The Ulam Numbers up to One Trillion

Philip Gibbs *and* Judson McCranie

All Ulam numbers up to one trillion are computed using an efficient linear-time algorithm. We report on the distribution of the numbers including the positions of the largest gaps and we provide an explanation for the previously observed period of clustering.

Introduction

The Ulam numbers were introduced in 1964 by Stanislaw Ulam as a sequence of integers with a simple additive definition that provide an apparently pseudo-random behaviour [1]. The sequence of Ulam numbers begins with the two integers $U_1 = 1$ and $U_2 = 2$. Each subsequent Ulam number U_n is the smallest integer that is the sum of two distinct smaller Ulam numbers in exactly one way. It continues 1, 2, 3, 4, 6, 8, 11, 13, 16, 18, 26, 28, 36, … In this study we have calculated all Ulam numbers less than one trillion. There are 73,979,274,540 Ulam numbers up to this limit.

The Ulam numbers form an infinite sequence. In the worst case an Ulam number U_n is the sum of U_{n-1} and U_{n-3} Therefore $U_n \leq N_{n-2}$ where $\{N_n\}$ are the numbers in Narayana's cows sequence: 1,1,1,2,3,4,6,9,13,19,… with the recurrence relation $N_n = N_{n-1} + N_{n-3}$ [2]. A lower bound can be set by observing that given five consecutive numbers greater than five, at most two can be Ulam numbers so $U_n \geq 2.5n - 7$. These limits imply that the asymptotic density $D = \lim_{n\to\infty} \frac{U_n}{n}$ of Ulam numbers in the positive integers is constrained by $0 \leq D \leq 0.4$. However, these bounds are not particularly good and whether or not the density is greater than zero is unsolved. Ulam proposed this question of density as a key problem about the sequence.

If the distribution of Ulam numbers could be treated as random, the expectation would be for the density to decrease asymptotically towards zero. The first indication that the situation is more complex came when computations indicated the density converges to a non-zero value which we now estimate as $D \cong 0.073979$ based on our count of Ulam numbers less than a trillion. It was also observed by David Wilson that once the sequence settled down, the Ulam numbers appear in clusters occurring with a period of about 22 [3].

A breakthrough arrived in 2015 when Stefan Steinerberger accidentally discovered a hidden signal in the Fourier analysis of the sequence at a frequency $\alpha \sim 2.57145$ [4]. He also observed that for all computed Ulam numbers except 2,3,47 and 69,

$$\cos(\alpha U_n) < 0$$

We can now confirm that this holds for all Ulam numbers less than one trillion. This observation is equivalent to the statement that when Ulam numbers are taken modulo a wavelength $\lambda = \frac{2\pi}{\alpha}$, the residues $[U_n]_\lambda$ lie in the central half of the range for all but the four exceptional Ulam numbers.

$$[U_n]_\lambda \in \left(\frac{\lambda}{4}, \frac{3\lambda}{4}\right), U_n > 69$$

The observation of clustering with a period of 22 can be attributed to the fact that $\lambda \cong \frac{22}{9}$ to a good approximation so that 9 wavelengths fit into a cluster of 22. However this explanation is incomplete. See the final section below for more details.

Philip Gibbs made a more general conjecture that the residues lie in the central third of the range plus a small margin ε either side with only a finite number of exceptions for any fixed ε no matter how small [5].

$$\exists\lambda: \forall\varepsilon > 0, \exists N: \forall n > N, [U_n]_\lambda \in \left(\frac{\lambda}{3} - \varepsilon, \frac{2\lambda}{3} + \varepsilon\right)$$

It is useful to use asymptotic inequalities as a shorthand for this statement as follows

$$\frac{\lambda}{3} \lesssim [U_n]_\lambda \lesssim \frac{2\lambda}{3}$$

It had been observed in 1996 by Calkin and Erdös that a sequence of numbers for which the residues of *all* elements lie in the central third of the range is sum-free [6]. It may therefore be useful to regard the Ulam sequence as an “almost” sum-free set (see e.g. [7]).

An Efficient Algorithm

Using a direct method of calculation, verifying each new Ulam number U_n when all smaller Ulam numbers are known takes a computation time of order (n) . Ulam numbers must be computed in an ascending sequence so it takes $O(N^2)$ time to compute Ulam numbers up to N (This conclusion assumes that the asymptotic density is non-zero). Prior to 2015 Judson McCranie had calculated all Ulam numbers up to 4.394 billion in this way.

The discovery that there is a natural wavelength λ in the Ulam sequence made it possible to construct a more efficient algorithm for generating the Ulam numbers with computational time of order $O(N)$ rather than $O(N^2)$ [8]. The correctness of the $O(N)$ estimate depends on the correctness of conjectures concerning the distribution of the Ulam numbers modulo λ but is confirmed empirically in the computations.

Using this algorithm implemented in Java it has now been possible to compute all Ulam numbers less than 1 trillion. The computation time taken to perform the calculation was in the order of 100 hours on a PC. The limitation to proceeding further is memory space. The computer used for the calculation had 32 GB of fast memory available, but for this computation a heap size of 64GB was needed. The use of virtual memory paged from disk space proved adequate for this requirement. To compute significantly further without using much more fast memory it would be necessary to use a more efficient database technology.

We are grateful to Donald Knuth for reviewing, improving and documenting the algorithm while rewriting it in CWEB [9]

The following table shows computed values for $K(N)$, the count of Ulam numbers less than N where N is a power of ten. The density D computed from these numbers is also shown. The only two powers of ten known to be Ulam numbers are a million and a hundred billion.

N	$K(N)$	D	Low Outliers	High Outliers
10	6	0.6	2	1
100	26	0.26	6	4
1,000	125	0.125	9	10
10,000	827	0.0827	16	15
100,000	7,584	0.07854	25	38
1,000,000	74,083	0.074083	58	62
10,000,000	740,368	0.0740368	115	105
100,000,000	7,399,353	0.07399353	243	198
1,000,000,000	73,976,842	0.073976842	445	412
10,000,000,000	739,778,038	0.0739778038	862	788
100,000,000,000	7,397,814,233	0.07397814233	1658	1626
1,000,000,000,000	73,979,274,540	0.073979274540	3226	3269

Table 1 - count of Ulam numbers and outliers

Distribution of Residues

Conjectures concerning the distribution of Ulam numbers modulo the underlying wavelength λ can be explored with our computation. For most Ulam numbers the residue lies in the middle third of the range. Those that lie outside are called outliers. Table 1 includes counts of the number of lower outliers with residues in the bottom third of the range and high outliers with residues in the top third.

The count of outliers is sensitive to the value assumed for λ in the computation. An accurate value of λ can be determined by adjusting its value so that the numbers of higher and lower outliers are equal. Alternatively the total number of outliers could be minimised. In this way we have obtained a very precise estimate.

$$\lambda \cong 2.443442967784743433$$

For the purposes of the computation it is useful to use a rational approximation to this number to avoid mistakes due to rounding errors. The results reported here are based on

$$\lambda = \frac{856371966}{350477575}$$

It is unlikely that any exact closed expression for λ is obtainable because its value depends on the appearance of outliers in the sequence of Ulam numbers and this seems to be a pseudo-random process.

Figure 1 shows the distribution of residues in the middle third of the range. Features of this plot are determined by the outliers [5,7]. For example, the space between the two peaks is determined by the smallest low and high outliers 2 and 3. The similarity between the higher ends of both comes from the appearance of Ulam pairs differing by 2. The complex bumps on the tops of the peaks are formed by the contributions of larger outliers

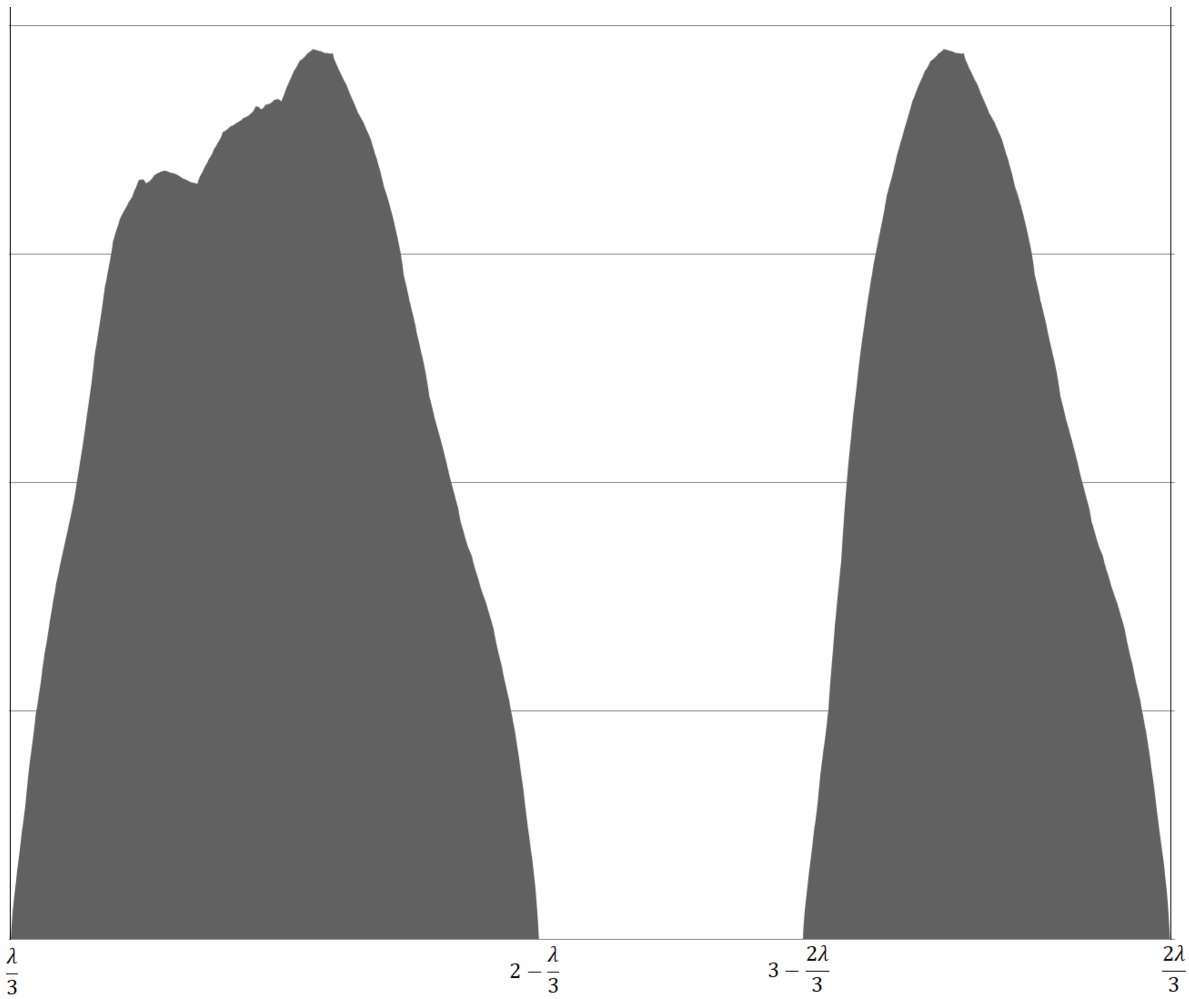

Figure 1 distribution of Ulam residues modulo λ

Outliers

An Ulam number that is not an outlier must be the sum of two Ulam numbers, at least one of which is an outlier. The outliers therefore appear very frequently in the sums, yet outliers themselves are very rare. By time we reach 1 trillion only about 1 in every ten million Ulam numbers is an outlier. The outliers that appear in sums most often are those whose residues are furthest from the central range. These are mostly small. Table 2 shows all the outliers less than 10,000 along with the percentage frequency that they appear in sums and the normalised residue.

$$r_n = \frac{[U_n]_\lambda}{\lambda}$$

The gaps 2, 3, 47, 69, 102, 339, 400, 1155, 2581, 9193 shown in bold in the table are maximal outliers with the property that there are (conjecturally) no larger outliers of the same type (high or low) that are further out from the centre of the residue range. These play an important part in structural features of the Ulam sequence such as the possible sizes of gaps (see below.)

n	U_n	% used	Type	r_n
2	**2**	36.93173	low	0.818517
3	**3**	13.69782	high	0.227776
6	8	1.64413	high	0.274069
8	13	0.04174	high	0.320362
13	36	4.84485	low	0.733309
15	**47**	11.70552	high	0.235153
17	53	0.22611	low	0.690705
18	57	0.00120	high	0.327739
20	**69**	10.13240	high	0.238842
25	97	0.50723	low	0.698082
27	**102**	8.42212	low	0.744375
42	206	0.22984	high	0.307268
53	273	2.92695	low	0.727592
56	316	0.00412	high	0.325711
59	**339**	5.95580	low	0.738659
61	356	0.22476	low	0.696055
67	**400**	0.55262	low	0.703432
78	483	0.00126	low	0.671894
123	983	0.51664	high	0.301184
128	1023	0.00139	low	0.671528
140	**1155**	0.20437	low	0.693660
152	1308	0.16642	high	0.310223
259	**2581**	0.82062	high	0.296396
296	3038	0.00239	high	0.327567
309	3205	0.00401	low	0.673750
357	3723	0.00036	low	0.669694
384	4118	0.00357	high	0.326834
544	6322	0.00001	high	0.332745
583	6862	0.00002	high	0.332378
707	8368	0.00821	low	0.675800
773	**9193**	0.09401	high	0.314129

Table 2 – smallest outliers

Ulam gaps

A gap g is the difference between consecutive Ulam numbers $g = U_n - U_{n-1}$. Knuth found a gap of size $g = 262$ in the Ulam numbers between 64420 and 64682 [11]. Since then, larger gaps have been noted [11] Table 3 shows all the record gaps for Ulam numbers less than 1 trillion. The sequence of gaps is A080288 in the online encyclopaedia of integer sequences while the start and end points are sequences A080287 and A080288.

Some integer values have never appeared as gaps (e.g. 6,11,14,16,18,21,… [12,13]) while others other appear only a few times (i.e. 1 appears 4 times, 4 appears 2 times, 9 appears 3 times.) A gap of $g = 1$ corresponds to consecutive Ulam numbers which appear only after 1, 2, 3 and 47. Some gap values are only seen very rarely and may first appear a long way into the sequence. E.g. a gap of $g = 35$ first appears before the Ulam number 483379914 [12]. This raises the question of whether gap values that have never yet been seen will appear eventually or not.

n	U_n	U_{n-1}	$g = U_n - U_{n-1}$
2	2	1	1
5	6	4	2
7	11	8	3
11	26	18	8
15	47	38	9
25	97	87	10
30	126	114	12
36	175	155	20
55	309	282	27
108	781	751	30
123	983	949	34
151	1296	1257	39
182	1594	1553	41
208	1900	1858	42
232	2247	2178	69
440	4878	4800	78
487	5484	5384	100
1510	18918	18796	122
2930	37699	37562	137
4953	64682	64420	262
18858	253034	252719	315
69181	934296	933709	587
21431880	289738726	289738117	609
24576524	332251032	332250401	631
49395954	667753657	667752899	758
51744076	699497942	699497052	890
71482878	966291200	966290117	1083
16614427387	224582903534	224582902442	1092
23647775834	319654122989	319654121875	1114
30985625147	418843013485	418843012121	1364
59359956445	802386466956	802386465583	1373

Table 3 – biggest Ulam gaps

Although nothing has been proven definitively about the distribution of gaps it is possible to understand much about which gaps are possible if we assume the correctness of the conjecture about the distribution of residues and that there are no missing outliers with residues far from the central region.

Since the number at the start and end of a gap has a residue lying asymptotically in the middle third of the range we know that the gap

$$g = U_n - U_{n-1},\ \frac{\lambda}{3} \lesssim [U_n]_\lambda \lesssim \frac{2\lambda}{3}$$

$$\Rightarrow [g]_\lambda \lesssim \frac{\lambda}{3} \quad \text{or} \quad [g]_\lambda \gtrsim \frac{2\lambda}{3}$$

From this we can conclude that any gap value with a residue in the middle third of the range can only occur a finite number of times. When they do appear one of the Ulam numbers at either the start or end of the gap (or both) must be an outlier. This explains why there are only a few examples of gaps of size 1,4,6,9 and 11, but there are other missing gap sizes that are not explained in this way, e.g. 14, 31, 36, 53 and 54.

Gaps of size 14 can be ruled out using a more involved argument. $[14]_\lambda = 0.72962012\lambda$ Therefore

$$g = U_n - U_{n-1} = 14$$

$$\frac{\lambda}{3} \lesssim [U_n]_\lambda \lesssim \frac{2\lambda}{3}$$

$$\Rightarrow \frac{4\lambda}{3} - [14]_\lambda \lesssim [U_{n-1}]_\lambda \lesssim \frac{2\lambda}{3}$$

$$\Rightarrow \frac{4\lambda}{3} + [12]_\lambda \lesssim [U_{n-1} + 2]_\lambda \lesssim [2]_\lambda - \frac{\lambda}{3}$$

Since $U_{n-1} + 2$ is inside the gap it cannot be an Ulam number and must therefore be equal to at least one other sum of two Ulam numbers including an outlier bigger than 2. However there are no known outliers with residues in a range that makes this possible.

Using similar arguments we have been able to account for appearance or non-appearance of all observed gap sizes up to $g = 263$. Admissible gap sizes become rarer as gap size increases but it is not known if there is a maximum gap size.

Clustering

It was mentioned in the introduction that the Ulam numbers have been observed to cluster over a period of around 22. David Wilson plotted the Ulam numbers as pixels on an image of size 108 by 108 showing that there are five clusters per row. This gives a period for the repetition of clusters of $W \cong 21.6 = \frac{108}{5}$. A more accurate estimate of the period computed previously is $W \cong 21.601584$ [3]. Nine signal wavelengths is $9\lambda \cong 21.990987$ so the clustering does not occur at an exact multiple of the wavelength. A closer approximation comes from $221\lambda \cong 540.000896$ which covers about 25 clusters in 5 times the size used by Wilson, but this still does not explain the exact frequency. Figure 2 shows the Ulam numbers plotted in rows of 540.

The full explanation comes from the observation that if the residue of a number falls in the range $[X]_\lambda \in [E - \delta, E]$ where $E = 2 - \frac{\lambda}{3}$ is the point at the end of the first density peak (figure 1) and $\delta = [5]_\lambda = 5 - 2\lambda$ then the next seven numbers at least have a range of residues that fall outside the high density ranges and therefore are not normally Ulam numbers. This creates a long gap between clusters of potential Ulam numbers. In the clusters themselves residues do fall in the high density ranges and Ulam numbers there are often but not always present. The average period between clusters will therefore be

$$W = \frac{\lambda}{5 - 2\lambda} \cong 21.6015840301253$$

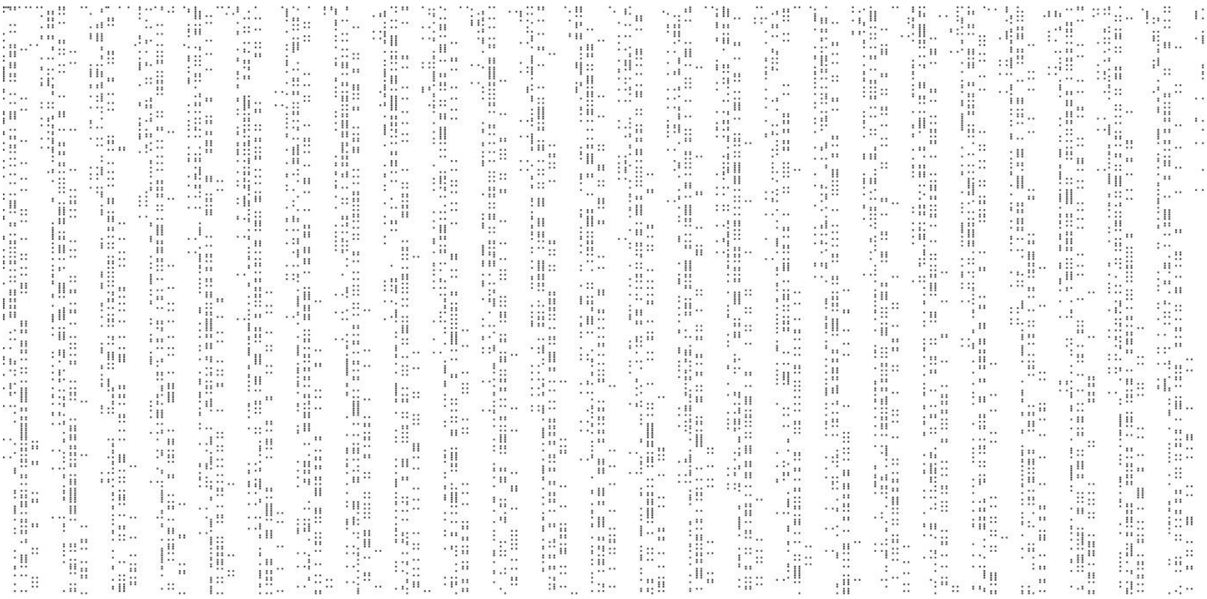

Figure 2 Ulam numbers plotted as pixels on an image

References

[1] S. Ulam, “Problems in Modern Mathematics”, New York: John Wiley & Sons, Inc, p. xi (1964)

[2] Sequence A000930 in “The On-Line Encyclopedia of Integer Sequences”

[3] See notes on sequence A002858 in “The On-Line Encyclopedia of Integer Sequences”

[4] S. Steinerberger, “A Hidden Signal in the Ulam Sequence”, Experimental Mathematics, Vol 26, issue 4, pages 460-467 (2016).

[5] P. Gibbs. “A Conjecture for Ulam Sequences”, viXra:1508.0045 (2015)

[6] N. J. Calkin, P Erdös, "On a class of aperiodic sum-free sets", Mathematical Proceedings of the Cambridge Philosophical Society, 120(1), 1-5. (1996)

[7] D. Ross, “The Ulam Sequence and Related Phenomena”, PhD Thesis at the University of Wisconsin, Madison, available at github (daniel3735928559)

[8] P. Gibbs, “An Efficient Method for Computing Ulam Numbers”, viXra:1508.0085 (2015)

[9] D. Knuth. Documentation of program Ulam-Gibbs, available at www-cs-faculty.stanford.edu/uno/programs.html (2016)

[10] D. Knuth The Art of Computer Programming, Volume 4A, Section 7.1.3. (2006)

[11] Sequence A080288 in “The On-Line Encyclopedia of Integer Sequences”

[12] Sequence A072540 in “The On-Line Encyclopedia of Integer Sequences”

[13] C. A. Pickover, “Wonders of Numbers, Adventures in Mathematics, Mind and Meaning” Oxford University Press, (2000)