hexagon:2611.01291v1

RECURSIVE ENTROPY AND OPTIMAL MODULAR COVERAGE OF HARMONIC SUBSET SUMS

PENGCHENG CHEN

Abstract. Let $s(n)$ be the logarithm of the number of distinct subset sums of $1, 1/2, \dots, 1/n$, and let $b(P)$ be the first height for which $s(n) \geq \log P$. An exact large-prime decomposition yields a recursion for s, regular variation of index one, and a pool of reciprocal mass $o(1)$ preserving asymptotically all the entropy. For a density-one set of primes of size P, the height $b(P)$ is the threshold separating the absence of nonempty zero sums from near-complete modular coverage. A four-point Fourier majorant for recursive two-element components yields ordinary full coverage, with uniform distinct-value mixing, above the explicit sufficient factor $\tau = 9200/4601 < 2$ of this threshold. A complementary augmented-block rational-reconstruction argument, combined with dissociated reciprocal sets, establishes uniform Bernoulli representation counts at the counting-optimal order of height and gives particularly small exceptional sets. At a slightly larger scale we obtain mixing stable under adversarial deletion. Exact Chinese remainder grid lifting gives an all-primes coverage bound, a height bound for arbitrary proper fractions, and a sharp order for the logarithm of the longest rational arithmetic progression of harmonic subset sums. Finally, the entropy threshold determines the order of the prime-square deletion loss, bounding its normalized constant between two and a number strictly below $2\sqrt{2}$. We prove that the lower constant is attained if and only if a density-one directed low-loss condition holds above the entropy threshold. The directed condition and the all-primes bound at height $O(\log p \log_2 p)$ remain open. Optimality refers to the counting-scale order on density-one primes, not to an established exact constant for ordinary full coverage.

Contents

1. Introduction and statements of results 2
2. Exact recursion and regular variation 5
2.1. The iterated-logarithm order 7
2.2. The top-level entropy pool 7
3. Prime-average collision estimates and a Fourier damper 8
4. The modular transition and the two-pool estimate 9
5. Uniform distinct-value coverage strictly below factor two 10
6. Density-one coverage at the counting scale 13
6.1. Augmenting the rectangular block by its outer primes 14
6.2. Dissociated high-entropy blocks and mixing 15
7. Rectangular blocks and rational reconstruction 16
8. Prime averages and deletion-stable Bernoulli mixing 18
9. A pointwise sufficient condition for full coverage 19
10. Exact rational grids and height 19
10.1. Dyadic completion 19
10.2. Height for arbitrary denominators 20
10.3. The counting lower bound and a progression obstruction 21

Date: 28 November 2026.

2020 *Mathematics Subject Classification.* 11B75, 11D68, 11N25, 11N37, 11T23.

Key words and phrases. Harmonic subset sums, entropy, Egyptian fractions, modular coverage, rational reconstruction, prime-square losses.

11. The least forbidden initial denominator 21
11.1. The least forbidden initial denominator 22
12. Uniform recursive coverage and deletion scaling 22
13. The large-prime deletion expansion 23
14. Prime-square loss and the directed threshold 24
14.1. An integral formula and the exact equivalence 25
15. The unresolved fixed-target problem 26
15.1. The uniform binary-coverage problem 26
Appendix A. A general entropy statement for rational projections 27
Appendix B. Related variants of the construction 27
B.1. Composite moduli 27
B.2. Coverage with bounded multiplicities 28
B.3. Height on most numerators of good prime denominators 28
Scope of the results 29
References 29

1. Introduction and statements of results

Let $\mathcal{H}(A) = \sum_{a\in A} 1/a$ for a finite set of positive integers, and let $H_n = \mathcal{H}([1,n])$. Write

$$\mathscr{S}(A) = \{\mathcal{H}(B) : B \subseteq A\}, \quad \mathscr{S}_n = \mathscr{S}([1,n]), \quad S(n) = |\mathscr{S}_n|, \quad s(n) = \log S(n).$$

We use the convention $S(0) = 1$, $H_0 = 0$, and $s(A) = \log|\mathscr{S}(A)|$. All logarithms are natural, and $\log_j$ denotes the j-fold iterated logarithm. Define the first entropy crossing by

$$b(P) = \min\{n \geq 1 : s(n) \geq \log P\}. \tag{1}$$

For a prime p and $q < p$ define the minimum deletion mass

$$d(p,q) = \min\{\mathcal{H}(D) : D \subseteq [1,q],\ \mathcal{H}(D) \equiv H_q \pmod p\}. \tag{2}$$

Equivalently, the retained set has residue zero. Rational congruences are used only when the denominators are units at the modulus. Put $d(p,0) = 0$. For each prime p, let $f(p)$ be the least integer $q < p$ for which all residues modulo p are represented by distinct reciprocal subset sums from $[1,q]$. Let $g(p)$ impose the additional requirement that each representation has reciprocal mass at most one. Let $z(p)$ denote the least $q < p$ admitting a nonempty zero-sum subset, with value $+\infty$ if no such subset exists. Set $G(x) = \max_{p\leq x} g(p)$ for $x \geq 2$. In particular $f(p) \leq g(p) \leq p-1$, since a nonzero class can be represented by a single reciprocal in $[1,p-1]$. A full cover at height q requires at least p distinct rational values, so $f(p) \geq b(p)$.

We also write $\mathcal{V}(x) = \mathscr{S}_{\lfloor x\rfloor}$ for real $x \geq 0$. Fix a cutoff $A_{\mathrm{cut}} > 3$ and define

$$\mathcal{C}_j(x) = \prod_{i\geq j:\, \log_i x \geq A_{\mathrm{cut}}} \log_i x, \qquad W(x) = \mathcal{C}_4(x),$$

where the iterates stop once the cutoff is crossed and an empty product is one. This convention is used uniformly throughout. Density one in primes means that for each fixed parameter the exceptional set in a dyadic interval $[P,2P]$ has size $o(P/\log P)$; such estimates also imply relative natural density one among all primes.

There are two probability laws in the coverage results. For a set of denominators A, the *distinct-value law* chooses a rational number uniformly from $\mathscr{S}(A)$ before modular reduction. The *Bernoulli law* chooses each denominator independently with probability $1/2$, thus weighting a rational value by its number of representations. The laws agree when all rational subset sums are

distinct, but in general they differ. For a prime p and a set D of unit denominators define the phase energy

$$E_D(p,h)=\sum_{a\in D}\left\|ha^{-1}/p\right\|^2,\qquad h\in\mathbb{F}_p, \tag{3}$$

where a^{-1} is taken modulo p and $\|\cdot\|$ is distance to the nearest integer. This notation is independent of a choice of integer representatives.

The distinct harmonic-subsums problem was studied by Bleicher and Erdős [1, 2]. The variable-depth lower bound of Bettin, Grenié, Molteni and Sanna [3] is used explicitly for the matching counting order and for the dissociated blocks; the entropy recursion and regular variation do not require it. The powersmooth interval lemma of Liu and Sawhney [4, Lemma 4.1 in arXiv v1] is used for the least forbidden initial denominator and the final height refinement. We use the prime number theorem with its classical error term, elementary finite Fourier analysis, and elementary rational reconstruction throughout. The entropy decomposition and the Bernoulli reconstruction address different aspects of the same counting scale; neither alone settles the directed threshold problem. No priority claim is made for the counting order pending a complete literature comparison. The published preprints [8, 9] give, respectively, an all-prime bounded-mass covering estimate at height $O(\log p\log_2 p(\log_3 p)^{12})$ and sharp entropy estimates for ordered disjoint representations with prescribed sums. These results concern different quantifiers and, in the second case, a different counting object from the number of distinct rational values used here. The present proofs remain self-contained apart from the indicated external inputs. For the explicit constant in Theorem 1.2, we additionally use the effective prime-counting inequalities of Rosser and Schoenfeld [7, (3.1), (3.2), (3.6)].

Theorem 1.1 (Entropy and its inverse)**.** *One has*

$$s(n)=\frac{n}{\log n}\left(\sum_{j\le\log n}\frac{s(j)}{j(j+1)}+O(1)\right), \tag{4}$$

and, for each fixed $t>0$,

$$s(\lfloor tn\rfloor)\sim ts(n),\qquad b(P^t)\sim tb(P). \tag{5}$$

For large x write $h(x)=\max\{j:\log_j x\ge A_{\mathrm{cut}}\}$. Then

$$s(n)\asymp\frac{n}{\log n}\prod_{j=3}^{h(n)}\log_j n,\qquad b(P)\asymp\frac{\log P\log_2 P}{W(P)}. \tag{6}$$

The lower bound in the first comparison uses [3].

Theorem 1.2 (Modular coverage)**.** *For fixed $0<t<1$, apart from $P^{t+o(1)}$ primes $p\in[P,2P]$, one has $d(p,q)=H_q$ for every $q\le\lfloor tb(p)\rfloor$. For fixed $t>1$ there is a pool of denominators in $[1,\lfloor tb(P)\rfloor]$, independent of p, with total mass $o(1)$ whose ordinary subset sums cover all but $o(p)$ residues for all but $O_t(Pe^{-c_t(\log_2 P)^2})$ primes in $[P,2P]$. For the same nonexceptional primes, signed subset sums from that pool cover every residue.*

For the explicit sufficient factor $\tau=9200/4601=1.9995653118\ldots\in(1,2)$ and every fixed $t>\tau$, the ordinary subset sums of the large-prime pool constructed below cover all *residues, with the same form of strong exceptional-set bound. This last statement holds even for the uniform distribution on its distinct rational subset sums: its reduced distribution ρ_p^{val} satisfies*

$$\sup_{h\in\mathbb{F}_p}|p\rho_p^{\mathrm{val}}(h)-1|\le e^{-3(\log_2 P)^2} \tag{7}$$

for sufficiently large P and every prime outside the exceptional set. In particular $g(p) \le (\tau+\varepsilon)b(p)$ and $z(p) \le (\tau+\varepsilon)b(p)$ for density-one primes and every fixed $\varepsilon > 0$, while $f(p) \ge b(p)$ for all primes and $z(p) \ge (1-\varepsilon)b(p)$ for density-one primes.

Theorem 1.3 (Counting-scale density-one coverage)**.** *There exist absolute constants $C, a, c > 0$ with the following property. For X sufficiently large put $L = \log X$, $U = \log L$, $F = \mathcal{C}_4(X)$, $Q = \lfloor CLU/F \rfloor$ and $t = \exp(aU/F)$. There exists a set $\mathcal{B}_X \subseteq [1,Q] \cap \mathbb{Z}$, independent of $p \in (X, 2X]$, with $\mathcal{H}(\mathcal{B}_X) = o(1)$, such that for all primes in this interval except at most*

$$O(X \exp(-ct))$$

we have, uniformly for $x \in \mathbb{F}_p$,

$$\#\{J \subseteq \mathcal{B}_X : \mathcal{H}(J) \equiv x \pmod p\} = \frac{2^{|\mathcal{B}_X|}}{p}(1 + O(e^{-t})).$$

Consequently on a density-one set of primes,

$$f(p) \asymp g(p) \asymp \frac{\log p \log_2 p}{\mathcal{C}_4(p)}.$$

The lower bound in this comparison holds for every sufficiently large prime. The constructed set need not lie in the upper half of $[1,Q]$.

Theorem 1.4 (Stronger mixing at a larger scale)**.** *There exist fixed $C, c > 0$ such that, with $q_p = \lfloor C \log p \log_2 p \rfloor$ and $I_p = [q_p/2, q_p] \cap \mathbb{Z}$, for all primes $X < p \le 2X$ except at most*

$$O\left(X \exp\left(-\frac{c \log X}{\sqrt{\log_2 X} \log_3 X}\right)\right),$$

we have, uniformly in $x \in \mathbb{F}_p$,

$$\#\{J \subseteq I_p : \mathcal{H}(J) \equiv x \pmod p\} = \frac{2^{|I_p|}}{p}\left(1 + O\left(e^{-c \log X/(\sqrt{\log_2 X} \log_3 X)}\right)\right).$$

In particular $g(p) \ll \log p \log_2 p$ on this density-one set of primes.

For an integer $b \ge 2$ define the worst-case height

$$D(b) = \max_{1 \le a < b} \min\{\max F : \mathcal{H}(F) = a/b\}.$$

The height question goes back to Bleicher–Erdős and Erdős–Graham [2, 5]; Yokota [6] proved the original $b(\log b)^{1+o(1)}$ bound. Liu–Sawhney [4, Theorem 1.5 in arXiv v1] proved $D(b) \ll b \log b (\log_2 b)^3 (\log_3 b)^{O(1)}$ and asked in their preprint (Section 1.2) for second-logarithm exponent $1 + o(1)$. Their result implies $f(p) \le \log p (\log_2 p)^{3+o(1)}$ by multiplying a representation of a/p by p and reducing modulo p: only denominators pj survive, since the height is less than p^2 for large p.

Theorem 1.5 (Coverage and height)**.** *Uniformly over sufficiently large primes p and integers b,*

$$\frac{\log p \log_2 p}{\mathcal{C}_4(p)} \ll f(p) \le g(p) \ll \frac{\log p \log_2 p \log_3 p}{\mathcal{C}_5(p)}, \qquad D(b) \ll \frac{b \log b \log_2 b \log_3 b}{\mathcal{C}_5(b)}.$$

In particular, $f(p) = \log p (\log_2 p)^{1+o(1)}$ and $D(p) = p \log p (\log_2 p)^{1+o(1)}$ along primes.

For an integer $N \ge 1$ let $t(N)$ be the least positive integer that cannot occur as the smallest denominator of a representation of one by distinct denominators not exceeding N (allowing $1 = 1/1$). Let $\mathcal{A}(q)$ be the maximum length of a rational arithmetic progression with positive common difference contained in $\mathcal{V}(q)$. Define also the weighted large-prime deletion loss

$$\mathcal{D}(N) = \sum_{\sqrt{N} < p \le N} \frac{d(p, \lfloor N/p \rfloor)}{p}.$$

Theorem 1.6 (Applications)**.** *For $L = \log N$, $U = \log_2 N$, $V = \log_3 N$,*

$$\frac{N\mathcal{C}_5(N)}{LUV} \ll t(N) \ll \frac{N\mathcal{C}_4(N)}{LU},$$
$$\log \mathcal{A}(q) \asymp \frac{q\mathcal{C}_4(q)}{\log q \log_2 q},$$
$$\mathcal{D}(N) = \frac{(U+V-\log \mathcal{C}_4(N))^2}{2L} + O(U/L).$$

Thus $t(N) = N/[\log N(\log_2 N)^{1+o(1)}]$ and

$$L\mathcal{D}(N) = U^2/2 + UV - U\log \mathcal{C}_4(N) + O(U).$$

The lower bound for $t(N)$ uses [4, Lemma 4.1, arXiv v1]*; the other results here do not.*

Theorem 1.7 (Prime-square loss and its sharp-threshold criterion)**.** *Define*

$$\Delta_2(N) = \sum_{N^{1/3}<p\le\sqrt{N}} \frac{d(p, \lfloor N/p^2 \rfloor)}{p^2}, \quad B = b(\sqrt{N}), \quad A_N = \frac{\sqrt{B}\log B}{\sqrt{N}\log N}. \tag{8}$$

Then, with τ as in Theorem 1.2,

$$2 \le \liminf_{N\to\infty} \frac{\Delta_2(N)}{A_N} \le \limsup_{N\to\infty} \frac{\Delta_2(N)}{A_N} \le 2\sqrt{\tau} < 2\sqrt{2}. \tag{9}$$

In particular

$$\Delta_2(N) \asymp \frac{(\log_2 N)^{3/2}}{\sqrt{N\log N\, W(N)}}. \tag{10}$$

Moreover, $\Delta_2(N) \sim 2A_N$ is equivalent to the assertion that for each fixed $t > 1$ and $\delta > 0$,

$$\#\{P \le p \le 2P : d(p, \lfloor tb(p) \rfloor) > \delta \log b(p)\} = o(P/\log P). \tag{11}$$

The assertion (11)*, and therefore the exact asymptotic with constant two, is not proved here.*

The proofs begin with the exact entropy decomposition, then treat the two distinct modular-mixing constructions. The exact grid and its arithmetic-progression obstruction follow. The prime-square application uses the precise entropy inverse and the quantitative exceptional-prime estimates, not simply an unquantified density-one statement. The open-problem section separates directed low-loss coverage from the uniform all-primes question. The appendix records the general rational-projection lemma and useful variants of the coverage construction.

2. Exact recursion and regular variation

Put $\mathcal{L}(n) = \operatorname{lcm}(1, \dots, n)$. We use the prime number theorem, including the classical error term $\pi(x) = \operatorname{Li}(x) + O(xe^{-c\sqrt{\log x}})$, and $\log \mathcal{L}(n) \sim n$.

Lemma 2.1 (Bounded-coefficient rigidity)**.** *Fix an integer $D \ge 4$, put $c = 1/4$ and, for all sufficiently large n (depending on D), let*

$$y_n = \frac{n}{c\log n} = \frac{4n}{\log n}, \quad k_r = \lfloor n/r \rfloor, \quad V_r = r[1, k_r] \quad (y_n < r \le n), \quad K_n = [1, n] \setminus \bigcup_{r>y_n} V_r.$$

Here and below r in these expressions is prime. These blocks are disjoint. Any relation $\sum_{a\le n} c_a/a = 0$ with integer $|c_a| \le D$ splits into zero relations on each V_r and on K_n. For $D = 4$ it is enough to take $n \ge n_0 = 2^{20}$.

Proof. First, for every integer $k \geq 1$,

$$\log \mathcal{L}(k) \leq 2k. \tag{12}$$

Indeed, the quotient $\mathcal{L}(k)/\mathcal{L}(\lceil k/2 \rceil)$ divides $\binom{k}{\lfloor k/2 \rfloor}$. For any prime power in $(\lceil k/2 \rceil, k]$, the corresponding term in the factorial valuation formula for the binomial coefficient is one; other terms are nonnegative. There is at most one such power of a given prime. Taking logarithms proves the induction step, since $2\lceil k/2 \rceil + k \log 2 \leq 2k$ for $k \geq 4$ by $\log 2 < 7/10$. The cases $1 \leq k \leq 3$ follow from $\mathcal{L}(1) = 1, \mathcal{L}(2) = 2, \mathcal{L}(3) = 6$.

Write $\ell = \log n$. For $n \geq 2^{20}$ we have $\ell > 40/3$, $\ell^2 < \sqrt{n}$ and $\sqrt{n} < y_n < n/3$. The square inequality follows from monotonicity of $x^2 e^{-x/2}$ for $x \geq 4$ and $20 \log 2 < 14$, $\sqrt{2^{20}} = 1024$. Consequently $k_r \leq \ell/4$ and, for coefficients bounded by four,

$$4\mathcal{L}(k_r) H_{k_r} \leq 4e^{2k_r} k_r \leq \ell\sqrt{n} < 4n/\ell < r.$$

For any larger fixed D the same strict inequality with D in place of four holds for sufficiently large n, since $\ell^2/\sqrt{n} \to 0$.

Because $y_n > \sqrt{n}$, distinct marker primes cannot divide the same integer in $[1, n]$. Multiplication of the original relation by r and reduction modulo r gives $\sum_{a \leq k_r} c_{ra}/a \equiv 0 \pmod r$: the denominators $a \leq k_r$ and all denominators outside V_r are units at r. Multiplying by $\mathcal{L}(k_r)$ gives an integer divisible by r and of absolute value strictly less than r. It is zero. Subtract all these block relations from the original relation to obtain the core relation. □

The sum map from the product of local distinct sumsets is one-to-one and onto, by Lemma 2.1. Thus

$$s(n) = s(K_n) + \sum_{r > y_n} s(k_r). \tag{13}$$

The same assertion holds for intersections of all the components with a specified subset of $[1, n]$. It is essential here that the product is of *distinct rational values*, not of arbitrary representations of those values. The least common multiple of the core divides $\prod_{r \leq y_n} r^{\lfloor \log n / \log r \rfloor}$. Its logarithm is $O_c(n/\log n)$, by the prime number theorem and the bound $O(\sqrt{n} \log n)$ for the contribution of higher powers. All core sums have this common denominator and size at most H_n, whence $s(K_n) = O_c(n/\log n)$.

Uniformly for $1 \leq j \leq \log n$, the prime number theorem with its error term gives

$$\pi(n/j) - \pi(n/(j+1)) = \frac{n}{j(j+1)\log n}\left(1 + O\left(\frac{\log_2 n}{\log n}\right)\right).$$

The relative error is uniform because the intervals have length at least a constant multiple of $n/(\log n)^2$. Using $s(j) \leq j \log 2$ in (13), the normalized total relative-error contribution is $O((\log_2 n)^2/\log n) = o(1)$. Extending the upper summation limit from $c \log n$ to $\log n$ adds only $O_c(1)$ to the normalized sum. A partially included terminal interval is similarly harmless. This proves (4).

The trivial bound $s(j) \leq j \log 2$ gives the first upper bound below; substituting it once more in the recursion gives the second. The lower bound follows from $s(j) \geq \pi(j) \log 2$, since sums of distinct primes' reciprocals are distinguished by reduction modulo each such prime. In detail, the sum of $1/(j \log j)$ up to $\log n$ is of order $\log_3 n$. Consequently

$$\frac{n \log_3 n}{\log n} \ll s(n) \ll \frac{n \log_2 n}{\log n}, \qquad s(n) \ll \frac{n}{\log n}(\log_3 n)^2. \tag{14}$$

For the second upper bound use $\sum_{2 \leq j \leq \log n} \log\log j/((j+1)\log j) = O((\log_3 n)^2)$, handling small terms separately.

Let $A(n) = \sum_{j \leq \log n} s(j)/(j(j+1))$. The lower bound just proved gives $A(n) \to \infty$. Replacing n by tn for fixed $t > 0$ changes A by $O_t(1/\log n)$, because the summation limit changes by

only a bounded number of terms. Formula (4) now proves regular variation. Monotonicity and $0 \le s(n+1)-s(n) \le \log 2$ give $s(b(P)) = \log P + O(1)$; regular variation yields the inverse formula in (5) by sandwiching with $(t \pm \varepsilon)b(P)$. In particular

$$\frac{\log P'}{\log P} \to 1 \quad \Longrightarrow \quad b(P')/b(P) \to 1, \tag{15}$$

uniformly whenever the ratio on the left tends to one uniformly. This follows directly by monotonicity and the two fixed-power bounds.

2.1. **The iterated-logarithm order.** For $A_*(x) = \prod_{j=3}^{h(x)} \log_j x$, the lower estimate $s(n) \gg (n/\log n)A_*(n)$ follows from [3, Theorem 1]: for $k \ge 4$ and $\log_k n \ge 3/2$ their lower bound is

$$(\log 2)\left(2 - \frac{3}{\log_k n}\right)\frac{n}{\log n}\prod_{j=3}^{k}\log_j n.$$

Here the theorem expressly permits variable k. Taking $k = h(n)$ bounds the parenthesis below by a fixed positive constant for large n.

For the upper estimate set $F(x) = \max_{2\le n\le x} s(n)\log n/n$. Inserting this bound into the recursion and using

$$\sum_{2\le j\le \log n}\frac{1}{(j+1)\log j} = \log_3 n + O(1)$$

gives

$$F(x) \le (\log_3 x + C)F(\log x)$$

for a fixed C and all sufficiently large x. In deriving this, the core error and the finite initial range can be absorbed using a positive lower bound for F. In particular, the coefficient of $\log_3 x$ is one, not a fixed constant greater than one. Iterate until the argument is bounded. The extra factor relative to the product of the iterated logarithms stays bounded, since

$$\prod_{j=3}^{h(x)}(1 + C/\log_j x) \le \exp\left(C\sum_{j=3}^{h(x)}(\log_j x)^{-1}\right) = O(1).$$

The last sum is bounded because the iterated logarithms, read in reverse starting from a value at least A_{cut}, grow at least geometrically. The choice of a fixed stopping argument rather than the chosen cutoff changes only a bounded final product. This proves the desired upper estimate.

For the inverse note from (14) that $b(P) = (\log P)^{1+o(1)}$. If $\log u/\log v$ stays between two positive constants, then $\log_2 u - \log_2 v = O(1)$ and subsequent iterates contract while they remain above the cutoff. The logarithmic difference of the common parts of their products A_* is bounded by a constant times the sum of reciprocal iterates just used. Once one sequence stops, the other has at most a bounded tail of bounded factors. Hence $A_*(u) \asymp A_*(v)$. Apply this with $u = b(P)$ and $v = \log P$. Since $A_*(\log P) = W(P)$ and $s(b(P)) = \log P + O(1)$, the inverse estimate in (6) follows. The cutoff product may jump by bounded factors at termination points; no precise asymptotic for the slow factor is being asserted. This completes Theorem 1.1.

2.2. **The top-level entropy pool.** Write $\mathcal{G}_n = \bigcup_{r>y_n} V_r$. The exact recursion and (14) give

$$s(\mathcal{G}_n) = (1-o(1))s(n), \qquad \mathcal{H}(\mathcal{G}_n) = \sum_{r>y_n}\frac{H_{k_r}}{r} \ll \frac{(\log_2 n)^2}{\log n} = o(1). \tag{16}$$

For the mass estimate use $H_{k_r} = O(\log_2 n)$ and partial summation for primes in $(y_n, n]$. For later use set $L = \log P$, $\lambda = \log L$, and $m \asymp b(P)$. Formula (14) implies

$$\log m \sim \lambda, \quad \frac{L\lambda}{\log\lambda} \ll m = o(L\lambda), \quad m/\log m = o(L), \quad L\log\lambda = o(m). \tag{17}$$

In particular the number of top-level markers is $O(m/\log m) = o(L)$ and their maximum individual entropy is $O(\log m) = o(L)$.

For reference, the upper half of Theorem 1.1 gives the uniform counting estimate

$$\log|\mathcal{V}(x)| \ll \frac{x}{\log x}\mathcal{C}_3(x) \tag{18}$$

for sufficiently large real x. This is the counting upper bound used throughout the later applications. The proof has one constant independent of the number of iterates. It is based on the classical large-prime counting mechanism of [2].

3. Prime-average collision estimates and a Fourier damper

For a real finite probability measure μ write $\widehat{\mu}(\theta) = \int e(\theta x)\, d\mu(x)$, where $e(z) = e^{2\pi i z}$. If its support is rational with denominators prime to p, write μ_p for its pushforward to $\mathbb{F}_p$ and $\widehat{\mu}_p(k) = \mathbb{E}e(kX/p)$, with X understood modulo p. Write $\widetilde{\mu}$ for its reflection.

Lemma 3.1 (Height and collision budget)**.** *Suppose a rational set U of size M has common denominator Q prime to all $p \in [P, 2P]$, and for distinct $u, v \in U$, $\log|Q(u-v)| \le Cm$. If μ_p is its uniform reduction, then*

$$\sum_{P\le p\le 2P}\sum_{k \bmod p} |\widehat{\mu}_p(k)|^2 \ll P\left(\frac{\Pi}{M} + \frac{m}{L}\right), \tag{19}$$

where Π counts the primes of the interval. More generally for a rational probability measure K of the same common denominator, with $\log|Qz| \le Cm$ on its nonzero support, one has

$$\sum_{P\le p\le 2P}\sum_{k \bmod p} \widehat{K}_p(k) \ll P\left(\Pi K(\{0\}) + \frac{m}{L}\right). \tag{20}$$

Proof. A nonzero integer of logarithmic size $O(m)$ has at most $O(m/L)$ distinct prime divisors in $[P, 2P]$. Parseval identifies the first inner sum as $p\mathbb{P}(u \equiv v \pmod p)$, with independent uniform samples. The diagonal contributes at most $2P\Pi/M$. For the off-diagonal interchange summations and apply the divisor bound. For the second formula, the inner sum is $pK\{z : z \equiv 0 \pmod p\}$. The zero atom supplies the first term; the same divisor budget bounds the others. When using (20) to bound Fourier coefficients below, we will also require $\widehat{K}_p(k) \ge 0$. □

For subset sums at height $m \asymp b(P)$ and their difference sets, the hypotheses follow with $Q = \mathcal{L}(m)$ because $\log Q = O(m)$ and the mass is bounded by H_m. If $M \ge P^{1+\delta}$ or $K(\{0\}) \le P^{-1-\delta}$ with fixed $\delta > 0$, the right sides are $O(P\lambda)$. Markov's inequality shows that outside $O(P\lambda e^{-\lambda^2})$ exceptional primes the respective inner energy is at most e^{λ^2}.

The next lemma supplies a small-mass pool reducing every nonzero Fourier coefficient at once. We state a slightly stronger form to handle the uniform distinct-value distribution later.

Lemma 3.2 (Simultaneous phase suppression)**.** *Take $K_* = \lceil 20\log\lambda\rceil$ and the denominator set*

$$\mathcal{D} = \{ra : 100L < r \le 200L,\ r \text{ prime},\ 1 \le a \le K_*\}.$$

Except for $O(P\exp(-cL\log\lambda/\lambda))$ primes $p \in [P, 2P]$, for every $1 \le k < p$,

$$E_p(k) = \sum_r \sum_{a\le K_*} \|k(ra)^{-1}/p\|^2 > \lambda^4. \tag{21}$$

The mass of $\mathcal{D}$ is $O(\log K_*/\lambda) = o(1)$ and its maximal denominator is $o(b(P))$. For the independent Bernoulli subset-sum law on this set, $|\widehat{\sigma}_p(k)| \le e^{-2\lambda^4}$ on good primes.

Proof. The mass and height statements follow by prime counting and (17). Suppose $E_p(k) \le \lambda^4$ for some k. Set $\epsilon = \lambda^3/\sqrt{L}$. At most L/λ^2 denominators have phase distance exceeding ϵ. A good marker r is one for which none of its K_* denominators is bad. Thus all but $O(L/\lambda^2)$ of the $\asymp L/\lambda$ markers are good.

Let z_d be the centered representative of $kd^{-1} \bmod p$, and write $dz_d = k + pu_d$. For good d, $|u_d + k/p| \le \epsilon d$ and $u_d \equiv -kp^{-1} \pmod d$. At a good marker r, $u_{ra} - u_r$ is divisible by r and has absolute value at most $\epsilon(a+1)r < r$. It is zero. Thus every u_r for a good marker is congruent to the same residue $b = -kp^{-1} \bmod \mathcal{L}(K_*)$. By the prime number theorem, $\lambda^{10} \le \mathcal{L}(K_*) \le \lambda^{40}$ for large P. For fixed b, each good marker permits at most

$$O(1 + \epsilon r/\mathcal{L}(K_*)) \le \sqrt{r}\,\lambda^{-6}$$

possible integers u_r; the unknown center $-k/p$ only enlarges the interval by one.

Choose a subset J of good markers whose product R_J first exceeds $4P^2$. Such a subset exists since their total logarithmic mass is $(100 + o(1))L$; then $R_J \ll LP^2$ and $|J| \gg L/\lambda$. For fixed $J, b, (u_r)_{r\in J}$, at most one pair (p,k) with $1 \le k < p$ is possible. Indeed equality of residues for two pairs implies $R_J \mid kp' - k'p$, and the integer on the right has absolute value less than $4P^2$. Equality of the reduced fractions forces equality of the pairs. The total number of possible vectors is at most $\sqrt{R_J}\lambda^{-6|J|}$. Summing over $2^{O(L/\lambda)}$ choices of J and at most λ^{40} values of b proves the claimed exceptional bound. Finally, $|(1 + e(u))/2| \le e^{-2\|u\|^2}$ gives the Bernoulli Fourier bound by multiplication. □

In applying this lemma at $m \asymp b(P)$, all markers of $\mathcal{D}$ belong to the top-level decomposition at height m, since $m/(c \log m) = o(L)$ and $k_r \ge K_*$. Removing the full blocks of these markers costs at most $O((L/\lambda)(m/L)) = o(L)$ in entropy. This is why we can keep an independent suppression resource without losing a positive proportion of entropy.

4. The modular transition and the two-pool estimate

Fix $0 < t < 1$, and take $m = \lfloor tb(2P) \rfloor$. Then $S(m) = P^{t+o(1)}$. For each nonzero harmonic subset sum x at height m, $\mathcal{L}(m)x$ is a positive integer of logarithmic size $O(m)$. It has at most $O(m/L) = P^{o(1)}$ prime divisors in $[P, 2P]$. The union over all such values has size $P^{t+o(1)}$. Outside this exceptional set, no nonempty subset of $[1, m]$ has residue zero. Hence $d(p, q) = H_q$ for every $q \le \lfloor tb(p) \rfloor$, proving the first part of Theorem 1.2. This also proves its lower bound for $z(p)$.

Now fix $t > 1$ and $m = \lfloor tb(P) \rfloor$. Remove the full marker blocks containing $\mathcal{D}$ from $\mathcal{G}_m$ and call the remaining pool G'. It has entropy $(t + o(1))L$ by (16) and the preceding observation. Take its uniform distinct-value distribution μ_p modulo p and convolve with the independent Bernoulli law σ_p on $\mathcal{D}$. The two resources are disjoint and their total mass is $o(1)$. Formula (19) and Lemma 3.2 give, outside $O_t(Pe^{-c_t\lambda^2})$ primes,

$$\frac{1}{p}\sum_{h \bmod p} |p(\mu_p * \sigma_p)(h) - 1|^2 = \sum_{k\ne 0} |\widehat{\mu}_p(k)\widehat{\sigma}_p(k)|^2 \le e^{-3\lambda^2}. \tag{22}$$

The number of omitted residues is at most $pe^{-3\lambda^2}$, and the total variation distance to uniform is at most $e^{-3\lambda^2/2}/2$. The support is a subset of ordinary sums. Since it exceeds half of $\mathbb{F}_p$, every residue is a difference of two support points. Removing their common denominators produces disjoint positive and negative sets of total mass $o(1)$. This is signed coverage, not necessarily a deletion representation. Below the entropy threshold even the full set of rational values has size $p^{t+o(1)}$, so the threshold for near-complete coverage is exactly one in this sense.

There is a useful simpler full-coverage proof for $t > 2$. Partition the remaining top-level blocks into two classes by assigning the next block to the class with the smaller current entropy. The discrepancy is at most the largest block entropy, $o(L)$; each class has entropy $(t/2 + o(1))L$. Their

uniform distinct-value laws $\mu_{1,p}, \mu_{2,p}$ each satisfy the collision estimate with energy at most e^{λ^2} on good primes. Fourier inversion and Cauchy–Schwarz give

$$|p(\mu_{1,p} * \mu_{2,p} * \sigma_p)(h) - 1| \le \max_{k\neq 0} |\widehat{\sigma}_p(k)| \left(\sum_k |\widehat{\mu}_{1,p}(k)|^2 \sum_k |\widehat{\mu}_{2,p}(k)|^2 \right)^{1/2} \le e^{-\lambda^2}.$$

Thus all residues have positive mass. The next section improves the factor two without requiring two independent high-entropy pools.

5. Uniform distinct-value coverage strictly below factor two

The recursive structure permits an absolute Fourier majorant whose zero atom is smaller than the one provided by the trivial Cauchy–Schwarz split. We give the required finite certificate and leaf count, so the proof here is independent of the companion work on absorption capacities.

Lemma 5.1 (Recursive two-element components, with an explicit constant)**.** *Apply the decomposition of Lemma 2.1 with $D = 4$ recursively to the children $[1, k_r]$, treating each core as a single component and stopping at height $n_0 = 2^{20}$. All distinct rational sums factor into independent local distinct-value components after rescaling. For any finite collection of roots of sizes greater than n_0, writing S_{ent} for the sum of their entropies and ℓ_2 for the number of two-element terminal children, one has*

$$\ell_2 \ge \frac{S_{\mathrm{ent}}}{184}. \tag{23}$$

Proof. We give numerical constants using the classical bounds [7, (3.1), (3.2), (3.6)]

$$\begin{aligned} \pi(x) &> \frac{x}{\log x}\left(1 + \frac{1}{2\log x}\right) && (x \ge 59), \\ \pi(x) &< \frac{x}{\log x}\left(1 + \frac{3}{2\log x}\right) && (x > 1), \\ \pi(x) &\le \frac{63x}{50\log x} && (x > 1). \end{aligned} \tag{24}$$

The last bound is a weakening of $1.25506x/\log x$. The elementary inequalities $2/3 < \log 2 < 139/200 < 7/10$, $\log 3 < 6/5$ and $\log(n_0 + 1) < 14$ will suffice. For example, expanding $\log((1+u)/(1-u)) = 2\sum_{j\ge 0} u^{2j+1}/(2j+1)$ at $u = 1/3, 1/2$ proves the first two bounds by summing a geometric tail; the last follows from $\log(1 + 2^{-20}) \le 2^{-20}$.

Consider an internal node $n > n_0$, and write $\ell = \log n$ and $y = 4n/\ell$. All primes $n/3 < r \le n/2$ are markers by Lemma 2.1, and their children have height two. From (24),

$$\pi(n/2) \ge \frac{n}{\ell}\left(\frac{1}{2} + \frac{7}{12\ell}\right), \qquad \pi(n/3) \le \frac{n}{\ell}\left(\frac{1}{3} + \frac{13}{12\ell}\right).$$

For the first estimate use $(1-u)^{-1} \ge 1 + u$ with $u = (\log 2)/\ell$ in (3.1). For the second put $b = \log 3$ and use $1 - b/\ell \ge 91/100$, valid since $\ell > 40/3$. In detail the coefficient of $1/\ell$ in the upper bound, after factoring n/ℓ, is at most

$$\frac{1}{3}\left(\frac{6/5}{91/100} + \frac{3/2}{(91/100)^2}\right) < \frac{13}{12}.$$

It follows that the number of two-element children at this node is at least

$$\pi(n/2) - \pi(n/3) \ge \frac{n}{\ell}\left(\frac{1}{6} - \frac{1}{2\ell}\right) \ge \frac{n}{8\ell}. \tag{25}$$

Next the core entropy satisfies the explicit bound

$$s(K_n) \le \frac{8n}{\ell}. \tag{26}$$

Indeed its least common denominator has logarithm at most $\vartheta(y) + \sqrt{n}\,\ell$, where $\vartheta(y) = \sum_{p \le y} \log p$. The latter term accounts for all higher powers at the at most $\sqrt{n}$ prime bases below $\sqrt{n}$. Since $\vartheta(y) \le \pi(y) \log y \le (63/50)y$, $\sqrt{n}\,\ell \le n/\ell$ and $\log(H_n + 1) \le \ell \le n/\ell$, counting the grid points in $[0, H_n]$ gives (26).

All terminal children at this node have $k_r \le n_0$ and therefore $r > a := \max(y, n/(n_0 + 1))$. We have $\log a \ge 9\ell/10$, since $\log(\ell/4) \le \ell/10$ follows from $\log v \le v/e < (2/5)v$ for $v > 0$. Also $\log(n/a) \le \log(n_0 + 1) < 14$. By finite prime partial summation, dropping the nonpositive lower-end term,

$$\sum_{a < r \le n} \frac{1}{r} \le \frac{\pi(n)}{n} + \int_a^n \frac{\pi(u)}{u^2}\, du \le \frac{63}{50\ell} \left(1 + \frac{10}{9} \log(n/a)\right).$$

Using $s(k_r) \le k_r \log 2$ and $k_r \le n/r$, the sum of the entropies of the terminal children at this node is at most

$$\sum_{r:\, k_r \le n_0} s(k_r) \le \frac{n}{\ell} \frac{63}{50} \frac{7}{10} \left(1 + \frac{140}{9}\right) < \frac{15n}{\ell}, \tag{27}$$

where the sum is over marker primes only. This estimate charges terminal children directly; a bound by their maximum entropy would give a much smaller numerical gain.

Every child is smaller than its parent, and the forest is finite. Exact entropy additivity, applied recursively, expresses S_{ent} as the sum of all internal core entropies and all terminal-child entropies. All roots are internal. Summing (26) and (27) over internal nodes gives $S_{\mathrm{ent}} \le 23 \sum n/\log n$. Summing (25) over the same nodes gives $\ell_2 \ge (1/8) \sum n/\log n$, which proves (23). Exact factorization follows from Lemma 2.1 at each node; scaling along the tree preserves all local cardinalities. The same relation splitting applies to differences of ordinary subset sums (integer coefficients of absolute value at most one). □

For a large height q let

$$J_0 = \left\lfloor \sqrt{s(q) \log q / q} \right\rfloor, \qquad U = \bigcup_{y_q < r \le q/J_0} V_r.$$

Then $J_0 \to \infty$. The discarded top-level entropy is $O(J_0 q/\log q) = o(s(q))$ by grouping r according to $k_r < J_0$ and using $s(j) \le j \log 2$. Thus $s(U) = (1 - o(1))s(q)$, and a common denominator D_U of its sums satisfies

$$D_U = \mathcal{L}(\lfloor c \log q \rfloor) \prod_{y_q < r \le q/J_0} r, \qquad \log D_U = O(q/J_0 + \log q) = o(q). \tag{28}$$

All retained children have sizes at least J_0; consequently Lemma 5.1 applies for large q. For any additional deletion of marker blocks of total entropy $o(s(q))$, the remaining roots still satisfy the same lemma.

Lemma 5.2 (Four-point majorant and global zero atoms)**.** *With the explicit constant $\sigma_0 = 1/2 + 1/9200 = 4601/9200$, the uniform law μ_U of distinct sums of U has two probability majorants K_1, K_2 supported on $\mathscr{S}(U) - \mathscr{S}(U)$, satisfying on the real line*

$$\widehat{K_i} \ge 0, \quad |\widehat{\mu}_U| \le \sqrt{\widehat{K_1}\widehat{K_2}}, \qquad K_i(\{0\}) \le e^{-\sigma_0 s(q)} \quad (i = 1, 2). \tag{29}$$

The same conclusion holds after deleting from U a union of marker blocks of total entropy $o(s(q))$.

Proof. A two-element leaf has denominators $w, 2w$ for some integer w, so multiplication by $2w$ identifies its four values with $\{0,1,2,3\}$. The modulus of the Fourier transform of their uniform distribution at angle θ is $|\cos(\theta/2)\cos\theta|$. Define

$$F(\theta) = \frac{49 + 33\cos\theta + 15\cos 2\theta + 3\cos 3\theta}{100}.$$

This is the Fourier transform of a symmetric probability law on $[-3,3] \cap \mathbb{Z}$, with atoms $49/100$ at zero, $33/200$ at each of ± 1, $15/200$ at each of ± 2, and $3/200$ at each of ± 3. It majorizes the modulus for all real θ. Here is a purely algebraic certificate for this last assertion. Put $z = |\cos(\theta/2)|$ and $f(z) = (24z^6 - 6z^4 + 7)/25$; the multiple-angle identities give $F(\theta) = f(z)$. On $0 \le z \le 1$ one has

$$f(z) - 5/18 = \frac{(6z^2-1)^2(12z^2+1)}{450} \ge 0, \tag{30}$$

$$2/27 - z^2(1-2z^2)^2 = \frac{(2-3z^2)(6z^2-1)^2}{27}. \tag{31}$$

When $z^2 \le 1/2$, the right side of (31) is nonnegative, and $2/27 < 25/324 = (5/18)^2$. Hence $|z(2z^2-1)| \le f(z)$. When $z^2 \ge 1/2$, the required difference is

$$\begin{aligned} 25(f(z) - z(2z^2-1)) &= (1-z)B(z), \\ B(z) &= 7 - 2z^3 + 32(z - z^3) + 32(z^2 - z^3) \\ &\quad + 24(z^3 - z^4) + 24(z^3 - z^5). \end{aligned} \tag{32}$$

Each parenthesized difference in $B(z)$ is nonnegative on $[0,1]$, and $7 - 2z^3 \ge 5$. Thus (32) is nonnegative. This proves the desired majorization and nonnegativity by polynomial identities, positive factors and exact rational inequalities, without a numerical trigonometric test.

For each two-element leaf take two copies of this law at its actual scale $1/(2w)$, both supported on the leaf difference set. For every other component with uniform law μ use the pair $(\mu * \widetilde{\mu}, \delta_0)$ or the reversed pair. Each pair dominates the modulus by the geometric mean of two nonnegative Fourier transforms. Convolve the first and second members respectively across the components. The zero atom of each global convolution is the product of the local zero atoms by rigidity of differences in Lemma 2.1. Write $S_U = s(U)$, or the entropy of the remaining union if further roots have been removed. The negative logarithm of the product of the two zero atoms is at least

$$S_U + 2\ell_2 \log(50/49) \ge \left(1 + \frac{2\log(50/49)}{184}\right) S_U,$$

because an ordinary component contributes its entropy, whereas a two-element leaf contributes $-2\log(49/100) = \log 4 + 2\log(50/49)$. The two-element leaves contribute equally to both capacities. Assign each other component's entropy to the smaller of the accumulated capacities. After each assignment the absolute discrepancy is at most the maximum assigned component entropy (by induction using $|d - a| \le \max(d, a)$ for $d, a \ge 0$). All components are contained in roots of height at most $c \log q$, hence this discrepancy is $O(\log q) = o(s(q))$. Thus, for both $i = 1, 2$,

$$-\log K_i(\{0\}) \ge \left(\frac{1}{2} + \frac{\log(50/49)}{184}\right) S_U - O(\log q) = \left(\frac{1}{2} + \frac{\log(50/49)}{184} - o(1)\right) s(q).$$

Since $\log(50/49) > 1/50$ (integrate $1/x$ on $[49, 50]$), the coefficient has a strict positive margin above $\sigma_0 = 1/2 + 1/9200$. Therefore (29) holds for this fixed rational σ_0 for all sufficiently large q. Removal of additional roots with entropy $o(s(q))$ preserves the same conclusion, since $S_U = (1 - o(1))s(q)$ and (23) applies to the remaining forest. In particular the threshold $1/\sigma_0 = 9200/4601$ in Theorem 1.2 is now explicit. No optimality of this sufficient factor is claimed. □

The ordinary Bernoulli damper alone does not prove the desired assertion about uniform distinct values on the original top-level pool. The following matching observation provides the necessary variant.

Lemma 5.3 (Uniform distinct-value damper)**.** *If $v \in A$ and X is uniform on the distinct sums of A, then for any real frequency θ,*

$$|\mathbb{E}e(\theta X)| \le e^{-\|\theta/v\|^2}.$$

The analogous finite-field estimate holds with θ/v replaced by kv^{-1}/p. For the uniform distinct-value law ν_p on the union of the full top-level marker blocks $100L < r \le 200L$ at height $q \asymp b(P)$,

$$\max_{k\neq 0} |\widehat{\nu}_p(k)| \le e^{-20\lambda^2} \tag{33}$$

on all good primes of Lemma 3.2.

Proof. Writing $T = \mathscr{S}(A)$ and $T_0 = \mathscr{S}(A \setminus \{v\})$, we have $T = T_0 \cup (T_0 + 1/v)$. Join each $x \in T_0$ to $x + 1/v$. The graph is a disjoint union of finite paths, with $|T_0| \ge |T|/2$ edges. There is a matching of at least $|T|/4$ edges. Group the complex sums in pairs and use $1 - |\cos(\pi u)| \ge 2\|u\|^2$ to obtain the first bound. The same reasoning in the modular Fourier sum groups rational pairs before reduction, and so is valid even if reductions collide. Apply this bound to each $v = ra$ with $a \le K_*$ in the corresponding full marker block. The geometric mean gives a bound by $\exp(-\sum_{a\le K_*} \|k(ra)^{-1}/p\|^2/K_*)$ for that block. Independent distinct-value laws across the blocks multiply, giving $\exp(-E_p(k)/K_*)$. Lemma 3.2 proves (33) for large P. □

Proof of the strict coverage assertion in Theorem 1.2. Set $q = \lfloor tb(P)\rfloor$ for fixed $t > 1/\sigma_0$, and $\tau = 1/\sigma_0 \in (1, 2)$. Use the entire top-level pool $\mathcal{G}_q$, with uniform law ρ_p^{val} after reduction on its distinct rational values. Split its marker blocks into the auxiliary blocks of Lemma 5.3, the set U of Lemma 5.2 with those auxiliary blocks excluded, and the remaining blocks. The excluded entropy is $o(L)$ by (17), so (29) holds with the same fixed σ_0 (chosen with strict margin). The three component laws are independent, and their convolution is supported on ordinary subset sums. The two Fourier majorants are auxiliary laws on differences. The two majorants have zero mass at most $P^{-1-\delta}$ for some fixed $\delta > 0$ since $s(q) = (t + o(1))L$. Their supports have the common denominator of (28), in particular with logarithmic height $O(q)$. Formula (20) and Markov give both nonnegative energies at most e^{λ^2} outside $O(P\lambda e^{-\lambda^2})$ primes. The real-line Fourier majorization implies the same inequality for every modular frequency: for a common denominator D prime to p, reduction and evaluation at k amount to evaluating the real transform at $kD\bar{D}/p$ with $D\bar{D} \equiv 1 \pmod p$. Thus Fourier inversion, (33), and Cauchy–Schwarz give

$$|p\rho_p^{\mathrm{val}}(h) - 1| \le e^{-20\lambda^2} \left(\sum_k \widehat{K}_{1,p}(k) \sum_k \widehat{K}_{2,p}(k)\right)^{1/2} \le e^{-19\lambda^2}.$$

This implies (7). The mass of the whole pool is $o(1)$ by (16). The count of primes excluded by the damper is also $O(Pe^{-c_t\lambda^2})$, because $L \log\lambda/\lambda \gg \lambda^2$. Therefore all residues have an ordinary representation at height q on the asserted density-one set. Since $\mathcal{G}_q$ has no denominator 1, representing -1 and appending 1 gives a nonempty zero sum at the same height. This proves the upper bounds for $z(p)$ and $g(p)$ using $b(p) \ge b(P)$. The lower bound $f(p) \ge b(p)$ follows simply from p distinct residues requiring at least p distinct rational values, since reduction cannot increase the number of values. The remaining assertions were proved in Section 4. □

6. Density-one coverage at the counting scale

Throughout this section $L = \log X$, $U = \log L$, $V = \log U$, $F = \mathcal{C}_4(X)$ and $Q = \lfloor CLU/F\rfloor$ for a sufficiently large fixed C. The constants in the asymptotic estimates may depend on C and the

fixed cutoff. We use two elementary properties of truncated products:

$$\mathcal{C}_j(\log x) = \mathcal{C}_{j+1}(x), \qquad \mathcal{C}_j(y) \asymp \mathcal{C}_j(x) \quad (x^{1/2} \le y \le x^2,\ j \ge 2). \tag{34}$$

For the latter, the second iterates differ by at most $\log 2$ and subsequent differences contract by a factor at most $1/A_{\mathrm{cut}}$ as long as both preceding iterates exceed the cutoff. Hence the sum of logarithms of the ratios of the common factors is bounded. At the first unmatched factor the larger iterate is at most $A_{\mathrm{cut}} + \log 2$; its next iterate lies below the cutoff, so at most one bounded factor is lost. This proves the comparison uniformly even when the depth changes. The same backward growth of the iterates gives $F \to \infty$ and $F = U^{o(1)}$. In particular $(U/F)/\log U \to \infty$.

6.1. **Augmenting the rectangular block by its outer primes.** Put

$$R = \{r \text{ prime} : 4L < r \le 8L\}, \qquad M = \lfloor Q/(8L) \rfloor,$$
$$S = \{s \text{ prime} : M/2 < s \le M\}, \qquad T_S = \prod_{s \in S} s.$$

Use the set

$$A = R \cup \{rs : r \in R,\ s \in S\}.$$

For all large X, $M \asymp_C U/F$, $M < 4L$, and $A \subseteq [1, Q]$ with all denominators distinct (each composite product has a unique prime factor in R). All denominators are smaller than p for $X < p \le 2X$. The PNT gives $|R| \sim 4L/U$, $|S| \asymp_C U/(FV)$ and

$$\log T_S \sim M/2 \sim CU/(16F). \tag{35}$$

Here $\log M \sim V$, since $\log F = o(V)$. In addition,

$$|A| = o(L), \qquad \mathcal{H}(A) = \sum_{r \in R} r^{-1}(1 + \sum_{s \in S} s^{-1}) = O(1/U).$$

Lemma 6.1 (Augmented-block rational reconstruction)**.** *For fixed sufficiently large C there are constants $a, c_1 > 0$ such that, for $t = \exp(aU/F)$,*

$$\#\{(p,h) : X < p \le 2X,\ p \text{ prime},\ 1 \le h < p,\ E_A(p,h) \le t\} \ll X \exp(-c_1 L/F).$$

Proof. Use least absolute integer residues $v_r \equiv hr^{-1} \pmod p$ and $u_{rs} \equiv h(rs)^{-1} \pmod p$. Write $rv_r = h + pk_r$ and set $\delta = 4\sqrt{t/|R|}$. If $E_A(p,h) \le t$, at most $|R|/16$ rows have $|v_r| > \delta p$. On every other row $|k_r| \le 8L\delta + 1$.

An inner phase exceeding $1/(4M)$ contributes more than $1/(16M^2)$ to the energy. Hence at most $16M^2 t = o(|R|)$ rows contain any such phase; here $t = L^{o(1)}$ because $F \to \infty$. On all remaining rows every su_{rs} lies in $[-p/4, p/4]$ and is congruent to v_r. The least absolute residue is unique, so $su_{rs} = v_r$, and $T_S \mid v_r$. Thus at least $7|R|/8$ rows satisfy both

$$|k_r| \le 8L\delta + 1, \qquad k_r \equiv -hp^{-1} \pmod{T_S}.$$

The inverse modulo T_S is legal since p is larger than every prime dividing it. Write κ for this common residue. For each good row and fixed κ the number of candidate integers k_r is at most $J_0 = O(1 + L\delta/T_S)$, uniformly in the row. By (35),

$$\log(L\delta) = U/2 + \tfrac{1}{2}\log(tU) + O(1), \qquad \log J_0 \le U/2 - c_2 U/F$$

for some $c_2 > 0$ when a is chosen sufficiently small in terms of C. Indeed $\log U = o(U/F)$ and $\log T_S \sim CU/(16F)$; also $\log(L\delta/T_S) = (1/2 + o(1))U$, so the additive one in J_0 causes no problem.

Choose $\ell = \lceil \log(4X^2)/\log(4L) \rceil + 1$ good rows, possible since $\ell = (2 + o(1))L/U$. Their prime product exceeds $4X^2$. For each selected row $k_r \equiv -hp^{-1} \pmod r$. If two pairs (p,h) and (p',h') have the same recorded values on these rows, their prime product divides $hp' - h'p$, whose absolute value is less than $4X^2$. Therefore $h/p = h'/p'$. Both fractions are reduced because p, p' are prime and h, h' are nonzero least positive representatives, so the two pairs are identical. Fix a

deterministic choice of ℓ good rows (or count every choice). There are at most $T_S 2^{|R|} J_0^\ell$ records. Now

$$\ell U/2 = L + O(L/U + U), \qquad \ell U/F \asymp L/F.$$

The logarithm of the record count is at most $L - c_3 L/F$ because $F = o(U)$ and $\log T_S = O(U/F)$. This proves the lemma. □

6.2. **Dissociated high-entropy blocks and mixing.**

Lemma 6.2 (Dissociated sets at the counting scale)**.** *For all large x there is a set $\mathcal{U}_x \subseteq [1, x] \cap \mathbb{Z}$ with pairwise distinct rational reciprocal subset sums, such that*

$$|\mathcal{U}_x| \gg \frac{x}{\log x} \mathcal{C}_3(x).$$

Proof. Use the set $\mathcal{U}$ defined in [3, Section 2]: $a \in \mathcal{U}$ if $1/a$ is not a signed sum with coefficients in $\{-1, 0, 1\}$ of reciprocals of smaller positive integers. Any nontrivial signed relation on $\mathcal{U} \cap [1, x]$ would contradict the defining property of its largest denominator. Thus the subset sums are distinct. The proof of [3, Theorem 1], specifically (7), (9), and (14) before applying Lemma 2, bounds $|\mathcal{U} \cap [1, N]|$ from below by

$$2\left(1 - \frac{1.4}{\log_k N}\right) \frac{N}{\log N} \prod_{j=3}^{k} \log_j N$$

when k is sufficiently large and $\log_k N \geq A_{\mathrm{cut}} > 3$. Take $N = \lfloor x \rfloor$ and the largest such k; the cutoff comparison (34) absorbs rounding. This uses the quantitative result of BGMS as an external input, not just the cardinality of $\mathcal{V}(x)$, which by itself would not ensure a dissociated subset of this size. □

Proof of Theorem 1.3. For large X, $Q \geq L$ and $\log Q \sim U$. By monotonicity $\mathcal{C}_3(Q) \geq \mathcal{C}_3(L) = F$. Lemma 6.2 therefore gives at least $c_4 CL$ dissociated denominators in $[1, Q]$, with $c_4 > 0$ independent of the chosen fixed C (for X sufficiently large depending on it). Delete A, which has $o(L)$ elements. Also delete all denominators up to $y = Q/\sqrt{U}$. Their number in a dissociated set is at most $\log |\mathcal{V}(\lfloor y \rfloor)|/\log 2$. The elementary counting bound $\log |\mathcal{V}(v)| \ll v \log_2 v/\log v$ follows by the entropy recursion (4), using the trivial bound $s(w) \leq w \log 2$ for each block. It gives

$$|\mathcal{U}_Q \cap [1, y]| \ll LV/(F\sqrt{U}) = o(L).$$

With fixed K large enough that $K \log 2 > 3$, take C sufficiently large to choose two disjoint sets H_1, H_2 of $\lceil KL \rceil$ remaining denominators each. Both have distinct rational subset sums and are disjoint from A. Their total reciprocal mass is at most $(2KL + 2)/y = O_{C,K}(F/\sqrt{U}) = o(1)$. Hence $\mathcal{B}_X = A \cup H_1 \cup H_2$ has the required real mass, uniformly in p.

For a finite set D use the Bernoulli transform $\Phi_{p,D}(h) = \prod_{a \in D}(1 + e_p(ha^{-1}))/2$, where $e_p(x) = \exp(2\pi i x/p)$. For $i = 1, 2$, orthogonality gives

$$\sum_{h \bmod p} |\Phi_{p,H_i}(h)|^2 = p 4^{-|H_i|} \sum_{J,J' \subseteq H_i} \mathbf{1}_{p | N_{J,J'}},$$

where $N_{J,J'} = \mathrm{lcm}(1, \ldots, Q)(\mathcal{H}(J) - \mathcal{H}(J'))$ is an integer. For $J \neq J'$ it is nonzero and $\log |N_{J,J'}| = O(Q)$ by the prime-power estimate for the lcm and the bound on the harmonic mass. Hence at most $O(Q/L)$ primes in $(X, 2X]$ can divide it. Summing over primes, separating the diagonal, yields

$$\sum_{X < p \leq 2X} \sum_{h \bmod p} |\Phi_{p,H_i}(h)|^2 \ll X^2 2^{-|H_i|} + XQ/L \ll XU/F. \tag{36}$$

This bound is independent of the specific choices of the two dissociated subsets.

Outside the exceptional primes of Lemma 6.1, for every $h \neq 0$ we have $|\Phi_{p,A}(h)| \leq e^{-2t}$, by the cosine estimate $|\cos \pi v| \leq e^{-2\|v\|^2}$. By disjointness and Cauchy–Schwarz, (36) yields

$$\sum_{p \text{ good}} \sum_{h \neq 0} |\Phi_{p,\mathcal{B}_X}(h)| \ll X(U/F)e^{-2t}.$$

Markov's inequality discards at most $O(X(U/F)e^{-t})$ additional primes to achieve absolute nonzero Fourier mass at most e^{-t}. Since $t/U^A \to \infty$ for each fixed A, $t = L^{o(1)}$, and $L/F \gg t$, both exceptional sets together have size $O(Xe^{-ct})$ for a fixed $c > 0$. Fourier inversion gives the claimed uniform representation count. In particular $g(p) \leq Q$ outside that set; (34) compares F with $\mathcal{C}_4(p)$. The inequality $f(p) \geq b(p)$ and Theorem 1.1 supply the matching lower bound, independently of the mixing theorem. Summing exceptions over dyadic intervals gives a density-one set relative to the primes, since $Xe^{-ct} = o(X/L)$; the smaller intervals below $\sqrt{X}$ can also be discarded. This proves Theorem 1.3. □

The saving in Lemma 6.1 is the joint use of the small height of k_r (from the denominator r) and the common residue modulo T_S (from the denominators rs). Without the outer-prime phases the number of choices per row is not compressed to just below $L^{1/2}$. This is a prime-averaged result; no pointwise lower bound at this energy threshold is asserted. Nor is the stronger deletion-stability of the next sections claimed for $\mathcal{B}_X$.

7. Rectangular blocks and rational reconstruction

Fix X, put $L = \log X$, $U = \log L$, $V = \log U$, $Q_0 = CLU$ (unrounded), and use primes $p \in (X, 2X]$. All denominators used below are contained in I_p for large X. In particular, the small $O(1)$ difference between $\log p$ and L creates no endpoint issue. Define the prime sets

$$R_0 = \{r : 8L < r \leq 12L\}, \quad S = \{s : CU/14 < s \leq CU/13\}, \quad A = \{rs : r \in R_0, s \in S\}.$$

All prime sets in the paper are explicitly prime-indexed. Put $r_0 = |R_0|$, $s_0 = |S|$, $s_- = CU/14$, $s_+ = CU/13$, and $T_S = \prod_{s \in S} s$. Then

$$r_0 \sim 4L/U, \qquad s_0 \sim CU/(182V), \qquad \log T_S = (C/182 + o(1))U.$$

Products in A are distinct by unique factorization, and lie between $4Q_0/7$ and $12Q_0/13$. In particular $|A| \asymp_C L/V$. We use the energy in (3) and least absolute integer residues for p.

Lemma 7.1 (Uniform weak gap)**.** *For fixed sufficiently large C and every $p \in (X, 2X]$, $h \neq 0$,*

$$E_A(p, h) \gg_C L/(U^2V).$$

Proof. Call a row r quiet if at least $3s_0/4$ of its phases have distance at most $1/(4s_+)$. For such s, writing u_{rs} for the least absolute residue of $h(rs)^{-1}$, the integers su_{rs} lie in $[-p/4, p/4]$ and are all congruent to hr^{-1}. Thus they coincide with an integer v_r divisible by each quiet s. Write $rv_r = h + k_r p$, $|k_r| = O(L)$. Two quiet rows have at least $s_0/2$ common quiet columns, and their k_r agree modulo each such prime. Their product exceeds $24L + O(1)$ for C sufficiently large, so the two integers k_r agree. All quiet rows therefore divide the same nonzero integer $h + kp$ of size $O(Lp)$, which has at most $(1 + o(1))L/U$ prime divisors in R_0. The remaining $\gg L/U$ rows each contribute at least $s_0/4$ phases of size $\gg 1/U$, giving the claim. □

Lemma 7.2 (Rational reconstruction with lists)**.** *Let $B_0 \geq 1$ be an integer. Denote by $\mathcal{R}_{B_0}$ the set of reduced fractions d/r with $r \in R_0$, $0 < |d| \leq 2r$, $r \nmid d$, and $dr^{-1} \pmod s$ admitting a representative in $[-2B_0, 2B_0]$ for at least $s_0/2$ primes $s \in S$. Let*

$$j_0 = \left\lceil \frac{\log(1000L^2)}{\log s_-} \right\rceil + 1.$$

For sufficiently large fixed C we have $j_0 \le \lceil s_0/2 \rceil$, and

$$|\mathcal{R}_{B_0}| \le \frac{\binom{s_0}{j_0}}{\binom{\lceil s_0/2 \rceil}{j_0}} (4B_0+1)^{j_0}.$$

In particular, for either $\beta = 1/2$ or $\beta = 3/4$ there exists a fixed $b_0 > 0$ sufficiently small that $B_0 = \lfloor b_0 U^\beta \rfloor$ satisfies

$$|\mathcal{R}_{B_0}| \le \exp(2\beta U - c_1 U/V). \tag{37}$$

Proof. Choose j_0 qualifying primes and one representative for each. Two reduced fractions having the same record satisfy $s \mid dr' - d'r$ for every recorded prime. The determinant has absolute value at most $576L^2$, less than the product of these primes. Hence it vanishes and the fractions coincide. Each fraction yields at least $\binom{\lceil s_0/2 \rceil}{j_0}$ records. This proves the bound. Now $j_0 = 2U/\log s_- + O(1)$ and $s_0/j_0 \asymp_C 1$. The binomial ratio is at most $\exp(O_C(j_0))$. For $B_0 = \lfloor b_0 U^\beta \rfloor$, choose b_0 so small that the factor per recorded prime (including the binomial ratio) is at most $e^{-2} s_-^\beta$. Then its logarithm times j_0 is at most $2\beta U - c_1 U/V$, since $\log s_- = V + O_C(1)$ and $V = o(U/V)$. This proves (37) with fixed constants. □

Here is the precise comparison of two frequencies used repeatedly. For a pair (p,h) with $1 \le h < p$, put $\kappa(p,h) = -hp^{-1} \pmod{T_S}$; all the inverses in this expression are legitimate. Let v_r and u_{rs} be the least absolute residues of hr^{-1} and $h(rs)^{-1}$ modulo p. Write

$$rv_r = h + pk_r, \qquad su_{rs} = v_r + pj_{rs}. \tag{38}$$

Thus $|k_r| \le r/2 + 1$, $|j_{rs}| \le s\|h(rs)^{-1}/p\| + 1/2$, and

$$k_r + rj_{rs} \equiv \kappa(p,h) \pmod{s}, \qquad k_r \equiv -hp^{-1} \pmod{r}. \tag{39}$$

Call a row calm if at least $3s_0/4$ of the columns have $|j_{rs}| \le B_0$. For two pairs $(p,h), (p',h')$ having the same κ and a common calm row, (39) shows that $d = k_r - k_r'$ satisfies the list condition for at least half the columns. If $r \nmid d$, then $0 < |d| \le r + 2 \le 2r$, and $d/r \in \mathcal{R}_{B_0}$.

Proposition 7.3 (Pointwise small spectrum and robust prime-average gap)**.** *There exist fixed positive constants c_0, c_1, δ such that the following hold.*

(1) *For every $p \in (X, 2X]$, if $\tau_1 = c_0 L/(UV)$, then*

$$\#\{h \ne 0 : E_A(p,h) \le \tau_1\} \le T_S = \exp(O_C(U)).$$

(2) *Put $\tau_2 = c_1 L/(\sqrt{U}V)$. Apart from $O(Xe^{-cL/V})$ primes in $(X, 2X]$, every nonzero h and every $D \subseteq A$ with $|D| \le \delta|A|$ satisfy*

$$E_{A\setminus D}(p,h) > \tau_2.$$

Proof. In either case the number of entries with $|j_{rs}| > B_0$ is at most $|D| + 4s_+^2 E_{A\setminus D}/B_0^2$, since for such an entry the phase distance is at least $B_0/(2s_+)$. Choose the threshold constants and δ so that this bound is at most $|A|/32$. It follows that at least $7r_0/8$ rows are calm. For (1) take $D = \varnothing$, $B_0 = \lfloor b_0\sqrt{U} \rfloor$. By (37), only $o(r_0)$ rows admit any fraction of the exceptional list. Two frequencies with the same κ then agree modulo r for at least $(3/4 - o(1))r_0$ rows. Their prime product exceeds p, forcing $h = h'$. This gives the claimed injection into the T_S residue classes.

For (2) take $B_0 = \lfloor b_0 U^{3/4} \rfloor$. For fixed κ and row r, choose one calm pair as representative if one exists. Each other calm pair has either the same k_r residue or corresponds, by subtraction from this representative, to a member d/r of the list. Therefore there are at most $1 + D_r$ possible k_r residues for calm pairs, where D_r is the number of list fractions of denominator r. We have $\sum_r D_r = |\mathcal{R}_{B_0}|$. Choose $\ell = \lceil \log(4X^2)/\log(8L) \rceil + 1$ calm rows; this is possible since $\ell = (2 + O(1/U))L/U$. Equality of their residues for two pairs implies that each of those primes divides $hp' - h'p$. Their product exceeds $4X^2$, so the reduced fractions $h/p, h'/p'$ coincide. Thus for a specified row set the

number of pairs is at most the product of the corresponding $1 + D_r$. Summing over row sets and κ, and using AM–GM, bounds the total number of bad pairs by

$$T_S 2^{r_0} \left(1 + \frac{|\mathcal{R}_{B_0}|}{\ell}\right)^{\ell}.$$

By (37), the last logarithm per factor is at most $U/2 - c'U/V$, since $\log \ell = U - V + O(1)$ and $V = o(U/V)$. As $\log T_S + r_0 \log 2 = O_C(U + L/U) = o(L/V)$, the number of pairs is $O(Xe^{-cL/V})$. No union over deletion sets is needed: the property counted is existence of enough calm rows. The claim follows. □

8. Prime averages and deletion-stable Bernoulli mixing

Use the prime block

$$B = \{b \text{ prime} : 4Q_0/5 < b < 9Q_0/10\}, \quad n = |B| = (C/10 + o(1))L.$$

It is disjoint from A. Write $F = A \cup B$. Its total reciprocal mass is $O_C(1/U)$ and $|F| \asymp_C L$. For any set D of invertible denominators and independent Bernoulli variables ϵ_a of parameter $1/2$, define

$$\phi_{p,D}(h) = \prod_{a \in D} (1 + \mathrm{e}^{2\pi i h a^{-1}/p})/2.$$

The independent difference variables $C_a = \epsilon_a - \epsilon'_a$ take $-1, 0, 1$ with probabilities $1/4, 1/2, 1/4$, and their transform is $|\phi_{p,D}(h)|^2$. We use both $|\cos \pi x| \le e^{-2\|x\|^2}$ and $\cos^2 \pi x \le e^{-\sin^2 \pi x}$.

Lemma 8.1 (Poisson majorant on average)**.** *For C sufficiently large and*

$$G_{p,B}(h) = \exp\left(-\tfrac{1}{8} \sum_{b \in B} \sin^2(\pi h b^{-1}/p)\right),$$

we have $\sum_{X < p \le 2X} \sum_{h \bmod p} G_{p,B}(h) \ll_C XU$.

Proof. Take independent $K_b = N_b^+ - N_b^-$ with $N_b^{\pm} \sim \mathrm{Pois}(1/32)$. Then $G_{p,B}$ is the transform of $\sum K_b/b$ modulo p. Truncate at $\sum |K_b| \le n$; the probability of the complementary event is $\exp(-cn)$ by domination by a Poisson variable of mean $n/16$. On the truncated set the rational sum is zero only if all coefficients vanish: multiplication by $P = \prod_{b \in B} b$ and reduction modulo each b proves this since $n < \min B$. The probability of the all-zero vector is $w(0)^n$ for a fixed $w(0) < 1$. With C large both errors are $O(X^{-3})$. Each nonzero truncated numerator with denominator P has logarithmic size $O_C(LU)$ and hence at most $O_C(U)$ prime divisors in $(X, 2X]$. Orthogonality says the sum over h is p times the zero-residue probability. Average this identity over p, bounding each nonzero numerator's weighted prime count by $O(XU)$. The two exceptional probabilities contribute $O(X^2 X^{-3})$. This proves the result. □

Theorem 8.2 (Deletion-stable mixing)**.** *With $\tau_{\mathrm{mix}} = \tau_2$ from Proposition 7.3 (decreasing its constant if necessary), there is a set of good primes in $(X, 2X]$, with complement $O(Xe^{-c\tau_{\mathrm{mix}}})$, such that for every good prime and every $D \subseteq F$ satisfying $|D \cap A| \le \delta|A|$ and $|D \cap B| \le \tau_{\mathrm{mix}}$,*

$$\sum_{h \ne 0} |\phi_{p,F \setminus D}(h)| \le e^{-\tau_{\mathrm{mix}}}. \tag{40}$$

For either Bernoulli or difference variables the residual sum modulo p therefore has point probabilities $p^{-1}(1 + O(e^{-\tau_{\mathrm{mix}}}))$, uniformly in D and the residue.

Proof. Restrict to primes satisfying Proposition 7.3(2) and $\sum_{h \ne 0} G_{p,B}(h) \le e^{\tau_{\mathrm{mix}}/2}$. The omitted primes number $O(Xe^{-cL/V} + XUe^{-\tau_{\mathrm{mix}}/2}) = O(Xe^{-c'\tau_{\mathrm{mix}}})$ by Lemma 8.1 and Markov's inequality. For a remaining prime, the A part after deletion contributes at most $e^{-2\tau_{\mathrm{mix}}}$ to each nonzero

coefficient. On the B part use $|\cos \pi x| \le e^{-\sin^2(\pi x)/2} \le e^{-\sin^2(\pi x)/8}$. Deleting at most τ_{mix} factors can increase the Poisson majorant by at most $e^{\tau_{\mathrm{mix}}/8}$. Consequently the left side of (40) is at most $e^{-2\tau_{\mathrm{mix}}+\tau_{\mathrm{mix}}/8+\tau_{\mathrm{mix}}/2} \le e^{-\tau_{\mathrm{mix}}}$. Squaring coefficients only improves the bound. Fourier inversion gives the conclusion. □

Theorem 1.4 follows: take $D = \varnothing$, note $F \subset I_p$, and convolve with the independent choices on $I_p \setminus F$; the sum of nontrivial Fourier magnitudes can only decrease. Since $\mathcal{H}(I_p) = \log 2 + o(1) < 1$, the representations have cost at most one.

9. A pointwise sufficient condition for full coverage

Return to the notation of Sections 7–8. Let $\Sigma_p(D)$ denote the set of reciprocal subset sums of D modulo p, and $\rho_p(D)$ the collision probability of two independent Bernoulli subset sums. The notation $\rho_p(D)$ refers to the maximum point probability of the *difference* distribution (its central value), not the maximum of the original Bernoulli distribution.

Proposition 9.1 (Support-set criterion)**.** *With τ_1 as in Proposition 7.3, suppose for a given $p \in (X, 2X]$ there is a set $D \subseteq I_p \setminus A$ such that*

$$|\Sigma_p(D)| \ge p e^{-4\tau_1+2}.$$

Then $g(p) \le 3q_p$. In particular, a sufficient condition is $\rho_p(D) \le p^{-1}e^{4\tau_1-2}$.

Proof. Let ν be uniform on $\Sigma_p(D)$ and μ the independent Bernoulli sum on A. By Parseval, $\sum_h |\widehat{\nu}(h)|^2 = p/|\Sigma_p(D)|$. Split the nonzero frequencies at $E_A(p,h) = \tau_1$. On the lower range, Lemma 7.1 and Proposition 7.3(1) bound the sum of squared convolution coefficients by $T_S e^{-cL/(U^2V)} = o(1)$. On the upper range it is at most $e^{-4\tau_1}p/|\Sigma_p(D)| \le e^{-2}$. Hence the squared coefficients including zero sum to less than 2. Cauchy–Schwarz shows that $S = \Sigma_p(A \cup D)$ has size greater than $p/2$. For every x, the sets S and $x - 3^{-1}S$ intersect, giving $S + 3^{-1}S = \mathbb{F}_p$. The blocks $A \cup D$ and $3(A \cup D)$ lie in disjoint intervals, and the total real reciprocal mass is at most $(4/3)H(I_p) = (4/3)\log 2 + o(1) < 1$, where $H(I_p) = \sum_{a \in I_p} 1/a$. All denominators are below p for large p. The collision-probability sufficient condition follows from $|\Sigma_p(D)| \ge 1/\rho_p(D)$. □

For the prime block B of Section 8, a concrete sufficient condition for uniform $O(\log p \log_2 p)$ coverage is

$$\max_{X < p \le 2X} \log\bigl(p\rho_p(B)\bigr) = o(L/(UV)). \tag{41}$$

Here L, U, V have the definitions of Section 7. The condition implies the collision-probability hypothesis of Proposition 9.1 for all sufficiently large primes. It is not proved in this paper. Ordinary prime averaging only controls the frequency of exceptional primes, not the worst isolated prime. For example, the same divisor argument as in Lemma 8.1, now with three-valued difference coefficients, gives $\sum_{X<p\le 2X} p\rho_p(B) \ll XU$: the rational zero event has probability $2^{-|B|}$ and every nonzero numerator has at most $O(U)$ prime factors in the dyadic interval. This does not imply (41).

10. Exact rational grids and height

10.1. **Dyadic completion.**

Lemma 10.1 (Smooth-denominator dyadic completion)**.** *Let M be sufficiently large and T a positive integer such that each prime power dividing T is at most M. Every $s/T \in [1/4, 1]$ is a sum of distinct reciprocals of positive integers at most $4M^3$.*

Proof. Write $T = 2^\nu T_o$, where $T_o = \prod_i d_i$ is the product of the maximal odd prime powers in T. These d_i are pairwise coprime and at most M. Take a power of two $B_* = 2^h$ with $M \le B_* < 2M$, and put $W = T_o B_*^2$, a multiple of T. For each d_i use the block of denominators $d_i B_* 2^j$, $1 \le j \le h$. After multiplication by W, their weights are $(T_o/d_i)2^{h-j}$. The binary sums of these powers of two give every integer from 0 to $B_* - 1$, hence every residue modulo d_i. Select one subset for each block so that the selected sum multiplied by W agrees with Ws/T modulo d_i; the other blocks vanish modulo d_i. By CRT the residual has denominator dividing B_*^2. The selected sum R_0 satisfies

$$0 \le R_0 \le B_*^{-1} \sum_i d_i^{-1} \le H_{\lfloor M \rfloor}/B_* = o(1).$$

The residual $s/T - R_0$ is in $[0,1]$, and has a finite binary expansion using denominators $2, 4, \dots, B_*^2$ if it is less than one (use the empty set for zero). If it equals one, use denominator 1. These denominators do not overlap any of the odd-prime-power blocks; those blocks also are mutually disjoint by their odd part. All denominators are at most $MB_*^2 < 4M^3$. □

Lemma 10.2 (Grid lifting)**.** *Suppose $\mathcal{R}$ is a set of primes in $(Y, 2Y]$, $P_0 = \prod_{r\in\mathcal{R}} r$, and $g(r) \le M$ for all $r \in \mathcal{R}$, where $M \to \infty$ and $M^3 = o(Y)$. For all sufficiently large Y,*

$$\{s/P_0 : s \in \mathbb{Z},\ 1/3 \le s/P_0 \le 1\} \subseteq \mathcal{V}(\lceil 2YM \rceil).$$

The represented fractions are exact, not merely congruent modulo a prime.

Proof. Let $T = \operatorname{lcm}(1, \dots, M)$, taking M integral. Fix $w = s/P_0$ in the displayed interval. For each r, use $g(r) \le M$ to choose $J_r \subseteq [1, M]$ with $\mathcal{H}(J_r) \le 1$ and

$$\mathcal{H}(J_r) \equiv s(P_0/r)^{-1} \pmod r.$$

The denominators rj, $j \in J_r$, belong to disjoint blocks, since $M < Y$. Let R be the sum of their reciprocals. On multiplying by $P_0 T$ and reducing modulo r, all other blocks vanish and $P_0 TR \equiv sT \pmod r$. Thus $w - R$ has denominator dividing T. Also $R \le \sum_{Y<r\le 2Y} r^{-1} = o(1)$, so $w - R \in [1/4, 1]$. Lemma 10.1 completes it with distinct denominators at most $4M^3 < Y$. These cannot overlap the initial blocks, which use denominators exceeding Y. The combined maximum is at most $2YM$. □

Take $\mathcal{R}$ to be the primes of $(Y, 2Y]$ that are good in Theorem 1.3. Its exceptional set has size $o(Y/\log Y)$. Hence

$$\begin{gathered} \log P_0 = (1 + o(1))Y, \qquad M = \lceil C' \log Y \log_2 Y / \mathcal{C}_4(Y) \rceil, \\ Z(Y) := \lceil 2YM \rceil \ll Y \log Y \log_2 Y / \mathcal{C}_4(Y). \end{gathered} \tag{42}$$

Here $M^3 = o(Y)$, as required. The lifted grid is independent of the target numerator. Choosing $Y = 2\log x$ gives $P_0 > 6x$ for all sufficiently large x. For $x = p$ prime, P_0 is prime to p. Given any residue a modulo p, choose $s \equiv aP_0 \pmod p$ with $P_0/3 \le s \le 2P_0/3$ (the interval has length greater than p). Grid lifting proves $g(p) \le Z(Y) \ll \log p \log_2 p \log_3 p / \mathcal{C}_5(p)$ by (34).

10.2. **Height for arbitrary denominators.** For an arbitrary integer b, take $Y = 2\log b$ and $P_0, Z = Z(Y)$ as in (42). Choose an integer x such that $s = aP_0 - bx \in [P_0/3, 2P_0/3]$; this is possible since $P_0 > 6b$. The grid represents s/P_0 by a set F with maximum at most Z. Put $z = x/P_0$, so

$$\frac{a}{b} = \frac{1}{b}\mathcal{H}(F) + z, \qquad \frac{a}{3b} \le z < \frac{a}{b} < 1.$$

Choose an integer y with $yz \in [1/3, 1]$ as follows. If $a > 3Z$, use $y = \lfloor 1/z \rfloor$, for which $yz \ge 1/2$ and $yZ \le 3bZ/a < b$. Otherwise choose a prime $y \in (1/(2z), 1/z]$ with $y > Z$ and $y \nmid b$. This

is possible by the PNT, uniformly since $1/z \geq b/(3Z)$, while the number of primes dividing b is $O(\log b)$. In either case represent $yz = yx/P_0$ by a set F' of denominators at most Z. Then

$$\frac{a}{b} = \sum_{u \in F} \frac{1}{bu} + \sum_{v \in F'} \frac{1}{yv}.$$

The two blocks are disjoint in the first case by the size bound. In the second case an equality $bu = yv$ would force $y \mid u$, impossible since $u \leq Z < y$. Each denominator is at most $3bZ$, since $y \leq 1/z \leq 3b/a$. This proves the height upper bound in Theorem 1.5. No coprimality of P_0 and b was needed: these are exact rational equalities.

10.3. **The counting lower bound and a progression obstruction.** The inequality $f(p) \geq b(p)$ and Theorem 1.1 give the stated all-primes counting lower bound. For a prime p, the upper bound for $D(p)$ proved above ensures $D(p) < p^2$ for sufficiently large p. In an exact representation of a/p with this height bound, multiply by p and reduce modulo p. Terms whose denominators are not divisible by p vanish, and terms with denominator pj become $1/j$ with distinct $j \leq D(p)/p < p$. Every nonzero class a is thus represented, with the empty set covering zero. It follows that $D(p) \geq pf(p)$. Combining with the upper bound proves Theorem 1.5 in full.

For the progression lower bound take $Y = cq\mathcal{C}_4(q)/(\log q \log_2 q)$ with $c > 0$ a sufficiently small fixed constant. Since $q^{1/2} \leq Y \leq q$ for all large q, (34) gives $\mathcal{C}_4(Y) \asymp \mathcal{C}_4(q)$ and $\log Y \asymp \log q$, $\log_2 Y \asymp \log_2 q$. The implied constants are uniform in this range, so (42) ensures $Z(Y) \leq q$ after choosing c small enough. Grid lifting gives a progression of length $\gg P_0$ contained in $\mathcal{V}(q) \cap [1/3, 1]$, and $\log P_0 \asymp Y$. Hence $\log \mathcal{A}(q) \gg q\mathcal{C}_4(q)/(\log q \log_2 q)$, matching the upper bound below. Conversely, let $w + ja/b \in \mathcal{V}(q)$ for $0 \leq j < K$, with a, b positive coprime integers. For $K \leq q$ there is nothing to prove. Otherwise $b \mid T_q := \operatorname{lcm}(1, \ldots, q)$, and $K \leq 1 + bH_q$. For each prime $r > \sqrt{q}$ dividing b, the common difference times T_q is nonzero modulo r; thus the progression times T_q covers all residues modulo r. Only denominators divisible by r contribute, and after dividing by T_q/r their weights are inverses of integers in $[1, \lfloor q/r \rfloor]$. Consequently $f(r) \leq q/r$. The lower bound just proved implies

$$r \ll q\mathcal{C}_4(q)/(\log q \log_2 q).$$

All prime factors of b are therefore below this threshold (including the primes at most $\sqrt{q}$); the additional prime powers contribute at most $O(\sqrt{q} \log q)$ to $\log b$. The Chebyshev bound and $K \leq 1 + bH_q$ give the asserted upper bound for $\log \mathcal{A}(q)$. In particular, constructing a single exact rational progression of length at least p requires

$$q \gg \frac{\log p \log_2 p \log_3 p}{\mathcal{C}_5(p)}.$$

Indeed, if q already exceeds $(\log p)^2$ this is immediate, and otherwise $\log q \asymp \log_2 p$ and $\mathcal{C}_4(q) \ll \mathcal{C}_5(p)$. This rules out a single-grid proof at scale $q \asymp \log p \log_2 p$, not arbitrary modular coverage at that scale.

11. The least forbidden initial denominator

We first establish a single exceptional-prime estimate for this problem. For $L = \log N$, $U = \log L$, choose

$$Q_- = \lfloor cLU/\mathcal{C}_4(N) \rfloor$$

with $c > 0$ sufficiently small. Estimate (18) and the comparison of iterates show $|\mathcal{V}(Q_-)| \leq N^{1/4}$ for large N. Each positive $v \in \mathcal{V}(Q_-)$ can be written as A_v/T_- with $T_- = \operatorname{lcm}(1, \ldots, Q_-)$ and $0 < A_v \leq T_- H_{Q_-}$, so $\log A_v = O(Q_-)$. Each such numerator has $O(Q_-/L)$ distinct prime factors exceeding N/Q_-. Thus at most

$$O(N^{1/4} Q_-/L) \tag{43}$$

primes $p > N/Q_-$ admit a nonempty reciprocal subsum from $[1, Q_-]$ which is zero modulo p. All denominators are less than p here.

11.1. **The least forbidden initial denominator.** For the upper bound, the PNT ensures a prime $p \in (N/Q_-, 2N/Q_-]$ outside the exceptional set (43). If a representation of one with first denominator p and all denominators at most N existed, multiplying it by p and reducing modulo p would give

$$1 + \sum_{p|n_i,\ i\geq 2} (n_i/p)^{-1} \equiv 0 \pmod{p}.$$

Indeed $p^2 > N$ and the other terms vanish. This is an impossible nonempty zero subsum from $[1, Q_-]$. Thus $t(N) \leq p \ll N\mathcal{C}_4(N)/(LU)$.

For the lower bound use the following published input, with the numbering and statement from [4, Lemma 4.1, arXiv v1]: for sufficiently large N, if every prime power dividing T is at most $c_0 N(\log N)^{-3}(\log_2 N)^{-4}$, then each fraction $s/T \in [1/3, 1]$ is a sum of distinct reciprocals of integers in $[N/16, N]$. We use this statement as an external theorem, not a reproof of its Fourier-analytic argument.

Choose the grid with $Y = 2\log N$, $P_0 > 6N$, $Z \ll LUV/\mathcal{C}_5(N)$. Every $1 \leq t \leq N/(100Z)$ is admissible. For $t = 1$ this is immediate. If $2 \leq t \leq N^{9/10}$, the external lemma directly represents $(t-1)/t$ on $[N/16, N]$, which is above t. For $N^{9/10} < t \leq N/(100Z)$ choose $s/P_0 \in [1/3, 2/3]$ with $s + P_0 = mt$ for some integer m. Represent s/P_0 by grid denominators F with maximum at most Z. Since $s/P_0 < 1$, $1 \notin F$. Then

$$\frac{1}{t} + \sum_{a\in F} \frac{1}{ta} = \frac{m}{P_0}.$$

The initial block consists of distinct denominators at most $N/100$, beginning with t. The residual $1 - m/P_0 = 1 - (1 + s/P_0)/t$ lies in $[1/3, 1]$. Its denominator divides the squarefree P_0 with primes at most $4\log N$. The external lemma supplies a disjoint completion on $[N/16, N]$. This proves $t(N) \gg N\mathcal{C}_5(N)/(LUV)$, giving the stated sharpened exponent for the problem recorded as Erdős problem 294 (the original formulation and earlier bound are in [5, 4]).

12. Uniform recursive coverage and deletion scaling

The exact CRT grid also proves a recursive bound for the worst prime. This will provide an unconditional fallback for exceptional primes in the square-loss application.

Lemma 12.1 (All-prime cover and recursion)**.** *There is an absolute $C > 0$ such that*

$$G(x) \ll \log x(\log_2 x)^3(\log_3 x)^4, \qquad G(x) \leq C\log x\, G(C\log x). \tag{44}$$

In particular, for each fixed nonnegative integer k,

$$G(x) \ll_k \left(\prod_{i=1}^{k+1} \log_i x\right)(\log_{k+2} x)^3(\log_{k+3} x)^4.$$

The first bound is weaker than Theorem 1.5, but the recursion is useful in its own right.

Proof. The first bound follows from Theorem 1.5; the finitely many small primes are harmless. For the recursion take $Y = 2\log x$, let $M = \max(2, G(2Y))$, and let $\mathcal{R}$ consist of all primes in $(Y, 2Y]$. The uniform upper bound already proved gives $M = O((\log Y)^2)$, hence $M^3 = o(Y)$; the counting lower bound and the PNT give $M \to \infty$. The product P_0 of these primes satisfies $\log P_0 \sim Y$, so $P_0 > 6x$. Lemma 10.2 represents every grid fraction in $[1/3, 2/3]$ at height at most $\lceil 2YM\rceil$. For $\lceil 2YM\rceil < p \leq x$, choose a grid numerator in the desired class after division by P_0; all the grid denominators are below p. For $p \leq \lceil 2YM\rceil$, one unit reciprocal (or the empty set for zero) suffices. Thus $G(x) \leq \lceil 2YM\rceil \leq C\log x\, G(C\log x)$ for sufficiently large x after enlarging C. Iterate a fixed

number of times and apply the first bound at the final argument, comparing iterated logarithms at constant multiples of their arguments. Constants may depend on k. □

Lemma 12.2 (Deletion scaling and exceptional primes)**.** *If $Z \le q < p$ and all residues modulo p are covered in $[1, Z]$ at mass at most one, then*

$$d(p, q) \le 2Z/q. \tag{45}$$

For $p > N^{1/3}$ and $q = \lfloor N/p^2 \rfloor < p$, one always has

$$\frac{d(p, q)}{p^2} \ll \frac{Z_0 \log(2Z_0)}{N}, \qquad Z_0 = \lceil (\log N)^2 \rceil, \tag{46}$$

when $q \ge 1$. The same statement holds with p^2 replaced by p and $q = \lfloor N/p \rfloor < p$.

Proof. Set $v = \lfloor q/Z \rfloor$, represent $vH_q \bmod p$ in $[1, Z]$, and multiply its denominators by v. This is valid since $v < p$ and gives the required deletion at cost at most $1/v$. Lemma 12.1 ensures a uniform cover by Z_0 for the specified primes, indeed for all primes up to N. If $q < Z_0$, use $d(p, q) \le H_q \ll \log(2Z_0)$ and $p^{-2} < (q + 1)/N$; otherwise use (45) and $p^2 q > N/2$. The argument for p is identical. □

13. The large-prime deletion expansion

We record the common large-prime deletion consequence of the coverage theorems. In applications to integer Egyptian fraction capacity, $H_N - \mathcal{D}(N)$ is the capacity obtained by separately optimizing the large-prime blocks. Finer capacity and higher-prime-power obstructions are not studied here.

Proof of the deletion formula in Theorem 1.6. Fix $a \in (0, 1)$ and put $m = \lfloor ab(N) \rfloor$. The number of nonzero harmonic subset sums at height m is $N^{a+o(1)}$ by regular variation. Each such sum has an integer numerator after multiplication by $\mathcal{L}(m)$ with logarithmic height $O(m)$. Hence at most $N^{a+o(1)}$ primes above N/m divide any of these nonzero numerators. Outside this exceptional set, $d(p, \lfloor N/p \rfloor) = H_{\lfloor N/p \rfloor}$ for $p > N/m$. The total harmonic loss omitted by exceptional primes in this range is at most $N^{a+o(1)} m H_m/N = o(1/\log N)$.

For an upper bound put $Z = \lceil Cb(N) \rceil$ for a sufficiently large fixed C. On dyadic intervals with lower endpoints between a fixed multiple of $\sqrt{N}$ and N, Theorem 1.2 gives ordinary full coverage with height at most Z and cost at most one outside $O(Pe^{-c(\log_2 P)^2})$ primes. When $\sqrt{N} < p \le N/Z$ is nonexceptional, Lemma 12.2 gives $d(p, \lfloor N/p \rfloor)/p \ll Z/N$. Summing gives $O(1/\log N)$. The exceptional primes, counted over these intervals, contribute $o(1/\log N)$: in an interval $(P, 2P]$ use $d \le H_{\lfloor N/p \rfloor} \ll \log N$, the weight $1/p \le 1/P$, and the exceptional-set estimate. There are $O(\log N)$ intervals, and $\log_2 P = \log_2 N + O(1)$ uniformly. Their aggregate contribution is $O((\log N)^2 e^{-c'(\log_2 N)^2}) = o(1/\log N)$. For $p > N/Z$ use $d \le H_{\lfloor N/p \rfloor}$.

For $z = m$ or $z = Z$, the PNT and partial summation give

$$\sum_{N/z < p \le N} \frac{H_{\lfloor N/p \rfloor}}{p} = \frac{(\log z)^2}{2 \log N} + O\left(\frac{\log z}{\log N}\right).$$

Indeed $H_{\lfloor N/p \rfloor} = \log(N/p) + O(1)$ uniformly; the classical PNT error is negligible compared with the displayed error, since $\log z = O(\log_2 N)$. The leading integral is $\int_0^{\log z} u/(\log N - u)\, du$. Both cutoffs have $\log z = \log b(N) + O(1)$, proving

$$\mathcal{D}(N) = \frac{(\log b(N))^2}{2 \log N} + O\left(\frac{\log b(N)}{\log N}\right). \tag{47}$$

Putting $L = \log N$, $U = \log_2 N$, $V = \log_3 N$, the inverse estimate gives $\log b(N) = U + V - \log \mathcal{C}_4(N) + O(1)$. Substitution proves the stated truncated iterated-logarithm expansion. Since

$(V-\log \mathcal{C}_4(N))^2 = O(V^2) = o(U)$, the equivalent centered expansion follows. In particular there is no $U \log_4 N/L$ term, whereas each fixed subsequent iterated logarithm starting with $U \log_5 N/L$ has coefficient -1 after subtracting the preceding terms. The cutoff terms within $O(U/L)$, including a constant at that scale, are not determined. □

14. Prime-square loss and the directed threshold

Keep the notation $B = b(\sqrt{N})$, $A_N = \sqrt{B}\log B/(\sqrt{N}\log N)$ from (8), and put $X = \sqrt{N/B}$. For $p \asymp X$ in any fixed multiplicative range, (15) gives $b(p)/B \to 1$ uniformly. The same range has $\lfloor N/p^2 \rfloor \asymp B$. The prime number theorem and partial summation give, for fixed $0 < a < b$,

$$\sum_{aX<p\le bX} p^{-2} \sim \frac{a^{-1}-b^{-1}}{X\log X}, \qquad \frac{\log B}{X\log X} \sim 2A_N. \tag{48}$$

Proposition 14.1 (The two determined regions)**.** *With the fixed τ of Theorem 1.2, one has*

$$\sum_{X<p\le\sqrt{N}} \frac{d(p,\lfloor N/p^2\rfloor)}{p^2} = (2+o(1))A_N, \tag{49}$$

$$\sum_{N^{1/3}<p\le X/\sqrt{\tau}} \frac{d(p,\lfloor N/p^2\rfloor)}{p^2} = o(A_N). \tag{50}$$

Proof. For the first assertion, fix $\eta > 0$ and $T > 1+\eta$. On $(1+\eta)X < p \le TX$ the local quotient q is at most $(1-\delta)b(p)$ for some fixed $\delta > 0$. The below-threshold result applied on finitely many dyadic intervals makes $d(p,q) = H_q = (1+o(1))\log B$ outside $o(X/\log X)$ primes. The mass on this interval has normalized lower limit at least $2(1/(1+\eta) - 1/T)$ by (48). On all $p > X$ use $d \le H_q \le \log B + O(1)$ and the inverse-square prime tail to bound the upper limit by two. Let $T \to \infty$ and then $\eta \downarrow 0$.

For the second assertion, fix $\eta > 0$ and first restrict to $p \le X/\sqrt{\tau+\eta}$. Cover $(N^{1/3}, \sqrt{N}]$ by dyadic intervals with lower endpoints $P \le \sqrt{N}$ and $\log_2 P = \log_2 N + O(1)$. Use the full cover at the fixed parameter $\tau + \eta/2$. For the nonexceptional primes of these intervals its denominator bound $Z = \lfloor(\tau+\eta/2)b(P)\rfloor$ is at most $(\tau+\eta/2)B \le q$ for large N in the restricted range. By (45), $d(p,q)/p^2 \ll B/N$. Summing over $p \le X$ gives $O(\sqrt{B}/(\sqrt{N}\log N)) = o(A_N)$.

The number of exceptional primes over the entire range is $O(\sqrt{N}e^{-c(\log_2 N)^2})$, by geometric summation of the dyadic lower endpoints. The uniform bound (46), not the trivial estimate with the smallest p^{-2}, shows that their aggregate loss is

$$O\left(\frac{Z_0\log(2Z_0)}{\sqrt{N}}e^{-c(\log_2 N)^2}\right) = o(A_N). \tag{51}$$

This is valid with constants depending on the fixed margin η. In the residual band $X/\sqrt{\tau+\eta} < p \le X/\sqrt{\tau}$, the bound $d \le \log B + O(1)$ and (48) give normalized upper limit $O(\eta)$. Letting $\eta \downarrow 0$ proves (50). □

Thus the exact remaining contribution is

$$\frac{\Delta_2(N)}{A_N} = 2 + \frac{1}{A_N}\sum_{X/\sqrt{\tau}<p\le X} \frac{d(p,\lfloor N/p^2\rfloor)}{p^2} + o(1). \tag{52}$$

Using $d \le \log B + O(1)$ in the transition band proves (9). The same proof using the simpler two-pool result gives the corresponding formulas with $\tau = 2$, as in the earlier threshold analysis. By

(6) and bounded-factor stability of the iterated product, $W(\sqrt{N}) \asymp W(N)$. Also $\log B \sim \log_2 N$, so (10) follows. In particular

$$\Delta_2(N) = \frac{(\log_2 N)^{3/2+o(1)}}{\sqrt{N \log N}},$$
$$\frac{\sqrt{N \log N}\,\Delta_2(N)}{\log_2 N} \longrightarrow \infty, \qquad \frac{\sqrt{N \log N}\,\Delta_2(N)}{(\log_2 N)^{3/2}} \longrightarrow 0.$$

Indeed $W(N) \to \infty$ and $\log W(N) = O(\log_5 N) = o(\log_3 N)$. This refutes a constant on the smaller $\log_2 N/\sqrt{N \log N}$ scale, while also showing that the uncorrected power $3/2$ needs its iterated factor. The weaker $3/2 + o(1)$ conclusion uses only (14) and the coverage bounds, not the external variable-depth estimate.

14.1. **An integral formula and the exact equivalence.** Define $D_p(t) = d(p, \lfloor tb(p) \rfloor)/\log b(p)$ for fixed $t > 0$ and sufficiently large p. We will use the comparison

$$d(p, q_2) \le d(p, q_1) + H_{q_2} - H_{q_1} \quad (0 \le q_1 \le q_2 < p). \tag{53}$$

A feasible retained set at q_1 can also be retained at q_2 by deleting all the additional denominators. In particular $D_p(t_2) \le D_p(t_1) + o(1)$ for fixed $0 < t_1 < t_2$. This is asymptotic monotonicity of the normalized loss, not monotonicity of d itself.

Proposition 14.2 (Conditional limiting profile)**.** *Suppose that there is a deterministic nonincreasing function $D(t)$ such that $D_p(t)$ converges to $D(t)$ in relative prime density for every continuity point $t > 0$. Then*

$$\frac{\Delta_2(N)}{A_N} \longrightarrow \int_0^\infty D(t) t^{-1/2}\, dt. \tag{54}$$

At continuity points, $D(t) = 1$ for $0 < t < 1$ and $D(t) = 0$ for $t > \tau$.

Proof. On a fixed compact range $a \le u = p/X \le b$ with $a > 0$, the local parameter $q/b(p)$ is $u^{-2} + o(1)$ uniformly. Use a finite partition whose endpoints correspond to continuity points of D and squeeze the normalized losses by (53). On each piece the exceptional primes are negligible by the assumed convergence and a finite dyadic covering. The prime weights in (48) make the normal part converge, as the partition is refined, to $2\int_a^b D(u^{-2})u^{-2}\, du$. The bounded monotone function is Riemann integrable away from zero and infinity. The change of variables $t = u^{-2}$ gives the truncated integral of (54). For $u > b$, the normalized tail is $O(1/b)$ by the proof of (49). For all sufficiently small a, the $u < a$ tail is $o(1)$ by (50), including its exceptional-prime bound. Expanding the compact range proves the formula. The claimed endpoint behavior follows from the two coverage theorems. □

Proof of the equivalence in Theorem 1.7. Assume (11). In (52) remove a band of width ηX next to X; its normalized loss is $O(\eta)$. On the remaining band $q/b(p) \ge 1 + c_\eta$ for some $c_\eta > 0$. Use (11) at $1 + c_\eta/2$ and then (53). For every fixed positive tolerance the normalized loss is within that tolerance outside a negligible prime set, whose contribution is negligible since $p \asymp X$. Letting the tolerance and then η tend to zero eliminates the transition sum.

Conversely suppose the directed condition fails. By (53) and the full cover above τ, we can take fixed $1 < t_0 < \tau$, $\delta > 0$, and a sequence $P \to \infty$ with $\gg P/\log P$ primes satisfying $D_p(t_0) > \delta$. Choose $c \in (1, t_0)$ and a small fixed $\xi > 0$ with $c/(1+\xi)^2 > 1$. Partition $[P, 2P]$ into finitely many intervals of multiplicative width at most $1 + \xi$. Some such interval $[U, (1+\xi)U]$ contains $\gg P/\log P$ bad primes. Set $N = \lfloor cb(P)U^2 \rfloor$. Then $\log\sqrt{N} \sim \log P$ and $B \sim b(P)$ by (15). On that interval,

$$\frac{N}{p^2 b(p)} = c(U/p)^2 + o(1)$$

is bounded away from both 1 and t_0. The primes in question are strictly in the transition band, since also $1 < N/(p^2B) < \tau$ with fixed margins. They have $q \le \lfloor t_0 b(p) \rfloor$, so (53) gives $d(p,q) \ge \delta \log b(p) - O(1) \ge (\delta/2)\log B$. Since $p \asymp P \asymp X$, their normalized contribution is bounded below by a fixed positive number. This contradicts $\Delta_2(N)/A_N \to 2$. If failure is initially at $t_0 \ge \tau$, choose a smaller fixed parameter in $(1,\tau)$ and apply the same comparison. This establishes necessity. Notice that the quantifier concerns all large N, not only a sparse sequence of N. □

Since $0 \le D_p(t) \le 1 + o(1)$ for fixed t, Markov's inequality also makes (11) equivalent to $(\log P/P)\sum_{P\le p\le 2P} D_p(t) \to 0$ for each fixed $t \in (1,\tau)$. Requiring ordinary full coverage for all $t > 1$ is stronger than this equivalence.

15. The unresolved fixed-target problem

One can reformulate the sharp constant criterion in two ways. First, the core mass $C_q = \mathcal{H}(K_q)$ has numerator and denominator of logarithmic size $O(q/\log q)$, hence of size $P^{o(1)}$ when $q \asymp b(P)$. Keeping the entire core and a suitable subset of $\mathcal{G}_q$ makes the remaining deletion mass $o(1)$ if one can represent $-C_q \bmod p$ using $\mathcal{G}_q$. This is a sufficient directed hit, not a consequence of near-complete coverage.

Second, for $0 < \delta < 1$ define the finite set

$$\mathcal{U}_\delta(q) = \{\mathcal{L}(q)(H_q - x) : x = \mathcal{H}(D),\ D \subseteq [1,q],\ x \le \delta \log q\}.$$

For large q all its members are positive integers, and for $p > q$,

$$d(p,q) \le \delta \log q \quad\Longleftrightarrow\quad p \mid \prod_{u\in\mathcal{U}_\delta(q)} u. \tag{55}$$

The set has logarithmic cardinality $(1+o(1))s(q)$ by (16), while each member has logarithmic height $O(q)$. At $q = \lfloor tb(P) \rfloor$ there are $P^{t+o(1)}$ distinct candidates. The directed statement (11) is equivalent to covering a density-one subset of $[P,2P]$ by their prime divisors for every fixed $t > 1$, $\delta \in (0,1)$. Indeed the change between $b(P)$ and $b(p)$ is uniform by (15); use a slightly smaller fixed t and (53) to transfer a low-loss condition to the larger height, then adjust the tolerance. This also proves the converse transfer. By the prime number theorem the density-one covering condition can equivalently be expressed as a sum of $\log p$ over the covered primes asymptotic to P.

Neither entropy nor modular convergence in mean square alone forces a prescribed hit. As an elementary model the uniform distribution on $\mathbb{F}_p \setminus \{h_0\}$ has normalized squared error $1/(p-1)$ from uniform, but is zero at h_0. This is a counterexample to a general inference, not a counterexample to (11) for harmonic sums. A bound on the total logarithmic mass of large prime factors of the candidate numerators would likewise not locate them in the required dyadic interval. The fixed-target question and stronger analytic estimates of signed correlations are reserved for a separate investigation, not used in any theorem above.

15.1. **The uniform binary-coverage problem.** The all-primes bound of Theorem 1.5 still contains the factor $\log_3 p/\mathcal{C}_5(p)$ beyond $\log p \log_2 p$. The matching arithmetic-progression obstruction rules out an argument based on a single exact rational progression of length at least p at the latter scale. It does not rule out a direct modular argument. The sufficient pointwise target (41) remains open; relative support expansion or a reformulation as a high-order average cannot by itself replace an absolute near-full support bound. In particular, the density-one mixing theorems cannot be applied to every isolated exceptional prime. The two open problems (11) and (41) have different quantifiers and neither is assumed in the results above.

Appendix A. A general entropy statement for rational projections

The elementary collision argument has an application outside harmonic sums. We state the assumptions explicitly because logarithmic entropy saturation alone should not be confused with equidistribution.

Proposition A.1. *Let U_P be a set of $P^{t+o(1)}$ distinct rationals for a fixed $t > 0$, with a common positive integer denominator Q_P satisfying $\log Q_P \le P^{o(1)}$. Suppose $\log |Q_P(u-v)| \le P^{o(1)}$ uniformly for distinct $u, v \in U_P$. On all but a relative-density-zero set of primes $p \in [P, 2P]$, reduction is defined and the collision entropy and Shannon entropy of the uniform rational law after reduction are*

$$H_2(\rho_p),\ H_{\mathrm{Sh}}(\rho_p) = (\min(t,1) + o(1)) \log P.$$

If $t < 1$, then also $|U_P \bmod p| = (1 - o(1))|U_P|$ in prime density.

Proof. At most $P^{o(1)}$ primes in the dyadic interval divide Q_P. For each off-diagonal pair, at most $P^{o(1)}$ primes can divide its nonzero integer difference. Averaging the collision probability over primes where reduction is defined, we obtain at most $|U_P|^{-1} + P^{-1+o(1)}$. By Markov, for every fixed $\epsilon > 0$ the collision probability is at most $P^{-\min(t,1)+\epsilon}$ except for a proportion tending to zero. By taking a diagonal sequence of fixed positive ϵ tending to zero, the upper bound holds with exponent $-\min(t,1) + o(1)$ outside a relative-density-zero exceptional set. The universal lower bound for the collision probability is $1/\min(p, |U_P|)$. Thus $H_2 = -\log \sum_h \rho_p(h)^2$ has the claimed exponent, and $H_2 \le H_{\mathrm{Sh}} \le \log \min(p, |U_P|)$ gives the Shannon statement. When $t < 1$ the expected number of colliding unordered pairs is at most $|U_P|^2 P^{-1+o(1)} = o(|U_P|)$. The loss in the number of distinct residues is bounded by that pair count. A final Markov estimate proves near-injectivity. □

Appendix B. Related variants of the construction

These variants are separate from the main strictly distinct, all-primes coverage question. We retain them for clarity rather than treating them as proofs of (41).

B.1. **Composite moduli.** For $M_0 \ge 2$ define $g_\times(M_0)$ analogously to $g(p)$, allowing only denominators prime to M_0 and $q < M_0$, with value $+\infty$ if full coverage fails. For sufficiently large M_0,

$$g_\times(M_0) \ll (\log M_0)^3 \log_2 M_0.$$

Indeed, write $L_0 = \log M_0$, choose $Y = C_0 L_0^3 \log L_0$ and let B_0 be the primes in $(Y, 2Y]$ not dividing M_0. The excluded primes number at most $L_0/\log Y$, so $|B_0| \sim Y/\log Y$. For each nonzero $h \pmod{M_0}$, if the phase distance of hb^{-1}/M_0 is at most c_0/L_0, then b divides a nonzero integer $h + jM_0$ with $|j| \le 2c_0 Y/L_0 + 1$. Each has at most $O(L_0/\log Y)$ prime divisors in the interval. Choosing c_0 small, fewer than half the primes in B_0 have such small phases. Hence the sum of squared phase distances is $\gg Y/(L_0^2 \log Y)$. For C_0 sufficiently large the Bernoulli transform at every nonzero frequency is bounded by M_0^{-2}. Fourier inversion on the cyclic group gives full coverage; the total real reciprocal mass is $O(1/\log Y) < 1$. This argument does not assume the group is a field.

For odd M_0, a binary construction gives $g_\times(M_0) \ll (\log M_0)^2$. Choose $Y = 3 \log M_0$, take all primes in $(Y, 2Y]$ not dividing M_0, and let their product be P_0. By the PNT and the bound on the sum of logarithms of prime divisors of M_0, $\log P_0 \ge (2 - o(1)) \log M_0$, so $P_0 > 6M_0$. Any $s/P_0 \in [1/2, 2/3]$ can be represented by the binary-block proof of Lemma 10.1 without the first common power of two: take a power of two $B_* \ge 2Y$ with $B_* < 4Y$, use denominators $r2^j$ for $1 \le j \le h$ where $B_* = 2^h$. Multiplication by $P_0 B_*$ gives local binary residues modulo r and total selected reciprocal mass $\le \sum_{Y < r \le 2Y} r^{-1} = o(1)$. The residual is a binary fraction of denominator B_*. All denominators are prime to M_0 because M_0 is odd. Choose $s \equiv xP_0 \pmod{M_0}$ in the grid

interval for each x. This yields the asserted bound. For even moduli the powers of two are not units, and that argument is unavailable.

More generally, the grid proof in Section 10 works for a composite modulus whose least prime factor exceeds $C_1 \log_2 M_0 \log_3 M_0$ for a sufficiently large fixed C_1. Take $Y = 3 \log M_0$ and discard from the good primes of $(Y, 2Y]$ those dividing M_0. Their product still exceeds $6M_0$. All local multipliers and all prime factors in the dyadic completion are at most $M = O(\log_2 M_0 \log_3 M_0)$; thus all constructed denominators are units. The grid then gives $g_\times(M_0) \ll \log M_0 \log_2 M_0 \log_3 M_0$ in this regime. Small moduli need not have full unit coverage; for example modulo 6 the units 1 and 5 only generate the subset sums 0, 1, 5. No all-composites $O(\log M_0 \log_2 M_0)$ estimate is claimed.

B.2. **Coverage with bounded multiplicities.** Let $g_K(p)$ be the analogue of $g(p)$ allowing each denominator a coefficient in $\{0, 1, \dots, K\}$, still with real total cost at most one. Thus $g_1(p) = g(p)$.

Proposition B.1 (Bounded multiplicities)**.** *There exist fixed constants $C_1, C_2 > 0$ such that, for every sufficiently large prime p,*

$$g_{\lceil C_1 \log_2 p \rceil}(p) \le C_2 \log p \log_2 p.$$

One can use denominators in $[Q/2, Q]$ with $Q = \lfloor C_2 \log p \log_2 p \rfloor$, with total cost $O(1/\log_3 p)$. A product measure on the permitted digits is equidistributed modulo p with relative error $e^{-c \log p/\log_3 p}$.

Proof. Now take $L = \log p$, $U = \log L$, $V = \log U$, and use the rectangular block A of Section 7 at this scale. On each denominator rs choose an independent uniform digit in $\{0, \dots, s-1\}$, not a Bernoulli variable. Lemma 7.1's proof actually gives $\gg L/V$ entries of phase size at least $1/(4s_+)$. The normalized digit factor is $\sin(\pi sx)/(s \sin(\pi x))$ in absolute value, with $x = \|h(rs)^{-1}/p\|$. On those entries sx is bounded below by a positive constant; if $sx \ge 1$ the factor is at most $1/2$, and if $sx \in [c_0, 1]$ it is uniformly bounded below one in absolute value by comparison with $|\sin(\pi sx)/(\pi sx)|$ as $s \to \infty$. Thus the transform of A has modulus at most $e^{-cL/V}$ at each nonzero frequency.

For two other disjoint rectangles choose large-prime intervals $(16L, 18L]$ and $(20L, 22L]$, with respective small-prime intervals $(CU/30, CU/22]$ and $(CU/35, CU/24]$ where $C = C_2$ is sufficiently large. Their products lie in $[Q/2, Q]$ for large p. For each rectangle select small primes S' with $T' = \prod_{s \in S'} s \le L/U^3$ and $\log T' = U - O(V)$: a greedy choice reaches within $O(V)$ of the logarithmic threshold since the full product has log larger than U. Similarly select large primes R' so that $P'T' \le p$ and $\log(P'T') = L - O(U)$, with $P' = \prod_{r \in R'} r$; the full large-prime interval has log product $(2 + o(1))L$. Each rectangle has $|S'| \asymp U/V$ and $|R'| \asymp L/U$. Choose independent uniform digits $0, \dots, s-1$ at rs. All its digit assignments give distinct sums modulo p: a difference has magnitude $\le |S'| \sum_{r \in R'} 1/r = O(1/V) < 1$ and common denominator $P'T' \le p$, hence a collision must be an exact rational equality. Reducing modulo each outer prime r leaves the integer $T' \sum_s c_{rs}/s$ with absolute value $\le T'|S'| < r$, so it is zero. Reducing modulo each s gives $s \mid c_{rs}$, hence $c_{rs} = 0$ since $|c_{rs}| < s$.

Consequently Parseval gives squared Fourier sum $p/(T')^{|R'|} \le e^{O(LV/U)}$ for either of these independent digit blocks, because $|R'| \log T' = L - O(LV/U)$. Cauchy–Schwarz with the first block bounds the total absolute nonzero Fourier mass of the combined digit law by $\exp(-cL/V + O(LV/U)) \le e^{-c'L/V}$. This proves the asserted mixing and coverage by Fourier inversion. All three rectangles have distinct denominators, but digits may exceed one; their maximum is $O_C(U)$ and the real total cost is at most the sum over rectangles of $|S'| \sum_{r \in R'} 1/r = O_C(1/V)$ (the same bound for the first rectangle). This proves the proposition without establishing the case $K = 1$. □

B.3. **Height on most numerators of good prime denominators.** The entropy-scale coverage result also improves the height for most numerators of a typical prime denominator. Fix $t > \tau$.

Outside the exceptional primes in Theorem 1.2, for $p \in [P, 2P]$ take $q = \lfloor tb(P) \rfloor$. For every integer $1 \le a < p$ with

$$a \ge C_3(\log_2 p)^3(\log_3 p)^4 \tag{56}$$

where C_3 is sufficiently large, there is an exact distinct-unit-fraction representation of a/p of height at most $pq \le tpb(p)$. To prove this, use the full distinct-value cover to find $J \subseteq [1, q]$ with $\mathcal{H}(J) \equiv a \pmod p$ and $\mathcal{H}(J) \le 1$ (in fact the mass is $o(1)$). Then

$$w = \frac{a - \mathcal{H}(J)}{p} > 0, \qquad \operatorname{den}(w) \mid \mathcal{L}(q), \qquad w < 1.$$

Take $y = \lfloor 1/w \rfloor$; then $yw \in [1/2, 1]$ and $1 \le y \le p/(a-1) < p$. With $M = \lceil C_4' q(\log q)^3(\log_2 q)^4 \rceil$ and sufficiently large fixed C_4', every prime power of $\operatorname{den}(yw)$ is at most the powersmooth threshold in the interval lemma of Liu–Sawhney [4, Lemma 4.1]. Represent yw by distinct reciprocals with denominators in $[M/16, M]$, and scale these denominators by y. We have $M < p$, and neither y nor any completing denominator before scaling is divisible by p. Thus the completion is disjoint from the initial denominators pj, $j \in J$. As $\log q \sim \log_2 p$ and $\log_2 q \sim \log_3 p$, condition (56), with a sufficiently large constant, ensures $yM \le pq$. This proves the claim. The smallest numerators are not covered by this argument; it is not an improved bound for the worst-case height $D(p)$ at the entropy scale.

Scope of the results

The paper uses the dissociated-set estimate of [3] with variable iterated-logarithm depth as an external input. The powersmooth interval representation theorem of [4] is used only for the least-forbidden-denominator lower bound and the preceding most-numerators refinement. The entropy recursion, regular variation, distinct-value Fourier majorization, augmented-block reconstruction, exact CRT grid and deletion-loss estimates have been organized as self-contained arguments subject to these stated inputs and the standard prime number theorem, with the explicit inequalities of Rosser–Schoenfeld used in Lemma 5.1 to fix the numeric coverage constant. This is a consolidated research manuscript for independent review; the proofs have not been independently refereed or formally certified here. Neither the exact constant in the prime-square loss nor the pointwise all-primes binary bound is asserted. The counting order is credited to the literature and is included with its proof for completeness. Further arithmetic-statistical and pointwise expansion results are deliberately kept separate from the present entropy and coverage theory.

References

[1] M. N. Bleicher and P. Erdős, The number of distinct subsums of $\sum_1^N 1/i$, *Math. Comp.* (1975), 29–42.

[2] M. N. Bleicher and P. Erdős, Denominators of Egyptian fractions. II, *Illinois J. Math.* (1976), 598–613.

[3] S. Bettin, L. Grenié, G. Molteni and C. Sanna, A lower bound for the number of Egyptian fractions, *Math. Comp.*, published online 2026, doi:10.1090/mcom/4190; arXiv:2509.10030.

[4] Y. P. Liu and M. Sawhney, On further questions regarding unit fractions, *Int. Math. Res. Not. IMRN* 2026, rnaf382; arXiv:2404.07113v1 (the interval lemma is Lemma 4.1 in this version).

[5] P. Erdős and R. L. Graham, *Old and new problems and results in combinatorial number theory*, Monographies de L'Enseignement Mathématique 28, Geneva, 1980.

[6] H. Yokota, *On a problem of Bleicher and Erdős*, J. Number Theory **30** (1988), 198–207. doi:10.1016/0022-314X(88)90017-0.

[7] J. B. Rosser and L. Schoenfeld, Approximate formulas for some functions of prime numbers, *Illinois J. Math.* **6** (1962), 64–94. doi:10.1215/ijm/1255631807.

[8] P. Chen, *Sharp Two-Term Capacity Asymptotics for Egyptian Fractions via Rational Grid Representations*, Hexagon preprint 2611.01225v1 (2026), `https://hexagonmath.org/2611.01225v1`.

[9] P. Chen, *Sharp Second-Order Entropy Asymptotics for Disjoint Egyptian Fraction Representations*, Hexagon preprint 2611.01226v1 (2026), `https://hexagonmath.org/2611.01226v1`.