

**ADDITIVE BASES WITH MANY REPRESENTATIONS AND NO MINIMAL
SUBBASIS
A NEGATIVE ANSWER TO ERDŐS PROBLEM #870**

DAVID CORNELIU TURTUREAN

Fulcrum*

davidct@mit.edu

ABSTRACT. An additive basis of order k is a set of positive integers from which every sufficiently large integer can be formed using at most k summands. Erdős Problem #870 asks whether a sufficiently large logarithmic lower bound on the number of such representations forces the basis to contain a minimal subbasis. We prove that it does not. For every $k \geq 3$ and every $C > 0$, we construct an order- k basis E for which every sufficiently large n has at least $C \log n$ representations, while no order- k subbasis of E is minimal.

The proof starts from the order-2 construction of Larsen and Larsen. For $k \geq 4$, an elementary finite residue-class gadget both amplifies the number of representations and preserves non-minimality. The case $k = 3$ cannot be reduced to the order-2 result by an external gadget, so we modify the randomized construction directly, building in enough local redundancy to withstand the finitely many ways an order-3 subbasis can recombine representations.

1. INTRODUCTION

Throughout, $\mathbb{N} = \{1, 2, 3, \dots\}$. We use the following convention throughout the paper: an additive representation of order k means a representation using at most k summands, not necessarily exactly k . One-summand representations are allowed, repeated summands are allowed, and the order of the summands is ignored.

More precisely, let $A \subseteq \mathbb{N}$ and let $k \geq 1$. A representation of n by elements of A of order at most k is a nondecreasing sum

$$n = a_1 + \dots + a_s, \quad 1 \leq s \leq k, \quad a_1 \leq \dots \leq a_s, \quad a_i \in A.$$

We write $R_{A,k}(n)$ for the number of these representations. Thus repeated summands are allowed, while two orderings of the same summands are counted only once. The set A is an *additive basis of order k* if $R_{A,k}(n) > 0$ for every sufficiently large n . A *subbasis* of A is a subset $D \subseteq A$ that is again a basis of order k in the same at-most- k sense. Such a basis is *minimal* if none of its proper subsets is again a basis of order k .

For exact two-summand representations we use the shorter notation

$$r_A(n) := \#\{(a, b) : a, b \in A, a \leq b, a + b = n\},$$

which counts exactly two summands. We also write $A(x) = |A \cap [1, x]|$. A set has density zero when $A(x) = o(x)$, and a set of integers is *cofinite* when its complement is finite.

In the order-2 case, the at-most-2 basis condition above is that $A \cup (A + A)$ is cofinite, not merely that $A + A$ is cofinite.

Date: June 2026.

*Also affiliated with the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139.

Incoming, Physical Superintelligence PBC.

Erdős Problem #870 asks whether, for every fixed $k \geq 3$, there is a constant $c(k) > 0$ with the following property: every order- k basis A satisfying

$$R_{A,k}(n) \geq c(k) \log n$$

for all sufficiently large n contains a minimal order- k basis [1]. We prove the precise negation.

Theorem 1.1. *For every integer $k \geq 3$ and every real number $C > 0$, there is a set $E \subseteq \mathbb{N}$ such that*

- (1) E is an additive basis of order k ;
- (2) $R_{E,k}(n) \geq C \log n$ for every sufficiently large n ; and
- (3) no subset of E is a minimal additive basis of order k .

The proof has one common source and two different deterministic reductions. Larsen and Larsen constructed an order-2 basis with logarithmically many representations and no minimal order-2 subbasis [4]. We first record the order-2 input in the at-most- k form just defined. When $k \geq 4$, this order-2 set is multiplied by a modulus and combined with finitely many small *fillers*. A specially chosen residue class then forces every subbasis of the resulting set to project to a subbasis of the order-2 set.

For $k = 3$, only one filler is available after reserving two summands for the order-2 core (a scaled copy of the order-2 set), and the residue argument no longer has enough room. We instead modify the random construction itself. The fragile integers in the Larsen–Larsen proof are often called *canaries*: they are deliberately given many representations, but every one of those representations is forced to use a prescribed control set. In the modified construction, one random center produces a finite cluster of nearby canaries. This lets one control all of the finitely many shifts that arise from the parity gadget for $k = 3$.

The deterministic reductions are given first, in Sections 3 and 4. The order-2 input is proved in Section 5, and the clustered input in Sections 6 and 7. Every specialized term used in the random construction is defined before it enters an argument. Appendix A records the correspondence with the accompanying Lean formalization.

2. THE TWO INPUT STATEMENTS

The first input is the Larsen–Larsen order-two construction in the form needed here. Its last clause is stated in the at-most-two language: every order-2 subbasis has a deletable element.

Proposition 2.1 (order-2 input). *There are a constant $\eta_2 > 0$ and a set $A \subseteq \mathbb{N}$ with the following properties.*

- (1) $A + A$ is cofinite.
- (2) $r_A(n) \geq \eta_2 \log n$ for every sufficiently large n .
- (3) $A(x) = o(x)$.
- (4) If $D \subseteq A$ and $D \cup (D + D)$ is cofinite, then some $d \in D$ satisfies

$$(D \setminus \{d\}) \cup ((D \setminus \{d\}) + (D \setminus \{d\})) \text{ is cofinite.}$$

Because order two means one or two summands here, a set D is an order-two basis exactly when $D \cup (D + D)$ is cofinite. Thus part (4) says that every order-two subbasis of A remains an order-two basis after deleting one of its elements.

The second input is designed for the parity argument in order 3. We first define the finite-shift expression that it controls. If $U, V \subseteq \mathbb{Z}_{\geq 0}$ are finite and $D \subseteq \mathbb{N}$, put

$$(1) \quad \Phi_{U,V}(D) := (U + (D \cup (D + D))) \cup (V + D).$$

In expanded form, $x \in \Phi_{U,V}(D)$ when at least one of the following holds:

$$\begin{aligned} x &= u + d, & u &\in U, d \in D, \\ x &= u + d_1 + d_2, & u &\in U, d_1, d_2 \in D, \\ x &= v + d, & v &\in V, d \in D. \end{aligned}$$

A pair (U, V) with $U \neq \emptyset$ will be called a *shift package*. The word “package” merely means that the finitely many shifts in U and V are to be handled simultaneously.

Proposition 2.2 (clustered order-3 input). *There is an absolute constant $\eta_3 > 0$ with the following property. Let \mathcal{P} be any finite list of shift packages, and let $P_0 \subseteq \mathbb{N}$ be finite. There is a set $A \subseteq \mathbb{N}$ such that*

- (1) $A \cap P_0 = \emptyset$;
- (2) $A + A$ is cofinite;
- (3) $r_A(n) \geq \eta_3 \log n$ for every sufficiently large n ;
- (4) $A(x) = o(x)$;
- (5) for every $(U, V) \in \mathcal{P}$, every $D \subseteq A$, and every finite $F_0 \subseteq D$, if $\Phi_{U,V}(D)$ is cofinite, then there is an element $d \in D \setminus F_0$ such that, for $D' = D \setminus \{d\}$,

$$\Phi_{U,V}(D') \text{ is cofinite} \quad \text{and} \quad D' \text{ is an additive basis of order 3.}$$

Part (5) is the form the parity reduction needs. It does not require D to be a basis, only that the shifted set $\Phi_{U,V}(D)$ be cofinite; it then returns a single deletion that keeps $\Phi_{U,V}(D')$ cofinite while turning D' into an order-3 basis.

3. THE FINITE-FILLER REDUCTION FOR $k \geq 4$

Fix $k \geq 4$ and write

$$h := k - 2.$$

The h filler summands will control the residue class, leaving at most two summands for the order-2 core. We need many filler choices in every residue, and one residue in which the number of fillers is forced to be at least h . Both requirements are supplied by the following elementary lemma.

Lemma 3.1 (finite filler gadget). *Let $h \geq 2$ and $L \geq 1$. There are integers $N < M$, a residue $\tau \in \{0, \dots, M-1\}$, and a finite set $F = [1, N] \cap \mathbb{N}$ such that:*

- (1) for every residue $r \pmod{M}$, there are at least L distinct multisets of exactly h elements of F whose sums are congruent to $r \pmod{M}$;
- (2) if a multiset of at most $h+1$ elements of F has sum congruent to $\tau \pmod{M}$, then its sum is exactly τ and it contains at least h elements.

Proof. Choose an integer $R \geq 4L + h$ and set

$$(2) \quad N := 4R, \quad M := (4h-1)R, \quad \tau := (4h-2)R, \quad F = [1, N] \cap \mathbb{N}.$$

Since $h \geq 2$, we have $N < M$ and $0 < \tau < M$.

We first produce many fillers in each residue. Put

$$B := 2L + h - 1, \quad U := hN + 1 - 2L.$$

The interval $[B, U]$ contains at least M consecutive integers, because

$$U - B + 1 = 4hR - 4L - h + 3 \geq (4h-1)R = M.$$

Consequently every residue $r \pmod{M}$ has a representative $t_r \in [B, U]$.

We claim that every integer $t \in [B, U]$ has at least L distinct representations as a sum of exactly h elements of F . Choose an integer S satisfying

$$(3) \quad 2L + 1 \leq S \leq 2N + 1 - 2L, \quad h - 2 \leq t - S \leq (h-2)N.$$

Such an S exists: the possible interval is

$$\max\{2L + 1, t - (h - 2)N\} \leq S \leq \min\{2N + 1 - 2L, t - (h - 2)\},$$

and the four required comparisons follow directly from $B \leq t \leq U$. Every integer in $[h - 2, (h - 2)N]$ is a sum of $h - 2$ elements of $[1, N]$ (for example, start with $h - 2$ ones and distribute the remaining amount one summand at a time). Thus $t - S$ has such a representation; for $h = 2$ this is the empty sum. Fix one resulting background multiset G .

It remains to write S in L different ways as a sum of two fillers. If $S \leq N + 1$, use

$$S = i + (S - i), \quad 1 \leq i \leq L.$$

Here $i \leq L$ while $S - i \geq L + 1$, and both terms lie in $[1, N]$. If $S > N + 1$, use, for $0 \leq i < L$,

$$S = (S - N + i) + (N - i).$$

The first terms lie at most at $N - L$, while the second terms lie at least at $N - L + 1$. Thus these are again L distinct unordered pairs in F . Adding the same background multiset G gives L distinct h -element multisets of sum t . Applying this to $t = t_r$ proves part (1).

For part (2), let s be the sum of at most $h + 1$ fillers and assume $s \equiv \tau \pmod{M}$. Then

$$s \leq (h + 1)N = (4h + 4)R < (8h - 3)R = \tau + M.$$

Because $0 < \tau < M$, the only nonnegative integer below $\tau + M$ that is congruent to τ is τ itself. Hence $s = \tau$. Finally, a sum of at most $h - 1$ fillers is at most

$$(h - 1)N = (4h - 4)R < \tau,$$

so a filler sum equal to τ must use at least h terms. \square

Proposition 3.2. *For every $k \geq 4$ and every $C > 0$, there is a set $E \subseteq \mathbb{N}$ satisfying all three conclusions of Theorem 1.1.*

Proof. Let A and η_2 be given by Proposition 2.1, and set $h = k - 2$. Choose L so large that

$$(4) \quad \frac{L\eta_2}{2} > C.$$

Apply Lemma 3.1 and define

$$(5) \quad E := MA \cup F, \quad MA := \{Ma : a \in A\}.$$

Since $M > N$ and $A \subseteq \mathbb{N}$, the scaled core MA and the filler set F are disjoint.

Basis property and representation count. Write $n = Mq + r$ with $0 \leq r < M$. Fix one of the h -element filler multisets from Lemma 3.1, and write its sum as

$$Mu + r$$

for a nonnegative integer u . For all sufficiently large q , the integer $q - u$ belongs to $A \cup (A + A)$. Scaling a representation of $q - u$ by M and adding the h fillers represents n by at most $h + 2 = k$ elements of E . Thus E is an order- k basis.

For the lower bound, use all L filler multisets in residue r . Their quotients u belong to a fixed finite set. Every exact pair representation $q - u = a + b$ gives

$$Mq + r = Ma + Mb + (\text{the chosen } h \text{ fillers}).$$

Different core pairs or different filler multisets give different nondecreasing representations: all core terms are at least M , while every filler is smaller than M . Hence, uniformly over the finitely many residues,

$$R_{E,k}(Mq + r) \geq \sum_{j=1}^L r_A(q - u_{r,j}).$$

For large q , every $q - u_{r,j}$ is large and

$$\log(q - u_{r,j}) \geq \frac{1}{2} \log(Mq + r).$$

Therefore

$$R_{E,k}(Mq + r) \geq \frac{L\eta_2}{2} \log(Mq + r) \geq C \log(Mq + r)$$

for all sufficiently large q .

Projection of an arbitrary subbasis. Let $T \subseteq E$ be an order- k basis, and define

$$D := \{a \in A : Ma \in T\}, \quad F_T := T \cap F.$$

Consider the progression

$$n_q := Mq + \tau.$$

For all sufficiently large q , choose a representation of n_q by at most $h + 2$ elements of T . Split its terms into core elements, which are multiples of M , and fillers from F_T . The filler sum is congruent to $\tau \pmod{M}$.

If the representation uses at most $h + 1$ fillers, the rigidity clause of Lemma 3.1 says that their sum is exactly τ and that there are at least h of them. Hence at most two core elements remain, and their quotients sum to q . A representation using $h + 2$ fillers and no core element has bounded total sum, so it cannot represent n_q for large q . It follows that

$$q \in D \cup (D + D)$$

for every sufficiently large q . Thus D is an at-most-two subbasis of A .

By Proposition 2.1, choose $d \in D$ such that

$$D' := D \setminus \{d\}$$

is still an at-most-two basis. We shall prove that

$$T' := T \setminus \{Md\}$$

is still an order- k basis.

Enough fillers survive in every residue. Fix a residue $r \pmod{M}$. We claim that some multiset of at most h elements of F_T has sum congruent to $r \pmod{M}$. Suppose not. Then every representation of $Mq + r$ by at most $h + 2$ elements of T uses at least $h + 1$ fillers and therefore at most one core element.

There are only finitely many filler multisets of cardinality $h + 1$ or $h + 2$ from the finite set F_T . For each fixed filler multiset, a representation with one core element forces q to lie in one fixed translate of D ; a representation with no core element can occur for at most one value of q . Consequently all sufficiently large q would lie in a finite union of translates of D .

This is impossible. Since $D \subseteq A$ and $A(x) = o(x)$, the set D has density zero. Every fixed translate of a density-zero set also has density zero, as does a finite union of such translates; a cofinite set has density one. The claim follows.

Choose, for each residue r , one surviving filler multiset G_r , drawn from F_T and of size at most h , and write

$$\sum_{g \in G_r} g = Mu_r + r.$$

Since D' is an at-most-two basis, every sufficiently large $q - u_r$ is a sum of at most two elements of D' . After scaling by M and adding G_r , we obtain a representation of $Mq + r$ by at most $h + 2 = k$ elements of T' . This works in every residue class, so T' is an order- k basis. Thus no order- k subbasis of E is minimal. \square

4. THE PARITY REDUCTION FOR ORDER 3

The order-3 construction uses one finite filler and two core summands. Its purpose is to turn representations of an integer q by the order-2 core into representations of both $2q$ and $2q + 1$.

Proposition 4.1. *For every $C > 0$, there is a set $E \subseteq \mathbb{N}$ satisfying all three conclusions of Theorem 1.1 with $k = 3$.*

Proof. Let η_3 be the constant from Proposition 2.2. Choose $J \geq 1$ so large that

$$(6) \quad \frac{J\eta_3}{2} > C,$$

and put

$$P := \{1, 2, \dots, J\}.$$

For every nonempty $U \subseteq P$ and every $V \subseteq P + P$, include the shift package (U, V) in a finite list \mathcal{P} . Apply Proposition 2.2 to \mathcal{P} and to the forbidden set P , obtaining a set A with $A \cap P = \emptyset$. Define the finite filler set

$$F := \{2p, 2p + 1 : p \in P\}$$

and the order-3 set

$$(7) \quad E := 2A \cup F.$$

The condition $A \cap P = \emptyset$ ensures that no even filler $2p$ is also a core element $2a$.

Basis property and representation count. Let $n = 2q + r$ with $r \in \{0, 1\}$. For any $p \in P$ and all sufficiently large q , the integer $q - p$ belongs to $A + A$. If $q - p = a + b$, then

$$(8) \quad 2q + r = 2a + 2b + (2p + r).$$

Thus E is an order-3 basis.

For each $p \in P$, every exact representation of $q - p$ gives a distinct representation in (8). The filler identifies p , and the core is disjoint from the filler set. Hence

$$R_{E,3}(2q + r) \geq \sum_{p \in P} r_A(q - p).$$

For large q , $\log(q - p) \geq \frac{1}{2} \log(2q + r)$ uniformly in $p \in P$. It follows from (6) that

$$R_{E,3}(2q + r) \geq C \log(2q + r)$$

for every sufficiently large q and both parities r .

The shift package determined by a subbasis. Let $T \subseteq E$ be an arbitrary order-3 basis. Its projected core is

$$D := \{a \in A : 2a \in T\}.$$

Define

$$(9) \quad U := \{p \in P : 2p + 1 \in T\},$$

$$(10) \quad V := \{p + p' : p, p' \in P, 2p + 1 \in T, 2p' \in T\}.$$

The set U is nonempty. Indeed, every representation of a sufficiently large odd integer by elements of T must use an odd filler; the core and the even fillers are all even, while a sum of three odd fillers is bounded because F is finite.

We now show that $\Phi_{U,V}(D)$ is cofinite. Consider a sufficiently large odd integer $2q + 1$ and a representation by at most three elements of T . Exactly one term is an odd filler, say $2p + 1$. Purely finite filler representations account for only finitely many q , so for large q one of the following occurs:

- (1) the remaining terms are one or two core elements, giving $q \in p + (D \cup (D + D))$;
- (2) the remaining terms are one even filler $2p'$ and one core element, giving $q \in (p + p') + D$.

By (9) and (10), these two alternatives say exactly that $q \in \Phi_{U,V}(D)$. Hence $\Phi_{U,V}(D)$ is cofinite. Apply Proposition 2.2 to this package. We obtain $d \in D$ such that, with $D' = D \setminus \{d\}$,

$$\Phi_{U,V}(D') \text{ is cofinite} \quad \text{and} \quad D' \text{ is an order-3 basis.}$$

Because $d \in A$ and $A \cap P = \emptyset$, the element $2d$ is not a filler. Set

$$T' := T \setminus \{2d\}.$$

Every sufficiently large even integer $2q$ is represented by at most three terms from $2D' \subseteq T'$, because D' is an order-3 basis. For odd integers, take a large $q \in \Phi_{U,V}(D')$. A witness in $U + (D' \cup (D' + D'))$ gives a representation using one odd filler and one or two doubled core terms. A witness in $V + D'$ gives, by the definition of V , a representation using one odd filler, one even filler, and one doubled core term. In both cases $2q + 1$ is represented by at most three elements of T' . Therefore T' is still an order-3 basis. Since T was arbitrary, E contains no minimal order-3 subbasis. \square

Theorem 1.1 now follows from Propositions 3.2 and 4.1, once the two input propositions have been established.

5. THE LARSEN–LARSEN INPUT IN AT-MOST-TWO FORM

We now prove Proposition 2.1. Most of the construction is exactly the one in [4]; the purpose of this section is to isolate the three points needed for the present application:

- (1) the final set has density zero uniformly in the cutoff x ;
- (2) a large canary is not itself an element of the final set;
- (3) the exactness argument for canary representations also excludes an accidental representation involving an element from an earlier stage.

Once these facts are recorded, the passage from exact two-summand representations to order two in the at-most sense is short.

5.1. The construction and its terminology. Set

$$(11) \quad X_n := 2^{2^n}, \quad I_n := [X_n, X_{n+1}) \cap \mathbb{N}.$$

Thus $X_{n+1} = X_n^2$. At stage n , every integer $x \in I_n$ is independently placed in a *raw sample* A_n with probability

$$(12) \quad p(x) = \min \left\{ 1, K \sqrt{\frac{\log x}{x}} \right\},$$

where K is a sufficiently large absolute constant. The union of all raw samples is denoted by $A^{(0)}$. An integer sampled at the current stage is called *fresh*; one fixed at an earlier stage is called *old*.

Conditional on the schedule (the data fixed by all earlier stages), a sparse set B_n of integers is chosen uniformly without replacement, independently of the fresh Bernoulli sample, in the *canary window*

$$(13) \quad W_n := \left[\frac{X_{n+1}}{6}, \frac{X_{n+1}}{4} \right) \cap \mathbb{N}.$$

The elements of B_n are the canaries. They are fragile test integers used to detect whether a proposed subbasis has been thinned too much. Every fresh raw element that participates in a representation of a canary is deleted. This removes all representations of each canary that existed before restorations are added.

The restoration step uses information from ten stages earlier. Certain integers $c \in I_{n-10}$ with many surviving pair representations are declared *robust*. From a fixed collection of representations

of such a c , one forms a finite *control set* S : it contains one summand from each of many selected pairs. A canary $b \in B_n$ is assigned to the data (c, S) , and for every $s \in S$ the integer

$$(14) \quad b - s$$

is added to the final set. The equality

$$(15) \quad b = s + (b - s)$$

is an *intended representation* of the canary. The summand s is old: it was fixed ten stages before b was chosen. The newly added term $b - s$ is called a *restoration element*.

The following estimates from [4] will be used repeatedly. Almost surely, for all sufficiently large n ,

$$(16) \quad |B_n| \ll X_n^{1/8},$$

$$(17) \quad |A_n| \ll X_n^{1+o(1)},$$

$$(18) \quad |A_n'' \setminus A_n'| \ll X_n^{1/4+o(1)},$$

where A_n' is the stage- n raw set after deletion and A_n'' is the stage set after restorations have been added. The union of all control summands used at stage n , which we denote by S_n , satisfies

$$(19) \quad S_n \subseteq \left[\frac{X_{n-10}}{4}, X_{n-9} \right), \quad |S_n| \ll X_n^{1/512+o(1)}.$$

The exponent $1/512$ is the ten-stage separation: $X_{n-9} = X_n^{1/512}$.

Larsen and Larsen prove that, with probability one, the canaries and robust integers together cover all sufficiently large integers, every large integer has $\gg \log n$ representations, and the only representations of a large canary are the intended ones in (15), subject to the additional old-summand check supplied below.

5.2. Uniform density zero.

Lemma 5.1. *Almost surely, the final Larsen–Larsen set A satisfies $A(x) = o(x)$.*

Proof. We first count the raw Bernoulli sample. On a dyadic interval $[y, 2y)$, the expected number of selected integers is

$$O\left(\sqrt{y \log y}\right).$$

A Chernoff upper-tail bound, followed by Borel–Cantelli over the dyadic intervals, gives almost surely

$$(20) \quad |A^{(0)} \cap [1, x]| = O\left(\sqrt{x \log x}\right)$$

for every sufficiently large x , not merely for $x = X_n$.

Now let $x \in I_N$. By (18), the total number of restoration elements created up to stage N is at most

$$\sum_{n \leq N} X_n^{1/4+o(1)} = X_N^{1/4+o(1)},$$

because the sequence X_n grows doubly exponentially. Restoration elements from later stages are larger than x once N is large. Since $x \geq X_N$, the restoration contribution is $o(x)$, and the raw contribution in (20) is also $o(x)$. Deletions only reduce the set, so the final set has density zero. \square

5.3. Canary exactness, including old summands. The canary calculation in [4] rules out an unintended representation using a restoration element and a fresh stage- n summand. We also need the same conclusion when the other summand was already present before stage n .

Lemma 5.2 (old-summand exclusion). *Almost surely, for all sufficiently large n , there are no distinct $b, b' \in B_n$, no $s \in S_n$, and no old element $a \in A''(n-1) := \bigcup_{j < n} A_j''$ satisfying*

$$(21) \quad b' = a + (b - s).$$

Proof. Condition on all choices made before the random canary set B_n is selected. Equation (21) is equivalent to

$$b' - b = a - s.$$

Thus a bad event occurs only if the difference set $B_n \ominus B_n$ meets

$$E_n := A''(n-1) - S_n.$$

The earlier stages contain

$$|A''(n-1)| \ll X_n^{1/2+o(1)}$$

elements: the raw Bernoulli contribution dominates, and the restoration contribution is smaller. Together with (19), this gives

$$|E_n| \ll X_n^{1/2+1/512+o(1)}.$$

For any fixed nonzero difference e , a uniformly chosen set of $O(X_n^{1/8})$ points in a window of length $\asymp X_n^2$ contains an ordered pair with difference e with probability $O(X_n^{1/4}/X_n^2)$. A union bound over E_n therefore gives

$$\mathbb{P}((B_n \ominus B_n) \cap E_n \neq \emptyset) \ll X_n^{-5/4+1/512+o(1)}.$$

The series over n converges. Borel–Cantelli proves the lemma. \square

The same difference estimate with A_n in place of $A''(n-1)$ gives the fresh-summand exclusion used in [4]; its stage probability is $X_n^{-3/4+1/512+o(1)}$, still summable. We also need to know that a canary is not itself available as a one-term representation.

Lemma 5.3. *Almost surely, every sufficiently large canary lies outside the final set A .*

Proof. A stage- n canary can enter A in only two ways. First, it might be selected in the raw Bernoulli sample. Since a canary has size $\asymp X_n^2$, (16) gives

$$\mathbb{P}(B_n \cap A_n \neq \emptyset) \ll X_n^{1/8} X_n^{-1} \sqrt{\log X_n} = X_n^{-7/8+o(1)}.$$

Second, one canary might coincide with a restoration element belonging to another canary. Such a coincidence has the form $b = b' - s$, so a canary difference belongs to S_n . The same fixed-difference estimate gives

$$\mathbb{P}((B_n \ominus B_n) \cap S_n \neq \emptyset) \ll X_n^{-7/4+1/512+o(1)}.$$

Both bounds are summable. Same-canary coincidence would require $s = 0$, which is impossible, and later stages add only much larger integers. The conclusion follows from Borel–Cantelli. \square

Combining the deletion rule, the fresh-summand estimate from [4], and Lemma 5.2, we obtain the exact form of canary fragility that will be used below:

$$(22) \quad \text{every sufficiently large canary } b \text{ has only the intended representations } b = s + (b - s).$$

5.4. From the exact order-2 construction to at most two summands.

Proof of Proposition 2.1. Larsen and Larsen prove that the final set has a cofinite two-fold sumset and that there is an absolute $\eta_2 > 0$ such that $r_A(n) \geq \eta_2 \log n$ for all sufficiently large n [4]. Lemma 5.1 gives density zero. It remains to prove the deletion statement in the at-most-two convention.

Let B be the union of the canary sets and C the union of the robust sets. The construction gives

$$(23) \quad B \cup C \text{ cofinite.}$$

Let $D \subseteq A$ and suppose $D \cup (D + D)$ is cofinite. By Lemma 5.3, every sufficiently large canary lies outside D . It must therefore lie in $D + D$.

We claim that every sufficiently large robust integer $c \in C$ has at least two exact representations by elements of D . Suppose instead that $r_D(c) \leq 1$. The construction selected many pair representations of c . At most one selected pair can lie entirely in D . Omit that pair, if it exists, and from every other selected pair choose one summand outside D . This produces one of the control sets S scheduled by the construction, with

$$S \subseteq A \setminus D.$$

Let b be the canary assigned to (c, S) . By (22), every representation of b uses an element of S , so $b \notin D + D$. For large c , the assigned canary b is also large, contradicting the preceding paragraph. This proves the claim.

Choose any $d \in D$ and put $D' = D \setminus \{d\}$. A fixed summand d can occur in at most one of two distinct representations of the same robust integer, except at the single irrelevant target $2d$. Hence every sufficiently large $c \in C$ still lies in $D' + D'$.

The same conclusion holds for sufficiently large canaries. Indeed, by (22), a large canary representation consists of an old control and its corresponding restoration element. For a fixed integer d , the control windows eventually lie above d , and the restoration windows also tend to infinity. Thus d occurs in only finitely many canary representations. Since every large canary was in $D + D$, all but finitely many remain in $D' + D'$ after d is removed.

By (23), the set $D' + D'$ is cofinite. In particular $D' \cup (D' + D')$ is cofinite, so D' is still an at-most-two basis. This proves part (4) and completes the proposition. \square

6. THE CLUSTERED CONSTRUCTION

We turn to Proposition 2.2. The underlying random set is still built stage by stage as in Section 5. The new feature is that one random center produces several nearby canaries, one for every shift that may appear in a prescribed finite list of packages.

6.1. What a cluster is designed to do. Fix a finite list

$$\mathcal{P} = \{(U_\lambda, V_\lambda) : \lambda \in \Lambda\}$$

of shift packages. Put

$$(24) \quad \Omega := \{0\} \cup \bigcup_{\lambda \in \Lambda} (U_\lambda \cup V_\lambda), \quad \Delta := (\Omega - \Omega) \setminus \{0\}.$$

Here $\Omega - \Omega$ is a finite subset of \mathbb{Z} . The shifts in Ω tell us which canaries to place around a center; the nonzero differences in Δ are precisely the possible accidental shifts between two members of the same cluster.

The terminology used below can be summarized as follows.

Term	Meaning
raw element	an element chosen by the Bernoulli sample before any deletion or restoration
robust integer	a noncanary integer retaining many specially chosen raw pair representations
center y	a random point used to create one finite family of nearby canaries
cluster at y	the set $\{y - \omega : \omega \in \Omega\}$
control set S	one selected summand from each of many robust pairs, with one pair omitted
restoration element	$y - \omega - s$, added so that $y - \omega = s + (y - \omega - s)$
Δ -clean summand	a summand s for which $s + \delta$ is absent from the final set for every $\delta \in \Delta$

Suppose a robust integer c has many disjoint pair representations. We shall select $T + 1$ of them. A control set chooses one summand from T of these pairs, leaving one pair unchosen. If a set D has at most one representation of c , then the omitted pair can be chosen to contain that possible D -representation; from every other pair one chooses a summand outside D . Thus one of the scheduled control sets satisfies

$$(25) \quad S \subseteq A \setminus D.$$

Now assign a center y to the data (λ, c, S) . For every $\omega \in \Omega$, declare $y - \omega$ to be a canary and add the restoration elements

$$(26) \quad y - \omega - s, \quad s \in S.$$

If these are the only representations of the canary $y - \omega$, then (25) implies

$$y - \omega \notin D + D.$$

In particular, $y - u \notin D \cup (D + D)$ for every $u \in U_\lambda$, and $y - v \notin D$ for every $v \in V_\lambda$. This gives

$$y \notin \Phi_{U_\lambda, V_\lambda}(D),$$

which contradicts cofiniteness. The whole construction is arranged to make this obstruction valid.

The role of Δ -cleanness is also visible immediately. A restoration from the same center gives

$$(y - \omega_0 - s) + a = y - \omega \implies a = s + \omega_0 - \omega.$$

If $\omega_0 = \omega$, this is the intended representation. Otherwise $\omega_0 - \omega \in \Delta$, so Δ -cleanness excludes the accidental summand a .

6.2. Two elementary estimates for random centers. We shall repeatedly use the following bounds. They are only sampling-without-replacement estimates; no feature of the additive construction enters them.

Lemma 6.1 (cluster estimates). *Let W be an interval of integers, let Y be a uniformly chosen K -element subset of W , and fix a finite set $\Omega \subseteq \mathbb{Z}$. Put*

$$B^\Omega := \{y - \omega : y \in Y, \omega \in \Omega\}.$$

Then the following hold.

(1) *For every set $S \subseteq \mathbb{Z}$ and every integer $\ell \geq 1$,*

$$(27) \quad \mathbb{P}(|B^\Omega \cap S| \geq \ell) \ll_{\ell, \Omega} \left(\frac{K|S + \Omega|}{|W|} \right)^{\lceil \ell/|\Omega| \rceil}.$$

(2) For every finite difference set $E \subseteq \mathbb{Z}$,

$$(28) \quad \mathbb{P}((Y \ominus Y) \cap E \neq \emptyset) \ll \frac{K^2|E|}{|W|},$$

where $Y \ominus Y = \{y - y' : y, y' \in Y, y \neq y'\}$.

Proof. If $B^\Omega \cap S$ contains ℓ points, then at least $\lceil \ell/|\Omega| \rceil$ distinct centers lie in $S + \Omega$. The probability that a fixed set of j points of W is contained in a uniformly chosen K -subset is $O_j((K/|W|)^j)$. Summing over the possible j -subsets of $S + \Omega$ gives (27).

For a fixed $e \neq 0$, there are at most $|W|$ ordered pairs $(w, w') \in W^2$ with $w - w' = e$, and the probability that a fixed ordered pair is contained in Y is $O(K^2/|W|^2)$. Summing first over such pairs and then over $e \in E$ proves (28). \square

6.3. Clean raw representations. Let

$$(29) \quad M_\Delta := 1 + \max(\{0\} \cup \{|\delta| : \delta \in \Delta\}).$$

For $m \in I_n$, a raw representation $m = a + a'$ is called *buffered* when

$$(30) \quad a, a' \in \left[\frac{X_n}{4} + M_\Delta, m - \frac{X_n}{4} - M_\Delta \right].$$

A summand s is *stage-safe* if, for every $\delta \in \Delta$, the two integers s and $s + \delta$ lie in the same stage interval I_j . Finally, a summand is Δ -*clean* if

$$(31) \quad s + \delta \notin A \quad (\delta \in \Delta)$$

in the final set.

The buffer makes cleanness local in time. If s occurs in a buffered representation of a target in I_n , then

$$\frac{X_n}{4} < s + \delta < X_{n+1} \quad (\delta \in \Delta).$$

Restoration elements from stages before n lie below $X_n/4$, while those from stages after n lie above X_{n+1} , once n is large. Thus only the stage- n Bernoulli sample and stage- n restorations can cause $s + \delta \in A$. The truth of (31) is therefore known by the end of stage n .

The following is the sole probabilistic input needed to guarantee enough robust integers. Its proof is deferred to Section 7, where the *losses* (raw pair representations removed by the construction) are separated into three classes.

Lemma 6.2 (clean representation supply). *Fix finite sets $\Omega \subseteq \mathbb{Z}_{\geq 0}$ and $\Delta \subseteq \mathbb{Z} \setminus \{0\}$, and let $L_0 < 159$. Suppose that at stage n :*

- (1) *the canary centers form a uniformly chosen set Y_n in a subinterval of W_n of length $\asymp X_n^2$;*
- (2) *$K_n := |Y_n| \leq X_n^{1/8}$;*
- (3) *the schedule and all old control summands are fixed before Y_n and the fresh Bernoulli sample A_n are chosen;*
- (4) *the union S_n of the old control summands satisfies*

$$(32) \quad S_n \subseteq \left[\frac{X_{n-10}}{4}, X_{n-9} \right), \quad |S_n| \ll X_n^{1/512+o(1)}.$$

The fresh Bernoulli sample and the center set are chosen independently, conditional on all earlier stages. If the Bernoulli constant K in (12) is sufficiently large in terms of Ω, Δ, L_0 , then, almost surely for all sufficiently large n , every

$$m \in I_n \setminus B_n^\Omega, \quad B_n^\Omega := \{y - \omega : y \in Y_n, \omega \in \Omega\},$$

has at least $L_0 \log m$ surviving buffered, stage-safe, Δ -clean raw representations after the stage- n deletion and restoration steps.

6.4. Scheduling the clusters. We now construct the set before imposing the finite avoidance condition P_0 . The construction is inductive. At the end of stage m , all choices through that stage are known. Put

$$(33) \quad T_m := 60 \log_2 X_m = 60 \cdot 2^m.$$

Call an integer $c \in I_m$ *robust* if it is not a stage- m canary and it retains at least $T_m + 1$ buffered, stage-safe, Δ -clean raw pair representations. Denote the set of such integers by C_m .

Choose $L_0 = 61/\log 2$, which is smaller than 159. For large m ,

$$T_m + 1 \leq L_0 \log c \quad (c \in I_m).$$

After increasing the Bernoulli constant in Lemma 6.2, every noncanary in I_m is robust. Hence, almost surely for all sufficiently large m ,

$$(34) \quad B_m^\Omega \cup C_m = I_m.$$

In particular, if

$$B := \bigcup_m B_m^\Omega, \quad C := \bigcup_m C_m,$$

then

$$(35) \quad B \cup C \text{ is cofinite.}$$

For every $c \in C_m$, choose, by a fixed deterministic rule, $T_m + 1$ of its surviving clean representations

$$(36) \quad c = a_i + a'_i, \quad 1 \leq i \leq T_m + 1.$$

Distinct unordered representations of one integer are automatically disjoint as pairs: if two such pairs share one summand, the other summand is forced.

Let \mathcal{S}_c be the family of all control sets obtained as follows. Choose one of the $T_m + 1$ pairs to omit; from each of the remaining T_m pairs choose one of its two summands. Thus every $S \in \mathcal{S}_c$ has cardinality T_m , and

$$(37) \quad |\mathcal{S}_c| \leq (T_m + 1)2^{T_m} = X_m^{60+o(1)}.$$

At stage $n = m + 10$, schedule one center for every triple

$$(\lambda, c, S), \quad \lambda \in \Lambda, \quad c \in C_m, \quad S \in \mathcal{S}_c.$$

Conditional on all earlier stages, choose distinct center positions uniformly in a slightly shortened version of W_n , shortened only enough to ensure that $y - \omega \in W_n$ for every $\omega \in \Omega$. For a scheduled center y , declare

$$(38) \quad \{y - \omega : \omega \in \Omega\}$$

to be its cluster of actual canaries, delete all pre-existing representations of these canaries, and add every restoration element in (26).

The number of centers is small enough for Lemma 6.2. Indeed,

$$(39) \quad K_n \leq |\Lambda| \sum_{c \in C_m} |\mathcal{S}_c|$$

$$(40) \quad \ll_{\mathcal{P}} X_m^{2+60+o(1)} = X_n^{62/1024+o(1)} < X_n^{1/8}.$$

Although there are exponentially many control sets, their union uses only the summands in the selected pairs. Therefore the union \mathcal{S}_n of all controls used at stage n satisfies

$$(41) \quad |\mathcal{S}_n| \leq \sum_{c \in C_m} 2(T_m + 1) \ll X_m^{2+o(1)} = X_n^{1/512+o(1)},$$

and every such control lies in the interval required by (32). This verifies the hypotheses of the clean-supply lemma at the next stages and closes the induction.

The same count also controls density. There are at most $|\Omega|K_n = O_{\mathcal{P}}(X_n^{1/8})$ canaries at stage n , and at most

$$(42) \quad |\Omega| T_m K_n \ll_{\mathcal{P}} X_n^{62/1024+o(1)} = O_{\mathcal{P}}(X_n^{1/4+o(1)})$$

restoration elements. The density proof of Lemma 5.1 applies unchanged, so the final clustered set has density zero.

6.5. Separating the clusters and making their representations exact. Before using the clusters, we verify three facts: distinct labels give distinct canaries, the canaries do not accidentally enter the set, and their only pair representations are the deliberately restored ones.

Lemma 6.3 (separation of clusters). *Almost surely, for all sufficiently large stages n , the map*

$$Y_n \times \Omega \longrightarrow \mathbb{Z}, \quad (y, \omega) \longmapsto y - \omega$$

is injective. Moreover, almost surely, simultaneously for every $e \in \mathbb{Z} \setminus (\Omega - \Omega)$, only finitely many pairs of canaries b, b' from the entire construction satisfy $b - b' = e$.

Proof. A collision between two different centers at stage n would give

$$y - y' = \omega - \omega' \in (\Omega - \Omega) \setminus \{0\}.$$

By (28), its probability is

$$O_{\Omega} \left(\frac{K_n^2}{|W_n|} \right) = O_{\Omega}(X_n^{-7/4}),$$

which is summable. Within one center, different shifts plainly give different integers. Canaries from sufficiently separated stages lie in disjoint windows, so only finitely many early cross-stage collisions remain.

Now fix $e \notin \Omega - \Omega$. Two canaries belonging to the same center cannot differ by e . For canaries from distinct centers at one stage, the equality

$$(y - \omega) - (y' - \omega') = e$$

forces $y - y'$ to be one of the finitely many numbers $e + \omega - \omega'$. Again (28) gives a summable bound $O_{\Omega, e}(X_n^{-7/4})$. Cross-stage pairs are eventually impossible by the separation of the windows. Thus the assertion holds almost surely for this fixed e . Since there are only countably many integers e , we may intersect these probability-one events and obtain the simultaneous statement. \square

Lemma 6.4 (exact canary representations). *Almost surely, every sufficiently large canary $y - \omega$ lies outside the final set A and has exactly the T_m representations*

$$(43) \quad y - \omega = s + (y - \omega - s), \quad s \in S,$$

attached to its own scheduled control set S .

Proof. At stage $n = m + 10$, every pre-existing representation of a new canary has a fresh stage- n summand; the construction deletes such fresh summands and thereby destroys every pre-existing representation without removing the old controls. It then inserts the elements in (26), so all representations in (43) are present. Distinct controls give distinct restoration elements, and Lemma 6.3 prevents two labels from naming the same canary. It remains to exclude unintended representations created by restoration.

Two stage- n restoration elements cannot sum to a stage- n canary. Each restoration element is at least $X_{n+1}/6 - o(X_{n+1})$, whereas every canary is at most $X_{n+1}/4 + O_{\Omega}(1)$. Restorations from later stages are larger than the canary. Thus an unintended representation must use exactly one new restoration element, say $y' - \omega_0 - s$, and one element a already available when stage n begins or freshly sampled at stage n .

If $y' = y$, then

$$(y - \omega_0 - s) + a = y - \omega \implies a = s + \omega_0 - \omega.$$

For $\omega_0 = \omega$ this is the intended representation. For $\omega_0 \neq \omega$, the difference $\omega_0 - \omega$ belongs to Δ , and the chosen control s is Δ -clean. Hence $a \notin A$.

It remains to rule out an interaction between different centers. Write $A''(n-1)$ for the set present after all stages before n , including old restorations. If a restoration from one center and an element $a \in A_n \cup A''(n-1)$ represent a canary from another center, then the two centers have a difference in

$$(44) \quad \mathcal{F}_n := ((A_n \cup A''(n-1)) - S_n) + \Omega - \Omega.$$

The standard size estimates from the Larsen–Larsen construction give

$$|A_n| \ll X_n^{1+o(1)}, \quad |A''(n-1)| \ll X_n^{1/2+o(1)}.$$

Together with (41), this yields

$$|\mathcal{F}_n| \ll X_n^{1+1/512+o(1)}.$$

Condition on the past and on the realized Bernoulli sample A_n . Then \mathcal{F}_n is fixed, while the center set Y_n remains uniformly random. By (28), the probability of any off-diagonal accident at stage n is at most

$$(45) \quad O_{\mathcal{P}} \left(\frac{X_n^{1/4} X_n^{1+1/512+o(1)}}{X_n^2} \right) = X_n^{-3/4+1/512+o(1)}.$$

This is summable. It covers old as well as fresh partner summands.

We next show that the canaries themselves are absent from A . A canary can enter the final set only by being selected in the Bernoulli sample or by coinciding with a restoration element. The first event has probability

$$\ll_{\mathcal{P}} K_n |\Omega| X_n^{-1} \sqrt{\log X_n} \ll_{\mathcal{P}} X_n^{-7/8+o(1)}.$$

A same-center coincidence would force a large control s to lie in the fixed set $\Omega - \Omega$, and is therefore impossible in all large stages. A coincidence involving different centers forces their difference to lie in

$$\mathcal{G}_n := -S_n + \Omega - \Omega, \quad |\mathcal{G}_n| \ll X_n^{1/512+o(1)}.$$

Its probability is

$$\ll_{\mathcal{P}} \frac{X_n^{1/4} X_n^{1/512+o(1)}}{X_n^2} = X_n^{-7/4+1/512+o(1)},$$

also summable. Borel–Cantelli now proves both the absence of large canaries from A and the exactness of (43). \square

6.6. Why every shifted subbasis has a deletable element. All probability-one conclusions used above hold simultaneously. Fix one realization with these properties. We can now prove the structural assertion in Proposition 2.2. Fix one package $(U, V) \in \mathcal{P}$, a set $D \subseteq A$, and a finite forbidden set $F_0 \subseteq D$, and suppose that $\Phi_{U,V}(D)$ is cofinite.

First, every sufficiently large robust integer $c \in C$ has at least two representations in $D + D$. Indeed, suppose that $r_D(c) \leq 1$. Among the $T_m + 1$ selected disjoint pairs for c , omit the possible pair contained in D ; if no such pair exists, omit any one pair. From every remaining pair choose a summand outside D . This gives a scheduled control set $S \subseteq A \setminus D$. Let y be the center assigned to (U, V, c, S) . By Lemma 6.4, for every $u \in U$ the canary $y - u$ lies outside D and every one of its pair representations uses an element of S . Consequently

$$y - u \notin D \cup (D + D).$$

Likewise, for every $v \in V$, the canary $y - v$ lies outside D . Hence $y \notin \Phi_{U,V}(D)$, contradicting the assumed cofiniteness.

The cofiniteness of $\Phi_{U,V}(D)$ also implies that D is infinite: a finite D would make all three sets in (1) finite. Choose

$$(46) \quad d \in D \setminus (F_0 \cup (\Omega - \Omega))$$

and put $D' := D \setminus \{d\}$.

Every sufficiently large robust integer remains in $D' + D'$. It had at least two D -representations, and at most one distinct unordered representation can use the fixed summand d .

We next show that the shifted cover survives. A witness for $x \in \Phi_{U,V}(D)$ can be lost only in one of two ways.

- (1) A singleton witness is deleted: $x - \theta = d$ for some $\theta \in U \cup V$. This accounts for only finitely many x .
- (2) A two-summand witness has the form $x - u = d + d_1$ with $u \in U$ and $d_1 \in D$.

For all sufficiently large x , the integer $x - u$ belongs to $B \cup C$ by (35). If $x - u \in C$, a second representation survives the deletion of d . Otherwise $x - u \in B$ is a canary. A fixed summand occurs in only finitely many intentional canary representations: it can be an old control only while the control interval still contains it, and it can be a restoration element only while the rapidly increasing canary window still contains it. Thus only finitely many canary witnesses use d . We have proved that

$$(47) \quad \Phi_{U,V}(D') \text{ is cofinite.}$$

It remains to prove that D' is an additive basis of order 3. Since D' is infinite, choose

$$e \in D' \setminus (\Omega - \Omega).$$

Lemma 6.3 says that only finitely many pairs of canaries differ by e . Therefore, for every sufficiently large $b \in B$, the integer $b - e$ is not in B . Since $B \cup C$ is cofinite, it follows that $b - e \in C$ for every sufficiently large $b \in B$. We already know that large members of C lie in $D' + D'$, so

$$b = e + d_1 + d_2 \in D' + D' + D'.$$

Large members of C themselves use only two summands. Since $B \cup C$ is cofinite, every sufficiently large integer is a sum of at most three elements of D' . This proves the deletion assertion in part (5) of Proposition 2.2.

6.7. Representation count and finite avoidance. A canary scheduled from a target in I_m is created at stage $n = m + 10$ and has exactly $T_m = 60 \cdot 2^m$ representations. Since such a canary is below $X_{m+11} = 2^{2^{m+11}}$, one has, for all sufficiently large m ,

$$(48) \quad T_m \geq \frac{15}{512 \log 2} \log b.$$

Every robust integer in C_m retains at least T_m clean raw representations. Since $c < X_{m+1}$ for $c \in C_m$,

$$T_m \geq \frac{30}{\log 2} \log c,$$

so the smaller constant in (48) also applies to robust integers. By (35), all sufficiently large integers are either canaries or robust. Thus, before imposing P_0 , the set satisfies

$$r_A(n) \geq \eta_3^* \log n, \quad \eta_3^* := \frac{15}{512 \log 2}.$$

Finally replace A by $\tilde{A} := A \setminus P_0$. For each fixed $p \in P_0$, at most one nondecreasing pair representation of a given integer uses p . Hence

$$r_{\tilde{A}}(n) \geq r_A(n) - |P_0| \geq \eta_3^* \log n - |P_0|.$$

With

$$(49) \quad \eta_3 := \frac{1}{2} \eta_3^* = \frac{15}{1024 \log 2},$$

we obtain $r_{\tilde{A}}(n) \geq \eta_3 \log n$ for all sufficiently large n , and in particular $\tilde{A} + \tilde{A}$ is cofinite. Density zero is unchanged.

The deletion property is unchanged as well. After finitely many stages, every control, canary, and restoration element lies above $\max P_0$; all arguments above concern only this eventual structure. If $D \subseteq \tilde{A}$, the element chosen in (46) already belongs to D and therefore lies outside P_0 . Relabeling \tilde{A} as A completes the proof of Proposition 2.2, subject only to the clean-supply lemma, which we now prove.

7. PROOF OF THE CLEAN REPRESENTATION SUPPLY

This is the technical verification behind Lemma 6.2. Its purpose is narrow: among the many raw pair representations supplied by the Larsen–Larsen random set, only a bounded number are lost for three additional reasons: failure of Δ -cleanness inside the Bernoulli set, failure caused by a restoration element, and deletion caused by a canary. Since the initial *supply* (the number of raw pair representations available to m) is at least $159 \log m$, bounded losses do not affect any fixed lower constant below 159.

Let

$$\mathcal{A}_n := A^{(0)}(n-1) \cup A_n$$

be the union of all raw Bernoulli samples available by stage n . In this section, a *raw representation* always means a pair $m = a + a'$ with $a, a' \in \mathcal{A}_n$; restoration elements are never counted as supply. The schedule, including the control union S_n , is fixed by the past. The fresh Bernoulli sample A_n and the random center set Y_n are independent conditional on that past. We shall condition on one or the other according to which makes the relevant forbidden set deterministic.

7.1. The initial raw supply. Larsen and Larsen’s lower-tail estimate [4, Lemma 2], with the Bernoulli constant chosen sufficiently large, gives almost surely, uniformly for every large n and every $m \in I_n$, more than $160 \log m$ raw representations whose smaller summand is at least $X_n/4$. Moving the endpoint inward by the fixed amount M_Δ removes only $O_\Delta(1)$ possibilities.

Stage-safety also has negligible cost. A summand can fail to be stage-safe only when it lies within M_Δ of one of the stage boundaries X_1, \dots, X_{n+1} . There are $O_\Delta(n)$ such integers, and $n = o(\log m)$. Consequently, almost surely for all sufficiently large n , every $m \in I_n$ has at least

$$(50) \quad 159 \log m$$

buffered, stage-safe raw representations. The matching upper-tail estimate also gives the uniform bound

$$(51) \quad \#\{\text{buffered raw representations of } m\} \ll_K \log X_n.$$

There is one useful monotonicity observation. A buffered summand used for a target in I_n is greater than $X_n/4$. Every raw element deleted at an earlier stage $j < n$ is smaller than $X_{j+1}/4 \leq X_n/4$. Hence no buffered old summand has already been deleted. We therefore need only account for cleanness failures and deletions occurring at stage n itself.

7.2. Raw Bernoulli points at forbidden shifts. If $\Delta = \emptyset$, this subsection and the following restoration-cleanness subsection are vacuous. We therefore assume in both that $\Delta \neq \emptyset$.

Fix $m \in I_n$ and $\delta \in \Delta$. Consider a raw representation $m = a + (m - a)$. It fails cleanness (we then call the representation *dirty*) on the first summand when the three points

$$a, \quad m - a, \quad a + \delta$$

all belong to \mathcal{A}_n . Apart from the degenerate equality $a + \delta = m - a$ and the diagonal equality $a = m - a$, each of which determines at most one a , these are three distinct Bernoulli points. From

(12), uniformly in $m \in I_n$,

$$(52) \quad \sum_a p(a)p(m-a)p(a+\delta) \ll_{K,\Delta} X_n^{-1/2+o(1)}.$$

To turn this expectation into a uniform bound, join two candidate witnesses when their sets of three Bernoulli points overlap. For fixed Δ , the resulting dependency graph has bounded degree D_Δ : every overlap is one of finitely many linear equations relating the two values of a . Thus H distinct dirty representations contain an independent subfamily of size at least

$$L = \lceil H/(D_\Delta + 1) \rceil.$$

For such a subfamily the Bernoulli events are independent, and the usual factorial-moment bound based on (52) gives

$$\mathbb{P}\{m \text{ has at least } H \text{ nondegenerate dirty representations}\} \ll_{H,K,\Delta} X_n^{-L/2+o(1)}.$$

Choose H so that $L \geq 5$. There are at most $X_n^{2+o(1)}$ targets in I_n , so the probability that any target violates this bound is $X_n^{-1/2+o(1)}$, which is summable because the X_n grow doubly exponentially. The same argument applies to the second summand and to every $\delta \in \Delta$. We conclude that, almost surely, only

$$(53) \quad O_\Delta(1)$$

raw representations of any large $m \in I_n$ fail cleanness because of another raw Bernoulli point.

7.3. Restoration elements at forbidden shifts. Let $t = s + \delta$ be a shifted companion of a buffered summand in a representation of $m \in I_n$. The buffer gives

$$X_n/4 < t < X_{n+1}.$$

Restorations from stages below n lie below $X_n/4$, while restorations from stages above n lie above X_{n+1} for all large n . Thus only a stage- n restoration can make t belong to A .

Let \mathcal{T}_m be the set of all shifted companions of both summands in all buffered, stage-safe raw representations of m . By (51),

$$(54) \quad |\mathcal{T}_m| = X_n^{o(1)}.$$

A stage- n restoration has the form $b - s$, where $b \in B_n^\Omega$ is a canary (the *top* of the pair $b = s + (b - s)$) and $s \in S_n$ is an old control. We prove first that one fixed top cannot be responsible for many dirty representations.

Put $C_\Delta := 2|\Delta|$. One fixed shifted target determines at most C_Δ non-diagonal raw representations of m : one chooses a shift and which side of the pair it came from, and the pair is then forced. Therefore, if one canary top dirties more than C_Δ representations, it must do so through two distinct controls $s_1 \neq s_2$.

Fix such a top b , two controls, two shifts, and the two side choices, and set

$$z_i := b - s_i - \delta_i, \quad w_i := m - z_i \quad (i = 1, 2).$$

The z_i are fresh-stage points, each with probability $X_n^{-1+o(1)}$. Depending on whether the complementary points w_i are fresh or old, the complete union bounds are as follows. The column ‘‘enumeration’’ lists all powers of X_n before the Bernoulli cost is applied.

Case	Enumeration	Bernoulli cost	Stage bound
both w_i fresh	$X_n^{1/8}$ tops, X_n^2 targets, and $X_n^{2/512+o(1)}$ control pairs	$X_n^{-3+o(1)}$	$X_n^{-7/8+2/512+o(1)}$
one fresh, one old	$X_n^{1/8}$ tops, $X_n^{1/2+o(1)}$ choices of the old point, and $X_n^{2/512+o(1)}$ control pairs	$X_n^{-5/2+o(1)}$	$X_n^{-15/8+2/512+o(1)}$
both w_i old	$X_n^{1/8}$ tops, $X_n^{1/2+o(1)}$ choices of one old point, and $X_n^{2/512+o(1)}$ control pairs	$X_n^{-2+o(1)}$	$X_n^{-11/8+2/512+o(1)}$

Two points explain the table. In the mixed case, choosing the old point determines m and hence the remaining complementary point. In the old–old case, the equation $z_1 + w_1 = z_2 + w_2$ makes w_2 a function of w_1 , so there is still only one old degree of freedom. Coincident unordered pairs do not count as two representations, and a diagonal pair $m = 2z_i$ pins m and gives a smaller bound. Every exponent in the last column is negative, and hence summable along the doubly exponential sequence X_n . It follows that, almost surely for all large n , each canary top dirties at most C_Δ non-diagonal raw representations of any target m .

We now condition in the opposite order: expose the Bernoulli sample A_n and leave the center set Y_n random. If m had at least

$$(55) \quad H_0 := C_\Delta(2|\Omega| + 1) + 1$$

restoration-dirty representations, then, after discarding the at most one diagonal representation, the preceding single-top bound would force at least $2|\Omega| + 1$ distinct canary tops in $\mathcal{T}_m + S_n$. By (54) and (32),

$$|\mathcal{T}_m + S_n| \ll X_n^{1/512+o(1)}.$$

At least three distinct centers are needed to supply $2|\Omega| + 1$ tops. Applying (27), the probability for one fixed m is therefore

$$\ll X_n^{(-15/8+1/512+o(1)) \cdot 3} = X_n^{-45/8+3/512+o(1)}.$$

After summing over all $m \in I_n$, the stage probability is

$$(56) \quad X_n^{-29/8+3/512+o(1)},$$

which is summable. Combining this event with the single-top event by a union bound, rather than conditioning one upon the other, proves that only $O_{\Delta, \Omega}(1)$ representations of any large m are dirtied by restoration elements.

7.4. Representations destroyed by the deletion step. Let

$$E_n := A''(n-1) \setminus A^{(0)}(n-1)$$

denote the old restoration elements. The standard high-probability size estimates in the Larsen–Larsen construction give

$$(57) \quad |\mathcal{A}_n| \ll X_n^{1+o(1)}, \quad |E_n| \ll X_n^{1/8+o(1)}, \quad |\mathcal{A}_n \cap (m - \mathcal{A}_n)| \ll \log X_n.$$

At stage n , only fresh raw elements are deleted. Such an element y is deleted only when it participates in a pre-existing representation $y + z = b$ of a new canary b .

First suppose the partner z is an old restoration element. Conditional on the Bernoulli sample, $A_n + E_n$ is fixed and has size at most $X_n^{9/8+o(1)}$. The first cluster estimate gives

$$(58) \quad \mathbb{P}(B_n^\Omega \cap (A_n + E_n) \neq \emptyset) \ll X_n^{1/8+9/8-2+o(1)} = X_n^{-3/4+o(1)}.$$

Thus this case occurs only at finitely many stages.

We may now assume that the partner z is raw. For a fixed noncanary $m \in I_n$, define the *dangerous top set*

$$(59) \quad \mathcal{D}_m := (\mathcal{A}_n \cap (m - \mathcal{A}_n)) + \mathcal{A}_n.$$

If deleting y destroys a raw representation $m = x + y$ through the canary witness $y + z = b$, then $b \in \mathcal{D}_m$. By (57),

$$|\mathcal{D}_m| \ll X_n^{1+o(1)}.$$

Choose $H = 2|\Omega| + 1$. If B_n^Ω contains H points of \mathcal{D}_m , at least three centers are involved, and (27) gives probability

$$X_n^{-21/8+o(1)}$$

for this fixed m . Summing over $m \in I_n$ gives the summable stage bound

$$(60) \quad X_n^{-5/8+o(1)}.$$

Therefore every large noncanary target is affected by only $O_\Omega(1)$ dangerous canaries.

For one fixed dangerous canary $b \neq m$, Larsen and Larsen's anti-clustering proposition [4, Proposition 5] applies to the system

$$x + y = m, \quad y + z = b, \quad x, y, z \in \mathcal{A}_n.$$

It gives at most 16 pairwise distinct solution triples. Hence one canary destroys at most 16 oriented raw representations of m , and the total number destroyed is $O_\Omega(1)$. Passing to nondecreasing pairs changes only the harmless absolute constant.

7.5. Completion of the clean-supply proof. The initial supply (50) is at least $159 \log m$. The Bernoulli-cleanness loss (53), the restoration loss controlled by (55), and the deletion loss from the previous subsection are each bounded independently of m and n . Thus the number of surviving buffered, stage-safe, Δ -clean raw representations is

$$159 \log m - O_{\Delta, \Omega}(1).$$

For every fixed $L_0 < 159$, this is at least $L_0 \log m$ once m is large. This proves Lemma 6.2 and completes the proof of the clustered input.

8. CONCLUDING PERSPECTIVE

The proof has two distinct parts. The order-2 random construction produces a sparse set with two complementary kinds of large integers: robust integers, which have many independent representations, and canaries, whose representations are numerous but deliberately controlled. This dichotomy is what makes one-element deletion possible in every subbasis.

For $k \geq 4$, no further randomness is needed. The finite filler gadget reserves two summands for the order-2 core, supplies the remaining residue classes, and uses density zero to ensure that a subbasis cannot discard all short fillers in any residue. For $k = 3$, the same idea has only one spare summand and therefore cannot be implemented by residues alone. Clusters solve precisely that problem: one random center controls all finitely many shifts created by the parity argument. With those two reductions in place, Propositions 3.2 and 4.1 prove Theorem 1.1 for every $k \geq 3$ and every prescribed constant $C > 0$.

APPENDIX A. GUIDE TO THE LEAN FORMALIZATION

The accompanying Lean 4 project formalizes the same at-most- k statement used in this paper. Its target is

$$\forall k \geq 3 \forall C > 0 \exists E \subseteq \mathbb{N} : \begin{cases} E \text{ is a basis of order at most } k, \\ R_{E,k}(n) \geq C \log n \text{ eventually,} \\ E \text{ contains no minimal order-}k \text{ subbasis.} \end{cases}$$

The source-level correspondence is as follows.

Defs.lean. Definitions of at-most- k representations, bases, logarithmic lower bounds, minimal subbases, stages, controls, and canaries; the file also defines `Erdos870Target`.

Order2Input.lean. The Larsen–Larsen order-two construction and the theorem `larsen_larsen_order2_input`, corresponding to Proposition 2.1.

ClusteredInput.lean. The finite-shift construction, culminating in `finite_shift_clustered_input_uniform`, corresponding to Proposition 2.2.

Probability files. The files `ProbabilityTools.lean` and `ClaudeProbCore.lean` contain the concentration infrastructure. The files `LLProp5AntiClustering.lean` and `LarsenLarsenL7BC.lean` contain the anti-clustering and Borel–Cantelli arguments.

Filler.lean. The deterministic deductions `kge4_from_order2` and `k3_from_clustered`, matching Sections 3 and 4.

MainTheorem.lean. The final theorem `main_theorem:Erdos870Target`; this file also contains `#printaxiomsmain_theorem`.

The project pins both the Lean toolchain and the Mathlib revision. The repository’s continuous integration builds `Erdos870` and checks the axiom dependencies of `main_theorem`; the axiom report is

[propext, Classical.choice, Quot.sound],

with no project-specific axiom and no admitted theorem. The formalization is not used as a substitute for any step in the paper; rather, it gives a machine-checkable audit of the definitions, the two deterministic reductions, and the probabilistic construction interfaces.

ACKNOWLEDGEMENTS

The natural-language proof was developed with the assistance of GPT-5.4-Pro, and subsequently GPT-5.5-Pro. The Lean 4 formalization was carried out with the assistance of GPT-5.5-Pro, Codex (GPT-5.5-xhigh), and Claude Code (Opus 4.8, maximum thinking). The author thanks [Fulcrum](#) for the compute resources that made this large project possible. The author is responsible for the final mathematical claims and exposition.

REFERENCES

- [1] T. F. Bloom, *Erdős Problem #870*, Erdős Problems website, <https://www.erdosproblems.com/870>.
- [2] P. Erdős and M. B. Nathanson, *Systems of distinct representatives and minimal bases in additive number theory*, in *Number Theory, Carbonale 1979*, Lecture Notes in Mathematics **751**, Springer, Berlin, 1979, 89–107.
- [3] P. Erdős and M. B. Nathanson, *Additive bases with many representations*, *Acta Arith.* **52** (1989), no. 4, 399–406.
- [4] D. Larsen and M. Larsen, *Robust additive bases without minimal subbases*, arXiv:2601.18507, 2026.