Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# Certified Boolean Quantization: Seven-Bit Counterexamples and Diffuse Gaussian Limits

**Sushanth Dasari**

sushanthdasari.sd@gmail.com

**Research Article**

**Keywords:** Boolean quantization, coding theory, coordinate influence, Gaussian limits, mutual information, verified computation

**Posted Date:** November 3rd, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-11585538/v1

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

**Additional Declarations:** No competing interests reported.

# Certified Boolean Quantization: Seven-Bit Counterexamples and Diffuse Gaussian Limits

Sushanth Dasari

Amity International School, Sector 46

Gurugram, Haryana 122002, India

Corresponding author: sushanthdasari.sd@gmail.com

**Abstract**

Coordinate selection is a natural benchmark for compressing independent binary variables observed through binary noise. We construct two explicit fourteen-input, seven-bit quantizers that beat this benchmark near complete noise, including an exactly balanced quantizer. We then analyze a diffuse block winner-take-all family whose maximum coordinate influence vanishes and whose Gaussian limit crosses the coordinate benchmark between seven and eight output bits. Directed arithmetic certifies finite eight-bit witnesses and their noise windows. We also prove influence asymptotics, robustness to unequal blocks, sufficiency of noisy scores, and uniform convergence of the full information curve.

**Keywords:** Boolean quantization; coding theory; coordinate influence; Gaussian limits; mutual information; verified computation

**MSC 2020:** 94A29, 60F05, 65G30

## 1 Introduction

Selecting $k$ input coordinates is a natural way to compress independent binary measurements to $k$ bits. It preserves those measurements exactly, but may discard collective structure that a multivariate representation retains. This paper studies explicit counterexamples to coordinate optimality and a separate requirement that no individual measurement substantially affect the output.

Let $X$ be uniform on $\{-1,1\}^n$, and let $Y_\rho$ be its coordinatewise $\rho$-correlated copy. Thus $\mathbb{P}(Y_{\rho,i} \neq X_i) = (1-\rho)/2$, independently across coordinates. For a deterministic quantizer $U = Q(X)$ with at most $2^k$ labels, the coordinate benchmark is

$$I_2((X_1,\dots,X_k);Y_\rho) = k\phi(\rho), \qquad \phi(\rho) = 1 - h_2((1-\rho)/2). \tag{1}$$

Here $I_2, H_2$ use base-two logarithms; $I_{\mathrm{nat}}, H_{\mathrm{nat}}$ use natural logarithms. Unless stated otherwise, $0 \le \rho \le 1$. The output budget is cardinality; when exact uniformity is required, we state it separately.

For a finite representation define

$$B(U) = \sum_{i=1}^{n} \mathbb{E}[\mathbb{E}(X_i \mid U)^2],$$
$$\mathrm{Inf}_i(Q) = \mathbb{P}(Q(X) \neq Q(X^{\oplus i})). \tag{2}$$

The quantity $B$ is both explained variance in squared-error reconstruction and the coefficient of $\rho^2/(2\ln 2)$ in noisy information. Coordinate selection has $B = k$. A sequence is diffuse when its maximum flip influence tends to zero.

## 1.1 Prior work and the question addressed here

Courtade and Kumar introduced the one-bit information problem [1]. Multibit improvements and the normalized high-output-budget optimum $\rho^2$ predate its resolution: see V. Chandar and A. Tchamkerten, *Most informative quantization functions*, unpublished 2014 workshop note, Corollary 1 (author-hosted version), following conditional-entropy methods of Witsenhausen and Wyner [5]. In September 2026 the one-bit conjecture was independently resolved by Mahdavifar and Beirami (arXiv:2609.26444v2) and OpenAI (*Sharp binary-information contraction on the discrete cube*). Mahdavifar and Beirami additionally give a shortened-Hamming eight-bit example and conjecture coordinate optimality for budgets two through seven. The seven-bit constructions below contradict only that multibit conjecture; they do not challenge either one-bit result.

Our finite codebook uses the classical $\mathbb{Z}_4$ Gray-image construction associated with Nordstrom–Robinson coding [2]. We claim neither a new code nor a new decoding principle. The results here are exact information-coefficient evaluations for specified quantizers, finite noise guarantees, and a balanced tie allocation preserving the strict seven-bit violation. The smallest unrestricted violating budget remains undetermined.

For the diffuse construction, the Gaussian connection has substantial precedents. Low-influence invariance and Gaussian partition methods were developed by Mossel, O'Donnell and Oleszkiewicz [3] and Isaksson and Mossel [4]. The entropy bound for $B$ is the entropic level-one inequality [6]; the explained-variance objective is a squared-error quantization objective [7]. Our block central-limit argument does not require an invariance principle or a conjecture about the optimal Gaussian partition. We use these established ideas to obtain explicit finite constructions and their influence and information limits.

## 1.2 Main results

Two fourteen-input quantizers with 128 nonempty labels satisfy

$$B(Q^{\mathrm{lex}}) = 7.0199613231\ldots > 7, \tag{3}$$
$$B(Q^{\mathrm{bal}}) = \frac{1837931}{262144} > 7, \qquad H_2(Q^{\mathrm{bal}}(X)) = 7. \tag{4}$$

They beat seven-coordinate selection throughout $0 < \rho \le 0.13$ and $0 < \rho \le 0.09$, respectively. Theorem 3.1 gives complete definitions and certificates. Adding independent reported coordinates extends the balanced violation to every $k \ge 7$.

To distribute sensitivity across many inputs, partition $Km$ independent signs into $K$ blocks and set

$$S_{j,m} = \frac{1}{\sqrt{m}} \sum_{r=1}^{m} X_{j,r}, \qquad Q_{K,m}(X) = \min \arg\max_{j \in [K]} S_{j,m}. \tag{5}$$

This block winner-take-all (WTA) quantizer uses deterministic ties. For independent standard Gaussians $G_1, \dots, G_K$, put

$$\mu_K = \mathbb{E} \max_{j \le K} G_j, \qquad \beta_K = \frac{K}{K-1} \mu_K^2. \tag{6}$$

We prove $B(Q_{K,m}) \to \beta_K$ and

$$\sqrt{m} \max_i \mathrm{Inf}_i(Q_{K,m}) \longrightarrow \frac{2\mu_K}{K}. \tag{7}$$

The exact family-wise transition is

$$\beta_{2^k} < k \quad (1 \le k \le 7), \qquad \beta_{2^k} > k \quad (k \ge 8). \tag{8}$$

Consequently every $k \ge 8$ admits WTA quantizers with arbitrarily small maximum influence and a strict information advantage near complete noise. This is a statement about this family, not a universal diffuse threshold. The finite seven-bit code quantizers above have substantial influences and give no seven-bit diffuse sequence.

Exact binomial formulas and independently implemented directed-arithmetic evaluators certify

$$B(Q_{256,428}) > 8.00000587080984, \tag{9}$$

$$B(Q_{256,1000}) > 8.01287871372770. \tag{10}$$

Their maximum influences are below 0.001207177 and 0.000756704, respectively. The first beats eight coordinates throughout $0 < \rho \le 0.002$; the second throughout $0 < \rho \le 0.098$. A majority baseline separates diffuse averaging from winner selection.

Finally, we prove exact sufficiency of noisy block scores and convergence of the entire information curve uniformly in $\rho$ for each fixed $K$. The limits extend to standardized unequal blocks when every block diverges. A diagonal WTA sequence gives a diffuse realization of the previously known high-output-budget optimum.

The model uses independent fair bits and independent binary noise. It is a mathematical benchmark, not a validated sensor model or a practical compression claim. The distinction between a short finite counterexample and a diffuse existence construction is central to the results.

## 2 Reconstruction and noisy information

The statistic $B$ has an exact reconstruction interpretation. Since the conditional mean minimizes squared error,

$$\min_g \mathbb{E}\|X - g(U)\|^2 = n - B(U). \tag{11}$$

Indeed, expand $X = (X - \mathbb{E}[X \mid U]) + \mathbb{E}[X \mid U]$ and use orthogonality of the two terms in $L^2$. Reporting $k$ coordinates has $B = k$ and risk $n - k$. Thus a certificate $B > k$ is simultaneously a strict improvement in centroid reconstruction and a high-noise information comparison. The absolute reduction is small in the examples; it is not a substantial reduction in error per input bit.

In this section $U$ may be randomized: the assumption is the Markov chain $U - X - Y_\rho$, so its internal randomization is independent of the channel noise given $X$. Let $a_i(u) = \mathbb{E}[X_i \mid U = u]$ and define

$$d(a) = \frac{1+a}{2}\ln(1+a) + \frac{1-a}{2}\ln(1-a), \tag{12}$$

with $0 \ln 0 = 0$. For $|a| \le 1$,

$$d(a) = \sum_{r \ge 1} \frac{a^{2r}}{2r(2r-1)}, \quad d(a) \ge a^2/2, \quad d(\rho a) \le \rho^2 d(a). \tag{13}$$

The series follows by integrating $d'(a) = \operatorname{arctanh}(a)$ and extends to the endpoints by continuity.

**Proposition 2.1** (Entropy and channel sandwich)**.** *For every finite $U$ as above,*

$$B(U) \le 2I_{\mathrm{nat}}(U;X) \le 2H_{\mathrm{nat}}(U), \tag{14}$$

$$\frac{\rho^2}{2\ln 2} B(U) \le I_2(U;Y_\rho) \le \rho^2 I_2(U;X). \tag{15}$$

*In particular, $B(U) \le 2 \ln K$ if $U$ has at most $K$ labels.*

*Proof.* Relative entropy to a product of fair signs dominates the sum of marginal relative entropies. Thus

$$D(P_{X|u} \| P_X) \ge \sum_i d(a_i(u)) \ge \frac{1}{2} \sum_i a_i(u)^2. \tag{16}$$

Averaging proves (14). The output is also a product of fair signs, and $\mathbb{E}[Y_{\rho,i} \mid U = u] = \rho a_i(u)$. Applying (16) to $Y_\rho$ proves the lower bound in (15).

For completeness, the upper bound admits a direct tensorization proof. For a fair sign $Z$, a finite auxiliary variable $V$, and a BSC output $W$, (13) gives $I_{\mathrm{nat}}(V;W) \le \rho^2 I_{\mathrm{nat}}(V;Z)$. Conditional on $Y_{<i}$, the sign $X_i$ is still fair and its channel noise is independent of $(U, Y_{<i})$. Therefore

$$I_{\mathrm{nat}}(U;Y_\rho) = \sum_i I_{\mathrm{nat}}(U;Y_i \mid Y_{<i})$$

$$\le \rho^2 \sum_i I_{\mathrm{nat}}(U; X_i \mid Y_{<i})$$
$$\le \rho^2 \sum_i I_{\mathrm{nat}}(U; X_i \mid X_{<i}) = \rho^2 I_{\mathrm{nat}}(U; X).$$

For the second inequality, $X_i$ is independent of both $X_{<i}$ and $Y_{<i}$, while $X_i - (U, X_{<i}) - (U, Y_{<i})$ is a Markov chain. Data processing applies term by term. Division by $\ln 2$ completes the proof. □

*Remark* 2.2. For a deterministic cell $A$ of measure $\alpha$, (16) yields $\sum_i \mathbb{E}[X_i \mid A]^2 \le 2\ln(1/\alpha)$, the entropic form of Chang's level-one bound [6]. Averaging over cells gives (14). The upper channel inequality is the standard uniform-product BSC contraction underlying [5] and the Chandar–Tchamkerten note discussed in Section 1; its proof is included to make our dependence explicit.

**Proposition 2.3** (Exact second-order coefficient)**.** *For fixed finite $n$ and $U$,*

$$I_2(U; Y_\rho) = \frac{\rho^2}{2\ln 2} B(U) + O_U(\rho^4), \qquad \rho \to 0. \tag{17}$$

*The remainder is not asserted to be uniform as $n$, $K$, or $m$ varies.*

*Proof.* For $p_u = \mathbb{P}(U = u) > 0$ define the likelihood ratio $f_u(x) = \mathbb{P}(U = u \mid X = x)/p_u$. In the Walsh basis $\chi_A(x) = \prod_{i \in A} x_i$,

$$f_u(x) = \sum_{A \subseteq [n]} \widehat{f}_u(A)\chi_A(x),$$
$$\widehat{f}_u(A) = \mathbb{E}[\chi_A(X) \mid U = u], \qquad \widehat{f}_u(\varnothing) = 1.$$

The likelihood ratio of $Y_\rho \mid U = u$ relative to uniform is $T_\rho f_u(y) = \sum_A \rho^{|A|} \widehat{f}_u(A)\chi_A(y)$. There are finitely many $u, y$, so these ratios are positive in a common neighborhood of zero. Expanding $(1+z)\ln(1+z) = z + z^2/2 + O(z^3)$ and using Walsh orthogonality gives the quadratic coefficient $\frac{1}{2}\sum_i a_i(u)^2$ in nats. Furthermore $T_{-\rho} f_u(y) = T_\rho f_u(-y)$, so its averaged entropy is an even analytic function of $\rho$. The cubic term vanishes, and the remainder is $O_U(\rho^4)$. Average over $u$ and convert to bits. □

**Corollary 2.4** (An explicit sufficient noise window)**.** *If $U = Q(X)$ takes at most $2^k$ values and $B(U) > k$, then it beats $k$ coordinates whenever*

$$0 < \rho < 1, \qquad B(U) > k\left(1 + \frac{\rho^2}{6(1-\rho^2)}\right). \tag{18}$$

*More generally, an endpoint check $\rho_0^2 B(U)/2 > k d(\rho_0)$ proves the improvement for every $0 < \rho \le \rho_0$.*

*Proof.* In (13), every denominator with $r \ge 2$ is at least 12, so $d(\rho) \le \rho^2/2 + \rho^4/[12(1-\rho^2)]$. Combine this with (15). For the second assertion, $d(\rho)/\rho^2$ is strictly increasing on $(0, 1)$ by its positive-coefficient series. □

## 3 Seven-bit counterexamples

We identify binary words with integers, using bit zero as the least significant bit, and use signs $(-1)^{x_j}$ when computing $B$. Over $\mathbb{Z}_4$, set

$$G = \begin{pmatrix} 1 & 0 & 0 & 0 & 2 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 2 & 1 & 3 \\ 0 & 0 & 1 & 0 & 1 & 3 & 2 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 3 & 2 \end{pmatrix}.$$

For $u \in \mathbb{Z}_4^4$, let $z = uG$ and define

$$w(u) = \sum_{j=0}^{7} g(z_j) 2^{2j}, \qquad (g(0), g(1), g(2), g(3)) = (0, 1, 3, 2).$$

Retain $w$ with bit zero equal to zero and delete its first two bits:

$$\mathcal{C} = \{\lfloor w(u)/4 \rfloor : w(u) \equiv 0 \pmod 2\}. \tag{19}$$

There are 128 distinct fourteen-bit words. Indeed, $u_0$ has two allowed values, the retained systematic symbols determine $u_1, u_2, u_3$, and the retained symbol $z_5 = u_0 + 2u_1 + 3u_2 + u_3$ determines $u_0$. Order the words as $c_0 < \cdots < c_{127}$. This is a specified shortening and puncturing of a classical Gray-image construction [2]; its identification with a named code is unnecessary for verification.

Define $Q^{\mathrm{lex}}(x)$ as the smallest index minimizing $d_H(x, c_a)$. For a balanced variant, first assign every input with nearest distance at most two to its unique nearest word. Each word receives 106 such inputs. Process the remaining 2816 inputs increasingly. A residual input can be assigned only to a nearest word, and each word has 22 residual slots. Find an augmenting path by breadth-first search, starting at the new input: visit candidate labels increasingly, then their currently assigned residual inputs increasingly; stop at the first label with a free slot. Reassign along the path. Call the resulting map $Q^{\mathrm{bal}}$. The exact certificate below verifies that no search fails. Online Resource 1 specifies the algorithm and certificate checks in detail.

**Theorem 3.1** (Explicit and balanced seven-bit violations)**.** *The preceding maps have 128 nonempty labels and satisfy*

$$B(Q^{\mathrm{lex}}) = \frac{1309967034269456355941480179}{186606018747259783459307520} > 7 + \frac{1}{64}, \tag{20}$$

$$B(Q^{\mathrm{bal}}) = \frac{1837931}{262144} > 7 + \frac{1}{100}. \tag{21}$$

*Every balanced cell has 128 inputs, so $H_2(Q^{\mathrm{bal}}(X)) = 7$ exactly. Their maximum influences are $737/1024$ and $6321/8192$, respectively. Moreover,*

$$I_2(Q^{\mathrm{lex}}(X); Y_\rho) > 7\phi(\rho) \qquad (0 < \rho \le 13/100),$$

$$I_2(Q^{\mathrm{bal}}(X); Y_\rho) > 7\phi(\rho) \qquad (0 < \rho \le 9/100).$$

*Proof.* The finite part is computer assisted and uses only integer enumeration and rational arithmetic. For each label $a$, compute

$$n_a = |Q^{-1}(a)|, \qquad s_{aj} = \sum_{x:Q(x)=a} (-1)^{x_j}.$$

The definition of conditional expectation gives

$$B(Q) = 2^{-14} \sum_{a=0}^{127} \frac{\sum_{j=0}^{13} s_{aj}^2}{n_a}. \tag{22}$$

The supplied generator reconstructs all codewords from $G$ and all labels from the stated rules. It enumerates all $2^{14}$ inputs, verifies nearest-word membership, and checks influences by all fourteen coordinate flips. For $Q^{\mathrm{lex}}$, the exact sum is (20). For $Q^{\mathrm{bal}}$, it verifies $n_a = 128$ for every $a$ and

$$\sum_{a,j} s_{aj}^2 = 14{,}703{,}448,$$

which gives (21). Each augmenting path is saved, allowing a second implementation to replay it, check capacities and nearest-word membership, and reconstruct the final balanced partition without trusting its lookup table.

Independent verifiers use reverse input traversal and unordered within-cell Hamming distances. For either map,

$$B(Q) = 14 - \frac{4}{2^{14}} \sum_{a} \frac{1}{n_a} \sum_{\{x,x'\} \subset Q^{-1}(a)} d_H(x, x'). \tag{23}$$

This identity follows by expanding $s_{aj}^2$: for a cell, $\sum_j s_{aj}^2 = 14n_a^2 - 4\sum_{\{x,x'\}} d_H(x, x')$. Both methods give the same exact values and label hashes. Online Resource 2 includes complete counts, moments, paths and verification programs, rather than relying on a reported floating-point value.

For the balanced map, Corollary 2.4 applies directly. At $r = 9/100$,

$$B(Q^{\mathrm{bal}}) - 7 - \frac{7r^2}{6(1-r^2)} = \frac{602947}{371458048} > 0.$$

The bracket decreases with $r$, proving its whole window. For the lexicographic map use the positive series for $d$. With $z = r^2$,

$$\frac{d(r)}{r^2} \le \sum_{j=1}^{20} \frac{z^{j-1}}{2j(2j-1)} + \frac{z^{20}}{42 \cdot 41(1-z)}. \tag{24}$$

At $r = 13/100$, exact rational comparison of (20) with twice seven times the right side gives a strict gap. Since $d(r)/r^2$ is increasing, the endpoint test in Corollary 2.4 gives every smaller positive correlation. □

**Table 1** Seven-bit constructions. Both use fourteen inputs and 128 labels; the noise windows are sufficient guarantees, not maximal intervals

| Tie rule | $B$ | Output entropy | Certified $\rho$ window |
|---|---|---|---|
| Lexicographic | $7.0199613231\ldots$ | $< 7$ | $(0, 0.13]$ |
| Balanced allocation | $1837931/262144$ | $7$ | $(0, 0.09]$ |

**Corollary 3.2** (Every larger output budget)**.** *For every integer $k \geq 7$, a deterministic map on $k+7$ independent fair bits has exactly $2^k$ equiprobable labels and beats $k$-coordinate selection throughout $0 < \rho \leq 0.09$.*

*Proof.* Apply $Q^{\mathrm{bal}}$ to fourteen inputs and report $k-7$ additional independent coordinates. Independence makes the output uniform and gives

$$I_2((Q^{\mathrm{bal}}(X_1^{14}), X_{15}^{k+7}); Y_\rho) = I_2(Q^{\mathrm{bal}}(X_1^{14}); Y_{\rho,1}^{14}) + (k-7)\phi(\rho).$$

For $k = 7$, omit the empty coordinate tuple. Theorem 3.1 proves the strict comparison. □

The codebook is classical, while the table concerns its specified information representations. Exact balance is an additional constraint: the second row shows that a seven-bit violation is not an artifact of allowing unequal output masses. These maps are not diffuse. Neither the direct-sum argument nor padding with unused inputs makes their nonzero coordinate influences tend to zero. No assertion about budgets two through six follows.

# 4 Block Quantization and Its Gaussian Limit

## 4.1 A coding and geometric interpretation

**Proposition 4.1** (Equal-distance code representation)**.** *Let $c^{(j)} \in \{-1, 1\}^{Km}$ be $+1$ on block $j$ and $-1$ on every other block. Then $Q_{K,m}$ is nearest-neighbor Hamming quantization to the $K$ codewords $c^{(j)}$, with the same lexicographic tie rule. Distinct codewords have Hamming distance $2m$.*

*Proof.* Writing $A_j = \sum_r X_{j,r}$ gives $\langle c^{(j)}, X\rangle = 2A_j - \sum_\ell A_\ell$. Since $d_H(c^{(j)}, X) = (Km - \langle c^{(j)}, X\rangle)/2$, minimizing Hamming distance is precisely maximizing $A_j$. Two codewords differ in exactly their two distinguished blocks. □

Thus the construction can be implemented by block addition and comparison, but it is not outside the class of code-based quantizers. After subtracting the common centroid, its codewords form a regular simplex. This also describes its Gaussian limit. The number of labels, rather than the number of input bits, is the rate parameter here.

## 4.2 Coefficient and influence limits

**Theorem 4.2.** *For each fixed $K \geq 2$,*

$$B(Q_{K,m}) \longrightarrow \beta_K. \tag{25}$$

*For every block $j$ and position $r$,*

$$\mathrm{Inf}_{(j,r)}(Q_{K,m}) = \frac{2}{\sqrt{m}} \mathbb{E}[S_{j,m} \mathbf{1}\{Q_{K,m} = j\}], \tag{26}$$

*and consequently*

$$\sqrt{m}\, \mathrm{Inf}_{(j,r)}(Q_{K,m}) \longrightarrow \frac{2\mu_K}{K}. \tag{27}$$

*The same limit holds for the maximum influence. For all $m$,*

$$\max_i \mathrm{Inf}_i(Q_{K,m}) \leq \min\left\{1, \frac{2(K-1)}{\sqrt{\pi m}}\right\}. \tag{28}$$

*Proof.* Write $U_m = Q_{K,m}$, $p_{j,m} = \mathbb{P}(U_m = j)$, and $b_{\ell j,m} = \mathbb{E}[S_{\ell,m} \mid U_m = j]$. Exchangeability within each block gives

$$\begin{aligned} \mathbb{E}[X_{\ell,r} \mid U_m = j] &= b_{\ell j,m}/\sqrt{m}, \\ B(U_m) &= \sum_j p_{j,m} \sum_\ell b_{\ell j,m}^2. \end{aligned} \tag{29}$$

The multivariate central limit theorem yields $S_m \Rightarrow G$. The Gaussian tie hyperplanes have probability zero, so $p_{j,m} \to 1/K$. Moreover $S_{\ell,m}\mathbf{1}\{U_m = j\}$ has convergent first moments: convergence in distribution follows from the zero boundary measure, and uniform integrability follows from $\mathbb{E}S_{\ell,m}^2 = 1$. Thus $b_{\ell j,m} \to \mathbb{E}[G_\ell \mid J = j]$, where $J = \arg\max_i G_i$.

Permutation symmetry gives $\mathbb{E}[G_j \mid J = j] = \mu_K$ and a common loser mean $\nu_K$. The Gaussian common mode $\sum_i G_i$ is independent of all score differences, hence of $J$. Its conditional mean is zero, implying $\nu_K = -\mu_K/(K-1)$. Substitution into (29) proves (25).

To prove (26), condition on all bits except the focal one, denoted $x$. As $x$ changes from $-1$ to $+1$, only the score of block $j$ increases. The output changes if and only if the indicator $f = \mathbf{1}\{U_m = j\}$ changes from zero to one; if block $j$ loses both times, the best competitor is unchanged. Thus

$$\begin{aligned} \mathbb{E}[xf \mid \text{other bits}] &= \tfrac{1}{2}(f(+1) - f(-1)), \\ \mathbb{E}[X_{j,r} f] &= \tfrac{1}{2}\, \mathrm{Inf}_{(j,r)}(Q_{K,m}). \end{aligned}$$

Summing over the $m$ exchangeable positions proves (26). The same first-moment convergence gives $\mathbb{E}[S_{j,m}\mathbf{1}\{U_m = j\}] \to \mu_K/K$, which proves (27), also for the maximum because $K$ is fixed.

Finally let $R$ be the sum of the other $m-1$ bits in block $j$ and $A_\ell$ a competing block sum. A label change requires $A_\ell \in [R-1, R+1]$ for some $\ell \neq j$. Parity forces $R - A_\ell = \pm 1$. This difference is a sum of $2m-1$ independent signs, and

$$\mathbb{P}(|R - A_\ell| = 1) = 2\binom{2m}{m} 4^{-m} \leq \frac{2}{\sqrt{\pi m}}.$$

For the last estimate, use $\binom{2m}{m}4^{-m} = (2/\pi)\int_0^{\pi/2}\cos^{2m}t\,dt$ and $\ln\cos t \le -t^2/2$; extension of the resulting Gaussian integral to $[0,\infty)$ gives $1/\sqrt{\pi m}$. A union bound proves (28). □

For $G \sim N(0, I_d)$ and a finite Gaussian partition $J = q(G)$ define

$$\Gamma(q) = \mathbb{E}\|\mathbb{E}[G \mid J]\|^2 = d - \mathbb{E}\|G - \mathbb{E}[G \mid J]\|^2. \tag{30}$$

This is the explained variance, equivalently the reduction in squared-error distortion from centroid reconstruction [7]. For the argmax partition it is $\beta_K$. Removing the independent common Gaussian mode leaves $K$ regular-simplex cones in dimension $K-1$. Neither (30) nor the simplex interpretation asserts global optimality over all $K$-cell partitions.

### 4.3 Unequal block sizes

Equal block lengths simplify exact finite evaluation but are unnecessary for the limiting geometry. For positive integers $m_1, \dots, m_K$, use standardized scores $S_j = m_j^{-1/2}\sum_{r=1}^{m_j} X_{j,r}$ and the same first-index tie rule.

**Proposition 4.3.** *For fixed $K$, as $\min_j m_j \to \infty$,*

$$B(Q_{K,\boldsymbol{m}}) \to \beta_K, \qquad \sqrt{m_j}\,\mathrm{Inf}_{(j,r)}(Q_{K,\boldsymbol{m}}) \to 2\mu_K/K$$

*for every $j, r$, and $\max_i \mathrm{Inf}_i \to 0$. No restriction on the ratios $m_j/m_\ell$ is required.*

*Proof.* The independent standardized block scores converge jointly to $N(0, I_K)$. Their squared moments equal one, so the truncated first moments used in Theorem 4.2 converge. Within-block exchangeability now gives $\mathbb{E}[X_{\ell,r} \mid U = j] = \mathbb{E}[S_\ell \mid U = j]/\sqrt{m_\ell}$; summing over the $m_\ell$ positions cancels the block lengths in $B$. The limiting centroid calculation is unchanged. Increasing one bit in block $j$ can change the label only by making that block the winner, even when other blocks use different scales. Thus

$$\mathrm{Inf}_{(j,r)} = \frac{2}{\sqrt{m_j}}\mathbb{E}[S_j \mathbf{1}\{U = j\}],$$

and convergence of the first moment proves the influence limit. There are finitely many blocks, so the maximum tends to zero. □

## 5 Certified finite constructions and comparison baselines

Let $A$ be a sum of $m$ fair signs. Its alphabet is $\mathcal{A}_m = \{-m, -m+2, \dots, m\}$ and

$$p(a) = 2^{-m}\binom{m}{(a+m)/2},$$

$$F(a) = \mathbb{P}(A \le a), \qquad F_-(a) = \mathbb{P}(A < a).$$

Also let $M(a) = \mathbb{E}[A\mathbf{1}\{A \le a\}]$ and $M_-(a) = \mathbb{E}[A\mathbf{1}\{A < a\}]$. For label $j$, set $e = j-1$, $\ell = K-j$, and define unnormalized moments

$$p_j = \sum_a p(a) F_-(a)^e F(a)^\ell, \tag{31}$$

$$w_j = \sum_a a p(a) F_-(a)^e F(a)^\ell, \tag{32}$$

$$v_j^- = \sum_a p(a) M_-(a) F_-(a)^{e-1} F(a)^\ell \quad (e > 0), \tag{33}$$

$$v_j^+ = \sum_a p(a) M(a) F_-(a)^e F(a)^{\ell-1} \quad (\ell > 0). \tag{34}$$

Terms involving an absent group of losers are omitted; zero powers, including $0^0$, are interpreted as one.

**Proposition 5.1** (Finite coefficient and influences)**.** *With the preceding notation,*

$$B(Q_{K,m}) = \frac{1}{m} \sum_{j=1}^{K} \frac{w_j^2 + e(v_j^-)^2 + \ell(v_j^+)^2}{p_j}. \tag{35}$$

*If $R$ is a sum of $m-1$ fair signs and $W_j(a) = F_-(a)^{j-1} F(a)^{K-j}$, then*

$$\mathrm{Inf}_{(j,r)}(Q_{K,m}) = \mathbb{E}[W_j(R+1) - W_j(R-1)] = 2w_j/m. \tag{36}$$

*Proof.* Condition on $A_j = a$. The event $U = j$ requires all earlier blocks to be strictly below $a$ and all later blocks to be at most $a$, giving (31). Multiplication by $a$ gives the winning first moment (32). For an earlier or later loser, replacing its cumulative probability by its truncated first moment gives (33) or (34). Dividing these moments by $p_j$ and using block exchangeability proves (35). Each $p_j$ is positive, for example on the event that block $j$ is all positive and the others all negative.

Conditional on $R$, increasing the focal bit changes the label precisely when block $j$ starts winning. The probability is the difference of its two winning probabilities, proving the first equality in (36); the second is (26). □

**Theorem 5.2** (Certified eight-bit examples)**.** *The following strict enclosures hold:*

$$\begin{aligned}
&8.00000587080984 < B(Q_{256,428}) < 8.00000587080986,\\
&8.01287871372770 < B(Q_{256,1000}) < 8.01287871372772,\\
&\max_i \mathrm{Inf}_i(Q_{256,428}) < 0.001207177,\\
&\max_i \mathrm{Inf}_i(Q_{256,1000}) < 0.000756704.
\end{aligned}$$

*Moreover $B(Q_{256,427}) < 7.99995311234123 < 8$. The maps at $m = 428$ and $m = 1000$ beat eight-coordinate projection throughout $0 < \rho \le 0.002$ and $0 < \rho \le 0.098$, respectively.*

*Proof.* The primary program `finite_certificate.py` evaluates (35) and the first expression in (36) using directed fixed-point integer intervals at scale $10^{70}$. Every multiplication and division is rounded outward by integer floor and ceiling. Exact binomial coefficients are enclosed before interval operations. An independent streaming implementation, `crosscheck_finite_arb.py`, uses Arb at 256-bit precision [8] and the exact truncated moments

$$M(2b-m) = -(m-b)p(2b-m), \quad M_-(2b-m) = -bp(2b-m).$$

Its interval results lie inside the primary enclosures for $m = 427, 428, 1000$. Both check total label mass; the Arb calculation also proves that block one has strictly largest influence. All displayed decimal bounds are compared as exact rationals or enclosed balls.

For $m = 428$, the lower endpoint exceeds $8[1 + \rho_0^2/(6(1-\rho_0^2))]$ at $\rho_0 = 1/500$, so Corollary 2.4 applies. For $m = 1000$ the endpoint $\rho_0 = 49/500$ satisfies the certified inequality

$$\frac{\rho_0^2 B(Q_{256,1000})/2 - 8d(\rho_0)}{\ln 2} > 1.65 \times 10^{-7}. \tag{37}$$

The endpoint part of Corollary 2.4 proves the whole interval. Equation (37) is a lower bound on the information advantage at the endpoint, not an evaluation of the full mutual information. □

The first example uses $256 \cdot 428 = 109{,}568$ input bits and the second 256,000. These are existence constructions, not improvements in input dimension over shortened-Hamming quantizers. The adjacent values at 427 and 428 certify a local crossing only; monotonicity or minimality in $m$ is not claimed. Exhaustive rational enumeration for $(K, m) = (2,2), (3,2), (3,3)$ independently checks the finite coefficient and maximum influence against the directed evaluator. Additional small-case checks of all influences, the code representation, posterior sufficiency and unequal blocks are included in Online Resource 2.

## 5.1 A diffuse baseline with the same output budget

For odd $L$, let $M_{k,L}$ report the majority signs of $k$ independent blocks of $L$ bits, using exactly $2^k$ possible labels. Put

$$c_L = 2^{-(L-1)} \binom{L-1}{(L-1)/2}.$$

**Proposition 5.3.** *The majority representation satisfies*

$$B(M_{k,L}) = kLc_L^2 \to 2k/\pi, \qquad \max_i \mathrm{Inf}_i(M_{k,L}) = c_L \to 0.$$

*Proof.* A bit changes its block's majority exactly when the other $L-1$ bits sum to zero, an event of probability $c_L$. Monotonicity gives $\mathbb{E}[X_i \,\mathrm{maj}(X)] = c_L$. The majority sign is balanced, so $\mathbb{E}[X_i \mid \mathrm{maj}(X)] = c_L \,\mathrm{maj}(X)$. Other block outputs are independent. Summing the squared conditional means gives $kLc_L^2$, and the central-binomial asymptotic yields the limits. □

**Table 2** Eight-bit comparisons in a common 109,568-input space. Decimal entries are rounded for display; interval and exact outputs accompany Online Resource 2

| Representation | Used inputs | $B$ | $\max \mathrm{Inf}_i$ |
|---|---|---|---|
| Coordinate selection | 8 | 8 | 1 |
| Shortened Hamming | 12 | 8.0625 | 0.875 |
| Eight block majorities | 109,560 | 5.093144 | 0.00681816 |
| 256-way WTA, $m = 428$ | 109,568 | 8.00000587 | 0.00120718 |

Table 2 compares the fixed eight-bit budget. Each finite map is embedded in 109,568 inputs by ignoring unused coordinates. Majority uses eight blocks of 13,695 bits, leaving eight coordinates unused. Thus the comparison fixes both available input and output budgets, while exposing the different influence levels; it is not an equal-influence optimality comparison. The shortened-Hamming map uses the syndrome rule in the Mahdavifar–Beirami preprint discussed in Section 1. Online Resource 1 restates that rule, and exhaustive rational enumeration independently gives $B = 129/16$ and maximum influence $7/8$.

The majority comparison shows why diffuseness alone is insufficient: both families distribute dependence, but independent sign decisions lose first-level information. WTA's competitive multiway partition retains enough explained variance to cross the coordinate benchmark. Conversely, the Hamming comparison prevents an efficiency claim: it achieves a larger coefficient with far fewer inputs, but substantial individual-coordinate influences.

Figure 1 reports a block-length sweep and an independently implemented evaluator using exact binomial truncated moments. The streaming evaluator uses $O(K)$ arithmetic objects rather than retaining all $O(Km)$ cumulative powers. This is a storage count, not a bit-complexity bound. Both programs agree by overlapping interval enclosures at $m = 427, 428, 1000$. The maximum influence and the entropy of the output are computed directly, with deterministic ties retained.

Figure 2 makes the scale of the guaranteed advantage explicit. Certification matters because the first coefficient excess is only about $5.87 \times 10^{-6}$ and because an asymptotic Gaussian advantage alone supplies no specific finite dimension or noise interval.

# 6 The Seven/Eight Coefficient Transition

Let $\varphi$ and $\Phi$ be the standard normal density and distribution function, with $\overline{\Phi} = 1 - \Phi$. Integration of the positive and negative tails of a maximum gives

$$\mu_K = \int_0^\infty f_K(t)\, dt, \qquad f_K(t) = 1 - \Phi(t)^K - \overline{\Phi}(t)^K. \tag{38}$$

For $K \geq 2$, $f_K$ is nonnegative and decreasing on $[0, \infty)$, since $f_K'(t) = -K\varphi(t)[\Phi(t)^{K-1} - \overline{\Phi}(t)^{K-1}] \leq 0$.

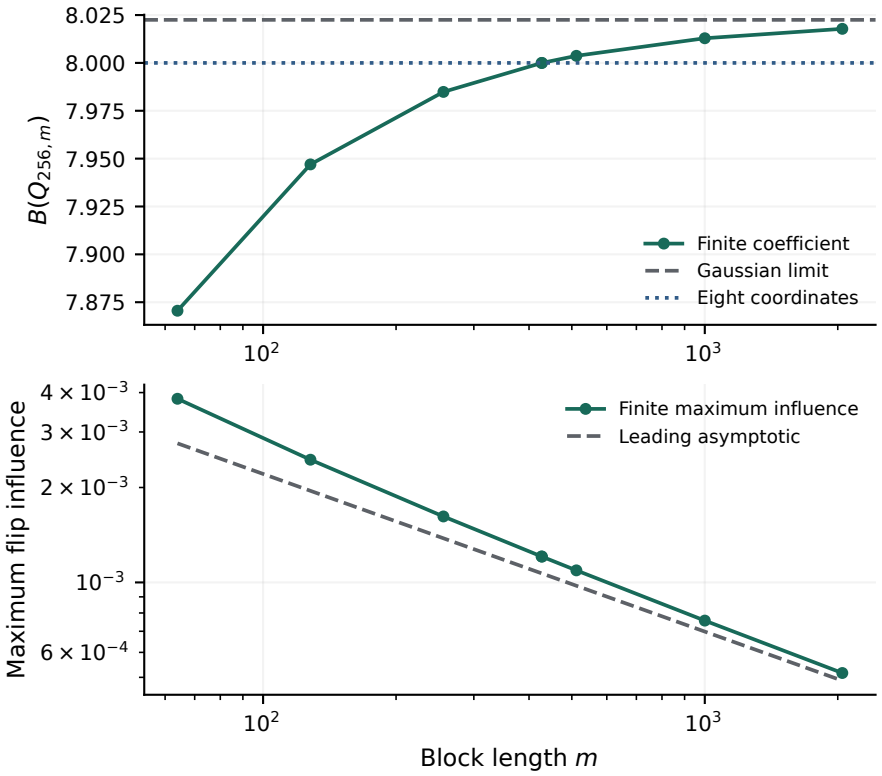



**Fig. 1** Certified finite coefficients and influences for 256-way WTA. Points come from 256-bit outward-rounded evaluation; connecting lines are visual guides. Dashed curves show the Gaussian coefficient limit and leading influence asymptotic. Arithmetic enclosure widths are smaller than the plotted symbols; no monotonicity in block length is inferred

**Theorem 6.1** (Exact transition within WTA)**.** *For every integer $k \geq 1$, the inequalities in* (8) *hold.*

*Proof.* The comparison $\beta_{2^k} \gtrless k$ is equivalent to $\mu_{2^k} \gtrless \tau_k$, where $\tau_k = \sqrt{k(1-2^{-k})}$. For $1 \leq k \leq 39$ we use a finite interval certificate, with a common mesh $h = 1/1000$ and $N = 12000$. Monotonicity gives

$$h \sum_{r=1}^{N} f_K(rh) \leq \mu_K \leq h \sum_{r=0}^{N-1} f_K(rh) + K\varphi(12). \tag{39}$$

Indeed $f_K \leq K\overline{\Phi}$ and

$$\int_{12}^{\infty} \overline{\Phi}(t)\, dt = \varphi(12) - 12\overline{\Phi}(12) \leq \varphi(12).$$

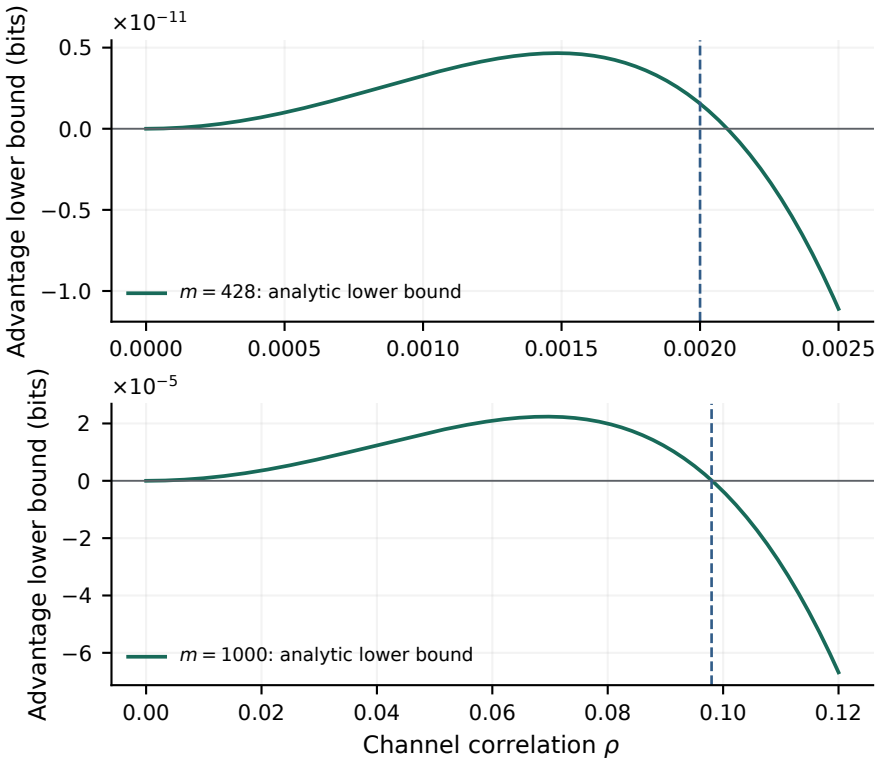



**Fig. 2** Analytic lower bounds on the advantage over eight coordinates, using certified rational lower bounds for $B$. Dashed vertical lines mark the certified endpoints 0.002 and 0.098. These are lower bounds, not full mutual-information evaluations; a negative lower bound does not establish inferiority

The script `certify_rate_transition.py` evaluates these sums with Arb ball arithmetic [8], using $\overline{\Phi}(t) = \operatorname{erfc}(t/\sqrt{2})/2$ and exact rational mesh points. It asserts that the upper enclosure is below $\tau_k$ for $k \le 7$, and the lower enclosure is above $\tau_k$ for $8 \le k \le 39$. In particular it proves

$$\mu_{128} < 2.595098 < 2.635396 < \tau_7,$$
$$\mu_{256} > 2.826363 > 2.822898 > \tau_8.$$

The code and complete outputs accompany the paper. No approximate quadrature value is needed for these assertions. As a check, $\beta_2 = 2/\pi < 1$ exactly.

For the infinite tail put $L = k \ln 2$, $K = 2^k$ and $T = (13/10)\sqrt{L}$. Integration by parts gives the elementary Mills bound

$$\overline{\Phi}(t) \ge \varphi(t)\frac{t}{1+t^2}, \qquad t > 0 : \tag{40}$$

indeed $\overline{\Phi}(t) = \varphi(t)/t - \int_t^\infty \varphi(s)s^{-2}ds \ge \varphi(t)/t - \overline{\Phi}(t)/t^2$. It follows that

$K\overline{\Phi}(T) \geq A(L)$, where

$$A(L) = \frac{(13/10)\sqrt{L}}{\sqrt{2\pi}[1 + (169/100)L]} e^{31L/200}. \tag{41}$$

Using $(1-q)^K \leq e^{-Kq}$ and $q^K \leq 2^{-K}$,

$$\mu_K \geq T f_K(T) \geq \frac{13}{10}\sqrt{L}\,[1 - e^{-A(L)} - 2^{-K}]. \tag{42}$$

For $L \geq 40 \ln 2 > 27$,

$$(\ln A)'(L) = \frac{31}{200} + \frac{1}{2L} - \frac{169/100}{1 + (169/100)L} > \frac{31}{200} - \frac{1}{2L} > 0.$$

The same certificate verifies $A(40 \ln 2) > 4.19$ and

$$\frac{13}{10}\sqrt{\ln 2}\,[1 - e^{-4.19} - 2^{-40}] > 1.06.$$

Since $2^{-K} \leq 2^{-40}$ for $k \geq 40$, (42) gives $\mu_{2^k} > 1.06\sqrt{k} > \tau_k$ for every such $k$. □

**Corollary 6.2.** *For every integer $k \geq 8$ and every $\eta > 0$, there exist finite $m$ and $\rho_0 > 0$ such that*

$$\max_i \mathrm{Inf}_i(Q_{2^k,m}) < \eta,$$
$$I_2(Q_{2^k,m}(X); Y_\rho) > k\phi(\rho) \quad (0 < \rho \leq \rho_0).$$

*Proof.* Theorems 4.2 and 6.1 let us choose $m$ with $B(Q_{2^k,m}) > k$ and the required influence bound. Corollary 2.4 then supplies a positive interval. □

Figure 3 locates the family-wise crossing. It concerns the order-$\rho^2$ coefficient, not an information comparison at every noise level and not an unrestricted threshold.

# 7 Sufficient Scores and the Full Information Limit

Define $T_{j,m} = m^{-1/2} \sum_r Y_{j,r}$ and $T_m = (T_{1,m}, \ldots, T_{K,m})$.

**Proposition 7.1** (Exact sufficiency)**.** *For every $K, m$ and $-1 \leq \rho \leq 1$,*

$$I_2(Q_{K,m}(X); Y_\rho) = I_2(Q_{K,m}(X); T_m). \tag{43}$$

*The same statement holds for any deterministic function of the noiseless block sums.*

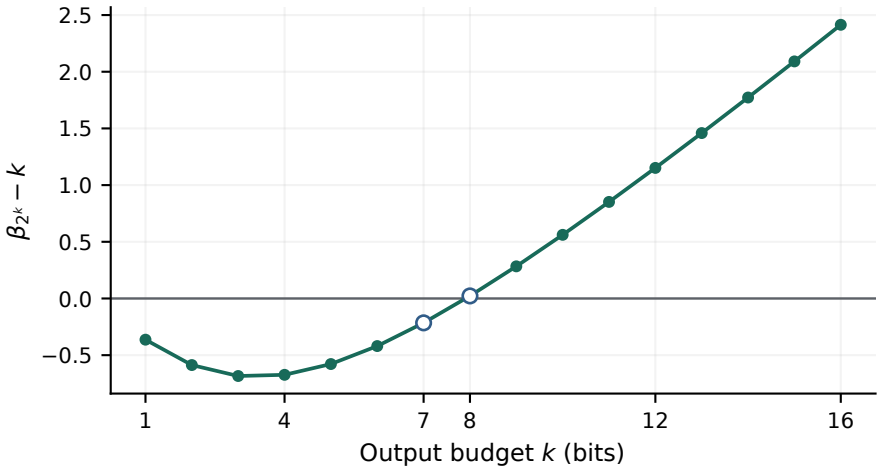



**Fig. 3** Gaussian coefficient excess over coordinate selection. The displayed values use floating-point quadrature for visualization; the signs for all integer budgets are established by interval sums and the analytic tail argument in Theorem 6.1

*Proof.* Conditional on $Y = y$, the input coordinates are independent and $\mathbb{P}(X_{j,r} = y_{j,r} \mid Y = y) = (1+\rho)/2$. The posterior law of each block sum depends only on the number of positive signs in that block of $y$, hence only on $T_{j,m}$. Different blocks remain conditionally independent. Therefore the posterior law of the quantizer depends on $y$ only through $T_m$. The quantizer and $Y$ are conditionally independent given $T_m$, which proves (43) by the chain rule. □

**Theorem 7.2** (Fixed-noise Gaussian information limit)**.** *Let $Z, W$ be independent $N(0, I_K)$ vectors and set $J_\rho = \arg\max_j(\rho Z_j + \sqrt{1-\rho^2}W_j)$. For every fixed $K \geq 2$ and $\rho \in [-1, 1]$,*

$$\lim_{m\to\infty} I_2(Q_{K,m}(X); Y_\rho) = \mathcal{I}_K(\rho) := I_2(J_\rho; Z). \tag{44}$$

*For $|\rho| < 1$, let $\sigma = \sqrt{1-\rho^2}$ and*

$$q_j(t) = \int_{\mathbb{R}} \varphi(z) \prod_{\ell \neq j} \Phi\left(z + \frac{\rho(t_j - t_\ell)}{\sigma}\right) dz. \tag{45}$$

*Then, with $0 \log_2 0 = 0$,*

$$\mathcal{I}_K(\rho) = \log_2 K + \mathbb{E}_Z \sum_{j=1}^{K} q_j(Z) \log_2 q_j(Z). \tag{46}$$

*At the endpoints, $\mathcal{I}_K(0) = 0$ and $\mathcal{I}_K(\pm 1) = \log_2 K$.*

*Proof.* Fix $|\rho| < 1$. Write $q_{m,j}(t) = \mathbb{P}(U_m = j \mid T_m = t)$ on the lattice of attainable score vectors. We first show that $q_{m,j}$ converges to $q_j$ uniformly on each compact set intersected with this lattice.

Consider any sequence of attainable $t_m \to t$ and select a sign vector $y_m$ with scores $t_m$. Conditional on $Y = y_m$, the centered variables $X_{j,r} - \rho y_{j,r}$ are independent, bounded in absolute value by two, and have variance $1 - \rho^2$. The multivariate Lindeberg central limit theorem therefore gives

$$S_m - \rho t_m \mid Y = y_m \quad \Rightarrow \quad N(0, (1-\rho^2) I_K).$$

The same law holds conditional on $T_m = t_m$, by the block-sum posterior invariance in Proposition 7.1. Consequently the posterior score vector converges to $\rho t + \sigma W$. Its tie hyperplanes have probability zero, implying $q_{m,j}(t_m) \to q_j(t)$. Conditioning on $W_j = z$ proves (45), and dominated convergence shows that $q_j$ is continuous in $t$.

If convergence were not uniform on a fixed compact set, there would be a subsequence and attainable $t_m$ in that set with a discrepancy bounded away from zero. A further subsequence has $t_m \to t$; the preceding argument and continuity give a contradiction.

The entropy function $h(v) = -\sum_j v_j \log_2 v_j$ is continuous, uniformly continuous, and bounded by $\log_2 K$ on the probability simplex. The unconditional vectors $T_m$ are tight and converge to $Z$. Compact-uniform convergence of the posterior probabilities implies

$$\mathbb{E}h(q_m(T_m)) - \mathbb{E}h(q(T_m)) \longrightarrow 0.$$

Since $h \circ q$ is bounded and continuous, $\mathbb{E}h(q(T_m)) \to \mathbb{E}h(q(Z))$. Also $H_2(U_m) \to \log_2 K$ by Theorem 4.2. Subtracting conditional entropy and using (43) proves (44)–(46). This establishes information convergence rather than inferring it from weak convergence alone.

At $\rho = 0$, the input and output are independent. At $\rho = \pm 1$, $Y$ determines $X$, so the finite information is $H_2(U_m) \to \log_2 K$. These observations prove the endpoint statements. □

Taking limits in (15) yields the useful bounds

$$\frac{\rho^2 \beta_K}{2 \ln 2} \le \mathcal{I}_K(\rho) \le \rho^2 \log_2 K. \tag{47}$$

Formula (46) characterizes the full limiting curve. It does not provide a closed form or a uniform-in-$K$ finite-$m$ error estimate. The posterior argument is specific to block-score quantization and does not use a continuity claim for mutual information under arbitrary weak convergence.

**Theorem 7.3** (Uniform convergence of the noise curve)**.** *For every fixed $K \ge 2$,*

$$\sup_{-1 \le \rho \le 1} |I_2(Q_{K,m}(X); Y_\rho) - \mathcal{I}_K(\rho)| \to 0.$$

*The same conclusion holds for standardized unequal blocks as* $\min_j m_j \to \infty$.

*Proof.* First, $\mathcal{I}_K$ is continuous on $[0,1]$. In the open interval this follows from the posterior formula and bounded continuity of entropy. At zero the posterior probabilities tend to $1/K$. At one, for almost every $z$ there is a unique largest coordinate, and the conditional winning probabilities tend to its point mass as the independent Gaussian perturbation vanishes. Dominated convergence of entropy gives continuity at both endpoints.

For each $m$, write $F_m(\rho) = I_2(U_m; Y_\rho)$. If $0 \le r \le s \le 1$, the channel of correlation $r$ is the channel of correlation $s$ followed by independent noise of correlation $r/s$ (with $s = 0$ handled directly). Data processing makes $F_m$ nondecreasing. Its pointwise limit is therefore nondecreasing too. Given $\epsilon > 0$, choose a finite grid $0 = t_0 < \cdots < t_N = 1$ on which the oscillation of the continuous limit in each interval is below $\epsilon$. For large $m$, pointwise convergence at all grid points gives $|F_m(t_i) - \mathcal{I}_K(t_i)| < \epsilon$. Monotonicity then bounds the error between grid points by $2\epsilon$. This proves uniform convergence on $[0,1]$. Replacing $Y_\rho$ by $-Y_\rho$ shows that both information curves are even in $\rho$.

For unequal blocks, posterior block sums have conditional mean $\rho t_j$ and variance $1 - \rho^2$ after their individual standardizations. The bounded triangular-array central limit argument in Theorem 7.2 applies whenever all block sizes diverge. It gives pointwise convergence along every such sequence. The same grid argument gives the asserted uniform convergence. □

The theorem supplies no convergence rate in $m$, and its fixed-$K$ hypothesis cannot be dropped when interchanging limits.

# 8 A consequence at large output budgets

The classical maximum expansion [9], proved with moment control in Online Resource 1, is

$$\beta_K = 2 \ln K - \ln \ln K - \ln(4\pi) + 2\gamma + o(1).$$

Consequently $\beta_{2^k}/(2k \ln 2) \to 1$. The information sandwich now yields an elementary constructive consequence of the known optimum in the Chandar–Tchamkerten note discussed in Section 1.

**Corollary 8.1.** *There exist integers $m(k)$, independent of $\rho$, such that* $\max_i \mathrm{Inf}_i(Q_{2^k,m(k)}) \to 0$ *and*

$$\sup_{0 \le \rho \le 1} \left| \frac{I_2(Q_{2^k,m(k)}(X); Y_\rho)}{k} - \rho^2 \right| \to 0.$$

*Proof.* For each $k$, choose $m(k)$ with $|B(Q_{2^k,m(k)}) - \beta_{2^k}| < 1/k$ and maximum influence below $1/k$. Theorem 4.2 ensures existence. For all $\rho$ simultaneously the sandwich gives

$$0 \le \rho^2 - \frac{I_2(Q_{2^k,m(k)}; Y_\rho)}{k} \le 1 - \frac{\beta_{2^k} - 1/k}{2k \ln 2} \to 0.$$

□

This is a diffuse realization of a prior optimum. It is neither a new value for that optimum nor an efficient input-dimension bound. The limit concerns an increasing output budget with unrestricted input size, not a compression ratio tending to one.

# 9 Interpretation and limitations

The seven-bit code constructions establish a finite failure of coordinate optimality, including under exact output balance. The WTA construction answers a different existence question with three simultaneous features: a fixed output budget, a strict information advantage near complete noise, and arbitrarily small influence of every coordinate. Its regular-simplex geometry, exchangeable block structure, and exact finite moments make WTA a tractable witness of this phenomenon. The baseline comparison shows both why winner selection is useful and why it is not a substitute for a short error-correcting-code quantizer when input dimension is the main cost.

The Gaussian eight-bit excess is approximately 0.02249. The shortened-Hamming comparator has excess $1/16$ and uses only twelve inputs. WTA's contribution is its diffuse realization with a certified coefficient–influence tradeoff. The present methods do not establish that this tradeoff is optimal. Nor do the two adjacent certificates at $m = 427, 428$ prove the first crossing among all block lengths.

**Proposition 9.1** (Finite-block endpoint obstruction)**.** *For every finite $m \geq 1$ and $K = 2^k \geq 2$, lexicographic WTA satisfies $H_2(U_m) < k$. Hence it is inferior to $k$ coordinates for all $\rho$ sufficiently close to one. When $B(U_m) > k$, the information difference has at least one zero in $(0, 1)$.*

*Proof.* By permutation symmetry, the probabilities that block one and block $K$ are unique maxima agree. Block one additionally wins every tie involving it at the maximum, while block $K$ wins only if it is unique. Such ties have positive probability (all blocks can be all positive), so $\mathbb{P}(U_m = 1) > \mathbb{P}(U_m = K)$. The output is not uniform, giving $H_2(U_m) < \log_2 K$. At $\rho = 1$ the information equals this entropy, whereas coordinate selection retains $k$ bits. Finite-alphabet information is continuous in the channel probabilities, proving the endpoint neighborhood. If $B > k$, Corollary 2.4 supplies positive differences near zero; continuity supplies an interior zero. □

Thus a positive coefficient does not promise superiority at all noise levels. The uniform limiting curve does not remove finite tie bias. Additional open questions concern the best finite dimension under an influence constraint, quantitative errors in the Gaussian approximation, and certified full-information evaluations at large $K$. The seven/eight transition proved here remains a WTA-family result; Theorem 3.1 precludes an unrestricted interpretation even for exactly balanced outputs. No universal Gaussian extremality or global low-influence threshold is required by any theorem in this paper.

## Statements and Declarations

**Funding** The author received no funding for this work.

**Competing interests** The author declares no relevant financial or non-financial competing interests.

**Data and code availability** No external empirical dataset was used. Online Resource 1 contains supplementary proofs and finite definitions. Online Resource 2 contains the verification and plotting scripts, dependency versions, certificate outputs and figure data. All reported computational claims can be regenerated from these files; there is no unavailable training data or stochastic search needed for verification.

**Author contribution** Sushanth Dasari is the sole author and is responsible for the research and final manuscript.

**Ethics and consent** Not applicable: the study involves no human participants or animals.

**Supplementary information** Online Resource 1 (`ESM_1.pdf`): supplementary proofs, posterior formula and explicit comparison constructions. Online Resource 2 (`ESM_2.zip`): reproducibility programs, certificates and figure data. Mathematical figures were generated from the stated formulas and computations using Matplotlib.

## References

[1] Courtade, T.A., Kumar, G.R.: Which Boolean functions maximize mutual information on noisy inputs? IEEE Trans. Inf. Theory 60(8), 4515–4525 (2014). https://doi.org/10.1109/TIT.2014.2326877

[2] Hammons Jr., A.R., Kumar, P.V., Calderbank, A.R., Sloane, N.J.A., Solé, P.: The $\mathbb{Z}_4$-linearity of Kerdock, Preparata, Goethals, and related codes. IEEE Trans. Inf. Theory 40(2), 301–319 (1994). https://doi.org/10.1109/18.312154

[3] Mossel, E., O'Donnell, R., Oleszkiewicz, K.: Noise stability of functions with low influences: invariance and optimality. Ann. Math. 171(1), 295–341 (2010). https://doi.org/10.4007/annals.2010.171.295

[4] Isaksson, M., Mossel, E.: Maximally stable Gaussian partitions with discrete applications. Isr. J. Math. 189, 347–396 (2012). https://doi.org/10.1007/s11856-011-0181-7

[5] Witsenhausen, H.S., Wyner, A.D.: A conditional entropy bound for a pair of discrete random variables. IEEE Trans. Inf. Theory 21(5), 493–501 (1975). https://doi.org/10.1109/TIT.1975.1055437

[6] Impagliazzo, R., Moore, C., Russell, A.: An entropic proof of Chang's inequality. SIAM J. Discrete Math. 28(1), 173–176 (2014). https://doi.org/10.1137/120877982

[7] Pollard, D.: Quantization and the method of $k$-means. IEEE Trans. Inf. Theory 28(2), 199–204 (1982). https://doi.org/10.1109/TIT.1982.1056481

[8] Johansson, F.: Arb: Efficient arbitrary-precision midpoint-radius interval arithmetic. IEEE Trans. Comput. 66(8), 1281–1292 (2017). https://doi.org/10.1109/TC.2017.2690633

[9] Leadbetter, M.R., Lindgren, G., Rootzén, H.: Extremes and Related Properties of Random Sequences and Processes. Springer, New York (1983). https://doi.org/10.1007/978-1-4612-5449-2