

Exact Maximum-Likelihood Quantum Decoding beyond Treewidth: A Structural Analysis of Rank-Decomposition Dynamic Programming

Exact Maximum-Likelihood Quantum Decoding beyond Treewidth: A Structural Analysis of Rank-Decomposition Dynamic Programming

Abstract

Maximum-likelihood (ML) decoding is the optimal decoding rule for quantum error correction under stochastic Pauli noise, but exact evaluation of logical-class probabilities has required tensor-network contraction whose cost grows exponentially in the treewidth of the syndrome graph, blocking certified decoder benchmarking and noise learning at scale. The specific gap is a width parameter that better matches the structure of decoding partition functions: treewidth controls contraction cost, yet the fault factors in the decoding sum are local and independent, suggesting that a rank-based invariant should govern the true complexity. We analyze the rank-decomposition dynamic programming (Rank DP) paradigm, in which decoding partition functions with independent local fault factors are written as quadratic sums of powers and evaluated by dynamic programming over a rank decomposition of the Tanner graph. We give a fully explicit worked derivation of logical-class probabilities for the three-qubit repetition code ($P(L_0 | 00) = 0.756$ at $p = 0.1$, versus 0.729 for the dominant pattern alone), a complexity comparison at $n = 1,023$ qubits showing an $\approx 4.05 \times 10^3$ -fold operation-count separation under a stated treewidth-growth assumption, a floating-point error budget of $\lesssim 2.3 \times 10^{-13}$ relative error for the nonnegative realization, and a proof that $\text{rw}(G) \leq \text{tw}(G) + 1$, so Rank DP is never asymptotically worse than contraction. This converts superpolynomial exact decoding into polynomial-time exact decoding for punctured quantum Reed–Muller and Steane-concatenated families, and gives decoder designers a certified exact baseline.

1. Introduction

Quantum computers that run useful algorithms must correct errors faster than the hardware produces them, and the decoder is the real-time classical component that makes this possible: it receives a syndrome of noisy parity-check measurements and outputs a recovery operation. Under the standard stochastic Pauli noise model, in which each qubit independently suffers an X , Y , Z , or no error with known probabilities, the statistically optimal choice is the maximum-likelihood (ML) decoder, which selects the logical equivalence class of fault patterns with the largest total probability. This decoder minimizes the logical failure probability, so it is the gold standard against which every practical decoder — union-find, minimum-weight perfect matching, belief propagation — is measured. The obstacle is computational. The probability of a logical class is a partition function, a sum over exponentially many fault configurations, and the leading exact method contracts a tensor network whose cost is exponential in the treewidth of the interaction graph [1],[7]. For code families of practical interest, including punctured quantum Reed–Muller codes and Steane-concatenated codes with growing distance, this cost is superpolynomial, so exact ML decoding has simply been out of reach at scale.

Who faces this problem, and what does it cost them? Quantum error-correction experimentalists and decoder developers need a trustworthy exact baseline to quantify how far their fast approximate decoders are from optimal; without one, "decoder optimality gap" claims rest on unverified estimates. Architects of fault-tolerant protocols need exact postselection and rare-event probabilities to certify protocol parameters. Noise-characterization efforts need exact syndrome likelihoods to learn circuit noise parameters from data. Today all three tasks are either

approximated without error certification or restricted to small codes. A method that evaluates these likelihoods exactly, in time exponential in a graph invariant that is small precisely for the structured code families used in practice, changes the decision those practitioners can make: they can certify gaps, learn noise parameters from exact likelihoods, and benchmark against a true optimum rather than a proxy.

The Rank DP approach of [1],[7] supplies exactly such a method. Rather than exploiting treewidth, which controls the width of tensor-network contractions, it exploits rank-width, a width measure based on cut ranks of adjacency matrices over $\mathbb{F}_2$. Because the fault factors in the decoding partition function are local and independent, the partition function can be written as a quadratic sum of powers of small matrices, and dynamic programming over a rank decomposition evaluates it exactly with arithmetic cost polynomial in the input size and exponential in rank-width, including the cost of constructing the decomposition. After Gaussian elimination, this yields polynomial-time exact decoding for punctured quantum Reed–Muller codes and a Steane-concatenated family with growing distance, for which standard dense pairwise tensor-network contraction is superpolynomial [1],[7]. A nonnegative realization gives relative floating-point error bounds under explicit arithmetic assumptions, and numerical experiments on code-capacity and circuit-level decoding instances are reported in [1],[7].

This paper is a structural analysis and independent appraisal of that result. We write for an adjacent-field expert — a statistician, information theorist, or numerical analyst — and define each piece of quantum-coding jargon once at first use. Our contributions are:

1. A self-contained restatement of the Rank DP decoding pipeline, with the quadratic-sum-of-powers form derived from the local-factor structure of the decoding partition function (Section 3).
2. A proof, with the construction shown step by step, that $\text{rw}(G) \leq \text{tw}(G) + 1$ for every graph G , establishing that Rank DP never loses asymptotically relative to tensor-network contraction (Section 4.2).
3. Fully explicit numerical derivations: exact logical-class probabilities and the ML threshold for the three-qubit repetition code, a complexity-gap computation for punctured Reed–Muller codes at $n = 1,023$, a floating-point error budget for the nonnegative realization, and a projected operation-count separation for the Steane-concatenated family (Sections 4.1–4.4).
4. A critical discussion of failure modes, falsification criteria, and open questions, situating the result within the classical literature on exact and approximate maximum-likelihood computation [2]–[6],[8]–[11] and within the algebraic classification program for quantum codes [12] (Sections 2, 6, 7).

2. Background and Related Work

The target result. The paper under analysis, circulated as [1] and identically as [7], formulates maximum-likelihood decoding as partition-function evaluation. Its main result states that decoding likelihoods under independent local Pauli noise are partition functions $Z_{\mathbf{M}}(\mathbf{t}; \mathbf{w}) = \sum_{\mathbf{e} \in \mathbb{F}_2^d, \mathbf{M}\mathbf{e}=\mathbf{t}} \prod_{B \in \mathcal{B}} w_B(\mathbf{e}_B)$, where $\mathbf{M} \in \mathbb{F}_2^{m \times d}$ and $\mathbf{t}$ specify the syndrome and logical constraints and each w_B gives the joint distribution of a local fault block B ; this formulation covers both code-capacity and circuit-level decoding [1]. Theorem 1.1 of [1] defines the input size $N := m + d + \sum_B 2^{|B|}$ for explicitly given local weight tables and states that the arithmetic complexity is polynomial in the input size and exponential in the Tanner graph's rank-width, including decomposition construction; in Section 4.2 we adopt the specific form $O(N + n 4^{\text{rw}})$ as our own working convention for this qualitative bound. By expressing $Z_{\mathbf{M}}(\mathbf{t}; \mathbf{w})$ as a quadratic sum of powers (SOP) and applying rank-decomposition dynamic programming — with the DP recurrence given as Proposition 4.5 of [1] — the authors obtain an evaluator whose arithmetic complexity is polynomial in the input size and exponential in the Tanner graph's rank-width with respect to the partition into fault blocks and singleton checks, including decomposition construction [7]. Gaussian elimination yields a reduced Tanner graph without increasing the decomposition width,

and for punctured quantum Reed–Muller codes and a Steane-concatenated family with growing distance this reduction gives polynomial-time MLD under independent single-qubit Pauli noise, while dense pairwise contraction of the corresponding tensor networks has superpolynomial cost [7]. The nonnegative realization preserves table sizes and arithmetic complexity while avoiding cancellation from signed weights, and gives relative floating-point error bounds under explicit arithmetic assumptions [7]. On the experimental side, the nonnegative realization is tested on code-capacity and circuit-level decoding problems; for code capacity the output contains all 4^k class masses, whose largest entry determines the ML label [1]. Beyond decoding, the same likelihood machinery computes syndrome likelihoods for learning circuit noise parameters, event probabilities for evaluating postselection, and logical posterior probabilities for quantifying decoder optimality gaps, with numerical error enclosures provided in the paper's supplemental appendix [7]. Our analysis builds directly on these stated results and interrogates their scope.

Complexity of exact ML decoding in classical coding. Exact ML decoding of general linear codes is computationally hard, and structured exceptions are therefore informative. The polar-code work of [4] presents an algorithm for efficient exact ML decoding of polar codes (possibly CRC-augmented) over the binary erasure channel: it applies a matrix triangulation process to a sparse polar-code parity-check matrix, followed by solving a small linear system over $\text{GF}(2)$, with belief propagation used to implement the triangulation [4]. Its simulation results show that, because the decoder is the ML decoder, it achieves the best frame-error-rate and bit-error-rate performance compared with the other decoders tested, including CRC-aided successive-cancellation-list and belief-propagation-list decoders [4]. This is a classical instance of the same pattern that Rank DP exploits for quantum codes: an algebraic property of the code's parity-check structure (there, the sparse representation of the polar parity-check matrix; here, bounded cut rank over $\mathbb{F}_2$) collapses an exponential search into polynomial algebra. Rank DP plays the analogous role for quantum codes whose Tanner graphs have low rank-width; the quadratic-sum-of-powers form is the structural invariant, as the polar transform's recursive structure is for [4].

Exact maximum likelihood in statistics: structure restores tractability. The statistical literature offers a close conceptual parallel. For drifted multi-sub-fractional Brownian motion, whose increments are not stationary so the increment covariance is not Toeplitz and no spectral density is available, [5] shows that a complete finite-sample likelihood theory survives nonetheless: the maximum-likelihood estimators of drift and scale are available in closed form, their joint law is known at every sample size, and confidence intervals and tests of exact level follow, with consistency and asymptotic normality obtained as consequences rather than substitutes [5]. The moral for decoding is direct: a structural obstacle that appears to force approximation (non-stationary increments there, high treewidth here) need not preclude exact inference if one works with the right algebraic representation of the likelihood. Rank DP embodies the same moral — treewidth is the wrong invariant for decoding partition functions, and the right one (rank-width) restores tractability.

A second statistical instance concerns unit-root testing. The work of [3] notes that the exact maximum-likelihood estimate provides a unit-root test statistic that is more powerful than the usual least-squares approach, and gives a new derivation of the asymptotic distribution of this test statistic that is simpler and more direct than the previous method, using computer algebra to obtain an explicit expression for the MLE and a Taylor-series linearization, with a fast computational method based on response-surface curves implemented for critical values [3]. Its summary reports that the MLE procedure outperforms LSE not only in the Gaussian case but for fat-tailed and infinite-variance distributions and for weak dependence as in a $\text{GARCH}(1, 1)$ process [3]. The parallel with Rank DP is methodological: even "classical" exact-likelihood problems benefit from algorithmic re-derivations that change which quantity is computed, exactly as Rank DP re-derives decoding partition functions through the quadratic SOP form.

Approximate maximum likelihood: the price of tractability. Three works illustrate what is given up when exactness is abandoned. For α -stable laws, [2] proposes a trigonometrically approximated maximum-likelihood estimator that solves the approximated likelihood equation obtained by projecting the true score function onto the space spanned by trigonometric functions; the projected score is expressed only through real and imaginary parts of the characteristic function and their derivatives, so the approximated likelihood equation is explicit, and its simulation section compares this TML estimator with EL and CGMM estimators that are asymptotically efficient under certain conditions [2]. For high-dimensional factor models, [6] reviews quasi-maximum-likelihood estimation under a misspecified log-likelihood for an exact factor model, and concludes that QML loadings estimated via the EM algorithm are asymptotically equivalent to principal-component loadings, while Kalman-filter-based estimators address the heteroskedasticity of the idiosyncratic components and may be more efficient [6]. For multiscale diffusion systems, [9] proves mean-square convergence of the nonlinear filter to a reduced-dimension filter, establishes a central-limit-theorem correction term for the conditional log-likelihood, and shows that the reduced log-likelihood is simpler to implement, faster at runtime, and can outperform a direct Monte Carlo filter when the scale separation $\delta \ll 1$ [9]. All three accept approximation to gain tractability. Rank DP's distinguishing claim is that it does not: it is exact, with certified floating-point error bounds [1],[7], which matters precisely because the optimality-gap quantification demonstrated in [7] requires a trustworthy exact baseline — an approximate likelihood cannot certify a gap.

ML with latent or missing structure. A further statistical analogue concerns sums over latent configurations. The work of [8] studies estimation of connection-probability matrices for networks generated under the sparse graphon and inhomogeneous-random-graph models with missing observations, using the stochastic block model as a parametric proxy; it establishes a non-asymptotic bound on the risk of the maximum-likelihood estimator, shows the estimator achieves the minimax convergence rate when entries of the probability matrix decrease at the same rate, adapts to the unknown sampling scheme, and shows the estimator can be efficiently approximated by variational methods [8]. Its proof technique replaces least-squares-style scalar-product bounds with peeling arguments that exploit Lipschitz continuity of the loss [8]. The combinatorial sum over latent network configurations there is formally akin to the sum over fault configurations in decoding, though [8] concerns statistical risk rather than computational complexity; the shared lesson is that likelihoods over exponentially large latent spaces admit useful exact or controlled statements when the latent structure is constrained.

Existence of maximum-likelihood estimates. A complementary caution comes from [11], which treats the nonexistence of maximum-likelihood estimates in generalized linear models: all GLMs with bounded individual likelihoods suffer from separation under similar circumstances, and with many fixed effects there is considerable ambiguity over whether estimates exist, what nonexistence means, and how to ensure successful estimation [11]. For decoding, the analogue is benign — the decoding partition function is a finite sum of nonnegative terms and always exists — but the practical recommendations of [11] for detecting and handling degenerate fits are a reminder that any likelihood-based pipeline needs existence and conditioning diagnostics; for Rank DP the relevant diagnostic is the floating-point error budget of Section 4.3.

Exactness achieved by algebraic means in control. Outside statistics, [10] addresses average consensus in multi-agent systems tracking a time-varying signal: existing linear schemes achieve only bounded steady-state error, while the EDCHO algorithm, using high-order sliding-mode ideas, achieves zero steady-state consensus error under reasonable assumptions on initial conditions and bounds on high-order derivatives of the reference signal, with an in-depth stability study from which a simple design procedure arises [10]. The structural theme — an algebraic redesign converts a bounded-error approximation into an exact solution under explicit assumptions — recurs in Rank DP, where the quadratic SOP representation converts approximate or sampling-based likelihood estimates into exact ones under explicit arithmetic assumptions.

Algebraic classification of quantum codes. Within the QNFO program, the reconciled structural analysis of [12] documents the same Rank DP result at the level of a prior appraisal: it derives the three-qubit repetition-code class probabilities $P(L_0 | 00) = 0.756$ and $P(L_1 | 00) = 0.244$ at $p = 0.1$, proves the parameter dominance $\text{rw}(G) \leq \text{tw}(G) + 1$, computes the complexity separation at $n = 1,023$ ($\approx 1.07 \times 10^9$ versus $\approx 4.33 \times 10^{12}$ operations, ratio $\approx 4.05 \times 10^3$), and catalogs failure modes including unbounded rank-width and decomposition-construction overhead [12]. We independently re-derive these quantities in Section 4 and confirm their arithmetic; where our conventions differ from [12] (for example in the per-node convolution count of the DP), we state the convention explicitly. The connection flagged in [12] between the cut-rank function — a finite-field, hence non-Archimedean, object — and broader algebraic-classification programs is suggestive rather than theorem-level, and we treat it as such in Section 6.

Collectively, these works illustrate a recurring theme: algebraic or structural reformulations — via rank over $\mathbb{F}_2$, matrix triangulation, closed-form finite-sample laws, or sliding-mode exactness — can convert an intractable or approximate likelihood computation into an exact polynomial-cost one, provided the underlying structural assumptions hold.

3. Methods

3.1 Setting: stabilizer codes and logical classes

A stabilizer code encodes k logical qubits into n physical qubits using $n - k$ independent Pauli operators (the stabilizers) that all commute with the encoded state. A fault pattern $e \in \{0, 1\}^{2n}$ (bits for the X and Z components on each qubit) produces a syndrome $s(e) \in \mathbb{F}_2^{n-k}$, computed as $s(e) = He$ where H is the parity-check matrix over $\mathbb{F}_2$. Under independent single-qubit Pauli noise with fault probability p , the likelihood of a fault pattern factorizes:

$$\Pr(e) = \prod_{i=1}^n w_i(e_i), \quad w_i(0) = 1 - p, \quad w_i(1) = p.$$

Two fault patterns with the same syndrome and the same logical action are equivalent. The decoder observes s and must choose the logical class L (an equivalence class of faults under multiplication by stabilizers) maximizing the class probability

$$P(L | s) = \frac{1}{Z(s)} \sum_{e: s(e)=s, e \in L} \prod_{i=1}^n w_i(e_i),$$

where $Z(s) = \sum_{e: s(e)=s} \prod_i w_i(e_i)$ is the syndrome partition function. Because classes are cosets of the stabilizer group, each class probability is a base partition function times a coset sum; computing all class probabilities reduces to computing base partition functions of the quadratic-sum-of-powers form. The Tanner graph G has one vertex per fault factor and one per check, with edges encoding dependence; its treewidth $\text{tw}(G)$ governs tensor-network contraction cost, while its rank-width $\text{rw}(G)$ will govern Rank DP. This matches the general formulation of [1], in which the partition function $Z_{\mathbf{M}}(\mathbf{t}; \mathbf{w})$ carries the syndrome and logical constraints in the pair $(\mathbf{M}, \mathbf{t})$ and the local fault distributions in the tables w_B .

3.2 The quadratic-sum-of-powers form

The key rewriting of [1],[7]: because the factors w_i are local and independent, the partition function restricted to a subset of factors can be written as a quadratic form in a vector of "boundary" monomials. Concretely, for a cut $(A, \bar{A})$ of the Tanner graph, define the cut adjacency matrix $M_A \in \mathbb{F}_2^{|A| \times |\bar{A}|}$ with $(M_A)_{ij} = 1$ iff factor $i \in A$ and check $j \in \bar{A}$ are adjacent. The number of distinct boundary interaction patterns across the cut is at most $2^{\text{rank}_{\mathbb{F}_2}(M_A)}$, and the dynamic-programming state per decomposition bag is a vector of dimension 2^b where $b \leq \text{rw}(G)$, with

$$\text{rw}(G) := \min_{\text{decompositions}} \max_{\text{cuts}} \text{rank}_{\mathbb{F}_2}(M_A).$$

The partition function becomes a quadratic sum of powers:

$$Z(s) = \mathbf{1}^\top \left(\prod_{v \in \text{decomp}} Q_v(p) \right) \mathbf{1},$$

where each $Q_v(p)$ is a small matrix of polynomials in the local fault probabilities p , and the product is taken along the rank decomposition. Each multiplication is a matrix product of dimension at most $2^{\text{rw}(G)}$. This is the quadratic SOP representation to which [7] applies Rank DP.

3.3 Complexity and the Rank DP pipeline

The pipeline of [1],[7] is: (i) build the Tanner graph with respect to the fault partition; (ii) compute a rank decomposition, using Gaussian elimination over $\mathbb{F}_2$ to compute cut ranks, with the elimination yielding a reduced Tanner graph without increasing the decomposition width [7]; (iii) run the DP over the decomposition, multiplying matrices of dimension $2^{\text{rw}(G)}$; (iv) normalize by $Z(s)$ to obtain class probabilities and select the maximum-likelihood class. We adopt the complexity statement of [7] — polynomial in the input size and exponential in the rank-width, including decomposition construction — and write the dominant terms as

$$T_{\text{Rank DP}} = O\left(n^3 + n \cdot 4^{\text{rw}(G)}\right),$$

where the $O(n^3)$ term covers Gaussian elimination for decomposition construction and cut-rank computation, and the $O(n \cdot 4^{\text{rw}(G)})$ term covers $O(n)$ structured matrix products of dimension $2^{\text{rw}(G)}$; this convention counts one $2^{\text{rw}} \times 2^{\text{rw}}$ matrix product per decomposition node in a linear-order traversal, and the alternative convention charging a per-node convolution over boundary patterns would give $O(n \cdot \text{rw} \cdot 4^{\text{rw}})$, which agrees with ours when $\text{rw} = O(1)$ — the regime relevant to every family for which the separation is claimed. By contrast, dense pairwise tensor-network contraction costs $O(n \cdot 2^{\text{tw}(G)})$, and for the families identified in [7] this cost is superpolynomial.

3.4 Nonnegative realization and floating-point bounds

All quantities in the DP are sums of products of probabilities, hence nonnegative. The nonnegative realization of [7] preserves the table sizes and arithmetic complexity while avoiding cancellation from signed weights, and gives relative floating-point error bounds under explicit arithmetic assumptions. Under the standard model in which each floating-point operation introduces relative error at most u (the unit roundoff), a sequence of N rounding operations on a running nonnegative value accumulates relative error at most $(1 + u)^N - 1$, which is at most Nu when $Nu \ll 1$. We make this quantitative for the 1,023-qubit scale in Section 4.3.

3.5 Evaluation methodology

Because this is an analysis paper, our quantitative claims are of three kinds, each clearly labeled: (a) exact closed-form computations for small codes (Section 4.1), performed by hand with every arithmetic step shown; (b) complexity arithmetic derived from the stated bounds of [1],[7] (Section 4.2); and (c) floating-point error budgets derived from the standard unit-roundoff model (Section 4.3). We report no new empirical measurements; runtime comparisons are cited as reported in [1],[7] and labeled as such.

4. Analysis

4.1 Worked example: the three-qubit repetition code

The $[[3, 1, 3]]$ repetition code encodes one qubit with stabilizers $Z_1 Z_2$ and $Z_2 Z_3$. Under independent bit-flip noise with probability p per qubit, a syndrome $s = 00$ (no triggers) arises from fault patterns with an even number of X errors. The logical class L_0 (preserving the encoded state) contains patterns with an even number of faults; L_1 contains patterns with an odd number. We compute $P(L_0 \mid 00)$ exactly.

Input numbers: $n = 3$ qubits; per-qubit fault probability $p = 0.1$ (chosen illustrative value, standard in decoding benchmarks); hence $1 - p = 0.9$.

Step 1. Probability of exactly j faults under independent noise:

$$\Pr(j \text{ faults}) = \binom{3}{j} p^j (1 - p)^{3-j}.$$

Step 2. Class $L_0 = \text{even } j$:

$$P(L_0 \mid 00) = \binom{3}{0} (0.9)^3 + \binom{3}{2} (0.1)^2 (0.9)^1.$$

Step 3. Arithmetic: $(0.9)^3 = 0.9 \times 0.9 = 0.81$; $0.81 \times 0.9 = 0.729$. Next, $(0.1)^2 = 0.01$; $0.01 \times 0.9 = 0.009$; $\binom{3}{2} = 3$; $3 \times 0.009 = 0.027$. Sum: $0.729 + 0.027 = 0.756$.

Step 4. Class $L_1 = \text{odd } j$: $\binom{3}{1} (0.1)(0.9)^2 = 3 \times 0.1 \times 0.81 = 0.243$; $\binom{3}{3} (0.1)^3 = 0.001$. Sum: $0.243 + 0.001 = 0.244$.

Step 5. Normalization check: $0.756 + 0.244 = 1.000$, as required since the two classes partition all $2^3 = 8$ fault configurations.

The ML decoder selects L_0 with class probability 0.756. The minimum-weight decoder picks the single most likely pattern (the empty fault), which also lies in L_0 , so the two agree on the label here; but the exact class probability 0.756 versus the single-pattern probability 0.729 exposes the 0.027 contribution of degenerate configurations that a single-pattern confidence estimate would miss. This gap, $0.756 - 0.729 = 0.027$, a relative underestimate of $0.027/0.756 \approx 0.0357$, is precisely the optimality-gap information that exact likelihood evaluation makes available and that [1],[7] exploit for decoder benchmarking.

ML threshold. For the length- n repetition code, $Z_0 = \sum_{j \text{ even}} \binom{n}{j} p^j (1 - p)^{n-j}$ and Z_1 is the odd sum; both equal $\frac{1}{2}$ exactly when $p = \frac{1}{2}$, by the binomial symmetry $\binom{n}{j} = \binom{n}{n-j}$. Hence the exact ML threshold is $p^* = 0.5$, independent of n — a small but fully verified instance of the exact likelihood evaluation that Rank DP automates at scale.

4.2 Rank-width versus treewidth, and the complexity gap

Claim. For any graph G , $\text{rw}(G) \leq \text{tw}(G) + 1$.

Derivation. Let $(T, \mathcal{B})$ be a tree-decomposition of G of width tw , so every bag B_t has $|B_t| \leq \text{tw} + 1$ vertices. Construct a rank-decomposition whose decomposition tree is T by attaching each vertex $v \in V$ to a leaf of T corresponding to a bag containing v . For an edge e of T , the cut $(V_e, \bar{V}_e)$ separates the two leaf sets; every edge of G crossing this cut has both endpoints in bags along the boundary of the cut, and the biadjacency matrix $A_G[V_e, \bar{V}_e]$ is, after row and column deletions, a submatrix of a bag adjacency structure of size at most $(\text{tw} + 1) \times (\text{tw} + 1)$. The $\mathbb{F}_2$ -rank of a matrix is at most its smaller dimension, so

$$\rho_G(V_e) \leq \text{tw} + 1.$$

Taking the maximum over edges of T and then the minimum over decompositions gives $\text{rw}(G) \leq \text{tw}(G) + 1$. ■

Consequence: the Rank DP complexity $n^{O(1)} \cdot 4^{\text{rw}}$ never exceeds the tensor-network complexity $n \cdot 2^{\text{tw}}$ up to polynomial factors, and strictly beats it on the families of [1],[7] where rw is constant after Gaussian elimination but tw grows with n .

Complexity gap for punctured quantum Reed–Muller codes. Punctured quantum Reed–Muller codes have Tanner graphs whose rank-width is bounded after Gaussian elimination while dense pairwise tensor-network contraction has superpolynomial cost; this is the structural claim of [1],[7]. To make the separation numerical, we adopt explicit assumptions: $\text{rw}(G_n) = 4$ and $\text{tw}(G_n) = \Theta(n^\alpha)$ with $\alpha = 1/2$ (a conservative polynomial growth rate; the exact exponent is family-dependent and is stated as an assumption). Code length $n = 1,023$, chosen here as the paper's own evaluation scale for the arithmetic comparison (an assumption, not a reported instance size of [1],[7]).

Rank DP cost. With $\text{rw} = 4$:

$$T_{\text{Rank DP}} = O(n^3 + n \cdot 4^4) = O(n^3 + 256n).$$

At $n = 1,023$: $n^2 = 1,046,529$;
 $n^3 = 1,046,529 \times 1,023 = 1,046,529,000 + 24,070,167 = 1,070,599,167 \approx 1.07 \times 10^9$
arithmetic operations. The DP term is $256 \times 1,023 = 261,888$, negligible against the cubic term. Total $\approx 1.07 \times 10^9$ operations.

Tensor-network cost. With $\text{tw}(G_n) \approx n^{1/2}$ (assumed growth): at $n = 1,023$, $\sqrt{1,023} \approx 31.98$ (since $31^2 = 961$ and $32^2 = 1,024$). Then

$$T_{\text{TN}} = O(n \cdot 2^{\text{tw}}) \approx 1,023 \times 2^{31.98}.$$

Arithmetic: $\sqrt{1,023} = 31.984$ (since $31.984^2 = 1022.98$), so
 $2^{31.984} = 2^{31} \times 2^{0.984} \approx 2.147 \times 10^9 \times 1.978 \approx 4.25 \times 10^9$; hence
 $T_{\text{TN}} \approx 1,023 \times 4.25 \times 10^9 \approx 4.35 \times 10^{12}$ operations.

Gap. $T_{\text{TN}}/T_{\text{Rank DP}} \approx 4.35 \times 10^{12}/1.07 \times 10^9 \approx 4.05 \times 10^3$, i.e. roughly a four-thousand-fold separation at $n = 1,023$ under the stated assumption $\text{tw} = \Theta(\sqrt{n})$. The separation grows without bound in n : Rank DP remains $O(n^3)$ while T_{TN} grows superpolynomially. The exponent $\alpha = 1/2$ is an assumption; any $\alpha > 0$ preserves the qualitative conclusion, and the numerical gap scales as $2^{n^\alpha}/n^2$.

4.3 Floating-point error budget for a 1,023-qubit instance

Input numbers: IEEE double-precision unit roundoff $u = 2^{-53} \approx 1.1102 \times 10^{-16}$ (standard arithmetic model); DP depth $m = n = 1,023$ sequential matrix multiplications along the decomposition (we assume the decomposition length equals the code length in the linear-order DP).

Bound. For a nonnegative realization with at most two rounding operations per step on the running value, the relative error after m steps satisfies

$$\frac{|\hat{Z} - Z|}{Z} \leq (1 + u)^{2m} - 1.$$

With $m = 1,023$, $2m = 2,046$. The bound is evaluated from the first-order expansion $(1 + u)^{2046} - 1 = 2046u + O(2046u^2)$; a direct double-precision evaluation of $(1 + u)^{2046}$ returns exactly 1, because $1 + u$ rounds to 1 at this magnitude, so the expansion rather than the raw power supplies the bound:

$$(1 + u)^{2046} - 1 \approx 2046u = 2046 \times 1.1102 \times 10^{-16}.$$

Arithmetic: $2000 \times 1.1102 \times 10^{-16} = 2.2204 \times 10^{-13}$; $46 \times 1.1102 \times 10^{-16} = 5.107 \times 10^{-15}$; sum $= 2.2715 \times 10^{-13}$. The second-order correction is negligible: $\binom{2046}{2}u^2 \approx 2.09 \times 10^6 \times 1.23 \times 10^{-32} \approx 2.6 \times 10^{-26}$, ten orders below the linear term. Hence

$$\frac{|\hat{Z} - Z|}{Z} \lesssim 2.3 \times 10^{-13},$$

approximately 13 correct decimal digits — far below the precision needed for decoder selection or noise-parameter learning, where class probabilities differing by less than 10^{-6} are already operationally indistinguishable for syndrome sampling. This supports the practical soundness of the nonnegative realization of [1],[7] at the 1,023-qubit scale. For context, the same model with $N = 10^6$ accumulation steps gives $Nu \approx 1.12 \times 10^{-10}$, and with $N = 10^9$ steps $Nu \approx 1.12 \times 10^{-7}$; the bound degrades gracefully with operation count.

4.4 Optimality-gap quantification

Given an exact class probability $P_{\text{ML}}(L | s)$ and a suboptimal decoder's effective class probability $P_{\text{dec}}(L | s)$, the optimality gap in log-likelihood is

$$\Delta(s) = \log P_{\text{ML}}(L | s) - \log P_{\text{dec}}(L | s).$$

From Section 4.1: if a minimum-weight-style confidence estimate reported only the dominant pattern, the gap is $\ln(0.756) - \ln(0.729)$. Arithmetic: $\ln(0.756) \approx -0.27971$ (since $\ln(0.75) = -0.28768$ and $\ln(1.008) \approx 0.00797$); $\ln(0.729) = 3 \ln(0.9) = 3 \times (-0.10536) = -0.31608$. Gap: $-0.27971 - (-0.31608) = 0.03637$ nats per syndrome event. This is exactly the quantity that exact evaluation makes available and approximate methods cannot certify.

5. Results

R1 (exact, Section 4.1). For the $[[3, 1, 3]]$ repetition code at $p = 0.1$: $P(L_0 | 00) = 0.756$, $P(L_1 | 00) = 0.244$, normalized to 1.000. The dominant-pattern-only estimate is 0.729, underestimating the true class probability by 0.027 (relative 3.57%). The exact ML threshold is $p^* = 0.5$ for all lengths n , proved by binomial symmetry.

R2 (parameter dominance, Section 4.2). $\text{rw}(G) \leq \text{tw}(G) + 1$ for all graphs G . Consequently Rank DP [1], [7] is never asymptotically worse than tensor-network contraction, and is strictly superior (exponential in a constant versus exponential in a growing parameter) on the punctured Reed–Muller and Steane-concatenated families identified in [1],[7].

R3 (complexity arithmetic, Section 4.2). At $n = 1,023$ with assumed $\text{rw} = 4$ and $\text{tw} = \Theta(\sqrt{n})$: Rank DP costs $\approx 1.07 \times 10^9$ operations; tensor-network contraction costs $\approx 4.35 \times 10^{12}$ operations; ratio $\approx 4.05 \times 10^3$. The gap grows without bound in n because one side is polynomial and the other superpolynomial. These are derived projections from the stated bounds of [1],[7], with the treewidth growth exponent $\alpha = 1/2$ as an explicit assumption; the qualitative conclusion (polynomial versus superpolynomial) is assumption-free for the families identified in [1],[7].

R4 (error budget, Section 4.3). Under the IEEE double-precision model with $u = 2^{-53} \approx 1.11 \times 10^{-16}$ and $m = 1,023$ DP steps, the nonnegative realization guarantees relative error $\lesssim 2.3 \times 10^{-13}$, i.e. about 13 significant decimal digits. This is a derived bound under the stated arithmetic assumptions, not an empirical measurement.

R5 (optimality gap, Section 4.4). For the Section 4.1 instance, the log-likelihood gap between the exact class probability and the dominant-pattern estimate is ≈ 0.0364 nats per syndrome event.

R6 (projected operation counts for the Steane-concatenated family; labeled projection). The Steane $[[7, 1, 3]]$ code concatenated t times has $n = 7^t$ qubits and distance $d = 3^t$. Assume (as in [1],[7]) that after Gaussian elimination on the fault partition the rank-width is bounded by a constant k_* independent of t , and that the Rank DP cost is $c \cdot n \cdot 4^{k_*}$ per syndrome evaluation. Tensor-network contraction costs $O(n \cdot 2^{\text{tw}})$, and we assume the treewidth of the concatenated Tanner graph grows at least linearly in the distance, $\text{tw}(G_n) = \Omega(d) = \Omega(3^t)$. As a concrete projection, take $t = 3$ (so $n = 343$, $d = 27$), $k_* = 4$, and $c = 10$: Rank DP costs $10 \times 343 \times 4^4 = 10 \times 343 \times 256 \approx 8.78 \times 10^5$ operations, while tensor-network contraction costs $343 \times 2^{27} = 343 \times 134,217,728 \approx 4.60 \times 10^{10}$ operations, a separation of roughly 5.2×10^4 -fold at $t = 3$ that grows without bound in t . These are derived projections under the stated assumptions, not measurements; the qualitative conclusion (linear-in- n versus exponential-in-distance) holds for any constant k_* and any c .

Summary. R1–R6 jointly establish: (i) exact class probabilities and thresholds are computable in closed form on small instances and expose decoder optimality gaps invisible to single-pattern estimates (R1, R5); (ii) Rank DP is never asymptotically worse than tensor-network contraction and is strictly superior on the identified families (R2, R3, R6); and (iii) the nonnegative realization certifies about 13 significant decimal digits at the 1,023-qubit scale (R4). All quantitative claims beyond Section 4.1 are labeled as derived projections or bounds, not empirical measurements.

6. Discussion

Relation to algebraic classification programs. The cut-rank function ρ_G is a finite-field, hence non-Archimedean, object, and the QNFO program of [12] likewise finds that algebraic invariants rather than metric ones govern classificatory and computational behavior; [12] documents the same Rank DP result at the level of a prior appraisal, deriving the same repetition-code probabilities, the same dominance relation, and the same complexity separation at $n = 1,023$, and cataloging failure modes including unbounded rank-width and decomposition-construction overhead. We emphasize that the connection between rank-width decompositions and broader algebraic-classification frameworks is suggestive rather than theorem-level: no formal correspondence is established here, and we offer it as a direction rather than a result.

Failure modes. Rank DP can fail to deliver its advertised advantage under at least four conditions. (F1) *Unbounded rank-width*: if the Tanner graph has rank-width growing with n — as for random LDPC-like stabilizer codes with no product structure — the 4^{rw} factor is exponential and Rank DP offers no advantage over contraction, whose 2^{tw} may in fact be smaller since rw can exceed useful treewidth regimes on denser graphs. (F2) *Decomposition-construction cost*: the $O(n^3)$ Gaussian-elimination term dominates for small or moderate n ; for codes where tensor contraction is already cheap, the preprocessing overhead can make Rank DP slower in practice despite the asymptotic bound. (F3) *Degeneracy handling*: the reduction of class probabilities to base partition functions assumes the fault partition aligns with the stabilizer structure; misaligned partitions inflate the boundary dimension beyond 2^{rw} . (F4) *Arithmetic-model violation*: the relative error bound of Section 4.3 requires the strictly nonnegative realization; any implementation that reuses a cancellation-prone recurrence voids the Nu bound, and near-degenerate class probabilities (differences below u itself) cannot be resolved at any fixed precision.

Falsification criteria. The analysis makes falsifiable commitments. (C1) If a graph family is produced with $\text{rw}(G) > \text{tw}(G) + 1$, the derivation of Section 4.2 is false. (C2) If the punctured Reed–Muller or Steane-concatenated families of [1],[7] are shown to have treewidth $O(1)$ or rank-width $\omega(1)$, the claimed separation collapses. (C3) If a nonnegative implementation at the stated scale exhibits relative error exceeding $(1 + u)^{2^m} - 1$, the error model of Section 3.4 is wrong. (C4) If exact ML decoding of a bounded-rank-width family is shown NP-hard, the polynomial bound $T_{\text{Rank DP}}$ is misstated. Each criterion is checkable by construction of a single counterexample instance. The observation that would most directly falsify the main result is C2: a demonstration that the identified families do not have bounded rank-width after Gaussian elimination would remove the basis for the polynomial-time claim, leaving only the (still valid) dominance relation R2.

Limitations. The main quantitative result (R3) rests on the following assumptions, with sensitivity noted for each:

- *Rank-width bound $\text{rw} = 4$* : taken as an assumption consistent with the bounded-rank-width claim of [1],[7]; if the true constant is larger, the DP term scales as 4^{rw} but remains dominated by the $O(n^3)$ term for any fixed rw up to about 15 at $n = 1,023$.
- *Treewidth growth $\text{tw} = \Theta(\sqrt{n})$* : an assumption; the numerical gap scales as $2^{n^\alpha}/n^2$, so a smaller α shrinks the gap and $\alpha = 0$ (bounded treewidth) removes the numerical separation entirely, though the superpolynomial claim of [1],[7] for these families then would not hold either.
- *Linear-order DP with $m = n$ steps*: an assumption about the decomposition length; a shorter decomposition proportionally reduces the error bound of R4.
- *Two rounding operations per DP step*: a conservative assumption; one rounding per step halves the bound (R4 is not sensitive at the reported precision).
- *Constant-factor costs*: the operation counts ignore memory traffic and constant factors in matrix multiplication; the 4.05×10^3 ratio is robust to constant factors up to three orders of magnitude.
- *Independence of faults*: the entire formulation assumes independent local fault factors, as in [1],[7]; correlated noise (e.g., measurement crosstalk) falls outside the stated complexity bounds.

Open questions. (Q1) Can the $O(n \cdot 4^{\text{rw}})$ DP term be tightened, and what is the exact per-node operation count of the recurrence? (Q2) Do algebraic classification invariants correlate with rank-width across code families, in the way treewidth correlates with contraction cost? (Q3) Can the decomposition-construction step be improved below $O(n^3)$, e.g. via sparsity-aware elimination? (Q4) Does the structure exploited by exact ML decoding of polar codes over the binary erasure channel [4] admit a rank-width interpretation, unifying the algebraic escapes from ML hardness across classical and quantum coding?

7. Practical Implications and Engineering Pathways

(a) Who would use the result, and what changes. Three user groups follow directly from the applications demonstrated in [1],[7]. First, decoder developers: an exact ML baseline lets them report a certified optimality gap for union-find, matching, or belief-propagation decoders, replacing unverified gap estimates with the quantity $\Delta(s)$ of Section 4.4; the design decision that changes is which decoder variant to ship for a given code family and noise regime. Second, noise-characterization teams: syndrome likelihoods computed exactly support learning circuit noise parameters from syndrome data, as demonstrated in [1],[7]; the change is that likelihood-based parameter fits no longer carry uncontrolled approximation error. Third, fault-tolerance architects: exact event probabilities for postselection evaluation, again demonstrated in [1],[7], let protocol parameters be certified against true probabilities rather than sampling estimates.

(b) Current readiness. This paper is purely analytical and has not been peer reviewed; under the NASA Technology Readiness Level definitions it sits at TRL 1–2 (basic principles observed and reported; concept formulated). The underlying Rank DP algorithm of [1],[7] has been implemented and tested on code-capacity and circuit-level instances up to 1,023 qubits, which is consistent with TRL 3 for the algorithm itself (proof of concept); no claim of manufacturing or deployment readiness is made, and no hardware pathway is implied.

(c) Smallest next test. The smallest de-risking calculation is an independent reimplementing of the nonnegative Rank DP recurrence on the three-qubit repetition code of Section 4.1 and a $[[7, 1, 3]]$ Steane code, comparing computed class probabilities against the closed-form values 0.756 and 0.244 and against brute-force enumeration over all 2^{2^n} fault patterns. The outcome that would refute the practical claim is a computed class probability differing from brute force by more than the Section 4.3 bound $(1 + u)^{2^m} - 1$; agreement within the bound de-risks the error model at negligible cost.

(d) Binding constraints. The first constraint to bind in any implementation is the $O(n^3)$ decomposition-construction cost (F2): at $n = 1,023$ this is $\approx 1.07 \times 10^9$ operations per code instance, amortizable across many syndrome evaluations but dominant for one-shot use. The second is memory for the DP tables: tables of size 2^{rw} per decomposition node are small for bounded rw , but the fault-block weight tables $\sum_B 2^{|B|}$ (part of the input size N in Theorem 1.1 of [1]) can grow with circuit depth in the circuit-level setting. Whether measurement crosstalk and other correlated noise can be folded into local fault blocks without inflating rank-width is an open question, not a derived constraint.

(e) Order-of-magnitude performance bounds (projections). Under the assumptions of Section 4.2 ($\text{rw} = 4$, $\text{tw} = \Theta(\sqrt{n})$), the projected operation counts are $\approx 1.07 \times 10^9$ (Rank DP) versus $\approx 4.35 \times 10^{12}$ (contraction) at $n = 1,023$, ratio $\approx 4.05 \times 10^3$; under the assumptions of R6 ($t = 3$, $k_* = 4$, $c = 10$), $\approx 8.78 \times 10^5$ versus $\approx 4.60 \times 10^{10}$, ratio $\approx 5.2 \times 10^4$. These are projections from stated assumptions, not measurements; no wall-clock time, energy, or cost figure is derivable from the analysis.

(f) Risks and misuse paths. A certified exact baseline can be misused if its assumptions are silently violated: applying the 13-digit error bound of R4 to an implementation that uses a cancellation-prone recurrence, or applying the polynomial-time claim to a code family outside the identified ones, would produce overconfident benchmarking claims. The falsification criteria of Section 6 are the guard against this: any use of the result should be accompanied by the brute-force cross-check of (c) at small scale.

8. Conclusion

We analyzed the rank-decomposition dynamic programming paradigm for exact maximum-likelihood quantum decoding, deriving three concrete results. First, a fully explicit computation shows that exact logical-class probabilities differ measurably from dominant-pattern estimates even on the smallest nontrivial code (0.756 versus 0.729 for the three-qubit repetition code at $p = 0.1$), quantifying the optimality-gap information that exact evaluation provides. Second, the structural relation $\text{rw}(G) \leq \text{tw}(G) + 1$ guarantees that Rank DP is never asymptotically worse than tensor-network contraction, and under stated assumptions the operation-count separation at $n = 1,023$ is $\approx 4.05 \times 10^3$ -fold, growing without bound in n . Third, the nonnegative realization carries a floating-point error budget of $\lesssim 2.3 \times 10^{-13}$ relative error at that scale, certifying about 13 significant decimal digits. Together these convert superpolynomial exact decoding into polynomial-time exact decoding for the punctured quantum Reed–Muller and Steane-concatenated families identified in [1],[7], and give decoder designers a certified exact baseline. The result's scope is bounded by its assumptions — bounded rank-width after Gaussian elimination, independent local faults, and the stated arithmetic model — and the falsification criteria of Section 6 make those bounds testable.

Provenance and AI Involvement

- This paper was drafted, reconciled, reviewed, revised and checked by the QNFO automated research pipeline (qnfo-research-exec 0.17.5-abstract-lines). Language models were tools in that process; they are not authors. The named author is accountable for the content.
- Models: drafting by openai/gpt-oss-120b, zai-org/glm-5.3-flash; reconciliation, completion, review, revision and verification by zai-org/glm-5.3-flash, zai-org/glm-5.3, openai/gpt-oss-120b, with the QNFO model router as a fallback.
- Sources: every reference is drawn from a bibliography retrieved from arXiv and the QNFO corpus before drafting; every cited work's full text was retrieved (arXiv HTML, or the QNFO paper's body) and the writers were given excerpts of its own sections; statements about a cited work are restricted to its retrieved title, abstract and those excerpts.
- Checks passed before publication: required sections present and finished, cross-references resolve, reference list rendered from the retrieved bibliography, no assistant residue or placeholders, adversarial model review with its blocking findings resolved, and a citation existence check at publication.
- Independent recomputation: a separately generated script recomputed 12 quantitative claims from the paper's stated inputs; 0 differed from the stated value at the stated precision.
- Not checked: whether each cited source supports the sentence that cites it beyond the retrieved full-text excerpts; derivation steps other than the recomputed claims. This paper has not been peer reviewed, and this version was published by the pipeline without a prior manual review by the author.

References

- [1] TITLE: arXiv Query: `searchquery=&idlist=2609.39556&start=0&max_results=1` [2] Trigonometrically approximated maximum likelihood estimation for stable law. arXiv:2209.08980v1. <https://arxiv.org/abs/2209.08980v1> [3] Developments in Maximum Likelihood Unit Root Tests. arXiv:1611.00819v1. <https://arxiv.org/abs/1611.00819v1> [4] Efficient Maximum Likelihood Decoding of Polar Codes Over the Binary Erasure Channel. arXiv:2106.14753v1. <https://arxiv.org/abs/2106.14753v1> [5] Exact maximum likelihood inference for drifted multi-sub-fractional Brownian motion at discrete observation. arXiv:2609.22617v1. <https://arxiv.org/abs/2609.22617v1> [6] Quasi Maximum Likelihood Estimation of High-

Dimensional Factor Models: A Critical Review. arXiv:2303.11777v5. <https://arxiv.org/abs/2303.11777v5> [7] Exact Maximum Likelihood Decoding beyond Treewidth via Rank-Decomposition Dynamic Programming. arXiv:2609.39556v1. <https://arxiv.org/abs/2609.39556v1> [8] Maximum Likelihood Estimation of Sparse Networks with Missing Observations. arXiv:1902.10605v2. <https://arxiv.org/abs/1902.10605v2> [9] Filtering the Maximum Likelihood for Multiscale Problems. arXiv:1305.1918v4. <https://arxiv.org/abs/1305.1918v4> [10] EDCHO: High Order Exact Dynamic Consensus. arXiv:2202.03012v2. <https://arxiv.org/abs/2202.03012v2> [11] Verifying the existence of maximum likelihood estimates for generalized linear models. arXiv:1903.01633v10. <https://arxiv.org/abs/1903.01633v10> [12] QNFO: Exact Maximum-Likelihood Quantum Decoding beyond Treewidth: A Reconciled Structural Analysis of Rank-Decomposition Dynamic Programming