SMALL MAXIMAL INTEGER SIDON SETS BY FINITE PHASE COLORINGS

WILL HOWE, GRANT HUFFMAN, JENSEN KOHLMEYER, AND LIAM KRUER

Abstract. We prove that every sufficiently large integer interval $\{1, \dots, N\}$ contains an inclusion-maximal Sidon set of cardinality $O(N^{1/3})$. The construction combines a quadratic Sidon seed with a periodically labeled family of anchors. Finite-dimensional bounds for affine sign patterns produce a fixed alphabet of integer translates and a periodic word that covers every relevant local window. A pair of separated alphabets supplies Sidon, sum-free tags for every prime, so the prime-selection step uses only Bertrand's postulate. The resulting seed blocks every point outside two intervals of length $O(p^2)$ and one residue class, where $p \asymp N^{1/3}$. A maximal completion adds only $O(p)$ elements. A Lean proof verifies the upper bound, including the uniform pattern counts and the placement argument, against the formal statement of Erdős problem 156.

hexagon:2611.00767v1

1. Introduction

A set of integers A is a Sidon set if $a + b = c + d$, with $a, b, c, d \in A$, implies $\{a, b\} = \{c, d\}$ as multisets. In particular, diagonal sums $2a$ are included in the definition. A Sidon subset of $[1, N] \cap \mathbb{Z}$ is maximal if adjoining any further point of that interval destroys the Sidon property. Write

$$s(N) = \min\{|A| : A \subseteq [1, N] \cap \mathbb{Z} \text{ is a maximal Sidon set}\}.$$

Maximality is with respect to inclusion.

Erdős, Sárközy, and Sós studied this problem in their work on sumsets of Sidon sets [1]. Ruzsa [2] proved the existence of maximal Sidon sets of size $O((N \log N)^{1/3})$. We establish the following upper bound by a construction whose uniformity statements are included in a Lean proof.

Theorem 1.1. *There are absolute constants $C > 0$ and $N_0 \in \mathbb{N}$ such that, for every integer $N \geq N_0$, there is a maximal Sidon set $A \subseteq [1, N] \cap \mathbb{Z}$ satisfying $|A| \leq CN^{1/3}$. Equivalently, $s(N) = O(N^{1/3})$.*

The exponent is forced by a counting argument.

Proposition 1.2. *Every maximal Sidon set $A \subseteq [1, N] \cap \mathbb{Z}$ satisfies $|A| \geq (N/3)^{1/3}$.*

Proof. The case $N = 0$ is immediate. Otherwise put $m = |A| \geq 1$. For every $x \in [1, N] \setminus A$, a nontrivial equality of pair sums in $A \cup \{x\}$ must take one of the forms

$$x + a = b + c \quad \text{or} \quad 2x = a + b, \qquad a, b, c \in A.$$

Indeed, equalities using x once on each side cancel to a trivial equality. Consequently the interval is contained in the union of A, $A + A - A$, and the set of integral midpoints of pairs from A. Thus $N \leq m + m^3 + m^2 \leq 3m^3$. □

Together, Theorem 1.1 and Proposition 1.2 give $s(N) \asymp N^{1/3}$. The companion certificate proves the upper bound.

Date: November 2, 2026.

2020 *Mathematics Subject Classification.* 11B13, 11B75, 05D40.

Key words and phrases. Maximal Sidon set, additive combinatorics, finite phase coloring, formal verification.

Organization and notation. For an integer z and a positive integer q, let $\langle z\rangle_q \in \{0,\dots,q-1\}$ be its least nonnegative residue. For a real number u, write $\{u\} = u - \lfloor u\rfloor$. All intervals below are intersected with $\mathbb{Z}$ when used as sets of marks.

We first work in $[0,N)$ and translate by 1 at the end. A quadratic seed C_p supplies controlled differences. A finite alphabet of translates covers all nonmultiples of p in a long interior interval. Two versions of this alphabet produce tags that preserve Sidonicity. A periodic choice of labels then makes a family of anchors cover the interior of the host interval. Only two short boundary intervals and the multiples of p remain available to a maximal completion.

2. Finite pattern selection

The construction uses finite families of patterns even though the arithmetic parameters range over infinitely many primes and targets. The following deliberately nonoptimal estimate agrees with the bound used in the certificate.

Lemma 2.1 (Affine sign patterns)**.** *Let $f_1,\dots,f_q$ be affine real functions on $\mathbb{R}^d$. The number of joint sign patterns, with each sign in $\{-,0,+\}$, is at most*

$$\big((d+2)(q+1)^{d+1}\big)^2. \tag{2.1}$$

The same bound applies to any finite-valued label determined by these signs.

Proof. Homogenize the affine functions on $\mathbb{R}^{d+1}$. The family of their positive supports has VC dimension at most $d+1$. To see the dimension bound, suppose a collection of the corresponding linear functionals is shattered. A nonzero linear dependence among them is impossible. Choose a realizing vector on which precisely the functionals with positive coefficients are positive, changing the sign of the dependence first if necessary. Evaluating the dependence then gives a strictly positive sum.

The Sauer–Shelah bound [3, 4] gives at most

$$\sum_{j=0}^{d+1}\binom{q}{j} \le (d+2)(q+1)^{d+1}$$

positive supports. Apply the same estimate to negative supports. The pair consisting of the positive and negative supports also determines all zero signs. □

Lemma 2.2 (Fractional-part tests)**.** *Suppose a label is determined by t comparisons*

$$\{f_i(v)\} < g_i(v), \qquad 1 \le i \le t,$$

where f_i, g_i are affine functions of $v \in \mathbb{R}^d$, and $|f_i(v)| \le B$ on the parameter set under consideration, with B a nonnegative integer. The number of labels is at most

$$\big((d+2)(2t(2B+1)+1)^{d+1}\big)^2.$$

No bound on the values of g_i is required.

Proof. For each integer $k \in [-B,B]$, record the signs of $f_i - k$ and $f_i - g_i - k$. The first collection determines $\lfloor f_i\rfloor$, including the case that f_i is an integer. The second determines whether $f_i - \lfloor f_i\rfloor < g_i$. There are $2t(2B+1)$ affine functions. Apply Lemma 2.1. □

Lemma 2.3 (A common hitting word)**.** *Fix $b \ge 2$. For each m, let $\mathcal{G}_m$ be a family of matrices*

$$(G_0,\dots,G_{m-1}), \qquad \varnothing \ne G_i \subseteq \{1,\dots,b\}.$$

If $|\mathcal{G}_m| \le Am^B$ for fixed constants A,B, then for all sufficiently large m there is a word $w \in \{1,\dots,b\}^m$ that meets every matrix: for each $G \in \mathcal{G}_m$, some i satisfies $w_i \in G_i$. One may also require this property simultaneously for every cyclic rotation of w.

Proof. Choose the letters independently and uniformly. A fixed matrix is missed with probability at most $(1-1/b)^m$. The union bound is less than 1 once $Am^B(1-1/b)^m < 1$. To include all rotations, use the bound $Am^{B+1}(1-1/b)^m$, which also tends to zero. □

3. The quadratic seed and exact differences

Let p be an odd prime and put

$$Q_p = 2p^2, \qquad c_p(t) = t + 2p\langle t^2\rangle_p, \qquad C_p = \{c_p(t) : 0 \le t < p\}, \qquad D_p = C_p - C_p. \tag{3.1}$$

Lemma 3.1. *The set C_p is Sidon, has p elements, and is contained in $[0, Q_p)$.*

Proof. If $c_p(i) + c_p(j) = c_p(k) + c_p(l)$, reduction modulo $2p$ gives $i + j = k + l$, since both sums lie in $[0, 2p)$. Subtracting this equality and dividing by $2p$ gives equality of the sums of the quadratic residues. In $\mathbb{F}_p$ we therefore have $i+j = k+l$ and $i^2+j^2 = k^2+l^2$. As p is odd, these imply $ij = kl$. The two pairs are roots of the same monic quadratic, so they agree as unordered pairs in $\mathbb{F}_p$, and hence as integer pairs in $[0, p)$. Distinct parameters give distinct residues modulo p, proving the cardinality statement. Finally, $0 \le c_p(t) \le (p-1) + 2p(p-1) < 2p^2$. □

Lemma 3.2 (Exact lift with two carries)**.** *Let $1 \le d < p$ and $1 \le H < p$. Define*

$$t = \langle (H + d^2)(2d)^{-1}\rangle_p, \quad \sigma = \mathbf{1}_{t<d}, \quad u = t - d + p\sigma, \quad \varepsilon = \mathbf{1}_{\langle t^2\rangle_p < H}.$$

Then $0 \le u < p$ and

$$c_p(t) - c_p(u) = d - p\sigma + 2pH - \varepsilon Q_p. \tag{3.2}$$

Proof. The definition of σ gives $t - u = d - p\sigma$ and $0 \le u < p$. In $\mathbb{F}_p$, $t^2 - u^2 = 2dt - d^2 = H$. The difference of the two least residues lies strictly between $-p$ and p. It is therefore H or $H - p$. The latter alternative occurs exactly when $\langle t^2\rangle_p < H$. Substitution in (3.1) proves the identity. □

In particular, if $\tau \in \{0, 1\}$ and $k \ge 0$, then

$$d + p\tau + 2p(H + k) = c_p(t) - c_p(u) + p(2k + \tau + \sigma) + \varepsilon Q_p. \tag{3.3}$$

Thus the six pairs $(s, \varepsilon) \in \{0, 1, 2\} \times \{0, 1\}$ contain a suitable choice in every row. The identity is in $\mathbb{Z}$.

4. A fixed covering alphabet

Definition 4.1. A covering alphabet consists of a finite label set L of size $b \ge 2$, integers $\delta_\ell \in [0, M]$, and $\varepsilon_\ell \in \{0, 1\}$, such that

(1) equality of two δ-pair sums determines the unordered pair of labels,
(2) for every odd prime p,

$$I_p(M) \setminus p\mathbb{Z} \subseteq D_p + \{p\delta_\ell + \varepsilon_\ell Q_p : \ell \in L\}, \quad I_p(M) = [2p(M+2), 2p(p-1)). \tag{4.1}$$

The interval in (4.1) may be empty.

Proposition 4.2 (Separated covering alphabets)**.** *Covering alphabets exist. Moreover, for every prescribed $G \ge 0$, one can choose a covering alphabet for which*

$$|\delta_i + \delta_j - \delta_k - \delta_l| > G \quad \text{whenever } \{i, j\} \ne \{k, l\} \text{ as multisets of labels.} \tag{4.2}$$

Proof. Fix a positive integer λ with $2\lambda - 4 > G$. For $0 \le i < m$, set $a_i = i + 2mi^2$ and $k_i = \lambda a_i$. The sequence (a_i) is Sidon. A pair-sum equality modulo $2m$ first determines the integer sum of the indices, and then the sum of their squares determines the unordered pair. Any selection of pairs $(s_i, \varepsilon_i) \in \{0, 1, 2\} \times \{0, 1\}$ gives

$$\delta_i = 2k_i + s_i, \qquad \varepsilon_i \in \{0, 1\}, \qquad M = 6\lambda m^3 + 2. \tag{4.3}$$

Here $\delta_i \le M$ and $k_i \le M$. For distinct unordered index pairs, the a-pair sums differ by at least 1, while the difference of the two s-pair sums has absolute value at most 4. This proves (4.2) for every selection.

Write a target in $I_p(M) \setminus p\mathbb{Z}$ uniquely as $y = d + p\tau + 2pY$, where $1 \le d < p$ and $\tau \in \{0, 1\}$. The interval bounds give $M + 2 \le Y < p$. Hence $H_i = Y - k_i$ lies in $[1, p)$ for every row. Let $r = (2d)^{-1} \in \mathbb{F}_p$ and $u = r(Y + d^2) \in \mathbb{F}_p$. The successful pair in (3.3) is determined by the tests

$$\begin{aligned} \{\langle u\rangle_p/p - k_i\langle r\rangle_p/p\} &< d/p, \\ \{\langle u^2\rangle_p/p + k_i\langle -2ur\rangle_p/p + k_i^2\langle r^2\rangle_p/p\} &< Y/p - k_i/p. \end{aligned} \tag{4.4}$$

The first test determines σ_i, and the second determines ε_i. Put $s_i = \tau + \sigma_i$.

Treat the eight normalized quantities in (4.4) as independent parameters in $[0, 1]^8$. This enlarges the pattern family. Since $k_i = O_\lambda(m^3)$, the arguments inside the fractional parts are bounded in absolute value by $O_\lambda(m^6)$. There are $2m$ tests, so Lemma 2.2 gives $O_\lambda(m^{126})$ matrices of successful choices, including both values of τ. Every row is a singleton. The formal bound is $200(16(3\lambda + 1)^2)^{18}m^{126}$.

By Lemma 2.3, for sufficiently large m there is a fixed word of six choices that meets every matrix. Fix this word and use (4.3). For any p and y under consideration, a row at which the selected word agrees with its successful choice supplies (3.3). All choices and constants are fixed before p and y are given. □

5. Tags for every prime

We next arrange that arbitrary labels can be attached to a Sidon ruler without creating new equalities of pair sums. The modulus below need not be prime.

Proposition 5.1 (Two alphabets suffice)**.** *There exist two covering alphabets L_1, L_2 and an integer $K \ge 1$ with the following property. For every integer $p \ge 0$, at least one alphabet admits a shift $T \in \{0, \dots, K - 1\}$ such that the residues*

$$v_\ell = 4p + T + \delta_\ell + 2p\varepsilon_\ell \pmod K \tag{5.1}$$

determine unordered label pairs from their pair sums and are sum-free:

$$v_i + v_j = v_k + v_l \implies \{i, j\} = \{k, l\}, \qquad v_i + v_j \ne v_k.$$

Proof. Choose any covering alphabet L_1 with bound M_1. By Proposition 4.2, choose L_2 with separation greater than $4M_1$ and bound M_2. Let b_1, b_2 be their sizes and put

$$K = 4(M_1 + M_2) + b_1^3 + b_2^3 + 1. \tag{5.2}$$

For either alphabet, first consider the unshifted residues $\delta_\ell + 2p\varepsilon_\ell$. A collision of distinct unordered label pairs gives

$$K \mid d + 2pe, \quad d = \delta_i + \delta_j - \delta_k - \delta_l, \quad e = \varepsilon_i + \varepsilon_j - \varepsilon_k - \varepsilon_l.$$

We have $d \ne 0$, $|d| \le 2M$, and $|e| \le 2$. Since $K > 2M$, such a collision must have $e \ne 0$. Suppose both alphabets have a collision, with differences (d_1, e_1) and (d_2, e_2). Eliminating p gives

$$K \mid e_2d_1 - e_1d_2, \qquad |e_2d_1 - e_1d_2| \le 4(M_1 + M_2) < K.$$

Thus $e_2d_1 = e_1d_2$. As $1 \le |e_1| \le 2$, we get $|d_2| \le |e_1d_2| = |e_2d_1| \le 4M_1$, contrary to the separation of L_2. At least one unshifted alphabet therefore has unique label pair sums modulo K.

For any b residues u_ℓ modulo K, translation by c can violate sum-freeness only if $c = u_k - u_i - u_j$ modulo K for some triple of labels. There are at most $b^3 < K$ forbidden shifts. Choose an allowed c, and take T to be the representative of $c - 4p$ modulo K. Translation preserves pair-sum uniqueness and gives the desired tags. □

This argument permits all primes p. The construction uses only Bertrand's postulate for prime selection.

6. A Sidon union with arbitrary anchor labels

For an odd prime r, define the increasing ruler $a_r(h) = 2rh + \langle h^2\rangle_r$ for $h \geq 0$. For $0 \leq h < r$ it is Sidon, by the proof of Lemma 3.1 with the high and low components interchanged. The residue sums are less than $2r$, so an equality of ruler pair sums first determines those residue sums modulo $2r$, then the integer sum of the indices, and finally their unordered pair in $\mathbb{F}_r$. Also, for every $h \geq 0$,

$$r < a_r(h+1) - a_r(h) < 3r. \tag{6.1}$$

Choose an alphabet and shift T from Proposition 5.1. For arbitrary labels ℓ_h, put

$$\begin{gathered} z_h = Ka_r(h) + 4p + T + \delta_{\ell_h} + 2p\varepsilon_{\ell_h}, \qquad A_h = pz_h, \\ B_h = p(Ka_r(h) + 4p + T). \end{gathered} \tag{6.2}$$

Thus $A_h = B_h + p\delta_{\ell_h} + \varepsilon_{\ell_h} Q_p$.

Lemma 6.1. *If p, r are odd primes and $0 \leq H < r$, then $S = C_p \cup \{A_h : 0 \leq h \leq H\}$ is Sidon for every choice of the labels.*

Proof. An equality $z_i + z_j = z_k + z_l$ first determines the unordered label pair modulo K. The sums of the label offsets are consequently equal as integers. Subtract them and divide by K to obtain an equality of ruler pair sums. The ruler is Sidon on $[0, r)$, so $\{i, j\} = \{k, l\}$. Moreover, $z_i + z_j = z_k$ is impossible by sum-freeness modulo K.

Every seed mark is less than $2p^2$, while every anchor is a multiple of p and at least $4p^2$. Hence a sum of two seed marks cannot equal a sum involving an anchor. If $c_p(t) + A_i = c_p(u) + A_j$, reduction modulo p gives $t = u$ and then $A_i = A_j$. Finally, an equality $A_i + A_j = c_p(t) + A_k$ forces $t = 0$ modulo p, so $c_p(t) = 0$ and $z_i + z_j = z_k$, which is impossible. These cases exhaust the possible equalities. □

7. A periodic word for all local windows

Fix a covering alphabet L, its bounds b, M, and an integer $K \geq 1$. The baselines B_h in (6.2) do not depend on the labels. We show that a single periodic word works for all arithmetic data. In the following proposition the formula for $a_r(h)$ is used for every positive integer r.

Proposition 7.1 (Uniform periodic coverage)**.** *There are $R \geq 2$ and a word $w \in L^R$ such that the labels $\ell_h = w_{h \bmod R}$ have the following property. For every odd prime p, every $r \geq 1$, every $0 \leq h < r$, every $T \geq 0$, and every integer x with $p \nmid x$, if*

$$x - B_{h+j} \in I_p(M) \qquad (0 \leq j < R), \tag{7.1}$$

then $x \in A_{h+j} + D_p$ for at least one $j \in \{0, \ldots, R-1\}$.

7.1. **The carry vector.** For a window starting at h, write

$$e_j = \left\lfloor \frac{\langle h^2\rangle_r}{r} + \frac{2h}{r} j + \frac{j^2}{r} \right\rfloor. \tag{7.2}$$

Direct subtraction gives the exact identity

$$a_r(h+j) - a_r(h) = 2(r+h)j + j^2 - re_j. \tag{7.3}$$

Since $h < r$, the three coefficients in (7.2) belong to $[0,1] \times [0,2] \times [0,1]$. For $j < R$, the expression being rounded is between 0 and $(R+1)^2$. Its floor is determined by $O(R^3)$ affine integer-level tests in three variables. Lemma 2.1 therefore gives $O(R^{24})$ possible carry vectors. The explicit bound used in Lean is

$$\#\{(e_j)_{j<R}\} \leq 25 \cdot 10^8 R^{24}. \tag{7.4}$$

7.2. **The exact good-label criterion.** Set $A = 2K(r+h)$, $B = Kr$, and $t_j = Aj + Kj^2 - Be_j$. Then $B_{h+j} = B_h + pt_j$. Write $x - B_h = d + pm_0$, where $d = \langle x \rangle_p \in \{1, \dots, p-1\}$. For each row and label put

$$n_{j\ell} = m_0 - t_j - \delta_\ell, \qquad \sigma_{j\ell} = \langle n_{j\ell} \rangle_2, \qquad H_{j\ell} = \frac{n_{j\ell} + \sigma_{j\ell}}{2}. \tag{7.5}$$

These are integers. Under (7.1), $0 < H_{j\ell} < p$. The lower bound on $x - B_{h+j}$ gives $m_0 - t_j \geq 2(M+2)$ and its upper bound gives $m_0 - t_j \leq 2p - 3$. Now use $0 \leq \delta_\ell \leq M$ and $0 \leq \sigma_{j\ell} \leq 1$.

Let $T_{j\ell} = \langle (H_{j\ell} + d^2)(2d)^{-1} \rangle_p$. The label ℓ is good in row j, meaning

$$x \in B_{h+j} + p\delta_\ell + \varepsilon_\ell Q_p + D_p,$$

if and only if

$$(T_{j\ell} < d) \iff \sigma_{j\ell} = 1, \qquad (\langle T_{j\ell}^2 \rangle_p < H_{j\ell}) \iff \varepsilon_\ell = 1. \tag{7.6}$$

For sufficiency, apply Lemma 3.2 and (7.5). For necessity, a representation by two seed marks fixes their parameter difference modulo p. Their actual parameter difference is $d - p\sigma'$ with $\sigma' \in \{0, 1\}$. The equality of differences then gives $n_{j\ell} + \sigma' = 2(H' + p\varepsilon_\ell)$, where H' is the actual difference of the two quadratic residues. Thus parity gives $\sigma' = \sigma_{j\ell}$, and $H' \equiv H_{j\ell} \pmod{p}$. The quadratic equation fixes the leading seed parameter $T_{j\ell}$. Since $H' \in (-p, p)$, its two possible values are $H_{j\ell}$ and $H_{j\ell} - p$. These give precisely the choices in (7.6).

7.3. **Counting the matrices.** For a fixed carry vector, the pattern $\sigma_{j\ell}$ depends only on the three parity bits of m_0, A, B, because K and the alphabet are fixed. There are at most eight parity patterns. Let $\rho = (4d)^{-1} \in \mathbb{F}_p$ and introduce the five integer features

$$F_{j\ell} = (1, j, j^2, e_j, \delta_\ell - \sigma_{j\ell}). \tag{7.7}$$

In $\mathbb{F}_p$, the parameter $T_{j\ell}$ is the linear combination of these features with coefficients

$$\rho(m_0 + 2d^2), \qquad -\rho A, \qquad -\rho K, \qquad \rho B, \qquad -\rho.$$

Normalize their least residues by p. The least residue of $T_{j\ell}$, divided by p, is the fractional part of the resulting real linear form. Its square residue is obtained from the twenty-five ordered products of the five features, using the normalized residues of the products of the corresponding field coefficients.

Permit these five linear coefficients and twenty-five quadratic coefficients to vary independently in $[0, 1]$. In addition, use five unrestricted real coefficients for $H_{j\ell}/p$, and one real parameter for d/p. Thus all good-label matrices are included in a family of fractional-part tests in $5 + 25 + 5 + 1 = 36$ real variables. Allowing independent coefficients enlarges the family. The five height coefficients need no bound in Lemma 2.2.

Each feature has absolute value at most

$$L = (R+1)^2 + M + 2 \leq (M+6)R^2 \qquad (R \geq 1).$$

The two forms inside the fractional parts have absolute values at most $5L$ and $25L^2$. With bR row-label positions, Lemma 2.2 bounds the number of matrices, for fixed carries and parity bits, by

$$\Big[38\big(4bR(2(25L^2 + 5L) + 1) + 1\big)^{37}\Big]^2. \tag{7.8}$$

This is $O_{b,M}(R^{370})$. Multiplying by (7.4) and the eight parity choices gives

$$\#\{\text{good-label matrices on an } R\text{-window}\} \leq C_{b,M} R^{394}. \tag{7.9}$$

For example, one may take

$$C_{b,M} = 25 \cdot 10^8 \cdot 8 \cdot 1444\big[4b(2(25(M+6)^2 + 5(M+6)) + 1) + 1\big]^{74}.$$

These estimates retain every equality case in the comparisons.

Proof of Proposition 7.1. Keep only those matrices in the counted family whose rows are all nonempty. Every actual window satisfying (7.1) gives such a matrix, by the covering property (4.1). Lemma 2.3, with all cyclic rotations included, supplies an $R \geq 2$ and a word meeting every retained matrix in every rotation, since

$$C_{b,M} R^{395} (1 - 1/b)^R < 1$$

for sufficiently large R. Fix this word. The labels on a window starting at h form the rotation beginning at $h \bmod R$. A good entry at which this rotation meets the matrix gives the asserted exact integer representation. □

8. Placement and maximal completion

Fix an alphabet and its periodic word, with period R. For the moment suppose that p, r are odd primes, $r \leq p$, and $0 \leq T < K$, and that

$$\begin{gathered} 4(M + K + 4) \leq p, \qquad 6KRr \leq p, \qquad R \leq r, \\ 10p^2 \leq N, \qquad N < B_{r-1}, \qquad \lfloor N/p \rfloor \leq p^2. \end{gathered} \tag{8.1}$$

By (6.1), the baselines are strictly increasing and

$$0 < B_{h+1} - B_h \leq \Delta, \qquad \Delta = 3Kpr, \qquad R\Delta \leq p^2/2. \tag{8.2}$$

Write $J = 2p(M + 2)$. The first condition in (8.1) gives $J \leq p^2/2$, and therefore

$$J + R\Delta \leq p^2 \leq 2p(p - 1). \tag{8.3}$$

Choose H as the largest integer in $[0, r - 1]$ for which

$$B_H + Q_p + pM + 1 \leq N. \tag{8.4}$$

This choice exists and satisfies $R - 1 \leq H < r - 1$. Indeed, $B_0 = 4p^2 + pT$, so (8.2) gives $B_{R-1} \leq 4p^2 + pT + R\Delta$. Together with $T < K \leq p$, $M \leq p$, and (8.1), this implies (8.4) at $R - 1$. The last index $r - 1$ fails because $N < B_{r-1}$.

All elements of $S = C_p \cup \{A_h : 0 \leq h \leq H\}$ lie in $[0, N)$, by (8.4). Lemma 6.1 shows that S is Sidon, and

$$|S| \leq p + H + 1 \leq p + r \leq 2p. \tag{8.5}$$

Lemma 8.1 (Interior points are blocked)**.** *If $x \notin S$, $p \nmid x$, and $B_{R-1} + J \leq x < B_H + J$, then $S \cup \{x\}$ is not Sidon.*

Proof. Let t be the largest index in $[0, H]$ with $B_t + J \leq x$. The bounds on x give $R - 1 \leq t < H$ and $x < B_{t+1} + J$. For $h = t - R + 1$ and $0 \leq j < R$, monotonicity and (8.2) give $J \leq x - B_{h+j} < J + R\Delta \leq 2p(p-1)$. Proposition 7.1 gives $x = A_{h+j} + c - d$ with $c, d \in C_p$. Hence $x + d = A_{h+j} + c$ is an equality of pair sums in $S \cup \{x\}$, and the unordered pairs differ because $x \notin S$. □

The unblocked points are thus contained in

$$[0, B_{R-1} + J] \cup [B_H + J, N] \cup p\mathbb{Z}. \tag{8.6}$$

The first interval has diameter at most $10p^2$. For the second, maximality of H gives

$$N < B_{H+1} + Q_p + pM + 1 \leq B_H + \Delta + Q_p + pM + 1.$$

Therefore $\max(0, N - B_H - J) \leq 10p^2$.

Lemma 8.2 (Cost of completion)**.** *Under* (8.1)*, some maximal Sidon subset of $[0, N)$ containing S has at most $14p$ elements.*

Proof. Extend S to an inclusion-maximal Sidon subset A of the finite host. Any point blocked by S remains blocked by every superset, so all new marks lie in (8.6).

If a Sidon set has m elements in an interval of diameter L, its $m(m-1)$ ordered nonzero differences are distinct. In particular $m(m-1) \le 2L+1$. For $L \le 10p^2$, this implies $m \le 5p$. For marks divisible by p, divide all marks by p. The resulting set remains Sidon and has diameter at most $\lfloor N/p \rfloor \le p^2$, so it has at most $2p$ elements. Restricting A to each exceptional set preserves Sidonicity. Using these three bounds, allowing overlap, and (8.5), gives

$$|A| \le |S| + 5p + 5p + 2p \le 14p.$$

Translation by 1 gives a maximal Sidon subset of $[1, N]$ with the same cardinality. In the certificate this step uses the corresponding translation of the exceptional residue class. □

9. Prime selection and proof of the main theorem

Fix the two alphabets and the modulus K from Proposition 5.1. Apply Proposition 7.1 to each alphabet with this K, obtaining periods R_1, R_2. Put

$$\begin{aligned} &R_* = R_1 + R_2, && M_* = M_1 + M_2, \quad L_* = 12KR_* + 16, \\ &n_0 = 40L_*^2 + 4(M_* + K + 4) + R_* + 2. \end{aligned} \tag{9.1}$$

These are absolute finite integers fixed independently of N.

Proof of Theorem 1.1. Let $N \ge n_0^3$ and choose $n \ge n_0$ such that $n^3 \le N < (n+1)^3$. Bertrand's postulate supplies primes

$$L_* n < p \le 2L_* n, \qquad n < r \le 2n. \tag{9.2}$$

Both primes are odd. Proposition 5.1 selects a compatible alphabet for this p and a shift $T < K$. Use its period $R \le R_*$ and bound $M \le M_*$.

We verify (8.1). Since $n \ge n_0$,

$$4(M + K + 4) \le n < p, \qquad R \le n < r \le p, \qquad 6KRr \le 12KR_*n \le L_*n < p.$$

Also $10p^2 \le 40L_*^2 n^2 \le n^3 \le N$. For $r \ge 2$ we have $a_r(r-1) \ge r^2$. Therefore $B_{r-1} \ge pKr^2 \ge 16n^3 > (n+1)^3 > N$, where $n \ge 2$ and $K \ge 1$. Finally $p \ge n+1$ gives $N < p^3$, hence $\lfloor N/p \rfloor \le p^2$.

Lemma 8.2 provides a maximal Sidon set in $[1, N]$ of cardinality at most $14p \le 28L_*n \le 28L_*N^{1/3}$. Thus one may take $C = 28L_*$ and $N_0 = n_0^3$. □

Remark 9.1. The proof is existential. The finite pattern arguments determine absolute alphabets and words. The resulting constants have not been optimized. No numerical search for these objects is used in establishing the theorem.

Appendix A. The Lean certificate

The original companion file `Erdos156.lean` contains the construction-specific auxiliary results and the theorem `Bounty.target`. The original challenge wrapper imports `Mathlib`, `FormalConjectures.ErdosProblems.156`, and `TaskSupport`. Its final declaration is

```
namespace Bounty
theorem target : fcTypeOfName% "Erdos156.erdos_156" :=
  target_iff_small_maximal_sets.mpr
    Construction.eventual_small_maximal_sets
end Bounty
```

The target type is checked against the explicit proposition

$$\text{True} \quad \Longleftrightarrow \quad (N \mapsto (\texttt{minMaximalSidonSet}(N) : \mathbb{R})) = O_{N\to\infty}(N \mapsto (N : \mathbb{R})^{1/3}).$$

The reduction `target_iff_small_maximal_sets` identifies this statement with the existence of eventually bounded maximal sets. The proof uses the construction above to supply these sets.

Verification record. The original source report gives Lean 4.35.0-rc2 and the following revisions.

Mathlib	`065356127b1dc0016f66b7283ce0ce2c4055aa55`
Challenge	`bf40872fe09f3c9c94e5b1480823a0622cc1698a`

The reported single-file source has 3,725 lines and SHA-256 digest

`e266058ad3e0e487ea88440470c6d8dbc0`
`f51999b08daa68c0fce5bd3bdc664b`.

The pinned portable project audit reports a successful statement bridge, source build, and kernel replay. It audited all 178 public theorems, including the eleven declarations with an inline `@[simp]` attribute omitted from the original count [8]. The final theorem depends exactly on `propext`, `Classical.choice`, and `Quot.sound`. The bibliography includes the standard references for Lean and Mathlib [5, 6].

Correspondence between the paper and certificate. In the table, the `SignPatterns`, `Window`, and `Construction` namespaces are nested in `Bounty`.

Mathematical component	Principal Lean declarations
Affine and fractional-part pattern bounds	`SignPatterns.affine_sign_determined_card_le` `SignPatterns.fractional_phase_card_le`
Finite selection and rotations	`Bounty.exists_hitting_word_permutations` `Bounty.exists_hitting_word_of_polynomial_bound`
Quadratic seed and carries	`Bounty.quadraticSeed_isSidon` `Bounty.quadratic_difference_canonical`
Separated covering alphabets	`Construction.exists_separatedAlphabet`
Tags for every prime	`Construction.one_of_two_tag_unique` `Construction.exists_two_tagged_alphabets`

Window counts and periodic word	`Window.floorFamily_card_polynomial`
	`Window.allWindowFamily_card_polynomial`
	`Window.CoverAlphabet.exists_ruler_word`

Sidon union and interior blocking	`Construction.seed_anchor_union_isSidon` `Construction.CoverScheme.blocks_interior`
Truncation and maximal completion	`Construction.CoverScheme.exists_truncation` `Construction.maximal_completion_shifted`

Prime selection and asymptotic bound	`Construction.choose_primes_at_scale`
	`Construction.eventual_small_maximal_sets`
	`Bounty.target`

Acknowledgments

OpenAI Codex provided substantial assistance with mathematical development, the Lean formalization, proof checking, and preparation of the exposition and LaTeX source. The formalization uses Lean and Mathlib, and the target statement is supplied by the Formal Conjectures project and the Conjectures.io task bundle [7].

References

[1] P. Erdős, A. Sárközy, and V. T. Sós, On sum sets of Sidon sets, I, *J. Number Theory* 47 (1994), no. 3, 329–347. DOI.

[2] I. Z. Ruzsa, A small maximal Sidon set, *Ramanujan J.* 2 (1998), 55–58. DOI.

[3] N. Sauer, On the density of families of sets, *J. Combin. Theory Ser. A* 13 (1972), no. 1, 145–147. DOI.

[4] S. Shelah, A combinatorial problem; stability and order for models and theories in infinitary languages, *Pacific J. Math.* 41 (1972), no. 1, 247–261. DOI.

[5] L. de Moura and S. Ullrich, The Lean 4 theorem prover and programming language, in *Automated Deduction—CADE 28*, LNCS 12699, Springer, 2021, 625–635. DOI.
[6] The mathlib Community, The Lean mathematical library, in *Proceedings of the 9th ACM SIGPLAN International Conference on Certified Programs and Proofs (CPP '20)*, ACM, 2020, 367–381. DOI.
[7] Formal Conjectures project and Conjectures.io, Erdős problem 156, formalized task bundle at revision `bf40872f`.
[8] Justin Sun Prize submission JSP-000154 and pinned manuscript and verification notes.

Email address: `howew@purdue.edu`

Email address: `huffma97@purdue.edu`

Email address: `jenwin@purdue.edu`

Email address: `kruerl@purdue.edu`