Structural Ambiguity Under Loop Flattening

Exact enumeration, universal envelopes, and a local-limit law

J. R. Landers

Abstract

A rectangular depth-k loop nest with ordered bounds $(n_1, \ldots, n_k)$ performs $N = \prod_i n_i$ innermost-body executions. Flattening can preserve this work exactly while erasing the factorization that gave the iteration space its shape. We measure the resulting structural ambiguity by the number $K_k(N)$ of ordered k-factorizations of N into integers at least two. We first derive an exact prime-exponent formula and show that $\lceil \log_2 K_k(N) \rceil$ is the sharp fixed-length metadata cost of recovering every compatible source shape. We then prove the factorization-free envelope

$$\log_2 K_k(N) \leq \inf_{s>1} \{s \log_2 N + k \log_2(\zeta(s) - 1)\},$$

which replaces a coarse $\mathrm{O}(\log k \log N)$ estimate by the correct linear scale in a uniform-random family.

The remaining slack has a more delicate geometry. Large-depth exact experiments falsify a natural conjecture that the zeta residual is merely $(r/2) \log_2 k$. The residual instead contains a positive linear large-deviation term because a scalar tilt remembers $\log N$ but not the individual prime exponents. For a fixed active prime set of size r, we restore those coordinates, solve the multivariate saddle in closed form up to one scalar equation, and prove

$$\log_2 K_k(N_k) = B_P(N_k, k) - \frac{r}{2} \log_2(2\pi k) - \frac{1}{2} \log_2 \det \Sigma_k + \mathrm{O}(k^{-1}).$$

For bounds uniformly distributed over $\{2, \ldots, 16\}$, the effective dimension is $r = 6$; exact counts through $k = 15{,}360$ recover $r_{\text{fit}} = 5.992933$ and agree with the leading local-limit correction to 0.000379 bits. The result separates universal work-only control from the prime geometry governing sharp structural recovery.

1 Introduction

Loop flattening turns a multidimensional iteration space into a single linear iteration count. In the ideal rectangular case,

$$(n_1, \ldots, n_k) \longmapsto N = \prod_{i=1}^{k} n_i.$$

This map may preserve every executed operation while destroying the visible boundaries between dimensions. For example, when $N = 12$ and $k = 2$, the ordered shapes

$$(2, 6), \quad (3, 4), \quad (4, 3), \quad (6, 2)$$

all produce the same work count. They are not interchangeable as source geometries: order determines coordinate maps, index expressions, and often the natural placement of parallelism.

Compiler infrastructures contain explicit loop-flattening transformations [5]. Work on loop rolling and rerolling studies the inverse problem using syntactic and semantic evidence retained in lower-level code [8, 10]. Here we isolate a prior question: if the observer receives only the product N and the depth k, how many source geometries remain possible? The answer is an ordered-factorization count. Its logarithm has a direct operational meaning as the least worst-case metadata needed to name the erased source shape.

This viewpoint exposes two different mathematical problems. The first is *universal control*: how large can the ambiguity be when only (N, k) is available? The second is *sharp asymptotics*: once the prime-exponent geometry of a family is known, how closely can its exact ambiguity be approximated? The first leads to a one-parameter zeta envelope; the second to a multivariate tilted lattice law.

Contributions.

- We formalize structural ambiguity under rectangular loop flattening as the exact ordered-factorization count $K_k(N)$ and prove the sharp recovery cost $\lceil \log_2 K_k(N) \rceil$.
- We organize three levels of control: a coarse work-only estimate, an exponent-aware stars-and-bars bound, and a factorization-free depth-sensitive zeta envelope.
- We conduct two complementary experiments: a random uniform-source study and a fixed-direction exact study reaching $k = 15{,}360$.
- We show that the scalar zeta residual is generally *linear plus logarithmic*, not purely logarithmic, and identify the linear coefficient as a large-deviation rate mismatch.
- We prove a uniform fixed-prime-support local-limit theorem with an explicit saddle and covariance. It identifies the effective dimension and yields logarithmic error of order k^{-1}.

The ordered-factorization ingredients are classical [12, 4]. Likewise, the local-limit mechanism belongs to established multivariate asymptotic theory [2, 7, 6]. The contribution is their explicit synthesis for transformation-induced source ambiguity, including the universal-to-sharp transition and the diagnosis of the zeta envelope's residual.

2 Model, exact ambiguity, and recovery

2.1 Observation model

A perfectly rectangular ordered loop nest of depth k is represented by

$$M = (n_1, \dots, n_k) \in \mathbb{N}_{\geq 2}^k.$$

Its idealized flattened observation is

$$F(M) = \prod_{i=1}^{k} n_i = N.$$

The model suppresses index arithmetic, memory traces, debug metadata, and compiler conventions. Such evidence can only reduce ambiguity; the present model measures what the product map itself erases.

Definition 1 (Depth-resolved ambiguity)**.** For $N > 1$ and $k \geq 1$, define

$$\mathcal{M}_{N,k} = \{(a_1, \dots, a_k) \in \mathbb{N}_{\geq 2}^k : a_1 \cdots a_k = N\}, \qquad K_k(N) = |\mathcal{M}_{N,k}|,$$

and define the structural information loss

$$L_k(N) = \log_2 K_k(N).$$

If

$$N = \prod_{j=1}^{r} p_j^{e_j}, \qquad e_j \geq 1,$$

then every candidate factorization corresponds to an $r \times k$ matrix of nonnegative prime exponents. Row j sums to e_j, and a zero column is forbidden because every factor must exceed one.

Theorem 2 (Exact structural ambiguity). *For $N = \prod_{j=1}^{r} p_j^{e_j} > 1$,*

$$K_k(N) = \sum_{q=1}^{k} (-1)^{k-q} \binom{k}{q} \prod_{j=1}^{r} \binom{e_j + q - 1}{e_j}.$$

Proof. First allow factors equal to one. Distributing the e_j indistinguishable copies of prime p_j across q labeled columns gives $\binom{e_j+q-1}{e_j}$ weak compositions. Prime rows are independent, so the number of exponent matrices on q available columns is the product over j. Inclusion–exclusion over the zero columns among the original k columns gives the displayed alternating sum. □

The maximum feasible depth is

$$\Omega(N) = \sum_{j=1}^{r} e_j,$$

because every factor consumes at least one prime occurrence. For a prime power the exact profile becomes especially transparent.

Corollary 3 (Prime-power depth law). *For a prime p and $1 \le k \le m$,*

$$K_k(p^m) = \binom{m-1}{k-1}.$$

Proof. Write each factor as p^{a_i}. The positive exponent vector satisfies $a_1 + \cdots + a_k = m$, so candidates are compositions of m into k positive parts. □

2.2 The operational bit cost

Suppose a metadata string must identify the original source among all members of $\mathcal{M}_{N,k}$.

Theorem 4 (Sharp fixed-length recovery cost). *The minimum fixed-length metadata capacity sufficient to recover every compatible source shape is*

$$B_{\min}(N,k) = \lceil \log_2 K_k(N) \rceil.$$

Under a budget of B bits, worst-case exact recovery is impossible whenever $K_k(N) > 2^B$.

Proof. A B-bit tag has at most 2^B values, so exact recovery requires an injection from $\mathcal{M}_{N,k}$ into those values. Conversely, enumerate the candidate shapes and encode the index of the source shape. □

This is a capacity result, not a runtime lower bound. Under a nonuniform source prior, the corresponding average description length is controlled by the posterior entropy $H(M \mid N, k)$ [9]; the theorem is prior-free and worst-case exact.

3 Three envelopes for the ambiguity

Exact evaluation of theorem 2 uses the prime exponents of N. We next separate what can be said with and without that arithmetic information.

3.1 A coarse work-only estimate

Label every prime occurrence of N and assign it to one of k factors. This overcounts valid exponent matrices, but gives

$$K_k(N) \le k^{\Omega(N)}.$$

Since every prime is at least two,

$$\Omega(N) \le \log_2 N,$$

and therefore

$$L_k(N) \le (\log_2 k)(\log_2 N). \tag{1}$$

The estimate is universal and elementary. Its two relaxations are costly: equal primes have been labeled, and factors equal to one have been admitted.

3.2 An exponent-aware relaxation

If the prime exponents are known, dropping only the nonempty-column condition gives

$$K_k(N) \le d_k(N) := \prod_{p^e \| N} \binom{e+k-1}{e}. \tag{2}$$

This generalized divisor count explains much of the looseness in equation (1), but it requires factoring N and remains exponentially loose in some fixed-direction regimes.

3.3 The factorization-free zeta envelope

The exact count has the Dirichlet generating function

$$(\zeta(s)-1)^k = \sum_{N \ge 1} K_k(N) N^{-s}, \qquad s > 1. \tag{3}$$

Every coefficient is nonnegative. Keeping only the Nth term yields the following bound.

Theorem 5 (Universal depth-aware envelope). *For every $N > 1$ and $k \ge 1$,*

$$L_k(N) \le B_\zeta(N,k) := \inf_{s>1} \left[s \log_2 N + k \log_2(\zeta(s) - 1) \right].$$

Consequently,

$$L_k(N) \le \min\{(\log_2 k)(\log_2 N),\ B_\zeta(N,k)\}.$$

Proof. From equation (3), $K_k(N)N^{-s} \le (\zeta(s)-1)^k$ for every $s > 1$. Rearrangement, base-two logarithms, and minimization prove the first claim. Taking the minimum with equation (1) proves the second. □

Let ρ be the unique real solution of $\zeta(\rho) = 2$:

$$\rho = 1.728647238998\ldots.$$

Choosing $s = \rho$ cancels the depth term and gives the simple universal corollary

$$L_k(N) \le \rho \log_2 N. \tag{4}$$

Unlike equation (1), this grows only linearly with $\log N$, uniformly in depth. The constant ρ also appears in maximal-order results for ordered factorizations [4].

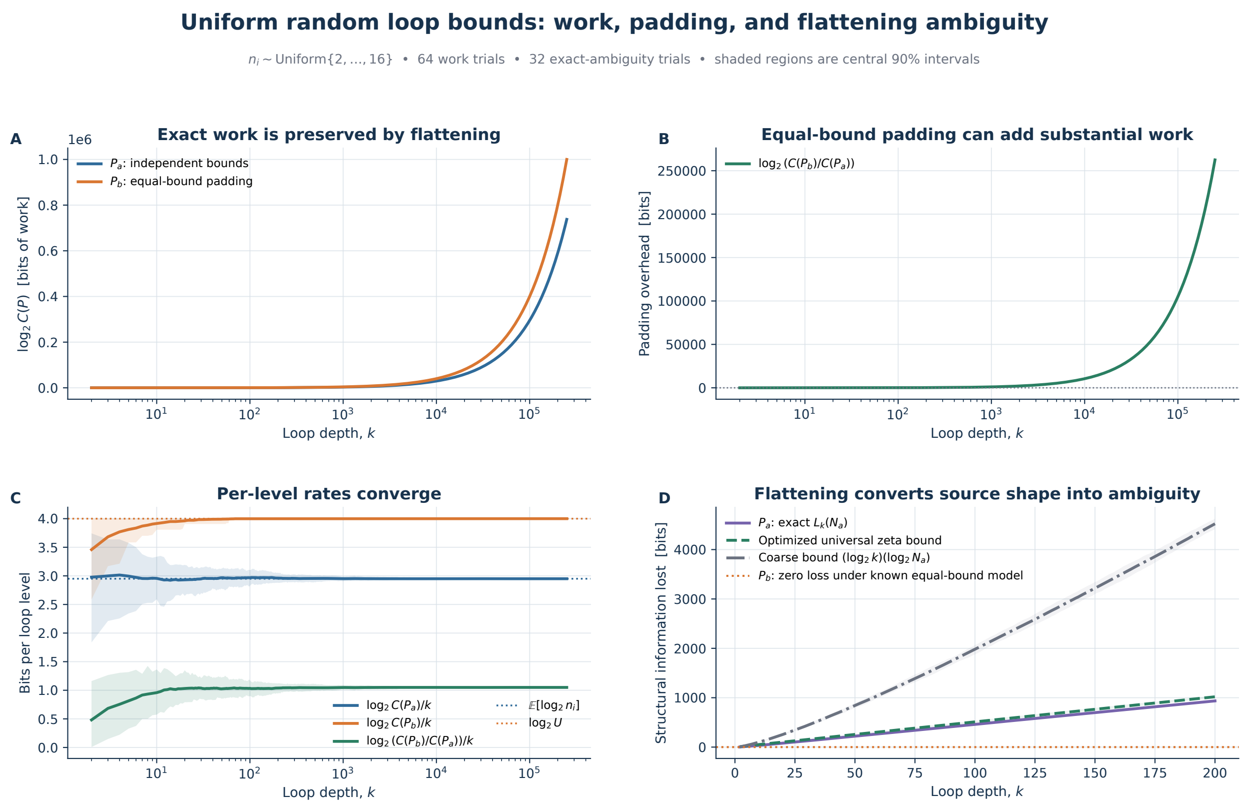

Figure 1: **Random uniform-source experiment.** Panel A shows work for the original and equal-bound-padded nests; flattening does not alter either curve. Panel B isolates padding overhead. Panel C shows convergent per-level rates. Panel D compares exact structural loss with the coarse and zeta envelopes. The zeta bound captures the linear scale that the coarse theorem misses.

4 Experiment I: where the bounds separate

4.1 Random-source protocol

We first sampled loop bounds independently from

$$n_i \sim \mathrm{Uniform}\{2, 3, \ldots, 16\}.$$

Each trial is one long sequence; depth-k observations use prefixes, so the curves represent a growing nest rather than unrelated programs. The independent-bound program has

$$C(P_a) = N_a(k) = \prod_{i=1}^{k} n_i.$$

For contrast, an equal-bound padded program uses $M_k = \max_{i \le k} n_i$ and has $C(P_b) = M_k^k$. Flattening P_a itself preserves $N_a(k)$ exactly. Padding is a separate transformation and can add substantial work.

The work experiment used 64 trials through $k = 250{,}000$. Exact ambiguity used 32 trials through $k = 200$, with arbitrary-precision inclusion–exclusion. Shaded bands in the figures are central 90% intervals.

At $k = 200$, the mean exact ambiguity was 935.335 bits. The optimized zeta bound was 1018.897 bits, while the coarse estimate allowed 4522.400 bits. Thus the zeta envelope used about 91.8% of its available capacity, compared with 20.7% for the coarse bound.

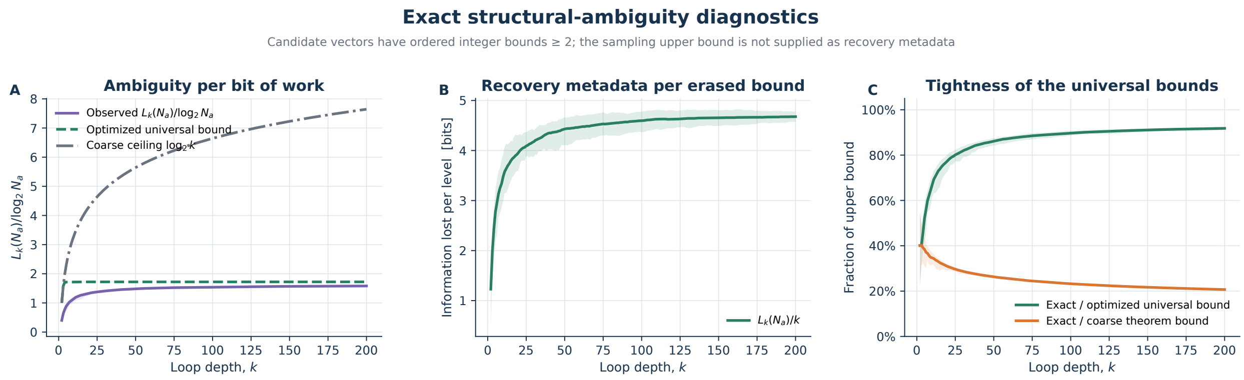

Figure 2: **Normalized ambiguity and bound utilization.** The exact loss per bit of work stabilizes, while the coarse normalized ceiling $\log_2 k$ diverges. The optimized zeta envelope remains close to the exact curve and is substantially tighter than both earlier estimates.

The asymptotic work rate is

$$\mathbb{E}[\log_2 n_i] = 2.9500093647 \quad \text{bits per level.}$$

For this scalar work rate, the zeta optimizer approaches

$$s^\star = 1.6241419368,$$

and the envelope approaches

$$\frac{B_\zeta}{k} = 5.0815475877 \quad \text{bits per level.}$$

The experiment therefore establishes that the universal bound has learned the right *scale*. It does not yet explain its residual.

5 The residual conjecture and its failure

5.1 Why a logarithmic correction looked plausible

Define the scalar partition function

$$Z(s) = \zeta(s) - 1$$

and the probability law

$$\mathbb{P}_s\{A = a\} = \frac{a^{-s}}{Z(s)}, \qquad a \geq 2.$$

For independent $A_1, \ldots, A_k$, coefficient extraction gives the exact identity

$$K_k(N) = N^s Z(s)^k \mathbb{P}_s\{A_1 \cdots A_k = N\}. \tag{5}$$

At the optimizing s, the first factor is precisely the exponential zeta envelope. The residual is the negative logarithm of a point probability. If that probability behaved like a centered r-dimensional lattice mass, one would expect

$$B_\zeta - L_k \stackrel{?}{=} \frac{r}{2} \log_2 k + \mathrm{O}(1).$$

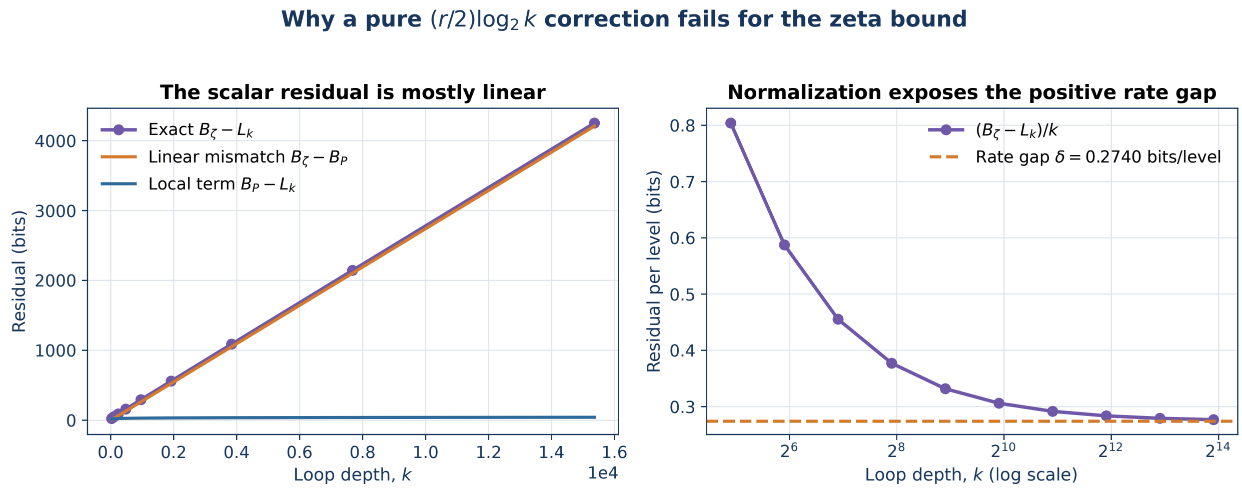

Figure 3: **The scalar conjecture fails in a structured way.** The zeta residual is dominated by a linear rate mismatch. Dividing by k reveals convergence to $\delta = 0.2739836276$ bits per level; the local term is lower order.

5.2 A fixed-direction test

Monte Carlo prefixes cause their normalized exponent vectors to fluctuate by $\mathrm{O}(k^{-1/2})$. To isolate the asymptotics, we repeated the deterministic block $(2, 3, \ldots, 16)$. Its product is

$$\prod_{a=2}^{16} a = 2^{15}3^6 5^3 7^2\, 11\, 13.$$

At every tested multiple of 15, the normalized exponent vector is exactly

$$\boldsymbol{\alpha} = \frac{\mathbf{e}}{k} = \left(1, \frac{2}{5}, \frac{1}{5}, \frac{2}{15}, \frac{1}{15}, \frac{1}{15}\right). \tag{6}$$

Exact counts were computed at geometrically spaced depths through $k = 15{,}360$.

The result is decisive. A log-only fit to $B_\zeta - L_k$ has a root-mean-square error of 773.745 bits. A linear-plus-log fit has error 0.0272 bits. The positive limiting rate is

$$\delta = 0.2739836276 \quad \text{bits per level.}$$

The conjecture was attached to the wrong envelope.

5.3 The missing geometry

The scalar statistic

$$\log N = \sum_p e_p \log p$$

is one projection of the exponent vector. Exact product equality requires matching every active prime exponent, and it also requires zero exponents for all primes absent from N. The zeta tilt gives positive mass to factors containing every prime. For a fixed smooth target, suppressing all outside prime coordinates is a large-deviation event, with an exponential rather than polynomial cost.

The remedy is not to force a local theorem onto the scalar partition function. It is to restore the exponent coordinates before tilting.

6 Prime-coordinate saddlepoint geometry

6.1 The multivariate coefficient

Fix the active prime set

$$P = \{p_1, \ldots, p_r\}, \qquad N = \prod_{j=1}^{r} p_j^{e_j}.$$

A factor of N corresponds to an exponent vector $\mathbf{v} \in \mathbb{N}^r \setminus \{\mathbf{0}\}$. Its generating function is

$$F(\mathbf{z}) = \sum_{\mathbf{v} \in \mathbb{N}^r \setminus \{\mathbf{0}\}} \mathbf{z}^{\mathbf{v}} = \prod_{j=1}^{r} (1 - z_j)^{-1} - 1. \tag{7}$$

Consequently,

$$K_k(N) = [\mathbf{z}^{\mathbf{e}}] F(\mathbf{z})^k. \tag{8}$$

For every $\mathbf{z} \in (0,1)^r$, positivity gives

$$K_k(N) \le \frac{F(\mathbf{z})^k}{\mathbf{z}^{\mathbf{e}}}.$$

Define the optimized multivariate envelope

$$B_P(N, k) = \inf_{\mathbf{z} \in (0,1)^r} \log_2 \frac{F(\mathbf{z})^k}{\mathbf{z}^{\mathbf{e}}}. \tag{9}$$

6.2 An explicit saddle

Let $\boldsymbol{\alpha} = \mathbf{e}/k$. The interior of the feasible mean cone is

$$\mathcal{D}_r = \left\{ \boldsymbol{\alpha} \in (0, \infty)^r : \sum_{j=1}^{r} \alpha_j > 1 \right\}.$$

The saddle equations are

$$\alpha_j = z_j \frac{\partial}{\partial z_j} \log F(\mathbf{z}).$$

They reduce to one scalar.

Lemma 6 (Scalar reduction of the vector saddle). *For every $\boldsymbol{\alpha} \in \mathcal{D}_r$, there is a unique $c > 1$ satisfying*

$$\prod_{j=1}^{r} \left(1 + \frac{\alpha_j}{c}\right) = \frac{c}{c-1}. \tag{10}$$

The unique positive saddle is

$$z_j = \frac{\alpha_j}{\alpha_j + c}. \tag{11}$$

At this saddle,

$$F(\mathbf{z}) = \frac{1}{c-1}.$$

Proof. Put $G(\mathbf{z}) = \prod_j (1 - z_j)^{-1}$, so $F = G - 1$, and set $c = G/(G-1)$. The saddle equation becomes $\alpha_j = c z_j/(1 - z_j)$, which rearranges to equation (11). Substitution into $G = c/(c-1)$ gives equation (10). Existence and uniqueness also follow from strict convexity of $\log F(e^{\boldsymbol{\theta}}) - \boldsymbol{\alpha} \cdot \boldsymbol{\theta}$ on the interior mean domain; equivalently, the left-minus-right logarithm in equation (10) crosses zero once when $\sum_j \alpha_j > 1$. □

The natural-log exponential rate is therefore

$$\beta_P(\boldsymbol{\alpha}) = -\log(c-1) - \sum_{j=1}^{r} \alpha_j \log z_j. \tag{12}$$

6.3 The tilted exponent law

At a saddle $\mathbf{z}$, define

$$\mathbb{P}_{\mathbf{z}}\{V = \mathbf{v}\} = \frac{\mathbf{z}^{\mathbf{v}}}{F(\mathbf{z})}, \qquad \mathbf{v} \in \mathbb{N}^r \setminus \{\mathbf{0}\}. \tag{13}$$

Then $\mathbb{E}_{\mathbf{z}} V = \boldsymbol{\alpha}$. Its covariance has a closed form.

Lemma 7 (Tilted covariance)**.** *For the law in equation (13),*

$$\Sigma(\boldsymbol{\alpha}) = \operatorname{diag}\left(\boldsymbol{\alpha} + \frac{\boldsymbol{\alpha}^2}{c}\right) - \frac{c-1}{c}\boldsymbol{\alpha}\boldsymbol{\alpha}^{\mathsf{T}}, \tag{14}$$

where vector squares are componentwise. This matrix is positive definite on $\mathcal{D}_r$.

Proof. Before conditioning away $\mathbf{v} = \mathbf{0}$, the coordinates are independent geometric variables with means $z_j/(1-z_j) = \alpha_j/c$. The probability of a nonzero vector is $1/c$. Conditioning and taking first and second moments gives equation (14). Positive definiteness follows because the supported vectors affinely span $\mathbb{R}^r$ and every supported vector has positive tilted mass. □

7 The fixed-prime-support local-limit law

Theorem 8 (Sharp ambiguity asymptotics)**.** *Fix r and a finite prime set $P = \{p_1, \ldots, p_r\}$. Let*

$$N_k = \prod_{j=1}^{r} p_j^{e_{j,k}}, \qquad \boldsymbol{\alpha}_k = \frac{\mathbf{e}_k}{k},$$

and suppose $\boldsymbol{\alpha}_k$ remains in a compact subset of $\mathcal{D}_r$. Let c_k, $\mathbf{z}_k$, and Σ_k be given by equations (10), (11) and (14). Then, uniformly on such compact subsets,

$$K_k(N_k) = \frac{F(\mathbf{z}_k)^k}{\mathbf{z}_k^{\mathbf{e}_k}(2\pi k)^{r/2}\sqrt{\det \Sigma_k}}\left(1 + \mathrm{O}(k^{-1})\right). \tag{15}$$

Equivalently,

$$L_k(N_k) = B_P(N_k, k) - \frac{r}{2}\log_2(2\pi k) - \frac{1}{2}\log_2 \det \Sigma_k + \mathrm{O}(k^{-1}). \tag{16}$$

Proof. Let $V_1, \ldots, V_k$ be independent copies of the law in equation (13). Coefficient extraction gives the exact change-of-measure identity

$$K_k(N_k) = \frac{F(\mathbf{z}_k)^k}{\mathbf{z}_k^{\mathbf{e}_k}} \mathbb{P}_{\mathbf{z}_k}\left\{\sum_{i=1}^{k} V_i = \mathbf{e}_k\right\}. \tag{17}$$

By construction, $\mathbb{E} V_i = \boldsymbol{\alpha}_k$, so the requested lattice point is exactly the mean of the sum.

The tilted law has exponential moments in a neighborhood of the origin because every $z_{j,k} < 1$. It is strongly aperiodic on $\mathbb{Z}^r$: the support contains each coordinate vector $\mathbf{e}_j$ and $2\mathbf{e}_j$. If its characteristic function had modulus one at a frequency $\mathbf{t}$, equality of the phases on these two support points would force $e^{it_j} = 1$ for every j.

Fourier inversion for the centered sum therefore has a unique dominant point at the origin. Uniformly on the stated compact subsets,

$$\log \mathbb{E} e^{i\mathbf{t}\cdot(V-\boldsymbol{\alpha}_k)} = -\frac{1}{2}\mathbf{t}^{\mathsf{T}}\Sigma_k\mathbf{t} + \mathrm{O}(\|\mathbf{t}\|^3).$$

Rescaling $\mathbf{t} = \mathbf{u}/\sqrt{k}$ gives the central Gaussian integral,

$$\mathbb{P}_{\mathbf{z}_k}\left\{\sum_{i=1}^{k} V_i = \mathbf{e}_k\right\} = \frac{1}{(2\pi k)^{r/2}\sqrt{\det \Sigma_k}}\left(1 + \mathrm{O}(k^{-1})\right).$$

The nominal $k^{-1/2}$ Edgeworth term is odd and vanishes at the centered lattice point; strong aperiodicity makes the complementary Fourier region exponentially smaller. Substitution into equation (17) proves equation (15). Taking base-two logarithms proves equation (16). □

This is a direct, explicit specialization of the multivariate lattice local-limit method [2, 3, 7]. The form matters here: the theorem identifies exactly which program-structure coordinates contribute the Gaussian penalty.

8 The corrected zeta residual

For a fixed exponent direction $\boldsymbol{\alpha}$, let

$$w = \sum_{j=1}^{r} \alpha_j \log p_j$$

and define the scalar natural-log rate

$$\beta_\zeta(w) = \inf_{s>1}\{sw + \log(\zeta(s) - 1)\}.$$

The multivariate rate is equation (12). Their difference in bits is

$$\delta(\boldsymbol{\alpha}, P) = \frac{\beta_\zeta(w) - \beta_P(\boldsymbol{\alpha})}{\log 2}.$$

Corollary 9 (Linear-plus-local residual)**.** *Under the fixed-direction hypotheses of theorem 8,*

$$B_\zeta(N_k, k) - L_k(N_k) = k\,\delta(\boldsymbol{\alpha}, P) + \frac{r}{2}\log_2(2\pi k) + \frac{1}{2}\log_2 \det \Sigma + \mathrm{O}(k^{-1}). \tag{18}$$

For every finite P in the interior regime, $\delta(\boldsymbol{\alpha}, P) > 0$.

Proof. Along a fixed direction, the scalar and multivariate rates are $\beta_\zeta(w)/\log 2$ and $\beta_P(\boldsymbol{\alpha})/\log 2$. Multiplying their difference by k and using equation (16) gives equation (18).

To see strict positivity, evaluate the multivariate partition function at $z_j = p_j^{-s}$. Then

$$F(p_1^{-s}, \ldots, p_r^{-s}) = \prod_{p\in P}(1 - p^{-s})^{-1} - 1 < \zeta(s) - 1.$$

The multivariate optimization is no larger than this restricted scalar evaluation. At the finite scalar minimizer, the displayed inequality is strict, so $\beta_P(\boldsymbol{\alpha}) < \beta_\zeta(w)$. □

The corollary precisely explains figure 3. The local-limit intuition was sound, but the scalar envelope first pays a linear rate for forgetting the exponent profile.

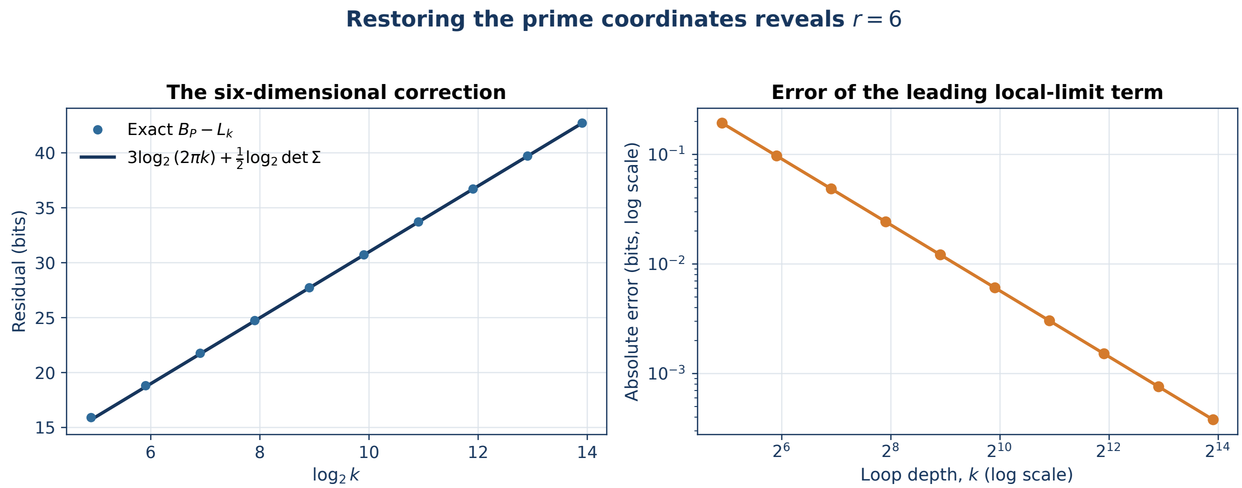

Figure 4: **The restored coordinates reveal the local law.** Left: exact $B_P - L_k$ is visually indistinguishable from the six-dimensional prediction. Right: the absolute error halves when k doubles, consistent with an $\mathrm{O}(k^{-1})$ logarithmic remainder.

Table 1: Exact fixed-direction residuals and the leading local-limit term.

k	$B_\zeta - L_k$	$B_P - L_k$	local error
30	24.1194	15.8999	0.193159
120	54.6333	21.7553	0.048509
480	159.2310	27.7189	0.012130
1,920	559.7584	33.7098	0.003032
7,680	2,143.9018	39.7075	0.000758
15,360	4,251.0957	42.7071	0.000379

9 Experiment II: recovering the dimension

For the fixed direction in equation (6), the saddle is

$$c = 1.5860207845$$

and

$$\mathbf{z} = (0.38669449,\ 0.20140776,\ 0.11198078,\ 0.07754850,\ 0.04033834,\ 0.04033834).$$

The covariance determinant is

$$\det\Sigma = 6.3769099904 \times 10^{-5},$$

with eigenvalues

$$0.06231,\ 0.06947,\ 0.13113,\ 0.20343,\ 0.42687,\ 1.29363.$$

All are positive, so the effective dimension is six.

Fitting $B_P - L_k = a\log_2 k + b$ over $k \geq 240$ gives

$$2a = 5.992933,$$

recovering $r = 6$ without supplying the dimension to the regression. The linear-plus-log fit to the zeta residual independently returns a linear coefficient 0.2739917 and logarithmic coefficient 2.97304, close to the theoretical 0.2739836 and $r/2 = 3$.

The theorem also permits controlled extrapolation beyond exact inclusion–exclusion. At $k = 10^9$, the multivariate local penalty is only 90.6782 bits, while the predicted zeta residual is $273{,}983{,}718.3$ bits. Nearly all of the latter is the scalar large-deviation mismatch.

10 Interpretation, boundaries, and scope

10.1 Which dimension is effective?

The effective dimension is the rank of the tilted covariance, not the loop depth and not the scalar dimension of $\log N$. In the interior fixed-support regime, every active prime density is positive and

$$\operatorname{rank} \Sigma = r.$$

This yields the $(r/2) \log_2 k$ penalty. Three boundary regimes require different asymptotics:

- If some $\alpha_j \to 0$, that prime coordinate disappears or enters a transition regime; a singular covariance cannot simply be inserted into equation (16).
- If $\sum_j \alpha_j \downarrow 1$, every factor asymptotically consumes exactly one prime occurrence. One linear constraint becomes deterministic, and the Gaussian rank is typically $r - 1$.
- If the active support grows, $r = r(k) \to \infty$, fixed-dimensional Fourier estimates no longer control the determinant or tails uniformly.

The growing-support case is the natural frontier for a sharper universal theorem. It is also where the scalar zeta envelope and prime-coordinate geometry must be reconciled rather than compared after the fact.

10.2 What the theorem does not say

The model concerns rectangular nests with constant integer bounds, known depth, known dimension order, and no early exits. It does not claim that every compiler transformation erases this much information: index arithmetic, memory accesses, symbolic expressions, provenance, and debug metadata may identify the source uniquely. Conversely, different source programs may be regarded as equivalent under a richer structural model.

Nor is $L_k(N)$ a computational hardness measure. It is a cardinality and coding quantity. Search may be easy even when the candidate class is large, and practical reconstruction may exploit evidence outside (N, k).

10.3 Padding and flattening are distinct

The equal-bound program in figure 1 deliberately illustrates a separate tradeoff. Replacing all bounds by their maximum can erase shape under a known model, but it adds work. Flattening the original rectangular domain can preserve work exactly while retaining ambiguity. Structural information loss and execution overhead are therefore independent axes; neither should be used as a proxy for the other.

11 Related work

Ordered factorizations. Ordered factorizations and their fixed-factor variants are established arithmetic objects. Sprittulla surveys combinatorial identities and average orders for exactly k factors [12]. Klazar and Luca study the maximal order of the unrestricted ordered-factorization function and the constant determined by $\zeta(\rho) = 2$ [4]. Our exact inclusion–exclusion formula is classical in substance. The present work assigns it an operational role as a transformation-specific ambiguity class and follows its coefficient asymptotics in a loop-source regime.

Multivariate coefficient asymptotics. Bender and Richmond developed multivariate central and local limit methods for asymptotic enumeration [2]; Hwang treated local large-deviation estimates via saddlepoint methods [3]. Pemantle and Wilson systematized multivariate coefficient asymptotics using the geometry of generating functions [7], and recent work makes central and local limit extraction increasingly effective [6]. Our theorem can be viewed as a particularly explicit power-of-a-generating-function case: conditioning nonzero geometric coordinates reduces the full saddle to one scalar c.

Program transformations and information. LLVM includes loop flattening as a canonical loop transformation [5]. Loop rolling and hardware rerolling reconstruct repeated structure from richer low-level evidence [8, 10]. Quantitative information flow supplies the broader language of residual uncertainty and channel capacity [11, 1]. Our hidden variable is source geometry rather than a secret value, and our observation channel is the product map.

12 Conclusion

Flattening a rectangular loop nest preserves a product and may erase a factorization. That elementary observation supports a complete quantitative story. Exact ordered-factorization counts measure the surviving source class; their logarithms give sharp worst-case reconstruction bits; and a zeta Dirichlet series yields a strong universal envelope without factoring the work count.

The remaining gap reveals the deeper geometry. A scalar work tilt cannot distinguish prime-exponent profiles, so its residual includes a linear large-deviation cost. Once the active prime coordinates are restored, the ambiguity becomes a centered multivariate lattice coefficient. Its saddle is explicit up to one scalar equation, its covariance identifies the effective dimension, and its residual is the expected Gaussian $(r/2)\log_2 k$ correction with a controlled remainder.

The sharp fixed-support theory is therefore settled. The next question is no longer whether a local correction exists, but whether it can be made uniform as the prime support grows. That is the point at which a genuinely sharper factorization-free theorem must either emerge or meet a structural obstruction.

A Additional exact identities

A.1 The ambiguity polynomial

The complete depth profile can be packaged as

$$\mathcal{A}_N(u) = \sum_{k=1}^{\Omega(N)} K_k(N)u^k.$$

With F from equation (7),

$$\mathcal{A}_N(u) = [\mathbf{z}^{\mathbf{e}}]\frac{uF(\mathbf{z})}{1-uF(\mathbf{z})}.$$

At $u=-1$,

$$\frac{-F}{1+F} = \prod_{j=1}^{r}(1-z_j) - 1,$$

so coefficient extraction gives the arithmetic signature

$$\sum_{k=1}^{\Omega(N)} (-1)^k K_k(N) = \mu(N),$$

where μ is the Möbius function. This is a useful consistency check for exact depth profiles.

A.2 Prime-power entropy profile

For $N = p^m$, corollary 3 gives

$$L_k(p^m) = \log_2 \binom{m-1}{k-1}.$$

Writing $q = (k-1)/(m-1)$ and applying Stirling's formula,

$$L_k(p^m) = (m-1)h_2(q) - \frac{1}{2}\log_2\bigl(2\pi(m-1)q(1-q)\bigr) + o(1),$$

where h_2 is binary entropy. The profile is symmetric and maximized near $k = (m+1)/2$. If depth is also erased,

$$\sum_{k=1}^{m} K_k(p^m) = 2^{m-1},$$

so exactly $m-1$ fixed-length bits are required to identify both the depth and the ordered exponent composition.

B Reproducibility and numerical method

The accompanying repository contains two independent experimental pipelines. The root-level uniform experiment generates the random-source work and exact-ambiguity data used in figures 1 and 2. The fixed-direction study contains the deterministic large-depth experiment, exact residual CSV, fitted constants, and figures 3 and 4.

Exact counts use theorem 2 with arbitrary-precision integers. For large k, the binomials are advanced by

$$\binom{e+q}{e} = \binom{e+q-1}{e}\frac{e+q}{q},$$

avoiding repeated construction from scratch. The scalar zeta saddle is solved by bounded one-dimensional minimization. The multivariate saddle uses bisection on equation (10); the covariance is then evaluated directly from equation (14). The accelerated exact recurrence was cross-checked against the original inclusion–exclusion implementation for every $2 \le k \le 40$.

References

[1] M. S. Alvim, K. Chatzikokolakis, A. McIver, C. Morgan, C. Palamidessi, and G. Smith. *The Science of Quantitative Information Flow*. Springer, 2020.

[2] E. A. Bender and L. B. Richmond. Central and local limit theorems applied to asymptotic enumeration II: multivariate generating functions. *Journal of Combinatorial Theory, Series A*, 34(3):255–265, 1983. https://doi.org/10.1016/0097-3165(83)90062-6.

[3] H.-K. Hwang. Large deviations of combinatorial distributions II: local limit theorems. *Annals of Applied Probability*, 8(1):163–181, 1998.

[4] M. Klazar and F. Luca. On the maximal order of numbers in the "factorisatio numerorum" problem. *Journal of Number Theory*, 124(2):470–490, 2007. https://arxiv.org/abs/math/0505352.

[5] LLVM Project. *LoopFlatten.cpp: flatten pairs of nested loops into a single loop*. LLVM source documentation, accessed July 2026.

[6] S. Melczer and T. Ruza. Central limit theorems via analytic combinatorics in several variables. *Electronic Journal of Combinatorics*, 31(2):P2.27, 2024. https://doi.org/10.37236/11732.

[7] R. Pemantle and M. C. Wilson. Asymptotics of multivariate sequences, part I: smooth points of the singular variety. *Journal of Combinatorial Theory, Series A*, 97(1):129–161, 2002. https://arxiv.org/abs/math/0003192.

[8] R. C. O. Rocha, P. Petoumenos, B. Franke, P. Bhatotia, and M. O'Boyle. Loop rolling for code size reduction. In *Proceedings of the 2022 IEEE/ACM International Symposium on Code Generation and Optimization (CGO)*, 2022.

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

[10] Z. D. Sisco, J. Balkind, T. Sherwood, and B. Hardekopf. Loop rerolling for hardware decompilation. *Proceedings of the ACM on Programming Languages*, 7(PLDI):420–442, 2023.

[11] G. Smith. On the foundations of quantitative information flow. In *Foundations of Software Science and Computational Structures (FoSSaCS)*, 2009.

[12] J. Sprittulla. Ordered factorizations with k factors. arXiv:1610.04826, 2016. https://arxiv.org/abs/1610.04826.