Yale University
Department of Computer Science

Approximating $x^2 \bmod N$ by a function $f : \mathbb{R} \to \mathbb{R}$

Renè Peralta Jatin Shah

YALEU/DCS/TR-1252
June 2003

Approximating $x^2 \bmod N$ by a function $f : \mathbb{R} \to \mathbb{R}$

Renè Peralta* Jatin Shah†

Abstract

In this paper, we identify and explain the patterns in the scatter diagram of the function $x^2 \bmod N$ when $0 \leq x \leq N$. The dominant patterns are parabolas and appear in the plots when $N + k$, where $k = 1, 2, \ldots$ is a small constant.

1 Introduction

The function $x^2 \bmod N$ has important applications to cryptology. It is the basis of Rabin's encryption scheme, of the cryptographically secure pseudo-random number generator of Blum, Blum, and Shub [1], and of the Quadratic Residuosity Assumption [4] and it's associated bit-commitment schemes. These applications are possible in part because the complexity of factoring N is probabilistic polynomial-time equivalent to inverting $x^2 \bmod N$. Intuitively, all this would seem to imply that the plot of this function should look like a scatter diagram for $x \in \Theta(N)$. Some versions of the latter assertion have indeed been proven (see [3], [6]). It is therefore quite surprising that plots of this function for specific N show a rich structure (such as seen in figure 1 below). This project aims at explaining some of the observed structure.

2 Notation

We define the operator $[t]_N : \mathbb{Z} \longrightarrow \{0, 1, \ldots, N-1\}$ to map t to the unique integer in the range $\{0, 1, \ldots, N-1\}$ congruent to t modulo N. We extend this notation so that it applies to residue classes modulo (a multiple of) N

*Department of Computer Science, Yale University, CT
†Department of Computer Science, Yale University, CT

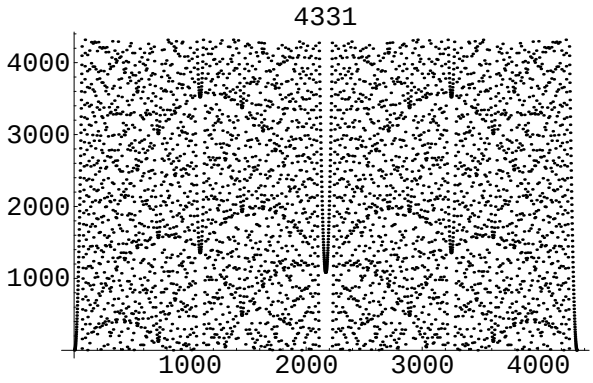

Figure 1: Plot of x^2 mod 4331.

as well as to integers. There is no ambiguity introduced here as $[x]_N$ takes the same value whether we consider x as an integer or an element of $\mathbb{Z}_N$ (or of $\mathbb{Z}_{kN}$). We define $[x^{-1}]_N = [y]_N$ where y is the multiplicative inverse of x modulo N. If $GCD(x, N) \neq 1$, then $[x^{-1}]_N$ is undefined. The notation $[\frac{a}{b}]_N$ will mean $[ab^{-1}]_N$, provided $[b^{-1}]_N$ exists. If b is not invertible modulo N, then the expression $[\frac{a}{b}]_N$ is disallowed. For example, $[\frac{6}{2}]_4$ is *not* equal to $[3]_4$.

Suppose $g : \mathbb{Z}_N \to \mathbb{Z}_N$, $f : \mathbb{R} \to \mathbb{R}$, and $\Lambda \subseteq \mathbb{Z}_N$. The notation $g =_\Lambda f$ will mean $g(x) = f(x)$ for $x \in \Lambda$. If this is the case, we say g is *embedded* in f (at Λ) and we say f is *host* to g (at Λ).

3 The functions $[(x_0 + si)^2]_N$

Let $\Lambda = \{L, L+1, \ldots, 0, \ldots H-1, H\}$. Suppose there exists s, x_0 such that for all $i \in \Lambda$ we have

$$0 \leq [s^2]_N \; i^2 + [2x_0 s]_N \; i + [x_0^2]_N < N. \tag{1}$$

Then $[(x_0 + si)^2]_N =_\Lambda ai^2 + bi + c$ where

$$\begin{array}{rcl} a & = & [s^2]_N \\ b & = & [2x_0 s]_N \\ c & = & [x_0^2]_N. \end{array} \tag{2}$$

Values of L, H that satisfy inequality (1) are:

$$\begin{array}{rcl} L & = & \left\lceil \frac{-b-\sqrt{b^2-4a(c-N)}}{2} \right\rceil \\ H & = & \left\lfloor \frac{-b+\sqrt{b^2-4a(c-N)}}{2} \right\rfloor \end{array} \tag{3}$$

Example 1 *For odd N, let $x_0 = \frac{N+1}{2}$ and $s = 1$. The reader can verify that*

$$[(x_0 + si)^2]_N = [(\frac{N+1}{2} + i)^2]_N =_\Lambda i^2 + i + c \tag{4}$$

where $c = \begin{cases} \frac{3N+1}{4} & : \quad [N]_4 = 1 \quad \textit{and } L, H \approx \sqrt{N}/2 \\ \frac{N+1}{4} & : \quad [N]_4 = 3 \quad \textit{and } L, H \approx \sqrt{3N}/2 \end{cases}$

Thus, the function $g(i) = [(\frac{N+1}{2} + i)^2]_N$ is embedded in the real-valued function $f(i) = i^2 + i + c$ for i in a range of size $\Theta(\sqrt{N})$ around 0. Changing variables via $x = x_0 + i$ we see that the graph of $[x^2]_N$ around $x = x_0 = \frac{N+1}{2}$ is embedded in a horizontal translation of the parabola $f(i) = i^2 + i + c$. Now, $f'(i) = 2i + 1 = 0$ gives us the vertex of the parabola. It occurs at $i = -0.5$ or, equivalently, at $x = \frac{N}{2}$. The vertical coordinate of the vertex is at $c - \frac{1}{4}$, which is $\frac{3N}{4}$ when $[N]_4 = 1$ and $\frac{N}{4}$ when $[N]_4 = 3$. Figure 1 shows the graph of x^2 mod 4331. As predicted, a parabolic pattern can be clearly seen around $x = \frac{4331}{2}$. Since 4331 is congruent to 3 modulo 4, the "height" of the parabolic pattern is close to $\frac{4331}{4}$. The reader will note that there are many other seemingly parabolic patterns in the graph. We will explain these later in this report.

3.1 Embeddings of $\left[\left(\frac{u}{2s} + r + si\right)^2\right]_N$

A large set of parabolic embeddings is described by letting $x_0 = [\frac{u}{2s} + r]_N$ to obtain

$$[(x_0 + si)^2]_N = \left[\left(\frac{u}{2s} + r + si\right)^2\right]_N \tag{5}$$

under the following restrictions

$$\begin{array}{lcl} GCD(2s, N) & = & 1 \\ u & \in & \{1, 2, \ldots, 2s-1\} \\ GCD(u, 2s) & = & 1 \\ r & \in & \{0, 1, \ldots, s-1\} \\ s & << & \sqrt{N}. \end{array} \tag{6}$$

The condition $s << \sqrt{N}$ guarantees $u + 2rs < N$. This in turn allows us to write

$$[x_0]_N = \frac{[\frac{-u}{N}]_{2s} N + u}{2s} + r.$$

Also, it is not hard to verify that $[(x_0+si)^2]_N = [(\frac{u}{2s}+r+si)^2]_N = ai^2+bi+c$ with

$$\begin{aligned} a &= s^2 \\ b &= u + 2sr \\ c &= [x_0^2]_N \end{aligned} \tag{7}$$

for integer i such that $0 \leq ai^2 + bi + c < N$. Thus, $[x^2]_N$ is embedded in a horizontal translation (and vertical scaling by a factor s) of the parabola $ai^2 + bi + c$, considered as a curve in the Euclidean plane. The magnitude of the horizontal translation is given by x_0 (see figure 2).

We now partition the set of parabolas characterized by equation (5). The dominant term in $[x_0]_N = \frac{[\frac{-u}{N}]_{2s} N + u}{2s} + r$ is $\frac{[\frac{-u}{N}]_{2s} N}{2s}$. As u takes on all values in $\mathbb{Z}^*_{2s}$, the value of $[\frac{-u}{N}]_{2s}$ takes on all values in $\mathbb{Z}^*_{2s}$ as well. Since $[x_0]_N$ determines the horizontal translation of the host parabola, we conclude the set of host parabolas can be partitioned into $\phi(2s)$ subsets, one for each value of u. Within each subset, the host parabolas (there are s of them) are located at approximately the same horizontal position.

We now consider the set of host parabolas determined by a given u. The vertex of the host parabola is at $i = i_0 = \frac{-b}{2a} = \frac{-(u+2sr)}{2s^2}$. Since $1 \leq u \leq 2s-1$ and $0 \leq r \leq s-1$, we have $-1 < i_0 < 0$. Thus $[(x_0 + si)^2]_N$ is closest to the vertex of the parabola when $i = 0$ or -1 (equivalently, when $x = x_0$ or $x_0 - s$). Therefore, the vertical coordinate of the vertex is approximately at $[x_0^2]_N$. (see figure 2).

Note

$$\begin{aligned} [x_0^2]_N &= \left[\frac{u^2}{4s^2} + r^2 + \frac{ur}{s}\right]_N \\ &= \left[V + r^2 + \left\lfloor \frac{ur}{s} \right\rfloor + \frac{[ur]_s}{s}\right]_N \\ &= \left[V + r^2 + \left\lfloor \frac{ur}{s} \right\rfloor + \left(\frac{[-N^{-1}]_s N + 1}{s}\right)[ur]_s\right]_N \\ &= \left[V + r^2 + \left\lfloor \frac{ur}{s} \right\rfloor + \left\lceil \frac{tN}{s} \right\rceil [ur]_s\right]_N \end{aligned} \tag{8}$$

where $V = \left[\frac{u^2}{4s^2}\right]_N$ and t is an integer strictly between 0 and s and independent of r. Therefore, as a function of r, the last expression is dominated by the term $\lceil \frac{tN}{s} \rceil [ur]_s$. As r ranges from 0 to $s-1$, and since $GCD(u, s) = 1$, $[ur]_s$ takes on all values between 0 and $s-1$ as well.

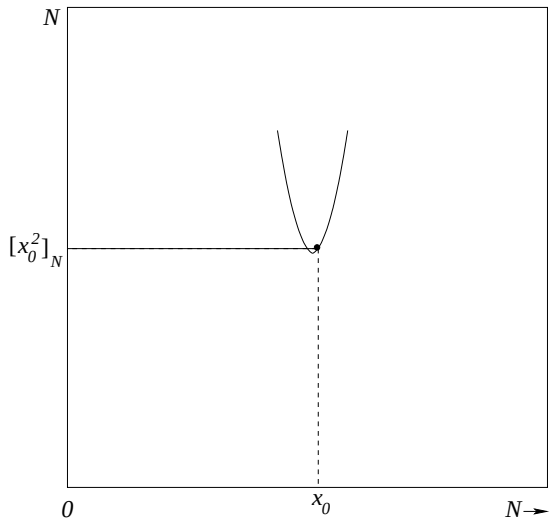

Figure 2: Location of host parabola when $x_0 = \left[\frac{u}{2s} + r\right]_N$. Note that the other $s-1$ parabolas for this value of u are not shown in the figure.

We summarize our observations as follows: the function $x^2 \bmod N$ is embedded into a rather large set of parabolas on the Euclidean plane. A subset of these host parabolas can be found using the change of variable $x = \left[\frac{u}{2s} + r + si\right]_N$ with parameters u, s, r satisfying (6). [1] The host parabola for $[(x_0 + si)^2]_N = \left[\left(\frac{u}{2s} + r + si\right)^2\right]_N$, expressed as a function of i, is $f(i) = s^2i^2 + (u + 2sr)i + \left[\left(\frac{u}{2s} + r\right)^2\right]_N$. The vertex of this parabola has horizontal coordinate close to $[x_0]_N = \frac{[\frac{-u}{N}]_{2s}N+u}{2s} + r$ and vertical coordinate close to $[x_0^2]_N$. The value of $[x_0^2]_N$, as a function of u, s, and r, is described by (8). Provided $s << \sqrt{N}$, there are $s\phi(2s)$ host parabolas described by this analysis. Equations (7), together with the condition $0 \leq ai^2 + bi + c < N$, tell us that there are order $\frac{\sqrt{N}}{s}$ points in each embedded parabola.

3.2 An application

Several powerful factoring algorithms like the continued fraction algorithm and the quadratic sieve algorithm are based on finding congruences of the form $x_i^2 \equiv y_i \bmod N$ such that all prime factors of y_i are smaller than

[1] An analogous set of parabolic embeddings of $[x^2]_N$ is found under the change of variables $x = \left[\frac{u}{s} + r + si\right]_N$ with s odd. The proof of this is essentially the same as above.

some parameter W. Numbers with this property are called W-*smooth*, or simply *smooth*. The set of primes smaller than W is called the *prime base*. If sufficiently many such congruences are found, they can be combined to produce a congruence of the form $X^2 \equiv Z^2 \bmod N$. If $X \neq \pm Z \bmod N$, then $GCD(X+Z, N)$ is a proper factor of N. For details of this general technique see Cohen [2, p. 477]. In this section, we give a simple method to generate congruences $x_i^2 \equiv y_i \bmod N$ so that y_i is a small number. Since, a small number is more likely to be smooth in a given prime base, we can use this method to generate congruences $x_i^2 \equiv y_i \bmod N$ such that y_i is smooth. We shall call this the *small squares problem* and the corresponding algorithm the *small squares algorithm*.

This method is based on the observation that the function $[(\frac{u}{2s}+r+si)^2]_N$ is embedded in s parabolas which are approximately at a vertical distance of $\frac{N}{s}$ from each other. We select the top-most parabola in this family and return the first point on this parabola which exceeds N.

If x_i is the x-coordinate of this point and $x_i^2 \equiv y_i \bmod N$, then we claim that y_i is $O(N^{5/8})$ (see figure 3). In fact, choosing any other parabola from the family will give us y_i with size $O(N^{5/8})$ as well[2]. However, choosing the top-most parabola will give us a smaller value on the average.

Thus, we have another easy way of generating congruences $x_i^2 \equiv y_i \bmod N$ where y_i is small. Potentially, this study could yield an alternative to the standard methods like quadratic sieve [8], multiple quadratic sieve [9], hypercube multiple quadratic sieve [7], etc.

The theorems that follow give the formula for the top-most parabola in the family of s parabolas and also show that y_i is $O(N^{5/8})$ (See page 10).

Theorem 2 *When $u = 1$, the top-most parabola in the family of parabolas given by equations (5), (6) and (7) is for*

$$r = r_0 = \left[N - \left[\frac{1}{4s} \right]_N \right]_s . \tag{9}$$

Proof: We are considering the family of parabolas given by equations (5), (6) and (7) with $u = 1$. Therefore, the vertices of parabolas in this family are close to $c = [x_0^2]_N$ where $x_0 = [\frac{1}{2s} + r]_N$. From equation (8), we have

$$[x_0^2]_N = \left[\frac{1}{4s^2} + r^2 + \frac{r}{s} \right]_N = \left[\frac{1}{4s^2} + r^2 + r \left(\frac{[-N^{-1}]_s N + 1}{s} \right) \right]_N .$$

[2]In this report, we only prove this statement for the case $u = 1$ and $s < N^{1/4}$.

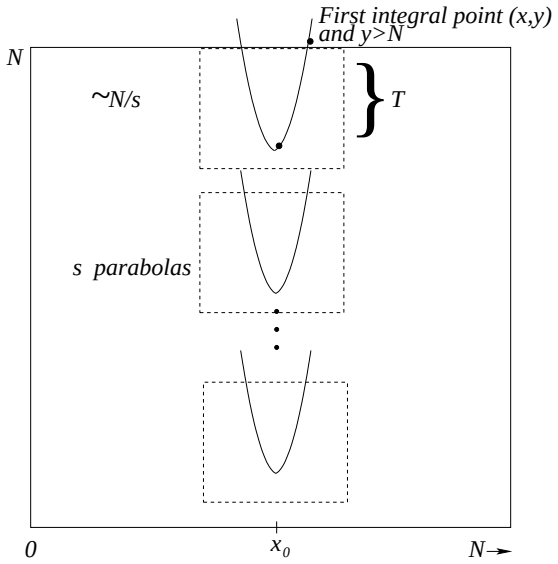

Figure 3: Generation of congruences $x_i^2 \equiv y_i \bmod N$ with small y_i.

Let $B(r) = r\left(\frac{[-N^{-1}]_s N+1}{s}\right)$ and $V = [\frac{1}{4s^2}]_N$. Assuming the value of r^2 is negligible modulo N, we have $c = [x_0^2]_N \approx [B(r) + V]_N$.[3]

The top-most parabola is for that value of r which maximizes c. We find a large c by setting $B(r)$ slightly less than $N - V$. In other words, we have to find a $r_0 \in \{0, \ldots, s-1\}$ such that $B(r_0) < N - V$ and $B(r_0) \geq B(r)$ for any $r \in \{0, \ldots, s-1\}$.

Since, $B(r) < N - V$, we have

$$r\left(\frac{[-N^{-1}]_s N + 1}{s}\right) = \frac{r[-N^{-1}]_s N}{s} + \frac{r}{s} < N - V.$$

Since $r/s < 1$, we can ignore r/s. Therefore, finding an r for which

$$\frac{r[-N^{-1}]_s N}{s} < N - V$$

holds and $B(r)$ is maximized should also suffice. Clearly, $B(r)$ is maximized when

$$r[-N^{-1}]_s = \left\lfloor \frac{(N-V)s}{N} \right\rfloor \tag{10}$$

[3]This will generally hold provided s is small enough.

and we have

$$r_0 = \left[(-N)\left\lfloor\frac{(N-V)s}{N}\right\rfloor\right]_s.$$

We shall now prove that

$$\left[(-N)\left\lfloor\frac{(N-V)s}{N}\right\rfloor\right]_s = \left[N - \left[\frac{1}{4s}\right]_N\right]_s.$$

We know that,

$$\begin{aligned}(-N)\left\lfloor\frac{(N-V)s}{N}\right\rfloor &= -(N-V)s + [(N-V)s]_N \\ &\quad (\text{by } n = m\lfloor n/m\rfloor + [n]_m).\end{aligned}$$

Therefore,

$$\begin{aligned}\left[(-N)\left\lfloor\frac{(N-V)s}{N}\right\rfloor\right]_s &= [-(N-V)s + [(N-V)s]_N]_s \\ &\quad (\text{Since } [s]_s = 0) \\ &= [[(N-V)s]_N]_s \\ &= [[Vs]_N]_s \\ &= \left[-\frac{1}{4s}\right]_N \\ &= \left[N - \left[\frac{1}{4s}\right]_N\right]_s.\end{aligned}$$

∎

This theorem immediately leads us to the following result about the vertical coordinate of the vertex of the top-most parabola.

Theorem 3 *If $u = 1$ and $N - T$ is the vertical coordinate of the vertex of the top-most parabola of the family of parabolas given by equations (5), (6) and (7), then T is approximately $\frac{N - \left[\frac{1}{4s}\right]_N}{s}$.*

Proof: Since $u = 1$, the vertices of the parabolas in the family given by equations (5), (6) and (7) are close to $c = [x_0^2]_N$ where $x_0 = [\frac{1}{2s} + r]_N$. Let $B(r) = r\left(\frac{[-N^{-1}]_s N+1}{s}\right)$ and $V = [\frac{1}{4s^2}]_N$. From equation (8), we have

$$[x_0^2]_N = \left[\frac{1}{4s^2} + r^2 + \frac{r}{s}\right]_N = \left[\frac{1}{4s^2} + r^2 + r\left(\frac{[-N^{-1}]_s N + 1}{s}\right)\right]_N.$$

We can assume that the value of r^2 is negligible modulo N, which gives us $c = [x_0^2]_N \approx [B(r) + V]_N$. Moreover, since $r/s < 1$, we have $B(r) = r\left(\frac{[-N^{-1}]_s N+1}{s}\right) \approx r\frac{[-N^{-1}]_s N}{s}$. For the top-most parabola we know, from equation (10), that $r[-N^{-1}]_s = \left\lfloor \frac{(N-V)s}{N} \right\rfloor$. Therefore

$$\begin{aligned} B(r) &\approx \left\lfloor \frac{(N-V)s}{N} \right\rfloor \frac{N}{s} \\ &\approx ((N-V)s - [(N-V)s]_N)\,\tfrac{1}{s} \\ &= N - V - \frac{[(N-V)s]_N}{s} \\ &= N - V - \frac{N - \left[\frac{1}{4s}\right]_N}{s}. \end{aligned}$$

Then $c \approx N - T \approx [B(r) + V]_N = N - \frac{N - \left[\frac{1}{4s}\right]_N}{s}$, and $T \approx \frac{N - \left[\frac{1}{4s}\right]_N}{s}$ as claimed. ∎

Theorem 4 *For a fixed u, if the small-squares algorithm[4] for the family of parabolas given by equations (5), (6) and (7) returns x as the small square, then $x^2 \bmod N$ is at most $s^2 + 2s\sqrt{T}$, where $N - T$ is the vertical coordinate of the vertex of the top-most parabola.*

Proof: Following equations (5), (6) and (7), let $(\left[\frac{u}{2s}\right]_N + r + si, ai^2 + bi + c)$ denote the parametric equation of the family of s parabolas chosen. Let us assume that the top-most parabola is seen for $r = r_0$. According to equation (7) we have,

$$\begin{aligned} a &= s^2 \\ b &= 2r_0 s + u \\ c &= \left[\frac{u^2}{4s^2} + r_0^2 + \frac{ur_0}{s}\right]_N. \end{aligned} \tag{11}$$

The vertical coordinate of the vertex of the top-most parabola is at

$$y = N - T = c - \frac{b^2}{4a}. \tag{12}$$

Let i_0 denote the parameter of the first point which exceeds N and therefore is the small square returned by this algorithm, i.e. i_0 denotes the first point when $ai^2 + bi + c > N$.

[4]We believe that it should be possible to generalize theorem 2 giving a small-squqres algorithm for any u.

Let

$$\alpha = \frac{b + \sqrt{b^2 - 4a(c-N)}}{2a} \text{ and } \beta = \frac{-b + \sqrt{b^2 - 4a(c-N)}}{2a}.$$

Using equation (12), this simplifies to

$$\alpha = \frac{b + \sqrt{4aT}}{2a} \text{ and } \beta = \frac{-b + \sqrt{4aT}}{2a}.$$

If α or β are integers then $i_0 = \beta + 1$ or $-\alpha - 1$. Otherwise $i_0 = \lceil \beta \rceil$ or $\lfloor -\alpha \rfloor = \lceil \beta \rceil$ or $-\lceil \alpha \rceil$ depending on the one which gives the smaller y-coordinate.

If β is an integer, then $i_0 = \beta + 1$. In this case, the value of small square is bounded by $ai_0^2 + bi_0 + c - (a(i_0 - 1)^2 + b(i_0 - 1) + c) = 2ai_0 - a + b = 2a\beta + a + b = a + \sqrt{4aT} = s^2 + 2s\sqrt{T}$. For the case when $i_0 = -\alpha - 1$, the value of the small square is also $s^2 + 2s\sqrt{T}$.

When α or β are not integers, we shall substitute i_0 in $ai^2 + bi + c$. For the case, $i_0 = \lceil \beta \rceil$, we have

$$\begin{aligned} ai_0^2 + bi_0 + c &= a\lceil \beta \rceil^2 + b\lceil \beta \rceil + c \\ &= \frac{(2a\lceil \beta \rceil)^2}{4a} + \frac{b(2a\lceil \beta \rceil)}{2a} + c. \end{aligned}$$

Since, $2a\lceil \beta \rceil = 2a\beta + [-2a\beta]_{2a}$ (where the $[\,]_N$ operator is extended for reals)[5], the value of small square is (after substitution and simplification)

$$\begin{aligned} ai_0^2 + bi_0 + c - N &= \frac{\sqrt{4aT}[-2a\beta]_{2a}}{2a} + \frac{[-2a\beta]_{2a}^2}{4a} \\ &\leq a + \sqrt{4aT} = s^2 + 2s\sqrt{T}. \end{aligned}$$

When $i_0 = -\lceil \alpha \rceil$, we also get the size of small square as $s^2 + 2s\sqrt{T}$. ∎

From theorem 3, the value of T for the top-most parabola is close[6] to $\frac{N - \left[\frac{1}{4s}\right]_N}{s}$. Substituting this value of T in $s^2 + 2s\sqrt{T}$, we get the size of the small square returned by the algorithm as $s^2 + 2\sqrt{s\left[-\frac{1}{4s}\right]_N} = O(N^{5/8})$ (if $s < N^{1/4}$).

If g is *embedded* in f (at Λ) and if Λ is a sufficiently large set, then we can observe the function f in a plot of the function g for $x \in \mathbb{Z}_N$. The

[5] In general, the $[\,]_N$ operator cannot be extended for reals unless we are considering the additive group only.

[6] $T = \frac{N - \left[\frac{1}{4s}\right]_N}{s} + O(s^2)$. If we choose a $s < N^{1/4}$, then $O(s^2)$ term can be ignored.

points which lie on the function f seen in the plot are precisely the points $\{(x, g(x))|x \in \Lambda\}$. Henceforth, the informal terminology, *f is observed in the plot of g* shall mean that *g is embedded in f at* Λ over some appropriately defined set Λ.

4 When $N \pm 1$ is smooth

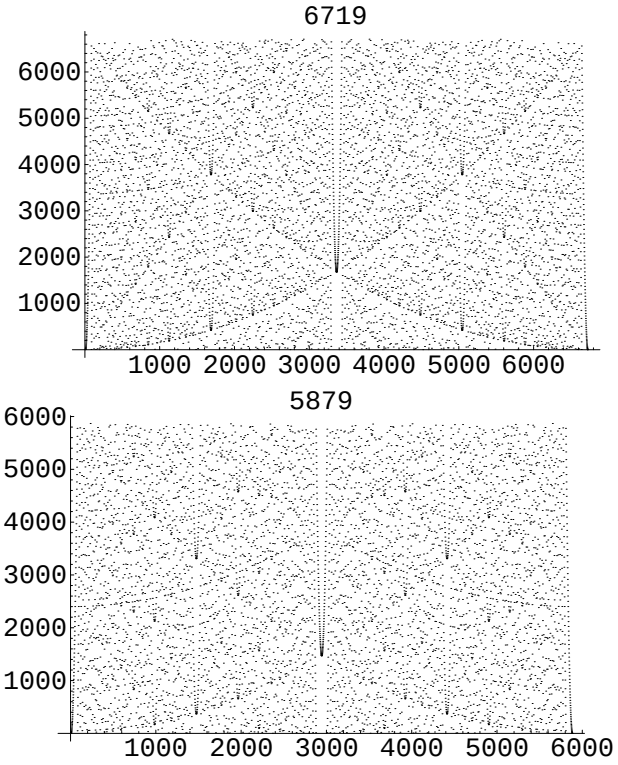

Figure 4: Plots of $y = x^2 \bmod N$ for $x = 0, \ldots, N-1$ when $N+1$ is a smooth number. Note that $6720 = 2^6.3.5.7$ and $5880 = 2^3.3.5.7^2$

In this section, we shall explain the two main parabolic patterns shown in Figure 4. These two patterns intersect at the vertex of the central parabola. Such patterns are observed in these plots when $N+1$ is a smooth number. The points on these patterns should be on or close to the parabolas $y = \frac{x^2}{N+1}$ and $y = \frac{(N-x)^2}{N+1}$ for $0 \leq x < N$. We shall prove this conjecture in theorems that follow only for the parabola $y = \frac{x^2}{N+1}$; analysis for the

other parabola is similar.

Theorem 5 *If N is such that $N+1$ is a smooth number; let $N+1 = \prod_{i=1}^{l} p_i^{\alpha_i}$, then the points $(x, [x^2]_N)$ with x-coordinate $x = \Psi u$ for $u = 0, \ldots, \sqrt{\frac{N}{\prod_{\{i|\alpha_i \ odd\}} p_i}}$ lie on the parabola $y = \frac{x^2}{N+1}$ where*

$$\Psi = \prod_{i=1}^{l} p_i^{\beta_i} \text{ and } \beta_i = \begin{cases} \alpha_i/2 & \text{if } \alpha_i \text{ even} \\ (\alpha_i+1)/2 & \text{if } \alpha_i \text{ odd} \end{cases}$$

Proof: It is easy to verify that $\Psi^2 = (N+1) \prod_{\{i|\alpha_i \text{ odd}\}} p_i$. Therefore, we have

$$\frac{x^2}{N+1} = \frac{\Psi^2 u^2}{N+1} = u^2 \prod_{\{i|\alpha_i \text{ odd}\}} p_i$$

and also

$$[x^2]_N = [\Psi^2 u^2]_N = \left[u^2(N+1) \prod_{\{i|\alpha_i \text{ odd}\}} p_i \right]_N = \left[u^2 \prod_{\{i|\alpha_i \text{ odd}\}} p_i \right]_N$$

Since $u < \sqrt{\frac{N}{\prod_{\{i|\alpha_i \text{ odd}\}} p_i}}$, $[x^2]_N = u^2 \prod_{\{i|\alpha_i \text{ odd}\}} p_i$. ∎

For example, if $N = 6719$, we have $N + 1 = 2^6.3.5.7$. If we choose $\Psi = 2^3.3.5.7 = 840$, then the points $(x = 840u, [x^2]_N)$ for $u = 0, \ldots, 7$ lie on the parabola $y = \frac{x^2}{6720}$.

We have now shown that the points with x-coordinate $x = \Psi u$ lie on the parabola $y = \frac{x^2}{N+1}$. However, there are several other points which are close but do not lie on this parabola. These points play a significant role in creating the pattern. In the following theorem we shall show that if $x^2 \equiv a \bmod (N+1)$ then the points with this x-coordinate lie on the parabola $y = \frac{x^2 + aN}{N+1}$. Later, we shall count the number of such points when a is a quadratic residue mod $N+1$.

Theorem 6 *The points $(x, [x^2]_N)$ with x-coordinate which satisfies $x^2 \equiv a \bmod (N+1)$, $a \in \mathbb{Z}_{N+1}$ and $x < \sqrt{N^2 + (1-a)N}$ lie on the parabola $y = \frac{x^2 + aN}{N+1}$.*

Proof: Let $x^2 = (N+1)k + a$. Since the x-coordinates of the points satisfy the equation $x^2 \equiv a \bmod (N+1)$, then $x^2 + aN \equiv a + aN \equiv 0 \bmod (N+1)$. Hence, $\frac{x^2+aN}{N+1}$ is an integer and $x^2 \equiv \frac{x^2+aN}{N+1} \bmod N$. Since, $x < \sqrt{N^2 + (1-a)N}$, $\frac{x^2+aN}{N+1} < N$ and therefore $[x^2]_N = \frac{x^2+aN}{N+1}$. ∎

Therefore, the solutions of the congruence $x^2 \equiv a \bmod (N+1)$ correspond to points $(x, [x^2]_N)$ which lie on the parabola $y = \frac{x^2+aN}{N+1} \approx \frac{x^2}{N+1} + a$. For small values of a, these points lie close to the parabola $y = \frac{x^2}{N+1}$. These points will give rise to the patterns observed in Figure 4. Suppose $N+1 = \prod_{i=1}^{l} p_i^{\alpha_i}$, then by the Chinese Remainder Theorem we know that the product of the number of solutions of the congruences $x^2 \equiv a \bmod p_i^{\alpha_i}$ for $i = 0, \ldots, l$, denote the number of points on the parabola $y = \frac{x^2+aN}{N+1}$. Hence, we will have several points lying close to the parabola and this leads to the formation of the specific pattern. The following theorems count the number of solutions of the congruences of the form $x^2 \equiv a \bmod p^e$ when p is a prime.

Theorem 7 *For an odd prime p, $x^2 \equiv a \bmod p^e$, $e \geq 1$, $a \in \mathbb{Z}_{p^e} - \{0\}$, has a solution iff $a^{\phi(p^e)/2} \equiv 1 \bmod p^e$. When a solution exists there are exactly 2 solutions.*

Proof: The theorem follows directly from a more general theorem on n^{th} power residues in Ireland and Rosen [5, p. 45]. ∎

Theorem 8 *For a prime p, $x^2 \equiv 0 \bmod p^e$, $e \geq 1$ has $p^{e/2}$ solutions if e is even and $p^{(e-1)/2}$ solutions if e is odd.*

Proof: The case $e = 1$ is clear, For $e \geq 2$, $x = 0$ is a solution. Let x_0 be a non-zero solution of the congruence $x^2 \equiv 0 \bmod p^e$ and p^l be the highest power of p in x_0. $p^e | x_0^2$ implies that $2l \geq e$. Therefore, x_0 is of the form $t_0 p^{e/2}$ if e is even and $t_0 p^{(e+1)/2}$ if e is odd for some non-zero t_0. Since $x_0 \in \mathbb{Z}_{p^e}$, the possible values of t_0 are $1, \ldots, p^{e/2} - 1$ if e is even and $1, \ldots, p^{\frac{e-1}{2}} - 1$ if e is odd. Thus, the number of solutions is as claimed. ∎

In particular, when $p = 2$ and $e \geq 1$ the number of solutions to the congruence $x^2 \equiv 0 \bmod 2^e$ is $2^{e/2}$ if e is even and $2^{(e-1)/2}$ if e is odd.

Lemma 9 *The number of solutions of the congruence* $x^2 \equiv 4a \bmod 2^e$, $e \geq 3$, $4a \in \{0, \ldots, 2^e - 1\}$ *is twice the number of solutions of* $x^2 \equiv a \bmod 2^{e-2}$.

Proof: If x_0 is a solution of the congruence $x^2 \equiv a \bmod 2^{e-2}$, then it is easy to verify that $2x_0$ and $2^{e-1} + 2x_0$ are distinct solutions of the congruence $x^2 \equiv 4a \bmod 2^e$. It is also not hard to see that distinct solutions of $x^2 \equiv a \bmod 2^{e-2}$ correspond to distinct solutions to $x^2 \equiv 4a \bmod 2^e$ according to the mapping given above.

We shall now prove that there are no more solutions to the congruence $x^2 \equiv 4a \bmod 2^e$. It can be shown that any solution to this congruence is even[7]. So, if $2x_0$ is a solution to $x^2 \equiv 4a \bmod 2^e$ then, it implies that $x_0^2 \equiv a \bmod 2^{e-2}$, thereby showing that we have counted every solution. ∎

Corollary 10 *The number of solutions of the congruence* $x^2 \equiv 2^k b \bmod 2^e$, $e \geq 3$, $k < e$ *is* $2^{k/2}$ *times the number of solutions of the congruence* $x^2 \equiv b \bmod 2^{e-k}$ *if* k *is even.*

Proof: Apply Lemma 9 $k/2$ times. ∎

Lemma 11 *There are no solutions to the congruence* $x^2 \equiv 2a \bmod 2^e$, $e \geq 2$, $2a \in \mathbb{Z}_{2^e}$, *when* a *is odd.*

Proof: If x_0 is a solution of the congruence $x^2 \equiv 2a \bmod 2^e$ then $x_0^2 = 2a + 2^e k$ for some k. Hence, x_0 has to be even. Let $x_0 = 2x_1$. On substitution and simplification, we have $2x_1^2 = a + 2^{e-1}k$ which implies that a is even. A contradiction. ∎

Now, we shall count the solutions of the congruence $x^2 \equiv a \bmod 2^e$. It is easy to verify that for $e \leq 2$, the solutions exists only when a is 0 or 1. When $e = 1$, the solutions are $x = 0$ for $a = 0$ and $x = 1$ for $a = 1$. When $x = 2$, the solutions are $x = 0, 2$ for $a = 0$ and $x = 1, 3$ for $a = 1$. The theorems that follow consider the case when $e \geq 3$.

Theorem 12 *The equation* $x^2 \equiv a \bmod 2^e$, $a \in \mathbb{Z}_{2^e}$ *and odd,* $e \geq 3$ *has solutions iff* $a \equiv 1 \bmod 8$. *When a solution exists there are 4 solutions.*

Proof: The theorem follows directly from a more general theorem in Ireland and Rosen [5, p. 46]. ∎

[7]Since $a \in \mathbb{Z}_N$ denotes a residue class we can classify it as even or odd only if N is even.

Theorem 13 *The equation $x^2 \equiv a \bmod 2^e$ with $a = 2^k b$ where b is odd and $e \geq 3$ has solutions iff k is even and $x^2 \equiv b \bmod 2^{e-k}$ is solvable. When a solution exists, there are $2^{k/2+2}$ solutions if $e - k \geq 3$ and $(e-k)2^{k/2}$ solutions if $e - k \leq 2$.*

Proof: We know from Corollary 10 that the number of solutions of $x^2 \equiv 2^k b \bmod 2^e$ is $2^{k/2}$ times the number of solutions of the congruence $x^2 \equiv b \bmod 2^{e-k}$. If the congruence is solvable then it follows from theorem 12 that it has 4 solutions when $(e-k) \geq 3$, 2 solutions when $e - k = 2$ and 1 solution when $e - k = 1$. Moreover, Lemma 11 implies that there will be no solutions when k is odd. ∎

Congruence ($a \neq 0$)	Number of solutions	Necessary and Sufficient Conditions
$x^2 \equiv a \bmod p^e$, p odd, $e \geq 1$	2	$a^{\phi(p^e)/2} \equiv 1 \bmod p^e$
$x^2 \equiv a \bmod 2^e$, $e = 1$	1	a odd
$x^2 \equiv a \bmod 2^e$, $e = 2$	2	$a \equiv 1 \bmod 4$
$x^2 \equiv a \bmod 2^e$, a odd, $e \geq 3$	4	$a \equiv 1 \bmod 8$
$x^2 \equiv 2^k b \bmod 2^e$, b odd, $e \geq 3$ $1 \leq (e-k) \leq 2$	$(e-k)2^{k/2}$	k even, $x^2 \equiv b \bmod 2^{e-k}$ has a solution
$x^2 \equiv 2^k b \bmod 2^e$, b odd, $e \geq 3$ $(e-k) \geq 3$	$2^{k/2+2}$	k even, $x^2 \equiv b \bmod 2^{e-k}$ has a solution

Congruence	Number of solutions	Conditions
$x^2 \equiv 0 \bmod p^e$	$p^{e/2}$	e even
$x^2 \equiv 0 \bmod p^e$	$p^{(e-1)/2}$	e odd

Table 1: Number of solutions of congruence $x^2 \equiv a \bmod p^e$, p prime.

Table 1 summarizes the results of theorems given above. Let us consider the specific example of $N = 6719$. In this case, $N + 1 = 6720 = 2^6.3.5.7$. We know that there are 2 solutions each to the congruences $x^2 \equiv 1 \bmod p$ where $p = 3, 5$ or 7. We also know that $x^2 \equiv 1 \bmod 2^6$ has 4 solutions. We can therefore conclude that the total number of points lying on the parabola $y = \frac{x^2+N}{N+1}$ is 32. Similarly, the total number of points lying on the parabola $y = \frac{x^2+4N}{N+1}$ is $2^3.8 = 64$.

Similarly, if $N - 1$ is a smooth number we observe the parabolas with concavities reversed as shown in Figure 5.

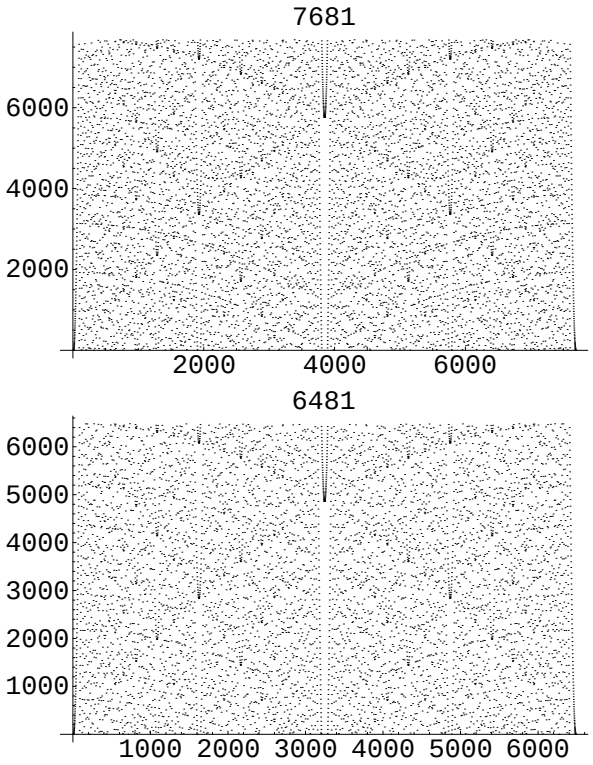

Figure 5: Plots of $y = x^2 \bmod N$ for $x = 0, \ldots, N-1$ when $N-1$ is a smooth number. Note that $7680 = 2^9.3.5$ and $6480 = 2^4.3^4.5$

5 When $N \pm c$ is smooth

The patterns shown in Figure 6 can be explained by generalizing the results of Section 4. In this case, if $N + c$ is a smooth number then the plots of $y = x^2 \bmod N$ have several points which lie on or close to the translations of parabolas $y = \frac{cx^2}{N+c}$ along the X and Y axes. We will prove this claim in the next theorem.

Theorem 14 *If c and $N + c$ are relatively prime[8] to N, then for every $0 \leq k \leq c$ the points $(x, [x^2]_N)$ with x-coordinate $x = \Psi u - k$ for $\frac{k}{\Psi} \leq u < \frac{N+k}{\Psi}$ and $\Phi u^2 c - 2k\Psi u + k^2 < N$ lie on the parabola $y - y_0 = \frac{c(x-x_0)^2}{N+c}$ where*

[8]Since, we are interested in using these results for improving the current algorithms for factoring, relative primality with N is a sufficient condition for our analysis.

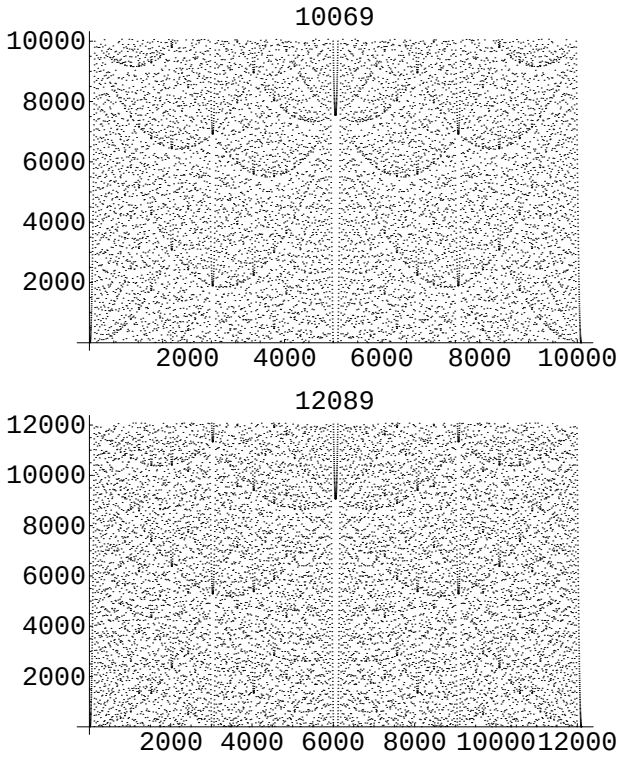

Figure 6: Plots of $y = x^2 \bmod N$ for $x = 0, \ldots, N-1$ when $N + c$ is a smooth number. Note that $10069 + 11 = 10080 = 2^5.3^2.5.7$ and $12089 + 7 = 12096 = 2^6.3^3.7$

$x_0 = \frac{kN}{c}$, $y_0 = \frac{[-k^2]_c N}{c}$ *and*

$$\Psi = \prod_{i=1}^{l} p_i^{\beta_i} \text{ and } \beta_i = \begin{cases} \alpha_i/2 & \text{if } \alpha_i \text{ even} \\ (\alpha_i + 1)/2 & \text{if } \alpha_i \text{ odd} \end{cases}$$

Proof: It is easy to verify that $\Psi^2 = (N + c)\Phi$ where $\Phi = \prod_{\{i|\alpha_i \text{ odd}\}} p_i$. Since, $0 \leq x < N$ and $x = \Psi u - k$, we get the bounds on u given in the

theorem. Now, we have,

$$\begin{aligned} y &= \frac{c(x-kN/c)^2}{N+c} + \frac{[-k^2]_c N}{c} \\ &= \frac{c(\Psi u - k - kN/c)^2}{N+c} + \frac{[-k^2]_c N}{c} \\ &= \frac{(\Psi uc - k(N+c))^2}{c(N+c)} + \frac{[-k^2]_c N}{c}. \end{aligned}$$

After expanding the square term, regrouping and some simplification we get,

$$y = \frac{\Psi^2 u^2 c}{N+c} - 2k\Psi u + k^2 + \frac{(k^2 + [-k^2]_c)N}{c}.$$

Since, $\Psi^2 = (N+c)\Phi$ and $[-k^2]_c + k^2 = cq$ for some positive integer q, we have, $y = \Phi u^2 c - 2k\Psi u + k^2 + qN$. We also know that $\Phi u^2 c - 2k\Psi u + k^2 < N$. Hence, it is an integer less than N.

It is not hard to verify that in this case,

$$x^2 \equiv y_0 + \frac{c(x-x_0)^2}{N+c} \bmod N.$$

Hence, the points $(x = \Psi u - k, [x^2]_N)$ lie on the parabola $y - \frac{[-k^2]_c N}{c} = \frac{c(x-kN/c)^2}{N+c}$. ∎

For example, if $N = 10069$, we have $N + 11 = 2^5.3^2.5.7$. If we choose $\Psi = 2^3.3.5.7 = 840$, then the points $(x = 840u - 3, [x^2]_N)$ for $u = 1, \ldots, 11$ lie on the parabola $y - 2N/11 = \frac{11(x-3N/11)^2}{N+11}$.

The following theorem characterizes the points that lie close to the parabola $y - \frac{[-k^2]_c N}{c} = \frac{c(x-kN/c)^2}{N+c}$. It is a more general form of Theorem 6.

Theorem 15 *If c and $N+c$ are relatively prime to N, then for every $0 \leq k \leq c$, the points $(x, [x^2]_N)$ with x-coordinate which satisfies $(x+k)^2 \equiv a \bmod N+c$, $a \in \mathbb{Z}_{N+c}$ and $x < x_0 + \sqrt{\frac{(N-y_0)(N+c)-aN}{c}}$ lie on the parabola $y - y_0 = \frac{c(x-x_0)^2 + aN}{N+c}$ where $x_0 = \frac{kN}{c}$ and $y_0 = \frac{[-k^2]_c N}{c}$.*

Proof: First, we shall show that if $(x+k)^2 \equiv a \bmod (N+c)$ then $y = y_0 + \dfrac{c(x-x_0)^2 + aN}{N+c}$ is an integer. In this case,

$$\begin{aligned} y &= \frac{c(x-kN/c)^2 + aN}{N+c} + \frac{[-k^2]_c N}{c} \\ &= \frac{(cx-kN)^2 + aN}{c(N+c)} + \frac{N[-k^2]_c}{c} \\ &= \frac{cx^2 - 2kxN + N^2 q + N[-k^2]_c + aN}{N+c} \\ &\quad (\text{where } k^2 + [-k^2]_c = cq \text{ for some } q \in \mathbb{Z}^+). \end{aligned}$$

Now, since $N \equiv -c \bmod (N+c)$ and $(x+k)^2 \equiv a \bmod (N+c)$, we have,

$$\begin{aligned} cx^2 - 2kxN + N^2 q + N[-k^2]_c + aN &\equiv cx^2 + 2kxc + c^2 q - c[-k^2]_c - ac \bmod (N+c) \\ &\equiv cx^2 + 2kxc + ck^2 - ac \bmod (N+c) \\ &\equiv c(x+k)^2 - ac \bmod (N+c) \\ &\equiv 0 \bmod (N+c). \end{aligned}$$

Hence, $y = y_0 + \dfrac{c(x-x_0)^2 + aN}{N+c}$ is an integer when $(x+k)^2 \equiv a \bmod (N+c)$. Since $x < x_0 + \sqrt{\frac{(N-y_0)(N+c)-aN}{c}}$, this integer is less than N.

It is also not hard to verify that in this case

$$x^2 \equiv y_0 + \frac{c(x-x_0)^2 + aN}{N+c} \bmod N$$

and hence the theorem follows. ∎

The above theorem says that the points $(x, [x^2]_N)$ such that $(x+k)^2 \equiv a \bmod (N+c)$ lie on the parabola $y - \dfrac{[-k^2]_c N}{c} = \dfrac{c(x-kN/c)^2 + aN}{N+c}$. Thus, if $N+c$ is a smooth number we can count the number of points that lie on this parabola in a manner similar to that described in section 4.

6 Future Work

In this report, we have been able to give certain embeddings of the function $[x^2]_N$. However, we have not been able to find all possible embeddings (see figure 7). In general, it would be interesting to be able to give all the embeddings of $[x^2]_N$ for a given N.

The embeddings described in sections 4 and 5 could provide some novel way of improving existing algorithms or developing new algorithms for factoring numbers.

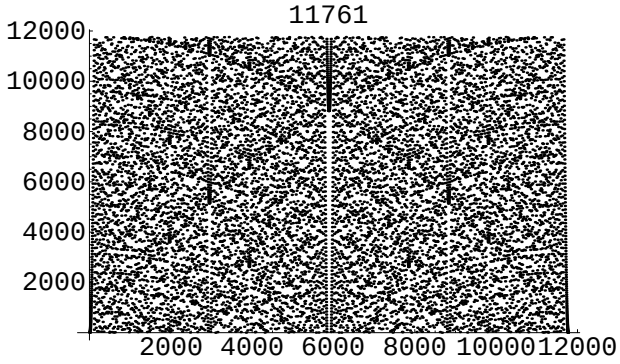

Figure 7: $11760 = 2^4.3.5.7^2$. In addition to the embedding explained in Section 4, we also see several other embedding similar to the one explained but translated by different amounts.

Acknowledgements

We would like to thank Dana Angluin for suggestions on an earlier version of this report.

References

[1] L. Blum, M. Blum, and M. Shub. A simple unpredictable pseudo-random number generator. *Siam Journal on Computing*, 15:364–383, 1986.

[2] Henri Cohen. *A Course in Computational Algebraic Number Theory.* Springer, New York, 2000.

[3] Ivan Damgård. On the randomness of legendre and jacobi sequences. In *Advances in Cryptology - Proceedings of CRYPTO 88*, pages 163–172, 1990.

[4] S. Goldwasser and S. Micali. Probabilistic encryption. *Journal of Computer and System Sciences*, 28:270–299, 1984.

[5] Kenneth Ireland and Michael Rosen. *A Classical Introduction to Modern Number Theory.* Springer - Verlag New York Inc., 1982.

[6] René Peralta. On the distribution of quadratic residues and non-residues modulo a prime number. *Mathematics of Computation*, 58(197):433–440, 1992.

[7] René Peralta. A quadratic sieve on the n-dimensional hypercube. In *Advances in Cryptology - Proceedings of CRYPTO 92*, volume 740 of *Lecture Notes in Computer Science*, pages 324–332, 1993.

[8] Carl Pomerance. The quadratic sieve factoring algorithm. In *Advances in Cryptology - Proceedings of EUROCRYPT 84*, volume 209 of *Lecture Notes in Computer Science*, pages 169–182. Springer-Verlag, 1985.

[9] Robert Silverman. The multiple polynomial quadratic sieve. *Mathematics of Computation*, 48(177):329–339, 1987.