1-\alpha)f}{1+2\sqrt{\alpha(1-\alpha)(1-f)}}.$$ □

*Proof of Proposition 13.* Write $\bar{\sigma}_\theta := (1-\theta)\sigma_0 + \theta\sigma_1$. Let $N_y := \mathbf{Pr}[\mathbf{y}=y]\bar{\sigma}_\theta^{-1/2}\sigma_y\bar{\sigma}_\theta^{-1/2}$ be the PGM observables for the ensemble indexed by $\mathbf{y}$, and let $\{M_0, M_1\}$ be the PGM observables for the composed ensemble indexed by $\mathbf{b}$. Since PGM is the Petz recovery map for cq channels (cf. Section A.1), by Petz functoriality (Fact 32(ii)), it follows that $\{M_0, M_1\}$ is a classical post-processing of $\{N_0, N_1\}$. More concretely,

$$\begin{aligned} M_b &= \mathbf{Pr}[\mathbf{b}=b]\bar{\sigma}_\theta^{-1/2}\tau_b\bar{\sigma}_\theta^{-1/2} \\ &= \sum_{y\in\{0,1\}} \mathbf{Pr}[\mathbf{b}=b, \mathbf{y}=y]\bar{\sigma}_\theta^{-1/2}\sigma_y\bar{\sigma}_\theta^{-1/2} \\ &= \sum_{y\in\{0,1\}} \mathbf{Pr}[\mathbf{b}=b \mid \mathbf{y}=y]N_y. \end{aligned}$$

Thus, the collective PGM first produces $\widehat{\mathbf{y}}$ using $\{N_0, N_1\}$ and then draws $\widehat{\mathbf{b}}$ from the posterior law of $\mathbf{b}$ given $\mathbf{y} = \widehat{\mathbf{y}}$, as asserted.

Now, the joint law of the intermediate input and PGM output is symmetric. Indeed, cyclicity of trace gives, for $y, \hat{y} \in \{0,1\}$,

$$\mathbf{Pr}[\mathbf{y}=y, \widehat{\mathbf{y}}=\hat{y}] = \mathbf{Pr}[\mathbf{y}=y]\mathbf{Pr}[\mathbf{y}=\hat{y}]\operatorname{Tr}\left(\sigma_y\bar{\sigma}_\theta^{-1/2}\sigma_{\hat{y}}\bar{\sigma}_\theta^{-1/2}\right) = \mathbf{Pr}[\mathbf{y}=\hat{y}, \widehat{\mathbf{y}}=y].$$

Put $\varepsilon := p_{\mathrm{e},\theta}(\sigma_0, \sigma_1)$. The two marginals of $(\mathbf{y}, \widehat{\mathbf{y}})$ are therefore both $\mathrm{Ber}(\theta)$, and the two off-diagonal probabilities are each $\frac{\varepsilon}{2}$. Thus, its $2 \times 2$ joint-distribution matrix, with rows indexed by $\mathbf{y}$ and columns by $\widehat{\mathbf{y}}$, is

$$J_{\mathbf{y},\widehat{\mathbf{y}}} = \begin{pmatrix} 1 - \theta - \frac{\varepsilon}{2} & \frac{\varepsilon}{2} \\ \frac{\varepsilon}{2} & \theta - \frac{\varepsilon}{2} \end{pmatrix}.$$

Let $R$ be the $2\times2$ probability matrix of the posterior sampler of $\mathbf{b}$ given $\mathbf{y}$, i.e., $R_{b,y} := \mathbf{Pr}[\mathbf{b} = b \mid \mathbf{y} = y]$. By Bayes' rule, we have that

$$R = \begin{pmatrix} 1 - \frac{\alpha\eta}{1-\theta} & 1 - \frac{\alpha(1-\eta)}{\theta} \\ \frac{\alpha\eta}{1-\theta} & \frac{\alpha(1-\eta)}{\theta} \end{pmatrix}, \qquad \det(R) = \frac{\alpha(1-\alpha)(1-2\eta)}{\theta(1-\theta)}.$$

Conditioned on $(\mathbf{y}, \widehat{\mathbf{y}})$, the original bit $\mathbf{b}$ and the fresh posterior draw $\widehat{\mathbf{b}}$ are independent, so their joint-distribution matrix is

$$J_{\mathbf{b},\widehat{\mathbf{b}}} = R J_{\mathbf{y},\widehat{\mathbf{y}}} R^{\mathsf{T}}.$$

For any pair of binary random variables $(\mathbf{x}, \mathbf{z})$, covariance is precisely the determinant of their $2 \times 2$ joint-distribution matrix. Indeed, if $J_{\mathbf{x},\mathbf{z}} = (p_{xz})_{x,z\in\{0,1\}}$, then

$$\begin{aligned} \mathrm{Cov}(\mathbf{x}, \mathbf{z}) &= p_{11} - (p_{11} + p_{10})(p_{11} + p_{01}) \\ &= p_{11} - p_{11}^2 - p_{11}p_{01} - p_{11}p_{10} - p_{10}p_{01} \\ &= p_{11}p_{00} - p_{10}p_{01} \\ &= \det(J_{\mathbf{x},\mathbf{z}}). \end{aligned}$$

Consequently,

$$\mathrm{Cov}(\mathbf{b}, \widehat{\mathbf{b}}) = \det(R)^2 \det(J_{\mathbf{y},\widehat{\mathbf{y}}}) = \left[\frac{\alpha(1-\alpha)(1-2\eta)}{\theta(1-\theta)}\right]^2 \left[\theta(1-\theta) - \frac{\varepsilon}{2}\right].$$

Since $\widehat{\mathbf{y}}$ has the same marginal as $\mathbf{y}$, posterior resampling gives $\widehat{\mathbf{b}} \sim \mathrm{Ber}(\alpha)$. Hence

$$\begin{aligned} p_{\mathrm{e},\alpha}(\tau_0, \tau_1) &= \mathbf{E}\left[(\mathbf{b} - \widehat{\mathbf{b}})^2\right] \\ &= 2\alpha(1-\alpha) - 2\,\mathrm{Cov}(\mathbf{b}, \widehat{\mathbf{b}}) \\ &= 2\alpha(1-\alpha) - \left[\frac{\alpha(1-\alpha)(1-2\eta)}{\theta(1-\theta)}\right]^2 \left[2\theta(1-\theta) - p_{\mathrm{e},\theta}(\sigma_0, \sigma_1)\right], \end{aligned}$$

which is (18). Setting $\alpha = \theta = \frac{1}{2}$ in (18) gives

$$p_{\mathrm{e}}(\tau_0, \tau_1) = \frac{1}{2}\left[1 - (1-2\eta)^2(1 - 2p_{\mathrm{e}}(\sigma_0, \sigma_1))\right]. \qquad \square$$

### A.8 Proposition 14: Qubit entropy in Pauli coordinates

In this subsection, we prove Proposition 14.

*Proof of Proposition 14.* Since $X^2 = Z^2 = I$ and $XZ = -ZX$, we see that

$$(xX + zZ)^2 = (x^2 + z^2)I.$$

Since the Hermitian matrix $xX + zZ$ is traceless, we therefore see that the eigenvalues of $xX + zZ$ are $\{-\sqrt{x^2+z^2}, +\sqrt{x^2+z^2}\}$. Therefore, the two eigenvalues of $\frac{1}{2}(I + xX + zZ)$ are $\frac{1}{2}(1 - \sqrt{x^2+z^2})$ and $\frac{1}{2}(1 + \sqrt{x^2+z^2})$. Substituting them into the definition of von Neumann entropy thus gives

$$S\left(\frac{1}{2}(I + xX + zZ)\right) = h\left(\frac{1-\sqrt{x^2+z^2}}{2}\right) = s(1 - x^2 - z^2). \qquad \square$$

### A.9 Propositions 15 and 16: Masking

In this subsection, we prove Propositions 15 and 16.

*Proof of Proposition 15.* First, observe that the average state is $\bar{\Sigma} := \frac{1}{2}(\Sigma_0 + \Sigma_1) = \frac{1}{2}\bar{\tau}_\alpha \oplus \frac{1}{2}\bar{\tau}_\alpha$. Let

$$N_0 := (1-\alpha)\bar{\tau}_\alpha^{-1/2}\tau_0\bar{\tau}_\alpha^{-1/2}, \qquad N_1 := \alpha\bar{\tau}_\alpha^{-1/2}\tau_1\bar{\tau}_\alpha^{-1/2},$$

be the PGM observables of the non-uniform binary ensemble $\{(1-\alpha, \tau_0), (\alpha, \tau_1)\}$. Since $\bar{\Sigma}^{-1/2} = \sqrt{2}(\bar{\tau}_\alpha^{-1/2} \oplus \bar{\tau}_\alpha^{-1/2})$ on its support, the PGM definition gives

$$M_0 = N_0 \oplus N_1, \qquad M_1 = N_1 \oplus N_0.$$

Thus, reading the block first leaves the non-uniform PGM in either block, with the outcome labels reversed in the second. Its error is

$$\begin{aligned}
p_e(\Sigma_0, \Sigma_1) &= \frac{1}{2}\operatorname{Tr}(\Sigma_0 M_1) + \frac{1}{2}\operatorname{Tr}(\Sigma_1 M_0) \\
&= \frac{1-\alpha}{2}\operatorname{Tr}(\tau_0 N_1) + \frac{\alpha}{2}\operatorname{Tr}(\tau_1 N_0) + \frac{\alpha}{2}\operatorname{Tr}(\tau_1 N_0) + \frac{1-\alpha}{2}\operatorname{Tr}(\tau_0 N_1) \\
&= (1-\alpha)\operatorname{Tr}(\tau_0 N_1) + \alpha\operatorname{Tr}(\tau_1 N_0) \\
&= \alpha(1-\alpha)\operatorname{Tr}(\tau_0\bar{\tau}_\alpha^{-1/2}\tau_1\bar{\tau}_\alpha^{-1/2}) + \alpha(1-\alpha)\operatorname{Tr}(\tau_1\bar{\tau}_\alpha^{-1/2}\tau_0\bar{\tau}_\alpha^{-1/2}) \\
&= 2\alpha(1-\alpha)\operatorname{Tr}(\tau_0\bar{\tau}_\alpha^{-1/2}\tau_1\bar{\tau}_\alpha^{-1/2}),
\end{aligned}$$

where cyclicity of the trace gives the last equality. This is also the PGM error of $\{(1-\alpha, \tau_0), (\alpha, \tau_1)\}$, proving (22). $\square$

*Proof of Proposition 16.* If the eigenvalues of $\tau_0$ and $\tau_1$ are $\{\lambda_i\}_i$ and $\{\mu_j\}_j$, respectively, then the nonzero eigenvalues of $\Sigma_0$ are $\{(1-\alpha)\lambda_i\}_i$ and $\{\alpha\mu_j\}_j$. Therefore

$$S(\Sigma_0) = S(\Sigma_1) = h(\alpha) + (1-\alpha)S(\tau_0) + \alpha S(\tau_1).$$

Likewise, the two copies of $\bar{\tau}_\alpha$ in $\bar{\Sigma} = \frac{1}{2}\bar{\tau}_\alpha \oplus \frac{1}{2}\bar{\tau}_\alpha$ give $S(\bar{\Sigma}) = 1 + S(\bar{\tau}_\alpha)$. Substitution in the Holevo identity proves (23). $\square$



ECCC ISSN 1433-8092
https://eccc.weizmann.ac.il