Hidden Signal in the Ulam sequence explained

Przemyslaw Chojecki
ulam.ai

January 20, 2026

Abstract

The classical (1,2)-Ulam sequence (a_n) is defined by the greedy rule that a_{n+1} is the smallest integer $> a_n$ representable *uniquely* as $a_i + a_j$ with $i < j \le n$. Extensive computations (Steinerberger) suggest a striking "hidden signal": there exists a frequency $\alpha \approx 2.571\ldots$ such that $\sum_{k \le N} \cos(\alpha a_k)$ grows linearly negative, and the phases $\alpha a_k \bmod 2\pi$ appear confined to a non-uniform distribution. This note isolates a clean, fully rigorous mechanism that explains *why such a signal is natural* once a large Fourier peak is present.

1 Introduction

The classical (1,2)-Ulam sequence (a_n) is defined by the greedy rule that a_{n+1} is the smallest integer $> a_n$ representable *uniquely* as $a_i + a_j$ with $i < j \le n$. Denote $A_n := \{a_1, \ldots, a_n\}$. A convenient reformulation is in terms of the ordered representation count

$$s_n(m) := \#\{(x, y) \in A_n^2 : \; x + y = m\}.$$

If m has a unique representation with distinct summands, then $s_n(m) = 2$ (coming from the two orders), and if additionally $m = 2x$ for some $x \in A_n$ then $s_n(m) = 3$. Thus, for the next Ulam term a_{n+1} one always has

$$s_n(a_{n+1}) \in \{2, 3\} \quad \text{and in particular} \quad s_n(a_{n+1}) \le 3. \tag{1}$$

Despite its simple definition, the (1,2)-Ulam sequence exhibits unexpectedly rigid global structure and remains poorly understood from a purely rigorous standpoint. A striking computational discovery due to Steinerberger is the presence of a "hidden signal": there appears to exist a real frequency $\alpha \approx 2.571\ldots$ for which the exponential sum $\sum_{k \le N} e^{\mathrm{i}\alpha a_k}$ grows linearly in N with a strong negative-real bias, and the phases $\alpha a_k \bmod 2\pi$ are far from equidistributed [2]. This phenomenon suggests that the greedy "unique sum" rule acts as a spectral filter, selecting integers with a persistent phase preference.

Ross initiated a systematic Fourier-analytic approach to Ulam-type sets, emphasizing that the defining rule severely suppresses additive triples $x + y = z$ relative to random sets and that such triple-scarcity should force substantial Fourier bias (in the spirit of additive-combinatorial inverse theorems) [1]. More recently, Clément and Steinerberger proved a complementary structural constraint in a different direction: among the first n consecutive ratios a_{k+1}/a_k one finds values arbitrarily close to 1 (quantitatively, $\min_{k \le n} a_{k+1}/a_k \le 1 + c \log n / n$ for an absolute constant c), giving a rigorous form of "small gaps" in the multiplicative sense [3].

The purpose of this note is to assemble a clean, fully rigorous *mechanism* that explains why a hidden Fourier signal is natural once one has a large (and sufficiently isolated) Fourier peak. The point is not to guess the frequency α or to assume heuristics about randomness; rather, we show that, conditional on a small set of global inputs (density/baseline size and spectral dominance), the observed stability is forced by deterministic identities.

The picture is a deterministic feedback loop:

- **Structure from scarcity (rigorous, conditional on density):** having very few solutions to $x + y = z$ forces a large Fourier coefficient once the set has positive density (or, more generally, once the DC baseline is not too small).
- **Phase trapping (rigorous):** a large Fourier peak forces any "unique-sum" integer to lie in a specific geometric trough for that frequency (up to a controllable error).
- **Availability (rigorous):** off-peak energy cannot make the error large everywhere; "good" points exist in any sufficiently long window.
- **Reinforcement (rigorous):** if the peak is aligned (empirically near angle π), choosing integers from the trough increases the peak magnitude, creating a stable feedback loop.

What remains open is to verify for the Ulam sequence the needed global inputs (density/baseline size, and a quantitative form of spectral dominance or energy localization). Given those inputs, the hidden signal becomes a theorem: the greedy rule enforces phase trapping, and trapping enforces reinforcement.

2 Fourier setup in a cyclic group (no wrap-around)

Fix $N \geq 1$ and choose an integer modulus $q > 2N$. We view subsets of $\{1, \ldots, N\}$ as subsets of $\mathbb{Z}/q\mathbb{Z}$. Because $q > 2N$, any equality $x + y \equiv m \pmod{q}$ with $x, y, m \in \{1, \ldots, N\}$ is equivalent to the integer equality $x + y = m$.

Let $\omega := e^{2\pi i/q}$. For a finite set $A \subset \mathbb{Z}/q\mathbb{Z}$, define the Fourier sums

$$S_A(r) := \sum_{x \in A} \omega^{rx} \qquad (r \in \mathbb{Z}/q\mathbb{Z}).$$

When $A = A_n$, write $S_n(r) := S_{A_n}(r)$.

2.1 Exact Fourier inversion for $s_n(m)$

The convolution identity gives, for all $m \in \mathbb{Z}/q\mathbb{Z}$,

$$s_n(m) = \frac{1}{q} \sum_{r \in \mathbb{Z}/q\mathbb{Z}} S_n(r)^2 \, \omega^{-rm}. \tag{2}$$

In particular, the $r = 0$ term is $S_n(0)^2/q = n^2/q$.

3 Scarcity of triples forces spectral bias (static)

For a finite set $A \subset \mathbb{Z}/q\mathbb{Z}$, define the additive triple count

$$T(A) := \#\{(x, y, z) \in A^3 : \ x + y = z\}.$$

For $A = A_n$, write $T_n := T(A_n)$.

3.1 Triple identity

Let $f = \mathbf{1}_A$. Using Fourier inversion,

$$T(A) = \frac{1}{q} \sum_{r \in \mathbb{Z}/q\mathbb{Z}} \widehat{f}(r)^2 \, \overline{\widehat{f}(r)} = \frac{1}{q} \sum_{r \in \mathbb{Z}/q\mathbb{Z}} S_A(r)^2 \, \overline{S_A(r)}. \tag{3}$$

The $r = 0$ term contributes $|A|^3 / q$.

3.2 A quantitative lower bound for the largest nontrivial Fourier coefficient

Let $A \subset \mathbb{Z}/q\mathbb{Z}$ with $|A| = n$ and define

$$M(A) := \max_{r \neq 0} |S_A(r)| .$$

Theorem 1 (Scarcity $\Rightarrow$ spectral bias). *Let $A \subset \mathbb{Z}/q\mathbb{Z}$ have size n, and let $T(A)$ be its triple count* (3)*. Then*

$$M(A) \ \geq \ \frac{n^2}{q} - \frac{T(A)}{n}. \tag{4}$$

Proof. From (3),

$$\frac{n^3}{q} - T(A) = \frac{1}{q} \sum_{r \neq 0} S_A(r)^2 \, \overline{S_A(r)}.$$

Hence

$$\left| \frac{n^3}{q} - T(A) \right| \leq \frac{1}{q} \sum_{r \neq 0} |S_A(r)|^3 \leq \frac{M(A)}{q} \sum_{r \neq 0} |S_A(r)|^2 .$$

By Parseval, $\sum_r |S_A(r)|^2 = q \sum_x f(x)^2 = qn$, and $S_A(0) = n$, so $\sum_{r \neq 0} |S_A(r)|^2 = qn - n^2 \leq qn$. Thus

$$\frac{n^3}{q} - T(A) \leq M(A) \, n,$$

which rearranges to (4). (The same bound follows if the left-hand side is negative.) ☐

Remark 2 (Specialization to Ulam truncations). For the Ulam set A_n, every a_k ($k \geq 3$) has a unique representation $a_k = a_i + a_j$ with $i < j < k$, so the number of distinct-summand triples is exactly $n - 2$. Counting ordered triples and allowing diagonals gives $T_n = O(n)$ (e.g. $T_n \leq 3n$ suffices). Under an additional *density-type* hypothesis $a_n \sim \lambda n$ (so one may take $q \asymp a_n \asymp n$), (4) forces $M(A_n) \gg n$, i.e. a linear-sized Fourier peak. This is the first (static) link: *scarcity of solutions to $x + y = z$ forces spectral bias once density is positive.* See Ross' thesis for stronger inverse-theorem formulations of this philosophy.

4 Phase trapping from a large Fourier peak (local, deterministic)

Fix n, choose a scan horizon N with $a_n \le N$, and pick $q > 2N$. Fix a nonzero frequency $r^* \in \mathbb{Z}/q\mathbb{Z}$ and write

$$S_n(r^*) =: M\,\mathrm{e}^{\mathrm{i}\theta}, \qquad M := |S_n(r^*)|\,.$$

Using symmetry $S_n(-r^*) = \overline{S_n(r^*)}$ and splitting (2) into $r \in \{0, \pm r^*\}$ plus the rest yields the *exact* decomposition

$$s_n(m) = \frac{n^2}{q} + \frac{2M^2}{q}\cos\Big(2\theta - \frac{2\pi r^* m}{q}\Big) + E_n(m), \tag{5}$$

where

$$E_n(m) := \frac{1}{q}\sum_{\substack{r\neq 0\\ r\neq \pm r^*}} S_n(r)^2\,\omega^{-rm}. \tag{6}$$

Theorem 3 (Deterministic phase trapping). *Let m be any integer (viewed in $\mathbb{Z}/q\mathbb{Z}$) such that*

$$s_n(m) \le K \quad \textit{and} \quad |E_n(m)| \le \tau\,\frac{M^2}{q}.$$

Then

$$\cos\Big(2\theta - \frac{2\pi r^* m}{q}\Big) \le \frac{qK - n^2}{2M^2} + \frac{\tau}{2}. \tag{7}$$

In particular, if the right-hand side equals $-1+\varepsilon$ with $0 < \varepsilon \le 1$, then the phase lies in a trough arc:

$$\Big|\Big(2\theta - \frac{2\pi r^* m}{q}\Big) - \pi\Big| \le \sqrt{2\varepsilon}.$$

Proof. Rearrange (5) and use the bounds on $s_n(m)$ and $E_n(m)$. The arc-width bound follows from $\cos(\pi \pm x) = -\cos x$ and $1 - \cos x \ge x^2/2$ for $x \in [0,\pi]$. ☐

Remark 4 (Ulam admissibility corresponds to $K = 3$). For $m = a_{n+1}$, (1) gives $s_n(m) \le 3$, hence $K = 3$ is valid throughout for the Ulam selection.

5 Availability (noise cannot block the trap everywhere)

To quantify how often the remainder $E_n(m)$ can be large, define the off-peak L^4 ratio

$$\eta := \frac{\sum_{r\neq 0,\pm r^*} |S_n(r)|^4}{M^4}. \tag{8}$$

Theorem 5 (Availability via Parseval / off-peak energy). *Let $\mathcal{I} \subset \mathbb{Z}/q\mathbb{Z}$ be any interval of length L, and let $\tau > 0$. Then*

$$\#\Big\{m \in \mathcal{I} :\ |E_n(m)| > \tau\,\frac{M^2}{q}\Big\} \le \frac{\eta}{\tau^2}\,L.$$

Equivalently, at least $(1-\eta/\tau^2)L$ points of $\mathcal{I}$ satisfy $|E_n(m)| \le \tau M^2/q$.

Proof. Normalize $\widetilde{E}_n(m) := qE_n(m)/M^2$. By Parseval,

$$\frac{1}{q}\sum_{m \bmod q}\Big|\widetilde{E}_n(m)\Big|^2 = \frac{1}{M^4}\sum_{r\neq 0,\pm r^*} |S_n(r)|^4 = \eta.$$

Apply Chebyshev on $\mathcal{I}$. ☐

6 Reinforcement (phase trap increases the peak)

Let $S \in \mathbb{C}$ with $|S| = M$ and write $S = M\mathrm{e}^{\mathrm{i}\theta}$. Let $u = \mathrm{e}^{\mathrm{i}\phi}$ be a unit vector.

Theorem 6 (Geometric reinforcement)**.** *If* $|\phi - \theta| \leq h \leq \pi$*, then*

$$|S + u| \geq M + \cos h \geq M - \frac{h^2}{2}.$$

Proof. Compute $|S + u|^2 = M^2 + 1 + 2M\cos(\phi - \theta)$ and use $\cos(\phi - \theta) \geq \cos h$. □

Remark 7 (Negative-real attractor)*.* Empirically, the dominant Fourier mode for Ulam truncations has argument close to π (negative real axis). Combined with Theorem 3, this forces ϕ close to π as well (a trough centered at π), and Theorem 6 then yields monotone linear growth of M along trapped steps.

7 A conditional "phase-locked loop" mechanism

Theorems 3, 5, and 6 combine into a deterministic feedback picture:

- **Phase filter:** uniquely representable sums require $s_n(m) \leq 3$, which forces m into a trough whenever $|E_n(m)|$ is not too large (Theorem 3).
- **Availability:** off-peak energy cannot make $|E_n(m)|$ large on most integers of a window (Theorem 5).
- **Reinforcement:** choosing m from the trough pushes the dominant Fourier sum further in the same direction (Theorem 6).

This yields a *conditional theorem*: once one verifies the required spectral dominance and non-sparsity inputs for the Ulam process, the signal becomes a deterministic consequence.

7.1 Capture vs. jump (a deterministic dichotomy)

A convenient corollary is that failure of capture forces large gaps.

Proposition 8 (Capture vs. jump)**.** *Fix* r^* *and define* M, η, E_n *as above. Let* $m = a_{n+1}$ *and let* $\tau > 0$*. If*

$$|E_n(m)| > \tau \, \frac{M^2}{q},$$

then m *must lie in the exceptional set counted in Theorem 5. In particular, if one considers the scan window* $\mathcal{I} = [a_n + 1, a_{n+1}]$ *of length* $L = a_{n+1} - a_n$*, then necessarily*

$$L \geq \frac{\tau^2}{\eta}.$$

Equivalently, if $a_{n+1} - a_n < \tau^2/\eta$*, then* $|E_n(a_{n+1})| \leq \tau M^2/q$ *and Theorem 3 forces* a_{n+1} *into the trough.*

8 What remains open and next results to attack

The framework above is fully rigorous *given* a suitable spectral regime. The remaining bottlenecks for the actual Ulam sequence are cleanly isolatable:

1. **Density / baseline input.** Phase trapping is strongest when the DC baseline n^2/q is large (e.g. when $q \asymp a_n$ and $a_n = O(n)$). Any lower bound of the form $n^2/a_n \gg 1$ infinitely often already yields recurring strong trapping episodes. Proving (or disproving) positive density remains a central open problem.

2. **Existence of a large Fourier coefficient.** One wants $M = |S_n(r^*)| \geq cn$ for some fixed $c > 0$ infinitely often (or eventually always). Theorem 1 gives this immediately under a positive-density hypothesis together with the Ulam triple scarcity $T_n = O(n)$; Ross' thesis develops stronger inverse-theorem statements in this direction.

3. **L^4-dominance / energy localization.** Our strongest dynamical consequences use the off-peak ratio η in (8). A natural next lemma is to show that for Ulam truncations, whenever M is large, η is not too large (i.e. Fourier energy is not spread over too many competing modes). This turns "noise" into a quantifiable small exceptional set.

4. **Stability and convergence of the maximizing frequency.** If the maximizing frequency is separated by a gap from the runner-up, then it cannot change quickly under the update $S_{n+1}(r) = S_n(r) + \omega^{ra_{n+1}}$ (each coefficient changes by at most 1 in magnitude). One can aim to prove long plateaus of stable maximizing frequency and subsequential convergence to a limiting α.

5. **Mixing/ergodicity inside the trough arc.** To reach arithmetic consequences (e.g. infinitely many gaps 2, or fine gap statistics), one needs that the phases explored within the trough arc are not confined to a single point. A next step is to deduce a quantitative "spread" statement from the greedy scan rule plus availability (Theorem 5).

Summary (what is rigorously established)

The results established here provide a deterministic feedback loop:

Triple scarcity $\Rightarrow$ *spectral bias (under density)* $\Rightarrow$ *phase trapping* $\Rightarrow$ *reinforcement*,

with *availability* ensuring that off-peak energy cannot obstruct the trap on most integers. This gives a rigorous conditional explanation of the hidden Fourier signal: once a dominant mode is present and not overly contaminated by off-peak energy, the greedy Ulam rule acts as a phase filter that locks onto and reinforces that mode.

References

[1] D. Ross, *The Ulam Sequence: Structure and Randomness*, PhD thesis, University of Illinois at Urbana–Champaign, 2016.

[2] S. Steinerberger, “The asymptotic distribution of Ulam numbers,” *Experimental Mathematics* (2015/2016; see published version and arXiv preprint for details).

[3] F. Clément and S. Steinerberger. *Small Gaps in the Ulam sequence.* arXiv:2501.16285, 2025. (Also appears in Comptes Rendus Mathématique.)