

Decryption-Failure Rate with Multidimensional Lattice Decoders: Unified Framework, Theory Refinement, and More Accurate Evaluation

Boyue Fang^{1†}, Songlin Li^{1†}, and Yunlei Zhao¹

College of Computer Science and Artificial Intelligence, Fudan University, Shanghai, China

byfang16@fudan.edu.cn, slli22@m.fudan.edu.cn, ylzhaofudan.edu.cn

Abstract. Multidimensional lattice coding could represent one of the future directions for constructing lattice-based key-encapsulation mechanisms (KEM). Its powerful error correction capability allows for much broader ranges of error noise and ciphertext compression, leading to more compact and robust KEM instantiations. However, multidimensional lattice decoders complicate decryption failure rate (DFR) analysis in two distinct ways: (1) a bounded-distance decoding (BDD) or packing certificate can occur much more often than an implemented message failure, and (2) polynomial multiplication can create nontrivial dependence among coordinates within a decoder block. In this work, we present a unified framework for analyzing the DFR of lattice-based KEMs using multidimensional lattice coding, incorporating decoder event, residual law, tail calculation, and ciphertext- or key-level aggregation. As case studies in this framework, we focus on BW-KEM, CTRU, DTRU, and their prime-degree variants, where CTRU is already in the standardization progress in China, BW-KEM and DTRU are the proposals to China’s Next-generation Commercial Cryptographic (NGCC) Algorithms Program.

We give distribution-free bounds from local boundary crossings to ciphertext events, as well as finite Gaussian bounds comparing radial and facet events. For BW-KEM, we extend the product limit to bounded centered coefficient distributions, with joint convergence on every fixed BW_{32} block. Under the Gaussian-product law, conditioning on terminal discrete noise gives an exact noncentral chi-square mixture for each block tail. Across the published parameters, its blockwise union sums are 6.537–7.621 bits lower on the $\log_2 p$ scale than the published Chernoff sums. At five diagnostic points, the discrepancy between the eight-block mixture sum and the implemented any-block BDD-exit probability is at most 1.116 bits, compared with 3.388 bits for the clipped Chernoff bound. Paired end-to-end experiments for CTRU and DTRU show that certificate exits can be substantially more frequent than implemented failures. At the CTRU diagnostic point, 2,396 packing exits contain 26 message failures, a 6.526-bit certificate-to-failure gap. Across six DTRU points, the gap ranges from 5.85 to 7.63 bits. For the prime-degree rings, we derive the adjacent-coefficient covariance and prove a joint normal limit for

^{**} Boyue Fang and Songlin Li contributed equally to this work.

fixed collections of nonadjacent product coordinates. A parity-separated layout removes these adjacent covariance terms within each decoder block.

Keywords: decryption failure rate · lattice-based cryptography · BW_{32} lattice coding · E_8 lattice coding · event alignment

1 Introduction

Lattice coding integrates modulation and error correction in the cryptosystem’s coefficient space, without the separate iterative decoder used by many conventional codes. Structured fixed-dimensional lattices such as E_8 and BW_{32} admit regular, data-independent block algorithms and scalable labelings, which facilitate constant-time implementations. Their correction margin can accommodate more cryptographic noise or a smaller modulus with stronger compression, improving concrete security or bandwidth [22,18,29,17,19].

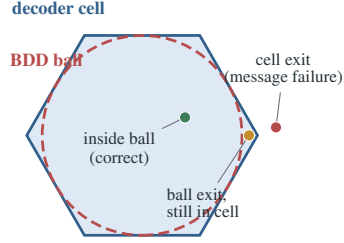
The corresponding decryption-failure calculation with multidimensional lattice decoders is more difficult. A multidimensional decoder cell can have many message-changing facets, and ring products couple the coordinates in each block. Existing analyses use different events, including BDD-ball exit and exact cell exit, and different residual models and tail calculations. Direct Monte Carlo cannot reach the probabilities used in deployed parameters. A sound analysis must therefore state what probability it computes, relate it to implemented failure, and justify extrapolation from observable to target probabilities. Figure 1 illustrates the distinction used throughout the paper: panel (a) separates a BDD certificate from implemented message failure, while panel (b) shows how the two events are measured on the same residual samples. These issues arise in both cryptographic research [28,10] and standardization. The public-key submission rules of China’s Next-generation Commercial Cryptographic (NGCC) Algorithms Program, for example, expressly require a detailed failure-probability analysis and its security implications [16].

Accuracy at this scale affects more than a correctness table. Correctness appears in Fujisaki–Okamoto security reductions [13], while observable failures can support key recovery through decryption-failure attacks, failure boosting, adaptive searches, and multitarget attacks [8,9,7,4,12]. An error of several bits in the failure exponent changes both a parameter decision and the estimated cost of exploiting failures.

We record every analysis as (A, Q, T, G) : decoder event, residual law, tail method, and ciphertext or key aggregation. This separates certificate conservatism from distributional and numerical error. The framework provides distribution-free directional bounds, a finite Gaussian comparison between radial and cell events, and paired experiments on common residuals. It is compatible with scalar decoders such as ML-KEM [24]. Our scheme analyses concern multidimensional lattice-code blocks.

We focus on three schemes that are relevant to China’s post-quantum standardization work: CTRU, DTRU, and BW-KEM.

(a) What the BDD condition certifies



(b) Measuring the gap with paired experiments

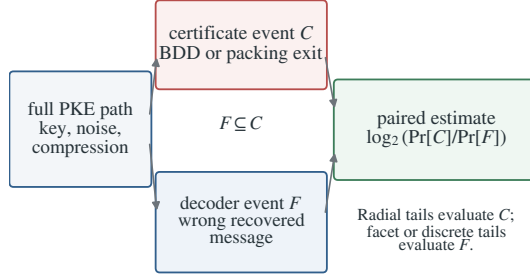


Fig. 1. The probability target of the paper. In (a), the BDD ball is inside the correct decoder cell; a ball exit can still decode correctly, whereas a cell exit changes the message. In (b), an end-to-end PKE experiment records the certificate event C and implemented failure event F on the same residuals, so $\log_2(\Pr[C]/\Pr[F])$ measures the event gap. Radial analyses target C ; facet analyses target their stated proxy events, and implemented-decoder measurements target F .

- **CTRU** decodes eight-dimensional E_8 blocks. We compare its packing certificate, the 224 message-changing facet directions of Zhou et al., and the implemented decoder [18,28].
- **DTRU** jointly decodes paired E_8 words in a 16-dimensional observation space, requiring their covariance to be retained [29].
- **BW-KEM** uses a 32-dimensional BW_{32} BDD. We revisit the product law and tail calculation for its fixed-radius event [17].
- **Prime-degree variants.** CTRU-Prime and DTRU-Prime use the noncyclotomic quotient $\mathbb{Z}_q[x]/(x^n - x - 1)$ with prime n . Following the NTRU Prime rationale, this choice avoids the specific rich subfield and automorphism structure of cyclotomic rings [3]; the relevant number-field objects are distinguished in section 2. Quotient reduction correlates adjacent coefficients, so the codeword layout affects each decoder block.

CTRU entered an industry-standard project in 2023 [6]; its design and a high-throughput CTRU-Prime implementation also appear in *Computer Standards & Interfaces* and *Journal of Software* [18,14]. In 2025, the Institute of Commercial Cryptography Standards launched a call for candidate public-key algorithms under the NGCC initiative [16]; DTRU and BW-KEM are candidates [29,17]. Their E_8 , paired- E_8 , and BW_{32} decoders, together with the prime-degree variants, form a complementary sample of events and residual structures.

Contributions.

1. We give a finite ratio between radial and facet-event probabilities, including corrections for tail approximation, facet overlap, and farther decision regions.

Directional bounds cover discrete residuals, and the (A, Q, T, G) record keeps the resulting probability objects distinct.

2. We extend BW-KEM’s product limit to bounded centered coefficients and prove joint convergence on each fixed block. Under its Gaussian product reference, a finite noncentral chi-square mixture exactly evaluates each published fixed-radius BDD block tail; ciphertext aggregation remains an eight-block union sum.
3. For CTRU and DTRU, we identify the relevant E_8 facets and derive the paired-block covariance used by the DTRU Gaussian reference. This yields a conservative message-facet event for CTRU, and covariance-preserving BDD and nearest-facet references for DTRU.
4. For the Prime variants, we derive the quotient-ring covariance and prove joint asymptotic independence for each fixed collection of nonadjacent product coordinates under bounded coefficient laws. A parity-separated codeword assignment removes adjacent pairs from every decoder block in the second-order model.
5. We use independent keys and evaluate the certificate and failure events on the same residuals in paired end-to-end PKE experiments. Figure 2 summarizes the empirical comparisons: panel (a) matches each CTRU estimate to its target event, while panel (b) measures the gap between the certificate and decoder events for CTRU and DTRU. The largest runs contain 2^{30} CTRU ciphertexts, 2^{30} DTRU ciphertexts, and independent 2^{32} -trial confirmations. CTRU has a 6.526-bit event gap and DTRU gaps of 5.85–7.63 bits on the stated diagnostic grids. On the measurable BW-KEM grid, the largest combined diagnostic discrepancy is 3.388 bits for clipped Chernoff and 1.116 bits for the mixture. Pair-tail, cumulant-generating function (CGF), and likelihood-ratio conditions state what is required to extrapolate.

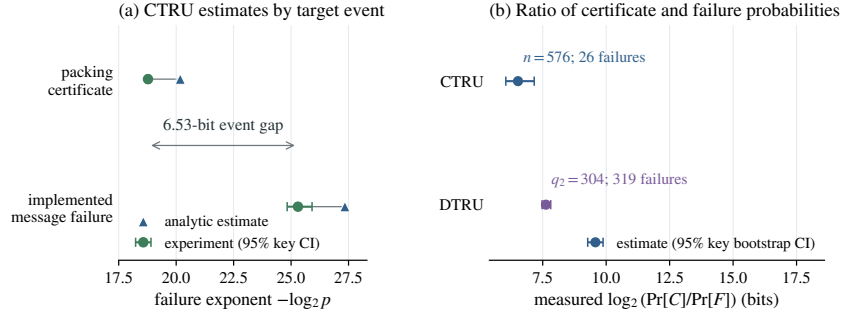


Fig. 2. Empirical comparisons for matched events. Panel (a) matches each CTRU calculation to its event at the diagnostic row using the events and models listed in table 3; section 3.4 separately gives the isotropic 2^{-20} geometric references. Panel (b) reports $\log_2(\Pr[C]/\Pr[F])$ for CTRU and DTRU with 95% bootstrap intervals obtained by resampling keys. Circles with error bars are experimental estimates. The remaining markers are analytic or Gaussian-reference calculations.

Related work. Prior DFR work studies dependent ring noise [10]. Zhou et al. give the CTRU facet analysis and methods for correlated tails [28]; DTRU and BW-KEM give the paired- E_8 and BW_{32} analyses [29,17]. Other work treats directional dependence in polynomial products [20], failure finding, and verified correctness [2,23]. The public permutation used for the Prime variants adopts the coefficient ordering proposed in [19].

Organization. Section 2 fixes notation and describes the schemes and their block decoders. Sections 3 to 6 then present the framework and scheme analyses; section 7 compares their evidence. The appendices provide supporting calculations and reproducibility information.

2 Preliminaries

We fix the notation and the parts of CTRU, DTRU, and BW-KEM that determine decryption correctness. The byte encodings and hash-domain conventions of the key-encapsulation mechanism (KEM) wrappers do not change the decoder observations and are omitted.

2.1 Notation

Write $\mathbb{Z}_q := \mathbb{Z}/q\mathbb{Z}$. For a monic polynomial ϕ of degree n , write $R(\phi) = \mathbb{Z}[x]/(\phi)$ and $R_q(\phi) = \mathbb{Z}_q[x]/(\phi)$, and identify $a = \sum_{i=0}^{n-1} a_i x^i$ with $(a_0, \dots, a_{n-1})$. For a residue class $a \in \mathbb{Z}_q$, $[a]_q$ is its selected centered integer representative in $[-q/2, q/2]$; selection and reduction extend coefficientwise. For a coefficient vector, we use the centered modular Euclidean norm $\|a\|_{q,2} := (\sum_i [a_i]_q^2)^{1/2}$, and I_d denotes the $d \times d$ identity matrix. Finally, $\lfloor z \rfloor := \lfloor z + 1/2 \rfloor$ denotes round-to-nearest with half-integer ties rounded upward. The centered binomial distribution B_η is the law of $\sum_{i=1}^\eta (a_i - b_i)$ for independent uniform bits. Its variance is $\eta/2$, and polynomial samples have independent coefficients unless stated otherwise. For $Z \sim \mathcal{N}(0, 1)$, we write $\Phi(x) = \Pr[Z \leq x]$ for the standard normal CDF, $\bar{\Phi}(x) = 1 - \Phi(x) = \Pr[Z \geq x]$ for its upper tail, and $\varphi(x) = e^{-x^2/2}/\sqrt{2\pi}$ for its density.

For $q > q' > 0$, the coefficient maps have types

$$\text{Comp}_{q \rightarrow q'} : \mathbb{Z}_q \longrightarrow \mathbb{Z}_{q'}, \quad \text{Decomp}_{q' \rightarrow q} : \mathbb{Z}_{q'} \longrightarrow \mathbb{Z}_q,$$

and are

$$\text{Comp}_{q \rightarrow q'}(a) = \left\lfloor \frac{q'}{q} [a]_q \right\rfloor \bmod q', \quad \text{Decomp}_{q' \rightarrow q}(c) = \left\lfloor \frac{q}{q'} [c]_{q'} \right\rfloor \bmod q. \quad (1)$$

The displayed integer formulas use the selected representatives before reduction modulo the target modulus. The maps extend coefficientwise to ring elements and vectors. BW-KEM writes $q' = 2^d$ and uses the names $\text{Compress}_{q,2^d}$ and $\text{Decompress}_{2^d,q}$. Each KEM applies a Fujisaki–Okamoto transform to the

PKE described below: decapsulation decrypts to m' , derives the coins again, and checks a re-encryption. For the Fujisaki–Okamoto wrappers considered here, an honest key mismatch can arise from failure to recover the encapsulated PKE message, apart from the explicitly neglected hash and key-derivation collision events. We do not use a wrapper-independent two-way equivalence beyond these scheme specifications.

2.2 Block codes

E_8 and paired- E_8 . Let C_8 be the binary $[8, 4, 4]$ code generated by 11110000, 00111100, 00001111, and 01010101. The Construction-A lattice $C_8 + 2\mathbb{Z}^8$ is the E_8 representation used by CTRU [18]. Under this normalization its minimum Euclidean distance is 2, hence its packing radius is 1; scaling by λ gives packing radius λ . At scale λ , the decoder returns the label of a closest point modulo the shaping lattice, and $\|z\|_2 < \lambda$ guarantees correct recovery. DTRU uses $C_{\text{pair}} = \{(u, u) : u \in C_8\} \subset \{0, 1\}^{16}$, whose minimum Hamming distance is eight. Its decoder tests all 16 labels using the sum of two eight-coordinate torus distances. At scale $(q+1)/2$, the sufficient condition is $\|z\|_{q,2}^2 < (q-1)^2/2$ [29].

BW_{32} . Let $b = 2^k$, let $a = (a_0, \dots, a_{N-1})$ be the coefficient vector to be partitioned, and put $d = N/32$ with $0 \leq j < d$. Define

$$W_b = \begin{pmatrix} 1 & 1 \\ 0 & 1+i \end{pmatrix}^{\otimes(k-1)}, \quad BW_b(\lambda) = \{\lambda v W_b : v \in \mathbb{Z}[i]^{b/2}\}, \quad (2)$$

$$a^{(j)} = (a_j, a_{j+d}, \dots, a_{j+31d}).$$

For an integer $\tau \geq k/2$, let $A_c = 2^\tau \lambda \mathbb{Z}[i]^{b/2}$ and let its standard fundamental parallelotope be

$$\mathcal{P}_c := \{(\alpha_1 + \beta_1 i, \dots, \alpha_{b/2} + \beta_{b/2} i) : \alpha_j, \beta_j \in \mathbb{R}, 0 \leq \alpha_j, \beta_j < 2^\tau \lambda, 1 \leq j \leq b/2\}.$$

BW-KEM uses the nested code $C_{b,\tau}(\lambda) := BW_b(\lambda) \cap \mathcal{P}_c$ [17]. The implemented decoder is recursive BDD, not maximum-likelihood decoding (MLD); the packing radius of $BW_b(\lambda)$ is $\sqrt{b/8}\lambda$. All rows below have $b = 32$ and $N = 256$, so $d = 8$ in equation (2) [17].

2.3 Schemes

CTRU. CTRU works in $R_q^{\text{tri}} = R_q(x^n - x^{n/2} + 1)$ with $8 \mid n$ and $q_2 < q$. It samples small f', g , sets $f = 2f' + 1$, and publishes $h = gf^{-1}$. If $M_q(m)$ places successive four-bit groups in scaled E_8 blocks, then, for fresh small r ,

$$c = \text{Comp}_{q \rightarrow q_2}(hr + M_q(m)), \quad m' = \text{PolyDecode}_{E_8}(fc \bmod q_2), \quad (3)$$

$$X = gr + (q/q_2)\varepsilon f.$$

The decoder uses scale $q_2/2$. There are $n/8$ consecutive blocks carrying $n/2$ message bits, and ε is the compression error lifted to modulus q .

DTRU. DTRU uses the same tricyclotomic ring and public key but samples r, e . For the binary paired- E_8 polynomial $w = \text{Encode}_{\text{pair}}(m)$,

$$c = \text{Comp}_{q \rightarrow q_2} \left(hr + e + \frac{q+1}{2}w \right), \quad m' = \text{Decode}_{\text{pair}}(f \text{Decomp}_{q_2 \rightarrow q}(c)),$$

$$\text{Err} = gr + f(e + \varepsilon) + f'w. \quad (4)$$

Each of the $n/16$ paired blocks joins two octets and carries four bits. Here ε is the compression–decompression error; the message-dependent term $f'w$ is retained in the analysis.

BW-KEM. BW-KEM is a module learning with errors (MLWE) construction over $R_q^{\text{neg}} = R_q(x^N + 1)$. For $A \in (R_q^{\text{neg}})^{\ell \times \ell}$ and small $s, e \in (R_q^{\text{neg}})^\ell$, the public key is $t = As + e$. Encryption samples $r, e_1 \in (R_q^{\text{neg}})^\ell$ and $e_2 \in R_q^{\text{neg}}$:

$$u = \text{Compress}_{q, 2^{d_u}}(A^T r + e_1),$$

$$v = \text{Compress}_{q, 2^{d_v}}(t^T r + e_2 + \text{PolyEncode}_{BW}(m)), \quad (5)$$

$$\text{Err}_{BW} = e^T r - s^T(e_1 + \varepsilon_u) + e_2 + \varepsilon_v.$$

Set $\tilde{u} := \text{Decompress}_{2^{d_u}, q}(u)$ and $\tilde{v} := \text{Decompress}_{2^{d_v}, q}(v)$. Decryption decodes $\tilde{v} - s^T \tilde{u}$ in the eight blocks of equation (2). In equation (5), $\varepsilon_u = \tilde{u} - (A^T r + e_1)$ and $\varepsilon_v = \tilde{v} - (t^T r + e_2 + \text{PolyEncode}_{BW}(m))$. Correctness is guaranteed when every block remains within the recursive-BDD radius.

Prime-degree variants. CTRU-Prime and DTRU-Prime replace the tricyclotomic ring by $R_q^{\text{pr}} = \mathbb{Z}_q[x]/(x^n - x - 1)$, where n is prime and $x^n - x - 1$ is irreducible modulo q . If α is a complex root, the number field is $K = \mathbb{Q}(\alpha)$, while the “large Galois group” refers to the Galois group of the Galois closure of K over $\mathbb{Q}$ [3], not to $\text{Aut}(K)$ or to the finite quotient R_q^{pr} . That rationale chooses a noncyclotomic prime-degree field to avoid the specific subfield and automorphism structure of cyclotomic rings; a large Galois group alone is not used here to assert the absence of all automorphisms or subfields. The calculation in section 6 uses only the quotient relation $x^n = x + 1$. The PKE flow and decoders stay the same.

For CTRU-Prime-761, $n_{\text{msg}} := 8\lfloor n/16 \rfloor = 376$ counts message bits; their encoding occupies the first $2n_{\text{msg}} = 752$ polynomial coefficients and therefore forms $752/8 = 94$ E_8 blocks. DTRU-Prime-1087 has 67 paired blocks joining positions $0, \dots, 535$ to $543, \dots, 1078$. Adjacent product coefficients can therefore contain common ordinary-convolution terms. Section 6 compares these original layouts with the parity-separated assignments defined there.

3 Unified Framework for DFR Analysis

3.1 Probability target and notation

Let $\Pi = (\text{KeyGen}, \text{Encaps}, \text{Decaps})$ be a KEM. In the honest correctness experiment, sample $(pk, sk) \leftarrow \text{KeyGen}$ and $(c, ss) \leftarrow \text{Encaps}(pk)$, and set $\mathcal{F}_{\text{hon}} =$

$\{\text{Decaps}(sk, c) \neq ss\}$. The one-shot correctness error is $\delta_\Pi = \Pr[\mathcal{F}_{\text{hon}}]$. The probability of a proven geometric superset is an upper bound on this event probability, not a replacement definition. For a fresh key K and honest coins R , using the residual and indicator notation formalized below, write

$$p_k = \mathbb{E}_R[D_k(\Psi_k(R))], \quad \bar{p} = \mathbb{E}_K[p_K], \quad w(t) = \Pr_K[p_K > t]. \quad (6)$$

Standard correctness is $\bar{p}$, but it does not determine repeated-use correctness or $w(t)$; the probability of at least one failure for N keys and s uses per key is $1 - (\mathbb{E}_K[(1 - p_K)^s])^N$. This formula assumes that the N keys are generated independently and, conditional on each key, the s honest encapsulations use independent fresh randomness; p_K is the corresponding key-conditioned one-shot failure probability. Key tails also matter in multi-target settings [8,4].

As noted in section 2, our end-to-end PKE runs track failure to recover the encapsulated message. For the Fujisaki–Okamoto wrappers considered here, this is the wrapper-level mechanism for honest key mismatch in our accounting, apart from the explicitly neglected hash or key-derivation collision events.

Failure-search attacks also change the ciphertext law and require an observable decapsulation failure. Thus an honest DFR does not by itself give an attack rate, but event alignment remains necessary: a proxy exit that decodes correctly supplies no failure signal. A d -bit DFR error changes the predicted work of an inverse-probability search by 2^d .

We describe each estimate by (A, Q, T, G) : its event, reference law, numerical procedure or bound, and key-level target. Comparisons that change one field have a precise interpretation; comparisons that change several fields are reported as combined model differences. For a probability $p \in (0, 1]$, write $E(p) := -\log_2 p$; we call $E(p)$ its failure exponent and reserve *log-probability* for $\log_2 p$.

We write K for a random key, k for a fixed key, R for honest encryption randomness, and $Z = \Psi_k(R)$ for the implemented residual. We use F for implemented message failure and C for a certificate event; the measurable key-conditioned implemented event is $\mathcal{F}_k$ below. The certificate event provides an upper bound on failure only after establishing $F \subseteq C$. A BDD exit lies outside the guaranteed decoding ball. The remaining notation follows section 2 or is introduced when needed.

3.2 Comparison record and protocol

The definitions below apply beyond CTRU, E_8 decoding, and Gaussian residuals.

3.2.1 Residual map, decoder event, and keys Let $K \sim P_K$, let Ω_k be the honest-encryption randomness sample space for key k with probability law $P_{R|k}$, and let $(\mathcal{Z}, \mathfrak{Z})$ be the measurable decoder observation space. A structured-lattice KEM defines a measurable deterministic residual map

$$\Psi_k : \Omega_k \longrightarrow \mathcal{Z}, \quad Z = \Psi_k(R), \quad R \sim P_{R|k}, \quad (7)$$

whose implemented residual law is $P_k := \mathcal{L}(\Psi_k(R) \mid K = k)$. Its measurable decoder-failure event is

$$\mathcal{F}_k := \{z \in \mathcal{Z} : \text{Decode}_k(\mu + z) \neq \mu\} \in \mathfrak{Z}, \quad D_k(z) := \mathbf{1}\{z \in \mathcal{F}_k\}. \quad (8)$$

Translation invariance is useful but not required; when it fails, the message is included in R . The implemented key-conditioned error is $p_k = P_k(\mathcal{F}_k) = \mathbb{E}_{Z \sim P_k}[D_k(Z)]$.

For analysis, a scheme may also use a score $S_k(z)$ with $\mathcal{F}_k = \{z : S_k(z) \geq 1\}$, a decomposition of $\mathcal{F}_k$ into simpler events, or a tractable reference probability measure Q_k on $(\mathcal{Z}, \mathfrak{Z})$. These devices are scheme-specific.

3.2.2 Decoder-dependent analysis requirements A scalar decoder has coefficientwise decisions; CTRU uses eight-dimensional E_8 blocks; DTRU jointly decodes paired E_8 observations in 16 dimensions; and BW-KEM uses 32-dimensional BW_{32} blocks. Decoder separability selects the natural scalar, facet, or radial event, while ring multiplication determines whether a marginal or joint residual law is required. These choices determine the event and residual law used for each scheme.

3.2.3 Separating event and distribution effects Changing the event and changing the residual model need not have additive effects. Let $u_{ab} > 0$ be four values obtained with the same aggregation, where $a \in \{0, 1\}$ chooses the event and $b \in \{0, 1\}$ chooses the residual model, and put $v_{ab} = -\log_2 u_{ab}$. Averaging over the two replacement orders gives

$$\Delta_A = \frac{1}{2}[(v_{10} - v_{00}) + (v_{11} - v_{01})], \quad \Delta_Q = \frac{1}{2}[(v_{01} - v_{00}) + (v_{11} - v_{10})]. \quad (9)$$

Direct expansion gives

$$\Delta_A + \Delta_Q = v_{11} - v_{00}. \quad (10)$$

If every u_{ab} is an exact probability under the same aggregation map, the two terms are the symmetric log-scale changes associated with replacing the event and residual law. The CTRU values in table 5 are instead ciphertext-level block-wise union sums. Equation (9) then gives a descriptive log-scale decomposition of the four sums, without a causal interpretation as exact ciphertext probabilities. In either case, equation (10) holds algebraically.

3.2.4 Comparison protocol For two events A and B under the same law Q ,

$$Q(A) - Q(B) = Q(A \setminus B) - Q(B \setminus A). \quad (11)$$

Equation (11) measures event replacement through paired disagreement. If $B \subseteq A$, only the certificate-only term remains; otherwise both directions must be measured. The identity applies keywise before G .

For an implemented pair $(P_k, \mathcal{F}_k)$ and a reference pair (Q_k, B_k) on the same decoder observation space, the triangle inequality gives

$$|Q_k(B_k) - P_k(\mathcal{F}_k)| \leq Q_k(\mathcal{F}_k \triangle B_k) + |Q_k(\mathcal{F}_k) - P_k(\mathcal{F}_k)|.$$

The terms are, respectively, event disagreement under Q_k and residual-law error on the implemented event. They can be averaged across keys for a linear unweighted mean; nonlinear targets such as repeated-use failure and $w(t)$ require applying G separately. The law term is event-specific: direct total-variation comparison of a discrete residual and a continuous Gaussian is vacuous, and bit-level claims need relative probability control. Our diagnostics therefore compare events on common residuals, then change the law with the event fixed. The independent-key indicator arguments in section 3.5 are scheme-independent; covariances, regions, and DFR values are not.

3.3 From local tails to ciphertext failure and smaller probabilities

Information at a measurable tail controls deployment parameters only under additional conditions, so we keep local tails separate from their union. For m local decisions, let $S_{k,j}(R)$ be a score with boundary one and define

$$N_{k,t} = \sum_{j=1}^m \mathbf{1}\{S_{k,j}(R) \geq t\}. \quad (12)$$

Here j may index a scalar, an E_8 or BW_{32} block, or a paired- E_8 decoder. Set $E_{k,t} = \{N_{k,t} > 0\}$, $A_t = \mathbb{E}[N_t]$, and $\nu_{2,t} = \mathbb{E}[N_t(N_t - 1)]$, conditionally on a key or including key generation. No independence is assumed. At $t = 1$, $E_{k,t}$ is the implemented failure or certificate event according to the score; shifted thresholds are diagnostic.

Theorem 1 (Bounds from local crossings to the union event). *For arbitrarily dependent local decisions,*

$$\max\{0, A_t - \nu_{2,t}/2\} \leq \Pr[N_t > 0] \leq \min\{1, A_t\}. \quad (13)$$

The pointwise inequality $n - \binom{n}{2} \leq \mathbf{1}\{n > 0\} \leq n$ proves the result on taking expectations; thus $\nu_{2,t} \ll A_t$ certifies a tight union bound.

Rare-event bounds and transport. For scalar facet, coefficient, or radial statistics, put $K_j(s) = \ln \mathbb{E}[e^{sY_j}]$ and $K_{ij}(s, r) = \ln \mathbb{E}[e^{sY_i + rY_j}]$. Markov's inequality gives

$$\Pr[Y_j \geq u] \leq e^{-I_j(u)}, \quad I_j(u) = \sup_{s \geq 0} \{su - K_j(s)\}, \quad (14)$$

$$\Pr[Y_i \geq u, Y_j \geq u] \leq \inf_{s, r \geq 0} e^{K_{ij}(s, r) - (s+r)u}. \quad (15)$$

Their sums control A_t and $\nu_{2,t}$; the lower endpoint additionally needs exact or certified lower local probabilities. If two CGFs differ by at most ε on the

optimized tilt interval, their Chernoff exponents differ by at most $\varepsilon/\ln 2$ bits. A change of law $P \ll Q$ instead uses, for $\alpha > 1$ and every event A , Hölder's bound

$$P(A) \leq \mathbb{E}_Q[(dP/dQ)^\alpha]^{1/\alpha} Q(A)^{1-1/\alpha}. \quad (16)$$

Thus fixed-law extrapolation needs uniform CGF control, whereas a change of law needs a likelihood moment or an equivalent finite-support bound.

3.4 BDD certificates versus message failure

An inscribed-ball exit is a correctness certificate, whereas crossing a message-changing decoder boundary is the failure event. For an arbitrary continuous, discrete, or mixed residual Z , let $H_i = \{z : a_i^T z \geq h_i\}$, $i = 1, \dots, m$, be the nearest message-changing facet events, with boundary atoms assigned by the implemented tie rule. Put $U_1 = \sum_i \Pr[Z \in H_i]$ and $U_2 = \sum_{i < j} \Pr[Z \in H_i \cap H_j]$.

Theorem 2 (Distribution-free message-facet interval). *For every law of Z ,*

$$\max\{0, U_1 - U_2\} \leq \Pr\left[Z \in \bigcup_i H_i\right] \leq \min\{1, U_1\}. \quad (17)$$

If the full failure event also contains a farther region $\mathcal{R}_{\text{far}}$, its probability is added to the upper endpoint.

This is theorem 1 with $N = \sum_i \mathbf{1}\{Z \in H_i\}$, $\Lambda = U_1$, and $\nu_2/2 = U_2$; a union bound adds $\mathcal{R}_{\text{far}}$.

For $Z = P + D$, terminal noise and rounding may instead be kept exact:

$$\Pr[Z \in H_i] = \mathbb{E}_D[\Pr[a_i^T P \geq h_i - a_i^T D \mid D]]. \quad (18)$$

The same identity holds for $H_i \cap H_j$, allowing exact finite-support convolution for D while approximating only P .

Finite Gaussian reference. Let

$$F = \bigcup_{i=1}^m \{x : u_i^T x \geq 1\} \cup \mathcal{R}_{\text{far}}, \quad \|u_i\|_2 = 1, \quad c_R = \inf_{x \in \mathcal{R}_{\text{far}}} \|x\|_2 > 1, \quad (19)$$

with $c_R = \infty$ if $\mathcal{R}_{\text{far}}$ is empty. In the normalized Gaussian reference of theorem 3, tF dilates this reference region; no such scaling is asserted for an arbitrary implemented region. For $Y \sim \mathcal{N}(0, I_d)$, define

$$\begin{aligned} P_B(t) &= \Pr[\|Y\|_2 \geq t], & S(t) &= m\bar{\Phi}(t), \\ D(t) &= \sum_{i < j} d_{ij}(t), & d_{ij}(t) &= \begin{cases} 0, & u_i^T u_j = -1, \\ \bar{\Phi}\left(t\sqrt{2/(1 + u_i^T u_j)}\right), & u_i^T u_j > -1, \end{cases} \end{aligned} \quad (20)$$

and $p_F(t) = \Pr[Y \in tF]$. Set

$$M_{d,m}(t) = \frac{\sqrt{2\pi}}{m 2^{d/2-1} \Gamma(d/2)} t^{d-1}, \quad (21)$$

$$\varepsilon_{\text{tail}}(t) = \log_2 \frac{P_B(t)}{S(t)} - \log_2 M_{d,m}(t). \quad (22)$$

Theorem 3 (Finite gap between BDD-ball exit and message failure).
If $D(t) < S(t)$, then

$$\begin{aligned} L_{d,m}(t) &= \log_2 M_{d,m}(t) + \varepsilon_{\text{tail}}(t) - \log_2 \left(1 + \frac{P_B(c_R t)}{S(t)} \right), \\ U_{d,m}(t) &= \log_2 M_{d,m}(t) + \varepsilon_{\text{tail}}(t) - \log_2 \left(1 - \frac{D(t)}{S(t)} \right), \\ L_{d,m}(t) &\leq \log_2 \frac{P_B(t)}{p_F(t)} \leq U_{d,m}(t). \end{aligned} \quad (23)$$

Proof. Let $p_0(t)$ be the union of the nearest half-spaces. The union bound and second Bonferroni inequality give $S(t) - D(t) \leq p_0(t) \leq S(t)$: for $u_i^T u_j > -1$, crossing both half-spaces implies $(u_i + u_j)^T Y \geq 2t$, yielding the stated $d_{ij}(t)$; opposite normals cannot both cross a positive threshold. The farther region contributes at most $P_B(c_R t)$, so $S - D \leq p_F \leq S + P_B(c_R t)$. Dividing $P_B(t)$ by these endpoints and using equations (21) and (22) proves equation (23).

The factor t^{d-1}/m is the finite geometric difference between radial mass and m prescribed directions; the other terms correct for marginal tails, pairwise overlaps, and farther regions. At a BDD tail of 2^{-20} , the (d, m, t) triples and gap intervals are $(8, 224, 6.543)$ and $[7.141, 7.168]$ bits for conservative CTRU E_8 facets, $(16, 3584, 7.645)$ and $[14.637, 14.732]$ for paired- E_8 DTRU facets, and $(32, 146816, 9.240)$ and $[28.971, 29.034]$ for BW_{32} minimal-vector geometry. Here $\mathcal{R}_{\text{far}}$ is empty. The normalized inner products are $\{-1, -1/2, 0, 1/2\}$ for CTRU and $\{-1, -3/4, -1/2, -1/4, 0, 1/4, 1/2, 3/4\}$ for DTRU; BW uses only the certified correlation bound $1/2$. The DTRU and BW values are geometric references, not implemented-decoder events.

3.5 Key averaging and statistical checks across independent keys

Calculations under a reference distribution must be checked against the implemented encryption and decoding path. Sample J independent keys, run m_j honest encryptions under key j , observe Y_j implemented failures, and put $R_j = Y_j/m_j$. Then

$$\hat{p} = \frac{1}{J} \sum_{j=1}^J R_j. \quad (24)$$

This estimator is unbiased for the key average even when the m_j differ. Keys, rather than pooled ciphertexts, are the independent statistical units, so the equal-allocation experiments resample the R_j over keys.

Key conditioning also precedes Gaussian tail evaluation. In general,

$$\mathbb{E}_K[Q_{\Sigma_K}(A)] \neq Q_{\bar{\Sigma}}(A), \quad \bar{\Sigma} = \mathbb{E}_K[\Sigma_K]. \quad (25)$$

The left side is a mixture of key-conditioned Gaussian probabilities; the right side is one Gaussian with averaged covariance. We label the latter as a single reference computed with averaged covariance, not as key averaging. On the CTRU diagnostic with 16 keys, averaging the key-conditioned probabilities of ball exits and facet unions gives exponents 19.028 and 26.574, whereas $Q_{\bar{\Sigma}}$ gives 20.144 and 28.198. Covariance averaging changes these two exponents by 1.116 and 1.625 bits at this diagnostic point. The sample is too small to estimate a key tail, but the two orders in equation (25) give different numerical values.

For paired proxy/implemented events, define

$$\Delta_j = (Y_j^{\text{proxy-only}} - Y_j^{\text{impl-only}})/m_j \in [-1, 1].$$

The mean of Δ_j estimates proxy bias and its interval is formed over keys. For a high-DFR threshold t , simultaneous per-key Clopper–Pearson intervals [5] bracket the number of keys with $p_j > t$; a second binomial interval bounds $\Pr_K[p_K > t]$.

For an i.i.d. Bernoulli confirmation run with common failure probability p , $\text{sd}(\hat{p})/p = \sqrt{(1-p)/(Np)}$. Thus, for $p = 2^{-20}$ and $N = 2^{32}$, the expected failure count is $Np = 4096$ and the relative standard error is 1.56%; if 4096 failures are observed, the exact 95% Clopper–Pearson interval is $[0.9696, 1.0311] \times 2^{-20}$, or -3.04% to $+3.11\%$ relative to p . This binomial calculation does not apply to experiments that reuse each key for multiple ciphertexts with heterogeneous p_K : such trials are only conditionally i.i.d. given the key, so keys remain the independent units and inference is based on the per-key rates R_j as above.

4 BW-KEM: Evaluation of the BDD Event

BW-KEM [17] deploys a recursive BDD for the 32-dimensional Barnes–Wall lattice BW_{32} . The published estimator bounds exit from its certified radius and sums the exit probabilities over eight blocks:

$$(A_{\text{BDD}}, Q_{\text{BW}}, T_{\text{Ch}}, G_{\text{8UB}}).$$

Here G_{8UB} denotes the eight-block union bound. We keep A_{BDD} and G_{8UB} fixed. First, we extend the product limit from the model with Gaussian coefficients in Theorem 6 of [17] to bounded centered laws and state joint convergence on a fixed decoder block. Second, we approximate the product term by the Gaussian law in equation (27) and replace the Chernoff step by a finite mixture that evaluates each block tail exactly under the resulting reference law Q_{BW} . A full Voronoi calculation would describe MLD, whereas the implemented diagnostic event here is BDD exit for Algorithm 2.

4.1 Extending the product limit to bounded centered coefficients

The primitive BW-KEM coefficients are centered binomials. The next result covers any bounded centered law and gives the joint form needed by one decoder block.

Theorem 4 (Joint Gaussian limit for negacyclic products with bounded coefficients). *For $j = 0, \dots, \ell-1$, let $f^{[j]}, g^{[j]} \in \mathbb{R}[x]/(x^N+1)$. Assume that all coefficients in these two families are mutually independent, that the f -coefficients are identically distributed with mean zero, variance $\sigma_f^2 > 0$, and bounded support, and impose the corresponding assumptions with variance $\sigma_g^2 > 0$ on the g -coefficients. Set*

$$h = \sum_{j=0}^{\ell-1} f^{[j]} g^{[j]}.$$

For a fixed m and any distinct indices $c_1, \dots, c_m \in \{0, \dots, N-1\}$, possibly depending on N ,

$$\left(\frac{h_{c_1}}{\sigma_f \sigma_g \sqrt{\ell N}}, \dots, \frac{h_{c_m}}{\sigma_f \sigma_g \sqrt{\ell N}} \right) \xrightarrow{d} \mathcal{N}(0, I_m) \quad (\ell N \rightarrow \infty). \quad (26)$$

Proof (Sketch). For a fixed Cramér–Wold vector, condition on the f -coefficients and write the projection as a triangular array in the independent g -coefficients. Bounded support gives the conditional Lindeberg condition. Frobenius orthogonality of distinct negacyclic shifts makes the conditional variance converge to $\|t\|_2^2$; its fluctuation is $O((\ell N)^{-1})$. Conditional Lindeberg–Feller and Cramér–Wold give equation (26). The quadratic-form estimate is proved in section A.1.

For one decoder block, decompose the residual in equation (5) as $P+D$, where $P = e^T r - s^T e_1 - s^T \varepsilon_u$ contains the polynomial-product terms and $D = e_2 + \varepsilon_v$ is terminal noise. If $\sigma_i^2 = \text{Var}(B_{\eta_i})$, the Gaussian block substitution is

$$X_0 \sim \mathcal{N}(0, \sigma_P^2 I_{32}), \quad \sigma_P^2 = \ell N \sigma_1^2 (\sigma_1^2 + \sigma_2^2 + \sigma_{\varepsilon_u}^2). \quad (27)$$

The reference law Q_{BW} is the law of $X_0 + D$ obtained by this substitution. Under Q_{BW} , the coefficient families contributing $e^T r$, $-s^T e_1$, and the modeled rounding product $-s^T \varepsilon_u$ are treated as independent when forming the displayed covariance; D retains the finite discrete distribution induced by e_2 and ε_v , its 32 coordinates are iid, and D is independent of X_0 . This defines a blockwise reference distribution, not the finite-parameter implemented residual law. No cross-block independence is assumed by the factor-eight union sum below. The product replacement is a finite-parameter Gaussian substitution, not an identity. Applying the Cramér–Wold argument jointly to the independent layers gives the block-diagonal limiting covariance; marginal convergence alone would not justify adding their variances.

4.2 Conditional noncentral mixture

Conditional on $V_D := \|D\|_2^2$, the normalized squared norm is $\chi_{32}^{\prime 2}(V_D/\sigma_P^2)$. Hence the eight-block union sum is

$$U_{\text{mix}} := 8 \sum_v \Pr[V_D = v] \Pr[\chi_{32}^{\prime 2}(v/\sigma_P^2) \geq r_{\text{BDD}}^2/\sigma_P^2]. \quad (28)$$

The exact mass function of V_D is computed without vector enumeration by

$$p_{j+1}(s) = \sum_z \Pr[D_1 = z] p_j(s - z^2), \quad 0 \leq j < 32.$$

Under the substitution in equation (27), this finite mixture evaluates each block tail exactly under the explicitly defined reference law Q_{BW} . It is not an exact finite-parameter DFR for the implemented BW-KEM residual. The factor eight remains a blockwise union sum, not the exact probability that at least one block exits.

For comparison, the original estimator evaluates the same reference event through the raw Chernoff union expression

$$U_{\text{raw}} := 8 \inf_{\theta > 0} \exp\{32\kappa(\theta) - \theta r_{\text{BDD}}^2\}, \quad \kappa(\theta) = \log \mathbb{E}[e^{\theta(X_{0,1} + D_1)^2}], \quad (29)$$

where $X_{0,1} \sim \mathcal{N}(0, \sigma_P^2)$; write $U_{\text{Ch}} := \min\{1, U_{\text{raw}}\}$ for the valid clipped Chernoff bound. For $r_n^2 = na$ and $\kappa'(\hat{\theta}) = a$, the nonlattice exact large-deviation expansion [1] gives

$$\Pr \left[\sum_{i=1}^n (X_{0,i} + D_i)^2 \geq na \right] = \frac{\exp\{-nI(a)\}}{\hat{\theta} \sqrt{2\pi n \kappa''(\hat{\theta})}} (1 + o(1)), \quad I(a) = \hat{\theta}a - \kappa(\hat{\theta}).$$

The optimized Chernoff value omits the displayed prefactor. Its multiplicative slack is therefore $\Theta(\sqrt{n})$, or $\frac{1}{2} \log_2 n + O(1)$ bits, under these conditions. The table below evaluates the finite $n = 32$ case directly.

At all six published rows, $\log_2 U_{\text{mix}}$ is 6.537–7.621 bits below $\log_2 U_{\text{raw}}$. The event, Gaussian product reference, terminal-noise distribution, and aggregation over eight blocks are identical in table 1, so the difference comes from the tail-evaluation method.

Table 1. BW-KEM BDD-radius calculations at the published parameter rows. Both columns fix (A, Q, G) and differ only in the method used to evaluate the tail probability: the original Chernoff expression or the conditional noncentral mixture.

Set	original Chernoff $\log_2 U_{\text{raw}}$	conditional mixture $\log_2 U_{\text{mix}}$	raw – mix (bits)
BW-KEM-512	−190.370	−196.907	6.537
BW-KEM-768	−226.792	−233.528	6.736
BW-KEM-1024	−222.092	−229.180	7.088
BW-KEM-512-s	−152.138	−159.259	7.121
BW-KEM-768-s	−171.152	−178.773	7.621
BW-KEM-1024-s	−209.942	−217.018	7.077

Table 2. BW-KEM comparison at $(d_u, d_v) = (8, 3)$. Simulation records whether at least one block exits the BDD radius. The analytic columns sum over the eight blocks under Q_{BW} ; $U_{\text{Ch}} = \min\{1, U_{\text{raw}}\}$ is the valid clipped Chernoff bound. The final column gives observed absolute log-differences from simulation and therefore combines the choice of residual law, aggregation, sampling, and numerical discrepancies.

(η_1, η_2)	trials	count	BDD sim. $\log_2 p$	raw Ch. $\log_2 U_{\text{raw}}$	valid Ch. $\log_2 U_{\text{Ch}}$	mixture $\log_2 U_{\text{mix}}$	absolute difference valid Ch./mix.
(4, 3)	2^{30}	2,169	−18.917	−15.529	−15.529	−20.033	3.388/1.116
(5, 4)	2^{30}	311,321	−11.752	−8.546	−8.546	−12.381	3.206/0.629
(6, 5)	2^{30}	8,111,083	−7.049	−3.917	−3.917	−7.381	3.131/0.333
(7, 6)	2^{28}	18,952,779	−3.824	−0.653	−0.653	−4.003	3.171/0.179
(8, 7)	2^{25}	9,426,877	−1.832	1.753	0.000	−1.684	1.832/0.148

Across the measurable grid, $\log_2 U_{\text{raw}}$ is 3.131–3.585 bits above the simulated $\log_2 p$. Because the last raw sum exceeds one, the largest difference for the valid clipped Chernoff bound is instead 3.388 bits; at (8, 7) its difference is 1.832 bits. The conditional mixture differs by at most 1.116 bits, with the absolute discrepancy decreasing to 0.148 bit at the highest-noise point. An independent run of 2^{32} trials at (7, 6) gives BDD log-probability −3.850, within 0.03 bit of the grid value. These diagnostic differences include changes in the residual law and aggregation map. Table 1 instead compares the two tail-probability methods under the same reference model.

5 CTRU and DTRU: E_8 and Paired- E_8 Decoding

CTRU decodes one eight-dimensional E_8 block, whereas DTRU jointly decodes a paired- E_8 observation block in 16 dimensions. We first compare the two CTRU calculations and their target events, then make the corresponding comparison for DTRU’s joint decoder.

5.1 CTRU analyses and the implemented message event

We compare two formally distinct analyses. The packing analysis developed in this paper derives position-dependent coordinate variances from the trinomial ring, retains the packing/BDD event, and applies a determinant-matched or Welch–Satterthwaite block tail. Zhou et al. [28] instead analyze 224 message-changing directions under an isotropic Gaussian law. The covariance expressions and facet enumeration used below follow that paper. The additional rows in table 3 pair each event with the same covariance model and specify the tail procedure used for that event.

CTRU residual and message invariance. The CTRU encryption equation and its lifted residual X are defined in equation (3). The end-to-end simulator also performs key generation, exact ring multiplication, compression, centered reduction, and message decoding.

The zero-message simulation is exact because the residual is invariant under translation by the encoded message. After compression the message component is $(q_2/2)s$ for a binary codeword $s \in \{0, 1\}^n$. The identity below is first taken in the integer quotient ring $\mathbb{Z}[x]/(x^n - x^{n/2} + 1)$ and only then reduced coefficientwise modulo q_2 :

$$\frac{q_2}{2}fs = q_2(f's) + \frac{q_2}{2}s \equiv \frac{q_2}{2}s \pmod{q_2}.$$

Thus every sample has the same residual after translation by any message. The implementation performs the integer products in a type wide enough to exclude overflow before centered modular reduction.

Gaussian reference with matched covariance. To compare the packing and facet events under the same residual law, we apply the covariance construction of Zhou et al. [28]. Let $M(a)$ be the matrix of multiplication by a in the CTRU ring and write the fixed-key proxy residual as $Z_k = M(g)r + cM(f)\varepsilon_k(r)$, where $c = q/q_2$. With $D_r = \text{Cov}(r)$, $D_{\varepsilon,k} = \text{Cov}(\varepsilon_k(r))$, and $\Gamma_k = \text{Cov}(r, \varepsilon_k(r))$, direct expansion gives

$$\begin{aligned} \Sigma_k &= M(g)D_rM(g)^T + c^2M(f)D_{\varepsilon,k}M(f)^T \\ &\quad + c\{M(g)\Gamma_kM(f)^T + M(f)\Gamma_k^T M(g)^T\}. \end{aligned} \tag{30}$$

The fixed-key Gaussian reference sets $D_r = \sigma_r^2 I$, $D_{\varepsilon,k} = \sigma_\varepsilon^2 I$, and $\Gamma_k = 0$; its covariance is therefore

$$\Sigma_k^{(0)} = \sigma_r^2 M(g)M(g)^T + c^2 \sigma_\varepsilon^2 M(f)M(f)^T. \tag{31}$$

At the measurable CTRU point, the mean relative Frobenius error is 3.30% for $\Sigma_k^{(0)}$ and 13.01% for a single reference computed with the covariance averaged over 16 independent keys. The latter is distinct from averaging key-conditioned tail probabilities as in equation (25).

Implemented failure and conservative message-changing ties. Zhou et al. identify the 224 message-changing minimum directions and exclude 16 shaping directions [28]. We express this geometry in the implemented coordinates and separate a closed, conservative facet event from the deterministic decoder event measured by the end-to-end PKE experiment.

Let $C_8 \subset \{0, 1\}^8$ be the extended $[8, 4, 4]$ Hamming code defined in section 2. Its weight enumerator is $1 + 14z^4 + z^8$. Let $\mathcal{H}$ be the 14 supports of its weight-four words. A centered received block is $x \in [-q/2, q/2]^8$. Let $\mathcal{F}_q^{\text{impl}}$ be the set of blocks that fail under the implemented deterministic tie rule. The following theorem defines a closed, conservative event that includes every message-changing tie.

Theorem 5 (Conservative message-changing facet event). *For the zero class,*

$$\mathcal{F}_q^{\text{cons}} := \left\{ x \in [-q/2, q/2]^8 : \exists H \in \mathcal{H} : \sum_{i \in H} |x_i| \geq q \right\}. \quad (32)$$

Equivalently, it is the union of $14 \cdot 2^4 = 224$ halfspaces

$$\mathcal{F}_q^{\text{cons}} = \bigcup_{a \in \mathcal{A}} \{x \in [-q/2, q/2]^8 : a^T x \geq q\}, \quad (33)$$

where $\mathcal{A}$ contains every vector supported on some $H \in \mathcal{H}$ with its four nonzero entries in $\{\pm 1\}$. Moreover,

$$\{x \in [-q/2, q/2]^8 : \exists a \in \mathcal{A} : a^T x > q\} \subseteq \mathcal{F}_q^{\text{impl}} \subseteq \mathcal{F}_q^{\text{cons}}.$$

Thus the two events can differ only on a message-changing tie hyperplane.

Proof. The fine lattice is $(q/2)(C_8 + 2\mathbb{Z}^8)$ inside the shaping lattice $q\mathbb{Z}^8$. A nonzero weight-four codeword with support H has nearest signed representatives $(q/2)a$; comparing squared distances with zero gives $a^T x \geq q$, and maximizing over signs gives equation (32). Strict inequality makes a different message closer, while equality is precisely a message-changing tie. Weight-eight representatives add no facets: summing their signed coordinates over the 14 weight-four supports counts each coordinate seven times, so any weight-eight inequality implies at least one of the 224 weight-four inequalities.

Gaussian reference calculations may use $\mathcal{F}_q^{\text{cons}}$ without a tie correction because the boundary has probability zero. End-to-end PKE experiments record $\mathcal{F}_q^{\text{impl}}$ directly.

The two-dimensional schematic in figure 1(a) illustrates this inclusion; the proof below establishes it for the eight-dimensional decoder.

Let $\mathcal{B}_1 := \{x \in [-q/2, q/2]^8 : \|x\|_2 \geq q/2\}$ and lift both $\mathcal{F}_q^{\text{cons}}$ and $\mathcal{B}_1$ to a full polynomial by taking a union over its $n/8$ blocks.

Theorem 6 (Packing sphere versus message failure). *For the conservative facet event,*

$$\mathcal{F}_q^{\text{impl}} \subseteq \mathcal{F}_q^{\text{cons}} \subsetneq \mathcal{B}_1.$$

Table 3. Probability objects compared in the CTRU study. The first two rows are the packing analysis with position-dependent variances developed here and the facet-based analysis of Zhou et al. [28]; the remaining rows evaluate the ball and facet events under the same covariance model.

Analysis	Event A	Law Q	Procedure T	Aggregation G
Packing (this work)	packing-ball exit $\mathcal{B}_1$	diagonal Gaussian; position- dependent variances	determinant-matched isotropic approximation or Satterthwaite	block union; one reference using averaged covariance
Facet (Zhou et al.)	224 message- changing facets	isotropic Gaussian with the scalar variance of Zhou et al.	224-term union bound	block union; one reference using averaged covariance
Ball under matched covariance	packing-ball exit $\mathcal{B}_1$	Gaussian with the covariance of Zhou et al.	quadratic-form inversion or saddlepoint	72-block union bound
Facet under matched covariance	conservative facet event $\mathcal{F}_q^{\text{cons}}$	Gaussian with the same covariance	finite Bonferroni or conditional mixture	72-block union bound
End-to- end PKE	implemented message event $\mathcal{F}_q^{\text{impl}}$	implemented discrete residual	direct simulation	mean of per-key any-block failure rates

Proof. If $x \in \mathcal{F}_q^{\text{cons}}$, some four-coordinate support has ℓ_1 norm at least q . Cauchy–Schwarz gives $\|x\|_2 \geq q/2$, proving inclusion. It is strict: take all eight absolute coordinates equal to an a satisfying $q/(4\sqrt{2}) < a < q/4$. Then $x \in \mathcal{B}_1$, while every weight-four support sum is below q .

5.2 Separating CTRU event and distribution effects

The first two rows below give the original packing calculation and the published facet calculation. Four further rows combine each event with each residual model. Moving from the original packing calculation to the ball event under the matched covariance model changes both the covariance construction and the tail method, so that difference is reported only as a combined discrepancy.

Ball tails as Gaussian quadratic forms. For one block let $X \sim \mathcal{N}(0, \Sigma)$, with eigenvalues $\lambda_1, \dots, \lambda_8 > 0$. Orthogonal diagonalization yields

$$\|X\|_2^2 \stackrel{d}{=} \sum_{i=1}^8 \lambda_i Z_i^2, \quad Z_i \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 1). \quad (34)$$

Imhof or Davies inversion gives a numerical reference [15,11]. At very small tail probabilities direct inversion can lose relative accuracy through cancellation; our implementation returns a warning whenever the quadrature error is not small relative to the tail.

The determinant-matching approximation used by the CTRU packing analysis replaces all eigenvalues by $g = (\prod_i \lambda_i)^{1/8}$. Its model mean is $8g$, which is at

most $\sum_i \lambda_i$ by the arithmetic–geometric mean inequality and is strictly smaller under heteroscedasticity. It is not a probability bound. Satterthwaite instead sets

$$a = \frac{\sum_i \lambda_i^2}{\sum_i \lambda_i}, \quad \nu = \frac{(\sum_i \lambda_i)^2}{\sum_i \lambda_i^2}, \quad (35)$$

so that $a\chi_\nu^2$ matches the first two moments [26]. We also use the Lugannani–Rice saddlepoint tail [21].

Gaussian facet unions. For $a \in \mathcal{A}$, let $H_a := \{x \in \mathbb{R}^8 : a^T x \geq q\}$ and

$$p_a = \bar{\Phi}\left(\frac{q}{\sqrt{a^T \Sigma a}}\right), \quad S = \sum_{a \in \mathcal{A}} p_a. \quad (36)$$

The analysis of Zhou et al. uses the isotropic specialization of S and then a block union bound. To distinguish marginal-bound looseness from covariance effects, we compute the full Gaussian facet-union probability under the same Σ .

As a numerical check, we use the union sampler of Owen et al. [25]. It selects facet a with probability p_a/S , samples the Gaussian conditional on H_a , and returns $W = S/N(X)$, where $N(X)$ counts the facets crossed by the draw. The estimator is unbiased for the facet union and is bounded by S . For CTRU, one draw requires one matrix–vector product and 224 threshold comparisons, irrespective of the inverse tail probability. Unbiasedness follows by expanding over the selected facet: $N(X)$ counts the conditional components that can generate the same union sample. The finite bracket used for the main geometric conclusion does not depend on this sampler.

From block events to ciphertext failure. Let F_b be the implemented message-failure event in block b and let $N_F = \sum_{b=1}^{n/8} \mathbf{1}_{F_b}$. The ciphertext event is $F_{\text{ct}} = \{N_F \geq 1\}$, irrespective of cross-block dependence. Hence

$$\sum_b Q(F_b) - Q(F_{\text{ct}}) = \mathbb{E}_Q[(N_F - 1)_+], \quad (37)$$

and consequently

$$\begin{aligned} \max_b Q(F_b) &\leq Q(F_{\text{ct}}) \leq \min\left\{1, \sum_b Q(F_b)\right\}, \\ 0 &\leq \sum_b Q(F_b) - Q(F_{\text{ct}}) \leq \sum_{b < b'} Q(F_b \cap F_{b'}). \end{aligned} \quad (38)$$

The first equality follows from $N_F - \mathbf{1}\{N_F \geq 1\} = (N_F - 1)_+$; the final inequality uses $(N_F - 1)_+ \leq N_F(N_F - 1)/2$. Equation (38) identifies the block-union overlap error, and pairwise intersections depend on the cross-block matrices $P_b \Sigma_{\text{full}} P_{b'}^T$. Independence gives $1 - \prod_b (1 - Q(F_b))$ only as a separate modeling assumption.

The same union sampler also applies to the full ciphertext event by indexing facets with pairs (b, a) and using the full covariance. The CTRU tables use the

block union bound from both CTRU analyses; they must therefore be read as blockwise union estimates for the stated events, not as exact cross-block Gaussian probabilities.

5.3 End-to-end CTRU evaluation

At diagnostic points where both events are measurable, we evaluate the ball and facet events under each residual model and compare the calculations with implemented end-to-end experiments over independently generated keys. Parameter screening, analytic comparisons, and implementation checks are reported separately.

Experimental objectives and reproducibility. The evaluation measures the looseness of the 224-term union bound and compares both analytic references with paired event frequencies from independently generated keys where direct failures are measurable. Event geometry, residual law, and numerical procedure are changed one at a time where possible; a comparison that changes several fields is reported as a combined discrepancy.

Run manifests, seeds, source hashes, and integrity checks are included in the artifact.

Fixed measurable CTRU diagnostic point. We use the fixed diagnostic point $(n, q, q_2, \eta_1, \eta_2) = (576, 3457, 512, 7, 8)$. The ring dimension n , large modulus q , encoder, and decoder are unchanged. We reduce the compression modulus from $q_2 = 1024$ to $q_2 = 512$ and increase the noise parameters to make both packing exits and implemented message failures measurable. This parameter change preserves the event geometry while making the residual-model and event discrepancies directly measurable. The packing analysis with position-dependent variances gives -20.177 with its determinant-matched isotropic approximation (-20.172 with Satterthwaite), whereas the facet analysis of Zhou et al. gives -27.342 . All results at this point are reported as diagnostic comparisons; calculations at the published parameters are presented separately.

For the published CTRU rows, the packing calculation with position-dependent variances uses the distributions specified by the construction: (B_4, B_4) for CTRU-576, (B_3, B_3) for CTRU-768, and (B_2, B_2) for CTRU-1024. With $q = 3457$ and $q_2 = 1024$, the corresponding estimates for the packing certificate are -190.812 , -210.294 , and -265.950 bits.

Gaussian comparison under matched covariance. We apply the covariance model of Zhou et al. to equation (3), then compute both events under that law. These are the matched-covariance rows in table 3; the analysis in this paper and the analysis of Zhou et al. use the same cross-block union bound in their respective rows.

Before attributing model differences, we checked the Gaussian ball calculation under the matched covariance model with three independent numerical procedures. Imhof inversion was implemented directly for the survival probability to

avoid subtracting a value close to one. Adaptive quadrature supplies an absolute error bound, which is propagated through the block union and converted to bits of failure exponent. Lugannani–Rice differs from Imhof by 0.006 bit and Satterthwaite by less than 0.001 bit; the propagated Imhof error bound is below 0.001 bit.

The Gaussian facet union can also be bounded deterministically, independently of the Monte Carlo estimator. Applying theorem 2 blockwise, with exact univariate Gaussian projections and bivariate Gaussian intersection bounds under each of the 72 block covariance matrices, gives

$$-28.205 \leq \log_2 p_{\text{facet}} \leq -28.198.$$

The interval width is approximately 0.007 bit. The dominant block is block 36; its covariance is exactly a scalar multiple of the identity, and it is dominant for both the ball and facet events. The aggregate Imhof ball value -20.150 therefore gives a certified geometric gap of 8.048 – 8.055 bits. The conditional-mixture estimate -28.199 lies inside the deterministic interval and is an independent numerical check.

Table 4. Gaussian-reference intervals for the matched CTRU-576 events under the same covariance law. The ball entry is an Imhof estimate with propagated numerical error below 0.001 bit; the facet and ratio entries are derived intervals.

Event or comparison	value or interval	error control	evaluation
$\log_2 p_{\text{ball}}$	-20.150	< 0.001 bit	Imhof
$\log_2 p_{\text{message facet}}$	$[-28.205, -28.198]$	deterministic	Bonferroni
$\log_2(p_{\text{ball}}/p_{\text{facet}})$	$[8.048, 8.055]$	includes Imhof error	derived interval

Table 5. Original CTRU-576 calculations and blockwise union sums for the four combinations of events and residual models, reported as $-\log_2 u$.

	original packing	published facet	v_{00} ball/cov.	v_{10} facet/cov.	v_{01} ball/iso.	v_{11} facet/iso.
$-\log_2 u$	20.177	27.342	20.144	28.199	19.311	27.342

For the four union sums in table 5, equation (9) assigns 8.043 bits to the event change, averaged over the two residual laws, and -0.844 bit to the residual-law change, averaged over the two events. The two changes sum to 7.199 bits, equal to $v_{11} - v_{00}$. The 7.165-bit difference between the original packing and published facet calculations is a distinct combined comparison because moving from the original packing calculation to the ball event under the matched covariance model changes both the covariance and the tail method. Across 16 fixed-key

calculations, the event and residual-law changes have medians 7.881 and -0.170 bits; the latter has interquartile range (IQR) $[-2.271, 0.491]$.

Confirmatory experiment at $n = 576$. The end-to-end run at the main diagnostic point uses 8,192 independently generated keys and 2^{17} ciphertexts per key, for 2^{30} ciphertexts in total. For each generated key, every ciphertext is processed through compression, decompression, and decoding, and the implemented failure and proxy events are recorded together.

Table 6. Final CTRU diagnostic experiment at $(576, 3457, 512, 7, 8)$. The pooled-binomial reference intervals are Clopper–Pearson intervals transformed to $\log_2 p$; their exact binomial interpretation requires a common iid Bernoulli success probability and is not an unconditional coverage statement for this heterogeneous-key experiment. The final column gives 20,000-replicate percentile cluster bootstrap intervals obtained by resampling 8,192 whole keys; the estimator is the unweighted mean of the per-key failure rates.

Event	Count	$\log_2 \hat{p}$	pooled binomial reference 95% interval	key bootstrap 95% interval
Implemented failure	26	-25.300	$[-25.914, -24.748]$	$[-25.913, -24.830]$
Packing-ball exit	2,396	-18.774	$[-18.832, -18.716]$	$[-18.852, -18.697]$
Packing-only	2,370	-18.789	—	—
Failure-only outside packing	0	—	—	zero by event inclusion

The measured packing probability is $2^{1.403}$ times the geometric prediction $2^{-20.177}$ from the packing analysis with position-dependent variances. Because the packing event is fixed while the residual law, aggregation map G , and order of key averaging differ between the analytic calculation and end-to-end experiment, this 1.403-bit difference is a combined model discrepancy with the event held fixed. It cannot be attributed to the residual distribution alone. In the paired data, all 26 implemented failures lie in the packing event, whereas 2,370 of the 2,396 packing exits decode correctly. The packing-exit probability is therefore about 92 times the implemented-failure probability, a 6.526-bit log-ratio. A separate 50,000-replicate percentile bootstrap resamples the 8,192 paired key-count vectors and computes the log-ratio within each resample; its 95% interval is $[6.046, 7.166]$ bits. Thus the measured event separation remains larger than six bits after accounting for between-key sampling variation. The 26 implemented failures occur on 26 distinct keys. All 50,000 log-ratio bootstrap resamples have a nonzero failure denominator. The percentile bootstrap does not provide a finite-sample coverage guarantee. The cluster bootstrap interval is the primary uncertainty interval for the observed implemented-failure rate. The pooled Clopper–Pearson interval is reported only as a binomial reference interval

and is not an exact unconditional interval for the key-averaged DFR under the observed key heterogeneity.

The events are also heterogeneous across keys: 1,673 of the 8,192 keys have at least one packing exit, and the ten percent most active keys account for approximately 64.4% of all packing exits. This is why the cluster bootstrap interval over keys, rather than a pooled ciphertext interval, is the primary uncertainty summary.

The cluster bootstrap interval for implemented failure excludes the facet estimate reported by Zhou et al. [28]: $2^{-27.342}$ is a factor of $2^{2.043}$ below the observation, so its failure exponent under the Gaussian reference is 2.043 bits larger than the observed implemented-failure exponent. By comparison, the position-dependent calculation differs from the measured probability of its packing event by 1.403 bits. The paired experiment records the two event frequencies on common residuals and measures the six-bit gap between certificate exits and implemented failures. Relative to the packing calculation with its stated law and aggregation, the 1.403-bit value remains a combined model discrepancy with the event held fixed; it is not assigned to a single causal source. The geometric inclusion itself is supplied by theorem 6, not by the zero observed failure-only count.

5.4 DTRU: paired- E_8 decoding in 16 dimensions

DTRU repeats each E_8 word across two octets and jointly decodes the resulting 16-dimensional observation, as described in section 2. Its residual Err is given in equation (4). Let I_j select the two octets decoded in block j . The BDD certificate in [29] exits when

$$C_j = \{\|\text{Err}_{I_j}\|_{q,2}^2 \geq (q-1)^2/2\}. \quad (39)$$

The minimum-distance argument in Theorem 1 of [29] proves $F_j \subseteq C_j$, where F_j is a message error. Thus, the BDD-exit probability upper-bounds the block message-failure probability, while the BDD ball is a strict subset of the correct paired- E_8 decoding cell.

The scalar calculation in [29] models each paired block by $\mathcal{N}(0, \sigma_{\text{Err}}^2 I_{16})$ and applies

$$\hat{p}_C^{\text{scalar}} = \frac{n}{16} \Pr\left[\chi_{16}^2 \geq \frac{(q-1)^2}{2\sigma_{\text{Err}}^2}\right]. \quad (40)$$

The DTRU-512 row has $q_2 = 2^{10}$ and reports failure exponent $E = 134.26$. The two octets are correlated by both message repetition and reduction modulo $x^n - x^{n/2} + 1$. We retain this dependence before evaluating the BDD tail. Let $z = e + \varepsilon$, let v_a denote the coefficient variance of a , let S_j be the exact covariance-count matrix of a product of two centered unit-variance polynomials restricted to I_j , and define $M_j = \mathbb{E}[(f'w)_{I_j}(f'w)_{I_j}^T]/v_{f'}$ for a uniform message. Under the independence assumptions on the primitive distributions used by the Gaussian reference, the covariance matrix for the paired block is

$$\Sigma_j = (v_g v_r + 4v_{f'} v_z) S_j + v_z I_{16} + v_{f'} M_j. \quad (41)$$

Both S_j and M_j are computed exactly from the ring reduction and the 16 codewords. If $\lambda_{j,1}, \dots, \lambda_{j,16}$ are the eigenvalues of Σ_j , the Gaussian BDD reference using the full block covariance sums the generalized chi-square tails $\Pr[\sum_i \lambda_{j,i} \chi_{1,i}^2 \geq (q-1)^2/2]$ over the $n/16$ blocks. Imhof inversion is used for the grid below; Lugannani–Rice differs by less than 0.003 bit at every row.

We also evaluate a nearest-facet event under the same Gaussian law with full block covariance. For a paired block, let $\mathcal{V}_4$ be the fourteen weight-four E_8 codewords and let $\mathcal{A}(v)$ be the centered torus representatives of the repeated word (v, v) that attain the smallest Euclidean norm in our enumeration. Across the 14 choices of v , this gives 3,584 distinct nearest halfspaces. Their finite union is

$$\mathcal{F}_j^G := \bigcup_{v \in \mathcal{V}_4} \bigcup_{a \in \mathcal{A}(v)} \{z \in \mathbb{R}^{16} : a^T z \geq \|a\|_2^2/2\}, \quad Z_j \sim \mathcal{N}(0, \Sigma_j). \quad (42)$$

The single weight-eight codeword contributes a farther halfspace family, whose one-dimensional Gaussian union bound is retained as an additive remainder. Because the enumeration does not establish that these families exhaust every farther Algorithm 2 decision region, the result is a nearest-facet Gaussian diagnostic, not a correctness bound. We evaluate the nearest-facet union with a conditional-mixture estimator. A facet is selected with probability proportional to its Gaussian marginal, and the sample is drawn conditionally on that facet. If H nearest facets contain the sample, its contribution is the first-order facet sum divided by H . The identity evaluates the finite union without direct hits at the target tail.

The end-to-end experiment uses DTRU-512 with $(n, q) = (512, 3457)$ and $(k_g, k_f, k_r, k_e) = (8, 4, 6, 3)$. It changes only q_2 from its published value so that both events can be measured. Each row samples 256 independent keys. For every key, the experiment applies the published PKE encryption map, the half-up integer compression rule, secret multiplication, and Algorithm 2 to independent uniform messages. A reference implementation uses schoolbook integer convolution, explicit tricyclotomic reduction, and exhaustive evaluation of all 16 decoder candidates. Regression tests compare its received words, residuals, decisions, and tie flags sample by sample with the FFT implementation using the certificate check in the large runs. Write $E_C := E(\hat{p}_C)$ and $E_F := E(\hat{p}_F)$. We use E_Σ for the Gaussian BDD exponent computed with the full block covariance and E_Σ^{near} for the corresponding finite nearest-facet Gaussian exponent.

The implemented decoder failure is consistently rarer than the BDD exit: $E_F - E_C$ increases from 5.85 to 7.63 bits. At $q_2 = 304$, a 50,000-replicate percentile bootstrap of paired key-count vectors gives $[7.48, 7.82]$ bits. All resamples contain nonzero failures, and no minimum-cost tie was observed. The scalar reconstruction uses the maximum coordinate variance $\sigma_{\text{Err}}^2 = \max_i \text{Var}(\text{Err}_i)$ and differs from the measured BDD exponent by 1.60–7.02 bits. Using Σ_j reduces this error to 0.30–1.39 bits. The nearest-facet Gaussian reference differs from the measured implemented-failure exponent by 0.70–2.25 bits across the same six points. On the six-point grid, $E_F = 0.878E_\Sigma^{\text{near}} + 0.770$ has a maximum leave-one-out interpolation error of 0.203 bit. This leave-one-row-out interpolation may

Table 7. DTRU-512 diagnostic-grid comparison. C/F gives measured ciphertext-level counts when any paired block exits the BDD certificate or fails Algorithm 2, respectively. E_Σ and E_Σ^{near} are Gaussian-reference calculations; the latter is the finite nearest-facet diagnostic under equation (41); its Monte Carlo numerical error is below 0.02 bit in every row.

q_2	trials	C/F counts	E_C	E_Σ	E_Σ^{near} (facet)	E_F	$E_F - E_C$
224	2^{26}	2,578,430 / 44,785	4.702	5.001	11.247	10.549	5.847
240	2^{26}	758,360 / 9,446	6.467	6.928	13.684	12.795	6.327
256	2^{27}	398,050 / 3,839	8.397	8.969	16.206	15.093	6.696
272	2^{28}	206,790 / 1,595	10.342	11.161	18.862	17.361	7.018
288	2^{29}	128,783 / 770	12.025	13.109	21.189	19.411	7.386
304	2^{30}	63,043 / 319	14.056	15.442	23.932	21.683	7.627

use points on both sides of the omitted row. These comparisons change both the scalar Gaussian model to a paired-block covariance model and the BDD event to the implemented decoder event, so the improvement is not assigned causally to either change alone.

The same covariance structure is present at the published $q_2 = 1024$ row. An implementation check over 16 keys and 2^{20} ciphertexts, comprising 2^{25} paired blocks, gives mean coordinate variance 23,735, compared with the analytic value 23,823, and a 6.08% relative Frobenius deviation between the two covariance matrices. The empirical eigenvalues span 8,167 to 57,226, while the analytic range is 8,423 to 51,388; the empirical condition number is 7.0, in contrast to one for the isotropic scalar model. Using the empirical covariance of the implemented block residuals as the Gaussian input gives a BDD exponent of 69.94. The analytic block covariance from equation (41) gives 76.29. Both are Gaussian extrapolations at the published parameters rather than finite-parameter DFR measurements; the published scalar exponent is 134.26. Applying equation (42) to the analytic block covariances gives a Gaussian-reference nearest-facet exponent of 90.999 bits at $q_2 = 1024$; the farther weight-eight remainder is below 2^{-171} . These deep-tail values depend on the two stated second-order covariance inputs; they do not estimate the implemented discrete failure probability at the published row.

An independent run of 2^{32} trials at $q_2 = 288$ gives $E_C = 11.963$ and $E_F = 19.315$, within 0.10 bit of their grid values. Sequential held-out prediction uses the Gaussian BDD exponent computed from the full block covariance as its predictor and has at most 0.35-bit error. The first target is $q_2 = 256$; an affine least-squares fit then uses all preceding, shallower rows. For targets 256, 272, 288, 304, the absolute errors are respectively 0.079, 0.255, 0.179, and 0.344 bit. This is the forward out-of-sample prediction test across the diagnostic grid, distinct from the ordinary leave-one-out interpolation above. Transfer to $q_2 = 1024$ requires control of the residual CGF over the intervening saddlepoint range.

6 Prime-Degree Variants and Parity-Separated Blocks

The prime-degree ring and the published CTRU-Prime and DTRU-Prime block layouts are specified in section 2. The remaining DTRU-Prime parameters are $(n, q, q_2) = (1087, 2017, 1024)$, $(k_g, k_f) = (2, 1)$, and $(k_r, k_e) = (2, 1)$ [29].

Zhou et al. showed that treating adjacent coefficients in the prime-degree quotient ring as independent can understate their Gaussian BDD reference [28]. We analyze the corresponding second-order covariance between adjacent coefficients at the encoding layer. For independent centered coefficient sequences a, b with variances v_a, v_b , let $t_s = \sum_{i+j=s} a_i b_j$ be their ordinary convolution and write $c = ab$ in the quotient ring. Reduction by $x^n = x + 1$ gives

$$c_k = t_k + \mathbf{1}\{k \leq n-2\}t_{n+k} + \mathbf{1}\{k \geq 1\}t_{n+k-1}.$$

The indicator factors are essential at $k = 0$ and $k = n-1$. The ordinary-convolution terms used by c_k and c_ℓ are disjoint unless $|k - \ell| = 1$. Counting the shared terms yields

$$\begin{aligned} \text{Var}(c_0) &= nv_a v_b, \\ \text{Var}(c_k) &= (2n - k)v_a v_b \quad (1 \leq k < n), \\ \text{Cov}(c_k, c_{k+1}) &= (n - k - 1)v_a v_b \quad (0 \leq k < n-1), \end{aligned} \tag{43}$$

and every other off-diagonal covariance is zero. The same count matrix applies to each independent centered product term in the residual, scaled by its single-layer variance product. A coefficientwise term contributes only to the diagonal.

The covariance calculation gives the adjacent terms, but zero covariance by itself does not prove independence. The next result gives the required joint limit under the bounded coefficient distributions used in the schemes.

Theorem 7 (Joint Gaussian limit in the prime-degree quotient ring).

For $j = 0, \dots, \ell-1$, let $f^{[j]}, g^{[j]} \in \mathbb{R}[x]/(x^n - x - 1)$. Suppose all coefficients in these two families are mutually independent and the two coefficient laws are centered, have bounded support, and have positive variances σ_f^2 and σ_g^2 , respectively. Set $h = \sum_j f^{[j]} g^{[j]}$ and define

$$\nu_c = \begin{cases} n, & c = 0, \\ 2n - c, & 1 \leq c < n. \end{cases}$$

For fixed m , choose indices $c_1, \dots, c_m$, possibly depending on n , such that $|c_a - c_b| \geq 2$ whenever $a \neq b$. Then

$$\left(\frac{h_{c_1}}{\sigma_f \sigma_g \sqrt{\ell \nu_{c_1}}}, \dots, \frac{h_{c_m}}{\sigma_f \sigma_g \sqrt{\ell \nu_{c_m}}} \right) \xrightarrow{d} \mathcal{N}(0, I_m) \quad (\ell n \rightarrow \infty). \tag{44}$$

Proof (Sketch). Represent h_c by its 0–1 convolution matrix A_c . Nonadjacent output indices have disjoint matrix support. After variance normalization, a fixed

Cramér–Wold projection is a conditional triangular array in the g -coefficients. Bounded support gives Lindeberg, while a bounded-degree quadratic-form estimate makes the conditional variance converge in probability. Its variance is $O((\ell n)^{-1})$. Conditional Lindeberg–Feller and Cramér–Wold prove equation (44). The matrix counts and variance estimate are proved in section A.2.

The spacing condition is sharp for this argument. From (43), the correlation between variance-normalized adjacent coefficients is

$$\frac{n - 1 - c}{\sqrt{\nu_c \nu_{c+1}}}.$$

For fixed $c \geq 1$ it tends to $1/2$ as $n \rightarrow \infty$, and for $c = 0$ it tends to $1/\sqrt{2}$. Adjacent coefficients therefore do not have an asymptotically independent standard-normal limit.

Equation (43) motivates a different codeword layout. At encoding, form blocks in the public order

$$\pi = (0, 2, 4, \dots, 1, 3, 5, \dots),$$

and scatter their coordinates to the listed polynomial positions; at decoding, gather in the same order before applying the block decoder. We call this the parity-separated layout. Every eight-coordinate block then lies inside one parity class for the 752-position CTRU layout, so the adjacent covariance terms identified in (43) are zero within the block’s second-order product covariance matrix. DTRU-Prime applies the same ordering separately to two 536-position decoder streams. Under the explicit floor convention for the odd dimension, the streams occupy positions $0, \dots, 535$ and $543, \dots, 1078$. A parity-boundary octet joins the largest even indices to the smallest odd indices and contains no consecutive pair; the two coordinates paired across streams differ by 543. Thus the identified adjacent terms within this block’s second-order product covariance matrix are also zero. The permutation leaves the ring arithmetic and message cardinality unchanged but changes the public codeword placement. We therefore analyze it as a variant of the codeword layout; integration into a scheme requires separate analyses of security, wire format, and integration into the scheme. For each fixed within-stream parity block, the relevant product coordinates satisfy the spacing condition of theorem 7 and hence have the stated joint Gaussian limit under the theorem’s assumptions. The layout removes adjacent pairs within each decoder block; it does not remove all finite-parameter or cross-block statistical dependence.

Finite Gaussian BDD references. For CTRU-Prime-761, the covariance is reconstructed directly from $(n, q, q_2, \eta_1, \eta_2) = (761, 4591, 1024, 2, 2)$. We diagonalize each 8×8 block, evaluate the resulting weighted chi-square tail with a modified saddlepoint approximation, and sum the 94 block tails. We use the same $q/2$ BDD radius as the correlation analysis of Zhou et al. If v_ε is the exact

unit-rounding variance, the covariance after encoding is

$$\Sigma_{ii} = d_i \mu + \alpha, \quad \Sigma_{i,i+1} = a_i \mu, \quad \mu = \frac{\eta_1 \eta_2}{4} + \left(\frac{2q}{q_2}\right)^2 v_\varepsilon \frac{\eta_1}{2}, \quad \alpha = \left(\frac{q}{q_2}\right)^2 v_\varepsilon,$$

where d_i and a_i are the diagonal and adjacent counts in (43). These formulas define the covariance used in the reported calculation.

The DTRU-Prime row has a narrower status: it is a calibrated second-order Gaussian sensitivity analysis, not an independently reconstructed covariance model or an implemented finite-parameter DFR. Its paper gives the scalar BDD result $2^{-180.72}$, but neither the coefficient-level construction of σ_{Err} nor an unambiguous implementation convention for $n/2$ at odd n is available. We condition on the zero message and use the residual $gr + 2f'(e + \varepsilon) + (e + \varepsilon)$ only to determine adjacent correlation coefficients. The marginal variance is fixed by calibration to the published scalar reference. We interpret the loop bound and second-stream offset by floor integer division: the 67 blocks pair $0, \dots, 535$ with $543, \dots, 1078$, while the remaining 15 coefficients lie outside these complete blocks. The calculation retains the displayed $n/16$ normalization.

Table 8. Second-order Gaussian BDD log-probabilities before and after the parity-separated layout. CTRU uses reconstructed second moments; DTRU is the zero-message, floor-division, calibrated sensitivity analysis that fixes the published scalar marginal. Neither row is a finite-parameter DFR.

Scheme	published scalar	contiguous correlated	parity separated	Gaussian-reference change (bits)
CTRU-Prime-761	−160.000	−167.424	−304.540	137.116
DTRU-Prime-1087	−180.720	−107.421	−180.720	73.299

For CTRU-Prime, zeroing the identified off-diagonal terms gives -304.844 in the original blocks and -304.540 after parity regrouping. Covariance removal contributes 137.420 bits, marginal regrouping offsets 0.304 bit, and the net change is 137.116 bits; the modified-layout value is 144.540 bits below the published -160 reference for the original layout. In the conditional DTRU model, zeroing gives -180.720 and parity separation leaves it unchanged, a 73.299-bit calibrated sensitivity at fixed marginal scale. Equality with the published DTRU value is imposed by that calibration and is not a scheme-level DFR improvement. The public RefinedDFR curvature expression gives -163.538 and -301.585 for the two CTRU layouts; our values use $K''(s) = 2 \sum_i \lambda_i^2 / (1 - 2s\lambda_i)^2$. With this corrected curvature, classical Lugannani–Rice changes the CTRU entries by at most 0.029 bit and the DTRU entries by at most 0.028 bit.

7 Comparison across schemes

The same distinctions apply elsewhere. The mixed-noise analysis of Scloud⁺, which uses a scaled BW_{32} code, reports BDD bounds of $2^{-133.39}$, $2^{-199.65}$, and $2^{-265.11}$ [27,28]. Coded P₂₄-Kyber1024 reports DFR below 2^{-281} under within-codeword independence [19]. Both specify the event, joint residual law, tail method, and block aggregation; we do not reevaluate them.

Our direct measurements cover the paired CTRU and DTRU event gaps, the BW-KEM BDD discrepancies, and aggregation at observable auxiliary thresholds. The Gaussian covariance calculations, Prime-layout references, and DTRU published-row deep tails are analytic or model-based; none of these results alone gives a deployment-tail guarantee.

Aggregation at observable thresholds. We move only the normalized decision threshold inward and test theorem 1 on the resulting auxiliary events. At thresholds with at least ten ordered pairs, we fit $\nu_{2,t} = CA_t^\beta$; $\beta > 1$ means that relative pair overlap decreases.

Table 9. Aggregation of auxiliary events at observable inward thresholds using 16 keys, 2^{20} confirmations, and a 2^{17} -sample pilot. Ranges are leave-one-threshold-out.

Local event	β (LOO range)	crossings/ events	event $\log_2 p$
CTRU-576 E_8 score	1.757 [1.736, 1.769]	1,264/1,260	−9.701
BW-KEM-512 BDD exit	1.820 [1.799, 1.830]	962/962	−10.090
DTRU-512 paired- E_8	1.701 [1.637, 1.805]	1,003/1,001	−10.033
score			
DTRU-512 BDD score	1.815 [1.772, 1.874]	1,000/999	−10.036

8 Conclusions

The (A, Q, T, G) record places heterogeneous DFR calculations on one scale by fixing the event, residual law, tail engine, and aggregation. Combined with the finite radial-to-facet ratio, distribution-free directional bounds, and the local-to-ciphertext sandwich, it gives a common route from decoder geometry to an auditable probability statement.

The scheme studies show why all four fields matter. The CTRU complete-map run records 2,396 packing exits but only 26 message failures, a 6.526-bit certificate gap. DTRU’s measured gap grows from 5.85 to 7.63 bits, while its paired-block covariance substantially improves both BDD and nearest-facet references. For BW-KEM, the bounded-coefficient product limit justifies a joint Gaussian product reference on each fixed block, and the finite noncentral mixture retains the terminal noise. On the measurable grid, its largest discrepancy is 1.116 bits, compared with 3.388 bits for the valid clipped Chernoff bound; across the six published rows, the fixed-reference Chernoff slack is 6.537–7.621 bits.

For the prime-degree variants, the quotient-ring reduction produces the adjacent covariance identified by Zhou et al.; contiguous codeword placement puts the correlated coordinates in the same decoder block. The parity-separated assignment removes this covariance from every E_8 octet at second order, while theorem 7 proves asymptotic joint independence for the dominant bounded-coefficient product terms. Under the stated Gaussian BDD references, the CTRU-Prime-761 certificate changes from $2^{-167.424}$ to $2^{-304.540}$ and the DTRU-Prime-1087 sensitivity changes from $2^{-107.421}$ to $2^{-180.720}$. These values quantify a modified coding layout; a finalized scheme assessment must also cover its security and format, higher-order dependence, and the implemented decoder event. More generally, observable-tail experiments support a deployment DFR when the pair tails and residual-law transfer are certified over the relevant saddlepoint range.

References

1. Bahadur, R.R., Rao, R.R.: On deviations of the sample mean. *The Annals of Mathematical Statistics* **31**(4), 1015–1027 (1960). <https://doi.org/10.1214/aoms/1177705674>
2. Barbosa, M., Kannwischer, M.J., Lim, T., Schwabe, P., Strub, P.Y.: Formally verified correctness bounds for lattice-based cryptography. In: *ACM CCS 2025*. pp. 156–169. ACM (2025). <https://doi.org/10.1145/3719027.3765218>
3. Bernstein, D.J., Chuengsatiansup, C., Lange, T., van Vredendaal, C.: NTRU Prime: Reducing attack surface at low cost. In: *Selected Areas in Cryptography (SAC 2017)*. LNCS, vol. 10719, pp. 235–260. Springer (2018). https://doi.org/10.1007/978-3-319-72565-9_12
4. Bindel, N., Schanck, J.M.: Decryption failure is more likely after success. In: *PQCrypto 2020*. LNCS, vol. 12100, pp. 206–225. Springer (2020). https://doi.org/10.1007/978-3-030-44223-1_12
5. Clopper, C.J., Pearson, E.S.: The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika* **26**(4), 404–413 (1934). <https://doi.org/10.1093/biomet/26.4.404>
6. Cryptography Standardization Technical Committee: 2023 annual cryptography industry standard formulation and revision tasks (commercial cryptography field). Standardization Task Notice No. 9 of 2023 (2023)
7. D’Anvers, J.P., Batsleer, S.: Multitarget decryption failure attacks and their application to Saber and Kyber. In: *PKC 2022, Part I*. LNCS, vol. 13177, pp. 3–33. Springer (2022). https://doi.org/10.1007/978-3-030-97121-2_1
8. D’Anvers, J.P., Guo, Q., Johansson, T., Nilsson, A., Vercauteren, F., Verbaudhede, I.: Decryption failure attacks on IND-CCA secure lattice-based schemes. In: *PKC 2019, Part II*. LNCS, vol. 11443, pp. 565–598. Springer (2019). https://doi.org/10.1007/978-3-030-17259-6_19
9. D’Anvers, J.P., Rossi, M., Virdia, F.: (one) failure is not an option: Bootstrapping the search for failures in lattice-based encryption schemes. In: *EUROCRYPT 2020, Part III*. LNCS, vol. 12107, pp. 3–33. Springer (2020). https://doi.org/10.1007/978-3-030-45727-3_1
10. D’Anvers, J.P., Vercauteren, F., Verbaudhede, I.: The impact of error dependencies on ring/mod-LWE/LWR based schemes. In: *Post-Quantum Cryptogra-*

- phy. LNCS, vol. 11505, pp. 103–115. Springer (2019). https://doi.org/10.1007/978-3-030-25510-7_6
11. Davies, R.B.: Algorithm AS 155: The distribution of a linear combination of chi-squared random variables. *Applied Statistics* **29**(3), 323–333 (1980). <https://doi.org/10.2307/2346911>
 12. Guo, Q., Nabokov, D., Johansson, T.: Unlocking the true potential of decryption failure oracles: A hybrid adaptive-LDPC attack on ML-KEM using imperfect oracles. *Cryptology ePrint Archive*, Paper 2026/070 (2026), <https://eprint.iacr.org/2026/070>
 13. Hofheinz, D., Hövelmanns, K., Kiltz, E.: A modular analysis of the Fujisaki–Okamoto transformation. In: TCC 2017, Part I. LNCS, vol. 10677, pp. 341–371. Springer (2017). https://doi.org/10.1007/978-3-319-70500-2_12
 14. Hu, X., Zou, H., Shen, S., Li, W., Zhao, Y.: CTRU-Prime high-throughput implementation based on CUDA core and tensor core. *Journal of Software* (Jan 2026). <https://doi.org/10.13328/j.cnki.jos.007559>
 15. Imhof, J.P.: Computing the distribution of quadratic forms in normal variables. *Biometrika* **48**(3/4), 419–426 (1961). <https://doi.org/10.1093/biomet/48.3-4.419>
 16. Institute of Commercial Cryptography Standards: Call for proposals for the next-generation public-key cryptographic algorithms (Oct 2025), https://www.niccs.org.cn/niccs/Notice/pc/content/content_1975896137741635584.html
 17. Li, S., Zou, H., Shen, S., Dong, Y., Zhao, Y.: BW-KEM: Robust and versatile MLWE-based KEM with Barnes–Wall lattices. In: PKC 2026. LNCS, vol. 16551, pp. 70–102. Springer (2026). https://doi.org/10.1007/978-3-032-26731-3_3
 18. Liang, Z., Fang, B., Zheng, J., Zhao, Y.: Compact and efficient KEMs over NTRU lattices. *Computer Standards & Interfaces* **89**, 103828 (2024). <https://doi.org/10.1016/j.csi.2023.103828>
 19. Liu, S., Sakzad, A.: Compact lattice-coded (multi-recipient) Kyber without CLT independence assumption. In: *Advances in Cryptology – ASIACRYPT 2025*. LNCS, vol. 16247, pp. 363–395. Springer (2025). https://doi.org/10.1007/978-981-95-5099-9_12
 20. Liu, Y., Ge, J., Zhang, Y., Wang, J., Lu, X.: A primer on dependency in polynomial product: Identify, exploit, and trim. *Cryptology ePrint Archive*, Paper 2026/802 (2026), <https://eprint.iacr.org/2026/802>
 21. Lugannani, R., Rice, S.: Saddle point approximation for the distribution of the sum of independent random variables. *Advances in Applied Probability* **12**(2), 475–490 (1980). <https://doi.org/10.2307/1426607>
 22. Lyu, S., Liu, L., Ling, C., Lai, J., Chen, H.: Lattice codes for lattice-based PKE. *Cryptology ePrint Archive*, Paper 2022/874 (2022), <https://eprint.iacr.org/2022/874>
 23. Majenz, C., Sisinni, F.: Provable decryption failure security for practical lattice-based PKE. In: *Selected Areas in Cryptography (SAC 2026)* (2026), <https://eprint.iacr.org/2025/1997>
 24. National Institute of Standards and Technology: Module-lattice-based key-encapsulation mechanism standard. Tech. Rep. FIPS 203, NIST (2024). <https://doi.org/10.6028/NIST.FIPS.203>
 25. Owen, A.B., Maximov, Y., Chertkov, M.: Importance sampling the union of rare events with an application to power systems analysis. *Electronic Journal of Statistics* **13**(1), 231–254 (2019). <https://doi.org/10.1214/18-EJS1527>

26. Satterthwaite, F.E.: An approximate distribution of estimates of variance components. *Biometrics Bulletin* **2**(6), 110–114 (1946). <https://doi.org/10.2307/3002019>
27. Wang, A., Zheng, Z., Zhao, C., Qiu, Z., Zeng, G., Yuan, Y., Mu, C., Wang, X.: Scloud+: An efficient LWE-based KEM without ring/module structure. In: *Security Standardisation Research*. LNCS, vol. 15559, pp. 147–174. Springer (2025). https://doi.org/10.1007/978-3-031-87541-0_7
28. Zhou, G., Wang, L., Cao, Y., Huang, B., Hou, S., Gu, D.: Refined evaluation methods of decryption failure rate in lattice-based public-key encryption with message encoding. *Cryptology ePrint Archive*, Paper 2026/1350 (2026), <https://eprint.iacr.org/2026/1350>
29. Zou, H., Li, S., Zheng, J., Hu, X., Wei, H., Ao, W., Dong, Y., Guo, W., Zhao, Y.: DTRU: A versatile, compact, simple, and robust NTRU KEM with double E_8 encoding. In: *European Symposium on Research in Computer Security (ESORICS 2026)* (2026), <https://eprint.iacr.org/2026/1701>

A Full Proofs of Theorems 4 and 7

This appendix gives the quadratic-form estimates used in the proofs of the two joint Gaussian limits.

A.1 Joint Gaussian limit for negacyclic products with bounded coefficients

Proof. For $c \in \{0, \dots, N-1\}$, let A_c be the signed permutation matrix such that the coefficient of x^c in $fg \bmod (x^N + 1)$ is $f^T A_c g$. Negacyclic reduction sends each coefficient product to exactly one output coordinate, with a sign determined by whether its degree wraps around N . Hence

$$\|A_c\|_{\text{op}} = 1, \quad \|A_c\|_F^2 = N, \quad \langle A_c, A_d \rangle_F = 0 \quad (c \neq d).$$

Fix $t = (t_1, \dots, t_m) \in \mathbb{R}^m$ and put $B = \sum_{a=1}^m t_a A_{c_a}$. The corresponding Cramér–Wold projection is

$$S_{\ell, N}(t) = \frac{1}{\sigma_f \sigma_g \sqrt{\ell N}} \sum_{j=0}^{\ell-1} (f^{[j]})^T B g^{[j]}.$$

Condition on all $f^{[j]}$. This is a triangular array in the independent g -coefficients. If the f -coefficients are bounded by M_f , each conditional array coefficient has absolute value at most

$$\frac{M_f \|t\|_1}{\sigma_f \sigma_g \sqrt{\ell N}},$$

because every column of each A_{c_a} has one nonzero entry. The maximum coefficient therefore tends to zero. Boundedness of the g -coefficients gives the conditional Lindeberg condition.

The conditional variance is

$$V_{\ell, N}(t) = \frac{1}{\sigma_f^2 \ell N} \sum_{j=0}^{\ell-1} (f^{[j]})^T B B^T f^{[j]}. \quad (45)$$

Frobenius orthogonality gives

$$\mathbb{E}[V_{\ell, N}(t)] = \frac{\|B\|_F^2}{N} = \|t\|_2^2.$$

For an iid centered vector F with bounded coordinates, there is a constant $K_f < \infty$, depending only on its second and fourth moments, such that

$$\text{Var}(F^T M F) \leq K_f \|M\|_F^2$$

for every symmetric matrix M . Independence of the ℓ blocks and equation (45) therefore imply

$$\text{Var}(V_{\ell, N}(t)) \leq \frac{K_f \|B B^T\|_F^2}{\sigma_f^4 \ell N^2}.$$

Moreover,

$$\|BB^T\|_F^2 \leq \|B\|_{\text{op}}^2 \|B\|_F^2 \leq \|t\|_1^2 N \|t\|_2^2.$$

Thus

$$\text{Var}(V_{\ell,N}(t)) \leq \frac{K_f \|t\|_1^2 \|t\|_2^2}{\sigma_f^4 \ell N} = o(1).$$

The conditional variance converges in probability to $\|t\|_2^2$. Conditional Lindeberg–Feller yields a centered normal limit with this variance for every fixed t . Taking expectations of the conditional characteristic functions and applying Cramér–Wold proves equation (26).

A.2 Joint Gaussian limit in the prime-degree quotient ring

Proof. For $c \in \{0, \dots, n-1\}$, let A_c be the 0–1 matrix defined by

$$h_c = \sum_j (f^{[j]})^T A_c g^{[j]}.$$

The reduction rule

$$x^{n+s} = x^{s+1} + x^s, \quad 0 \leq s \leq n-2,$$

shows that every coefficient product contributes either to one output coordinate or to two adjacent output coordinates. Direct counting gives

$$\|A_c\|_F^2 = \nu_c, \quad \nu_0 = n, \quad \nu_c = 2n - c \quad (1 \leq c < n).$$

Each row and column contains at most two ones, so $\|A_c\|_{\text{op}} \leq 2$. A coefficient product can occur in two output matrices only when their indices are adjacent. Hence the supports of A_c and A_d are disjoint when $|c - d| \geq 2$.

Fix $t = (t_1, \dots, t_m) \in \mathbb{R}^m$ and define

$$C_t = \sum_{a=1}^m \frac{t_a A_{c_a}}{\sqrt{\nu_{c_a}}}.$$

The spacing assumption and the preceding counts imply

$$\|C_t\|_F^2 = \|t\|_2^2, \quad \|C_t\|_{\text{op}} \leq \frac{2\|t\|_1}{\sqrt{n}}. \quad (46)$$

The normalized Cramér–Wold projection is

$$S_{\ell,n}(t) = \frac{1}{\sigma_f \sigma_g \sqrt{\ell}} \sum_{j=0}^{\ell-1} (f^{[j]})^T C_t g^{[j]}.$$

Condition on all $f^{[j]}$. Since each column of A_c has at most two ones, boundedness of the f -coefficients makes the largest coefficient in this conditional triangular

array $O((\ell n)^{-1/2})$. Boundedness of the g -coefficients then gives the conditional Lindeberg condition.

The conditional variance equals

$$V_{\ell,n}(t) = \frac{1}{\sigma_f^2 \ell} \sum_{j=0}^{\ell-1} (f^{[j]})^T C_t C_t^T f^{[j]}. \quad (47)$$

Its expectation is $\|C_t\|_F^2 = \|t\|_2^2$. Using the quadratic-form variance bound from the preceding proof and independence across the ℓ blocks gives

$$\text{Var}(V_{\ell,n}(t)) \leq \frac{K_f \|C_t C_t^T\|_F^2}{\sigma_f^4 \ell}.$$

By equation (46),

$$\|C_t C_t^T\|_F^2 \leq \|C_t\|_{\text{op}}^2 \|C_t\|_F^2 \leq \frac{4\|t\|_1^2 \|t\|_2^2}{n}.$$

Consequently,

$$\text{Var}(V_{\ell,n}(t)) \leq \frac{4K_f \|t\|_1^2 \|t\|_2^2}{\sigma_f^4 \ell n} = o(1).$$

Thus the conditional variance converges in probability to $\|t\|_2^2$. Conditional Lindeberg–Feller, followed by expectation over the conditioning variables and Cramér–Wold, proves equation (44).

B Supplementary Scheme Tables and Plots

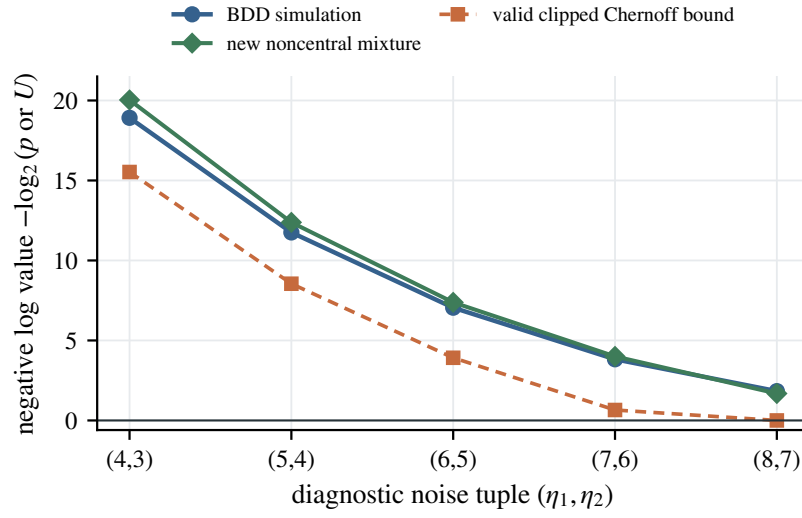


Fig. 3. BW-KEM diagnostic grid. The analytic curves use Q_{BW} and a sum over eight blocks; the simulation uses the implemented residual law and the implemented event that at least one block exits. Table 1 compares the two methods for evaluating the tail probability under the same reference model.

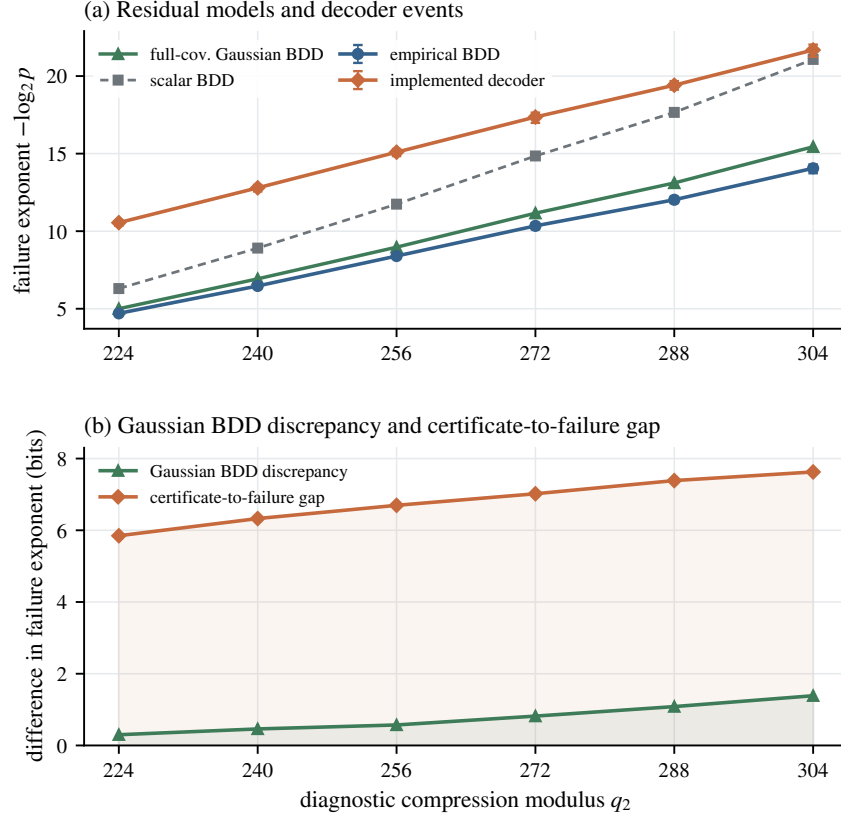


Fig. 4. DTRU diagnostic grid. Panel (a) compares scalarized Gaussian references and references using the full block covariance with empirical BDD and decoder events. Panel (b) displays the Gaussian BDD discrepancy relative to the implemented decoder separately from the measured certificate-to-failure gap; it does not assign the combined discrepancy to a single cause. The intervals resample whole keys.

C Technical Details for the Finite Event Gap

Theorem 3 gives the finite interval used in the main text. This appendix refines it into an exact remainder identity, records the closed form used for numerical evaluation, and examines nonisotropic laws.

Set

$$\delta_{\cap}(t) = \frac{S(t) - p_0(t)}{S(t)}, \quad \delta_R(t) = \frac{p_F(t) - p_0(t)}{S(t)}. \quad (48)$$

Then

$$\log_2 \frac{P_B(t)}{p_F(t)} = \log_2 M_{d,m}(t) + \varepsilon_{\text{tail}}(t) + \varepsilon_{\text{event}}(t), \quad (49)$$

$$\varepsilon_{\text{event}}(t) = -\log_2(1 - \delta_{\cap}(t) + \delta_R(t)), \quad (50)$$

$$0 \leq \delta_{\cap}(t) \leq \frac{D(t)}{S(t)}, \quad 0 \leq \delta_R(t) \leq \frac{P_B(c_R t)}{S(t)}. \quad (51)$$

Consequently,

$$-\log_2 \left(1 + \frac{P_B(c_R t)}{S(t)} \right) \leq \varepsilon_{\text{event}}(t) \leq -\log_2 \left(1 - \frac{D(t)}{S(t)} \right). \quad (52)$$

With

$$\varepsilon_{\text{rem}}(t) := \varepsilon_{\text{tail}}(t) + \varepsilon_{\text{event}}(t), \quad (53)$$

the ratio and the complete remainder interval are

$$\log_2 \frac{P_B(t)}{p_F(t)} = \log_2 M_{d,m}(t) + \varepsilon_{\text{rem}}(t), \quad (54)$$

$$\varepsilon_{\text{tail}}(t) - \log_2 \left(1 + \frac{P_B(c_R t)}{S(t)} \right) \leq \varepsilon_{\text{rem}}(t) \leq \varepsilon_{\text{tail}}(t) - \log_2 \left(1 - \frac{D(t)}{S(t)} \right). \quad (55)$$

For even dimension $d = 2k$, define

$$C_k(t) = \frac{\sum_{j=0}^{k-1} (t^2/2)^j / j!}{(t^2/2)^{k-1} / (k-1)!}, \quad \mu(t) = \frac{\bar{\Phi}(t)}{\varphi(t)}, \quad \varphi(t) = \frac{e^{-t^2/2}}{\sqrt{2\pi}}. \quad (56)$$

The integer-order incomplete-gamma identity gives

$$P_B(t) = \frac{\Gamma(k, t^2/2)}{\Gamma(k)} = e^{-t^2/2} \sum_{j=0}^{k-1} \frac{(t^2/2)^j}{j!}. \quad (57)$$

Using $\bar{\Phi}(t) = \varphi(t)\mu(t)$ and the definitions in equations (21) and (56), direct cancellation yields

$$\begin{aligned} \frac{P_B(t)}{m\bar{\Phi}(t)M_{2k,m}(t)} &= \frac{C_k(t)}{t\mu(t)}, \\ \varepsilon_{\text{tail}}(t) &= \log_2 C_k(t) - \log_2(t\mu(t)). \end{aligned} \quad (58)$$

This is an identity at finite t .

The Mills-ratio bounds used in the main text follow by integration by parts. Since

$$\bar{\Phi}(t) = \frac{\varphi(t)}{t} - \int_t^\infty \frac{\varphi(x)}{x^2} dx,$$

the integral is nonnegative and at most $\bar{\Phi}(t)/t^2$. Therefore

$$\frac{t}{1+t^2} \leq \mu(t) = \frac{\bar{\Phi}(t)}{\varphi(t)} \leq \frac{1}{t}. \quad (59)$$

It follows that $t^2/(1+t^2) \leq t\mu(t) \leq 1$. Substitution into equation (58) proves

$$\log_2 C_k(t) \leq \varepsilon_{\text{tail}}(t) \leq \log_2 C_k(t) + \log_2(1+t^{-2}). \quad (60)$$

No asymptotic expansion is needed for this bound.

Restoring a decoder radius r and an isotropic Gaussian scale σ gives $t = r/\sigma$ and

$$\log_2 M_{d,m}(r/\sigma) = (d-1)\log_2(r/\sigma) + \log_2 \frac{\sqrt{2\pi}}{m 2^{d/2-1} \Gamma(d/2)}. \quad (61)$$

Combining this formula with equation (49) yields the scaled finite identity

$$\begin{aligned} \log_2 \frac{\Pr[X_\sigma \in B^{(r)}]}{\Pr[X_\sigma \in F^{(r)}]} &= (d-1)\log_2(r/\sigma) + \log_2 \frac{\sqrt{2\pi}}{m 2^{d/2-1} \Gamma(d/2)} \\ &\quad + \varepsilon_{\text{tail}}(r/\sigma) + \varepsilon_{\text{event}}(r/\sigma), \end{aligned} \quad (62)$$

where $X_\sigma \sim \mathcal{N}(0, \sigma^2 I_d)$. For the scalar symmetric reference, $d = 1$ and $m = 2$, so the geometric factor and both corrections are zero.

For an anisotropic Gaussian $X \sim \mathcal{N}(0, \Sigma)$, the ball and facet events begin with different quadratic costs,

$$I_{\text{ball}} = \frac{r^2}{2\lambda_{\max}(\Sigma)}, \quad I_{\text{facet}} = \min_i \frac{h_i^2}{2a_i^T \Sigma a_i}.$$

Different costs give an exponential separation. Equal costs still do not determine the probability ratio because whitening maps the ball to an ellipsoid and leaves the active directions to be identified.

For an arbitrary residual law, d , m , r , and a covariance matrix do not determine the ratio. A law supported on a message-changing ray may satisfy $\Pr[B^{(r)}] = \Pr[F^{(r)}]$. If the radial event strictly contains the failure event, a law supported on $B^{(r)} \setminus F^{(r)}$ instead has positive radial probability and zero failure probability. Mixtures place either example at an arbitrarily small tail probability. These examples explain why the main theorem is stated for a fixed reference law and why the scheme comparisons separately assess directional residual tails on diagnostic data.

D CTRU Covariance Diagnostic

For $0 \leq i, j, k < n$, let ρ_{kij} be the coefficient of x^k in the canonical reduction of x^{i+j} modulo $x^n - x^{n/2} + 1$, and define $M(a)_{kj} = \sum_i \rho_{kij} a_i$. Then coefficient vectors satisfy $ab = M(a)b$. For a fixed key $k = (f, g, h)$, the compression error is a deterministic function $\varepsilon_k(r)$ of the encryption randomness. Put

$$D_r = \text{Cov}(r), \quad D_{\varepsilon,k} = \text{Cov}(\varepsilon_k(r)), \quad \Gamma_k = \text{Cov}(r, \varepsilon_k(r)).$$

Proposition 1 (Conditional and averaged reference covariance). *Let $c = q/q_2$. The conditional covariance of the proxy residual is*

$$\begin{aligned} \Sigma_k &= M(g)D_r M(g)^T + c^2 M(f)D_{\varepsilon,k} M(f)^T \\ &\quad + c\{M(g)\Gamma_k M(f)^T + M(f)\Gamma_k^T M(g)^T\}. \end{aligned} \quad (63)$$

The analytic fixed-key reference sets $D_{\varepsilon,k} = \sigma_\varepsilon^2 I_n$ and $\Gamma_k = 0$:

$$\Sigma_k^{(0)} = \sigma_r^2 M(g)M(g)^T + c^2 \sigma_\varepsilon^2 M(f)M(f)^T. \quad (64)$$

If its conditional means are zero, averaging the reference covariance gives $\Sigma_{\text{full}} = \mathbb{E}_K[\Sigma_K^{(0)}]$. For iid centered key coefficients,

$$[\Sigma_{\text{full}}]_{uv} = \sigma_r^2 \sigma_g^2 K_{uv} + c^2 \sigma_\varepsilon^2 (\mathbf{1}\{u = v\} + 4\sigma_f^2 K_{uv}), \quad K_{uv} = \sum_{i,j} \rho_{uij} \rho_{vij}. \quad (65)$$

The identity follows by expanding the covariance of $M(g)r + cM(f)\varepsilon_k(r)$. It does not assume that the implemented compression error is independent of the key or encryption randomness; equation (64) is the reference model used only in the Gaussian comparison under matched covariance.

In the diagnostic below, $\hat{\Sigma}_k$ is the sample covariance of the implemented block residuals for key k . The rounding-variance deviation is $(\hat{\sigma}_{\varepsilon,k}^2 / \sigma_\varepsilon^2) - 1$. The product/rounding entry is $\|\hat{\Gamma}_k + \hat{\Gamma}_k^T\|_F / \|\hat{\Sigma}_k\|_F$, where $\hat{\Gamma}_k$ is the empirical blockwise cross-covariance between the product and rounding components.

Table 10. Fixed-key covariance diagnostic at $(n, q, q_2, \eta_1, \eta_2) = (576, 3457, 512, 7, 8)$. Rows are means over 16 keys; intervals are 95% key-level Student- t intervals.

Diagnostic	Mean	95% key interval
$\ \hat{\Sigma}_k - \Sigma_k^{(0)}\ _F / \ \hat{\Sigma}_k\ _F$	3.303%	[3.274, 3.332]%
$\ \hat{\Sigma}_k - \Sigma_{\text{full}}\ _F / \ \hat{\Sigma}_k\ _F$	13.009%	[10.913, 15.105]%
$\ \Sigma_k^{(0)} - \Sigma_{\text{full}}\ _F / \ \Sigma_{\text{full}}\ _F$	12.195%	[10.340, 14.051]%
rounding-variance deviation from the uniform reference	0.007%	[-0.013, 0.027]%
product/rounding cross-covariance relative norm	1.896%	[1.870, 1.921]%

E Proofs and Numerical Details for the Scheme Comparisons

Proof (Proposition 1). For fixed k , write $X = M(g)r + cM(f)\varepsilon_k(r)$. Expanding the covariance of this sum gives equation (63), including both cross terms. The independent coefficientwise rounding reference gives equation (64). Averaging that reference over keys gives Σ_{full} . Finally, $\mathbb{E}[M(g)M(g)^T]_{uv} = \sigma_g^2 K_{uv}$, while $M(f) = I_n + 2M(f')$ and centeredness removes the two linear terms in f' . Substitution proves equation (65).

Proposition 2 (Gaussian sampler for a union of facets). *Let $H_a = \{x : a^T x \geq q\}$ for $a \in \mathcal{A}$, $p_a = \Pr_{X \sim \mathcal{N}(0, \Sigma)}[H_a]$, and $S = \sum_{a \in \mathcal{A}} p_a$. Select $J = a$ with probability p_a/S , sample $X \sim \mathcal{N}(0, \Sigma) \mid H_a$, and put $N(X) = \sum_{b \in \mathcal{A}} \mathbf{1}\{X \in H_b\}$. Then*

$$W(X) = \frac{S}{N(X)} \quad (66)$$

is unbiased for $\Pr[X \in \cup_a H_a]$ and satisfies $0 < W \leq S$.

Proof (Details for proposition 2). The mixture density is $g(x) = \phi_\Sigma(x)N(x)/S$ on the facet union and zero outside. Hence

$$\mathbb{E}_g[W] = \int_{\cup_a H_a} \frac{S}{N(x)} \frac{\varphi_\Sigma(x)N(x)}{S} dx = \Pr_{X \sim \mathcal{N}(0, \Sigma)}[X \in \cup_a H_a],$$

and $0 < W \leq S$ follows from $N(x) \geq 1$. This is the union estimator of [25] specialized to Gaussian halfspaces.

For a selected facet a , draw $Y = a^T X$ from its upper-truncated normal law. Independently draw the conditional residual with covariance $\Sigma - \Sigma a a^T \Sigma / (a^T \Sigma a)$ and combine the two components. The result has law $\mathcal{N}(0, \Sigma)$ conditioned on H_a .

For the weighted quadratic form in equation (34), the Lugannani–Rice calculation uses $K(s) = -\frac{1}{2} \sum_i \ln(1 - 2s\lambda_i)$. Solving $K'(\hat{s}) = r^2$ and setting

$$w = \sqrt{2(\hat{s}r^2 - K(\hat{s}))}, \quad u = \hat{s}\sqrt{K''(\hat{s})}$$

gives the approximation $\Pr[\|X\|^2 \geq r^2] \approx \Phi(-w) + \varphi(w)(u^{-1} - w^{-1})$.

Table 11. Numerical stability of the Gaussian ball calculation with matched covariance. The first three columns are $\log_2$ ciphertext probabilities under the same block union. The last column is the propagated Imhof quadrature error bound in bits of failure exponent.

n	Lugannani–			LR – I	Satt – I	error bound
	Imhof	Rice	Satterthwaite			
576	−20.150	−20.144	−20.150	0.006	< 0.001	< 0.001

F Parameter Rows and Diagnostic Overrides

Table 12 transcribes the rows with multidimensional decoders used in the analytic tables from the CTRU, BW-KEM, and DTRU papers. For each scheme, the parameters are listed in the following order:

$$\begin{aligned} \text{BW-KEM: } & (N, \ell, q, \eta_1, \eta_2, d_u, d_v), \\ \text{CTRU: } & (n, q, q_2, \chi_f, \chi_g), \\ \text{DTRU: } & (n, q, q_2, k_g, k_f, k_r, k_e). \end{aligned}$$

These values are checked against the analysis registries by the parameter-audit script.

Table 12. Published parameter rows used by the analytic calculations.

Scheme	set	parameter tuple
BW-KEM	512	(256, 2, 3329, 3, 3, 9, 4)
	512-s	(256, 2, 3329, 4, 4, 9, 5)
	768	(256, 3, 3329, 2, 2, 9, 4)
	768-s	(256, 3, 3329, 3, 3, 9, 6)
	1024	(256, 4, 3329, 2, 2, 9, 5)
	1024-s	(256, 4, 3329, 3, 3, 10, 5)
CTRU	576	(576, 3457, 1024, B_4, B_4)
	768	(768, 3457, 1024, B_3, B_3)
	1024	(1024, 3457, 1024, B_2, B_2)
DTRU	512	(512, 3457, 1024, 8, 4, 6, 3)
	768	(768, 3457, 1024, 4, 3, 4, 2)
	1024	(1024, 3457, 1024, 5, 2, 3, 2)
	1536	(1536, 3457, 1024, 3, 2, 2, 1)
	2048	(2048, 3457, 1024, 5, 1, 2, 1)

The CTRU tuple (576, 3457, 512, 7, 8) is the fixed diagnostic point for the comparison in section 5.3. It retains the published ring dimension, large modulus, encoder, and decoder while changing the compression modulus and using elevated noise strengths solely to make both decoder events measurable. The remaining experiments at observable failure probabilities use the overrides listed in table 13; unlisted fields remain those of the corresponding published 512-level row.

Table 13. Diagnostic overrides used only for end-to-end event checks.

Scheme	override	purpose
CTRU-576	$q_2 = 512; (\eta_1, \eta_2) = (7, 8)$	packing/implemented-event pair
BW-KEM-512	$(\eta_1, \eta_2) = (4, 3), \dots, (8, 7);$ $(d_u, d_v) = (8, 3)$	BDD tail-probability grid
DTRU-512	$q_2 \in \{224, 240, 256, 272, 288, 304\}$	implemented/BDD tail grid

G Reproducibility and Run Integrity

The anonymous artifact supplied with the submission provides access to the files named below. Hashes authenticate retrieved files but do not provide access to them. The public ledger redacts local data roots, and the release manifest was checked for usernames and absolute author paths before packaging.

The released artifact includes the 8,192-key CTRU confirmation, end-to-end event checks for BW-KEM and DTRU, and a common blockwise threshold experiment at the published parameters with 16 independent keys and 2^{20} confirmation ciphertexts per scheme, preceded by a 2^{17} pilot. The samples from the reference distribution contain 196,608 BW-KEM blocks across the six published BW-KEM rows. The run ledger records parameter sources, software and geometry hashes, covariance inputs, seeds, and merge checks.

The final CTRU confirmation (run prefix **pkc2027-v2-paper**) uses 256 deterministic shards, 8,192 keys, and 2^{30} ciphertexts. Its experiment source has SHA-256 prefix **43d48cbb5b39**. The merger rejects missing or duplicate shards, parameter or source-hash mismatches, and violations of the paired event-count identities.

The v5 release contains five BW-KEM and six DTRU grid runs; all 11 run manifests and the batch manifest have status **complete**. The result archive has SHA-256 prefix **7c561a48e69e**, all 6,081 inner checksum entries verify, and the server-code archive has SHA-256 prefix **4dd04f04b5e2**. The independent 2^{32} -trial scale-confirm archive has SHA-256 prefix **995326e824fe**. Its source hash and seeds differ from the grid, so its counts are reported as replications rather than pooled estimates. Full hashes and run paths are recorded in **artifact/FINAL_RELEASE_SHA256.txt** and the release ledger.

H Asymptotic Corollary of the Finite Ratio

The earlier small-noise expression follows from the finite identity in theorem 3; it is not used in any numerical result.

Corollary 1 (Prefactor for radial and facet-event probabilities). *Under the hypotheses of theorem 3, suppose that the number of nearest facets is finite,*

the normals u_i are distinct, and $c_R > 1$ whenever $\mathcal{R}_{\text{far}}$ is nonempty. Then, as $t \rightarrow \infty$,

$$\frac{P_B(t)}{p_F(t)} \sim \frac{\sqrt{2\pi}}{m 2^{d/2-1} \Gamma(d/2)} t^{d-1}. \quad (67)$$

Equivalently, for $X_\sigma \sim \mathcal{N}(0, \sigma^2 I_d)$ and fixed decoder radius $r > 0$, the same expression holds with $t = r/\sigma$ as $\sigma/r \rightarrow 0$.

Proof. The incomplete-gamma and Mills-ratio expansions give $P_B(t)/(m\bar{\Phi}(t)) \sim M_{d,m}(t)$, or equivalently $\varepsilon_{\text{tail}}(t) \rightarrow 0$. For distinct nonopposite normals, $u_i^T u_j < 1$, so every term in $D(t)/S(t)$ has a strictly larger Gaussian exponent than $\bar{\Phi}(t)$; opposite normals have an empty positive-tail intersection. Hence $D(t)/S(t) \rightarrow 0$. If $\mathcal{R}_{\text{far}}$ is nonempty, $c_R > 1$ similarly implies $P_B(c_R t)/S(t) \rightarrow 0$. The finite bounds in equation (55) therefore yield $\varepsilon_{\text{rem}}(t) \rightarrow 0$, and equation (54) proves the claim.

The scope of the corollary is the stated isotropic Gaussian reference law. The main-text intervals supply finite numerical results; transfer to a structured discrete residual requires separate control of the directional tails.