

## FIRST-HIT ANALYSIS OF ALGORITHMS FOR COMPUTING QUADRATIC IRREGULARITY

JOSHUA HOLDEN

ABSTRACT. The author has previously extended the theory of regular and irregular primes to the setting of arbitrary totally real number fields. It has been conjectured that the Bernoulli numbers, or alternatively the values of the Riemann zeta function at odd negative integers, are uniformly distributed modulo  $p$  for every  $p$ . This is the basis of a well-known heuristic, given by Siegel, estimating the frequency of irregular primes. So far, analyses have shown that if  $\mathbf{Q}(\sqrt{D})$  is a real quadratic field, then the values of the zeta function  $\zeta_D(1 - 2m) = \zeta_{\mathbf{Q}(\sqrt{D})}(1 - 2m)$  at negative odd integers are also distributed as expected modulo  $p$  for any  $p$ . We use this heuristic to predict the computational time required to find quadratic analogues of irregular primes with a given order of magnitude. We also discuss alternative ways of collecting large amounts of data to test the heuristic.

### 1. INTRODUCTION

Let  $\mathbf{Q}(\sqrt{D})$  be a real quadratic field with  $D$  a positive fundamental discriminant. In several previous papers the author has defined an analogue for the theory of regular and irregular primes in this setting, based on the following definition:

**Definition 1.** Let  $\zeta_D$  be the zeta function for  $\mathbf{Q}(\sqrt{D})$ , and let  $\delta$  be equal to  $p - 1$  unless  $D = p$ , in which case  $\delta = (p - 1)/2$ . We say that  $p$  is  $D$ -regular if  $p$  is relatively prime to  $\zeta_D(1 - 2m)$  for all integers  $m$  such that  $2 \leq 2m \leq \delta - 2$  and also  $p$  is relatively prime to  $p\zeta_D(1 - \delta)$ . The number of such zeta-values that are divisible by  $p$  will be the *index of  $D$ -irregularity* of  $p$ .

(More generally, we may refer to the concept as “quadratic irregularity”; see [12, 13, 14] for more details and extensions to any totally real number field.)

According to a well-known theorem of Kummer,  $p$  divides the order of the class group of  $\mathbf{Q}(\zeta_p)$  if and only if  $p$  divides the numerator of a Bernoulli number  $B_{2m}$  for some even  $2m$  such that  $2 \leq 2m \leq p - 3$ . Such primes are called irregular; the others are called regular. In [13], building on work of Greenberg and Kudo, the author proved that in the setting we have described above, Kummer’s criterion can be extended to give information about whether  $p$  divides the class number (that is, the order of the class group) of  $\mathbf{Q}(\sqrt{D}, \zeta_p)$ . To be exact, we have:

---

Received by the editor November 21, 2000 and, in revised form, September 16, 2002.

2000 *Mathematics Subject Classification.* Primary 11Y40, 11Y60, 11Y16, 11R42; Secondary 11B68, 11R29, 94A60, 11R18.

*Key words and phrases.* Bernoulli numbers, irregular primes, zeta functions, quadratic extensions, cyclotomic extensions, class groups, computational number theory, cryptography.

©2003 American Mathematical Society

**Theorem 1** (Greenberg, Holden). *Assume that  $p$  does not divide  $D$ . Then  $p$  divides the class number of  $\mathbf{Q}(\sqrt{D}, \zeta_p)$  if and only if  $p$  is not  $D$ -regular.*

The main focus of this paper is in finding large  $p$  which are irregular for some  $D$ . This may be useful for cryptography, in that one common way of constructing public-key cryptographic systems is to use the difficulty of finding discrete logarithms in suitable abelian groups. In order to make sure that the discrete logarithm problem is computationally hard, one needs to know something about the structure of the group involved, e.g., that it is divisible by a large prime. Theorem 1 shows that if  $p$  is a large  $D$ -irregular prime and  $p$  does not divide  $D$ , then the class group of  $\mathbf{Q}(\sqrt{D}, \zeta_p)$  may be suitable for cryptography. We will come back to this in Section 5.

## 2. SEARCH ALGORITHMS

Suppose, for instance, that we want to find a prime  $p$  of a specified size dividing  $\zeta_D(1 - 2m)$  for some  $m$  such that  $2 \leq 2m \leq \delta - 2$  or dividing  $p\zeta_D(1 - \delta)$ . (In practical algorithms  $D$  will be much smaller than  $p$ , so we may focus on the case where  $\delta = p - 1$ . See Section 5 for the details.) More specifically, we might fix a real number  $c$  greater than 1 and then look for  $m$  and  $D$  such that  $P \leq p \leq cP$  and  $p$  divides  $\zeta_D(1 - 2m)$  for some positive  $m$  less than or equal to  $(p - 1)/2$ . (In practice,  $c = 2$  would probably be the most common choice, since that would be equivalent to specifying the size of  $p$  in number of bits.)

The algorithm that we will use to carry out this search is described in [14]. The algorithms there fall into two basic types. The first type calculates  $\zeta_D(1 - 2m)$  in a range of  $m$  for each  $D$  before going on to the next  $D$ , and it calculates each value in time  $O(m^{O(1)}D^{1+o(1)})$  when amortized over both  $D$  and  $m$ . The second type calculates  $\zeta_D(1 - 2m)$  in a range of  $D$  for each  $m$  before going on to the next  $m$ . If one keeps a table of intermediate values as described in Section 3 of [14], this algorithm can calculate each value in time  $O(m^{O(1)}L(D)^{O(1)})$  when amortized over both  $D$  and  $m$ , where  $L(x)$  is a subexponential function corresponding to a choice of factoring routine used in the calculation, e.g.,  $L(x) = e^{c(\log x)^{1/3}(\log \log x)^{2/3}}$  for the number field sieve. The facts that the amortized times are subpolynomial in  $D$  and that a range of  $D$  is calculated for each  $m$  suggest that this algorithm is preferable.

In fact, one might suppose that one could always use  $m = 1$  and search until an appropriate  $D$  is found without ever going on to the next  $m$ . However, one other factor needs to be taken into account. The size of the numerator of  $\zeta_D(1 - 2m)$  is  $O(m(\lg m + \lg D))$  bits (see [12]), so  $\zeta_D(1 - 2m)$  is much more likely to have large prime factors for large  $m$  than for small  $m$ . The same issue comes up for  $D$ , of course, but to a lesser degree and in a way which does not greatly affect this algorithm, since a large number of values of  $D$  are used for each  $m$ .

To make sure that we can avoid getting stuck in a range where the values of the numerator of  $\zeta_D(1 - 2m)$  are too small, we will give our algorithm parameters  $M_1$ ,  $D_1$ , and  $D_2$  such that we always have  $m \geq M_1$  and  $D_1 \leq D \leq D_2$ . In Section 3 we will explain some conjectures which imply that for each pair  $(D, m)$ , the chance that a given prime between  $P$  and  $cP$  divides  $\zeta_D(1 - 2m)$  is approximately  $2/P$ . Given this, the Prime Number Theorem implies that the chance that some prime between  $P$  and  $cP$  divides  $\zeta_D(1 - 2m)$  is approximately  $2(c - 1)/(\log P)$ .

TABLE 1. Time in minutes to find the first suitable pair  $(D, m)$  with given parameters

$c$	$P$	$D_2$	$(D, m)$	$p$	min.
1.01	$10^5$	$10^4$	(4156, 2)	100391	27
1.1	$10^5$	$10^4$	(697, 2)	106681	33
1.6	$10^5$	$10^4$	(205, 2)	113173	81
2	$10^5$	$10^4$	(184, 2)	164999	82
2	$10^6$	300	(40, 3)	1264807	169
2	$10^6$	$10^3$	(380, 2)	1017299	191
2	$10^6$	$10^4$	(380, 2)	1017299	191
2	$2 \cdot 10^6$	$10^4$	(317, 2)	2027569	569

Now if we are trying  $D_1 \leq D \leq D_2$  for each  $m$ , we see that the chance that we find a suitable  $D$  for any given  $m$  is approximately

$$\left(\frac{3}{\pi^2}\right) \frac{2(D_2 - D_1)(c - 1)}{\log P},$$

since the asymptotic density of fundamental discriminants in the integers is  $3/\pi^2$ . However, if we consider the typical case  $c = 2$ , then we see that we only have to choose  $D_2 - D_1$  to be on the order of magnitude of  $\log P$  for the expected probability of success on any given  $m$  to be 1! Thus with  $D_2 - D_1$  reasonably large, the expected time using this strategy is

$$O\left(\left(\frac{\log P}{c - 1}\right) L(D)^{O(1)} + P\right),$$

where the added term of  $P$  accounts for the time it takes to check whether each  $p$  divides each computed  $\zeta_D(1 - 2m)$ . Table 1 provides some actual timing examples of this algorithm, running on a Pentium III computer using the Linux operating system and the GP-Pari interpreted language (see [2]). In all cases  $D_1 = 5$  and  $M_1 = 2$ .

### 3. THE HYPOTHESES: CONJECTURES AND PREVIOUS RESULTS

The hypotheses mentioned in Section 2 stem from the conjecture, made (not very explicitly) by Siegel in [18], that the numerators of the Bernoulli numbers  $B_{2m}$  were uniformly distributed modulo  $p$  for any odd prime  $p$ . Siegel used the conjecture to derive a conjectural density for the irregular primes. (Lehmer seems to have done the same thing in [16] but only gives the density.) Siegel's hypothesis was used more generally by Johnson ([15]) and independently by Wooldridge ([21, Chap. III]) to predict the density of primes with a given index of irregularity, that is such that  $p$  divides a given number of the Bernoulli numbers  $B_2, \dots, B_{p-3}$ . It also comes in handy for predicting many other values that are related to irregular primes, such as the order of magnitude of the first prime of a given index of irregularity (see, for example, [19]). Since  $B_{2m} = -\zeta(1 - 2m)(2m)$ , it is equivalent to say that the values of  $\zeta(1 - 2m)$  are uniformly distributed modulo  $p$ , where  $\zeta(s)$  is the Riemann zeta function.

Little or no progress has been made on proving Siegel's hypothesis, but a great deal of data has been collected, especially in regard to the prediction of Johnson and Wooldridge. Specifically, this prediction says that as  $p \rightarrow \infty$ , the probability that  $p$  has index of irregularity  $r$  goes to

$$\left(\frac{1}{2}\right)^r \frac{e^{-1/2}}{r!}.$$

In other words, the index of irregularity follows the Poisson distribution with mean  $1/2$ . (In addition to the original sources, the details may be found in Section 5.3 of [20].) Note that this prediction does not rely on the full strength of Siegel's hypothesis, but merely on the weaker hypothesis that the Bernoulli numbers are 0 modulo  $p$  with probability  $1/p$ . The assumptions made in Section 2 relate only to predictions about indices of irregularity based on this weaker hypothesis.

Wagstaff, in [19], computed  $u_r(x)$ , the fraction of primes not exceeding  $x$  with index  $r$  of irregularity for each  $r$  between 0 and 2 and for all  $r \geq 3$  grouped together, and he compared this distribution to the predicted distribution for each multiple  $x$  of 1000 up to 125000. Wagstaff performed a chi-squared test for goodness of fit on this set of data, using the hypothesized Poisson distribution above. He reported that the value of the test statistic  $X_{Obs}^2$  "fluctuated usually between 0.1 and 1.0 and had the value 0.29 at  $x = 125000$ . It was 0.03 at  $x = 8000$ " [19]. These results correspond to  $p$ -values of .992, .801, .962, and .999, respectively.<sup>1</sup>

Buhler, Crandall, Ernvall, and Metsänkylä hold the record for computations with irregular primes, having found all the irregular primes below four million as described in [7]. They do not seem to have done a chi-squared analysis, but they tabulate the values of  $u_r(x)$  for  $x = 4000000$  and  $r$  between 0 and 7. A chi-squared test using the same classes as before ( $r = 0$ ,  $r = 1$ ,  $r = 2$ , and  $r \geq 3$ ) has  $X_{Obs}^2 = 1.02$ , for a  $p$ -value of .796. Earlier, in [8], Buhler, Crandall, and Sompolski tabulated the same data for  $x = 1000000$ . The same chi-squared test has  $X_{Obs}^2 = 0.78$ , for a  $p$ -value of .855.

Unfortunately, the only way to collect data to test Siegel's hypothesis is to investigate  $B_{2m}$  for larger and larger  $m$ , which is very computationally intensive (see [1] or [10] for details).

However, in the more general number field case, there are many more dimensions to the problem. We start by restricting our attention to the case of  $k$  an abelian totally real number field. Then we know that

$$\zeta_k(s) = \prod_{\chi \in \hat{G}} L(s, \chi),$$

where  $\hat{G}$  is the character group of  $G = \text{Gal}(k/\mathbf{Q})$  and  $L(s, \chi)$  is the  $L$ -function associated with the character  $\chi$ . Note that  $L(s, 1) = \zeta(s)$ , so the Riemann zeta function

---

<sup>1</sup> $X_{Obs}^2$  will be used here to denote the chi-squared test statistic computed from observed data. This should not be confused with  $\chi$ , which will be used later to denote a character. The  $p$ -value here is the probability that a chi-squared random variable with the appropriate number of degrees of freedom is at least as large as the observed value. It can be interpreted as the chance that a random sample taken from the predicted distribution would deviate from the distribution as a whole at least as much as the observed data did. Thus this set of data matches the prediction quite well. We will be using statistical language in this paper even though the data sets do not come from random variables and are in fact deterministic. Thus, all of the statistical results in this paper should be taken with a very large grain of salt.

is a factor of the zeta function for  $k$  (see [9], e.g., for more details). Certainly it seems likely that for a fixed (totally real) number field  $k$  and character  $\chi$  the values of the numerator of  $L(1 - 2m, \chi)$  are uniformly distributed modulo  $p$  as  $m$  varies. (It is known that these values are rational numbers.) We also hypothesize that these values for different  $\chi$  are independent, which implies that the numerators of  $\zeta_k(1 - 2m)$  are distributed modulo  $p$  like the product of  $|G|$  independent integer variables, each of which is uniformly distributed modulo  $p$ . We will refer to this as the “product distribution”, for lack of a better term. However, it also is reasonable to conjecture that for a fixed  $m$  the values of  $\zeta_k(1 - 2m)$  are distributed according to the product distribution modulo  $p$  as  $k$  varies. More precisely, if we fix  $m$  and the degree of  $k$ , we expect the values to be distributed according to the product distribution modulo  $p$  as the discriminant of  $k$  varies. Alternatively, if we fix  $m$  and the discriminant of  $k$ , we expect the values to be distributed according to the product distribution modulo  $p$  as the degree varies.

In this paper we will be considering the former situation. As in the previous sections, we fix the degree at 2, and we let  $k = \mathbf{Q}(\sqrt{D})$  be a real quadratic field with zeta function  $\zeta_D(s)$ . In this case

$$\zeta_D(s) = L(s, 1)L(s, \chi) = \zeta(s)L(s, \chi)$$

where  $\chi(s) = \left(\frac{D}{s}\right)$ , the Kronecker symbol, where appropriate.

In addition to the above definitions we will make one more set:

**Definition 2.** Let  $\chi$  be as above and let  $\delta$  be as in Section 1. We will say that  $p$  is  $\chi$ -regular if  $p$  is relatively prime to  $L(1 - 2m, \chi)$  for all integers  $m$  such that  $2 \leq 2m \leq \delta - 2$  and also  $p$  is relatively prime to  $pL(1 - 2m, \chi)$ . The number of such  $L$ -values that are divisible by  $p$  will be the *index of  $\chi$ -irregularity of  $p$* .

Saying that the values of  $\zeta_k(1 - 2m)$  are distributed according to the product distribution and that the values of  $\zeta(1 - 2m)$  are uniformly distributed is the same as saying that the values of  $L(1 - 2m, \chi)$  are uniformly distributed modulo  $p$ . Then we can make the same prediction about the indices of  $\chi$ -irregularity that Johnson and Wooldridge made about the indices of irregularity in the rational case. We briefly investigated this issue in [12], where there are tables of the analogue of  $u_r(x)$  (using the index of  $\chi$ -irregularity) for  $x = 1000$ ,  $r$  from 0 to 4, and  $D = 5, 8, 12$ , and 13. (We shall refer to this analogue as  $u_{\chi,r}(x)$ .) The values of  $X_{Obs}^2$  are not included, but using the same methodology as before (with classes  $r = 0$ ,  $r = 1$ ,  $r = 2$ , and  $r \geq 3$ ), they are 3.32, 1.74, 1.15, and 2.54. The corresponding  $p$ -values are .344, .627, .765, and .469, respectively.

Since we hypothesized that the values of  $L(1 - 2m, \chi)$  for different  $\chi$  are also independent, we can attempt to view each value of  $D$  as an independent set of “trials” of our “experiment”. In this case it would make sense to total the values of  $u_{\chi,r}(x)$  for the four values of  $D$  and compare them to the predicted values, that is, consider the observed and predicted values of  $\sum_D u_{\chi,r}(x)$  for each class of  $r$ . We might expect that this would give us a better  $p$ -value because of the larger “sample size”. However, in this case  $X_{Obs}^2 = 3.53$  and the  $p$ -value is .316, which is worse than any of the results for the values of  $D$  taken separately! This may be due to some small second-order bias which is common to each sample and thus is reinforced when they are pooled together. If this is true, then the “trials” cannot in fact be regarded as independent. We will consider this further in the next section.

## 4. THE HYPOTHESES: NEW RESULTS

In the course of testing the algorithms in [14], we collected more data in addition to that above. Table 2 shows the number of primes less than  $x$  which have  $\chi$ -index of irregularity  $r$ , for various values of  $x$  and  $r$  and  $D = 5$ . (Two of the primes listed under  $r \geq 3$  had index 4; none had index larger than 4.) We compared the observed and predicted distributions, using the methodology and classes above, for primes below  $x$  where  $x$  was 1000, 2000, 3000, 4000, and 5000, and we found  $X_{Obs}^2$  values of 3.32, 5.03, 2.51, 1.73, and 2.10 and  $p$ -values of .344, .170, .473, .630, and .552, respectively. These values are much less than those previously observed, but they are still well above the value of .050 commonly used by statisticians as a threshold for accepting the predicted model.<sup>2</sup>

TABLE 2. Results for  $D = 5$  and  $p < 5000$ 

$r$	$p < 1000$		$p < 2000$		$p < 3000$		$p < 4000$		$p < 5000$	
	obs.	pred.	obs.	pred.	obs.	pred.	obs.	pred.	obs.	pred.
0	112	101.29	202	183.17	274	260.20	341	332.99	422	405.16
1	43	50.65	78	91.59	123	130.10	159	166.49	186	202.58
2	11	12.66	18	22.90	28	32.53	44	41.62	51	50.65
$\geq 3$	1	2.40	4	4.35	4	6.17	5	7.90	9	9.61

Other data was obtained using the philosophy, described in Section 2, of computing the values of  $L(1 - 2m, \chi)$  for large numbers of  $D$  and relatively small values of  $m$ . As in the discussion of  $D = 5, 8, 12$ , and 13 above, we present the total across the different discriminants, under the assumption that values obtained for different  $D$  are independent. Table 3 presents the data for all  $D < 5000$  and  $p < 100$ . (Forty-one of the values had index 4, seven had index 5, none had index larger than 5.) The value of  $X_{Obs}^2$  for the totals is 81.1 and the  $p$ -value is .000. One would have to conclude that our predictions should be rejected for this set of data.

TABLE 3. Results for  $D < 5000$  and  $p < 100$ 

$r$	obs.	pred.
0	21864	22068.01
1	11596	11034.01
2	2529	2758.50
$\geq 3$	395	523.48

However, if we view the data broken down by prime, as in Table 4, we see that a large part of the contribution to  $X_{Obs}^2$  is from small primes. The values of  $p$  shown in the table were selected with an eye towards showing both a trend toward smaller chi-squared values as  $p$  increases and also some of the exceptions.

<sup>2</sup>It should be noted that the predicted numbers of primes are on the borderline of acceptability. The statistics textbook [17], for example, says: "There is no general agreement regarding the minimum value of expected frequencies, but values of 3, 4, and 5 are widely regarded as minimum. Some writers suggest that an expected frequency could be as small as 1 or 2, so long as most of them exceed 5" (p. 357). We have chosen to leave the classes as they stand in order to make them parallel to those of Wagstaff mentioned above.

TABLE 4. Results for  $D < 5000$ ; selected values of  $p$ 

$r$	0	1	2	$\geq 3$	$p$ -value
pred.	919.50	459.75	114.94	21.81	
$p = 3$	876	640	0	0	.000
$p = 5$	956	500	60	0	.000
$p = 7$	895	530	89	2	.000
$p = 11$	876	497	131	12	.008
$p = 13$	947	467	91	11	.010
$p = 17$	950	452	95	19	.175
$p = 23$	933	462	106	15	.387
$p = 37$	913	468	108	27	.605
$p = 47$	911	476	109	20	.775
$p = 67$	915	466	114	21	.986
$p = 79$	859	487	144	26	.003
$p = 97$	909	468	122	17	.623

It is not clear whether it makes sense to apply our hypotheses for small primes. In the case of Bernoulli numbers originally considered by Siegel, there is only one data point for each prime so it is impossible even to collect enough data while restricting ourselves to small primes. In Table 4, however, we have gathered many “trials” for each  $p$ . There are at least two possible reasons why this might be problematic for small primes. First, it is not clear whether Siegel’s hypothesis should apply at all. On the other hand, the previous section gave evidence that different values of  $D$  may not produce independent results. In that case, it may not be reasonable to analyze the data this way regardless of the size of the prime. We hope that more data will make the situation clearer. In particular, it should be possible to statistically test our assumptions about independence.

## 5. PRACTICAL NOTES

The hypotheses that  $p$  divides  $\zeta(1 - 2m)$  with probability  $1/p$  and that  $p$  divides  $L(1 - 2m, \chi)$  with the same probability clearly imply that  $p$  divides  $\zeta_D(1 - 2m) = \zeta(1 - 2m)L(1 - 2m, \chi)$  with probability  $(2/p) - (1/p^2)$ , or approximately  $2/p$  for very large  $p$ , as we claimed in Section 2. Also, for the algorithms in that section one does not really have to worry about the possibility that  $\delta$  is not  $p - 1$ , since this would require  $D = p$ . However, we showed that  $D_2 - D_1$  can be on the order of magnitude of  $\log p$ , so the case of  $D = p$  can only arise as the result of what can only be called bad planning.

As mentioned in the introduction, one use for the algorithms of this paper may be to find  $D$  and  $p$  such that the class group of  $\mathbf{Q}(\sqrt{D}, \zeta_p)$  can be used for cryptographic protocols. In [5], Buchmann and Paulus introduced a one way function based on class groups of number fields and noted that such a function could be used to implement Diffie-Hellman key exchanges and ElGamal signature schemes, to take two examples. These ideas are expanded on in [3], which introduces a signature scheme called RDSA which is based on taking  $p$ -th roots in the class group of a number field or other abelian group. Here  $p$  is a random prime number which (one assumes) does not divide the order of the group. One advantage of this signature scheme is that it is unnecessary and in fact undesirable to know precisely the order

of the group, a situation which frequently occurs with class groups and in fact is generally true for the class groups found with the algorithms above.

Given that the order of the class group is unknown, the question of which class groups are suitable for these protocols is addressed in [11], which gives two necessary conditions on the class number:

- The class number must be sufficiently large. This should make it difficult to determine the class number or discrete logarithms using exhaustive search, Pollard Rho, Baby-Step-Giant-Step, Hafner-McCurley, or index calculus.
- The class number must have at least one sufficiently large prime divisor. This should make it difficult to find discrete logarithms using a Pohlig-Hellman attack.

As we have seen, the algorithms of this paper allow us to find a class number with a prime divisor as large as desired, and thus with the class number itself as large as desired.

The drawback is that the amount of time and space needed to carry out the cryptographic protocols in these groups can also be very large. The papers [5] and [11] explain how to represent the objects necessary to compute with. Since elements in the class group are equivalence classes of ideals in a ring of integers, we need to store a  $\mathbf{Z}$ -basis for the ring of integers of  $\mathbf{Q}(\sqrt{D}, \zeta_p)$ . As noted in [5] and explained in more detail in [6], this requires  $(\log |\Delta|)^{O(1)}$  bits of storage, where  $\Delta$  is the discriminant of  $\mathbf{Q}(\sqrt{D}, \zeta_p)$ . Unfortunately, it is not hard to show that  $\Delta = D^{p-1}p^{2p-4}$  if  $p$  does not divide  $D$  or  $D^{p-1}p^{p-3}$  if  $p$  does divide  $D$  (but  $p \neq D$ ) (see [20], for instance). Thus the  $\mathbf{Z}$ -basis requires  $(p \log D)^{O(1)}$  bits of storage. Furthermore, as explained in [11], an ideal class should be represented by a member of the class which is LLL-reduced, that is, by one which corresponds to an LLL-reduced lattice under Minkowski's embedding. Such a representation requires  $(n + \log |\Delta|)^{O(1)}$  bits of storage, where  $n$  is the degree of the field and  $\Delta$  is as before (see [6] and [4] for details). In our case  $n = p - 1$  (assuming  $p \neq D$ ) so this again requires  $(p \log D)^{O(1)}$  bits of storage. Of course, this means that the time it takes to carry out the basic algorithms for the class group is also generally going to be exponential in the size of  $p$ . Whether this situation is bad enough to preclude the use of our fields is not yet clear.

## 6. CONCLUSION AND FUTURE WORK

Much of the future work described in [14] still remains to be done; in particular many improvements could be made in the implementations of the algorithms and perhaps in the algorithms themselves. However, the results already seem encouraging. With a faster implementation, the use of the search algorithm of Section 2 to find class groups large enough for secure cryptography seems quite feasible, although this should be tested in practice. More importantly, an implementation of one or more cryptographic protocols needs to be done using the class groups we have described in order to determine whether secure cryptography can be done sufficiently quickly in these groups.

The data collected in Section 4 is also encouraging, but clearly more is necessary. The author hopes to implement and run his algorithms on a true supercomputer in the near future. The data produced by this will undoubtedly give a clearer picture of the phenomena so far observed, perhaps leading to refinements of our hypotheses.

## ACKNOWLEDGMENTS

The author would like to thank Don Burdick of the Institute of Statistics and Decision Sciences at Duke University for his help in making sense of the data sets presented in Sections 3 and 4. He would also like to thank Carl Pomerance for suggesting that these data sets were worthy of being presented to a statistician in the first place and the anonymous referee for help with statistical terminology and other helpful suggestions.

## REFERENCES

1. Eric Bach, *The complexity of number-theoretic constants*, Inform. Process. Lett. **62** (1997), 145–152. MR **98g**:11148
2. C. Batut, K. Belabas, D. Bernardi, H. Cohen, and M. Olivier, *User's guide to PARI-GP*, Laboratoire A2X, Université Bordeaux I, version 2.0.9 ed., May 13, 1998, <<http://www.parigp-home.de>>, <<ftp://megrez.math.u-bordeaux.fr>>.
3. Ingrid Biehl, Johannes Buchmann, Safuat Hamdy, and Andreas Meyer, *A signature scheme based on the intractability of extracting roots*, Tech. Report Technical Report No. TI-1/00, Darmstadt University of Technology, 2000, <<http://www.informatik.tu-darmstadt.de/TI/Mitarbeiter/amy/Welcome.html>>.
4. Johannes Buchmann and Volker Kessler, *Computing a reduced lattice basis from a generating system*, <<http://www.informatik.tu-darmstadt.de/TI/Veroeffentlichung/reports/>>, 1992.
5. Johannes Buchmann and Sachar Paulus, *A one way function based on ideal arithmetic in number fields*, Advances in Cryptology—CRYPTO '97 (Burton S. Kaliski, Jr, ed.), Lecture Notes in Computer Science, vol. 1294, Springer-Verlag, 1997, pp. 385–394. MR **99a**:94041
6. Johannes Buchmann and Oliver van Sprang, *On short representations of orders and number fields*, <<http://www.informatik.tu-darmstadt.de/TI/Veroeffentlichung/reports/>>, 1992.
7. J. Buhler, R. Crandall, R. Ernvall, and T. Metsänkylä, *Irregular primes and cyclotomic invariants up to four million*, Math. Comp. **59** (1992), 717–722. MR **93a**:11106
8. J. P. Buhler, R. E. Crandall, and R. W. Sompolski, *Irregular primes to one million*, Math. Comp. **59** (1992), 717–722. MR **93a**:11106
9. J. W. S. Cassels and A. Fröhlich (eds.), *Algebraic number theory*, Academic Press, 1986, Reprint of the 1967 original. MR **88h**:11073
10. Sandra Fillebrown, *Faster computation of Bernoulli numbers*, J. Algorithms **13** (1992), 431–445. MR **94d**:68044
11. Tobias Hahn, Andreas Meyer, Stefan Neis, and Thomas Pfahler, *Implementing cryptographic protocols based on algebraic number fields*, Tech. Report Technical Report No. TI-24/99, Darmstadt University of Technology, 1999, <<http://www.informatik.tu-darmstadt.de/TI/Mitarbeiter/amy/Welcome.html>>.
12. Joshua Holden, *Irregularity of prime numbers over real quadratic fields*, Algorithmic Number Theory: Third International Symposium; Proceedings (J. P. Buhler, ed.), Springer Lecture Notes in Computer Science, vol. 1423, Springer-Verlag, 1998, pp. 454–462. MR **2000m**:11113
13. ———, *On the Fontaine-Mazur Conjecture for number fields and an analogue for function fields*, J. Number Theory **81** (2000), 16–47. MR **2001e**:11111
14. ———, *Comparison of algorithms to calculate quadratic irregularity of prime numbers*, Math. Comp. **71** (2002), 863–871. MR **2003d**:11183
15. Wells Johnson, *Irregular primes and cyclotomic invariants*, Math. Comp. **29** (1975), 113–120. MR **51**:12781
16. D. H. Lehmer, *Automation and pure mathematics*, Applications of Digital Computers (Walter F. Freiberger and William Prager, eds.), Ginn and Company, Boston, 1963, pp. 219–231. MR **27**:2136
17. Douglas C. Montgomery and George C. Runger, *Applied statistics and probability for engineers*, second ed., John Wiley & Sons, Inc., 1999.
18. Carl Ludwig Siegel, *Zu zwei bemerkungen Kummers*, Nachr. Akad. Wiss. Göttingen Math.-Phys. Kl. II **6** (1964), 51–57. MR **29**:1198

19. Samuel S. Wagstaff, Jr., *The irregular primes to 125000*, Math. Comp. **32** (1978), 583–591. MR **58**:10711
20. Lawrence C. Washington, *Introduction to cyclotomic fields*, second ed., Graduate Texts in Mathematics, vol. 83, Springer-Verlag, 1997. MR **97h**:11130
21. K. Wooldridge, *Some results in arithmetical functions similar to Euler's phi-function*, Ph.D. thesis, University of Illinois at Urbana-Champaign, 1975.

DEPARTMENT OF MATHEMATICS, DUKE UNIVERSITY, DURHAM, NORTH CAROLINA 27708

*Current address:* Department of Mathematics, Rose-Hulman Institute of Technology, Terre Haute, Indiana 47803

*E-mail address:* [holden@rose-hulman.edu](mailto:holden@rose-hulman.edu)

*URL:* <http://www.rose-hulman.edu/~holden>