Collision bounds for UMASH: closing the projection gap

Thomas Dybdahl Ahle

September 2026

Abstract

UMASH combines carryless and integer multiplication, a polynomial accumulator, and an invertible finalizer. Its published analysis does not justify the projection of the compressor's two words modulo $2^{61}-1$: an almost-universal or XOR-universal bound cannot in general be multiplied by the maximum size of a projection fibre. We supply the missing analysis using exact congruence masks, binary lifting, integer and modular quadratic counts, and conditional product estimates. For ideal keys and a fixed seed, the implemented 64-bit hash has collision probability less than $58\lceil L/512\rceil/2^{61}$ on distinct strings of at most $8L$ bytes. A finer envelope has a certified length-normalized score of 56.18 bits. The C architecture's fingerprint, with two independent polynomial multipliers and shared compressor keys, satisfies $(81/128)\left\lceil L/2^{23}\right\rceil^2 2^{-83}$; its finer envelope scores 83.99 bits through 2^{46} words. These establish both published numerical headlines without asserting the disputed $162/2^{64}$ intermediate estimate. A separate block-swap example refutes the 83-bit claim for the Python reference's shared-multiplier fingerprint. Exact finite certificates and the precise boundary of Lean verification are recorded explicitly.

1 Model, implementation, and the resulting bounds

Put $q = 2^{64}$ and $p = 2^{61}-1$. A word is an integer in $[0, q)$; a pair of words is written $(z_{\mathrm{lo}}, z_{\mathrm{hi}})$. All messages and the common 64-bit seed are fixed before keys are chosen. Initially the 34 compressor key words are independent and uniform. We subsequently condition them to be distinct, obtaining uniform sampling without replacement. Polynomial multipliers are independent of these words and, for fingerprints, independent of each other. Unless specified otherwise their range is $\{2, \dots, p-1\}$. Every final bound also holds for the uniform range $\{1, \dots, p-1\}$ allowed by the C initializer. This last statement concerns the ideal uniform distribution on that range, rather than the distribution produced by the initializer's entropy expansion and key repair.

We analyze Khuong's literal arithmetic construction[4]. The relation to Apple's encode–hash–combine approach[3] is in the second compressor; no theorem about HalftimeHash is needed here. The mathematical proofs and finite certificates consolidated below are those of the accompanying corrected analyses and the earlier projection studies[1].

1.1 Compression and encoding

For word polynomials over $\mathbb{F}_2$, write $a \odot b$ for the unreduced carryless product. A nonfinal chunk (a, b) is compressed as

$$\mathrm{PH}(a, b; k, l) = (a \oplus k) \odot (b \oplus l).$$

For the final chunk use ordinary multiplication:

$$A = (a+k) \bmod q, \quad B = (b+l) \bmod q, \quad AB = qH + Y, \quad U = (H+\tau) \bmod q, \quad \mathrm{ENH}_\tau = (Y, U \oplus Y). \tag{1}$$

The XOR of the low word into the high word is part of the definition. It cannot be removed before modular projection. For a block of b bytes, $1 \le b \le 256$, its tag is $\tau = \mathrm{seed} \oplus (b \bmod 256)$. Two valid tags therefore have ordinary signed difference of magnitude at most 255.

A block has $n = \lceil b/16 \rceil \le 16$ expanded chunks. The primary compressor XORs its $n-1$ PH products and its final ENH value. Number chunks from zero. Let T shift each word left by one bit, separately, discarding overflow. The secondary compressor XORs the same ENH value, a shuffled version of each PH product, and

a checksum PH product. At PH position $i < n-2$ the shuffler is $T + T^{n-1-i}$; at $i = n-2$ it is T. Here addition of binary linear maps means XOR. Thus the possible maps are

$$S_1 = T, \qquad S_k = T + T^k \quad (2 \le k \le 15). \tag{2}$$

The checksum is the XOR of all data chunks and their original key pairs, including the final chunk before its additive treatment. The extra words at indices 32 and 33 key its PH product. The checksums have equal differences to the data checksums when chunk counts agree.

Messages longer than eight bytes use expanded chunks. A 9–15 byte message contributes its first and last eight bytes. Otherwise ordinary full chunks are used, and a final partial chunk is replaced by the last 16 bytes of the message. Grouping these chunks gives 256-byte blocks, with only the last block possibly shorter. The block count and final tag recover the byte length: the final size is the unique representative in $[1, 256]$ of its tagged residue. The overlapping reads then recover every byte. Consequently the ordered list of expanded block/tag tuples is injective on long messages.

1.2 The literal polynomial accumulator

If a compressor produces (a_i, b_i) for block i, the reference updates

$$z \longmapsto ((f^2 \bmod p)(z + a_i) + f b_i) \bmod (8p), \qquad z_0 = 0. \tag{3}$$

Modulo p, this evaluates a polynomial of degree at most $2n$ with coefficients $a_1, b_1, \ldots, a_n, b_n$ on powers $2n, 2n-1, \ldots, 1$. There is no constant term. The C implementation uses congruent unsigned arithmetic and the two stored pairs $(f_i^2 \bmod p, f_i)$ in `poly[2][2]`. The medium and long fingerprint paths use pair i for output i.

The long finalizer is $F(z) = z \oplus \mathrm{rotl}(z, 8) \oplus \mathrm{rotl}(z, 33)$. It is bijective. Indeed, if R denotes one-bit rotation over $\mathbb{F}_2$, then $N = R + I$ has $N^{64} = 0$. Since $1 + X^8 + X^{33}$ takes value one at $X = 1$, $F = I + NQ(N)$ for a polynomial Q. Its inverse is the finite geometric sum in $NQ(N)$. Actual hash equality thus implies equality of the polynomial evaluations modulo p. Section 7 also uses the stronger congruence modulo $8p$ retained in (3).

1.3 Statements and score convention

For a collision envelope $e(L)$ and output width b, define its score on a set $\mathcal{L}$ of positive word caps by

$$s_b(e; \mathcal{L}) = \inf_{L \in \mathcal{L}} \log_2 \frac{L}{\max(2^{-b}, \min(1, e(L)))}. \tag{4}$$

Both strings have at most $8L$ bytes. This score describes an upper envelope; it does not assert an attaining collision pair.

Theorem 1.1 (Primary hash)**.** *Set*

$$A_4 = \frac{205}{q - 561}, \qquad S_4 = \frac{435}{q - 561} + \frac{2}{p - 2}, \qquad \rho(L) = \min\left(1, \frac{2\lceil L/32 \rceil}{p - 2}\right).$$

For every positive integer L and distinct messages of at most $8L$ bytes, the primary collision probability is at most

$$I(1) = \frac{1}{q - 561}, \qquad I(L) = \min\{1, \max(A_4 + (1 - A_4)\rho(L), S_4)\} \quad (L \ge 2), \tag{5}$$

and is strictly less than $58\lceil L/512 \rceil / 2^{61}$. The score of (5) on all positive integer caps is $\log_2(2/S_4) = 56.1830163767\ldots$.

Theorem 1.2 (Two independent multipliers)**.** *Let*

$$C_0 = 3125, \quad C_1 = 729632, \quad J = 1416246956032, \qquad a = \frac{732757}{q - 561}, \quad b = \frac{J}{q(q - 561)}.$$

The actual fingerprint collision probability is at most

$$E(1) = \frac{1}{q(q - 561)}, \qquad E(L) = b(1 - \rho(L))^2 + a\rho(L)(1 - \rho(L)) + \rho(L)^2 \quad (L \ge 2). \tag{6}$$

It is strictly less than

$$\frac{81}{128}\left\lceil \frac{L}{2^{23}} \right\rceil^2 2^{-83}. \tag{7}$$

The score of (6) *is greater than* 83.99 *on* $L \le 2^{46}$*, greater than* 88.63 *on* $L \le 2^{23}$ *(64 MiB), and* $68.99999999999914\ldots$ *on* $L \le 2^{61}$*, covering the full 64-bit byte-length domain.*

The last score is displayed as 69.0 to one decimal place; it is slightly below the exact integer 69. Neither theorem assumes the original projection estimate $162/q$. For the generic field-root assembly, any block constant $C \le 255$ would already imply the published primary coefficient 64; 162 is a stronger intermediate target, not a necessary target. The special suffix argument below establishes the headline directly.

2 The projection gap and concrete counterexamples

The published calculation multiplies the raw $2/q$ compressor bound by $\lceil q/p \rceil^2 = 81$, obtaining $162/q$[4, pp. 6 and 19]. A raw equality bound alone controls neither arbitrary differences nor the union of events after a many-to-one map. Even XOR universality does not justify multiplying by a largest fibre size, since congruence modulo an odd number admits many different XOR differences.

The finite example recorded with the original gap analysis makes the distinction explicit. Take five equally likely keys, indexed by $u \in \{1, 2, 4, 8, 16\}$, and two 6-bit outputs $H(x) = u$, $H(y) = u + 31$. Their XOR differences are respectively $33, 35, 39, 47, 63$. Every fixed XOR target has probability at most $1/5$, but reduction modulo 31 makes the outputs equal with probability one. The largest fibre of that projection has size three; the proposed inference would incorrectly give $3/5$.

There is also a counterexample to a proposed uniform 128-key conditional PH count at the actual width. For

$$X(v) = u \odot v \oplus A, \qquad Y(v) = u' \odot v \oplus B,$$

exact division of candidate binary-polynomial differences gives the following numbers of free words v for which both projected words agree:

u	u'	A	B	Count
$2^{61}+1$	$2^{61}+3$	0	0	183
$2^{62}+1$	$2^{62}+3$	1	2^{61}	242

These slices are realizable through the literal compressor with a 32-byte message and seed 32, including ENH. They refute a uniform conditional $128/q$ assertion. They do not refute an averaged bound on all keys or either theorem above. Likewise, the scaled all-key example $[(0,0),(0,0)]$ versus $[(1,1),(0,0)]$ at $(w,p) = (8,31)$ has projected probability $138985472/2^{32}$; this is a fully specified finite example, not a general bound at width 64.

3 Congruence masks and exact signed digits

Define

$$\begin{aligned} D &= \{x \oplus y : 0 \le x, y < q,\ x \equiv y \pmod p\}, \\ P(d) &= \{(d + jp)/2 : -8 \le j \le 8,\ d + jp \ge 0,\ d + jp \text{ even},\ ((d + jp)/2)\ \&\ d = (d + jp)/2\}. \end{aligned} \tag{8}$$

For general width w and modulus p, replace eight by $\lfloor (2^w - 1)/p \rfloor$.

Lemma 3.1 (Exact pattern criterion)**.** *If* $d = x \oplus y$*, then* $x - y = 2(x\ \&\ d) - d$*. Thus* $x \equiv y \pmod p$ *if and only if* $x\ \&\ d \in P(d)$*.*

Proof. At each differing bit, subtraction contributes its weight positively when it belongs to x, negatively otherwise; equal bits cancel. This proves the identity. The difference is jp with $|j| \le 8$ exactly under (8). Conversely each submask $t \in P(d)$ is realized by $x = t$, $y = d \oplus t$, with arbitrary common bits outside d. □

Let $\mathcal{S}(n, w)$ be the support masks of representations of n by w digits in $\{-1, 0, 1\}$. The following recursion is complete, with base cases evaluated first:

$$\begin{aligned}
&\mathcal{S}(0,0) = \{0\}, \quad \mathcal{S}(n,0) = \varnothing \ (n \neq 0), \quad \mathcal{S}(n,w) = \varnothing \ (|n| \geq 2^w), \\
&\mathcal{S}(n,w) = \{2s : s \in \mathcal{S}(n/2, w-1)\} \quad (n \text{ even}), \\
&\mathcal{S}(n,w) = \{2s+1 : s \in \mathcal{S}((n-1)/2, w-1) \cup \mathcal{S}((n+1)/2, w-1)\} \quad (n \text{ odd}).
\end{aligned} \tag{9}$$

Indeed the least significant digit is forced to zero in the even case and is 1 or -1 in the odd case. Removing it proves soundness and completeness by induction. Negation preserves support, so $D = \bigcup_{j=0}^{8} \mathcal{S}(jp, 64)$. Evaluation with integers gives:

Quantity	Value
$\lvert\mathcal{S}(jp, 64)\rvert$, $j = 0, \ldots, 8$	$1, 184, 123, 361, 62, 355, 121, 178, 1$
$\lvert D\rvert$	852
Masks with $\lvert P(d)\rvert = 1, 2, 4, 8$	$1, 435, 357, 59$
$\sum_{d \in D} \lvert P(d)\rvert$	2771
Masks with bit 0 equal to 0, 1	248, 604
Masks with bit 63 equal to 0, 1	248, 604

The 248 masks with top bit zero include the zero mask: there are 247 nonzero such masks. Writing $N_r = \lvert D \cap 2^r\mathbb{Z}\rvert$ and $T_r = \sum_{d \in D \cap 2^r\mathbb{Z}} \lvert P(d)\rvert$, we have

r	0	1	2	3	$r \geq 4$
N_r	852	248	64	2	1
T_r	2771	615	127	3	1

(10)

An independent completeness check is

$$\sum_{d \in D} |P(d)| 2^{64 - \mathrm{wt}(d)} = q + 2 \sum_{j=1}^{8} (q - jp) = 8q + 72. \tag{11}$$

Each summand counts a disjoint set of ordered congruent pairs. The right side counts all such pairs independently. Equivalently, eight residue classes have nine representatives and the other $p - 8$ classes have eight. A sound list attaining this total is therefore complete.

Lemma 3.2. *$D \cap 16\mathbb{Z} = \{0\}$, and every nonzero $d \in D$ has its highest set bit at least 60. The only nonzero mask divisible by eight is $q - 8$; exactly 184 masks have valuation one and 62 have valuation two.*

Proof. If $16 \mid x \oplus y$, their low four bits agree, so $16 \mid x - y = jp$. Since p is odd and $|j| \leq 8$, $j = 0$, hence $x = y$. For a nonzero mask, $p \leq |x - y| \leq x \oplus y$, proving the highest-bit assertion. The remaining statements follow by subtracting adjacent entries of (10) and checking $q - 8 \in D$, for example with the congruent words $8p$ and zero. □

4 PH slices and the ENH carry and fold counts

For two final chunks, translate the first message's keys to independent uniform A, B, and write

$$A' = (A + \delta) \bmod q, \quad B' = (B + \eta) \bmod q, \quad AB = qH + Y, \quad A'B' = qH' + Y'.$$

When data differ, exchange operands so $\delta \neq 0$ and $r = \nu_2(\delta) \leq \nu_2(\eta)$, taking $\nu_2(0) = 64$. Put $R = 2^r$, $Q = q/R$. The valuation of an additive word difference equals the valuation of its XOR difference: both are determined by the lowest differing bit.

Lemma 4.1 (PH slices)**.** *A nonzero PH input difference has probability at most $1/q$ at any prescribed full XOR target. If its minimum coordinate valuation is s, its low XOR difference is uniform on $2^s\mathbb{Z}/q\mathbb{Z}$. With arbitrary fixed, possibly unequal, XOR offsets, projected PH collision has probability at most C/q, where $C = 604^2 = 364816$. If an input coordinate difference is odd, $17/q$ suffices.*

Proof. Expose the key opposite a nonzero coordinate difference d, conditioning the other key. The full difference has form $d \odot V \oplus c$, injective in V, since $\mathbb{F}_2[X]$ is an integral domain. If $d = X^s g$ with $g(0) = 1$, truncated multiplication by g is invertible modulo X^{64-s} by successive coefficient recovery. Its image is exactly the words divisible by 2^s, each with 2^s preimages; the affine constant has the same divisibility.

For odd d, fix a signed multiple jp, $-8 \le j \le 8$, as the integer low-output difference. At bit i, the borrow and all contributions from earlier bits are known, and the difference has coefficient one on V_i. Thus jp determines at most one V. For even d, bit zero of the full difference is fixed, and bit 127 is fixed because carryless products have degree at most 126. There are at most 604 eligible masks in each word. Each of the 604^2 full targets has at most one preimage. These are bounds on one injective slice, not independence assertions about its words. □

Lemma 4.2 (Low additive difference)**.** *$(Y' - Y) \bmod q$ is uniform on $R\mathbb{Z}/q\mathbb{Z}$, with mass R/q at each point. With any common low XOR offset, low projection equality has probability at most RT_r/q.*

Proof. Modulo q, the difference is $\delta B + \eta A + \delta\eta$. For fixed A, multiplication by δ has exactly R preimages per attainable value. Its offset is divisible by R. On a projected equality, $d = Y \oplus Y'$ is divisible by R and the projected first word has a pattern $t \in P(d)$. For a fixed common offset M, this specifies the additive target

$$Y' - Y = d - 2((M \,\&\, d) \oplus t) \pmod q.$$

Sum the R/q point bound over the T_r mask-pattern pairs. For $r = 0, 1, 2, 3$ the numerators are $2771, 1230, 508, 24$. □

Lemma 4.3 (Uniqueness of an ENH lift)**.** *Fix A, a wrap indicator $\beta = \lfloor (B + \eta)/q \rfloor$, and a mask-pattern pair (m, t) for high projection equality after a common XOR offset M. There is at most one B satisfying this condition and $Y = Y'$.*

Proof. Let $\alpha = \lfloor (A + \delta)/q \rfloor$, $d = (\delta - \alpha q)/R$, and $e = (\eta - \beta q)/R$. Then d is odd, $A' = A + Rd$, $B' = B + Re$, and

$$A'B' - AB = R(dB + eA + Rde).$$

On low equality, the high XOR mask is $m = U \oplus U'$. The first projected high pattern gives

$$U \,\&\, m = z(B) = ((AB) \,\&\, m) \oplus (M \,\&\, m) \oplus t.$$

This equality uses the fold in (1): high outputs differ from U, U' by the common $M \oplus Y$. With $\Delta\tau = \tau' - \tau$, a necessary congruence is

$$dB + eA + Rde \equiv Q(m - 2z(B) - \Delta\tau) \pmod q. \tag{12}$$

It follows first modulo qQ from the product identity and then modulo q. If two solutions first differ at bit $j < 64$, multiplication by A, masking, and fixed XOR preserve equality below bit j. Therefore $z(B_1) - z(B_2)$ is divisible by 2^j. The right-side difference in (12) is divisible by 2^{j+1}; the left-side difference has valuation exactly j because d is odd. This is impossible modulo 2^{j+1}. □

Corollary 4.4. *ENH-only projected primary collision, with arbitrary fixed common PH offsets and arbitrary tags, has probability at most $5542/q$ at every valuation.*

Proof. For $r \le 3$, use Lemma 4.2; all four numerators are at most 5542. For $r \ge 4$, projected low equality implies $Y = Y'$ by Lemma 3.2. On the collision event the map $(A, B) \mapsto (A, \beta, m, t)$ is injective by Lemma 4.3. Its codomain has at most $q \cdot 2 \cdot 2771$ elements. Division by q^2 proves the result, including zero operands, all wraps, and a zero second increment. □

Lemma 4.5 (Full additive product target)**.** *For $\delta \neq 0$ and arbitrary η, every prescribed value of $A'B' - AB$ modulo q^2 has probability at most $1/q$.*

Proof. Fix A, hence distinct A, A'. Within each B-wrap interval the difference is affine with nonzero slope $A' - A$, and distinct values differ by a nonzero integer of magnitude below q^2. Across the wrap, for $B_1 < B_2$, the difference of the two values is

$$(A' - A)(B_2 - B_1) - qA'.$$

It lies strictly between $-q^2$ and zero, including $A = 0$ or $A' = 0$. Thus at most one B hits a residue across both intervals together. Average over A. □

Lemma 4.6 (Fresh points and tag-only comparisons)**.** *A fresh ENH pair after any fixed tag and XOR offset has projected point mass at most* $(82q-81)/q^2 < 82/q$*. If expanded data agree and valid tags differ, primary projected collision has probability less than* $1/(8q)$*.*

Proof. A projected pair has at most 81 raw representatives. Undoing the offset, fold, and tag is bijective. At most one product target is zero, with $2q-1$ operand pairs; each nonzero target has at most $q-1$. The count is at most $(2q-1)+80(q-1)$.

For the second assertion let $d = \tau' - \tau$, $0 < |d| \le 255$. A nonzero high mask $m \in D$ has highest bit $k \ge 60$. Moving by signed d around the word circle must cross one of the 16 boundaries at multiples of 2^{60}. There are at most $16|d| \le 4080$ possible H. Put $g(m) = 2^{k+1} - m$. Every signed difference with XOR mask m has absolute value, and circular distance, at least $g(m)$: lower bits cannot cancel the highest bit, and $q - m \ge g(m)$. Hence $g(m) \le 255$, so bits $8, \dots, k$ and at least one lower bit are set. Thus $\mathrm{wt}(m) \ge 54$. For each H, at most $8 \cdot 2^{10} = 8192$ low words satisfy the high pattern criterion. Product zero cannot qualify because its mask is $\tau \oplus \tau' \le 255$.

Every integer $1 \le n < 2^{128}$ has fewer than 2^{36} divisors. To see this, for primes $\ell \ge 17$ use $(a+1)^4 \le 16^a \le \ell^a$. For $\ell = 2, 3, 5, 7, 11, 13$, the maxima of $(a+1)^4/\ell^a$ on $0 \le a \le 127$ are

$$81/2, \quad 256/27, \quad 81/25, \quad 16/7, \quad 16/11, \quad 16/13.$$

Their product is $127401984/25025 < 16^4$; these six finite maxima are checked by integer comparisons. Consequently $d(n)^4 < 16^4 n$ and $d(n) < 2^{36}$. Each nonzero product target has at most $d(n)$ ordered factorizations. The total collision count is less than $4080 \cdot 8192 \cdot 2^{36} = 255 \cdot 2^{53} < q/8$. Divide by q^2 and average any independent common PH offset. □

5 Quadratic counts and the constants 1123 and 3125

The counts in this section are uniform over fixed messages; they do not assert extremality of an experimentally chosen family.

Lemma 5.1 (Integer quadratic intervals)**.** *If* $F(t) = at^2 + bt + c$ *has integer coefficients and* $a \ne 0$*, and* J *is a union of at most* m *disjoint half-open intervals of length at most* ℓ*, then*

$$\#\{t \in \mathbb{Z} : F(t) \in J\} \le 2\sqrt{m\ell/|a|} + 2m.$$

On one monotone side of the vertex, the bound is $\sqrt{m\ell/|a|} + m$*.*

Proof. Complete the square. On either side, the inverse is a translated square root scaled by $|a|^{-1/2}$. Moving disjoint intervals left to remove gaps increases their square-root lengths because $\sqrt{x+s} - \sqrt{x}$ decreases with x. The lengths then telescope to at most $\sqrt{m\ell}$. There are at most m preimage intervals on each side. An interval of length d contains at most $d+1$ integers. This also covers an endpoint at the vertex. □

Lemma 5.2 (Bits of a modular quadratic)**.** *Let* X *be uniform modulo* 2^n*, and* $F(X) = bX^2 + cX + d$ *with* b *or* c *odd. Any specified assignment to* h *selected output bits has probability at most* $\min(1, 4 \cdot 2^{-\lfloor h/2 \rfloor})$*.*

Proof. If c is odd, each root modulo two lifts uniquely at every subsequent bit: changing X by 2^j changes F by 2^j modulo 2^{j+1}. Thus every output has at most two preimages, giving 2^{1-h}.

Otherwise b is odd and c is even. Complete the square modulo 2^n to obtain $b(X+a)^2 + d'$. Conditional on $\nu_2(X+a) = t$ with $2t < n$, the odd part squared is uniform on the odd squares. At width at least three these are precisely the residues 1 modulo eight, with four roots each. This follows by induction: among four roots modulo 2^j, translating a root by 2^{j-1} changes the next square bit, and each selected root has two lifts. Smaller widths have a fixed odd square. Thus F is uniform on a coset modulo $2^{\min(n, 2t+3)}$, and the conditional assignment probability is at most $2^{-\max(0, h-2t-3)}$.

The valuation mass is 2^{-t-1}. For $h = 2m$, summing $t < m-1$ and bounding the remaining mass gives at most $4 \cdot 2^{-m} - 2^{-2m+2} \le 4 \cdot 2^{-m}$. For $h = 2m+1$ the analogous sum is at most $3 \cdot 2^{-m}$. Cases $h \le 3$ and $2t \ge n$ are covered by the trivial bound and this tail mass. □

Lemma 5.3 (High and low ENH target counts)**.** *For two nonzero ENH increments and* $h = \mathrm{wt}(e)$*, write* e_{63} *for the top bit. Then*

$$\Pr[Y = Y',\ U \oplus U' = e] \le K_H(e)/q,$$
$$K_H(e) = \min\{2^{1+h-e_{63}},\ 276\,2^{64-h},\ 16(2^{\lceil h/2 \rceil} + 1)\}. \tag{13}$$

For minimum valuation $r \geq 1$ and $2^r \mid e$,

$$\Pr[Y \oplus Y' = e] \leq K_{L,r}(e)/q, \quad K_{L,r}(0) = 2^r,$$
$$K_{L,r}(e) = \min\{2^r 2^{h-e_{63}},\ 65\,2^{64-h},\ 4 \cdot 2^r 2^{\lceil h/2 \rceil}\} \quad (e \neq 0). \tag{14}$$

Proof. For the quadratic term in (13), fix $z = U \,\&\, e$, one of 2^h patterns. Low equality specifies

$$A'B' - AB \equiv q(e - 2z - \Delta\tau) \pmod{q^2}.$$

There are at most two integer representatives in $(-q^2, q^2)$ and four operand-wrap rectangles. In each, the nonzero signed increments $a = A' - A$, $b = B' - B$ impose $aB + bA + ab = c$. If soluble, all integer points are $A = A_0 + (a/g)t$, $B = B_0 - (b/g)t$, where $g = \gcd(|a|, |b|)$. Their product is a quadratic with nonzero integer highest-degree coefficient $-ab/g^2$. The pattern on U permits 2^{64-h} high words, hence that many product intervals of length q. Lemma 5.1, summed over patterns, representatives, and rectangles, gives $16(2^{h/2} + 1)/q$.

The sparse term counts at most $2^{h-e_{63}}$ additive high differences modulo q; the elementary separate-wrap product count is at most $2/q$ per difference. For the dense term, $U + U' = e + 2(U \,\&\, \neg e)$ gives at most 2^{64-h} sums modulo q. Each permits two integer values of $H + H'$. On low equality, $AB + A'B' = q(H + H') + 2Y$ lies in an interval of length $2q$. Fixing A and a B-wrap gives slope $A + A' > 0$, hence at most $2q/(A + A') + 1$ points. Since $1/(A + A') \leq (1/A + 1/A')/4$ when both are positive, and the two possible zero-factor positions contribute at most two, the sum of these reciprocals is at most 34 using the 64 dyadic harmonic intervals. Two wraps and two high sums therefore cost at most $4(2q \cdot 34 + q) = 276q$.

For (14), write $\delta = Ra$, $\eta = Rb$, a odd, and change variables bijectively from (A, B) to (A, C) with

$$C = aB + bA + Rab \pmod{q}, \qquad aY = -bA^2 + (C - Rab)A \pmod{q}. \tag{15}$$

Fix $t = Y \,\&\, e$. The XOR event requires $C \equiv (e - 2t)/R \pmod{q/R}$, with R lifts. For $e \neq 0$, let $v = \nu_2(e)$. Every lift has valuation $v - r$. The common valuation of b and $C - Rab$ is $s = \min(\nu_2(b), v - r)$. After division by 2^s, at least one quadratic coefficient is odd. All h selected bits survive because $v \geq s + r > s$. Lemma 5.2 bounds the number of A for each lift and pattern by $4q2^{-\lfloor h/2 \rfloor}$. Summing gives the third term. The additive low distribution and the $2^{h-e_{63}}$ possible differences give the first term, and give R/q for zero.

Finally,

$$(Y + Y')/2 \equiv (A + \delta/2)(B + \eta/2) + \delta\eta/4 \pmod{q/2}.$$

An XOR mask of weight h permits at most 2^{64-h} half-sums. For two independent residues modulo 2^n, a product has point mass at most $(n + 2)/2^{n+1}$: each first-factor valuation contributes at most 2^{-n-1} and the zero first factor contributes 2^{-n}. With $n = 63$ this is $65/q$, proving the middle term. □

Lemma 5.4 (Truncated low counts)**.** *At width $n \geq 4$, put $M = 2^n$ and $Z = AB \oplus ((A + \delta)(B + \eta))$, with products reduced modulo M and minimum increment valuation $r < n$. Then:*

condition	*bound*
$r = 0$	*Z uniform on all M words*
$Z = 0$	$\Pr[Z = 0] = 2^r/M$
$r \geq 1,\ \nu_2(e) = r$	$\Pr[Z = e] \leq 2^{r+1}/M$
$r = 1,\ \nu_2(e) = 2$	$\Pr[Z = e] \leq 5/M$
δ, η *even*	$\Pr[Z = M - 8] \leq (n + 12)/M$

The second increment may be zero.

Proof. For $r = 0$, condition on A. At each successive bit of B the two products have opposite coefficients on the new bit; all addition carries are already determined. This is a triangular bijection. The zero assertion is Lemma 4.2. For $\nu_2(e) = r \geq 1$, use (15) at width n. Each of the R lifts C is odd, so the linear coefficient is odd. For each Y there are at most two roots per lift, giving $2R/M$.

For $r = 1$, $\nu_2(e) = 2$, C has valuation one. If b is even, Y must be even; division by two gives at most four A modulo M for each of two lifts, giving $4/M$. If b is odd, A, B have the same parity. The even-even part, after division by four, has odd first increment and contributes $1/M$. In the odd-odd part the completed square has an odd root, with at most four possibilities. Indeed $z^2 = 1$ forces one of $z - 1, z + 1$ to have valuation one, giving at most four residues. Counting $M/2$ odd Y and two lifts contributes at most $4/M$.

For the dense mask, $(Y + Y')/2$ has eight consecutive cyclic targets modulo $K = M/2$, after a fixed translation. Among eight such residues, four are odd, two have valuation one, one has valuation two,

and one is divisible by eight. A nonzero product target of valuation v has probability $(v+1)/M$; zero has probability $(n+1)/M$, by the valuation sum used above. The eight targets therefore cost at most $(4+4+3+n+1)/M = (n+12)/M$. □

Theorem 5.5 (PH bound 1123)**.** *If at least one PH chunk differs in two same-count blocks, primary projected collision has probability at most* $1123/q$*, for arbitrary other data and tags.*

Proof. Let s be the minimum nonzero PH coordinate valuation. The sum of all PH low differences is uniform on $2^s\mathbb{Z}/q\mathbb{Z}$: expose a key attaining s; every other difference and the affine constant lie in the same subspace. This uniform variable is independent of the ENH difference $E = Y \oplus Y'$. With $M = 2^s$,

$$q \Pr[Z \oplus E \in D] = M \sum_{d \in D} \Pr[E \bmod M = d \bmod M]. \tag{16}$$

For $s = 0$ use $17/q$. For $s = 1, 2, 3$, count masks by their low s bits, $n_s(c)$. For each pair of increment residues, the numerator in (16) is

$$\frac{1}{M} \sum_{a,b=0}^{M-1} n_s\big(ab \oplus ((a+\delta)(b+\eta)) \bmod M\big),$$

where each product is truncated before XOR. Enumerating all $4 + 16 + 64 = 84$ increment pairs gives integer sums $1704, 3408, 6816$ before division by M, hence numerator 852 in each width.

For $s \geq 4$ and both ENH increments divisible by 16, the total low mask is divisible by 16 and must vanish. After conditioning ENH and all but one nonzero PH slice, at most 604 high masks remain: bit 63 of the conditioned ENH high XOR difference fixes their top bit. Injectivity gives $604/q$. Otherwise the minimum ENH valuation r is below four. Apply Lemma 5.4 at width s in (16); the complete numerator counts are

r	0 mask	$\nu_2(d) = 1$	$\nu_2(d) = 2$	$d = q - 8$	total
0	uniform truncated difference				852
1	2	$184 \cdot 4$	$62 \cdot 5$	$s + 12$	$1060 + s \leq 1123$
2	4	0	$62 \cdot 8$	$s + 12$	$512 + s \leq 575$
3	8	0	0	16	24

Every nonzero even mask has valuation at most three, so it remains nonzero on truncation. The table counts full masks even if their prefixes repeat; this is exactly the sum in (16). All cases are exhausted. □

Lemma 5.6 (Classes of lifting functions)**.** *For* $r \geq 2$*, the function* $f_{m,t}(l) = m - 2((l \,\&\, m) \oplus t) \pmod{2^r}$ *is determined exactly by*

$$(m \bmod 2^{r-1},\ (m - 2t) \bmod 2^r). \tag{17}$$

A common input XOR translation preserves these equivalence classes. For (m, t) *from* (8)*, their number is*

$$F_4 = 36, \qquad F_r = 26r - 75 \quad (5 \leq r \leq 63).$$

Proof. As an affine function of separate bits l_j, the constant is $m - 2t$ and the coefficient at a set bit j of m is $\pm 2^{j+1}$. Coefficients for $j \geq r - 1$ vanish. At $j = r - 2$ the signs coincide; below that, their signs follow from $2t \equiv m - (m - 2t) \pmod{2^{r-1}}$. Thus (17) determines the function. Evaluation at zero and basis inputs recovers the constant and all set bits below $r - 1$, proving the converse. XOR translation is a bijection. For the count, enumerate every pair in (8), reduce its two labels, and count distinct pairs at each of the sixty widths $4, \ldots, 63$. The independent certificate computes instead the constant and all affine coefficients; the two partitions coincide. This is a finite calculation at every stated width, not an extrapolation of the affine formula for F_r. □

Theorem 5.7 (ENH bound 3125)**.** *With common PH data and different final data, primary projected collision is at most* $3125/q$*, for arbitrary fixed common offsets and tags. At* $r \geq 4$ *the numerator is at most* $\min(2^r, 2F_r - 1)$.

Proof. At $r \leq 3$ use Lemma 4.2. For $r \geq 4$, projection forces $Y = Y'$. In (12) only f modulo R is used because $QR = q$. Thus all pairs in a class give the same equation, including after a common XOR translation. The uniqueness proof applies to a class, allowing at most one B per A, wrap, and class.

The zero label contains only $(m, t) = (0, 0)$: its constant requires $R \mid jp$, $|j| \leq 8$, so $j = 0$; then $m = 2t$ and $t \,\&\, m = t$ force $m = t = 0$. This class implies a fixed product difference modulo q^2 and costs only q pairs by Lemma 4.5. The other $F_r - 1$ classes cost at most $2q$ each. Hence the numerator is $2F_r - 1 \leq 3125$. Separately, raw low equality has probability $2^r/q$. All common PH keys can be conditioned first and averaged last. □

Together with Lemma 4.6, the disjoint block cases (different counts; a PH change; only ENH data change; only a valid tag change) have constants $82, 1123, 3125, 1/8$. Their maximum is 3125. This bound will supply the primary marginal for the fingerprint, independently of the sharper primary argument that follows.

6 Sharper primary counts: 151 and 205

6.1 Small ENH valuations

Lemma 6.1. *At width $n \geq 4$, if an XOR target has valuation equal to the minimum increment valuation $r \geq 1$, its probability is exactly $2^r/2^n$. At $r = 1$ a valuation-two target costs at most $3/2^n$. The dense target $2^n - 8$ costs at most $4/2^n$ for $r = 1$ and $6/2^n$ for $r = 2$.*

Proof. Use (15) with $M = 2^n$. For a valuation-r target all lifts C are odd. If b is even, the quadratic is a permutation: its linear coefficient is odd, its quadratic coefficient even, and each bit lifts uniquely from the unique root modulo two. If b is odd, there are two roots for every even target and none for an odd target. Each possible even Y still has R lifts. Both cases give exactly RM operand pairs.

At $r = 1$ and target valuation two, if b is even, division by two gives either at most two A for even Y, or four for Y divisible by four, according as $b/2$ is even or odd. Two lifts C give at most $2M$ pairs. If b is odd, the even-even contribution is $1/M$ as before. On odd-odd pairs, $c = C - 2ab$ is divisible by four and

$$(bA - c/2)^2 = (c/2)^2 - abY \pmod{M}.$$

Write the target as $4v$, v odd, and $Y = 1 + 2y_1 + 4y_2 \pmod 8$. Since $C = 2v - (Y \,\&\, (4v)) \pmod{M/2}$, the right side modulo eight is $(v - ab)^2 - ab(1 + 2y_1)$. Only one value of y_1 gives an odd square, which must be one modulo eight. Thus at most $M/4$ odd Y, two lifts, and four roots contribute $2/M$; the total is $3/M$. This calculation also holds at $n = 4$.

For the dense target set $K = M/2$, $u = \delta/2$, $v = \eta/2$, $T = A + u \pmod K$, $S = B + v \pmod K$. The product TS lies in eight consecutive residues and also $uS + vT = 4 \pmod 8$. Put $H_0 = 2^{4-r}$, $h = 4 - r$; dividing gives $aS + bT = H_0/2 \pmod{H_0}$ with a odd. For each $T = t \pmod{H_0}$ there is one $S = s \pmod{H_0}$, and $8/H_0$ permitted product targets. If $t \neq 0$, fixing T permits at most $2^{\nu_2(t)}$ values of S per target; if $t = 0$, fix $S = H_0/2 \pmod{H_0}$ instead and use 2^{h-1}. Since

$$\sum_{t=1}^{H_0-1} 2^{\nu_2(t)} + 2^{h-1} = (h+1)H_0/2,$$

the probability is at most $8(h + 1)/(H_0M)$, giving four or six over M. □

Lemma 6.2 (Small-valuation block counts)**.** *If PH data agree, or every nonzero PH coordinate difference is divisible by 16, the primary bounds at ENH valuations $r = 0, 1, 2, 3$ are respectively*

$$17/q, \qquad 19/q + 64/q^2 < 20/q, \qquad 14/q + 32/q^2 < 15/q, \qquad 16/q.$$

The $r = 0$ bound allows arbitrary PH differences.

Proof. For $r = 0$, condition the PH keys and A; successive bits of B, including known carries and subtraction borrows, give at most one solution per signed target jp. For the other cases the two conditioned PH low offsets agree modulo 16. There are $2q + 16$ congruent ordered word pairs whose XOR valuation is one (differences jp with $j = \pm 2, \pm 6$), and $q + 8$ of valuation two ($j = \pm 4$). Their first-word residues are evenly distributed modulo two or four as used below.

At $r = 1$, valuation-one pairs have at most two ENH preimages when b is even, or four restricted to even Y when b is odd. Either costs $4/q + 32/q^2$. For valuation-two pairs with b even the costs are at most four preimages on even Y, or eight on $4 \mid Y$, giving $2/q + 16/q^2$. With b odd, odd-odd pairs have at most eight preimages and occupy half the $q + 8$ output pairs, giving $4/q + 32/q^2$. Bound the entire even-even projected event separately: dividing products by four leaves an odd-first-increment problem and fixed low two output bits allow at most five signed j, giving $5/q$ unconditionally. Zero and $q - 8$ contribute $2/q$ and $4/q$, by Lemma 6.1 and the per-mask version of (16). The worse total is $19/q + 64/q^2$.

At $r = 2$, valuation-two pairs cost $4/q + 32/q^2$, and zero and $q - 8$ cost four and six over q. At $r = 3$ only zero and $q - 8$ remain, each costing $8/q$. This argument permits a zero second increment and overcounts some even-even events harmlessly. □

6.2 Unused randomness in a PH pair

Lemma 6.3 (Conditional PH product)**.** *Let both PH coordinate differences be nonzero, $a = gu$, $b = gv$ in $\mathbb{F}_2[X]$, with $g = \gcd(a, b)$ and $t = \max(\deg u, \deg v)$. Put $N = 64 - t$, $h = uv$. For independent word polynomials U, V, the difference is gT where $T = uV + vU + gh$ is uniform on degree $< 64 + t$. Conditional on T, the first product is a fixed polynomial plus*

$$SZ + hZ^2, \qquad S = T + gh, \qquad \deg Z < N, \tag{18}$$

with Z uniform; $\deg h \le 2t$ *and* $\nu_2(h) \le t$.

Proof. Coprimality gives kernel $(U, V) = (uZ, vZ)$ for the map $uV + vU$. The word restrictions are exactly $\deg Z < 64 - t$, so the rank is $64 + t$, the dimension of the entire possible image. Expanding $(U_0 + uZ)(V_0 + vZ)$ gives (18). One of u, v is odd, proving $\nu_2(h) \le t$. If only one coordinate changes, the separate conditional model is simply $U \odot V$ with V fixed and U uniform. □

Here is a complete specification of the finite ranks used next. In an output row n of (18), the coefficient of input bit i is $S_{n-i} + h_{n-2i}$. For a low row $n < 64$, let $v = \nu_2(S)$ and e_0, e_1 be the least even and odd set positions of h, with absent positions infinite. Only retain $n < N + v$. The largest possible input indices are $n - v$ and $(n - e_{n \bmod 2})/2$ when defined. Discard equality; otherwise the larger is an uncancelled extreme. For $t \le 11$ the square indices here are below N. For a high row $n = 64 + j$, let $d = \deg S$ and H_0, H_1 be the greatest even and odd set positions of h. Discard $n < d$ and indices outside $[0, N)$. The smallest indices are $n - d$ and $(n - H_{n \bmod 2})/2$; discard equality and otherwise retain the smaller. Rows with distinct extremes are linearly independent: the extreme in a nontrivial linear combination cannot cancel. If the selected mask rows have certified rank k, any translated set of acceptable patterns has probability at most $\min(1, |P(m)|2^{-k})$.

Lemma 6.4 (PH finite tables)**.** *The primary bound for a PH change is* $151/q$. *In the case of minimum PH valuation $s \in \{1, 2, 3\}$ and even ENH increments it is less than* $147/q$.

Proof. We specify the complete maximizations so that the numerical assertions are reproducible independently. Select a PH chunk attaining s and condition all other original keys. Let C be the conditioned low XOR offset; it is even. A low target mask m must agree with C below bit s.

For a two-coordinate change and $m \ne C$, put $k = \nu_2(m \oplus C)$, $v_T = k - s$. The target determines T modulo X^{64-s}, with 2^{s+t} possible full values, each of mass 2^{-64-t}. If $b = \nu_2(h)$, the possible valuations of S are $\min(v_T, s + b)$ when unequal, and $s + b + 1, \ldots, 64 + t$ when equal, the endpoint including zero. For $t \le 9$, enumerate the least even/odd positions in $[0, 2t]$, either possibly absent, with minimum at most t. Take the smallest low-row rank above over all these choices and valuations. Let $w(m, k)$ be its clipped pattern bound, and assign weight one to $m = C$. The required numerator is at most

$$2^s \max_{C \text{ even}} \sum_{m \in D,\ m \equiv C \ (2^s)} \begin{cases} 1 & m = C, \\ w(m, \nu_2(m \oplus C)) & m \ne C. \end{cases} \tag{19}$$

This is a finite maximum over all centers C, not selected examples. At bit k, split the masks sharing the chosen prefix into I_0, I_1. The recursion is

$$V(I, k) = \max_{b=0,1} \left(\sum_{m \in I_{1-b}} w(m, k) + V(I_b, k + 1) \right).$$

An empty set contributes zero; an occupied leaf at bit 64 contributes one. Starting at each even prefix of length s gives every center. All $s = 1, 2, 3$, $t = 0, \ldots, 9$ entries in (19) are strictly below 145. For $t \ge 10$, the full difference rank $64 + t$ and the 248 even low masks and 604 eligible high masks give $248 \cdot 604/2^t \le 4681/32 < 147$.

For one-coordinate changes, conditional on $v = \nu_2(V)$, the low product bits at and above v are independent uniform bits. For $m \ne C$, the target fixes $v = \nu_2(m \oplus C) - s$. Assign weight one to $m = C$ and use in the same recursion

$$w_L(m, v) = \frac{\max_z \#\{t \in P(m) : t \bmod 2^v = z\}}{2^{\mathrm{wt}(m \gg v)}}. \tag{20}$$

The resulting maxima for $s = 1, 2, 3$ are $78.75, 81.125, 84.53125$.

If all PH and ENH low differences are divisible by 16, a projected low equality is raw equality. For a two-coordinate change with $t \geq 2$, at most 604 full targets remain, costing $604/2^t \leq 151$. For $t = 0, 1$, let $D_0 = \deg g$, $4 \leq D_0 \leq 63 - t$. The possible h are 1 for $t = 0$ and $X, 1 + X, X + X^2$ for $t = 1$. For conditioned high offset C and $m \neq C$, set $j = \text{bitlength}(m \oplus C)$. Necessarily

$$\deg T = 64 - D_0 + j - 1.$$

If this is at least $64 + t$, discard the target. Otherwise $\deg S$ is the larger of this degree and $D_0 + \deg h$ when unequal, and any of $-1, \ldots, D_0 + \deg h - 1$ when equal. Minimize the high-row rank over the listed possibilities. Maximize the corresponding pattern sum by the same recursion read from bit 62 down to zero, with either fixed top bit, and divide by 2^t. All 119 pairs (t, D_0) give a maximum below 27 (the largest is 26.25).

For a one-coordinate change, the determined nonzero V has degree $64 - \deg d + j - 1 \geq j$. Its high product is uniform on its lowest $\deg V$ bits. Use

$$w_H(m, j) = \frac{\max_z \#\{t \in P(m) : t \gg j = z\}}{2^{\text{wt}(m \& (2^j - 1))}}. \tag{21}$$

This decreases with j: freeing a selected bit at most doubles the numerator and doubles the denominator. The high recursion gives maxima 8 and 11.5 for the two top bits.

These computations use dyadic integers and all masks of Section 3; an independent calculation forms the literal rows from basis monomials and enumerates every occupied leaf and every first empty branch of the center recursion. They give the same maxima. Finally an odd PH change costs $17/q$; an odd ENH change also costs $17/q$; $s \geq 4$ with ENH valuation 1–3 uses Lemma 6.2. The remaining cases are exactly the low and high tables just proved, completing the $151/q$ bound. □

6.3 High ENH counts with a common PH product

The quadratic high bound has a useful strengthening:

$$K(e) = \min\{2^{\text{wt}(e) - e_{63}},\ 276\, 2^{64 - \text{wt}(e)},\ 6(2^{\lceil \text{wt}(e)/2 \rceil} + 1)\}. \tag{22}$$

Then $\Pr[Y = Y', U \oplus U' = e] \leq K(e)/q$ for two nonzero increments. The first term now uses Lemma 4.5. For the last term, a wrap rectangle with signed increments a, b has side lengths $q - |a|, q - |b|$. Its product-difference range has length less than $|a|(q - |b|) + |b|(q - |a|) \leq q^2$, so it permits only one integer representative. In the two same-sign rectangles use both sides of the quadratic vertex; in the two opposite-sign rectangles both operands move in the same direction and their nonnegative product is monotone. The interval count is therefore $2 + 2 + 1 + 1 = 6$, proving (22).

For one changed word and valid tags use, with $g(e) = 2^{\text{bitlength}(e)} - e$,

$$\kappa(e) = \min\left\{2^{\text{wt}(e)},\ 5 + 2^{71 - \text{wt}(e)},\ 5 + \left\lceil \frac{4q}{g(e) - 255} \right\rceil \text{ if } g(e) > 255\right\}. \tag{23}$$

Here is a proof of its three branches. Low equality forces $B = mq/2^r$, $0 \leq m < 2^r$. In either A-wrap interval the tagged high additive difference is $((\delta - \alpha q)/2^r)m + \Delta\tau \pmod q$, with odd coefficient. Each specified difference therefore costs at most $1/q$ after summing interval lengths. There are at most $2^{\text{wt}(e)}$ possible differences. For a nonzero B and a fixed interval, $2z = e - s \pmod q$ has at most two submask solutions, allowing at most $2^{65 - \text{wt}(e)}$ high words; each has at most $\lceil q/B \rceil$ preimages A. Summing the two intervals, $m \geq 1$, and $B = 0$ gives at most $[1 + 4 \cdot 2^{r - \text{wt}(e)}(H_{2^r - 1} + 1)]/q \leq [1 + 2^{71 - \text{wt}(e)}]/q$. Finally circular distance at least $g(e)$ and tag gap at most 255 require $B > g(e) - 255$. Enumerating the high patterns and their unique B in each interval gives at most $2\lceil q/(g(e) - 255) \rceil / q$. Each estimate is no larger than the corresponding branch of (23).

Theorem 6.5 (ENH with a common PH chunk)**.** *If PH data agree, at least one PH chunk exists, and final data differ, primary projected collision has probability less than* $205/q$*. For one changed final word its high-valuation part is less than* $134/q$*.*

Proof. One independent common PH product supplies the high-pattern weight

$$W(m) = \min\left(1, \frac{1}{q} + \sum_{j=0}^{63} \frac{2^j}{q} w_H(m, j)\right). \tag{24}$$

Indeed its first factor is zero with probability $1/q$, and has degree j with probability $2^j/q$. In the latter case its high product is uniform on its lowest j bits and zero above, by distinct extreme rows. The numerator of (21) counts the compatible upper patterns after any fixed XOR translation.

For $r \le 3$ use Lemma 6.2; for $4 \le r \le 7$, raw low equality costs at most $128/q$. At higher valuations low equality is necessary. For each nonzero m, first conditioning the ENH keys gives $K(m)W(m)/q$, or $\kappa(m)W(m)/q$ for one changed word. Independently, conditioning PH first and using lifting gives $2|P(m)|/q$. Take the smaller of these two justified counts for each mask. The zero mask costs $1/q$. The complete rational sums are

$$1 + \sum_{m \neq 0} \min(2|P(m)|, K(m)W(m)) = \frac{944062127676554878353}{2^{62}} < 205, \tag{25}$$
$$1 + \sum_{m \neq 0} \min(2|P(m)|, \kappa(m)W(m)) = 133.78125 < 134.$$

All 851 terms are explicit functions of (8), (21), (22), and (23). Extra common PH chunks are fixed offsets. No incompatible conditional bounds are multiplied. □

7 The final one-chunk block and the modulus $8p$

The remaining primary case is a common exact encoded prefix followed by differing final blocks, each consisting of one ENH chunk.

Lemma 7.1 (Accumulator divisibility)**.** *On projected final-block identity with minimum ENH valuation $r \ge 4$, write the raw high-word difference as jp, $-8 \le j \le 8$. Actual hash collision requires $8 \mid fj$.*

Proof. The common prefix gives exactly equal accumulators, and projected low equality forces $Y = Y'$. Subtracting the literal final updates (3) leaves fjp modulo $8p$. Equality of final hashes, by invertibility of F, requires $8p \mid fjp$, equivalently $8 \mid fj$. Reducing the accumulator to $\mathbb{F}_p$ alone would lose this restriction. □

For nonzero j define

$$d_j = 2^{\max(0, 3 - \nu_2(j))}, \qquad n_j = \left\lfloor \frac{p-1}{d_j} \right\rfloor - \left\lfloor \frac{1}{d_j} \right\rfloor. \tag{26}$$

The required multiplier event has probability exactly $n_j/(p-2)$. For $j = 0$ use one. The $j = 0$ class costs $1/q$ by Lemma 4.5.

Lemma 7.2 (Weighted suffix count)**.** *For $r \ge 8$, the joint event of projected final-block identity and actual hash collision has probability less than $435/q$ over the IID compressor words and the independent multiplier.*

Proof. Use (12) with common high offset zero. Fix A, its wrap, a B-wrap, and a residue $B_0 \pmod{16}$. Then $AB \pmod{16} = l = AB_0 \pmod{16}$. On words with these four low bits fixed, the remaining affine function $f_{m,t} \pmod{R}$ has label

$$b_m = m \,\&\, (R/2 - 1) \,\&\, \neg 15, \quad b_t = t \,\&\, b_m \,\&\, (R/4 - 1), \quad c = m - 2((l \,\&\, m) \oplus t) \pmod{R}. \tag{27}$$

Its constant is c. Each remaining selected bit i contributes $\pm 2^{i+1}$; the sign is specified by b_t, except at $i = r - 2$ where the signs coincide. This proves exact equality of the restricted functions.

The odd coefficient d in (12) determines B_0 successively, since the varying right side has an extra factor $2Q$. Group all nonzero mask-pattern pairs by (B_0, b_m, b_t, c). The uniqueness lemma gives at most one full B per group. That candidate has one actual mask-pattern pair, fixed by the compressor words independently of f. Its multiplier probability is bounded by the maximum $n_j/(p-2)$ among the group's members. Thus group maxima are valid even though several values of j may share a label.

For $8 \le r \le 60$, $Q \ge 16$, and the prefix equation is $dB + eA = 0 \pmod{16}$. Therefore $B_0 = uA \pmod{16}$ for $u = -e/d$, and $l = uA^2 \pmod{16}$. Let $G_r(l)$ be the sum of maximum n_j over labels (27) at l, omitting $m = 0$. Since each A-wrap interval has length divisible by 16, a valid numerator is

$$B_r = 1 + \frac{2}{16(p-2)} \max_{0 \le u < 16} \sum_{a=0}^{15} G_r(ua^2 \bmod 16). \tag{28}$$

For $r = 61, 62, 63$, $Q = 8, 4, 2$. Write $\delta = RD$, $\eta = RE$, enumerate odd $1 \le D < Q$, $0 \le E < Q$, and the tag gap modulo $16/Q$. For each wrap pair and $a \pmod{16}$, solve the four-bit equation and sum the label

maxima, denoting this sum $T(\alpha, \beta, a)$. Only $\beta = 0$ occurs if $E = 0$. With $l_0 = Q - D$, $l_1 = D$, the complete numerator is

$$B_r = 1 + \frac{1}{16Q(p-2)} \max_{D,E,\Delta\tau \bmod 16/Q} \sum_{\alpha,\beta,a} l_\alpha T(\alpha, \beta, a). \tag{29}$$

Every full-width context reduces to one of these finite contexts. Nonzero masks have valuation at most three, so their functions cannot be the zero class modulo R. Exact integer evaluation of all 56 widths gives the values displayed below; the independent check evaluates affine functions on zero and basis inputs and tries all sixteen residues instead of using the prefix solver.

r	Numerical value of B_r	Strict upper bound
$8, \ldots, 60$	$\simeq 37 + 7.5(r-8)$, maximum approximately 427	428
61	$1001888787503350020012/(p-2) \simeq 434.5$	435
62	$\simeq 420.1875$	421
63	$\simeq 423.6875$	424

The decimal entries describe the exact rational certificate values; they do not replace them. In particular the 61 row is approximately 434.5, with ceiling 435. All rows are strictly below 435. Division by q for the free word and averaging A prove the claim. □

For $r \le 3$, the plain no-PH projected bounds are $17, 18, 20, 24$ over q: the additive target jp must have $2^r \mid j$, so there are $2\lfloor 8/2^r \rfloor + 1$ targets, each of mass $2^r/q$. These differ from Lemma 6.2's bounds and are all sufficient. For $4 \le r \le 7$ use $2^r/q \le 128/q$. Outside projected identity, the common prefix cancels in the field comparison, leaving a nonzero polynomial of degree at most two. Hence the special suffix collision probability is at most

$$\frac{435}{q-561} + \frac{2}{p-2} = S_4.$$

The numerator 435 bounds collision together with identity, averaged over the multiplier. It is not an unweighted block projection constant. If multipliers are instead uniform on $\{1, \ldots, p-1\}$, the nonzero j weights decrease except at $8 \mid j$, where both are one; $f = 1$ fails every other divisibility condition. The root denominator increases to $p - 1$. All conclusions therefore hold a fortiori.

8 Closed joint compressor cases

All bounds in this section use IID compressor words. A joint event is the equality, modulo p, of both words of both compressors.

8.1 Independent checksum pattern weights

For $v \in D$ define

$$\begin{aligned} f_H(v) &= q^{-1} + \sum_{j=0}^{63} 2^{j-64} 2^{-\operatorname{wt}(v\&(2^j-1))}, \\ f_L(v) &= q^{-1} + \sum_{j=0}^{63} 2^{-j-1} 2^{-\operatorname{wt}(v \gg j)}, \\ W_H(v) &= \min(1, |P(v)| f_H(v)), \qquad W_L(v) = \min(1, |P(v)| f_L(v)). \end{aligned} \tag{30}$$

With equal checksums, after all original keys are fixed the common checksum PH is a product of two independent uniform words. Conditioning its first factor by degree shows that any selected high bits below that degree are independent uniform bits; conditioning by valuation proves the corresponding assertion for low bits at or above that valuation. The zero first factor contributes $1/q$. This proves f_H, f_L as pointwise pattern bounds, and summing over $P(v)$ proves the conditional weights in (30). Any other common offset merely translates the pattern.

Theorem 8.1 (ENH-only joint bound)**.** *With equal chunk counts, all PH chunks equal, and different final data, joint probability is at most* $4721784/q^2 < 2^{-105}$ *at every valuation.*

Proof. The checksum difference is exactly the final input XOR difference, with minimum valuation r. Conditional on all original keys, the checksum PH low difference is uniform on $2^r\mathbb{Z}/q\mathbb{Z}$. At $r \le 3$, the conditional secondary necessary mask event costs $N_r 2^r/q$. Multiply by the primary $2^r T_r/q$ from Lemma 4.2. The four numerators are $2360892, 610080, 130048, 384$.

For $r \ge 4$, a primary collision forces raw low equality. A secondary collision then forces its checksum PH low difference to zero and its high difference to one of 852 fixed targets, after XOR with the fixed ENH difference. Lemma 4.1 gives conditional probability at most $852/q$. Corollary 4.4 gives $5542/q$ for the primary, so the product is $5542 \cdot 852/q^2 = 4721784/q^2$. This is a probability times a uniform conditional bound. Using Theorem 5.7 would improve the high-valuation numerator to 2662500; the stated constant preserves the separately certified joint theorem. □

Theorem 8.2 (One PH chunk and two final words)**.** *For equal counts and equal data checksums, exactly one differing PH chunk, and two differing final words of minimum valuation $r > 0$, joint probability is at most* $170906186782/q^2 < 2^{-90}$.

Proof. The common checksum products cancel in the raw differences. If S is the changed PH shuffler, those differences are $\Delta P + \Delta E$ and $S\Delta P + \Delta E$. Since S is strictly triangular, $I + S$ is invertible. In either word, masks u, v determine

$$x = (I + S)^{-1}(u \oplus v), \qquad e = u \oplus x. \tag{31}$$

Equal checksums make the PH coordinate differences equal to the ENH coordinate XOR differences. At $r \ge 4$, both low masks must be zero, so $\Delta P_{\text{lo}} = 0$ and $Y = Y'$. The high PH target additionally satisfies $x < 2^{63}$. At $r \le 3$ we need only the low masks in $D_r = D \cap 2^r\mathbb{Z}$.

PH and ENH use disjoint original keys. For small r, their fixed low targets cost respectively $2^r/q$ and $K_{L,r}(e)/q$. After those keys are fixed the secondary pattern costs $W_L(v)$. For large r, the full PH target costs $1/q$, the ENH event $K_H(e)/q$, and the conditional pattern $W_H(v)$. Hence the exact numerators are

$$C_{r,k} = 2^r \sum_{u,v \in D_r} K_{L,r}\big(u \oplus (I + S_k)^{-1}(u \oplus v)\big) W_L(v), \tag{32}$$

$$C_{H,k} = \sum_{\substack{u,v \in D \\ x=(I+S_k)^{-1}(u\oplus v)<2^{63}}} K_H(u \oplus x) W_H(v). \tag{33}$$

Every mask pair and all fifteen maps are included. Solve the inverse bit by bit, or use the nilpotent geometric inverse, and evaluate all terms as dyadic rationals. The maxima, always at $k = 2$, have ceilings

case	$r = 1$	$r = 2$	$r = 3$	$r \ge 4$
ceiling	27196168693	13437880698	268435521	170906186782.

These sixty finite sums, together with their complete masks and exact fractions, constitute the certificate. Each conditional step was uniform over the previously exposed keys; no common PH offset was assumed for the two messages. The largest numerator is below 2^{38}. □

For $r = 0$, odd triangular ENH lifting and the PH low image are both uniform; the invertible map (31) gives numerator 852^2. If exactly one final word changes, the ENH low XOR difference is uniform on multiples of 2^r. A direct proof conditions on $\nu_2(B) = t$: after removing 2^t, odd multiplication makes $(AB, \delta B)$ uniform with the second coordinate of valuation r; the map $(z, c) \mapsto (z, z \oplus (z + c))$ is a bijection on that valuation class. Each attainable nonzero target, and zero, has mass $2^r/q$. For $r \le 3$ the joint numerator is $(N_r 2^r)^2$. For $r \ge 4$, replace K_H in (33) by κ and the pattern weight by the valid coarser $\min(1, 16 f_H)$. Its fifteen exact sums have maximum ceiling 111924178297, less than 2^{37}.

8.2 At least two PH changes and equal checksums

Lemma 8.3. *Equal-count, equal-checksum blocks with at least two differing PH chunks have joint projected probability at most*

$$\frac{345763417}{2^{116}} = \frac{J}{q^2}, \qquad J = 1416246956032.$$

Proof. For two selected PH positions, their shuffler difference is T^h times an invertible polynomial in T, for $2 \le h \le 15$. Restricting multiplication by any nonzero w-bit polynomial $d = X^v g$, $g(0) = 1$, to the low $w - h$ bits of each output word has rank at least $w - h$. Explicitly set $r_0 = \max(w - h - v, 0)$ and

$r_1 = \min(v, w-h)$. Select input bits $0, \dots, r_0-1$ and $w-v, \dots, w-v+r_1-1$; use output rows $v, \dots, v+r_0-1$ and $w, \dots, w+r_1-1$. The resulting matrix is block triangular with unit triangular diagonal blocks, of total size $w-h$.

Leave one key word free in each selected PH chunk, fixing their companions and all other original keys. For specified primary and secondary raw targets, eliminating one PH difference prescribes a shifted value of the other, costing at most $2^{-(64-h)}$. The remaining full target costs at most 2^{-64}. Thus each joint target has probability at most $2^{-(128-h)}$.

The identity for the secondary difference minus T times the primary difference shows that, in each word, $v \oplus (u \ll 1)$ has fixed low h bits. The constant includes the generally nonzero $(I+T)\Delta E$ and all other conditioned terms. Define

$$N_h = \max_c \#\{(u,v) \in D^2 : (v \oplus (u \ll 1)) \bmod 2^h = c\}.$$

For $h \geq 5$ call a secondary mask long when bits $3, \dots, h-1$ are all one, and let S_h be the same maximum restricted to non-long v; for $h < 5$ set $S_h = N_h$.

For a long mask, the pattern equation $2(x \,\&\, v) = v + jp$, reduced modulo 2^h, forces bits $3, \dots, h-2$ of x all zero or all one, since $p \equiv -1 \pmod{2^h}$ and $|j| \leq 8$. There are at most sixteen assignments to the low $h-1$ bits. The common checksum product, conditional on all original keys, has low k point bound $(k+2)2^{-k-1}$ and high k point bound $2^{-k} + k/q$. The former follows by summing first-factor valuations, the latter by summing degrees and the zero factor. Taking $k = h-1$ gives a uniform additional cost $\rho_h = \min(1, (h+1)2^{4-h})$ for a designated long secondary word.

At most S_h^2 joint targets have neither secondary mask long, and at most N_h^2 occur altogether. Therefore the probability is at most

$$2^{-(128-h)} \min\{N_h^2, S_h^2 + \rho_h N_h^2\}. \tag{34}$$

Every ordered pair in D^2 is counted for each h. The complete maxima are

h	1	2	3	4	5	6	7	8
N_h	514608	278664	153492	104060	99768	96425	93140	89913
S_h	514608	278664	153492	104060	2966	2916	2866	2816

h	9	10	11	12	13	14	15
N_h	86744	83633	80580	77585	74648	71769	68948
S_h	2766	2716	2666	2616	2566	2516	2466.

The maximum of (34) is at $h = 15$ and equals $(2466^2 + 68948^2/128)2^{-113} = 345763417/2^{116}$. The $h = 1$ row is a harmless enclosure; the actual parameter starts at two. This supplies the claimed constant with the exact source correspondence and counts above. □

For tag-only pairs, the boundary argument of Lemma 4.6 confines H to at most 4080 values. Its point mass is less than $48/q$: the count is at most $q(H_{q-1}+2)$, and $H_{q-1} < 1 + \log q < 1 + 64(7/10)$, since $e^{7/10} > 1 + 7/10 + (7/10)^2/2 + (7/10)^3/6 > 2$. Conditional on original keys, the high mask forces bits 8–60 set. The checksum weight for those selected bits is at most $16 f_H(\{8, \dots, 60\}) = 61 \cdot 2^{-52}$. Multiplying yields the joint bound $11946240 \cdot 2^{-116}$, or numerator $11946240 \cdot 4096$ over q^2.

8.3 The secondary marginal and complete ledger

Lemma 8.4. *Every distinct valid block/tag pair has secondary projected probability at most C_1/q, $C_1 = 2C = 729632$.*

Proof. If keyed checksums differ, condition all original keys; the twisting PH has two distinct inputs and fixed possibly unequal offsets, so Lemma 4.1 gives C/q. For equal counts and equal checksums with a PH change, leave one opposite PH word free. The raw slice is injective in that word and lies in the binary space with bit 127 zero. Every S_k in (2) has the same kernel as T, spanned by the two word top bits. Its intersection with that space has two elements, zero and 2^{63}. Each shuffled full target has at most two preimages. Bit zero in each shuffled word is zero, so at most 604^2 targets meet the prescribed parity restrictions after other terms are fixed. This gives $2C/q$, irrespective of the common checksum value.

At unequal counts, the keyed checksum difference includes a fresh original key pair and vanishes with probability q^{-2}. Bound this exceptional event by its entire probability and apply C/q otherwise, giving $q^{-2} + C/q < 2C/q$. With PH agreement and changed ENH data, the checksum changes. For tag-only differences, the preceding necessary boundary event costs $195840/q < 2C/q$, without conditioning a key-dependent checksum. These cases are exhaustive. □

Disjoint block configuration	Primary $\times q$	Secondary $\times q$	Joint $\times q^2$
Different chunk counts	82	$C + 1/q$	$1 + 82C$
Same count, different data checksums, a PH change	1123	C	$1123C$
Same count/checksum, at least two PH changes	1123	$2C$	J
Same count/checksum, one PH and one ENH word change	1123	$2C$	2^{37}
Same count/checksum, one PH and two ENH words, $r = 0$	1123	$2C$	852^2
Same count/checksum, one PH and two ENH words, $r > 0$	1123	$2C$	170906186782
Same count, all PH equal, final data differ	3125	C	4721784
Same expanded data, valid tags differ	1/8	195840	$11946240 \cdot 4096$

The entries are safe non-strict upper bounds. For different counts, multiply the primary event by the uniform conditional C/q outside checksum equality, and add q^{-2}. For different checksums and a PH change, multiply $1123/q$ by that conditional bound. All remaining joint entries were proved above. Equal checksums with exactly one PH change require a final-data change; otherwise that sole change cannot cancel in the checksum. Thus the rows are disjoint and exhaustive, and their column maxima are $C_0 = 3125$, C_1, and J. They are combined by a maximum, not a sum.

9 From compressor counts to complete hashes

9.1 Short messages and field comparisons

For a short message, let lo, hi be the first and last four-byte reads at lengths 4–8. Below length four, let lo be the first byte for odd length and zero otherwise, and hi the last two-byte read at lengths two or three and zero otherwise. The packed word is

$$2^{32}\text{hi} + ((\text{hi} + \text{lo}) \bmod 2^{32}).$$

Its high part recovers hi and subtraction recovers lo; the reads cover the message. Packing is therefore injective at each length. The short mixer successively uses XOR right shifts 30, 27, 31, the odd multipliers `BF58476D1CE4E5B9` and `94D049BB133111EB`, and at the middle XOR adds the noise word $(\text{seed} + k_s) \bmod q$. Right-shift XOR maps and odd multiplications are invertible. A fixed short input consequently hashes uniformly under IID words; equal-length distinct inputs never collide. Different lengths have primary probability $1/q$.

For the short fingerprint, the noise indices are s and $s + 4$. For a fixed message these are distinct. For two unequal lengths s, t, the two noise constraints have edges $\{s, t\}$ and $\{s + 4, t + 4\}$. They are distinct edges of a forest, even when they share a vertex. Exposing the independent vertex words from a root gives exactly two factors $1/q$, hence short joint probability q^{-2}.

For a long message, use its coefficient polynomial from (3); for a short one use the constant obtained by applying F^{-1} to its short output and reducing modulo p. These polynomials are independent of the polynomial multipliers. For comparisons involving a long message, write Z_i for the event that comparison polynomial i vanishes identically. Equal numbers of long blocks have a differing aligned block/tag tuple by encoding injectivity; identity requires equality of its two coefficients. Choose that tuple from the messages before sampling keys and apply the ledger.

If the block counts differ, the longer message's first two coefficients lie above the shorter polynomial's degree and must both vanish. The primary point cost is $82/q$. Conditional on original keys, the secondary checksum product is fresh. A prescribed projected value has at most nine low representatives and five high representatives after fixing its top bit, since $p \geq 8$. There are at most 45 raw product targets, at most one zero. The carryless-product point bound is thus $(46q - 45)/q^2 < 46/q$, conditionally. This gives joint cost $82 \cdot 46/q^2$.

For short/long comparisons, identity forces the short constant to be zero. Exactly nine word representatives reduce to zero, giving marginals $9/q$ and joint $81/q^2$ from the fixed short input's independent components. This is a necessary event; its possible dependence on the long computation is immaterial. It follows that

$$\Pr(Z_0) \leq C_0/q, \quad \Pr(Z_1) \leq C_1/q, \quad \Pr(Z_0 \cap Z_1) \leq J/q^2. \tag{35}$$

The actual short/short probabilities are used separately. No assertion about their reduced constant-polynomial identity is needed.

9.2 Distinct keys and roots

The probability that 34 IID words are all distinct is at least $1 - \binom{34}{2}/q = 1 - 561/q > 0$. Conditioning gives uniform distinct tuples, so every compressor-key event of probability b_0 becomes at most $b_0/(1 - 561/q)$. Apply this factor once to a joint event. The independent multipliers retain their original distribution.

The primality of p has an elementary certificate. Its predecessor factors as

$$p - 1 = 2 \cdot 3^2 \cdot 5^2 \cdot 7 \cdot 11 \cdot 13 \cdot 31 \cdot 41 \cdot 61 \cdot 151 \cdot 331 \cdot 1321.$$

Trial division verifies these small primes, and witness 37 satisfies $37^{p-1} = 1 \pmod p$ and $\gcd(37^{(p-1)/\ell} - 1, p) = 1$ for each listed prime factor ℓ. For any prime divisor s of p, the order of 37 modulo s is therefore $p - 1$; thus $p - 1 \mid s - 1$, forcing $s = p$. A nonzero degree-d polynomial over this field has at most d roots, by division by $X - \alpha$ and induction. With at most $\lceil L/32 \rceil$ blocks, the conditional root probability is at most $\rho(L)$.

9.3 Proof of the primary theorem

For equal long block counts the following exhaustive choice uses fixed message data. If a differing aligned tuple has different chunk counts, use $82/q$. Otherwise, if a differing tuple has at least two chunks, use $151/q$ for a PH change, $205/q$ for an ENH-only change, or $1/(8q)$ for a tag-only change. If no such tuple exists, every differing tuple has one chunk and must be final; all earlier blocks agree exactly. A data change there is the special suffix case of Section 7; a tag-only change still uses Lemma 4.6. Unequal block counts and short/long comparisons have constants 82 and nine as above. Thus every ordinary case has identity probability at most A_4, and probability at most $A_4 + (1 - A_4)\rho(L)$. The special case has bound S_4. This proves (5).

For $H = \lceil L/512 \rceil$, $\lceil L/32 \rceil \le 16H$, and exact rational comparison gives

$$57 < 2^{61}\left(\frac{205}{q - 561} + \frac{32}{p - 2}\right) < 58, \qquad 56 < 2^{61} S_4 < 57.$$

Both terms of the maximum are strictly below $58H/2^{61}$, proving Theorem 1.1. The margin of the first scaled coefficient below 58 is approximately 0.375. For $L \ge 2$, $\lceil L/32 \rceil \le L/2$, so the ordinary envelope divided by L is at most $A_4/2 + (1 - A_4)/(p - 2) < S_4/2$. The special term attains $S_4/2$ at $L = 2$. At $L = 1$ the score is larger; clipping and the output floor do not change this. Hence the score is $\log_2(2/S_4)$, with exact certificates $(2/S_4)^{50} > 2^{2809}$ and $(2/S_4)^{100} < 2^{5619}$.

The weaker but separately useful constant 3125 gives $A = 3125/(q - 561)$ and envelope $A + (1 - A)\rho(L)$ for $L \ge 2$, coefficient $423\,\lceil L/512 \rceil / 2^{61}$, and score $53.38\ldots$. The original complete constant 364816 similarly gives coefficient 45635 and score $46.52\ldots$. All these statements, and primary-bound inheritance by any fingerprint containing that primary component, remain valid.

9.4 Proof of the fingerprint theorem: the product of root bounds

Condition on the entire compressor key. If $z_i = \mathbf{1}_{Z_i}$, independent multipliers give the pointwise bound

$$[\rho + (1 - \rho)z_0][\rho + (1 - \rho)z_1] = \rho^2 + \rho(1 - \rho)(z_0 + z_1) + (1 - \rho)^2 z_0 z_1. \tag{36}$$

This includes all four zero/nonzero polynomial configurations. A zero field polynomial need not force equality modulo $8p$, but its conditional probability is safely at most one. Averaging (36) and using (35) after the single distinct-key correction proves (6). Completed compressors and their identity events have not been assumed independent. The short/short probability is $1/[q(q - 561)]$, below every other branch. The same proof uses a larger root denominator, hence holds a fortiori, for uniform nonzero multipliers including one.

Let $H = \lceil L/2^{23} \rceil$ and $r_0 = 2^{19}/(p - 2)$. Then $\rho(L) \le r_0 H$ and

$$E(L) \le b + a\rho + \rho^2 \le H^2(b + ar_0 + r_0^2).$$

The exact certificate is

$$2^{83}(b + ar_0 + r_0^2) = \frac{6155518206291691470964434247450439200925290553582341324 8}{9807971461541688369693481200556472928541357065351095405 5} < \frac{81}{128}.$$

The short branch also satisfies this inequality, proving Theorem 1.2. The fingerprint proof uses the complete 3125 primary marginal and the joint ledger, and does not depend on the $8p$ suffix improvement.

10 Length domains and the shared-multiplier refutation

10.1 Exact score minimization

Let $t = \lceil L/32 \rceil$ and $m = p - 2$. Before ρ reaches one, write

$$E(L) = w + vt + ut^2, \quad w = b, \quad v = 2(a - 2b)/m, \quad u = 4(1 - a + b)/m^2.$$

All three coefficients are positive. On the first long-input plateau, $E(L)/L$ is largest at $L = 2$; on later plateaus it is largest at $L = 32t - 31$. Polynomial division gives

$$\frac{w + vt + ut^2}{32t - 31} = \frac{u}{32}t + \frac{v}{32} + \frac{31u}{1024} + \frac{w + 31v/32 + 961u/1024}{32t - 31}.$$

Its second derivative is positive, so on any finite interval of plateau indices its maximum is at an endpoint. Exact rational endpoint comparisons therefore establish the following minima, without searching all message lengths:

Word-cap domain	Minimizing L	Score of E
$1 \le L \le 2^{23}$	2	$88.6347780628\ldots$
$1 \le L \le 2^{46}$	$2^{46} - 31$	$83.99999997\ldots$
$1 \le L \le 2^{61}$	$2^{61} - 31$	$68.99999999999914\ldots$
All positive integers	$2^{65} - 95$	$64.9999999999999999975\ldots$

For the unrestricted domain the last unsaturated plateau starts at $2^{65} - 95$ and the first saturated one at $2^{65} - 63$; an exact cross-product selects the former. The difference is about $4.7 \cdot 10^{-52}$ in the error-per-word ratios and has no practical significance. The 128-bit output floor is always smaller than E.

The coarse bound (7) scores only $83.66\ldots$ through 2^{46} words; the 83.99 figure belongs to the fine envelope (6). The published coefficient-one formula scores $82.99999983\ldots$ on that domain, and about 68 on the full 64-bit byte domain. The 88.63 figure applies specifically through 64 MiB. For all $L \ge 2$, setting $r = \rho(L)$,

$$A_4 + (1 - A_4)r - E(L) = (1 - r)\{r(1 - a) + A_4 - b(1 - r)\} \ge 0,$$

since $a < 1$ and $A_4 > b$. Thus the fine fingerprint envelope also dominates the inherited primary envelope pointwise.

10.2 A block-swap pair for the reference fingerprint

The Python reference stores only one multiplier. Calling it with the same key and two compressor selectors therefore defines a different fingerprint family from the C architecture.

Proposition 10.1. *Let A be 256 zero bytes and B be 256 bytes of value one. The distinct 512-byte strings $A\|B$ and $B\|A$ collide in both reference components for every compressor key and seed whenever $f = p - 1$. Their reference fingerprint collision probability is at least $1/(p - 2) > 2^{-83}$.*

Proof. Both blocks are full, so their tags are the common seed independently of position. At $f = p - 1$, $f^2 \bmod p = 1$, and the literal update becomes

$$z \longmapsto z + \mathrm{low} + (p - 1)\mathrm{high} \pmod{8p}.$$

The two block contributions commute, separately for each compressor. The finalizer inputs, hence outputs, coincide. There are $(q)_{34} = \prod_{j=0}^{33}(q - j)$ distinct-word compressor keys in this guaranteed multiplier fibre, out of $(p - 2)(q)_{34}$ keys altogether. Its probability is exactly $1/(p - 2)$, although the entire collision event may be larger. For this length the published bound is 2^{-83}, smaller by a factor greater than 2^{22}. □

For independent multipliers in $\{2, \ldots, p - 1\}$, the simultaneous $(p - 1, p - 1)$ fibre has probability $(p - 2)^{-2}$ and does not refute Theorem 1.2. If one is permitted, $f = 1$ also makes updates commute. The guaranteed commuting fibre for two independent uniform nonzero multipliers is therefore $(2/(p - 1))^2$, about 2^{-120}. This is a lower bound from a specified fibre, not an assertion that other multiplier choices cannot collide.

11 Scope and machine-checked status

These are information-theoretic bounds for ideal full keys, a fixed common seed, and messages chosen independently of keys. They do not establish the distribution of the Salsa20-derived keys, the OH-word repair sampler, cryptographic security after disclosure of keys, or a cross-seed collision guarantee. The mathematical all-length envelopes are unrestricted in L; actual C inputs must also meet the platform's byte-length and object requirements.

Source inspection identifies the folded ENH, checksum construction, separate word shifts, and independent polynomial pairs used by the C architecture. The archived bridge checks compare 1776 C fingerprints with literal arithmetic at all sizes 0–273 and further boundaries through 8192 bytes, including extreme multipliers and tags, in both long-input configurations. That bridge disables dynamic dispatch; separate archived comparisons cover dynamic dispatch and the streaming interface. These are implementation checks, not a formal compiler or machine-instruction refinement theorem.

The machine-checked statements[2] distinguish the following results. Probabilities there are exact finite uniform counts in nonnegative rationals.

- The complete 364816 primary bound, coefficient 45635, both IID and distinct-key all-message theorems, linear fingerprint inheritance, and the sharp $4721784/q^2$ ENH-only joint theorem are closed Lean theorems. The earlier complete development contains 454 audited declarations.
- The continuation has 830 audited declarations in total. The original positive-valuation PH+ENH obligation below 2^{-87} is closed, with numerator 2058363321984. For the sharper paper numerator 170906186782, the literal reduction, analytic bounds, and all 45 low-word cases are checked; fifteen high-word integer-sum inequalities remain outside the completed Lean result.
- The 3125 ENH argument and parts of the finite arithmetic are checked, but the full 1123 PH and sharp tag-only probability bounds, hence the unconditional 3125 all-message theorem, are not yet closed in Lean. Their assembly statements retain these premises explicitly.
- Lemma 7.1 is checked on the literal modulo-$8p$ update, including the common-prefix form. The full 151, 205, and weighted 435 bounds and their all-message assembly remain paper proofs with exact certificates. The arithmetic implication from the stated envelope to coefficient 58 is checked.
- The product of independent root bounds and the rational 81/128 comparison are checked. The needed secondary and joint compressor assembly, the full fingerprint theorem, and the reference block-swap proposition are not reported as closed Lean theorems.

The recorded successful builds and axiom audits use Lean 4.24.0 and Mathlib revision `f897ebcf72cd16f8 9ab4577d0c826cd14afaafc7`. Reported declarations use only subsets of `propext`, `Classical.choice`, and `Quot.sound`; no additional mathematical axioms or unproved placeholders occur in the reported modules. This status does not reduce any paper theorem: it specifies which parts have additionally been checked by a proof assistant.

A Reproducibility and exact certificates

All paths below are relative to the post's `records/` directory. This source embeds its BibTeX database, so it builds with `latexmk -pdf`. The numerical proof dependencies use integer or rational arithmetic; displayed decimal scores are not premises. For a claimed interval $(k/100, (k+1)/100)$ and exact score ratio $R = N/D$, the check is $N^{100} > 2^k D^{100}$ and $N^{100} < 2^{k+1} D^{100}$.

Directory	Certificate programs and their mathematical role
`umash-corrected/checks/`	`triangular.py`, `certify_constants.py`: complete masks and patterns, independent addition automaton, primality, lifting and initial envelope. `exhaustive_lift.cpp`: finite-width algebra checks.
`umash-corrected/proof2/checks/`	`certify_open_ph_enh.py`, `verify_certificate.py`: all sixty sums in (32)–(33), divisor estimate and completeness mass. `validate_quadratics.cpp`: independent quadratic checks.

Directory	Certificate programs and their mathematical role
`umash-corrected/proof3/checks/`	`certify_primary.py`, `verify_certificate.py`: the 84 convolution cases, every lifting-function class at widths 4–63, the 1123/3125 ledger, and the 423 envelope. `validate_primary.cpp`: truncated counts and class uniqueness checks.
`umash-corrected/proof4/checks/`	`prepare_masks.py`; `ph_low_table.cpp`, `verify_ph_low.cpp`; `certify_ph_linear.py`, `certify_ph_high.py`: the complete PH rank tables. `certify_enh_weights.py`: (25). `certify_tail.py`: (28)–(29). `verify_certificates.py`: independent reconstruction. `certify_envelope.py`: coefficient 58 and score. `validate_new.cpp`: algebra checks.
`umash-corrected/proof5/checks/`	`certify_fingerprint.py`, `verify_certificate.py`: the joint ledger, the two-PH table, exact headline comparison, all score endpoints, and block-swap fibre. `check_c_model.py`, `c_bridge.c`: source/arithmetic comparisons.

Each of the five UMASH check directories includes `run_xeon.sh` and a source/certificate manifest. The exact JSON outputs accompany the named scripts. The independent verifiers check complete masks from (11), rather than accepting a sampled list. The PH maxima range over every center through the finite prefix recursion; the suffix counts include every reduced increment, wrap, and tag context. These are certificates for all cases of the stated bounds. Scaled exhaustive checks support the algebra and code used to compute them; their frequencies are not extrapolated to width 64. In particular the smallest scaled widths do not all share the production ratio $\lfloor (2^w - 1)/p \rfloor = 8$.

The complete rational values and certificate data supporting all five sets of bounds are included in the companion records[1].

References

[1] Thomas Dybdahl Ahle. Corrected UMASH bounds: proofs and exact certificates. https://thomasahle.com/blog/adversarial-examples-for-hashes/records/README.html, September 2026. Paths relative to the post's records directory; complete reproduction map in Appendix A.

[2] Thomas Dybdahl Ahle. UMASH formalization: certified statements and remaining mathematical interfaces. https://thomasahle.com/blog/adversarial-examples-for-hashes/records/machine-checked.html, September 2026.

[3] Jim Apple. HalftimeHash: Modern hashing without 64-bit multipliers or finite fields. arXiv:2104.08865v2, 2021.

[4] Paul Khuong. UMASH: a fast almost universal 64-bit string hash. https://github.com/backtrace-labs/umash/blob/master/umash.pdf, February 2022. Literate whitepaper, 23 February 2022.