MATCHED COLUMN SETS IN REPETEND TABLES

R. J. Booth
Private Researcher
clamjet@hotmail.co.uk

D. J. Horton
Private Researcher

Received: 2/29/24, Revised: 10/21/24, Accepted: 3/7/25, Published: 3/26/25

Abstract

Let n be a positive integer. A *repetend* is the recurring sequence of L digits when m/n is written in some base b coprime to n. Consider all bases less than n which have identical repetend length L, and write their repetends as rows of a matrix (a *repetend table*). If $m = 1$, we prove that, with certain exceptions, the multiset of integers in one column (corresponding to position within the repetends) is equal to the multiset for another column, and determine the number $P(L)$ of such pairings – this depends on L but not on n. Denoting those general column matches as "Class A", we also demonstrate two further classes B and C of column matches under special, but infinitely occurring, conditions. Most of the theorems rely on a closed form expression for each repetend digit, the "Repetend Digit Formula", which appears to be new. This expression also allows us to present a much shorter proof of some results of Lewittes. An interesting side result from research into "sporadic" column matches is a new number-theoretic result, namely a new formula for the number of elements of given order in $\mathbb{Z}_n^*$. There is one appendix, which gives details of a novel hybrid residue number system which we used to compute with n between 26.5 and 65.5 bits in size.

1. Introduction

The patterns emerging from expressing the reciprocal of positive integers as a recurring decimal have produced much literature over time spanning three centuries. Many of the papers concentrate on the sums of subdivisions within the recurrence, known as Midy's theorem [10] of 1836, with various generalizations starting with Ginsberg – see Section 5 for a table of other papers considered. Most of these authors do not use the term "repetend" to denote the recurring digits, yet it is present in the Oxford Dictionary and dates back at least to Heal [7] in 1887.

DOI: 10.5281/zenodo.15091059

Our study here is of properties of repetends of the same number but written in the differing bases which give identical lengths of the repeated sequences of digits. These properties include some fascinating relationships between subsets of their digits. This provides a new area of research which is complementary to the interesting studies on Midy sums.

A key innovation is a formula for calculating any repetend digit independently of the others, in contrast to the sequential nature of the long division algorithm. This allows tests of equality of repetend digits to be performed, enabling proofs of the major features in this paper. It also allows a much shorter proof to be given, in Section 5, of one of Lewittes's [8] theorems.

The following four definitions help to provide a structure for the succeeding analysis.

• Given a positive integer n (sometimes called a *modulus*), a *base* is any positive integer which is coprime to n.

• A *small base* is any base less than the modulus n.

• Given a base b and a positive integer m which is less than and coprime to n, the *repetend* is the minimal set of recurring digits in m/n when written in base b.

• The *repetend length*, denoted by L, is the number of digits in the repetend.

This research has been motivated by the second author, who has been fascinated by repetends since at least 1990. For the $m = 1$ case, he carried out a search on all numbers n up to 1000 and each $L < n$, and recorded a number of interesting statistics and features, most notably linkages between repetends on different bases whereby one repetend digit column can equal a different column as a multiset – leading to the "matched column sets" phrase in the title. These features are described in Theorem 2, with proofs by the first author, mainly using standard group theory on $\mathbb{Z}_n^*$, the group of integers coprime to n with multiplication as the group operation. During the period of joint research, the second author found two further instances of matched column sets (Classes B and C, the original being Class A), whose proofs were more difficult (Theorems 4 and 6). Also, the first author noticed and proved a further feature regarding the rows of the repetend tables (see Theorem 3).

The Class A instances of the matching columns feature give rise to an infinite sequence of integers denoted $P(L)$, which is the number of pairs of matching columns, dependent on L but not n. We have proved numerous results about $P(L)$, such as severe restrictions on solutions to $P(L) \equiv 0 \bmod 3$ and even more severe on $P(L) \equiv 3 \bmod 6$; these are the subject of a follow-on paper in preparation.

In theory there could exist undiscovered algebraic instances of matched columns outside of Classes A, B and C, or even totally sporadic instances. Section 6 first uses a combination of algebra and computation to exclude specific cases, where the algebra includes Theorem 10 which provides a new formula for the number

of elements of given order in $\mathbb{Z}_n^*$. It then uses probabilistic number theory on the remaining cases, to estimate the likelihood that sporadic instances exist. The resulting estimate is less than two in a million.

The Class B paired columns require events of heuristic probability 2^{-m} to occur, and for some values of m which are still computationally feasible, the size of the successful moduli exceed a natural limit of $2^{26.5}$ below which they are computable with straightforward arithmetic. So to compute with these larger moduli we devised algorithms using a novel hybrid residue number system, and this is described in Appendix A.

2. Theory behind the Features

We derive some general theory first, before the first feature is proved in Lemma 3. All logarithms (written log(.)) in this paper are natural logarithms. Also we use the convention that $x = y \bmod n$ means that $0 \le x < n$, whereas $x \equiv y \bmod n$ is a congruence which implies nothing about the size of x or y.

If r is the repetend considered as an integer in base b^L, then we may write a fundamental equation

$$m/n = \sum_{i=1}^{\infty} r b^{-iL}. \tag{1}$$

The relationship between b, L, and n is described in the following result.

Lemma 1. *The order (or index) of the element b in the group $\mathbb{Z}_n^*$ is L, also typically denoted by* $\mathrm{ord}_n(b)$.

Proof. In Equation (1), the sum is an infinite geometric sum equal to $r/(b^L - 1)$, and hence after cross-multiplying,

$$m(b^L - 1) = nr$$

which implies, since $\gcd(m, n) = 1$, that

$$b^L \equiv 1 \bmod n.$$

Any L satisfying this equation gives a repetend of L digits, but L is defined as the minimum value for this to hold, so $L = \mathrm{ord}_n(b)$. □

Note that the order of $\mathbb{Z}_n^*$ is $\phi(n)$ and so by Lagrange's Theorem L must divide this number.

We now introduce the following notation which will apply throughout the paper.

Notation. Given a base b and an integer x coprime to L, we denote by b_x the small base $b^x \bmod n = b^{x \bmod L} \bmod n$, and by r_x the repetend for m/n as an

integer in base b_x^L, so that the left to right base L digits are $r_{x,0} \dots r_{x,L-1}$, with

$$r_x = \sum_{j=0}^{L-1} r_{x,j} b_x^{L-1-j}.$$

Note that $b_1 = b$ if and only if b is a small base. Also note that we use the index 0 for the first base b digit, whereas the cited papers tend to use 1, that being natural because the first digit is the coefficient of b^{-1}. However, the equations in Lemma 5 and Theorem 2 prove to be simpler if the first index is chosen to be 0, and starting with 0 facilitates mod L arithmetic on the repetend indices. The following defines when b_x has order (i.e., repetend length) L, as proved in the text following Equation (3) in Lewittes [8].

Lemma 2. *The base b_x has order L if and only if* $\gcd(x, L) = 1$.

Lemma 3 *For any modulus n, the number of small bases with repetend length $L > 2$ is even.*

Proof. Since $L - 1$ is coprime to L, by Lemma 2, for any small base b of order L there exists another small base b_{L-1} with that order. If these two bases are unequal then the bases are paired and therefore the number of them is even. The condition that they be unequal is, by Lemma 1, that $b^{L-1} = b^{-1} \neq b \bmod n$, or $b^2 \neq 1 \bmod n$, which is true for b of order $L > 2$. □

Example 1. The modulus $n = 37$ has one small base (36) with $L = 2$, two bases (10 and 26) with $L = 3$, two bases each with $L = 4$ and 6, four bases with $L = 12$, six bases each with $L = 9$ and 18, and twelve bases with $L = 36$.

Since this example has prime n, it may be noticed that in each case the number of small bases is $\phi(L)$ – see the book by Gallian [3, Theorem 4.4] for details.

Lemma 4. *If $\mathbb{Z}_n^*$ is cyclic then the number of small bases which have repetend length L is $\phi(L)$.*

We shall need the following theorem which is due to Gauss and is proved in, for example, Ore [11].

Theorem 1. *The group $\mathbb{Z}_n^*$ is cyclic if and only if n equals 2, 4, p^m, or $2p^m$, where p is any odd prime and m is any positive integer.*

Corollary 1. *If n is prime, and there is one small base which produces a repetend length L, then there are $\phi(L) - 1$ other bases with that length, where ϕ is Euler's totient function.*

Proof. Since n is prime, Theorem 1 implies that the group is cyclic and then the result follows from Lemma 4. □

Example 2. Modulus $n = 11$ has $4 = \phi(5)$ small bases with repetend length $L = 5$. These are: $3(\text{repetend } 1/11 = 0.\overline{00211}_3), 4(1/11 = 0.\overline{01131}_4), 5(1/11 = 0.\overline{02114}_5)$, $9(1/11 = 0.\overline{07324}_9)$. This accords with Lemma 4.

We now study the generation of base b repetend digits r_j by long division of m by n. The following lemma is vital to the proofs of many theorems below, yet it seems to be novel in the literature on repetends (see Section 5 for more information).

Note that to compute the quotients r_j in the long division, the lemma also computes the remainders $c_{j+1} < n$.

Lemma 5. *If $m/n = E_k = \sum_{i=0}^{k-1} r_i b^{-i-1} + c_k/(nb^k)$ for $k = 0, \ldots, \infty$, then a solution to this is*

$$c_j = mb^j \bmod n,\ r_j = (bc_j - c_{j+1})/n = \lfloor bc_j/n \rfloor \text{ for each } j \geq 0.$$

Proof. We claim that the equation $bc_j/n = r_j + c_{j+1}/n$ holds. It may be derived either by appealing to the well known long division algorithm in which the quotients and remainders are derived successively by multiplying the preceding remainder c by b and then dividing by n, or formally by simple manipulation of the equation $E_{j+1} = E_j$.

Since $m/n = E_0 = c_0/n$, then $c_0 = m$. Then by choosing r_j in the equation above to be the integer part of bc_j/n, each $c_{j+1} \in \{0, \ldots, n-1\}$. Also, by multiplying the equation by n, we have $bc_j = r_j n + c_{j+1},$ so $c_{j+1} = bc_j \bmod n$.

Since $c_0 = m$, then $c_j = mb^j \bmod n$ by induction. Then the solution for the r's is

$$r_j = bc_j/n – c_{j+1}/n = \lfloor bc_j/n \rfloor \in \{0, \ldots, b-1\}.$$

□

Note also that if $m = 1$, then $c_j = b_j$ according to the prior notation for b_j. The first form of r_j above will prove crucial for proving various results below.

Further below we shall present matrices of repetends with common m/n but to different bases b_x. If we let $B = b_x$, since $B_y = b_x^y \bmod n = b^{xy} \bmod n = b_{xy \bmod L}$, then in Lemma 5, the replacements $b \to B \to b_x$ make b_j become b_{xj}, so

$$r_{x,j} = (b_x c_{xj} - c_{x(j+1)})/n = \lfloor b_x c_{xj}/n \rfloor. \tag{2}$$

We call this the *repetend digit formula* (or RDF for short), noting that if $m = 1$ then c can be replaced by b.

The following corollary explains a feature which is noticeable when n and L are both prime, namely a correspondence between the remainders and the small bases.

Corollary 2. *If n and L are both prime, and if $m = 1$ or m is a base with that repetend length, then the set of the remainders c_j (excluding 1) is identical to the set of all the bases for that length.*

Proof. By Theorem 1 $\mathbb{Z}_n^*$ is cyclic, so b in Lemma 5 generates the unique cyclic subgroup of order L. By assumption, m is in that subgroup and therefore equal to b^u for some u. Then each $c_j = mb_j = b_{u+j}$, and apart from $c_{L-u} = 1$ is a small base of order L. Thus the set of remainders c_j excluding 1 is the same as the set of small bases b_j. □

Corollary 3. *For a given position in the repetend, if a base is repeatedly increased by n, then the digit there increases each time by a constant which is less than n.*

Proof. If $b = an + s$ with $0 < s < n$ then $r_j = \lfloor sc_j/n + ac_j \rfloor = \lfloor sc_j/n \rfloor + ac_j$ by Lemma 5, where $c_j \equiv m(an+s)^j \equiv ms^j \bmod n$. So for each increment by 1 of a, the repetend digit r_j increases by $c_j < n$ (thereby giving an arithmetic progression). □

Example 3. Illustrating Corollary 3, in Table 1 the four given repetends of 1/7, including the classic $0.\overline{142857}$... for base 10, have values within each column j which increase by $3^j \bmod 7 = 1, 3, 2, 6, 4, 5$, respectively, as j runs from 0 to 5.

		j					
a	b	0	1	2	3	4	5
0	3	0	1	0	2	1	2
1	10	1	4	2	8	5	7
2	17	2	7	4	14	9	12
3	24	3	10	6	20	13	17

Table 1: The repetends of 1/7 to bases 3, 10, 17, 24

We now consider the "matched column sets" feature, which consists of relationships between repetends from different bases, starting with an example calculation.

Example 4. Use Equation (2) to compute $r_{2,5}$ for $n = 29,\ m = 1,\ L = 7,\ b_1 = 7$: $x = 2,\ b_2 = 20,\ j = 5,\ xj = 3 \bmod 7,\ c_3 = 24,\ r_{2,5} = \lfloor 20{\cdot}24/29 \rfloor = \lfloor 16.5517 \rfloor = 16.$

That value 16 can be seen in Row 2, Column 5 of Table 2. Using the notation, with 'a' to represent 10 etc., the repetends from Rows 1 and 2 are $0.\overline{0145536}$... and $0.\overline{0dfh4gb}$..., respectively. Table 3 similarly gives data for $n = 31,\ L = 15$.

Now the "matched column sets" feature is the rather remarkable observation that in numerous cases the multiset of integers in one column (say j) is identical to the multiset in another column (say i). This can be seen in Table 2, where the columns which match are 2 with 4 and 3 with 5; in Table 3 the column matches are 2:8 and 7:13.

Yet the columns are permutations of each other rather than being identical, and this is somewhat remarkable because the bases, which constrain the size of the elements in each row, are different. So, for example, the 5 in Column 4 of Table 2 is necessarily less than $b_1 = 7$, but is matched by the 5 in Column 2, which is only forced to be less than $b_4 = 23$.

					j			
x	b_x	0	1	2	3	4	5	6
1	7	0	1	4	5	5	3	6
2	20	0	13	15	17	4	16	11
3	24	0	19	20	16	13	5	19
4	23	0	18	5	12	15	19	19
5	16	0	8	13	3	13	12	11
6	25	0	21	13	19	20	17	6

Table 2: $r_{x,j}$ for $n = 29,\ m = 1,\ b = 7,\ L = 7,\ \phi(L) = 6$, column matches 2:4, 3:5

									j							
x	b_x	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	7	0	1	4	0	3	1	0	6	2	1	5	4	3	4	2
2	18	0	10	8	2	5	14	9	5	4	1	2	16	4	11	11
4	14	0	6	4	7	3	2	3	8	8	1	11	4	0	12	9
7	28	0	25	8	3	17	4	14	12	18	1	22	16	7	6	9
8	10	0	3	2	2	5	8	0	6	4	5	1	6	1	2	9
11	20	0	12	18	1	5	16	2	11	12	5	3	4	10	6	9
13	19	0	11	12	4	17	3	1	4	5	9	15	6	2	8	11
14	9	0	2	5	4	5	7	2	2	8	1	1	4	0	5	2

Table 3: $r_{x,j}$ for $n = 31,\ b = 7,\ m = 1,\ L = 15,\ \phi(L) = 8$, matches 2:8, 7:13

The explanation for this feature of matched column sets, as in the title of this paper, is given in Theorem 2 below, and the specification of which pairs of columns match lies in the following lemma.

Lemma 6. *If $m = 1$, and x and j are coprime to L, then $r_{xj,1/j} = r_{x,j}$, where subscripts are computed mod L.*

Proof. The inverse of j mod L exists because j is coprime to L, and Equation (2) with $c_y = b_y$ (as $m = 1$) implies

$$r_{xj,1/j} = \lfloor b_{xj} b_{xj(1/j)}/n \rfloor = \lfloor b_{xj} b_x/n \rfloor = r_{x,j}. \qquad \square$$

Thus it is columns j and j^{-1} mod L which match.

Remark 1. This relationship is easy to see in Tables 1 and 2; but our original research counted columns 1-up instead of 0-up, which made the relationship harder to spot!

Theorem 2. *Let n be an integer, $\mathcal{B}$ be the set of small bases giving a repetend length L for $1/n$, and A be the matrix whose rows are the repetend digits r_j of $1/n$*

for each base $b \in \mathcal{B}$, with column index j and $0 \le j < L$. Then if j is coprime to L and $j^2 \not\equiv 1 \bmod L$, column j matches distinct column $1/j \bmod L$. If L is the product of 2^e and f distinct odd prime powers then the number of pairs of matched columns is $P(L) = (\phi(L) - 2^{f+g(e)})/2$ where $g(e) = \max(0, \min(2, e-1))$. Further, the final column $(L-1)$ possesses pairs of identical repetend digits.

Proof. Consider any row labelled by its base $b = b_1 \in \mathcal{B}$. The digit in column j of this row is $r_{1,j}$, which by Lemma 6 and the assumptions on j, equals $r_{j,1/j}$ in the distinct column $1/j \bmod L$ and the row labelled by $b_j = b_1^j \bmod n$. Hence there is a 1-1 match between the digits in the distinct columns j and $1/j \bmod L$.

The number of pairs of matching columns is half the number of eligible j values. Excluded from these are the square roots of unity mod L which form a group to whose order each of the f odd prime divisors of L contributes a factor of 2, and a divisor 2^e contributes the factor 1, 1, 2, and 4, respectively, for the cases $e = 0,\ e = 1,\ e = 2,\ e \ge 3$. Hence the number of matching pairs of columns is

$$P(L) = (\phi(L) - 2^{f+g(e)})/2 \text{ where } g(e) = \max(0, \min(2, e-1)). \tag{3}$$

Finally, if $j = L - 1 = -1 \bmod L$, then $1/j = 1/(-1) = -1 = j \bmod L$, so by Lemma 6, $r_{1,L-1} = r_{L-1,L-1}$. Thus column $L-1$ has pairs of equal repetend digits. □

$P(L)$, the number of paired columns for any given repetend length L, is an infinite sequence whose properties could be of interest. It is the subject of a paper currently in preparation, which contains interesting information about $P(L)$, such as proofs that its range excludes "most" 3 mod 6 integers.

Remark 2. If the matrix A in Theorem 2 is restricted to the rows corresponding to the cyclic subgroup formed by increasing powers of a single base b, then the last column is palindromic. This is visible in Tables 2 and 3. Example 7 is the only one with proper subgroups, and shows how the column matching and palindromic feature occur within the subgroups.

Example 5. In Theorem 2 take $n = 29,\ L = 7,\ b_1 = 7,\ j = 5$, so $1/j = 3 \bmod 7$, which shows that Columns 5 and 3 of Table 2 match. Take $x = 4$, so $xj = 6 \bmod 7$, then $r_{4,5} = r_{6,3} = \lfloor b_4 b_6/29 \rfloor = \lfloor 23 \cdot 25/29 \rfloor = 19$. For the number of paired columns, $L = 2^0 7^1$, so $f = 1$ and $e = 0$, and since $\phi(L) = 6$, we have $P(L) = (6-2)/2 = 2$.

Example 6. In Theorem 2 take $n = 31, L = 15, b_1 = 7, j = 7$, so $1/j = 13 \bmod 15$, which shows Columns 7 and 13 of Table 3 match. Take $x = 7$, so $xj = 4 \bmod 15$, then $r_{7,7} = r_{4,13} = \lfloor b_7 b_4/31 \rfloor = \lfloor 28 \cdot 14/31 \rfloor = 12$. For the number of paired columns, $L = 2^0 3^1 5^1$, so $f = 2$ and $e = 0$, and since $\phi(L) = 8$, we have $P(L) = (8-4)/2 = 2$.

Note that in Table 3, in addition to the palindromic Column 14 from Theorem 2, Columns 4 and 11 are also interesting, each having just three unique entries out

of the eight possible. Part of the explanation is that $4^2 = 11^2 = 1 \bmod 15$. So in Column 4, $r_{x,4} = \lfloor b_x b_{x4}/31 \rfloor = \lfloor b_{4x} b_{1x}/31 \rfloor = \lfloor b_{4x} b_{(4x)4}/31 \rfloor = r_{4x,4}$. That explains why the eight values are split into four pairs. The reason why 5 occurs four times instead of twice is that $b_3 = 343 = 2 \bmod 31$, so if $y = x + 3$ then $b_y = 2b_x$, leading to, for example, $r_{8,4} = \lfloor b_8 b_2/31 \rfloor = \lfloor 10 \cdot 18/31 \rfloor = \lfloor 20 \cdot 9/31 \rfloor = \lfloor b_{11} b_{14}/31 \rfloor = r_{11,4}$.

The next example illustrates some features in the case of non-cyclic $\mathbb{Z}_n^*$.

Example 7. In Table 4 there are 24 bases of order 5 which split into 6 distinct multiplicative subgroups of order 4. Columns 2 and 3 match, and it can be seen that the matching digits occur within each subgroup.

b	r_0	r_1	r_2	r_3	r_4	b	r_0	r_1	r_2	r_3	r_4
16	0	0	14	14	5	26	0	2	11	23	19
256	0	238	80	14	229	126	0	57	92	11	115
246	0	220	14	76	229	251	0	229	23	183	115
86	0	26	76	80	5	201	0	146	183	92	19
31	0	3	15	10	8	36	0	4	25	23	25
136	0	67	35	15	45	196	0	139	136	25	129
91	0	30	10	23	45	181	0	119	23	125	129
71	0	18	23	35	8	191	0	132	125	136	25
56	0	11	22	33	45	81	0	23	69	41	43
111	0	44	89	22	67	236	0	202	125	69	121
166	0	100	33	133	67	141	0	72	41	74	121
221	0	177	133	89	45	146	0	77	74	125	43

Table 4: The 24 repetends for $n = 275 = 5^2 \cdot 11$ with $L = 5$

The next example was chosen for two reasons: to use the "nice" base 10, and to have $L = 25$ which is the only value to achieve $P(L) = 9$ (and as mentioned earlier is a rare 3 mod 6 value of $P(L)$). Consequently, n has to be quite large, both a divisor of $10^{25} - 1$ and with 10 of full order 25, and from Table 9 of [13] the smallest such is the prime 21401.

This choice was serendipitous in the sense that it led to the discovery of a new feature of matching digits across rows (as opposed to columns). It was noticed because of the presence of unusually small numbers in each row, even when the limit, the value of the base, is quite large.

Example 8. With the parameters as in Table 5, the $20 = \phi(25)$ bases $10^i \bmod n$ with $5 \nmid i$ are 10, 100, 1000, 10000, 15554, 5733, 14528, 16874, 18122, 10012, 14516, 16754, 6122, 18418, 12972, 1314, 2994, 8539, 21187, 19261. The smallest value outside Column 0 has been underlined in each of the Rows 3 to 24, and an example of a large matching pair is shown in italics in Rows 17 and 18.

**	00000	00001	02*13	03*17	04*19	00005	06*21	07*18	08*22	09*14	00010	11*16	12*23
01	00000	00000	00000	00000	00004	00006	00007	00002	00006	00007	00008	00008	00004
02	00000	00000	00046	00072	00067	00088	00046	00078	00028	00060	00061	00039	00090
03	00000	00046	00726	00788	00467	00828	00606	00139	00900	00004	00672	00678	00846
04	00000	04672	06788	04678	02860	06139	09000	00467	02678	08467	08286	00613	09900
06	00000	11304	07276	09427	13998	10462	13170	13385	15398	07267	13760	04449	06206
07	00000	01535	04488	00802	00267	05071	04933	05159	04166	03888	03520	00026	04520
08	00000	09862	04155	13075	03891	12037	14382	10558	11373	05796	09772	09854	02032
09	00000	13304	10228	00078	14288	10360	07884	11445	06732	12263	13981	15186	11454
11	00000	15345	07230	12302	01112	12190	05184	00084	12291	16309	16032	02535	04854
12	00000	04683	09010	08477	09911	08857	03994	07894	01400	06796	06147	02682	00614
13	00000	09846	00006	11364	00067	12028	00678	04152	06782	12492	09764	08798	10550
14	00000	13116	00782	14418	12176	10286	13209	16586	07838	00007	13882	07828	10155
16	00000	01751	01639	06060	04792	04118	00856	02864	00286	00375	05416	00002	05268
17	00000	15850	14522	00008	*11163*	16294	00086	01130	15596	00860	11308	08616	08606
18	00000	07862	10984	06061	05175	10748	08806	00006	00796	06068	08725	12842	03710
19	00000	00080	00891	00352	00000	00806	01028	00892	00006	00183	01088	01036	00061
21	00000	00418	02576	02030	02360	02014	00001	01194	01814	02343	02648	02176	00013
22	00000	03407	00524	02442	05791	07554	02287	03990	00003	08453	05242	07348	06684
23	00000	20975	02964	01300	18233	17555	14370	17940	16705	05675	14252	00990	00009
24	00000	17334	19068	07685	02694	11826	01182	11674	16576	05509	15959	15078	13064

**	13*02	14*09	00015	16*11	17*03	18*07	19*04	00020	21*06	22*08	23*12	00024
01	00006	00007	00008	00002	00008	00006	00000	00006	00001	00003	00009	00009
02	00000	00004	00067	00026	00078	00084	00067	00082	00086	00006	00013	00099
03	00782	00860	00613	00990	00000	00467	00267	00884	00678	00286	00061	00399
04	00046	07267	08846	07828	06061	03990	00004	06726	07884	06782	08606	01399
06	00726	12263	12888	02176	00072	10558	12176	09550	00007	04166	10550	00955
07	01639	05675	03856	02682	00352	00002	03891	04750	02287	02678	04854	03475
08	06788	06796	08920	00678	12302	00892	00067	12852	08806	00006	11454	12503
09	14522*	00007	14928	01036	00788	07894	02360	11350	13209	16705	04520	04827
11	10984	08467	15016	00008	08477	17940	14288	11126	13170	15596	00846	14187
12	07276	06068	06734	08616	04678	02864	00467	08296	00046	07838	00004	06791
13	00891	03888	08912	09854	02030	11445	05791	12841	14370	12291	13064	06791
14	04488	02343	14821	15078	11364	00078	04792	11270	01028	11373	06684	14187
16	04155	05509	05072	04449	02442	04152	02860	03758	05184	00028	03710	04827
17	08606	02576	12492	12389	07348	14418	13385	18233	15261	04933	16576	05268
18	10228	00060	07964	08798	09427	11674	*11163*	11476	00606	01814	10155	03475
19	00524	00375	01162	00613	01300	01130	01112	00883	01182	00796	00614	00955
21	02964	00183	02480	02535	00802	00139	02694	01838	00856	01400	02032	01399
22	07230	05796	05744	00039	00685	01194	05175	07075	03994	06732	06206	00399
23	19068	08453	13008	12842	06060	16586	09911	18743	14382	15398	09900	00099
24	09010	16309	17039	15186	13075	05159	13998	12956	09000	00900	00090	00009

Table 5: $n = 21401,\ L = 25,\ b_1 = 10$; leading zeroes used; $x * y$ means the column is x and is paired with y

Each smallest value is less than 10, a feature whereby each row has a match with Row 1, and more generally with any other row, as in the following theorem. Whereas Theorem 2 requires the numerator m to be 1, for row matching the constraint on m is less severe, as follows.

Theorem 3. *If m is a power of a base, then each pair of rows has an element which is common to the two rows. Further, in each row there are two adjacent digits, the first being zero, which are not part of the matching pairs.*

Proof. Since m is a power of a base for the repetend length, let the base be b and $m = b^u \bmod n$. Then in Lemma 5 $c_j = b^{j+u} \bmod n = b_{j+u}$ by definition, and by Equation (2) with $c_{xj} = b_{xj+u}$, we have $r_{x,j} = \lfloor b_x b_{xj+u}/n \rfloor$. To make $r_{x,j}$ equal $r_{y,i}$ by matching the "opposite" subscripts, we need $x = yi + u$ and $y = xj + u$. Since the row numbers x and y are already required to be coprime to L, there is a solution $i = (x - u)/y \bmod L$ and $j = (y - u)/x \bmod n$.

Now, given a fixed row x, the other row index y may not be 0 nor x. The equation for j then means it cannot be $-u/x \bmod L$ nor $(x-u)/x = 1+(-u/x) \bmod L$, which are two adjacent columns. The digit at the first is $\lfloor b_x b_{x(-u/x)+u}/n \rfloor = \lfloor b_x/n \rfloor = 0$. □

Example 9. Table 6 shows an example where the numerator m is not unity. The

					j			
x	b_x	0	1	2	3	4	5	6
1	7	6_a	0	**1**	4_b	5_c	5_d	3_e
2	20	17_f	4_b	16_g	11_h	0	13	15_i
3	24	20_j	16_g	13_k	5_c	19_l	0	19
4	23	19_m	19_l	0	**18**	5_d	12_n	15_i
5	16	13_o	12_n	11_h	0	**8**	13_k	3_e
6	25	**21**	13_o	19_m	20_j	17_f	6_a	0

Table 6: $r_{x,j}$ for $n = 29,\ b = 7,\ m = 25 \equiv b^6,\ L = 7$

15 matching digits are subscripted by equal letters, from a to o and ordering left to right and top to bottom. The four unique digits (following 0 in each case) are written in bold.

3. Class B: A Second Class of Column Matches

Theorem 2 provides sufficient conditions for column matches to occur. Regarding necessity, non-systematic computer experiments had suggested that two converses of Theorem 2 are true. The weaker converse is that a third column cannot match

the pair of columns j and $1/j$, and the stronger converse is that a pair i, j matches only if $ij = 1 \bmod L$. However, a more systematic search contradicted the stronger converse, by revealing two new classes of column match which we shall call Class B and Class C (with Class A used for examples under Theorem 2 above). All three classes are based on algebraic identities, and in fact both Class B and Class C depend on the cyclotomic polynomial $x^2 - x + 1$ dividing $x^6 - 1$. However, that does not preclude the possibility of a "sporadic match" in which random numerical close approximations cause a match. Section 6 presents a heuristic probability model which suggests that the probability of any sporadic matches is low – less than two in a million.

The definition of *Class B* is as follows. Let $L = 12a + 6$ with $a > 0$, and n be a prime power with the prime congruent to 1 mod L. Then a certain proportion of repetends of $1/n$ with length L have columns $2a + 1$ and $4a + 2$ matching and columns $8a + 4$ and $10a + 5$ matching (as proved later).

Example 10. The smallest Class B example is given in Table 7, with $n = 199$.

		j																
x	b_x	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	19	0	1	15	8	16	13	8	14	16	18	17	3	10	2	5	10	4	2
5	141	0	99	127	75	114	110	75	14	123	140	41	13	65	26	30	65	126	17
7	156	0	122	45	72	141	16	72	18	126	155	33	110	83	14	139	83	137	29
11	37	0	6	32	19	33	3	19	26	7	36	30	4	17	3	33	17	10	29
13	24	0	2	21	11	5	4	11	2	6	23	21	2	12	18	19	12	21	17
17	21	0	2	4	11	6	2	11	3	18	20	18	16	9	14	18	9	17	2

Table 7: The six repetends of 1/199 with $L = 18$

Note that the Class B matching Columns 3:6 and 12:15 have their elements in exactly the same order, i.e., $r_{x,3} = r_{x,6}$. The Class B theorem below explains this.

Remark 3. In Class B, $a = 0$ could in theory be a valid case, except that $L = 6$ is closely analyzed in Section 6, where Lemmas 40 and 39 show that for any n, Columns 1:2 and 4:5 do not match.

Theorem 4 (Class B)**.** *Let $L = 12a + 6$ with $a > 0$, n be a prime power with the prime congruent to 1 mod L, and b be a small base giving repetend length L. Let $b_i = b^i \bmod n$ and $k_i = (b_i + b_{i+L/3} + b_{i+2L/3})/n$. Then if (and only if) $k_i = 1$ for all i coprime to L, columns $2a+1$ and $4a+2$ match and columns $8a+4$ and $10a+5$ match. The proportion of such n's is conjectured to be 2^{-s}, where*

$$s = \begin{cases} \phi(L)/3 \text{ if } a \equiv 1 \bmod 3 \\ \phi(L)/2 \text{ otherwise} \end{cases}.$$

Proof. Let $d = L/6 = 2a+1,\ e = 2d$. Since $b_1 = b$ has order L, b_d has order 6 and, since n is a prime power, is a zero of the 6^{th} cyclotomic polynomial x^2-x+1 modulo n. Hence $b_d^2 - b_d + 1 = 0 \bmod n$. Then $b_e \equiv b_d^2 \equiv b_d - 1 \bmod n$, so $b_e = b_d - 1$. Since b_{-d} is the other element of order 6 and b_{-e} is the other element of order 3, $b_{-e} = b_{-d} - 1$ in similar fashion and the following analysis for the pair (d, e) also applies to the pair $(-d, -e) \equiv (L-d, L-e) \equiv (10a+5, 8a+4)$.

Now let $D_i = r_{i,d} - r_{i,e} = (b_i b_{di} - b_{(d+1)i} - b_i b_{ei} + b_{(e+1)i})/n$, by Equation (2). Since i is coprime to L and hence to 6, b_{di} has order 6, so $b_{ei} = b_{di} - 1$ as above. Then $nD_i = b_i b_{di} - b_{(d+1)i} - b_i(b_{di}-1) + b_{(e+1)i} = b_i - b_{(d+1)i} + b_{(e+1)i} = 0$ if and only if $b_i + b_{i+ei} < n$. The last inference there is true because $b_{(d+1)i} \equiv b_i + b_{(e+1)i} \bmod n$ is forced, but if the RHS is greater than n then $b_{(d+1)i} < n$ implies D_i has to be 1.

Now, since $e = L/3$, by definition $k_i = (b_i + b_{i+e} + b_{i+2e})/n$, an integer since $b_i + b_{i+e} + b_{i+2e} \equiv b_i(1 + b_e + b_{2e}) \equiv b_i(1 + b^e + b^{2e}) \equiv b_i(b^{3e}-1)/(b^e-1) \equiv 0 \bmod n$, as $b^{3e} \equiv b^L \equiv 1 \bmod n$. Since i is coprime to L, the set $\{(e+1)i, (2e+1)i\} \bmod L$ equals $\{i + (i \bmod 3)e, i + (2i \bmod 3)e\} \bmod L$, which is $\{i+e, i+2e\} \bmod L$. Therefore $k_i = (b_i + b_{(e+1)i} + b_{(2e+1)i})/n$, and so

$$D_i = (b_i - b_{(d+1)i} + b_{(e+1)i})/n = k_i - b_{(d+1)i}/n - b_{(2e+1)i}/n.$$

Since D_i and k_i are both integers and the b_*/n terms are each less than 1, $D_i = k_i - 1$.

Define a "triple" T_i to be $\{i, i+e, i+2e\} \bmod L$. Then if $T_{i_1} \dots T_{i_s}$ is a minimal set of triples which covers all values i coprime to L, then all the D_i's are 0 if and only if all s of the k_{i_j}'s are 1.

This s is the same as in the statement of the theorem, and we now prove the claim about its value. We need to know how many members of T_i are coprime to L, given that member i is. If $i + je = i + jL/3$ is *not* coprime to L, then for some prime $p,\ p \mid L,\ p \mid i + jL/3,\ p \nmid i$. So $jL/3 \equiv (i + jL/3) - i \equiv 0 - i \not\equiv 0 \bmod p$, so $p \nmid L/3$. But since $p \mid L$, the only possibility is $p = 3$ and $9 \nmid L = 6(2a+1)$, hence $a \not\equiv 1 \bmod 3$. In this case $j \equiv -i/(4a+2) \bmod 3$ specifies the unique $i + je \bmod L$ which does not contribute to covering the $\phi(L)$ values coprime to L. Hence T_i covers 3 values coprime to L if $a \equiv 1 \bmod 3$ and 2 otherwise, and the number s of them is as claimed since there are $\phi(L)$ values to cover.

The heuristic probability that a random eligible n is in Class B is 2^{-s} for the following reason. Each k_i can take only the values 1 or 2; $k_i = 1$ if and only if $b_i + b_{i+e} < n$, since b_{i+2e} is determined by those two values. A heuristic model that b_i and b_{i+e} are independently and uniformly distributed between 2 and $n-2$ implies that their sum exceeds n with probability 1/2. Hence, over the set of s independent triples there is a 2^{-s} heuristic probability that all the k_i's are 1. □

Table 8 displays details of Class B cases for a such that $s \le 24$, ordered by s. (The first missing a, 14, has $s = \phi(174)/2 = 28$.) For each a, the target number H of hits was chosen to be as large as possible up to 10000 without taking too long to run, in order to help test the statistics. Up to 5 a's were run in parallel, consuming about 11% each of available CPU power on a modern 6-core laptop. (A single run at a time consumed about 15% CPU.) Therefore the number of prime candidates

was $2^s H$. The proportion h/H of actual hits to expected, should be close to 1. In fact it was always slightly below 1 for $8 \le s \le 20$, but not significantly so against a Poisson distribution of mean H. The last four rows achieved 6 hits (expected 6); fortunately the expensive $s = 24$ runs did not require H to be raised above 1.

It is more efficient not to test for prime power examples, but where we did for $a = 1$, the following prime powers exhibited Class B:

$$37^2,\ 163^2,\ 37^3,\ 19^4,\ 631^2,\ 757^2,\ 919^2,\ 991^2,\ 1117^2,\ 37^4,\ 127^3.$$

The largest successful n contained in a search starting from zero was 27544840363, which is a 35-bit number, with modular multiplication in theory requiring use of numbers up to 70 bits in length. Fast enough computation in 'R' was a challenge whose details are described in Appendix A.

a	s	H	h	h/H	time	smallest	largest
1	2	10000	9855	0.985	29.56s	199	3347947
2	4	10000	10156	1.016	2.55m	2131	20175751
3	6	10000	9839	0.984	21.54m	5503	135647989
4	6	10000	10072	1.007	31.64m	12421	208531261
7	8	1600	1584	0.990	29.68m	258211	176238001
5	10	400	384	0.960	20.44m	447877	145087867
6	12	100	90	0.900	25.22m	855271	174542629
10	12	100	97	0.970	28.12m	1529893	269612029
8	16	100	89	0.890	13.68h	20260873	4418602159
9	18	25	23	0.920	14.65h	405145171	4597467241
13	18	25	21	0.840	16.82h	77221999	7552690129
12	20	10	9	0.900	1.35d	430129351	8869214251
16	20	10	9	0.900	1.29d	236180143	11140938271
11	22	3	1	0.333	1.43d	2980689601	2980689601
17	24	1	1	1.000	2.72d	3835537861	3835537861
19	24	1	3	3.000	2.31d	2285793901	27544840363
22	24	1	1	1.000	2.56d	11152477711	11152477711

Table 8: Class B statistics, with a = class B index, $s = -\log_2(\text{hit rate})$, H = target number of hits, h = actual number of hits

4. Class C: A Third Class of Column Matches

For this section we introduce some particular notation and constraints which encapsulate a third Class (C) of matching columns. In addition, we explain how a number of pairs of "almost matching" columns arise, labelled Class $\tilde{\text{C}}$.

The definition of *Class C* is as follows. With n and L as specified in Table 9, there exists a non-Class A pair of matching columns with particular indices d, e.

Variable	Definition
p	is a prime congruent to 1 mod 6
j	is a positive integer
i	is an integer $\geq 2j$
n	$= p^i$ is the modulus
Q	$= p^j$, (so 1 mod 6)
a	$= (Q-1)/6$
L	$= 6Q$ is the repetend length
b, B	are small bases for which n has repetend length L
l, h	are integers arising in $B^2 = B - 1 + lp^h \bmod n$ proved in Lemma 8
M	$= lp^h$
d, e	are the indices of a pair of matching (or near-matching) columns
y	is an integer which is ± 1 mod 6
s	$= y$ mod 6 with $s = \pm 1$, for brevity sometimes written as $+$ or $-$
m_s	is a multiplier describing the matching of columns d and e
D_s	$= n(r_{m_s y,e} - r_{y,d})$ measures the difference in column d and e digits
(u, v)	$\bmod(Q, 6)$ is the unique number in $[0, L) = u$ mod Q and v mod 6

Table 9: Notation and constraints for Class C

Remark 4. It is evident from the notation in Table 9, with 6 and $B^2 \equiv B - 1$ appearing, that like Class B, Class C instances are possible through the nature of the 6^{th} cyclotomic polynomial. However, Class B and Class C cannot occur for the same n, because the former requires that $n \equiv 1 \bmod L$, which in the latter requires that $p^i \equiv 1 \bmod 6p^j$, which implies that p divides $p^i - 1$, which is impossible.

Example 11. This smallest example of Class C/$\tilde{\text{C}}$ has $p = 7,\ i = 2,\ j = 1$. The repetend length is $L = 42$ and the smallest base for this is $b = 3$. In the partial repetend Table 10 we display the pairs of columns $d : e = 16 : 4,\ 10 : 40$, and $8 : 19$, first in straight order r_{x*} and then in sorted order (denoted by 's' in the column heading), with second column differences accentuated by underlining.

Columns 16:4 are the pure Class C case, with equal sorted columns. Columns 10:40 also have $d \equiv e \equiv 4 \bmod 6$ as explained later, and are a Class $\tilde{\text{C}}$ case with 2 differences of +1, 8 of 0, and 2 of -1. Columns 8:19 are another Class $\tilde{\text{C}}$ case, with $d = Q + 1$ as explained later, and have 6 +1 differences (i.e., half of the rows), which accords with Theorem 8 below.

The background to Class C/$\tilde{\text{C}}$ is as follows. In Class A, for any two matching columns there is a linear relationship between the row of one column and the matching row in the other column. The same is true for Class C, except that the relationship differs depending on a certain ± 1 mod 6 value. Specifically, if $r_{y,d}$ (almost) matches $r_{x,e}$, then $x = m_s y \bmod L$ where $y \equiv s = \pm 1 \bmod 6$, and the

x	b_x	16	4	10	40	8	19	16s	4s	10s	40s	8s	19s
1	3	1	1	0	0	2	2	35	35	42	42	26	26
5	47	22	15	42	35	10	11	30	30	35	35	21	22
11	12	10	2	0	3	7	8	22	22	24	24	20	21
13	10	2	0	6	5	3	4	15	15	22	23	12	13
17	26	8	1	4	23	2	2	15	15	19	18	10	11
19	38	3	30	35	24	1	13	10	10	13	13	7	8
23	40	30	35	13	18	20	21	8	8	6	6	4	4
25	17	15	8	3	13	12	3	3	3	4	5	4	4
29	5	0	3	2	0	4	4	2	2	3	3	3	3
31	45	35	10	22	42	21	2	1	1	2	1	2	2
37	24	15	22	19	1	4	22	1	1	0	0	2	2
41	33	1	15	24	6	26	26	0	0	0	0	1	2

Table 10: Six columns (and sorted columns) of the repetend table for $1/49$, $L = 42$

multiplier depends on s. (Note from the example that y must be coprime to 6 and p must be a base giving repetend length L.) We now calculate $D_s = n(r_{m_s y,e} - r_{y,d})$ in the following lemma.

Lemma 7. *Let $B = b_y$. Then*

$$\begin{aligned} D_s &= b_{ym_s} b_{ym_s e} - b_{ym_s(e+1)} - b_y b_{yd} + b_{y(d+1)} \\ &= B_{m_s} B_{m_s e} - B_{m_s(e+1)} - B_1 B_d + B_{d+1}. \end{aligned}$$

Proof. This follows from two uses of the Repetend Digit Formula (2) with different parameters. □

It is clear that the magnitude of D_s is in general dominated by its first and third terms, which are products of two B_k's. The second and fourth terms are instrumental in determining whether there is an exact match, i.e., $D_s = 0$.

Here is a summary of the remainder of this section. We first prove some basic results in Lemmas 8 to 11, which lead to the proofs of Class C and $\tilde{\text{C}}$ statements regarding the relationships of e and m_s to d which provide small values of D_s. These only occur when $d \equiv \pm 2 \bmod 6$, and are specified in Lemmas 12, 14 and 15. Theorem 5 then enumerates the Class C/$\tilde{\text{C}}$ cases, and Theorem 6 derives the single special Class C case, i.e., $D_s = 0$, from Lemma 12. Finally, Theorems 7 and 8 explain why, for four of the Class $\tilde{\text{C}}$ cases, the difference D_s/n in the repetend digits is $+1$ in half the cases and 0 in the other half.

Lemma 8. *Base B satisfies $B^2 = B - 1 + M \bmod n$, where $M = lp^h$ with $h = i - j$ and l coprime to p. Also, $n \mid QM \mid M^2$ (also implying $Q \mid M$).*

Proof. The notation implies that B has order $6Q \bmod n$. Let $B^6 \equiv a \bmod p^i$, so a has order $Q = p^j$. Then $1 \equiv B^L \equiv a^{p^j} \bmod p^i$ by the notation. Hence $a^{p^j} \equiv 1 \bmod p$, but by Fermat's Little Theorem and induction, $a^{p^j} \equiv a \bmod p$, so $a \equiv 1 \bmod p$. Now we may write $a = 1+kp^h$ for some $h > 0$ and $k \not\equiv 0 \bmod p$. Then $a^{p^j} \equiv (1+kp^h)^{p^j} \equiv 1+kp^{h+j} \bmod p^{2h+j}$, which is consistent with $a^{p^j} \equiv 1 \bmod p^i$ if and only if $h \geq i-j$, which is ≥ 1. But the order of a implies that $a^{p^{j-1}} \not\equiv 1 \bmod p^i$, so $1 \not\equiv 1+kp^{h+j-1} \bmod p^{2h+j-1} \equiv 1+kp^{h+j-1} \bmod p^i$ since $2h+j-1 \geq i+h-1 \geq i$. Hence $h+j-1 < i$, which implies equality in $h \geq i-j$ above, i.e., $h = i-j$ as claimed. Now

$$kp^h = a - 1 \equiv B^6 - 1,$$

so $B \bmod p^h$ has order dividing 6. But the order must be exactly 6 since $6Q$ is the order of $B \bmod p^i$. Therefore, as in Class B, B satisfies the cyclotomic polynomial $B^2 - B + 1 \equiv 0 \bmod p^h$, so $B^2 = B - 1 + lp^h \bmod p^i$ for some integer l. If $p \mid l$ then since $h < i$, $p^{h+1} \mid B^2 - B + 1 \mid B^6 - 1 \equiv kp^h \bmod p^i$, and this contradicts $p \nmid k$, so $p \nmid l$.

For the final result, $p^h \mid M$ by definition, and since $j \leq i/2 \leq i - j = h$, it follows that $p^i = p^{h+j} \mid p^{2h}$, and so $(p^i = n) \mid (lp^{h+j} = QM) \mid (l^2p^{2h} = M^2)$. □

Many of the calculations in this section will make use of Table 11, which is derived from Lemma 8 in the ring mod $n = p^i$, as follows:

$$\begin{array}{lllll}
B^2 & & = B - 1 & +M, \\
B^3 & = B^2 - B + BM & = -1 & +(B+1)M, \\
B^6 & = (-1 + (B+1)M)^2 & = 1 & -2(B+1)M, \\
B^{6x} & = (1 - 2(B+1)M)^x & = 1 & -2(B+1)xM, \\
B^{6x+1} & = B - 2x(B^2+B)M & = B & -2(2B-1)xM, \\
\vdots & & &
\end{array}$$

From time to time we shall use without comment the implication from the first line above that $B^2M \equiv (B - 1 + M)M \equiv (B-1)M \bmod n$, and similarly for $C \equiv B^{-1} \bmod n$.

k	$B^{6x+k} \bmod n$
0	$1 - 2(B+1)xM$
1	$B - 2(2B-1)xM$
2	$B - 1 + (1 - 2(B-2)x)M$
3	$-1 + (B+1)(2x+1)M$
4	$-B + (2B-1)(2x+1)M$
5	$-B + 1 + (B - 3 + 2(B-2)x)M$

Table 11: $B^{6x+k} \bmod n$ (B of order $L = 6Q$)

Lemma 9. *Base B satisfies $B^{(B^{-1},2) \bmod (Q,6)} \equiv B^{(1-B,2) \bmod (Q,6)} \equiv B-1 \bmod n$.*

Proof. By Lemma 8, we have $Q \mid M$, so since $-1 \equiv 5 \bmod 6$, Table 11 Row 5 implies that $B^{-1} \equiv 1 - B \bmod Q$. Now let $(1 - B, 2) \bmod (Q, 6) = 6x + 2$. Then $x \equiv (1 - B - 2)/6 \bmod Q$, and again by Lemma 8, $n \mid QM$, so x only needs to be known modulo Q. Then by Row 2 of Table 11,

$$\begin{aligned} B^{6x+2} &\equiv B - 1 + (1 - 2x(B - 2))M \\ &\equiv B - 1 + (3 + (B + 1)(B - 2))M/3 \\ &\equiv B - 1 + (3 + B - 1 - B - 2)M/3 \\ &\equiv B - 1 \bmod n. \end{aligned}$$

□

Lemma 10. *If $y \equiv s \bmod 6$, then $b^y \equiv b^s \bmod Q$.*

Proof. Let $y = 6x + s$. Then

$$b^y = b^{6x+s} = (b^6)^x b^s \equiv b^s \bmod Q,$$

since $b^6 \equiv 1 \bmod M$ by Row 0 of Table 11, and this is $1 \bmod Q$ since $Q \mid M$ by Lemma 8. □

Lemma 11. *Let $d \equiv 2f \bmod 6$ where $f = \pm 1$, $B = b_y$ with $y \equiv s = \pm 1 \bmod 6$, and $k_s \equiv (b^{sf} d, f) \bmod (Q, 6)$. Then $B_{k_s} = B_d + 1$.*

Proof. We first calculate B_d. Letting $d = 6w + 2f$, note that $w \equiv (d - 2f)/6 \bmod Q$ is sufficient for the arithmetic involving wM, since $n \mid QM$. Let $C = B^f$, which like B is an element of order $L \bmod n$, and hence satisfying the arithmetic of Table 11. Now $f^2 = 1$, so $B = C^f$ and

$$\begin{aligned} B_d \equiv B^d &= C^{fd} = C^{6wf+2} \\ &\equiv C - 1 + (1 - 2wf(C - 2))M \quad \text{(by Row 2 of Table 11)} \\ &\equiv C - 1 + (3 - f(C - 2)(d - 2f))M/3 \\ &\equiv C - 1 - ((fd - 2)C - 2fd + 1)M/3 \bmod n. \end{aligned}$$

For B_{k_s}, again use $B = C^f$ and recall that $b^{sf} \equiv b^{yf} \equiv B^f \equiv C \bmod Q$, and then $k_s \equiv (Cd, f) \bmod (Q, 6)$. Hence $B^{k_s} = C^{fk_s} = C^{(dfC,1) \bmod (Q,6)} \equiv C^{6x+1}$ where $x = (dfC - 1)/6 \bmod Q$, so using Row 1 of Table 11,

$$\begin{aligned} B_{k_s} &\equiv C - 2(dfC - 1)(2C - 1)M/6 \\ &\equiv C - (2df(C - 1) - (df + 2)C + 1)M/3 \\ &\equiv C - ((fd - 2)C - 2fd + 1)M/3 \bmod n. \end{aligned}$$

Comparing the two calculations, with identical M terms, and knowing that $C \neq 0$, the result follows. □

In the following, "size constraints" refers to the fact that the sum of -1, 0, or +1 and a number of signed values B_z each of which is greater than 1 and less than $n-1$, if known to be 0 mod n, can take only a limited number of values. For example, $B_i + B_j - B_k + 1$ is greater than $4 - n$ and less than $2n - 2$ and can therefore only be 0 or n given that it is a multiple of n.

The following lemma specifies certain (d, e) pairs of Class C/$\tilde{\text{C}}$ columns.

Lemma 12. *Given the conditions of Lemma 11, plus the restriction $Q \nmid d$, let $d_Q \equiv d \bmod Q$ and*

$$m_s = \begin{cases} 1 & \text{if } f = +1,\ s = -1 \\ k_s & \text{otherwise,} \end{cases}$$

$$e \equiv \begin{cases} k_+^{-1} \bmod L & \text{if } f = +1 \\ (d^{-1}, 4) \bmod (Q, 6) & \text{if } f = -1. \end{cases}$$

$$\text{If } f = +1, d_Q \equiv \pm 1 \text{ then} \quad D_s = B_1 + B_{d+1} - B_{k_+^s + 1} \qquad \in \{0, n\}.$$

$$\text{If } f = -1, \text{ then} \quad D_s = B_1 + B_{d+1} - B_d - B_{k_s(e+1)} - 1$$
$$= B_1 + B_{d+1} - B_{k_s} - B_{k_s(e+1)} \qquad \in \{-n, 0, n\}.$$

Proof. From Lemma 7 with rearranged terms to put the biggest two first,

$$D_s = B_{m_s} B_{m_s e} - B_1 B_d - B_{m_s(e+1)} + B_{d+1} \text{ (and is a multiple of } n). \tag{4}$$

We shall use the result of Lemma 11 several times without specific reference.

Consider $f = -1$ first, so $m_s = k_s$. Then $B_{m_s} = B_d + 1$; we shall also show that $B_{m_s e} = B_1 - 1$. Since $B = b_y$, Lemma 10 gives $B \equiv b^s \bmod Q$. Therefore $m_s e = k_s e \equiv (b^{sf} d, f)(d^{-1}, 4) \equiv (B^{-1}, -4) \equiv (1 - B, 2) \bmod (Q, 6))$. Then by Lemma 9, $B^{m_s e} \equiv B - 1 \bmod n$. Since $B^z \equiv B_z \bmod n$ and $0 < B_z < n$ for any z, it follows that $B_{m_s e} = B_1 - 1$ by size constraints. Then Equation (4) becomes

$$D_s = (B_d{+}1)(B_1{-}1){-}B_1 B_d{-}B_{m_s(e+1)}{+}B_{d+1} = B_1{+}B_{d+1}{-}B_d{-}B_{k_s(e+1)}{-}1. \tag{5}$$

as claimed. By size constraints, the only possible multiples of n for D_s are $-n,\ 0, +n$.

Next consider $f = +1$. If $s = +1$ then $m_s = k_s = k_+ \equiv e^{-1} \bmod L$ so $m_s e = 1$ and, by Lemma 11, $B_{m_s} = B_{k_s} = B_d + 1$. Hence Equation (4) becomes

$$D_+ = (B_d + 1)B_1 - B_1 B_d - B_{1+k_+} + B_{d+1} = B_1 + B_{d+1} - B_{k_+^s + 1}$$

as claimed.

But if $s = -1$, Equation (4) gives $D_- = B_1 B_e - B_1 B_d - B_{e+1} + B_{d+1}$ since $m_s = 1$ in this case. The constraint that $d \equiv \pm 1 \bmod Q$ is now used. This implies that d is either $Q + 1$ or $3Q - 1$ (since Q is 1 mod 6), and $d \equiv d^{-1} \bmod Q$. But by assumption,

$$e \equiv k_+^{-1} \bmod L \equiv (b^f d, f)^{-1} \bmod (Q, 6) \equiv (b^{-1} d, 1) \bmod (Q, 6) = k_- = k_s$$

by the definition of k_s when $f = 1$. Now using Lemma 11, $B_e = B_{k_s} = B_d + 1$, so

$$D_- = B_1(B_d+1) - B_1B_d - B_{e+1} + B_{d+1} = B_1 + B_{d+1} - B_{e+1} = B_1 + B_{d+1} - B_{k_+^s+1}$$

as claimed, since $e \equiv k_+^{-1} = k_+^s$. □

The following results are useful for later analysis.

Lemma 13. *If $f = \pm 1$ then:*

a. $B_{2fQ} = B_{fQ} - 1$,
b. $B_{fQ} + B_{-fQ} = n + 1$,
c. $B_{3fQ} = n - 1$,
d. $Q^2 \equiv Q \bmod L$.

Proof. Let $C = B_{fQ}$, which has order 6 since B has order $L = 6Q \bmod n$. Therefore

$$0 \equiv C^6 - 1 \equiv (C-1)(C+1)(C^2+C+1)(C^2-C+1) \bmod n.$$

Just as in Lemma 8, it follows that $C^2 \equiv C - 1 \bmod n$, so $B^{2fQ} \equiv B^{fQ} - 1 \bmod n$. Part (a) follows from size constraints. Then multiplying this by B^{-fQ} gives $B_{fQ} \equiv 1 - B_{-fQ} \bmod n$ and then part (b) follows from size constraints. Part (c) derives from C^3 having order 2, and -1 is the unique element of order 2 in this group.

Finally, $L = 6Q$ so $Q(Q-1) \equiv 0 \bmod L$ since $6 \mid Q - 1$; part (d) then follows. □

Lemma 14. *Let $B = b_y$. If $f = \pm 1$ and $d \equiv 2fQ$, $e \equiv fQ \bmod L$, and $m_s = 1$ for both signs of s, then $D_s = B_1 + B_{d+1} - B_{e+1} \in \{0, n\}$.*

Proof. By Lemma 7 with $m_s = 1$,

$$\begin{aligned} D_s &= B_1B_e - B_{e+1} - B_1B_d + B_{d+1} \\ &= B_1(B_{fQ} - B_{2fQ}) + B_{d+1} - B_{e+1} \\ &= B_1 + B_{d+1} - B_{e+1} \text{ (using Lemma 13(a))} \\ &\in \{0, n\} \text{ (by size constraints).} \end{aligned}$$

□

We now count all the instances of Class C/$\tilde{\text{C}}$ implied by the preceding lemmas.

Theorem 5. *Given the definition of $P(.)$ in Theorem 2, the number of Class C and Class $\tilde{\text{C}}$ instances described in this section is $P(Q) + 6$.*

Proof. With $f = -1$ in Lemma 12, $d \equiv 4 \bmod 6$ and $e \equiv (d^{-1}, 4) \bmod (Q, 6)$. This provides a pair of columns d, e whose sorted values differ by $D_s/n \in \{-1, 0, 1\}$, and this small value qualifies as Class $\tilde{\text{C}}$. Each distinct pair $(d, e) \bmod 6Q$ with fixed value 4 mod 6, and such that $d^2 \not\equiv 1 \bmod Q$, corresponds to a unique pair, $(d \bmod Q, d^{-1} \bmod Q)$ of Class A matches with parameter $L = Q$ as studied in Theorem 2. Therefore the number of pairs is the same, namely $P(Q)$.

Moving on to the case $f = +1,\ d \equiv \pm 1 \bmod Q$, this implies that $d \equiv 2 \bmod 6$ and hence d is either $Q+1$ or $3Q-1$. For each of these two cases $e \equiv k_+^{-1} \bmod 6Q$ is well defined, and in addition, the Class A match between columns e and $e^{-1} \equiv k_+$ provides another Class $\tilde{\text{C}}$ match with column d. So there are four instances arising from this case.

Finally, Lemma 14 yields the two instances $d = 2Q,\ e = Q$ and $d = 4Q,\ e = 5Q$. Therefore the total number of Class $\tilde{\text{C}}$/C instances is $P(Q) + 6$. □

Table 12 provides the details of the six special cases in the above theorem. The calculations for e in the first four rows of the table, where $f = +1$, are as follows. Lemmas 11 and 12 imply that $e \equiv k_+^{-1} \equiv (bd, 1)^{-1} \bmod (Q, 6)$. If $d = Q + 1$ then $e \equiv (b^{-1}, 1) \equiv (1 - b, 1) \bmod (Q, 6)$ or the class A inverse $(b^1, 1) \bmod (Q, 6)$. If $d = 3Q - 1$ instead, its value mod Q is negated, so the mod Q part of e is also negated. Further properties of these cases are studied later.

We have previously asserted that a Class C case exists, and provided Example 11 for it, in which the difference between sorted columns d and e is not just small, but zero. We now prove the theorem which establishes the d which achieves this.

d	e	Relevant lemmas
$3Q - 1$	$(b - 1, 1) \bmod (Q, 6)$	Lemmas 12, 15
$3Q - 1$	$(-b, 1) \bmod (Q, 6)$	Lemmas 12, 15
$Q + 1$	$(1 - b, 1) \bmod (Q, 6)$	Lemma 12
$Q + 1$	$(b, 1) \bmod (Q, 6)$	Lemma 12
$2Q$	Q	Lemma 14
$4Q$	$5Q$	Lemma 14

Table 12: Six special Class $\tilde{\text{C}}$ cases

Theorem 6 (Class C)**.** *If* $d \equiv (b^2, 4) \bmod (Q, 6)$ *and* $e \equiv (b^4, 4) \bmod (Q, 6)$ *then* $D_s = 0$ *and the sorted columns* d *and* e *are identical.*

Proof. We shall assume that $d \equiv 4 \bmod 6$ and make use of Rows 2 and 3 of Table 11, implying that $b^2 \equiv b - 1,\ b^3 \equiv -1 \bmod Q$.

In the case of $s = +1$, we shall solve d to make two particular terms in Equation (5), $D_s = B_1 + B_{d+1} - B_d - B_{k_s(e+1)} - 1$, cancel. Size constraints then prevent the remaining two from adding up to n or $-n$, implying that $D_+ = 0$. Then we shall show that two different terms cancel in the $s = -1$ case.

With $s = +1$, we use Lemma 11's result $B_{k_+} = B_d + 1$ with $f = -1$ (the $d \equiv 4 \bmod 6$ case) and solve $k_+ = d + 1$, so that $B_{d+1} - B_d - 1 = 0$:

$$d + 1 = k_+ \equiv d/b \bmod Q, \text{ so } d \equiv b/(1 - b) \equiv b/(-b^2) \equiv b^2/(-b^3) \equiv b^2 \bmod Q.$$

Hence, if $d \equiv 4 \bmod 6$, then $B_{d+1} = B_d + 1$ if and only if $d \equiv b^2 \bmod Q$, and then $D_+ = B_1 - B_{k_s(e+1)} = 0$ because of size constraints. Using Lemma 12 with $f = -1$ we derive $e \equiv (d^{-1}, 4) \bmod (Q, 6) \equiv (b^4, 4) \bmod (Q, 6)$ as in the theorem statement.

For $s = -1$, write $C_1 = b_y$ where $y \equiv -1 \bmod 6$. Then letting $A_1 = C_{-1} = b_{-y}$, note that $-y \equiv +1 \bmod 6$. Therefore A satisfies the same relations as B did in the $+1 \bmod 6$ case. In particular, $d \equiv b^2 \bmod Q$ if and only if $A_{d+1} = A_d + 1$ if and only if $C_{-d-1} = C_{-d} + 1$. Multiplying by C^{d+1} gives $1 \equiv C^1 + C^{d+1}$, so $C_1 + C_{d+1} = n + 1$. It follows that $D_- = C_1 + C_{d+1} - C_d - C_{k_-(e+1)} - 1 = n - C_d - C_{k_-(e+1)}$, which is zero by size constraints.

We note that the value of k_- is not used above, but $k_- \equiv bd \equiv b^3 \equiv -1 \bmod Q$. Despite the "if and only if" clauses, it cannot be concluded that D_s is identically zero only when $d \equiv b^2 \bmod Q$ (or $b^4 \bmod Q$ if swapped with e), because Equation (5) could be zero through other means. □

For some of the Class $\tilde{\text{C}}$ cases, more can be said on the distribution of the "errors" between the two near matching columns, in particular for the six cases of Table 12, where D_s/n can only be 0 or 1. The remainder of this section studies these cases, and includes a fairly complicated proof that in four cases the 0's and 1's are equal in number.

We deal first with the first two rows of Table 12, having $d = 3Q - 1$. These cases have different errors than do the final four, which is why they are placed first, and the explanation is much simpler.

Lemma 15. *If $d = 3Q - 1$ then $D_s = n$.*

Proof. Let $B = b_y$. From Lemma 12, $D_s = B_1 + B_{d+1} - B_{k_+^s+1}$. But $d + 1 = 3Q$ and $B_{3Q} = n - 1$ by Lemma 13(c). This term is too large to allow D_s to be zero, so it must be the only other alternative, n. □

Thus in these two cases, instead of Class A where the sorted columns match exactly, we have them differing by $+1$ in every row. The four other cases have evenly distributed errors, which is explained via the following general theorem on the difference between two linear functions when reduced modulo a prime power.

Theorem 7. *Let $Q = p^j$ be a prime power, and let $c_i(z) = g_i z + f_i + e_i \bmod Q$ for $i = 1, 2$, with g_1, g_2, $g_2 - g_1$ all coprime to p, and all quantities in $\mathbb{Z}_Q$ except for $e_i \in (0, 1)$, with $e_1 < e_2$ and $f_1 \le f_2$. Then the sign of $d(z) = c_2(z) - c_1(z)$ takes the values -1 and +1 an equal number of times over the following sets:*

a. $\{z : 1 \le z \le Q - 1\}$,

b. $\{z : 1 \le z \le Q - 1,\ p \nmid z\}$ in the case that either $f_1 \equiv 0$ or $f_1 \equiv f_2 \bmod p$.

(Note that case (b) is only relevant if $j > 1$.)

Proof. Let $d(z) = c_2(z) - c_1(z)$, which lies in $[-Q + 1, Q - 1]$, and let Z denote the set of z which are coprime to p. Let the integer $C_i(z) = g_i z + f_i \bmod Q$, which lies in $[0, Q - 1]$. Since the fractional e_i cannot cause wrapping modulo Q, $c_i(z) = C_i(z) + e_i$. We then let $D(z) = C_2(z) - C_1(z)$, which lies in $[-Q + 1, Q - 1]$,

so $D(z) \equiv (g_2 - g_1)z + f_2 - f_1 \bmod Q$. Note that since $g_2 - g_1$ is coprime to Q, this linear function with integer coefficients is invertible modulo Q, so that for any $y \in [0, Q-1]$ there is a unique $z \in [0, Q-1]$ such that $D(z) \equiv y \bmod Q$. Therefore $D(z) = y$ or $y - Q$ since $y + Q \geq Q > D(z)$.

Let z_0 denote the special case $D(z_0) \equiv 0 \bmod Q$ – in fact $D(z_0) = 0$ because $D(z) \geq -Q + 1 > -Q$. Now, $d(z) = D(z) + e_2 - e_1$, so $d(z_0) = e_2 - e_1 > 0$.

A further constraint is that since g_i is coprime to Q, similar invertibility ensures that each possible value of $C_i(z) \in [0, Q-1]$ is attained, so $\sum_{z=0}^{Q-1} C_i(z) = \sum_{y=0}^{Q-1} y$, which is independent of i. This implies that $\sum_{z=0}^{Q-1} D(z) = 0$.

Now consider $D(z) = y > 0$ with $0 \leq z < Q$. As $D(z)$ is invertible, $Q - y$ must equal $D(z')$ for some $0 \leq z' < Q$. Then either $D(z') = -y$ or $D(z') = -y + Q$. In the former case $D(z) + D(z')$ contributes $y + (-y) = 0$ to the overall sum. In the latter case, $D(z) + D(z')$ contributes $y + (-y + Q) = Q$. But in this case, for the sum to come to 0, this must be counterbalanced by a negative pair $D(w) = -u < 0$ and $D(w') = u - Q < 0$ contributing $-Q$ to the sum. Hence $D(z)$ is 0 at $z = z_0$, negative for $(Q-1)/2$ of the values of z, and positive for the other $(Q-1)/2$.

Since $d(z) = D(z) + e_2 - e_1$ and $0 < e_2 - e_1 < 1$, its pattern of signs is the same as for $D(z)$ except for the case z_0, where it is the positive value $e_2 - e_1$, meaning that positives exceed negatives by 1. But $z = 0$ is not in the set of z's for case (a), so is excluded, and its value is $d(0) = c_2(0) - c_1(0) = f_2 - f_1 + e_2 - e_1 > 0$; this excluded positive reduces their number to equal the negatives.

Thus case (a) has been proved. Case (b) is only relevant if Q is not prime, i.e., $j > 1$, and then it has to be proved that the further exclusions $p \mid z$ are balanced between positive and negative, so that the non-exclusions also balance. Let $Q' = Q/p$ and write these excluded cases $p \mid z$ as $z = z'p$ for $1 \leq z' < Q'$. Let $c_i'(z') = c_i(z'p)/p$, noting that $c_2(z'p) > c_1(z'p)$ if and only if $c_2'(z') > c_1'(z')$. Now $c_i'(z') = g_i z' + (f_i + e_i)/p \bmod Q'$, and this may be written as $g_i' z' + f_i' + e_i' \bmod Q'$ where $g_i' = g_i$, $f_i' = \lfloor (f_i + e_i)/p \rfloor$ and $e_i' = \{(f_i + e_i)/p\}$, where $\{.\}$ denotes fractional part. Now this theorem Case a. can be applied to the $'$ quantities Q', e_i', f_i', g_i' provided first that $0 < g_1' \neq g_2' > 0$, which is true since $g_i' = g_i$, and second that $e_2' > e_1'$. This latter requires that $\{(f_2 + e_2)/p\} > \{(f_1 + e_1)/p\}$. Since $0 < e_i < 1$, we have $f_i + e_i \bmod p = (f_i \bmod p) + e_i$. Now $\{x/p\} = (x \bmod p)/p$, so $e_2' > e_1'$ if and only if $(f_2 \bmod p) + e_2 > (f_1 \bmod p) + e_1$, and since $e_2 > e_1$, this is assured if either f_1 or $f_2 - f_1$ is congruent to 0 modulo p. Thus case (b) is proved. □

We now exploit this theorem to prove results on the final four cases of Table 12. Each of these four has D_s of the form $B_1 + B_{d+1} - B_u \in \{0, n\}$ for some u, where $B = b_y$. We can write $D_s = n + B_1 - (n - B_{d+1}) - B_u$. Then since $B_u < n$,

$$D_s = n \text{ if and only if } B_1 > n - B_{d+1} = (-B_{d+1} \bmod n). \tag{6}$$

Now, using the notation, let

$$\begin{aligned} x &= (y - s)/6, \\ z &= x - sa = (y - s)/6 - s(Q-1)/6 = (y - sQ)/6. \end{aligned} \tag{7}$$

Note that $B = b_y$ depends on y and hence on z, which is congruent to 0 mod p if and only if $y \equiv sQ \equiv 0 \bmod p$. But these are excluded values since y must be coprime to p. So the excluded values of z conform to the restrictions of Theorem 7b. Then write

$$\begin{aligned} (-B_{d+1} \bmod n)/M &= e_1 + f_1 + g_1 z \\ B_1/M &= e_2 + f_2 + g_2 z \end{aligned} \tag{8}$$

with all non-negative coefficients. The aim is to prove that these coefficients satisfy the requirements for Theorem 7. As part of that, the following result is needed.

Lemma 16. *Let n be a power of a prime which is congruent to 1 mod 6, and let b be an integer of order 6 mod n. Then, $b \not\equiv 0, 1/2, \pm 1, 2 \bmod n$.*

Proof. Since b is of order 6 mod n, it is clearly not 0 or ± 1 mod n. If $b \equiv 2^{\pm 1}$ then $2^6 = 64 \equiv 1 \bmod n$, so $n \mid 63$. Since $n \equiv 1 \bmod 6$, the only solution is $n = 7$. But 2 and $1/2$ have order 3 mod n, not 6, a contradiction. □

Theorem 8. *For each of the last four rows of Table 12, the sorted column e in the repetend table has entries which are the same as the sorted column d in half the cases, and are one less in the other half of cases.*

Proof. In the following, recall that subscripts to bases are taken modulo $L = 6Q$, and all calculations with bases are implicitly done modulo n.

Since $y = 6x + s$ by Equation (7), any c gives $B_c = b_{yc} = b_{(6x+s)c} = b_{6(z+sa)c+sc}$, which can be converted by row sc mod 6 of Table 11 into the form $u + (v + wz)M$ with $u, v \geq 0$ and $u < M$. Then $u/M + v$ can be evaluated by setting $z = 0$, equivalent to $x = sa$ and $y = 6sa + s = sQ$, and from this $e_k = u/M,\ f_k = v$ where $k = 1, 2$ according to the case being considered. The same is true for any $n - B_c$ in place of B_c.

First consider e_2 in Equation (8); f_2 is not actually needed for our purposes, because f_1 turns out to equal either 0 or f_2. Now e_2 derives from $B_1 = b_{6sa+s} = b_{sQ}$, which depends only on s. The mod M part of the tabled formula provides e_2, and is $(b \bmod M)/M$ for $s = +1$ (at Row 1) and $(1 - b \bmod M)/M$ for $s = -1$ (at Row 5). If either of those values was 0, $1/M$, or $2/M$, then b would have to be one of $0, 1, 2, -1 \bmod M$. But since $Q \mid M$, the same would be true mod Q, yet by Lemma 16 those four values of b do not occur. Hence $e_2 \geq 3/M$ for either s.

Next consider $e_1,\ f_1$ as derived from $n - B_{d+1}$ and $y = sQ$ in each of the three cases of d in the final four rows of Table 12, using $B = b_y = b_{sQ}$.

- $d = Q + 1$: $n - B_{d+1} = n - b_{sQ(Q+2)} = n - b_{s(Q+2Q)} = n - b_{3sQ} = 1$ using Lemma 13(cd). Therefore $e_1 = 1/M < 3/M \leq e_2,\ f_1 = 0$.
- $d = 2Q$: $n - B_{d+1} = n - b_{sQ(2Q+1)} = n - b_{s(2Q+Q)} = 1$ as above, and $e_1 = 1/M < 3/M \leq e_2,\ f_1 = 0$.

			$d = Q+1$			$d = 2Q$			$d = 4Q$		
s	$g_2/2$	X_2	$g_1/2$	X_1	X_0	$g_1/2$	X_1	X_0	$g_1/2$	X_1	X_0
$+1$	$1-2b$	$1/2$	$-2-2b$	-1	∞	$-1-b$	-1	2	$2-b$	2	-1
-1	$-2+b$	2	$-2-2b$	-1	0	$-1-b$	-1	$1/2$	$-1+2b$	$1/2$	-1

Table 13: $g_1/2$ and $g_2/2$ with the "illegal" values of $b = X_0, X_1, X_2$

- $d = 4Q$: $-B_{d+1} = -b_{sQ(4Q+1)} = -b_{5sQ} = -b_{-sQ}$ since $6sQ \equiv 0 \bmod L$. But with $y = sQ,\ B_1 = b_{sQ}$ and so by Lemma 13(b), $B_1 - (-B_{d+1}) \equiv 1 \bmod n$, hence $e_2 - e_1 = 1/M,\ f_2 - f_1 = 0$, so $e_1 < e_2,\ f_1 = f_2$.

We have now established that in each of the three cases for d, either f_1 or $f_1 - f_2$ is $0 \bmod Q$ and hence $0 \bmod p$, and $e_1 < e_2$, as required by Theorem 7b. It remains to show that $p \nmid g_1 g_2 (g_1 - g_2)$, where g_i derives from Equation (8) and Table 11. The g_i values are written in Table 13 as functions of b, and the X_i's denote values of b which would make $g_i = 0$ or $g_1 - g_2 = 0$ and thereby exclude them from qualifying for Theorem 7b.

Since all the X values in the table are either ∞ or precluded by Lemma 16, there is no b of valid order which causes $p \mid g_1 g_2 (g_1 - g_2)$ in any of the cases. Therefore Theorem 7b applies in all three given cases for d.

We now show how g_2 and the $d = 2Q$ case for g_1 are derived for the $s = -1$ row of Table 13; other parts are similar. Note that the g coefficients do not change under the transformation between x and $z = x - sa$, so we use x. (Also, as g_k derives from the M coefficients of Table 11, those parts may be reduced mod Q because $n \mid QM$ by Lemma 8.)

In the case considered, $B_1 = b_y = b_{6x-1} = b_{6(x-1)+5}$ and Row 5 of Table 11 gives $g_2 = 2b-4$, which is 0 at $b = X_2 = 2$. Now $-B_{d+1} = -b_{(6x-1)(d+1)}$ and since $-(d+1)$ is 3 (mod 6), Row 3 gives $-g_1 = 2(b+1)(d+1) = 2(b+1)(2Q+1) \equiv 2(b+1) \bmod Q$, or $g_1/2 \equiv -1 - b \bmod Q$. This is 0 at $b = X_1 = -1$. Then since $g_1 = g_2$ if and only if $-2b - 2 = 2b - 4$ if and only if $b = 1/2 \pmod{Q}$, $X_0 = 1/2$ follows. □

Example 11 at the beginning of this section can now be seen to illustrate three things:

- in Columns 8s and 19s: the first case of Theorem 8, with $Q = 7,\ d = Q + 1 = 8,\ b = 3,\ e \equiv (-2, 1) \bmod (7, 6) = 19$;
- in Columns 16s and 4s: Theorem 6 with $d \equiv (3^2, 4) \bmod (7, 6) = 16,\ e \equiv (3^4, 4) \bmod (7, 6) = 4$;
- in Columns 10s and 40s: Lemma 12 with $f = -1,\ d = 10,\ e \equiv (\frac{1}{d}, 4) \bmod (Q, 6) = 40$.

5. The RDF and Contemporary Papers on Midy Sums

The Repetend Digit Formula (RDF) is an example of necessity being the mother of invention – we needed it to prove various results relating single repetend digits to each other. In contrast, Midy sums, mentioned in all the papers in the table below, combine repetend digits, and proofs without the RDF are available. However, we show here that in at least one case (Lewittes) the RDF enables shorter proofs of such results, and that some authors came close to discovering the RDF.

First we summarize the generality or otherwise, with respect to modulus, base, numerator, and number of Midy groups, of seven notable papers listed in date order – but we do not comment on their varying degrees of complexity. In this section, to help the reader compare our methods with those of Lewittes, we have chosen to alter our notation to match that of Lewittes [8], except that we continue to number from 0 instead of 1. Thus, the fraction is now x/N, the base is B, the repetend digits are $a_0, \ldots, a_{dk-1}$, and the Midy sum is of the d groups of k base B digits.

Paper	N composite?	$B \neq 10$?	$x \neq 1$?	d general?
Midy 1836 [10]	X	X	X	2
Ginsberg 2004 [5]	X	X	X	3
Gupta and Sury 2005 [6]	X	X	X	✓
Martin 2007 [9]	✓	X	X	✓
Lewittes 2007 [8]	✓	✓	✓	✓
Garcia-Pulgarin and Giraldo 2009 [4]	✓	✓	✓	✓
Dang et al 2021 [2]	✓	X	✓	✓

Table 14: List of papers on Midy sums

We now prove a theorem in which part (a) is Lewittes [8] Theorem 1 and part (b) is part of his Theorem 2.

Theorem 9. *If $x/N = \sum_{i=0}^{\infty} a_i B^{-i-1}$ has repetend length $e = dk$, then on letting*

$$A_j = \sum_{i=0}^{k-1} a_{jk+i} B^{k-1-i}; \quad S_d(x) = \sum_{j=0}^{d-1} A_j; \quad x_j = xB^j \bmod N; \quad R_d(x) = \sum_{j=0}^{d-1} x_{jk},$$

we have

a. $S_d(x) = R_d(x)(B^k - 1)/N$,

b. if $\gcd(B^k - 1, N) = 1$ *then* $S_d(x) \equiv 0 \bmod B^k - 1$.

Proof. Note that A_j is the j^{th} digit for the repetend in base $C = B^k$, and that the period of A is $e/k = d$, so $A_d = A_0$.

Let $c_j = xC^j \bmod N$, so $c_j = (xB^{kj} \bmod n) = x_{jk}$. By Lemma 5 with $m \to x$, $n \to N$, $r_j \to A_j$, $b \to C$, the base C digits of the repetend, which must equal A_j

for $j = 0, 1, \ldots, d-1$, are $A_j = Cc_j/N - c_{j+1}/N$. Now $C^d = B^{kd} = B^e = 1 \bmod N$, so $c_d = c_0$. Then

$$S_d(x) = \sum_{j=0}^{d-1}(Cc_j/N - c_{j+1}/N) = \frac{C-1}{N}\sum_{j=0}^{d-1} c_j \text{ since } c_d = c_0.$$

Since $C = B^k$ and $c_j = x_{jk}$, we then have $S_d(x) = (B^k - 1)R_d(x)/N$, which proves part (a). For part (b), if N has no factors in common with $B^k - 1$ then it divides $R_d(x)$, and so $S_d(x)$ is a multiple of $B^k - 1$. □

Note that the statement of this theorem takes 7 lines, as does Theorem 1 of Lewittes, but the proof takes 10 lines versus about 52 lines in Lewittes' preamble from the top of page 4 with multi-line equations counted as just one line. The principal difference in the proofs is the exploitation here of the generality of the RDF Equation (2) by replacing B with $C = B^k$.

Regarding hints of the RDF in other papers, our equation for c_j in Lemma 5 may effectively be seen at Lewittes [8] Equation (3), but the one for r_j is not visible there. However, Gupta and Sury [6] come somewhat closer: at one point in the proof of their Theorem 1, with base 10 and prime modulus, they have an unlabelled equation for $U_1 \ldots U_i$ which becomes

$$a_0 a_1 \ldots a_{i-1} = \lfloor B^i/N \rfloor \tag{9}$$

when generalized into the present notation and with k set to 1.

Lemma 17. *Equation (9) implies the RDF.*

Proof. This equation, with i replaced by $i + 1$ and reversed, gives

$$\lfloor B^{i+1}/N \rfloor = a_0 \ldots a_i = B(a_0 \ldots a_{i-1}) + a_i = B\lfloor B^i/N \rfloor + a_i.$$

Hence $a_i = \lfloor B^{i+1}/N \rfloor - B\lfloor B^i/N \rfloor$. Now, $\lfloor B^x/N \rfloor = (B^x - B_x)/N$ for any non-negative integer x, so

$$a_i = (B^{i+1} - B_{i+1} - B(B^i - B_i))/N = (BB_i - B_{i+1})/N$$

and this is the RDF. □

6. The Probability of Sporadic Column Matches

Sections 2 to 4 introduced and analyzed Classes A, B, and C, of column matches in repetend tables where the numerator $m = 1$. A natural question to ask is whether any column matches outside these classes might occur. In this section we develop a heuristic probability model for the distribution of repetend digits and the chances that some specific pair of them might match, and thence that pairs of columns might match (permuted).

This leads to an assessment that if the number of rows in the table is less than four then the probability would be non-negligible. We therefore exclude such cases by algebraic analysis, which proves quite complex in the case of two rows.

Notation. The following variables are used in this section. Let n_r be the number of rows (small bases of order $L \bmod n$) in the repetend table. Let $n_c = L - P(L)$ be the number of columns excluding one from each of the $P(L)$ pairs of Class A matching columns. In the case $n_r = 2$, let C_j denote the 2-element column j. Let $C_i \sim C_j$ denote that columns i and j match, and $C_i \not\sim C_j$ denote that they do not. Let $D_j = r_{L-1,j} - r_{1,j}$ and $S_j = r_{L-1,j} + r_{1,j}$: the second row has index $L-1$ since $b_{L-1} \equiv b_1^{-1} \bmod n$ is the second base of order L.

We say that two columns i, j are a "sporadic match" if they do not conform to Classes A, B, or C from Sections 2-4 and, for each x coprime to L, there is a y also coprime to L such that $r_{x,j} = r_{y,i}$, and that the mapping between x and y is 1-1. This section shows that the heuristic probability of a sporadic match diminishes rapidly as L increases, but it does not rule out the existence of further special cases arising from algebraic conditions such as those for Classes B and C.

Lemma 18. *Let U and V be distinct bases of order $L \bmod n$. Under a heuristic assumption that U and V are independent random integers uniformly distributed between 2 and $n-1$, the distribution function of the product UV is*

$$\begin{aligned} P[UV \le z] &\simeq F(z) \\ &= \frac{z \log \frac{z+1}{3} - \frac{3z}{4} + \frac{3}{2}}{(n-2)^2} && \text{if } z < n, \\ &= \frac{(n - \frac{1}{2})\frac{z}{n-1} - 3n + \frac{15}{4} + z \log(\frac{(n-1)(n-\frac{1}{2})}{z})}{(n-2)^2} && \text{if } z \ge n. \end{aligned}$$

Proof. First consider $z < n$. Then U cannot exceed $z/2$ since $V \ge 2$. (Notationally, we allow real number limits to summations, which later get converted to integers by floor or ceiling functions.) Then

$$\begin{aligned} (n-2)^2 P[UV \le z] &= \sum_{u=2}^{z/2} \sum_{v=2}^{z/u} 1 \\ &= \sum_{u=2}^{z/2} \lfloor (z/u) - 1 \rfloor. \end{aligned}$$

At this point we approximate the floor function with the obvious smooth function, and approximate the sum with an integral extending 1/2 past each integer

summation limit. So

$$\begin{aligned}(n-2)^2 P[UV \le z] &\simeq \int_{3/2}^{(z+1)/2} (\frac{z}{u} - \frac{3}{2}) du \\ &= z \log \frac{z+1}{3} - \frac{3z}{4} + \frac{3}{2}.\end{aligned} \tag{10}$$

In a similar vein, if $z \ge n$, then

$$\begin{aligned}(n-2)^2 P[UV \le z] &= \sum_{u=2}^{n-1} \sum_{v=2}^{\min(n-1, z/u)} 1 \\ &= \sum_{u=2}^{z/(n-1)} \sum_{v=2}^{n-1} 1 + \sum_{z/(n-1)}^{n-1} \sum_{v=2}^{z/u} 1 \\ &= (n-2)\lfloor z/(n-1) - 1 \rfloor + \sum_{\lceil z/(n-1) \rceil}^{n-1} \lfloor z/u - 1 \rfloor \\ &\simeq (n-2)(\frac{z}{n-1} - \frac{3}{2}) + \int_{z/(n-1)}^{n-\frac{1}{2}} (\frac{z}{u} - \frac{3}{2}) du \\ &= (n-2)(\tfrac{z}{n-1} - \tfrac{3}{2}) + z \log(\tfrac{(n-1)(n-\frac{1}{2})}{z}) - \tfrac{3}{2}(n - \tfrac{1}{2} - \tfrac{z}{n-1}) \\ &= (n - \tfrac{1}{2})\tfrac{z}{n-1} - 3n + \tfrac{15}{4} + z \log((n-1)(n-\tfrac{1}{2})/z).\end{aligned} \tag{11}$$

Then the approximation to $P[UV \le z]$ defined by Equations (10) and (11) is $F(z)$ as per the lemma statement. □

Lemma 19. *The probability density function of $F(z)$ is*

$$f(z) = F'(z) = \frac{1}{(n-2)^2} \begin{cases} \log((z+1)/3) + 1/4 - 1/(z+1), & \text{if } z < n \\ \log((n-1)(n-\frac{1}{2})/z) + 1/(2(n-1))), & \text{if } z \ge n \end{cases} \tag{12}$$

Proof. This is a straightforward differentiation of $F(z)$ defined in Lemma 18. □

Lemma 20. *The approximate heuristic probability that the repetend digits $r_{x,j}$ and $r_{y,i}$ are equal is*

$$\begin{aligned} p &= \sum_{r=0}^{n-1} g_r^2, \text{ where} \\ g_r &= nf\left(n\left(r + \frac{1}{2}\right)\right). \end{aligned} \tag{13}$$

Proof. Consider the event that $r_{x,j} = r$. This implies that $nr \le b_x b_{xj} < n(r+1)$, and by Lemma 18 with $U = b_x$ and $V = b_{xj}$ the heuristic probability of this is

$$F(n(r+1)) - F(nr) \quad \simeq nf(n(r+1/2)) = g_r.$$

Then the joint probability that $r_{x,j} = r$ and $r_{y,i} = r$ is g_r^2, so the probability that $r_{x,j} = r_{y,i}$ is $\sum_{r=0}^{n-1} g_r^2$. □

From this we derive an estimate of the expected number of pairs of columns (from a given set) all of whose digits match.

Lemma 21. *For any given n and L, let S_d be the heuristic expected number of sporadic pairs of matching columns out of $n_c - d$ columns. Then*

$$S_d = \binom{n_c - d}{2} \prod_{k=1}^{n_r} (1 - e^{-kp})$$

where p is given by Lemma 20.

Proof. Start with row $x = 1$, column j, where there are n_r possibilities for row y on column i, and the probability of a match is $1 - e^{-n_r p}$, using the Poisson approximation to a binomial. Continuing, when $n_r - k$ matches have been made and k remain to be made, the probability of the next one occurring is $1 - e^{-kp}$. Hence the probability of a complete match for columns i and j is $\prod_{k=1}^{n_r}(1 - e^{-kp})$, and $\binom{n_c-d}{2}$ is the number of pairs of columns under consideration. □

In the above lemma, $d = 0$ corresponds to considering all the $n_c = L - P(L)$ columns excluding the Class A matches. Now $d = 1$ corresponds to also excluding Column 0, because it is all 0 so unlikely to match any other column. Then $d = 2$ further excludes Column 1, because it has a unique distribution since $r_{x,1} = \lfloor b_x^2/n \rfloor$, and so it is also less likely to match other columns.

Here is an analysis of the behaviour of S_d as n grows, first obtaining an asymptotic expression for p.

Lemma 22. *With p as in Lemma 20, $p = \frac{2}{n} + O(\log n/n^2)$ as $n \to \infty$.*

Proof. By Equation (13) and the top line of (12) ($z \in (0, n^2)$ means that $z > n$ dominates for large n),

$$\begin{aligned} ng_r &= \frac{1}{2n-2} + \log\left(\frac{(n-1)(n-1/2)}{n(r+1/2)}\right) \\ &= \log(n) - \log(r+1/2) + O(1/n), \end{aligned}$$

and then using Lemma 20

$$\begin{aligned} n^2 p = \sum_{r=0}^{n-1} (n g_r)^2 &= \sum_{r=0}^{n-1} \left((\log n - \log(r+1/2))^2 + O(\log n/n) \right) \\ &\simeq \int_0^n (\log n - \log s)^2 ds + O(\log n) \\ &= \int_0^\infty (t)^2 (n e^{-t}) dt + O(\log n) \quad [\text{with } t = \log(n/s)] \\ &= 2n + O(\log n). \end{aligned}$$

Dividing by n^2 yields the specified result. □

Remark 5. Fitting $n^2 p - 2n$ against $\log n$ on $n \in \{10, 30, 100, 300, 1000, 3000, 10000\}$ via linear regression shows that $p \simeq 2/n + (9.46 - 4.54 \log n)/n^2$ with an absolute relative error $|\hat{p}/p - 1|$ bounded above by $0.7/n$ on that set.

Lemma 23. *Let S_d be as in Lemma 21. Then*

$$\log S_d < \log \left(\binom{n_c - d}{2} \right) + n_r(\log p + \log n_r - 1) + \frac{1}{2}(\log n_r + \log(2\pi)) + O(1/n_r).$$

Proof. From Lemma 21, and using the inequality $1 - \exp(-x) < x$ for $x > 0$ and Stirling's formula $z! \sim (z/e)^z \sqrt{2\pi z}$, we have

$$\begin{aligned} \log S_d &= \log \binom{n_c - d}{2} + \sum_{k=1}^{n_r} \log(1 - \exp(-kp)), \\ \log S_d - \log \binom{n_c - d}{2} &< \sum_{k=1}^{n_r} \log(kp) \\ &= n_r \log p + \log n_r! \\ &= n_r(\log p + \log n_r - 1) + \tfrac{1}{2}(\log n_r + \log(2\pi)) + O(1/n_r). \end{aligned}$$

□

The gradient of S_d is negative against n as $\log p \sim \log 2 - \log n$, positive against n_c, and negative against n_r because $n_r < n$ so $(\log p + \log n_r - 1) \sim \log(\frac{2n_r}{ne}) < 0$. Later we show that by elimination of some possibilities by algebra, and others by computer evaluation, the sum of S_0 over all infinitely many remaining possibilities is very small.

Example 12. Let $n = 103$, $L = 6$, $n_r = 2$, $n_c = 6$. Then $S_0 < 0.0066$ by Lemma 23. But see Theorem 12, case $L = 6$, which shows that column matches are impossible in this case.

The development thus far has been theoretical with many (reasonable) approximations. But could these combine to produce discernible errors in the distribution of repetend digit matches? We now provide expectation and variance for this, and test the predictions against several examples.

Lemma 24. *Let H be the number of repetend digit matches between all pairs of the $n_c - d$ columns under consideration, and let $N = n_r(n_c - d)$. Then*

$$E[2H] = \sum_r N^{(2)} g_r^2,$$
$$\mathrm{Var}[2H] = N^{(2)} \sum_r g_r^2(-(4N-6)g_r^2 + 4(N-2)g_r + 2).$$

where g_r is defined in Lemma 20 and $x^{(k)}$ is the fairly common notation to denote $x(x-1)\dots(x-k+1)$.

Proof. Let m_r be the random variable which is the number of times the repetend digit r occurs in the allowed columns. Then we assume that m_r has a Binomial distribution with parameters N and g_r as defined at Equation (13). The probability generating function $E[s^{m_r}]$ is easily shown to equal $(sg_r + 1 - g_r)^N$, and differentiating k times and setting s to 1 gives

$$E[m_r^{(k)}] = N^{(k)} g_r^k. \tag{14}$$

Then

$$H = \sum_r m_r(m_r-1)/2,$$
$$E[2H] = \sum E[m_r^{(2)}] = \sum N^{(2)} g_r^2. \tag{15}$$

Calculation of the standard deviation of H requires moments up to the 4th of m_r. Now $4H^2 = \sum_r \sum_s m_r(m_r-1)m_s(m_s-1)$.

When $r \neq s$, ignoring the small negative covariance from high m_r making high m_s less likely out of the $N - m_r$ remaining digits, Equation (14) shows that

$$E[m_r(m_r-1)m_s(m_s-1)] = N^{(2)} g_r^2 N^{(2)} g_s^2 \tag{16}$$

whereas when $r = s$,

$$\begin{aligned} E[m_r^2(m_r-1)^2] &= E[m_r^{(4)} + 4m_r^{(3)} + 2m_r^{(2)}] \\ &= N^{(4)} g_r^4 + 4N^{(3)} g_r^3 + 2N^{(2)} g_r^2. \end{aligned} \tag{17}$$

Using (15), (16), (17), and the identity $\sum_r \sum_{s\neq r} x_r x_s = (\sum_r x_r)^2 - \sum_r x_r^2$,

$$\begin{aligned} \mathrm{Var}[2H] &= E[4H^2] - E[2H]^2 \\ &= (\textstyle\sum_r N^{(2)} g_r^2)^2 - \sum_r (N^{(2)})^2 g_r^4 + N^{(4)} g_r^4 + 4N^{(3)} g_r^3 + 2N^{(2)} g_r^2 \\ &\quad -(\textstyle\sum_r N^{(2)} g_r^2)^2 \\ &= N^{(2)} \textstyle\sum_r g_r^2(-(4N-6)g_r^2 + 4(N-2)g_r + 2). \end{aligned} \tag{18}$$

□

We now give statistics on how H behaves in practice versus theory. We use $d = 2$ for this, so that we exclude Columns 0 and 1 because they have different probability distributions from the other columns. We also tabulate $l_d = -\log_{10} S_d$ for $d = 0, 1, 2$.

An R program was written to evaluate the observed H_o and expected H_e values of H in the case $d = 2$, along with the standard deviation $\sqrt{\text{Var}[2H]/4}$. The results of nine examples are displayed in Table 15, ordered by increasing n_r and then by increasing n.

Note that the highest relative error for H_o against the model is $(94-63)/23 = 1.35$ standard deviations, which is unremarkable, and $H_o > H_e$ in 5 of the 9 cases. Therefore our heuristic model for repetend digit hits appears to be quite a good one. Note also that $S_0 = 10^{-l_0}$ decreases fairly rapidly as n and n_r increase.

We now move on to Theorem 12 regarding sporadic column matches with $n_r < 4$, which requires a way of calculating n_r from the inputs n and L. This is provided in the following theorem (but see also Toth [15] for a nice alternative formula, albeit requiring an implicit enumeration of subsets of divisors).

Theorem 10. *Let $n = \prod_{i=1}^{m} q_i$ where the q_i are prime powers of increasing characteristic, and let $i_0 = 0$ if $8 \mid n$ and $i_0 = 1$ otherwise. Let $L = \prod_p p^{f_p}$. For $i \geq 2 - i_0$ let $\phi(q_i) = \prod_p p^{e_{p,i}}$. In the case that $i_0 = 0$, so $q_1 = 2^k$ with $k \geq 3$, let $e_{2,0} = 1$ and $e_{2,1} = k - 2$. Also let $S_p = \{i_0 \leq i \leq m : e_{p,i} \geq f_p\}$. Then the number of elements of order L in $\mathbb{Z}_n^*$ is*

$$E(L, n) = \prod_{p \mid L} \left(\left(p^{f_p |S_p|} - p^{(f_p - 1)|S_p|} \right) \prod_{i \notin S_p} p^{e_{p,i}} \right).$$

Proof. Standard theory gives $\phi(n) = \prod_{i=1}^{m} \phi(q_i)$. For $i \geq 2 - i_0$ let $G_i = \mathbb{Z}_{q_i}^*$. By Theorem 1 and the exclusion of $i = 1$ if $i_0 = 0$, the group G_i is cyclic. But if

n	L	$P(L)$	n_r	n_c	H_o	H_e	s.d.	l_2	l_1	l_0
11	10	1	4	9	31	39	18	0.95	0.81	0.68
13	12	0	4	12	92	98	34	0.72	0.63	0.55
17	8	0	4	8	28	25	13	1.59	1.44	1.32
37	12	0	4	12	29	35	14	2.30	2.21	2.14
19	9	2	6	7	12	10	8	3.25	2.95	2.73
29	14	2	6	12	94	63	23	3.21	3.10	3.00
17	16	2	8	14	290	300	77	2.37	2.28	2.20
31	15	2	8	13	157	134	39	4.15	4.05	3.97
23	11	4	10	7	73	47	21	4.64	4.42	4.24

Table 15: Statistics relating to H and S_d

$i_0 = 0$, then $\mathbb{Z}_{q_1}^* \simeq \mathbb{Z}_2 \times \mathbb{Z}_{2^{k-2}}$, so let $G_0 = \mathbb{Z}_2$ and $G_1 = \mathbb{Z}_{2^{k-2}}$. Then in all cases $\mathbb{Z}_n^*$ is isomorphic to the direct product of the cyclic groups $G_{i_0}, \ldots, G_m$, and each $|G_i| = \prod_p p^{e_{p,i}}$ with $e_{p,i}$ as specified.

Let x be an element of order L, if one exists – otherwise see Remark 6. For each prime p and each index i let H_i be the p-subgroup of G_i. There must be at least one index i such that the order of x in H_i is p^{f_p}. For each i we now find the proportion of values which x can take in H_i. First suppose that i is not in S_p, i.e., $e_{p,i} < f_p$; then x does not have order p^{f_p} in H_i, so can take any of the $p^{e_{p,i}}$ values in that subgroup, giving proportion 1.

For other values of i, those in S_p, the order of x in H_i can either be p^{f_p}, enabling p^{f_p} to maximally divide L, or smaller, not so enabling. The number of elements which have order dividing p^{f_p} in this product of subgroups H_i is $\prod_{i \in S_p} p^{f_p}$, and $\prod_{i \in S_p} p^{f_p - 1}$ is the number of elements of order less than p^{f_p} (unless $f_p = 0$, in which case it is 0). The difference is the number of elements with order equal to the desired p^{f_p}. Then the proportion of elements of desired order in this subgroup is $y_p = (p^{f_p|S_p|} - \lfloor p^{(f_p-1)|S_p|} \rfloor) / \prod_{i \in S_p} p^{e_{p,i}}$. The number of elements of order L, $E(L, n)$, is then the product of all these proportions times $\phi(n) = \prod_{i=i_0}^{m} |G_i| = \prod_{i,p} p^{e_{p,i}}$, so

$$E(L,n) = \prod_p \left(\left(\prod_{i=i_0}^{m} p^{e_{p,i}} \right) \left((p^{f_p|S_p|} - \lfloor p^{(f_p-1)|S_p|} \rfloor) / \prod_{i \in S_p} p^{e_{p,i}} \right) \right)$$
$$= \prod_p \left((p^{f_p|S_p|} - \lfloor p^{(f_p-1)|S_p|} \rfloor) \prod_{i \notin S_p} p^{e_{p,i}} \right).$$

Now, if $f_p = 0$ then $e_{p,i} \geq f_p$ for each i so the inside product above is over the null set and the overall term for that p is 1, because $p^{0|S_p|} - \lfloor p^{(-1)|S_p|} \rfloor = 1 - 0$. Hence the product need only be taken over p such that $f_p > 0$, i.e., $p \mid L$, as in the theorem statement of $E(L, n)$, and in this case the floor function is the identity. □

Remark 6. If any $|S_p| = 0$, then $p^{f_p|S_p|} - p^{(f_p-1)|S_p|} = 1 - 1 = 0$ and $E(L, n) = 0$. This can occur if $p \mid L$ but $p \nmid \phi(n)$, agreeing with Lagrange's Theorem that there is no subgroup of order L in this case, or if $p^r \mid L$ with $r \geq 2$ while $p^r \nmid \phi(q_i)$ for any i, in which case there *might* be a subgroup of order p^r yet no element of order p^r.

Corollary 4. *The integer $E(L, n)$ is a multiple of $\phi(L)$.*

Proof. The value $\phi(L)$ contains a factor $h = \phi(p^{f_p}) = p^{f_p-1}(p-1)$ for each prime power $p^{f_p} || L$. The corresponding term in $E(L, n)$ from Theorem 10 is

$$k = (p^{f_p|S_p|} - p^{(f_p-1)|S_p|}) \prod_{i \notin S_p} p^{e_{p,i}} = p^{(f_p-1)|S_p|}(p^{|S_p|} - 1) \prod_{i \notin S_p} p^{e_{p,i}}.$$

Since $|S_p| > 0$, $p^{f_p-1} \mid p^{(f_p-1)|S_p|}$ and $(p-1) \mid (p^{|S_p|} - 1)$, so $h \mid k$ and the result is proved. □

The next examples demonstrate inequality in Corollary 4, i.e., $E(L,n) > \phi(L)$.

Example 13. Let $n = 35,\ L = 12,\ \phi(L) = 4,\ E(L,n) = 8$. Evaluation of cases shows there are 8 bases with repetend length 12: $2^{\{1,5,7,11\}} \bmod 35 = \{2, 32, 23, 18\}$ and $3^{\{1,5,7,11\}} \bmod 35 = \{3, 33, 17, 12\}$. This agrees with Theorem 10 as follows: $m = 2,\ q_1 = 5,\ q_2 = 7;\ \phi(q_1) = 4,\ e_{2,1} = 2,\ e_{3,1} = 0;\ \phi(q_2) = 6,\ e_{2,2} = e_{3,2} = 1;$ $f_2 = 2,\ f_3 = 1,\ S_2 = \{1\},\ S_3 = \{2\};\ E(L,n) = (2^{2\cdot 1} - 2^{1\cdot 1})(2^1)(3^{1\cdot 1} - 3^{0\cdot 1})(3^0)$, which is $8 = 2\phi(L)$.

Example 14. Let $n = 2^k \geq 8,\ L = 2,\ E(L,n) = 3$. In this case, $m = 1,\ i_0 = 0,$ $e_{2,0} = 1,\ e_{2,1} = k - 2 \geq 1 = f_2$, so $S_2 = \{0, 1\}$ and $E(L,n) = (2^2 - 2^0)1 = 3$, which is $3\phi(L)$.

Example 15. Let $n = 2^k \geq 16,\ L = 2^{f_2} \geq 4$. Then $m = 1,\ i_0 = 0,\ e_{2,0} = 1$, and $e_{2,1} = k - 2$ which must be at least $f_2 \geq 2$ to make S_2 non-empty and equal to $\{1\}$. Then $E(L,n) = (2^{f_2} - 2^{f_2 - 1})2 = 2^{f_2} = L = 2\phi(L)$.

It is possible to characterize exactly when equality occurs in Corollary 4.

Theorem 11. *In the notation of Theorem 10, $E(L,n) = \phi(L)$ if and only if for each $p \mid L$, $|S_p| = 1$ and $e_{p,i} = 0$ if $i \notin S_p$.*

Proof. In the proof of Corollary 4, $k = h$ if and only if

$$p^{(f_p - 1)|S_p|}/p^{f_p - 1} = (p^{|S_p|} - 1)/(p - 1) = 1 \text{ and } \prod_{i \notin S_p} p^{e_{p,i}} = 1.$$

It is clear that this occurs only under the conditions of this theorem. □

Note that it would also be possible to argue from the standpoint that the elements of order L must be in a unique cyclic subgroup of order L.

Corollary 5. *If $E(L,n) = \phi(L)$ and $2 \mid L$ then either $n = 4, L = 2$, or $n = tq$ and $L \mid \phi(q)$ where $t = 1$ or 2 and q is an odd prime power.*

Proof. With $p = 2$ in Theorem 11, since every $\phi(q_j)$ is even except for the possibility $q_1 = 2$, the unique divisibility by 2 implies that either $m = 2$ with $q_1 = 2$ and q_2 an odd prime power, so $n = 2q_2$, or $m = 1$. In the latter case, n is either an odd prime power, as in the statement of this corollary, or 2^k, so $L \mid 2^{k-1}$. In this latter case, $k = 1$ does not allow $2 \mid L$, and $k \geq 3$ in Theorem 10 gives $e_{2,1} \geq e_{2,0} = 1$ which does not satisfy Theorem 11. This only leaves the possibility $n = 2^2,\ L = 2$. □

This corollary means that if $E(L,n) = \phi(L)$ and L is even, then $\mathbb{Z}_n^*$ has to be cyclic. But if L is odd, this is not the case, since the odd primes dividing L can be shared among different $\phi(q_i)$'s.

Corollary 6. *If $L = 2$ then $E(L,n)$ is a power of 2 minus 1.*

Proof. Since $L = 2$, the prime $p = 2$ is the only one in the outside product in the Theorem 10 equation. Also, $L = 2 = 2^{f_2}$ so $f_2 = 1$.

Then $E(L, n) = (2^{|S_p|} - 1) \prod_{i \notin S_2} 2^{e_{2,i}}$. But the last term is unity since for each i either $e_{2,i} \geq 1 = f_2$, so $i \in S_2$, or $e_{2,i} = 0$ so $2^{e_{2,i}} = 1$. □

Now that we know how $n_r = E(L, n)$ relates to n and L, we can prove Theorem 12 that if the number of rows $n_r \leq 3$ then there are no sporadic column matches. The following series of lemmas, ending at Lemma 40, will prepare for this theorem.

Lemma 25. *Column $L - 1$ contains no zeroes.*

Proof. If any row x has final element zero, then under the notation of Section 2, $r_x = \sum_{j=0}^{L-2} r_{x,j} b_x^{L-1-j} = b_x z$ say, where z is an integer. But then by Equation (2), if $m = 1$ as mandated in this section, then $b_x^L - 1 = n b_x z$, so $b_x^{L-1} - 1/b_x = nz$. This implies that b_x divides 1, a contradiction. □

Lemma 26. *If $L \leq 2$ there are no column matches.*

Proof. Since L is the number of columns, if $L = 1$ then trivially no column matches are possible. If $L = 2$, since Column 0 is all zero, the only possible match, between Columns 0 and 1, implies that Column 1 is zero. But then each repetend is 0, which is impossible. □

Lemma 27. *If $n_r = 1$ or $n_r = 3$ then there are no column matches.*

Proof. If $n_r = 1$ then 1 is a multiple of $\phi(L)$ by Corollary 4, so $\phi(L) = 1$ and hence $L = 1$ or 2. Then the result follows from Lemma 26.

If $n_r = 3$ then by Corollary 4, 3 must be a multiple of $\phi(L)$, which is either 1 or even. Since 3 is not a multiple of an even number, we must have $\phi(L) = 1$, and the arguments of the $n_r = 1$ case apply to show there are no column matches. □

We now move on to $n_r = E(L, n) = 2$, and Corollary 4 requires that $\phi(L) \mid 2$. In fact $\phi(L) = 2$ is necessary, which occurs just for $L = 3, 4, 6$.

Lemma 28. *If $n_r = 2$ and columns i and j match then: $D_i = \pm D_j$, and $S_i = S_j$.*

Proof. Let $s = L - 1$, so $D_i = r_{s,i} - r_{1,i}$. There is a match if and only if either $r_{1,i} = r_{1,j}$ and $r_{s,i} = r_{s,j}$ or $r_{1,i} = r_{s,j}$ and $r_{s,i} = r_{1,j}$. In the first case $D_i = D_j$, in the second case $D_i = -D_j$, and in both cases $S_i = S_j$. □

Lemma 29. *If $L \equiv 0 \bmod 2$ and $\mathbb{Z}_n^*$ is cyclic then $b_{j+L/2} = n - b_j$ for any j, and $C_j + C_{j+L/2} = C_{L/2}$, $S_j + S_{j+L/2} = S_{L/2}$.*

Proof. The relation in S clearly follows from the relation in C, and for that, any base of order L may be considered to be b_1, so we may just consider Row 1. Being cyclic and of even order, $\mathbb{Z}_n^*$ has exactly one element of order 2, namely -1. Since $b_1^{L/2}$ has order 2, it is $-1 \bmod n$ and therefore $b_1^{j+L/2} \equiv -b_j \bmod n$ for any j, and

this implies $b_{j+L/2} = n - b_j$. Three applications of the Repetend Digit Formula (2) give a Row 1 identity, and hence a column identity, as:

$$\begin{aligned} n(r_{1,j} + r_{1,j+L/2} - r_{1,L/2}) & \\ &= b_1 b_j - b_{j+1} + b_1 b_{j+L/2} - b_{1+j+L/2} - b_1 b_{L/2} + b_{1+L/2} \\ &= b_1 b_j - b_{j+1} + b_1(n - b_j) - (n - b_{j+1}) - b_1(n-1) + (n - b_1) \\ &= 0. \end{aligned}$$

□

Lemma 30. *If $n_r = 2$ and $L = 3$ then there are no column matches.*

Proof. We know that C_0 is all 0, from the RDF and $b_* < n$, and from Theorem 2 that Column 2 consists of pairs of identical digits, so if $n_r = 2$ then $D_0 = D_2 = 0$. To prove that $C_0 \not\sim C_1 \not\sim C_2$, we show $D_1 \neq 0$ and apply Lemma 28. We have:

$$D_1 = r_{2,1} - r_{1,1} = (b_2^2 - b_{4 \bmod 3})/n - (b_1^2 - b_2)/n = (b_2 - b_1)(b_1 + b_2 + 1)/n.$$

Since $b_1 + b_2 + 1 > 0$, the value D_1 can only be zero if $b_1 = b_2$, which is impossible since they are distinct cube roots of 1 mod n.

$C_2 \not\sim C_0$ since the latter is all zero and the former has no 0's by Lemma 25. □

Lemma 31. *Let b be a small base and let $b^2 = b_2 + un$ with $2 \leq b_2 \leq n-2$. Then $b \neq u + 1$.*

Proof. If $b = u + 1$ then $0 < b_2 - 1 = b^2 - un - 1 = u(u + 2 - n)$. Since $u \geq 0$, this inequality implies $u > n - 2$, so $b > n - 1$, a contradiction. □

Lemma 32. *If $n_r = 2$ and $L = 4$ then there are no column matches.*

Proof. With $L = 4$ in Theorem 10, $f_2 = 2$ so $E(L, n) = (2^{2|S_2|} - 2^{|S_2|}) \prod_{i \notin S_2} 2^{e_{2,i}}$, and this equalling $n_r = 2$ implies $|S_2| = 1$ and that if $i \notin S_2$ then $e_{2,i} = 0$ (for otherwise $E(L, n) \geq 4$). Such an i implies that $2 \nmid \phi(q_i)$, which implies $q_i = 2$. Any other i has to be in the single element set S_2, so either $m = 1$ and n is a prime power or $m = 2$ and n is twice an odd prime power.

In fact, in the prime power case, n cannot be 2^k, because if $k \leq 3$ there are no elements of order 4, and if $k \geq 4$ Example 15 with $f_2 = 2$ shows that $E(L, n) = 4$.

Therefore $n = tp^k$ where p is a 1 mod 4 prime and t is 1 or 2; by Theorem 1, $\mathbb{Z}_n^*$ is cyclic.

By Lemma 29, $b_2 = n-1$, $b_3 = n - b_1$, and $C_1 = C_2 - C_3$. Now define the integer u by the equation $b_2 = b_1^2 - un$. We compute the repetend digits for Columns 2 and 3 and deduce Column 1:

$$\begin{aligned} nr_{1,3} = nr_{3,3} \quad &= b_1 b_3 - b_4 = b_1(n - b_1) - 1 && = n(b_1 - u - 1), \\ nr_{1,2} \quad &= b_1 b_2 - b_3 = b_1(n-1) - (n - b_1) && = n(b_1 - 1), \\ nr_{3,2} \quad &= b_3 b_6 - b_9 = (n - b_1)(n-1) - b_1 && = n(n - b_1 - 1). \end{aligned}$$

Thus the repetends for Columns 1 to 3 are given in the following array,

$$\begin{array}{rccc} & 1 & 2 & 3 \\ b_1: & u & b_1-1 & b_1-u-1 \\ b_3: & n-2b_1+u & n-b_1-1 & b_1-u-1 \end{array},$$

and the six column comparisons are as follows:

- $C_3 \not\sim C_0$ by Lemma 25 since $C_0 = 0$,
- $C_1 \not\sim C_2$ since $S_3 > 0$ and $C_1 = C_2 - C_3$ implies $S_1 = S_2 - S_3 < S_2$, then applying Lemma 28,
- $C_0 \not\sim \{C_1, C_2\} \not\sim C_3$ since $D_0 = D_3 = 0$ but $D_1 = D_2 = n - 2b_1$ from the table, so to equal $\pm D_3 = 0$ for Lemma 28 requires $n = 2p^k,\ b_1 = p^k$, which is not coprime to n. □

Remark 7. Instead of using earlier lemmas to prove the above lemma, it would be possible, for each column pair and for each of the two permutations for matching, to use two simultaneous equations to eliminate u, and then find a contradiction for the solved value of b_1. The same would apply to proofs below. However, in the case of $L = 6$ it would involve $15 \times 3 \times 2 = 90$ calculations.

Lemma 33. *If $n_r = 2$ and $L = 6$ then either $n = 3^k t \geq 9$ or $n = p^k t$, where p is a prime 1 mod 6 and $t = 1$ or 2.*

Proof. In Theorem 10, $L = 6$ implies $f_2 = f_3 = 1$, and since $E(6, n) = n_r = 2 = \phi(L)$, Theorem 11 implies that in Theorem 10, $|S_2| = |S_3| = 1$. We let p denote an odd prime.

Now if $8 \mid n$ then $i_0 = 0,\ e_{2,0} = 1 \geq f_2,\ e_{2,1} = k - 2 \geq 1 \geq f_2$ so $|S_2| \geq 2$, a contradiction. Therefore $i_0 = 1$. If $m \geq 3$ then q_2, q_3 are odd so $\phi(q_2), \phi(q_3)$ are even, so $e_{2,2}, e_{2,3} \geq 1 = f_2$, implying $|S_2| > 1$, a contradiction. If $m = 1$ then $n = q_1 = p^k$. If $m = 2$ then $|S_2| = 1$ implies that $q_1 = 2,\ q_2 = p^k$. In both the cases $m = 1$ and $m = 2$, to get elements of order 3, either $9 \mid p^k$ or $p \equiv 1 \bmod 6$. □

We separate the proofs for $p = 3$ and $p > 3$ because they have different properties with respect to the cyclotomic polynomial $b^2 - b + 1 \mid b^6 - 1$: it turns out that if $p = 3$ then $b_1^2 - b_1 + 1 \equiv 3 \bmod n$ and $b_1 + 1 = n/3$, whereas if $p > 3$ then $b_1^2 - b_1 + 1 \equiv 0 \bmod n/t$.

Lemma 34. *If $n = 3^k t$ with $k \geq 2$, $1 \leq t \leq 2$ and $L = 6$ then there are no column matches.*

Proof. Consider $b = n/3 - 1$, which is coprime to 3 and to t (and of order 1 mod t of course). In the following, note that $(n/3)^2 \equiv 3^{2k-2} = 3^k 3^{k-2} \equiv 0 \bmod 3^k$, so squared $n/3$ terms disappear. Then $b^2 \equiv -2n/3 + 1 \equiv n/3 + 1 = b_2$. Then by Lemma 29, $b_3 = n - b_0 = n - 1$, which confirms that b has order 6, $b_4 = n - b_1 = 2n/3 + 3$,

and $b_5 = n - b_2 = 2n/3 - 1$. Using the Repetend Digit Formula, the repetends can easily be shown to be as in the following array:

	1	2	3	4	5
b_1 :	$n/9-1$	$n/9-1$	$n/3-2$	$2n/9-1$	$2n/9-1$
b_5 :	$4n/9-2$	$4n/9-1$	$2n/3-2$	$2n/9$	$2n/9-1$

where by inspection, there are no repeats in the b_5 row, so two columns j, k could only match if $r_{1,j} = r_{5,k}$. This does occur for $j = 4,\ k = 5$, but $r_{5,4} = 2n/9 \neq r_{1,5}$ prevents those columns from matching. □

The final case to consider for Theorem 12 is $p \equiv 1 \bmod 6,\ L = 6,\ n_r = 2$. Within this there are three sub-cases, because $t = 2$ causes a bifurcation: though $b^2 - b + 1$ is congruent to 0 mod p^k, it is not congruent to 0 mod 2 for any b. However, $b^3 + 1 = (b + 1)(b^2 - b + 1) \equiv 0 \bmod 2$ provided b is odd. Therefore if $b < p^k$ is a root mod p^k, the root mod $2p^k$, which we call b_1, is b if it is odd, and $b + p^k$ if b is even. We then unify the three sub-cases by defining:

$$h = \begin{cases} 0 & \text{if } t = 1, \\ +\frac{1}{2} & \text{if } t = 2 \text{ and } b_1 < p^k, \\ -\frac{1}{2} & \text{if } t = 2 \text{ and } b_1 > p^k. \end{cases}$$

This device (with h for "half") means that $b_1 + hn \in [1, n-1]$ is even when $t = 2$. The following six lemmas cover, without explicitly including the parameters in their statements, all the cases with $n = tp^k,\ t = 1$ or $2, p \equiv 1 \bmod 6,\ L = 6,\ n_r = 2$.

Lemma 35. *The following equations hold for some integer u:*

$$\begin{aligned} b_2 &= b_1 - 1 + hn \\ b_3 &= n - 1 \\ b_4 &= n - b_1 \\ b_5 &= n - b_2 = n - b_1 + 1 - hn \\ b_1^2 &= un + b_1 - 1 + hn \\ b_2^2 &= un - b_1 + (hn + 2b_1 - 1)hn \\ b_1 b_2 &= un - 1 + (b_1 + 1)hn. \end{aligned}$$

Proof. Since $b_1^2 \equiv b_2 \bmod n$ by definition, let $u = (b_1^2 - b_2)/n$. Since $b = b_1$ has order 6 mod n, and $\mathbb{Z}_n^*$ is cyclic, $b^3 \equiv -1 \bmod tp^k$, and therefore $p \mid (b+1)(b^2 - b + 1)$. If p divides both factors then it divides $(b+1)^2 - (b^2 - b + 1) = 3b$. Since $\gcd(b, p) = 1$, then $p \mid 3$, a contradiction. Now p cannot divide $b + 1$ alone, for then b has order 2, so $b^2 - b + 1 \equiv 0 \bmod p^k$, implying $b_2 \equiv b_1 - 1 \bmod p^k$. If $h = 0$, i.e., $t = 1$, then $b_2 = b_1 - 1$. But if $t = 2$ then b_2 must be odd in order that $b_2 + 1 \equiv 0 \bmod t$. Since $b_1 - 1$ is even (for the same reason), $p^k = n/2$ must be added to or subtracted from $b_1 - 1$ to give $0 < b_2 < n$. Hence $b_2 = b_1 - 1 + hn$ in all three cases.

Next, $b_3 \equiv b_1 b_2 \equiv -1 \bmod p^k$. If $t = 1$ then $b_3 = n - 1$, which also works to make b_3 odd when $t = 2$ and n is even. Similarly $b_4 \equiv b_3 b_1 \equiv -b_1 \bmod p^k$ yields $b_4 = n - b_1$, and likewise $b_5 = n - b_2$.

For $b_1^2,\ b_2^2,\ b_1 b_2$, we first derive u from $un = b_1^2 - b_2$, and then $b_1^2 = un + b_1 - 1 + hn$. Then

$$\begin{aligned} b_2^2 &= b_1^2 + 1 + h^2 n^2 - 2b_1 + 2b_1 hn - 2hn \\ &= un + (b_1 - 1 + hn) + 1 + h^2 n^2 - 2b_1 + 2b_1 hn - 2hn \\ &= un - b_1 + (hn + 2b_1 - 1)hn, \\ b_1 b_2 &= (un + b_1 - 1 + hn) - b_1 + b_1 hn \\ &= un - 1 + (b_1 + 1)hn. \end{aligned}$$

□

Lemma 36. *The differences D_i satisfy $D_0 = D_5 = 0$ and*

$$\begin{aligned} D_2 &= n - b_1 - b_2 = 1 - 2b + (1 - h)n \\ D_1 &= (1 - h)D_2 \\ D_3 &= D_2 \\ D_4 &= hD_2. \end{aligned}$$

Proof. Column C_0 is zero, and C_5 has equal digits by Theorem 2, so $D_0 = D_5 = 0$. The RDF and Lemma 35, with ... representing some omitted algebra, give

$$\begin{aligned} nD_2 &= n(r_{5,2} - r_{1,2}) = b_5 b_4 - b_3 - b_1 b_2 + b_3 = (n - b_2)(n - b_1) - b_1 b_2 \\ &= (n - b_1 - b_2)n; \\ nD_1 &= b_5^2 - b_4 - b_1^2 + b_2 = (n - b_2 - b_1)(n - b_2 + b_1) + b_2 - (n - b_1) \\ &= n(n - 2b_2) + b_2^2 - b_1^2 + b_2 + b_1 - n = \ldots = (1 - h)(n - b_1 - b_2); \\ nD_3 &= b_5 b_3 - b_2 - b_1 b_3 + b_4 = (n - b_2 - b_1)(n - 1) - b_2 + n - b_1 \\ &= (n - b_1 - b_2)n; \\ nD_4 &= b_5 b_2 - b_1 - b_1 b_4 + b_5 = (n - b_2)(b_2 + 1) - b_1(n - b_1) - b_1 \\ &= n(b_2 - b_1 + 1) + b_1^2 - b_2^2 - b_1 - b_2 = \ldots = (n - b_1 - b_2)hn. \end{aligned}$$

□

We now examine the 15 combinations of column pairs, with the understanding that a set of x columns compared with a set of y columns covers all xy combinations.

Lemma 37. *Column C_0 does not match C_5 (Case 1).*

Proof. This follows since $C_0 = 0$, and C_5 contains no zeroes by Lemma 25. □

Lemma 38. *Columns $\{C_1, C_2, C_3\}$ do not match $\{C_0, C_5\}$ (Cases 2-7).*

Proof. The differences $D_0 = D_5 = 0$, so by Lemma 28, a match in these six cases requires $D_j = 0$ for some $1 \le j \le 3$. But by Lemma 36, since $(1-h) > 0$, any such $D_j = 0$ implies $D_2 = 0$, which implies $b \equiv 1/2 \bmod p^k$, contradicting Lemma 16. □

Lemma 39. *Column C_4 does not match columns $\{C_0, C_5\}$ (Cases 8-9).*

Proof. Since $D_4 = hD_2$ by Lemma 36, and $D_2 \neq 0$ by the proof of the previous lemma, if $h \neq 0$ then $D_4 \neq 0$, and as before C_4 cannot match C_0 or C_5. And if $h = 0$, forcing $D_4 = 0$, we need $0 \neq r_{1,4} \neq r_{1,5}$. Now,

$$nr_{1,4} = b_1b_4 - b_5 = b_1(n-b_1) - n + b_2 = (b_1-1)n - (b_2+un) + b_2 = (b_1 - 1 - u)n,$$

and this is non-zero by Lemma 31. Finally, $nr_{1,5} = b_1b_5 - b_6 = b_1(n - b_1 + 1) - 1$, and this equals $b_1n - (un + b_1 - 1) + b_1 - 1 = (b_1 - u)n \neq nr_{1,4}$. □

Lemma 40. *No pair of columns within $\{C_1, C_2, C_3, C_4\}$ matches (Cases 10-15).*

Proof. By Lemma 28, $C_i \sim C_j$ requires $D_i = \pm D_j$; by inspection of the cases of Lemma 36, and knowing that $h \in \{0, \frac{1}{2}, -\frac{1}{2}\}$, this is ruled out except for three cases, which we now examine.

`Case 1:` Column 2 versus 3. By Lemma 29, $S_2 + S_5 = S_3$, and since $S_5 > 0$ by previous analysis of Column 5, $S_2 \neq S_3$, so $C_2 \not\sim C_3$ by Lemma 28.

`Case 2:` Column 1 versus 2 and 3 when $h = 0$. In this case $D_1 = D_2 = D_3$ prevents use of Lemma 28. But $D_4 = D_5 = 0$ is useful as it implies $S_j = 2r_{1,j}$ for $j = 4, 5$. Then since $S_1 + S_4 = S_2 + S_5$ by Lemma 29, $S_2 - S_1 = S_4 - S_5 = 2(r_{1,4} - r_{1,5}) \neq 0$ since $C_4 \not\sim C_5$ by Lemma 39. The other relation $C_1 \not\sim C_3$ is proved from $S_1 + S_4 = S_3$ by Lemma 29, and $S_4 > 0$ since $C_4 \not\sim C_0$ has been proved, so $S_1 < S_3$.

`Case 3:` Column 1 versus 4 when $h = 1/2$. Now $D_1 = D_4$, so a match occurs if and only if $r_{1,1} = r_{1,4}$. By the RDF and Lemma 35, this implies that $0 = n(r_{1,4} - r_{1,1}) = b_1b_4 - b_5 - b_1^2 + b_2 = b_1(n - 2b_1) + 2b_2 - n = (b_1 - 1)n + 2(b_2 - b_2 - un) = (b_1 - 1 - 2u)n$. Solving for u then implies $b_1^2 = b_2 + un = (b_1 - 1 + n/2) + (b_1 - 1)n/2 = b_1(1 + n/2) - 1$. Since n is even, dividing by b_1 implies $b_1 \mid 1$, contradicting Lemma 16. □

This concludes the proofs of the 15 cases for $n = tp^k,\ t = 1$ or $2, p \equiv 1 \bmod 6$, $L = 6,\ n_r = 2$, and Lemmas 25 to 40 combined prove the following theorem.

Theorem 12. *If the number of rows $n_r \le 3$ then there are no sporadic column matches in the repetend table among all the $n_c = L - P(L)$ possible columns.*

Now fix $n_r = E(L, n)$, and note from Corollary 4 that it is a multiple of $\phi(L)$. Therefore, the given n_r implies a limited set for $\phi(L)$, which in turn implies a limited set for L. The set of n's compatible with L is infinite, and though it tends to be relatively sparse, we assume that there is an integer N such that all $n \le N$ have been ruled out from sporadic column matches by direct computation, and (pessimistically) that all $n > N$ are compatible with L. This, in combination with

Theorem 12 above, which rules out sporadic matches for small n_r, leads to a useful upper bound on the probability of sporadics, given in Theorem 13 below. We first derive an upper bound on the size of the set of compatible L's.

Lemma 41. *The number of possible L's given $n_r > 3$ is at most $T(n_r)$ where*

$$T(x) = \begin{cases} 2 & \text{if } x \text{ is a power of 2 minus 1} \\ 0 & \text{if } x \text{ is otherwise odd} \\ 8 & \text{if } x = 4, 6 \\ 12 & \text{if } x = 8 \\ 15x/4 & \text{if } 8 < x < 48 \text{ and even} \\ 35x/8 & \text{if } 48 \le x < 480 \text{ and even} \\ 77x/16 & \text{if } 480 \le x < 5760 \text{ and even.} \end{cases} \tag{19}$$

Proof. By Corollary 6, $L = 2$ if and only if n_r is a power of 2 minus 1, and otherwise n_r is never odd. This justifies the first two lines in Equation (19).

If $n_r = 4$, $\phi(L) \in \{1, 2, 4\}$ so $L \in \{2, 3, 4, 5, 6, 8, 10, 12\}$, with 8 elements.

If $n_r = 6$, $\phi(L) \in \{1, 2, 3, 6\}$ so $L \in \{2, 3, 4, 6, 7, 9, 14, 18\}$, also with 8 elements.

If $n_r = 8$, $\phi(L) \in \{1, 2, 4, 8\}$ so $L \in \{2, 3, 4, 5, 6, 8, 10, 12, 15, 16, 20, 24\}$, with 12 elements. These results justify lines 3 and 4 of Equation (19) – they are not best possible since detailed use of Theorem 10 can show some of these (n_r, L) pairs are impossible.

Let $p_1 = 2,\ p_2 = 3, \ldots$ be the list of all primes and write $L = \prod_{i=1}^{f} p_{j(i)}^{e_i}$ where each $e_i > 0$. Then using Corollary 4,

$$\frac{n_r}{L} \ge \frac{\phi(L)}{L} = \frac{\prod_{i=1}^{f} p_{j(i)}^{e_i - 1}(p_{j(i)} - 1)}{\prod_{i=1}^{f} p_{j(i)}^{e_i}} = \prod_{i=1}^{f}(1 - \frac{1}{p_{j(i)}}) \ge \prod_{i=1}^{f}(1 - \frac{1}{p_i}) = N_f/T_f \tag{20}$$

where $N_f = \prod_1^f (p_i - 1)$, $T_f = \prod_1^f p_i$. Following the numerators of the above sequence also shows that $\phi(L) \ge N_f$, and the denominators show that $L \ge T_f$.

f	1	2	3	4	5	6
T_f	2	6	30	210	2310	30030
N_f	1	2	8	48	480	5760
T_f/N_f	2	3	3.75	4.37	4.81	5.21

Table 16: The first six values of the increasing sequences T_f, N_f and T_f/N_f

Then, given n_r but not L, if $N_g \le n_r < N_{g+1}$ then $N_f \le \phi(L) \le n_r < N_{g+1}$, so $f \le g$. We summarize all this in:

$$L \ge T_f; N_f \le N_g \le n_r < N_{g+1}. \tag{21}$$

Hence by (20), $L \le n_r T_f/N_f \le n_r T_g/N_g$. Since the number of L's is less than the maximum L, lines 5, 6 and 7 of Equation (19) are derived from the values of T/N in Table 16 with the column indices 3, 4, and 5. □

Remark 8. Though not needed, it is possible to extend Lemma 41 to any general $x \ge 5760$ by using a lower bound on the value of the n^{th} prime, such as given by Robin [14]. Details are available on request to the first author.

Next we prove a lemma needed for Theorem 13.

Lemma 42. *If $m, s > 0$, then $\sum_{n=s+1}^{\infty} n^{-m} < ((m-1)s^{m-1})^{-1}$.*

Proof. It follows by integrating n from s instead of summing from $s+1$, and because the series is decreasing, each term n^{-m} is less than its integral over the interval $(n-1, n)$. □

In bounding the expected number of sporadic column matches, in the following theorem, we use $S_0(n_c, n_r, n)$ from Lemma 21, which allows Columns 0 and 1 to be included even though the reasoning following that lemma explains why they are less likely to match other columns.

Theorem 13. *The expected number of sporadic column matches for cases where $n > N$ is less than*

$$S(N) = \sum_{n_r=4}^{N-1} U(n_r) \frac{N}{n_r - 1} \left(\frac{2n_r}{eN}\right)^{n_r} + \sum_{n_r=N}^{\infty} U(n_r) \frac{n_r}{n_r - 1} \left(\frac{2}{e}\right)^{n_r}, \tag{22}$$

where $U(x) = T(x)(T(x)+2)(T(x)+1)\sqrt{\pi x/2}$.

Proof. By Lemmas 21 and 23

$$S_0(n_c, n_r, n) = \binom{n_c}{2} \left(\frac{2n_r}{en}\right)^{n_r} \sqrt{2\pi n_r}.$$

Note that $n_r \le \phi(n) < n$. Since $n_r \le 3$ has been ruled out by Theorem 12, and since $n_c = L - P(L) \le L \le T(n_r)$, the expected number of sporadic column matches is

$$\sum_{n_r=4}^{\infty} \sum_{L:\phi(L)|n_r} \sum_{n>N, L|\phi(n)} S_0(n_c, n_r, n) \le \sum_{n_r=4}^{\infty} T(n_r) \sum_{n>\max(N,n_r)} S_0(T(n_r), n_r, n)$$

$$= \sum_{n_r=4}^{\infty} T(n_r) \sum_{n>\max(N,n_r)} (T(n_r)+2)(T(n_r)+1)(\frac{2n_r}{en})^{n_r} \sqrt{2\pi n_r}/2.$$

We now split the outer sum into two parts: $4 \le n_r < N$ and $n_r \ge N$. The first sum is

$$\sum_{n_r=4}^{N-1} U(n_r)(2/e)^{n_r} \sum_{n=N+1}^{\infty} \left(\frac{n_r}{n}\right)^{n_r} < \sum_{n_r=4}^{N-1} U(n_r)(2/e)^{n_r} \frac{n_r^{n_r}}{(n_r-1)N^{n_r-1}}$$

$$= \sum_{n_r=4}^{N-1} U(n_r) \frac{N}{n_r-1} \left(\frac{2n_r}{eN}\right)^{n_r}$$

using Lemma 42 with $m \to n_r$, $s \to N$. This sum, which is the first sum in Equation (22), can be evaluated directly by computer.

The second sum is similar, but has limits $n_r \ge N$ and $n \ge n_r+1$ so the Lemma 42 parameters are $m \to n_r$ and $s \to n_r$. This makes the sum less than

$$\sum_{n_r=N}^{\infty} U(n_r) \frac{n_r}{n_r-1} \left(\frac{2}{e}\right)^{n_r}.$$

This is the second sum in Equation (22), so the theorem is proved. Though an infinite sum, the terms diminish exponentially and become zero to machine precision within a few thousand terms, in particular fewer than the 5760 which is the limit on n_r covered by Lemma 41, so computation of all non-zero (machine precision) terms is easily achievable. □

We initially performed exhaustive testing for all $n \le N = 1000$, and found no sporadic column matches. Hence we were able to use Theorem 13 to assert that the expected number of such matches is less than 3.2×10^{-5} (and virtually all of that amount arises from the case $n_r = 4$). We then tried all moduli up to $N = 2500$: at this point, the expected number of sporadic matches is less than 2.1×10^{-6} (but would have been 0.65 if $n_r = 2$ had not been ruled out). We therefore make the following conjecture.

Conjecture 1: There are no sporadic column matches.

This does not rule out the existence of algebraic column matches beyond Classes A, B, and C, though they would have to satisfy fairly severe conditions for them not to occur for any $n < 2500$.

Acknowledgement. We wish to thank anonymous referees for pointing out a number of relevant and interesting references.

References

[1] L. Childs, *A Concrete Introduction to Higher Algebra*, Springer, New York, 1979.

[2] S. Dang, S. Nanoti, and A. Tripathi, Extensions of Midy's theorem for periodic decimals, *Integers* **21** (2021), #A26.

[3] J. Gallian, *Contemporary Abstract Algebra (8th ed.)*, Cengage Learning, Boston, 2012.

[4] G. Garcia-Pulgarin and H. Giraldo, Characterizations of Midy's property, *Integers* **9** (2009), #A18.

[5] B.D. Ginsberg, Midy's (nearly) secret theorem – an extension after 165 years, *College Math. J.* **35** (2004), 26-30.

[6] A. Gupta and B. Sury, Decimal expansion of 1/p and subgroup sums, *Integers* **5** (2005), #A19.

[7] W.E. Heal, Some properties of Repetends, *Ann. of Math.*, **3** (4) (Aug. 1887), 97–103.

[8] J. Lewittes, Midy's theorem for periodic decimals, *Integers* **7** (2007), #A2.

[9] H. W. Martin, Generalizations of Midy's Theorem on repeating decimals, *Integers* **7** (2007), #A3.

[10] E. Midy, De quelques proprietes des nombres et des fractions decimals periodiques, College of Nantes, France, 1836.

[11] O. Ore, *Number theory and its history*, McGraw-Hill, New York, 1948.

[12] K. C. Posch and R. Posch, Modulo Reduction in Residue Number Systems, *IEEE Trans. Parallel Distrib. Syst.* (1995) **6** (5).

[13] H. Riesel, *Prime Numbers and Computer Methods for Factorization*, Birkhauser, 1994 (2nd edition).

[14] G. Robin, Estimation de la fonction de Tchebychef Θ sur le k-ieme nombre premier et grandes valeurs de la fonction $\omega(n)$ nombre de diviseurs premiers de n, *Acta Arith.* (1983) **52**, 367-389.

[15] L. Toth, On the number of cyclic subgroups of a finite Abelian group, *Bull. Math. Soc. Sci. Math. Roumanie* **55** (103) No. 4 (2012), 423-428.

Appendix A: Class B Arithmetic via a Hybrid Residue Number System

Introduction

Class B primes (or prime powers – see Section 3) take the form $p \equiv 1 \bmod L$ where $L = 12a + 6$ and a is an integer which needs to be fairly small to have a reasonable chance 2^{-s} of finding examples. The value s is specified as a function of $f = \phi(L)$, with L being the repetend length, as in Theorem 4. For each a, we can estimate how large the first successful p will be as follows. Ignoring prime powers, which become proportionally rarer as the limit increases, the expected number of Class B primes below a bound B is roughly

$$H(a, B) = B/((\log(B) - 1)f2^s)$$

from the prime number theorem and the restriction that p is not just coprime, but unity, relative to L.

If we consider $a = 10$ in Table 8, just before the leap from $s = 12$ to $s = 16$, we find that $H(10, 2000000) = 1.00$, so the expected number of hits below that bound of two million is approximately 1. In fact the first hit is about three-quarters of that. But the 97^{th} and final hit computed is at $269612029 \sim 2^{28.01}$. We shall see that that is too large for single precision computation within the R language. At the bottom of that table, where $a = 22$, the limit from $H(22, 2.8e10) \sim 1$ is $2^{34.70}$, too large for products to fit in the 64-bit word common to modern micro-processors. Therefore special code was needed, and though we could, with some effort, have got more efficient code by rewriting in the C language and using a multiprecision package such as NTL, we decided to continue with the interpreted language R and see what we could achieve.

The Scale of the Challenge

An "average" prime in the $a = 22, s = 24$ case would be around half the limit computed above, so $2^{33.7}$. Based on an experiment to find 1000 candidate primes of the correct form and of that size, the mean work done per candidate is as follows.

First there are 6.07 GCD tests against the product of primes up to 37, to weed out many composites.

Then there are 3.79 single Fermat tests (does $2^n \equiv 2 \bmod n$?) including the successful one which produces a candidate which is probably prime. Of course, this test can pass composites (pseudoprimes), and indeed among the 1000 candidates there was one: $13954784761 = 19421 \cdot 718541$. However, such pseudoprimes are detected later in the process, for example by k_i in Theorem 4 not being an integer.

Finally, a suitable base b is sought of order L mod p, and the k_i's evaluated. This is done by finding the order of small integers z. For each z, having earlier factorized $p-1$, $z^{(p-1)/y}$ is evaluated with prime power factors of $p-1$ systematically put into and removed from y until the exact order x of z is found. If $L \nmid x$ then the next z is tried, but otherwise $z^{x/L}$ is the required element of order L. Typically only a small number of z's need to be tried.

In the experiment, the order-finding part took 87% of the work, so improving the speed of finding candidates would have a marginal effect. But in any case, both primality testing and order-finding employ modular exponentiations and hence modular multiplications (MMs). Speed (and without saying, accuracy) of the latter is therefore of paramount importance; in the experiment 1563 MMs per candidate were performed. Each candidate took 9.7ms, so $2^s = 2^{24}$, the expected number per success, would take 45 hours on a single processor.

The Hybrid RNS

The `double` arithmetic type in R uses IEEE floating point (FP) numbers with a 52-bit mantissa. With an implied leading 1 for any number (except zero), this gives 53 bits of precision. Our experiments confirmed that it was possible to multiply two numbers less than $2^{26.5}$ and get a precise result. Rather than use 26-bit packets and pack and unpack these whilst performing standard multiplication, we decided to use

a Residue Number System (RNS) with 2 or 3 prime moduli; the identical working over those primes made for relatively efficient code since R can act on vectors. The hybrid aspect of the algorithm comes from also carrying an FP estimate of the size of each number.

Given k prime moduli $P_1 > \ldots > P_k$ close in size just below 2^c for some c such as 26, we write $\mathbf{P}$ as the vector of these primes, and use the following notation regarding an integer x. The true integer value is x, x' is a FP approximation to it, and bold $\mathbf{x} = (x_1, \ldots, x_k)$ is the RNS representation. Square brackets $[\cdot]$ will represent reduction modulo $\mathbf{P}$ of either a scalar or vector of the correct length, so $\mathbf{x} = [x]$. Operations on bold variables such as $\mathbf{x}$ are implicitly done modulo $\mathbf{P}$. We use $\mathrm{CRT}(\cdot)$ to denote the inverse function to $[\cdot]$, derived from the Chinese Remainder Theorem and imposing the limits $\pm M_0'/2$ where $M_0' \sim M_0 = \prod_1^k P_j$. Thus $\mathrm{CRT}(\mathbf{x}) = x$, and it additionally uses the stored data $\mathbf{M} = M_0/\mathbf{P}$ and $\mathbf{R} = [1/\mathbf{M}]$. (The variable R is used to suggest "reciprocal".) Note that $\mathbf{M}$ is known exactly since each element is a product of $k - 1 \leq 2$ primes each of which is less than 26.5 bits.

With this notation, the Chinese Remainder Theorem algorithm [1] (Chapter 14) can be written as:

$$\mathrm{CRT}(\mathbf{x}) = (\mathbf{xR}) \cdot \mathbf{M} \bmod M_0. \tag{A1}$$

We use $\mathrm{CRT}'(\mathbf{x})$ to denote a FP approximation to this.

The details of the number system used depend on the size of the modulus n in accordance with Table A1.

$\log_2 N$	k	c	$\log_2 E$
26.5	-	-	-
51.0	2	26.0	-
62.0	3	25.0	-
64.0	3	26.0	-
65.5	3	26.5	64.63

Table A1: limits on $\log_2 n$ given c and k

Here, N is the maximum value of n appertaining to each row. The first row simply uses FP numbers, so the number k of Chinese primes is irrelevant. In the last row E stands for extra error correction: if $E < n < N$ then an extra step is used to ensure that modular multiplications stay on track.

Note that $k = 2$ primes are easily sufficient for the Class B calculations on moduli up to $2^{34.7}$ mentioned above, but we wanted to see how far the method could be pushed with 3 primes, achieving 65.5 bits as detailed below. We did not implement the next possibility, 4 primes, but our inequalities suggest that it would support moduli up to 74 bits. But it appears to be a case of diminishing returns of modulus size against number of primes, so researchers interested in larger moduli should consider algorithms such as those of Posch and Posch [12] which do not have limits on k. They favour Montgomery-style arithmetic and base extension by doubling the number of Chinese primes. Their Figure 3 suggests that they perform 7 k-fold

modular multiplications, whereas we take 4 (at Step 3 (2), Step 4, and Step 6). We have not implemented their algorithm, so do not know which is faster in practice at 65 bits.

The Hybrid RNS Modular Multiplication Algorithm

This algorithm takes inputs $\mathbf{a}, a', \mathbf{b}, b', \mathbf{n}, n'$, and outputs $\mathbf{z}, z'$ where $z = ab \bmod n$. With exact arithmetic, $z = ab - hn$ where $h = \lfloor ab/n \rfloor$, and setting $\mathbf{z} \leftarrow \mathbf{ab} - h\mathbf{n}$, we have $z' = z = \mathrm{CRT}(\mathbf{z}) = (\mathbf{zR}) \cdot \mathbf{M} \bmod M_0$ by Equation (A1).

But for larger n, the arithmetic is not exact, and many refinements can be needed to determine h. The FP estimate $h' = \lfloor a'b'/n' \rfloor$ can differ from h, so a correction using $\mathrm{CRT}(\mathbf{ab} - h'\mathbf{n})$ is applied in Steps 3 to 6. This can still leave a substantial error, so a correction using mod P_1 arithmetic is performed in Step 7. If $c > 25$ (hence $n > 2^{62}$) it proves necessary to test modulo P_2 too, in Step 8, and apply a further correction if needed. Even after this the result can be just above n, and Step 9 tests and corrects for this. Finally, if $n > E$, further error checking is done in Step 10 to see whether the small value l in Step 5 needs changing, and if so then it gets changed and Steps 6 to 9 are redone.

In the algorithm below, for intermediate variables we use vectors lettered between $\mathbf{u}$ and $\mathbf{y}$, and earlier letters for scalars.

Error Analysis

In this section, we analyze the floating point errors which occur in each step. Table A2 gives values for the error bounds for various values of c and $m = \log_2 n$, and these explain the thresholds chosen in Table A1.

For each scalar variable, such as a for example, with its FP approximation a', use $e_a = a' - a$ to denote the error in the approximation, and E_a for a bound on its magnitude. IEEE-754 standard at https://ieeexplore.ieee.org/document/8766229 provides for double-precision arithmetic with a format of a 53-bit mantissa using 52 bits of data and an implied leading 1. Let $\delta = 2^{-53}$. Through rounding in IEEE-754, if $a > 2^{53}$ then, by the theory of that document and confirmed by experiment,

$$|e_a| \le \delta 2^{\lceil \log_2 a \rceil - 1}. \tag{A2}$$

However, in our arithmetic Step 7 involves an addition step which doubles the error limit, so we assume, and this is borne out by experiment, that

$$|e_a| \le E_a = \delta 2^{\lceil \log_2 a \rceil} = 2^{\lceil \log_2 a \rceil - 53}. \tag{A3}$$

We will actually demonstrate this bound for the output z, which often becomes the a in a later multiplication. This formula also applies to e_b. We choose to reduce the bound on e_n by 1 bit by the expense of calculating e_n exactly and modifying n' if needed to reduce $|e_n|$. Now in Step 1, $h' = (a + e_a)(b + e_b)(n + e_n)^{-1} + e^*$, where

The Algorithm

1.	$h' \leftarrow \lfloor (a'b')/n' \rfloor$
2.	$\mathbf{y} \leftarrow [h']$ (if $h' > 2^{64}$ then compute l, v from $h' \sim 2^l v$ and use $[2^l][v]$)
3.	$\mathbf{x} \leftarrow \mathbf{ab} - \mathbf{yn}$ (so $\mathrm{CRT}(\mathbf{x}) = ab - h'n$, only reduced mod n if $h' = h$)
4.	$\mathbf{w} \leftarrow \mathbf{xR}$
5.	$g' \leftarrow\sim \mathbf{w} \cdot \mathbf{M} - lM_0 \in (-M_0/2, M_0/2)$ where, to reduce FP errors when $k = 3$, $g' \sim ((w_1M_1 + w_2M_2 - l_2M_0) + w_3M_3 - l_3M_0) - l_4M_0$ for each $l_j \in \{0, 1\}$; g' is an estimate of $ab - h'n$ but $\notin (0, n')$ under certain FP error patterns
6.	Let $q \leftarrow \lfloor g'/n' \rfloor, f' \leftarrow\sim g' - qn', \mathbf{z} \leftarrow \mathbf{x} - q\mathbf{n}$ (revising Step 3 and in practice avoiding arithmetic if $q = 0$)
7.	$q_1 \leftarrow \lfloor (f' - z_1)/P_1 + \frac{1}{2} \rfloor$, $d' \leftarrow\sim d = z_1 + q_1P_1 = f' + j$, where $\lvert j \rvert < P_1/2$; $d \equiv z_1 \bmod P_1$ is a closer estimate of $\mathrm{CRT}(\mathbf{z})$; set $z' = d'$ for tentative output $(\mathbf{z}, z')$
8.	If $c > 25$ then do the following, letting $e \leftarrow P_1 - P_2 > 0$: let $d_2 = z_1 + (q_1 \bmod P_2)(P_1 \bmod P_2) = (z_1 + e(q_1 \bmod P_2)) \bmod P_2$; if $d_2 \neq z_2$ then solve i in $z_2 = (d_2 + iP_1) \bmod P_2 = (d_2 + ie) \bmod P_2$, so $i \leftarrow (z_2 - d_2 \bmod^{\pm} P_2)/e$ where $\bmod^{\pm}$ reduces to a centred interval; set $z' \leftarrow d' \leftarrow\sim z_1 + (q_1 + i)P_1$
9.	$\mathbf{u} \leftarrow \mathbf{z} - \mathbf{n}$; if $\mathbf{u}$ is of the form (j, j, j) then $\mathbf{z} \leftarrow \mathbf{u}, z' \leftarrow z' - n'$
10.	If $n > E$, compute bounds on the error in h' knowing n', a', b'; if l can then possibly be in error, refine the error bounds successively by computing errors in a', b', n', and if the possibility of error remains, compute $l_{10} \in \{-1, 0, +1\}$ such that $\mathbf{w} \cdot \mathbf{M} - (l + l_{10}M_0 - qn)$ is inside $(-M_0/2, M_0/2)$, and if $l_{10} \neq 0$ then redo Steps 6 to 9 (see the Error Analysis section for further details)

e^* is the error from the FP multiplication, division, and flooring. Then

$$\begin{aligned} e_h &= h' - h \\ &= (ab + e_ab + e_ba + e_ae_b)(1 - e_n/n + e_n^2/n^2 + \ldots)/n + e^* - (ab - z)/n \\ &= (e_ab + e_ba)/n - abe_n/n^2 + e^* + z/n + O(1/n). \end{aligned} \tag{A4}$$

Then ignoring $O(1/n)$ terms and knowing that $0 < a, b < n$,

$$\lvert e_h \rvert < \lvert e_a \rvert + \lvert e_b \rvert + \lvert e_n \rvert + \lvert e^* \rvert + 1 \leq \frac{5}{2}\delta 2^{\lceil m \rceil} + \lvert e^* \rvert + 1.$$

For e^*, $a'b'$ attracts a FP error up to $\delta 2^{\lceil 2m \rceil - 1}$, by the IEEE mandated rounding. When that result is divided by n', that error gets so divided also, giving $\delta 2^{\lceil 2m \rceil - m - 1}$, and a new error up to $\delta 2^{\lceil m \rceil - 1}$ occurs on the output which is $\sim a'b'/n' < 2^m$. Hence

$$\begin{aligned} \lvert e^* \rvert &< E^* = \delta(2^{\lceil m \rceil} + 2^{\lceil 2m \rceil - m})/2, \\ \lvert e_h \rvert &< E_h = \delta(3 \cdot 2^{\lceil m \rceil} + 2^{\lceil 2m \rceil - m - 1}) + 1. \end{aligned} \tag{A5}$$

Now in Step 5, $g' = \mathrm{CRT}'(\mathbf{x})$ is the estimate of $ab - h'n = \mathrm{CRT}(\mathbf{x})$. Also, $w_i M_i < M_0 < 2^{3c}$ for each i, and its error is less than $r = \delta 2^{\lceil 3c \rceil}/2$. When the first two are added together their error bounds add, and being up to $2M_0$, the error from the addition is double the individual ones, for a total of $4r$. The subtraction of $l_2 M_0$ ensures the result is less than M_0 again, so the addition of $w_3 M_3$ adds r for its own error plus $2r$ as before for the sum. The total is now $7r$, but there are up to three subtractions of M_0, each introducing an identical error $e_{M_0} = M_0' - M_0$ (total $\to 7r + 3e_{M_0}$).

We evaluate e_{M_0} as follows. Let $C = \lceil 3c \rceil$. Then M_0 has C bits and may be written $M_0 = 2^{C-53}x - y$ with $2^{52} < x < 2^{53}$, $-2^{C-54} < y < 2^{C-54}$; x is exactly the FP mantissa, and y is the FP error $M_0' - M_0$. Then $e_{M_0} = y = M_0 \bmod^{\pm} 2^{C-54}$, where $\bmod^{\pm}$ means reduction into a centred interval. In practice we only use the values $c = 25,\ 26,\ 26.5$, and using the top three primes below each of these bounds gives, respectively,

$$e_{M_0} = 116571,\ 6075,\ 15426731 \text{ for } c = 25, 26, 26.5. \tag{A6}$$

The overall error bound from Step 5 is then

$$e_g = g' - \mathrm{CRT}(\mathbf{x}),\ \text{with } |e_g| < E_g = 7r + 3e_{M_0},\ \text{where } r = \delta 2^{\lceil 3c \rceil - 1}. \tag{A7}$$

In Step 6 g' gets reduced mod n' to give f', and the same quotient q subtracts qn from $\mathbf{x}$ to give $\mathbf{z}$. So we have:

$$\begin{aligned} f' &= (g' - qn')' \\ &= (\mathrm{CRT}(\mathbf{x}) + e_g - qn')' \\ &= (\mathrm{CRT}(\mathbf{z}) + e_g - q(n - n'))' \\ &= \mathrm{CRT}(\mathbf{z}) + e_g - qe_n + e_6 \\ &\quad \text{(where } e_6 \text{ is FP error from this step 6 calculation)}, \\ e_f &= f' - \mathrm{CRT}(\mathbf{z}) \\ &= e_g - qe_n + e_6. \end{aligned} \tag{A8}$$

Bounds on $e_g,\ e_n$ have already been given, so now we need to study q and e_6. We have $q = \lfloor g'/n' \rfloor \sim (g + e_g)/n' = (ab - h'n + e_g)/n' = (ab - hn - e_h n + e_g)/n'$. Here, $0 < ab - hn = z < n$, the true answer, so this term is small. Likewise by Equation (A7), $|e_g/n'| < 8\delta 2^{3c-1}/2^m \le 2^{-27}$ which occurs at $m = 51$ in Table A1, and is negligible. Therefore

$$q \sim -e_h n/n' \sim -e_h. \tag{A9}$$

For e_6, this is the FP error arising from $g' - qn'$, where the result is smaller in magnitude than g'. IEEE-754 requires that the subtraction be done in higher precision and then rounded to 53 bits, so the error bound is

$$E_6 = \delta 2^{\lceil \log_2(|q|n') \rceil - 1} = \delta 2^{\lceil m + \log_2(E_h) \rceil - 1} \tag{A10}$$

from Equation (A2). In Equation (A8), $|qe_n| \le E_h E_n$ by Equation (A9), so

$$|e_f| \le E_f = E_g + E_n E_h + E_6. \tag{A11}$$

The expression $e_n e_h$ contains, through Equation (A4), a quadratic term $-abe_n^2/n^2$ which is negative, and this causes the upper and lower bounds for e_f to be slightly different in magnitude; this is observed in practice – see Table A2 for the differing bounds.

Next, in Step 7 (the mod P_1 correction), the calculation of q_1 implies that the value $d = v_1 + q_1 P_1$ in full precision has the correct value mod P_1 and a size close to f'. Therefore d is initially posited to be equal to CRT($\mathbf{z}$). Then $e_d = d' - d$ arises from error in multiplication of q_1 by P_1 and error in addition of v_1. The former is, by Equation (A2) and $z' = d' \sim f'$, bounded by $\delta 2^{\lceil \log_2 z \rceil - 1}$, and the latter suffers the same zeroization of its lower bits, so the error is doubled, giving the value at Equation (A3) with a replaced by z.

If the CRT instantiation of interest is relatively rare, as occurs when the modulus n is updated, then by performing additional calculations modulo some number greater than the known error bound, the output error can be exactly determined, and then reduced to achieve the smaller bound in Equation (A2).

Now, if $n < 2^{65.5}$, then $e_d < \delta 2^{\lceil \log_2 z \rceil} < 2^{-53+66} = 8192 << P_1/2$, so error in d' is not a problem. However, $d = f' + j$ is subject to error $|e_f| < E_f$, so if $E_f > P_1/2 \simeq 2^{c-1}$ then d might be in error. Table A2 shows this to be the case for $c > 25$.

So if $c > 25$, Step 8 performs mod P_2 work to test the assumption that d is CRT($\mathbf{z}$). Now $d \bmod P_2 = v_1 + (q_1 \bmod P_2)(P_1 \bmod P_2) = (v_1 + e(q_1 \bmod P_2)) \bmod P_2$. If this equals v_2, then all is well. Otherwise, q_1 needs to be changed, say by adding i, as given in the Step 8 definition. Since $f' + j + iP_1 = f + j + (e_f + iP_1)$ with $|e_f + iP_1| < P_1/2$, the small value $-e_f/P_1$ roughly equals i, with its bound $E_f/P_1 < 9$ in all cases in Table A2. After this, d does equal CRT($\mathbf{z}$), and the bound on e_d is unchanged because $q_1 + i$ is very close to q_1, so

$$|e_d| < E_d = \delta 2^{\lceil \log_2 n \rceil}. \tag{A12}$$

Step 9 is needed because $q = \lfloor g'/n' \rfloor = \lfloor (g + e_g)/(n + e_n) \rfloor$, and this can differ by ± 1 from the correct value $q_0 = \lfloor g/n \rfloor$ which is needed to reduce the RNS value $\mathbf{x}$ properly. Step 9 prevents $\mathbf{z}$ from representing a value slightly bigger than n. This is important because the Fermat test and the order finding function respectively require comparisons with 2 and 1, so outputs $n + 2$ and $n + 1$ are undesirable. We could also test and correct for a value slightly smaller than zero, but do not since the code works fine with such numbers.

All calculations to this point assume that $l = l_2 + l_3 + l_4$ is the correct value to determine q in Step 6. But a change of ± 1 in l changes q by roughly $\pm M_0/n$. From Equation (A9), $|q| \sim |e_h| < E_h$, so if $E_h > M_0/(2n)$ then q slightly below E_h cannot be correctly distinguished from $q - M_0/(2n) > -E_h$, and an error can occur.

Step 10 corrects this by first converting FP values to CRT, taking differences with the original CRT, and converting these delta CRTs to integers (simple when the absolute value is less than P_3). For each new $e_.$ evaluated, the bounds on e_h can be re-evaluated, and if small enough Step 10 can be aborted. Otherwise, now knowing e_a, e_b, e_n exactly, $k = (e_a b + e_b a)/n - abe_n/n^2$ from Equation (A4) is known to high accuracy. Replacing l in Step 10 by $l + l_{10}$ causes q to change to $q + l_{10}M_0/n + O(1)$, so Equations (A4) and (A9) imply $e_h \sim k + e^* \sim -q - l_{10}M_0/n$ and therefore $|q + k + l_{10}M_0/n| \sim |-e^*| < E^*$. It is a requirement of the algorithm that $E^* < M_0/(2n)$, so $l_{10} = \lfloor (-q-k)n/M_0 + \frac{1}{2} \rfloor$ is the correct value – of course, this is often zero.

c	m	E_{m0}	E_n	E_a	B^*	E^*	E_h	nE_h^+	nE_h^-	E_g	E_6	E_f^+	E_f^-
25	61.51	16.83	8	9	12.49	9.27	10.89	18.89	17.77	23.84	19	23.935	23.870
25	61.99	16.83	8	9	12.01	9.01	10.81	18.81	17.59	23.84	19	23.933	23.872
25	62.01	16.83	9	10	11.99	9.99	11.81	20.81	19.58	23.84	20	*24.094*	23.866
26	62.01	12.57	9	10	14.99	9.99	11.81	20.81	19.58	26.81	20	26.843	26.811
26	63.99	12.57	10	11	13.01	11.01	12.81	22.81	21.59	26.81	23	26.989	26.870
26	64.01	12.57	11	12	12.99	11.99	*13.81*	24.81	23.58	26.81	24	27.285	26.859
26.5	64.01	23.88	11	12	14.49	11.99	13.81	24.81	23.58	28.94	24	29.066	28.955
26.5	64.64	23.88	11	12	13.86	12.19	*13.87*	24.87	23.72	28.94	25	29.112	28.997
26.5	65.49	23.88	12	13	13.01	12.78	14.75	26.75	25.44	28.94	27	29.507	29.172
26.5	65.51	23.88	12	13	12.99	*13.27*	14.89	26.89	25.77	28.94	27	29.529	29.144

Table A2: $\log_2$ error bounds affecting the algorithm

Table A2 displays the various $\log_2$ error bounds as functions of c and m. A negative superscript such as in E_f^- denotes a negative bound, with the log of the absolute value displayed. The B^* column gives the value 2^{3c-m-1}, which needs to be greater than E^* to avoid occasional errors. Four entries are in italics, since they indicate when optional steps are needed or when the value of c changes, as follows.

1. $c = 25,\ m = 62.01,\ \log_2 E_f+ = 24.094 > c - 1$, indicating that $m > 62$ requires Step 8, though we remark that so many extreme events have to occur to attain $e_f > 2^{24}$ that in a 10-hour run with $m = 62.1$ and $c = 25$, $e_f < 2^{23.94}$ was observed, so no errors occurred. Table A1 switches c from 25 to 26 at this point; that does not avert the need for Step 8, but does delay the need for Step 10.

2. $c = 26,\ m = 64.01,\ \log_2 E_h = 13.81 > \log_2 B^* = 12.99$ indicates that Step 10 would be needed, so Table A1 chooses $c = 26.5$ for $m \geq 64$, again to delay the need for Step 10.

3. Similarly for $c = 26.5,\ m = 64.64,\ E_h = 13.87$ is too large so Step 10 is needed for $m > 64.63$. A straightforward calculation using B^* and Equation

(A5) shows that $E_h = B^*$ at $m = 64 + \log_2 \frac{4\sqrt{2}-1}{3} = 64.6344$.

4. $c = 26.5,\ m = 65.51,\ \log_2 E^* = 13.27 > \log_2 B^* = 12.99$, which shows that, even with Step 10, $m > 65.5$ would suffer occasional errors, so this is the upper limit for this algorithm.