MATHEMATICS OF COMPUTATION
Volume 00, Number 0, Pages 000–000
S 0025-5718(XX)0000-0

SQUARE FORM FACTORIZATION

JASON E. GOWER AND SAMUEL S. WAGSTAFF, JR.

This paper is dedicated to the memory of Daniel Shanks.

Abstract. We present a detailed analysis of SQUFOF, Daniel Shanks' Square Form Factorization algorithm. We give the average time and space requirements for SQUFOF. We analyze the effect of multipliers, either used for a single factorization or when racing the algorithm in parallel.

1. Introduction

SQUFOF, or SQUare FOrm Factorization, is an integer factoring algorithm invented by Daniel Shanks more than thirty years ago.

For each size of integer, there is a fastest general purpose algorithm (among known methods) to factor a number of that size. At present, the number field sieve (NFS) is best for integers greater than about 10^{120} and the quadratic sieve (QS) is best for numbers between 10^{50} and 10^{120}, etc. As new algorithms are discovered, these ranges change. On a 32-bit computer, SQUFOF is the clear champion factoring algorithm for numbers between 10^{10} and 10^{18}, and will likely remain so. It can split almost any composite 18-digit integer in less than a millisecond. The SQUFOF algorithm is extraordinarily simple, beautiful and efficient. Further, it is used in many implementations of NFS and QS to factor small auxiliary numbers arising when factoring a large integer.

Although Shanks [16], [18] described other new algorithms for factoring integers, he published nothing about SQUFOF. He did lecture [12] on SQUFOF and he explained its operation to a few people. Some works of others, such as [3], [11], [2], and [19], discuss the algorithm, but none contains a detailed analysis. After Shanks died in 1996, H. C. Williams discovered some of Shanks' unpublished hand-written manuscripts [15], [14], [13], and eventually they appeared on the web [6].

The manuscript [15] is the closest Shanks ever came to a full description and analysis of SQUFOF. In [15], Shanks described the algorithm and began a heuristic argument for the following statement. Let N be a product of k distinct odd primes with $N \equiv 3 \bmod 4$. Then the average number of forms that SQUFOF must examine before finding a proper square form (one leading to a non-trivial factor of N) is

$$\frac{3\left(\sqrt{2}+2\right)\log 2}{2\left(2^k-2\right)}\sqrt[4]{N}\,.$$

1991 *Mathematics Subject Classification.* 11A11, 11E16, 11R11, 11Y05.

Key words and phrases. Integer factorization, binary quadratic form.

This paper is based on the Ph.D. thesis [5] of the first author, supervised by the second author. Both authors are grateful for the support of the CERIAS Center at Purdue University and by the Lilly Endowment Inc.

The manuscript also contains a discussion of how to decide whether a square form is proper or not, but there is no proof for why this decision is always correct. Shanks also discusses the use of multipliers as a way to overcome a failure to factor N, and the possibility of racing multipliers. It is clear from [15] and from discussions Shanks had with the second author and others that Shanks knew a lot of the content of this paper and much more about SQUFOF.

We give a detailed description and analysis of SQUFOF, and determine its time and space complexity. In Theorem 4.22, we complete the heuristic argument started by Shanks in [15], derive the average number shown above, and extend the argument to the cases $N \equiv 1,\ 2 \bmod 4$. The only variable-length storage SQUFOF uses is a queue data structure. In Theorem 4.24, we estimate the average number of entries placed in this queue. Theorems 5.4 and 5.8 give the time and space complexity of SQUFOF when multipliers are used to factor N. We give a detailed description of the process for deciding which square forms are proper, show how to modify it when multipliers are used, and prove that it works in all cases. We study SQUFOF as if it were a random walk on the principal cycle of binary quadratic forms of discriminant N or $4N$. Our theorems about the complexity of SQUFOF are proved using reasonable and perhaps provable assumptions about this random walk.

In Section 2 we provide a minimum background for the sequel. We describe the algorithm in Section 3 and give some examples of it. Then in Section 4 we derive the average time and space requirements for the basic algorithm. Section 5 presents the time and space requirements for SQUFOF with multipliers. We give in Section 6 the results of some experiments, which provide evidence that our simplifying assumptions are reasonable. Finally, we conclude in Section 7 with some questions for future research.

We thank Arunkumar Navasivasakthivelsamy and Rupak Sanjel for writing some programs we used in the experiments. We are grateful to an anonymous referee for improving the clarity of the paper.

2. Background

2.1. **Binary Quadratic Forms.**

We begin with a brief survey of binary quadratic forms. For a more detailed account of the theory see [2] or [3].

2.1.1. *Basic Definitions.* Let $f(x, y) = ax^2 + bxy + cy^2$, a *binary quadratic form* in the variables x and y. The constants a, b, and c will be taken in $\mathbb{Z}$. The *discriminant* of f is defined to be $b^2 - 4ac$. A discriminant Δ is called *fundamental* if either Δ is odd and square-free; or Δ is even, $\Delta/4$ is square-free, and $\Delta/4 \equiv 2$ or $3 \bmod 4$. The form f is called *primitive* if $\gcd(a, b, c) = 1$.

We will frequently write $f = (a, b, c)$, or just $(a, b, *)$, where c can be computed if we know the discriminant of f. We shall also write $f = (a, *, *)$ whenever b and c are either unknown or irrelevant. Note that if Δ is the discriminant of the form f, then $\Delta \equiv 0$ or $1 \bmod 4$, and $b \equiv \Delta \bmod 2$.

The form f is said to *represent* $m \in \mathbb{Z}$ if there exists $x_0, y_0 \in \mathbb{Z}$ such that $f(x_0, y_0) = ax_0^2 + bx_0y_0 + cy_0^2 = m$. The representation is *primitive* if $\gcd(x_0, y_0) = 1$.

We say that two forms f_1 and f_2 are *properly equivalent*, or just *equivalent*, if we can find $\alpha, \beta, \gamma, \delta \in \mathbb{Z}$ such that $\alpha\delta - \beta\gamma = 1$ and $f_1(x, y) = f_2(\alpha x + \beta y, \gamma x + \delta y)$. We write $f_1 \sim f_2$ when f_1 and f_2 are equivalent. If $\alpha\delta - \beta\gamma = -1$, then we say that f_1 and f_2 are *improperly equivalent.* Let $\Gamma = \mathrm{SL}_2(\mathbb{Z})$ be the classical modular group and define the action of Γ on the set of binary quadratic forms by

$$\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \cdot f(x,y) = f(\alpha x + \beta y, \gamma x + \delta y) \ .$$

Then $f_1 \sim f_2$ if and only if f_1 and f_2 are equivalent modulo the action of Γ. We make special note of the equivalence: $(a, b + 2na, a + nb + c) \sim (a, b, c)$ for any $n \in \mathbb{Z}$, using the matrix $\begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix}$.

The number of classes of forms of discriminant Δ will be written $h^+(\Delta)$ or just h^+. It can be shown that $h^+(\Delta)$ is finite.

Forms with negative discriminant are called *definite*, while forms with positive discriminant are called *indefinite.* We will be concerned only with indefinite forms.

Any form (k, kn, c) is called *ambiguous.* There exists an ambiguous form (k, kn, c) of discriminant Δ for each divisor k of Δ. We also refer to any form (a, b, a) as ambiguous since it is equivalent to $(b + 2a, b + 2a, a)$.

2.1.2. *Indefinite Forms.* Let Δ be any non-square positive integer. Each class of indefinite forms of discriminant Δ contains a set of canonical representatives, called *reduced* forms. The form $f = (a, b, c)$ is called *reduced* if $\left|\sqrt{\Delta} - 2|a|\right| < b < \sqrt{\Delta}$. It is not hard to see that f is reduced if and only if $\left|\sqrt{\Delta} - 2|c|\right| < b < \sqrt{\Delta}$, and that the number of reduced forms of a given discriminant is finite. For any indefinite form $f = (a, b, c)$ with $ac \neq 0$ we define the *standard reduction operator* by

$$\rho(a,b,c) = \left(c, r(-b,c), \frac{r(-b,c)^2 - \Delta}{4c} \right) , \tag{2.1}$$

where $r(-b, c)$ is defined to be the unique integer r such that $r + b \equiv 0 \bmod 2c$ and

$$\begin{aligned} -|c| < r \le |c| \quad &\text{if} \quad \sqrt{\Delta} < |c| \ , \\ \sqrt{\Delta} - 2|c| < r < \sqrt{\Delta} \quad &\text{if} \quad |c| < \sqrt{\Delta} \ . \end{aligned}$$

$\rho(f)$ is called the *reduction* of f and the result of n applications of ρ is written $\rho^n(f)$. It will be convenient to define the *inverse reduction operator* by

$$\rho^{-1}(a,b,c) = \left(\frac{r(-b,a)^2 - \Delta}{4a}, r(-b,a), a \right) ,$$

where $r(-b, a)$ is defined as in the definition of ρ. Note that if the discriminant of f is Δ, then the discriminants of both $\rho(f)$ and $\rho^{-1}(f)$ are Δ.

If f is reduced, then both $\rho(f)$ and $\rho^{-1}(f)$ are reduced. If f is not reduced, then $\rho^n(f)$ is reduced for some finite n. Similarly f can be reduced after a finite number of applications of ρ^{-1}. The identities $\rho(\rho^{-1}(f)) = \rho^{-1}(\rho(f)) = f$ hold only when f is reduced. The unique reduced form $(1, b, c)$ is called the *principal* form.

We say that (a, b, c) and (c, b', c') are *adjacent* if $b + b' \equiv 0 \bmod 2c$. More specifically, we say that (a, b, c) is adjacent to the left of (c, b', c') and (c, b', c') is adjacent to the right of (a, b, c). It is easy to see that there is a unique reduced form adjacent to the right and to the left of any given reduced form, these forms being $\rho(a, b, c)$ and $\rho^{-1}(a, b, c)$, respectively. We now see that within each equivalence class of forms of discriminant $\Delta > 0$ there are *cycles* of reduced forms. The cycle that contains the principal form is called the *principal cycle*. The number of reduced forms in any cycle is always even.

The two forms (a, b, c) and (c, b, a) are said to be *associated*. If the form f_1 and its associate f_2 are in different cycles, then this will be the case for all forms in either cycle, and in this case the two cycles are said to be *associated cycles*. Furthermore, any cycle which contains an ambiguous form (called an *ambiguous cycle*) contains exactly two ambiguous forms and is its own associate. Conversely, a cycle which is its own associate contains exactly two ambiguous forms. The principal cycle is ambiguous since it contains the principal form $(1, b, c)$.

The form $(a, -b, c)$ is the *opposite* of the form (a, b, c). A form (a, b, c) is improperly equivalent to both its associate and its opposite. Hence, (a, b, c) is properly equivalent to the associate of its opposite: $(a, b, c) \sim (c, -b, a)$. Likewise, the opposite and the inverse of (a, b, c) are properly equivalent: $(a, -b, c) \sim (c, b, a)$.

If (a, b, c) is a form of discriminant Δ which represents the integer r, then $s^2 \equiv \Delta \bmod 4r$ has a solution. Conversely, if a solution to $s^2 \equiv \Delta \bmod 4r$ exists, then r is represented by some form of discriminant Δ.

Let $\left(\frac{r}{s}\right)$ be the Jacobi symbol and define the quadratic characters $\chi(r) = \left(\frac{-1}{r}\right)$ and $\psi(r) = \left(\frac{2}{r}\right)$. The *generic characters* of a discriminant Δ are

$$\begin{aligned} \left(\frac{r}{p}\right) &\quad \text{for all odd primes } p \text{ that divide } \Delta\ , \\ \chi(r) &\quad \text{if } \Delta \text{ is even and } \Delta/4 \equiv 3, 4, 7 \bmod 8\ , \\ \psi(r) &\quad \text{if } \Delta \text{ is even and } \Delta/4 \equiv 2 \bmod 8\ , \\ \chi(r)\cdot\psi(r) &\quad \text{if } \Delta \text{ is even and } \Delta/4 \equiv 6 \bmod 8\ , \\ \chi(r) \text{ and } \psi(r) &\quad \text{if } \Delta \text{ is even and } \Delta/4 \equiv 0 \bmod 8\ . \end{aligned}$$

These characters are multiplicative functions from $\mathbb{Z}$ to $\{\pm 1\}$. Suppose the discriminant Δ has n generic characters. Then for some arbitrary ordering we have a vector-valued function from $\mathbb{Z}$ to the n-tuples with ± 1 entries. The n-tuple corresponding to an integer r is called the *assigned value* of r. It can be shown that all integers r which are representable by forms of a given equivalence class possess the same assigned values of generic characters. The set of classes of forms possessing the same assigned values of generic characters is called a *genus* of forms. The genus for which the assigned value is $(1, 1, \ldots, 1)$ is called the *principal genus*. The principal genus contains the principal form. An integer r is representable by some class of forms of discriminant Δ if and only if the assigned values of the generic characters of r match the assigned values of characters of some genus of discriminant Δ. This is true if and only if the congruence $s^2 \equiv \Delta \bmod 4r$ is solvable.

The number of ambiguous classes (including the principal class) is equal to one-half the number of possible genera. If Δ is a fundamental discriminant, then we know that the product of the assigned values for the characters for any genus is $+1$ and that exactly half of the possible genera exist.

2.1.3. *Composition of Forms.* We now define *composition* of forms. Let $f_1 = (a_1, b_1, c_1)$ and $f_2 = (a_2, b_2, c_2)$ be two forms with the same discriminant. Let $\beta = (b_1 + b_2)/2$, $m = \gcd(a_1, \beta)$, and $n = \gcd(m, a_2)$. Solve $a_1 x + \beta y = m$ for x and y and

$$mz/n \equiv x\left(\frac{b_2 - b_1}{2}\right) - c_1 y \bmod a_2/n \quad \text{for } z\ .$$

Then the composition of f_1 and f_2, written $f_1 \circ f_2$ is

$$\left(a_1 a_2/n^2, b_1 + 2a_1 z/n, *\right) \ ,$$

where the third coefficient may be determined by the discriminant formula. Although the composition is not unique, all compositions of given forms f_1 and f_2 are equivalent. The class of $f_1 \circ f_2$ depends only on the classes of f_1 and f_2, and the classes form a group under composition. We note that even if f_1 and f_2 are reduced, their composition need not be reduced.

As a special case, we present the formula for the square $f^2 = f \circ f$ as follows. Suppose $f = (a, b, c)$, $n = \gcd(a, b)$, and y is a solution for $by/n \equiv 1 \bmod a/n$. Then f^2 is equivalent to

$$\left(a^2/n^2, b - 2acy/n, *\right) \ .$$

Note that if $\gcd(a, b) = 1$, then

$$(a, b, -ac)^2 \sim \left(a^2, b, -c\right) \ .$$

Moreover, g is equivalent to an ambiguous form if and only if $g \circ g$ is equivalent to the principal form. This implies that the square of $g \circ (a, b, -ac)$ is equivalent to $\left(a^2, b, -c\right)$.

Also note that if f is a square form on the principal cycle, then f must have a square root on the principal cycle. To see this, let $f^{1/2}$ be any square root of f. If neither $f^{1/2}$ nor $\rho^n\left(f^{1/2}\right)$ for all $n > 0$ is on the principal cycle, then $f^{1/2}$ must be equivalent to some ambiguous form other than the principal form, say g. Then $f^{1/2} \circ g$ is equivalent to the principal form, and its square is equivalent to f. Finally, we can reduce this form to an equivalent form on the principal cycle.

Observe that

$$(1, b_1, c_1) \circ (a_2, b_2, c_2) \sim (a_2, b_2, c_2) \ ,$$

and that

$$(a, b, c) \circ (a, -b, c) \sim (a, b, c) \circ (c, b, a) \sim (ac, b, 1) \ .$$

In other words, under composition, the principal class is the identity and the associate is the inverse. Also composition is commutative and associative. Thus the set of equivalence classes of forms of a given discriminant is an abelian group under composition.

2.2. **Periodic Continued Fractions.** Let $N > 0$ be a positive integer, not a square. The *simple continued fraction expansion* of $\sqrt{N}$ is given by

$$\sqrt{N} = q_0 + \cfrac{1}{q_1 + \cfrac{1}{q_2 + \cdots}} \ .$$

We will always abbreviate the expansion as $[q_0, q_1, \dots]$. The expansion is periodic beginning with q_1, meaning that for some $j > 0$ we will have $a_i = a_{i+j}$ for all $i > 0$, where j is the period of the continued fraction. In this case, we will write $\sqrt{N} = [q_0, \overline{q_1, \dots, q_j}]$.

The q_i are called the *partial quotients* of the continued fraction. The rational number $[q_0, q_1, \dots, q_n]$ is called the n^{th} *convergent* of the continued fraction. Define

$$A_n = \begin{cases} 1 & \text{if } n = 0 \ , \\ q_0 & \text{if } n = 1 \ , \\ q_n A_{n-1} + A_{n-2} & \text{if } n \geq 2 \ , \end{cases}$$

and

$$B_n = \begin{cases} 0 & \text{if } n = 0 \ , \\ 1 & \text{if } n = 1 \ , \\ q_n B_{n-1} + B_{n-2} & \text{if } n \geq 2 \ . \end{cases}$$

Then $[q_0, q_1, \ldots, q_n] = A_{n+1}/B_{n+1}$ for $n \geq 0$.

We define the n^{th} *complete quotient* by

$$x_n = \begin{cases} \sqrt{N} & \text{if } n = 0 \ , \\ 1/\left(x_{n-1} - q_{n-1}\right) & \text{if } n \geq 1 \ . \end{cases}$$

It can be shown that $x_n = (P_n + \sqrt{N})/Q_n$ for $n \geq 0$, where

$$P_n = \begin{cases} 0 & \text{if } n = 0 \ , \\ q_0 & \text{if } n = 1 \ , \\ q_{n-1}Q_{n-1} - P_{n-1} & \text{if } n \geq 2 \ , \end{cases} \tag{2.2}$$

and

$$Q_n = \begin{cases} 1 & \text{if } n = 0 \ , \\ N - q_0^2 & \text{if } n = 1 \ , \\ Q_{n-2} + (P_{n-1} - P_n)q_{n-1} & \text{if } n \geq 1 \ . \end{cases} \tag{2.3}$$

The q_n can be computed using

$$q_n = \begin{cases} \left\lfloor \sqrt{N} \right\rfloor & \text{if } n = 0 \ , \\ \\ \left\lfloor \dfrac{q_0 + P_n}{Q_n} \right\rfloor & \text{if } n > 0 \ . \end{cases} \tag{2.4}$$

Some important facts that we shall need are as follows.

$$(-1)^n Q_n = A_n^2 - B_n^2 N \ ,$$

$$\frac{A_n + B_n\sqrt{N}}{\sqrt{Q_n}} = \frac{A_{n-1} + B_{n-1}\sqrt{N}}{\sqrt{Q_{n-1}}} \cdot \frac{\sqrt{N} + P_n}{\sqrt{Q_{n-1}Q_n}} \ ,$$

$$N = P_n^2 + Q_n Q_{n-1} \ ,$$

$$0 \leq P_n, \, Q_n < 2\sqrt{N} \ .$$

See [11] for a proof of these facts. The first integer factoring algorithm with subexponential time complexity was based on continued fractions. See [1] for details. Shanks discovered SQUFOF [12], [15] while investigating the "failures" of the continued fraction factoring algorithm.

2.3. **Real Quadratic Number Fields.** Let $N \neq 1$ be a square-free integer, and define

$$\Delta = \begin{cases} 4N & \text{if } N \equiv 2, 3 \bmod 4 \ , \\ N & \text{if } N \equiv 1 \bmod 4 \ . \end{cases}$$

Any finite extension of $\mathbb{Q}$ is called a *number field*. The extension $\mathbb{Q}(\sqrt{N})/\mathbb{Q}$ is called the *quadratic number field* of radicand N and discriminant Δ. We note in passing that $\mathbb{Q}(\sqrt{N}) = \mathbb{Q}(\sqrt{\Delta})$.

Let K be any number field. The ring of integers $\mathcal{O}_K$ of K is the integral closure of $\mathbb{Z}$ in K. When $K = \mathbb{Q}(\sqrt{N})$ we have $\mathcal{O}_K = \mathbb{Z}[\omega]$, where

$$\omega = \begin{cases} \sqrt{N} & \text{if } N \equiv 2, 3 \bmod 4 \ , \\ \dfrac{1+\sqrt{N}}{2} & \text{if } N \equiv 1 \bmod 4 \ . \end{cases}$$

The odd rational primes p fall into three categories according to the value of the Legendre symbol

$$\left(\frac{\Delta}{p}\right) = \begin{cases} 0 & \text{if } p \text{ is a \textit{ramified} prime ,} \\ 1 & \text{if } p \text{ is a \textit{split} prime ,} \\ -1 & \text{if } p \text{ is a \textit{inert} prime .} \end{cases}$$

When $N \equiv 1 \bmod 4$, the rational prime 2 is split whenever $N \equiv 1 \bmod 8$, and inert whenever $N \equiv 5 \bmod 8$. The ramified primes are precisely those that divide Δ. A consequence of the Chebotarev density theorem (see [9]) is that the density of primes that split in $\mathbb{Q}(\sqrt{\Delta})$ is $1/2$. Since there are only finitely many ramified primes, it follows that the density of inert primes is also $1/2$.

A *fractional ideal* is a subset $\mathfrak{a}$ of $\mathbb{Q}(\sqrt{\Delta})$ such that

(1) for any $\alpha, \beta \in \mathfrak{a}$ and any $\lambda, \mu \in \mathbb{Z}[\omega]$ we have $\lambda\alpha + \mu\beta \in \mathfrak{a}$.
(2) there exist a fixed $\nu \in \mathbb{Z}[\omega]$ such that for every $\alpha \in \mathfrak{a}$ we have $\nu\alpha \in \mathbb{Z}[\omega]$.

Two fractional ideals $\mathfrak{a}, \mathfrak{b}$ are *equivalent* if there is some $\alpha \in \mathbb{Z}[\omega]$ such that $\mathfrak{a} = (\alpha)\mathfrak{b}$, and *narrowly equivalent* if there is some $\alpha \in \mathbb{Z}[\omega]$ with positive norm such that $\mathfrak{a} = (\alpha)\mathfrak{b}$. Both types of equivalences are indeed equivalence relations. The first equivalence leads to the *class group* I/P, where I is the set of fractional ideals and P is the set of principal ideals. Narrow equivalence leads to the *narrow class group* I/P^+, where P^+ is the set of principal ideals with positive norm. The *class number* of $\mathbb{Q}(\sqrt{\Delta})$ is the order of I/P, while the *narrow class number* is the order of I/P^+, written $h(\Delta)$ and $h^+(\Delta)$, respectively. It is no coincidence that we use the same symbol to denote both the number of classes of forms of discriminant Δ and the narrow class number of $\mathbb{Q}(\sqrt{\Delta})$, as it can be shown (see [3]) that they are equal.

2.4. **The Infrastructure of the Class Group.** The theories of binary quadratic forms, continued fractions, and real quadratic number fields are closely related (see [4] or [3].) First, there is a correspondence between binary quadratic forms of discriminant $N > 0$ and the fractional ideals of $\mathbb{Q}(\sqrt{N})$ defined by

$$(a, b, c) \longleftrightarrow \left(a\mathbb{Z} + \left(\frac{-b+\sqrt{N}}{2} \right) \mathbb{Z} \right) \alpha \ ,$$

where α is any element of $\mathbb{Q}(\sqrt{N})^\times$ such that $N(\alpha) = \operatorname{sign}(a)$. Under this correspondence, composition of forms corresponds with ideal multiplication.

There is also a correspondence between binary quadratic forms and continued fractions. The definitions for P_n and Q_n in Section 2.2 satisfy

$$N = P_n^2 + Q_{n-1}Q_n \quad \text{for all } n \ ,$$

and so the binary quadratic form

$$F_n = \left((-1)^{n-1}Q_{n-1}, 2P_n, (-1)^n Q_n\right)$$

has discriminant $4N$. In fact, the sequence of forms $F_0, F_1, \ldots$ constitutes the principal cycle of forms of discriminant $4N$, where $F_0 = \left(1, 2q_0, q_0^2 - N\right)$.

Shanks defined the *infrastructure* of the class group [17] collectively as the inner structure within each cycle of reduced forms determined by ρ, the standard reduction operator. Originally, Shanks defined the infrastructure distance between the form F_n and the principal form by the equation

$$d_n = \log\left(A_n + B_n\sqrt{N}\right) ,$$

but this metric did not have all the desirable properties that one would like it to have, so he later [15] changed it to

$$d_n = \log\left(\frac{A_n + B_n\sqrt{N}}{\sqrt{Q_n}}\right) .$$

In [8], Lenstra independently proposed this same metric in a slightly different form as follows. Let $f = (a, b, c)$ be a form of discriminant Δ. Then

$$d\left(f, \rho(f)\right) = \frac{1}{2}\log\left|\frac{b+\sqrt{\Delta}}{b-\sqrt{\Delta}}\right| .$$

That these two definitions agree follows from the facts at the end of Section 2.2.

Now by the laws of Khinchin, Gauss-Kuzmin, and Lévy [7], we can approximate d_n by

$$\log\left(\frac{A_n + B_n\sqrt{N}}{\sqrt{Q_n}}\right) \approx \frac{\pi^2}{12\log 2}\, n , \tag{2.5}$$

where the constant $\pi^2/\left(12\log 2\right)$ is approximately 1.19.

More generally, one can define the infrastructure distance $d(f, g)$ between two reduced quadratic forms by the following. Let $\mathfrak{a}, \mathfrak{b}$ be the ideals corresponding to f, g respectively. If f and g are narrowly equivalent, then we can find γ with $N(\gamma) > 0$ such that $\mathfrak{a} = \gamma\mathfrak{b}$. Define the infrastructure distance between f and g by

$$d(f, g) = \frac{1}{2}\log\left|\frac{\gamma}{\sigma(\gamma)}\right| ,$$

where σ is the automorphism of $\mathbb{Q}(\sqrt{N})$ taking $\sqrt{N}$ to $-\sqrt{N}$. With this definition, if f is reduced, one can show that the distance between the reduction of f^2 and the principal form is twice the distance between f and the principal form. To see this, let $\mathfrak{a}$ be the fractional ideal corresponding to f and let 1 denote the principal form. Writing $\mathfrak{a} = \gamma \cdot 1$, we have $\mathfrak{a}^2 = \gamma^2 \cdot 1$, and

$$d(f^2, 1) = \frac{1}{2}\log\left|\frac{\gamma^2}{\sigma(\gamma^2)}\right| = \frac{1}{2}\log\left|\frac{\gamma}{\sigma(\gamma)}\right|^2 = 2\, d(f, 1) .$$

More generally, let $\mathfrak{b}_1 = \gamma_1\mathfrak{a}_1$ and $\mathfrak{b}_2 = \gamma_2\mathfrak{a}_2$, so that $\mathfrak{b}_1\mathfrak{b}_2 = \gamma_1\gamma_2\mathfrak{a}_1\mathfrak{a}_2$. If reduced forms f_i, g_i correspond to $\mathfrak{a}_i, \mathfrak{b}_i$, respectively, then we have

$$d\left(g_1 \circ g_2, f_1 \circ f_2\right) = d\left(g_1, f_1\right) + d\left(g_2, f_2\right) . \tag{2.6}$$

Note that the forms $g_1 \circ g_2$ and $f_1 \circ f_2$ need not be reduced. See Proposition 5.8.4 in Cohen [3] for the correction needed when they are not reduced.

Now suppose F_n is a square form on the principal cycle. Then we know that a square root of F_n must also lie on the principal cycle at a distance $d_n/2$ from the principal form. But using the approximation (2.5), this form will be very close to $F_{n/2}$. Likewise, the square roots of F_n in other cycles are a distance $d_n/2$, all in the same direction, from an ambiguous form. Also note that Equation (2.6) can be used to show that an inverse square root of F_n is at a distance $d_n/2$ in the reverse direction.

3. The Description of the Algorithm

We now describe the algorithm in detail. We begin with a description of the fastest and most practical version. A more general version follows.

3.1. **Continued Fractions Description.** In the analysis that follows, we will assume that N is a square-free positive integer. Our experience in factoring millions of integers with SQUFOF suggests that the algorithm works equally well when N is not square-free, but we don't know how to extend our analysis to that case.

In most implementations of SQUFOF, we work with binary quadratic forms of discriminant $\Delta = 4N$. Unfortunately, if we do this when $N \equiv 1 \bmod 4$, then Δ is not a fundamental discriminant. Although the algorithm works for non-fundamental discriminants, the analysis of SQUFOF presented below will assume that Δ is fundamental. Therefore, if $N \equiv 1 \bmod 4$ then we replace N with $2N$. We may now assume that $N \equiv 2$ or $3 \bmod 4$ for the remainder of this subsection. Finally, take $\Delta = 4N$ which is then always a fundamental discriminant.

The principal form is $F_0 = (1, 2q_0, q_0^2 - N)$. We compute the forms on the principal cycle by

$$F_n = \rho^n(F_0) = \left((-1)^{n-1}Q_{n-1}, 2P_n, (-1)^n Q_n\right) ,$$

where P_n and Q_n are calculated according to (2.2), (2.3), and (2.4). We seek a *square form* $(*, *, c^2)$, which can only occur when n is even. Suppose we have found a square form $F_n = (-Q, 2P, S^2)$, where $Q > 0$. Define $F^{-1/2} = (-S, 2P, SQ)$, an inverse square root of F_n under composition of forms. This form may not be reduced so let $G_0 = (-S_{-1}, 2R_0, S_0)$ be its reduction, where

$$R_0 = P + S\left\lfloor \frac{q_0 - P}{S} \right\rfloor , \qquad S_{-1} = S, \qquad S_0 = \frac{N - R_0^2}{S} .$$

Using $R_m = t_{m-1}S_{m-1} - R_{m-1}$, $S_m = S_{m-2} + t_{m-1}(R_{m-1} - R_m)$, and $t_m = \left\lfloor \frac{q_0 + R_m}{S_m} \right\rfloor$, for $m \geq 1$, which are completely analogous to (2.2), (2.3), and (2.4), we generate a (hopefully) new sequence of forms

$$G_m = \left((-1)^{m-1}S_{m-1}, 2R_m, (-1)^m S_m\right) .$$

Now suppose we find m such that $R_m = R_{m+1}$. We expect this to happen at approximately $m \approx n/2$ for reasons explained at the end of Section 2.4. At this m we will have $R_m = t_m S_m/2$ and $N = R_m^2 + S_{m-1}S_m$, which gives

$$N = S_m\left(S_{m-1} + S_m \frac{t_m^2}{4}\right) ,$$

a possible factorization of N. We call the square form F_n *improper* if this factorization is trivial. If a non-trivial factor of N is found, then F_n is a *proper* square form.

One should note that all computations other than those of F_0 and G_0 are with numbers less than $2\sqrt{N}$ in magnitude. So if N is taken to be no larger than double the word size of the computer, then all computations (except F_0 and G_0) will be with single precision integers.

Three main issues arise at this point. First, we will need to test every other form for squareness. This is not a major obstacle since there are fast algorithms to test for squareness. A more serious issue is the possibility of finding only a trivial factorization. We could return to the last square form F_n, but this is time consuming. Instead, we will keep track of certain forms and use them in a test for proper square forms. Finally, there may be no proper square forms at all on the principal cycle. If this is so, then we can try SQUFOF on mN, for some small m. We shall see later that this is reasonable.

3.2. **Identifying Proper Square Forms.**

We begin with a few facts about square roots of square forms.

Proposition 3.1. *Suppose that a is a positive odd integer, b is a positive integer, $\gcd(a,b) = 1$, and that $\left(a^2, 2b, -c\right)$ is a square form on the principal cycle of discriminant $4N$ with $c > 0$. Then $(-a, 2b, ac)^2 \sim \left(a^2, 2b, -c\right)$.*

Proof. This follows directly from the definition of composition of forms, noting that $\left(a^2, 2b, -c\right)$ is equivalent to any form $\left(a^2, 2\beta, *\right)$, where $\beta \equiv b \bmod a^2$. □

Let $b = \left\lfloor \sqrt{N} \right\rfloor$ and $\mathbf{1} = (1, 2b, c)$ denote the principal form. Let $b' = \left\lfloor \sqrt{N} \right\rfloor$ or $\left\lfloor \sqrt{N} \right\rfloor - 1$, whichever is odd, and let $\mathbf{2}$ denote the reduced ambiguous form $(2, 2b', c')$. By $-\mathbf{1}$, $-\mathbf{2}$ we mean the forms $(-1, 2b, -c)$, $(-2, 2b', -c')$, respectively. It is easy to see that $\pm\mathbf{1} \circ (\alpha, 2\beta, \gamma) \sim (\pm\alpha, 2\beta, \pm\gamma)$ and that $\pm\mathbf{2} \circ (\alpha, 2\beta, *) \sim (\pm 2\alpha, 2\beta, *)$, when α is odd and $\pm\mathbf{2} \circ (\alpha, 2\beta, *) \sim (\pm\alpha/2, 2\beta, *)$, when α is even.

Proposition 3.2. *Suppose that a is a positive odd integer, b is a positive integer, $\gcd(a,b) = 1$, and that $F_n = \left(a^2, 2b, -c\right)$ is a square form on the principal cycle of discriminant $4N$, with $c > 0$. Some form $(\alpha, 2\beta, *)$ appears on the principal cycle at position $m < n$ with $\alpha \in \{\pm a, \pm 2a\}$ and $\beta \equiv b \bmod a$ if and only if $(-a, 2b, ac)$ is equivalent to one of the ambiguous forms $\pm\mathbf{1}$, $\pm\mathbf{2}$.*

Proof. First suppose that the form $(a, 2\beta, *)$ appears as form F_m on the principal cycle with $m < n$. This form is equivalent to $(a, 2b, -ac)$, and $-\mathbf{1} \sim (a, 2b, -ac) \circ -\mathbf{1} \sim (-a, 2b, ac)$, so we are done. Similarly, if $F_m = (-a, 2\beta, *) \sim (-a, 2b, ac)$, then $(-a, 2b, ac) \sim \mathbf{1}$.

Now suppose that the form $(2a, 2\beta, *)$ appears as F_m. Then $-\mathbf{2} \sim (2a, 2\beta, *) \circ -\mathbf{2} \sim (-a, 2\beta, *) \sim (-a, 2b, ac)$. Finally, if $F_m = (-2a, 2\beta, *)$, then $\mathbf{2} \sim (-2a, 2\beta, *) \circ \mathbf{2} \sim (-a, 2\beta, *) \sim (-a, 2b, ac)$.

Finally, suppose that there is no form $(\alpha, 2\beta, *)$ with $\alpha \in \{\pm a, \pm 2a\}$ with $\beta \equiv b \bmod a$ which appears on the principal cycle as a form F_m with $m < n$. The square root $f = (-a, 2b, ac)$ cannot be equivalent to $\mathbf{1}$, since if it is then we can find a multiple of $2a$ that we can add to $2b$ to get an equivalent reduced form $(a, 2\beta, *)$ on the principal cycle with $\beta \equiv b \bmod a$. But then this form is a square root of F_n and hence must appear on the principal cycle before F_n, since $d(f, \mathbf{1}) = d(F_n, \mathbf{1})/2$. Therefore f cannot be equivalent to $\mathbf{1}$.

In fact, if $f \sim g$ with $g \in \{\pm\mathbf{1}, \pm\mathbf{2}\}$, then $f \circ g \sim \mathbf{1}$, and $f \circ g = (\alpha, 2\beta', *)$ for $\alpha \in \{\pm a, \pm 2a\}$ and $\beta' \equiv b \bmod a$. But then $f \circ g$ is a square root of F_n, and $f \circ g$ is equivalent to some reduced form $(\alpha, 2\beta, *)$ on the principal cycle with

$\beta \equiv \beta' \equiv b \bmod a$. As before this form must appear on the principal cycle before F_n, a contradiction. So it must be that f is not equivalent to any of the forms $\pm\mathbf{1}, \pm\mathbf{2}$. □

We now describe Shanks' method for determining when a square form is proper. For each form F_m that is examined, we perform the following test. Define $L = 2\sqrt{2\sqrt{N}}$. If Q_m is even and less than L, then put the pair $\left(Q_m/2, \overline{P_m}\right)$ into a queue, where $\overline{P_m}$ is the least positive residue of P_m modulo $Q_m/2$. If Q_m is odd and less than $L/2$, then put the pair $\left(Q_m, \overline{P_m}\right)$ into the queue, where $\overline{P_m}$ is the least positive residue of P_m modulo Q_m. If we come to the square form $F_n = \left(-a, 2b, c^2\right) \sim \left(c^2, -2b, -a\right)^{-1} \sim (-c, -2b, ac)^{-2} \sim (ac, 2b, -c)^2$, then we search the queue in the order that items are put into the queue for the pair $(c, 2b \bmod c)$, taking $c > 0$. Proposition 3.2 says that if this pair is in the queue, then the form $(ac, 2b, -c)$ is equivalent to one of the forms $\pm\mathbf{1}, \pm\mathbf{2}$; hence the square form is improper. If on the other hand the pair $(c, 2b \bmod c)$ is not in the queue, then Proposition 3.2 says that $(ac, 2b, -c)$ is not equivalent to one of the forms $\pm\mathbf{1}, \pm\mathbf{2}$; hence the square form is proper.

Note that the quantities placed in the queue will have one-quarter the precision of N. Hence, the queue entries will be relatively small and easy to work with. Also note that if we have found a square form $F_n = \left(-a, 2b, c^2\right)$ and also the pair $(c, 2b \bmod c)$ in the queue, then we may delete this pair along with all other pairs that precede it in the queue. This is possible since if we find another square form F_m with $n < m$, then any of its square roots appearing on the principal cycle must appear after the discovered square root for F_n because of the infrastructure explained in Section 2.4.

3.3. **The Algorithm.** In the following description of the algorithm, the variable N is the integer to factor, S remembers q_0, q holds the current q_i, P and P' hold two consecutive values of P_i, $\hat{Q}$ and Q hold two consecutive values of Q_i, and t is a temporary variable used in updating Q. The formulas (2.2), (2.3), and (2.4) are used to advance from one form to the next. Finally, B is an upper bound on the number of forms tested for being square forms before the algorithm gives up.

```
1. Initialize:
      Read the odd positive integer $N$ to be factored. If $N$ is the square of
   an integer, output the square root and stop. If $N \equiv 1 \bmod 4$, then set
   $D \leftarrow 2N$; otherwise, set $D \leftarrow N$. In any case, set $S \leftarrow \left\lfloor \sqrt{D} \right\rfloor$, $\hat{Q} \leftarrow 1$,
   $P \leftarrow S$, $Q \leftarrow D - P \cdot P$, $L \leftarrow \left\lfloor 2\sqrt{2\sqrt{D}} \right\rfloor$, $B \leftarrow 2 \cdot L$, and $i \leftarrow 0$.
      At this point the principal form is $(1, 2P, -Q) = (1, 2q_0, -(D - q_0^2))$.
2. Cycle forward to find a proper square form:
      Steps 2a through 2e are repeated for $i = 1, 2, 3, \ldots$.
   2a: Set $q \leftarrow \lfloor (S + P)/Q \rfloor$ and $P' \leftarrow q \cdot Q - P$ .
   2b: If $Q \leq L$, then:
         If $Q$ is even, put the pair $(Q/2, P \bmod (Q/2))$ onto the QUEUE; oth-
         erwise, if $Q \leq L/2$, then put the pair $(Q, P \bmod Q)$ onto the QUEUE.
   2c: Set $t \leftarrow \hat{Q} + q \cdot (P - P')$, $\hat{Q} \leftarrow Q$, $Q \leftarrow t$, and $P \leftarrow P'$.
         Here the current form is $(-\hat{Q}, 2P, Q)$ if $i$ is even, and it is $(\hat{Q}, 2P, -Q)$
         if $i$ is odd.
```

2d: If i is odd, go to Step 2e. If Q is not the square of an integer, go to Step 2e. Otherwise, set $r \leftarrow \sqrt{Q}$, a positive integer. If there is no pair (r, t) in the QUEUE for which r divides $P - t$, then go to Step 3. If $r > 1$ and there is a pair (r, t) in the QUEUE for which r divides $P - t$, then remove all pairs from the beginning of the QUEUE up to and including this pair and go to Step 2e. If $r = 1$ and there is a pair $(1, t)$ in the QUEUE, then the algorithm fails because we have traversed the entire principal period of quadratic forms of discriminant $4N$ without finding a proper square form.

2e: Let $i \leftarrow i + 1$. If $i > B$, give up. Otherwise, go to Step 2a.

3. Compute an inverse square root of the square form:

Here we have found a square form $F = (-\hat{Q}, 2P, r^2)$. Its inverse square root is $F^{-1/2} = (-r, 2P, r\hat{Q})$.

Set $\hat{Q} \leftarrow r$, $P \leftarrow P + r \cdot \lfloor (S - P)/r \rfloor$, and $Q \leftarrow (D - P \cdot P)/\hat{Q}$. (This last division is exact.)

Now the reduced inverse square root is the form $(-\hat{Q}, 2P, Q)$.

4. Cycle in the reverse direction to find a factor of N:

4a: Set $q \leftarrow \lfloor (S + P)/Q \rfloor$ and $P' \leftarrow q \cdot Q - P$.

4b: If $P = P'$, then go to Step 5.

4c: Set $t \leftarrow \hat{Q} + q \cdot (P - P')$, $\hat{Q} \leftarrow Q$, $Q \leftarrow t$, and $P \leftarrow P'$ and go to Step 4a.

5. Print the factor of N:

If Q is even, set $Q \leftarrow Q/2$. Output the factor Q of N.

The algorithm fails if the QUEUE overflows. For virtually all successful factorizations, a QUEUE size of 50 is adequate. In Theorem 4.24, we show that the average maximum queue size is about 1 or 2.

Step 4 is executed approximately half as many times as Step 2.

In Step 2b, Q almost always exceeds L. Also, Q is almost never a square in Step 2c. Thus, the time spent inserting pairs into the QUEUE and searching for them in it is negligible compared to the total time for Step 2.

3.4. **Examples.** We give one complete example of the SQUFOF algorithm. We will factor $N = 22117019 = D$. Note that $\sqrt{D} \approx 4702.873483$ and $2\sqrt{2\sqrt{D}} \approx 193.948447$. Thus $P = S = 4702$, $Q = D - P^2 = 8215$, and $L = 193$ in Step 1.

Table 1 shows the forms computed in Step 2 of SQUFOF and their infrastructure distances for $N = 22117019$. (The algorithm begins with form F_1, but F_0 is shown here because it is Shanks' origin for infrastructure distance. Of course, the SQUFOF algorithm computes no infrastructure distances.) In the algorithm descriptions, the form is represented by $\hat{Q}$, P, Q. Note that $\hat{Q}$ and Q in the algorithm are the absolute values of the numbers shown in Table 1, and that the factors of 2 in the middle coefficients of the forms do not appear in P of the algorithm.

At the end of Table 1 we have found the square form $F_{18} = (-6314, 2 \cdot 1737, 55^2)$. Since nothing has been placed into the queue, this is a proper square form. We compute its inverse square root as $(-55, 2 \cdot 1737, 347270)$ and reduce it to get $G_0 = (-55, 2 \cdot 4652, 8653)$.

Table 2 shows the forms computed in Step 4 of SQUFOF and their infrastructure distances for $N = 22117019$. In the algorithm descriptions, the form is represented

	F_i			
i	$(-1)^{i-1}Q_{i-1}$	$2 \cdot P_n$	$(-1)^i Q_i$	d_i
0	-8215	$2 \cdot 4702$	1	0.000000
1	1	$2 \cdot 4702$	-8215	4.642125
2	-8215	$2 \cdot 3513$	1190	5.608235
3	1190	$2 \cdot 3627$	-7531	6.631593
4	-7531	$2 \cdot 3904$	913	7.820150
5	913	$2 \cdot 4313$	-3850	9.390610
6	-3850	$2 \cdot 3387$	2765	10.298666
7	2765	$2 \cdot 2143$	-6338	10.790510
8	-6338	$2 \cdot 4195$	713	12.222178
9	713	$2 \cdot 4361$	-4346	13.860983
10	-4346	$2 \cdot 4331$	773	15.456075
11	773	$2 \cdot 4172$	-6095	16.864302
12	-6095	$2 \cdot 1923$	3022	17.298591
13	3022	$2 \cdot 4121$	-1699	18.658072
14	-1699	$2 \cdot 4374$	1757	20.316978
15	1757	$2 \cdot 4411$	-1514	22.037595
16	-1514	$2 \cdot 4673$	185	24.912057
17	185	$2 \cdot 4577$	-6314	27.062220
18	-6314	$2 \cdot 1737$	3025	27.449888

TABLE 1. Step 2 (Cycle forward) for $N = 22117019$.

	G_i			
i	$(-1)^{i-1}S_{i-1}$	$2 \cdot R_n$	$(-1)^i S_i$	d_i
0	-55	$2 \cdot 4652$	8653	2.607155
1	8653	$2 \cdot 4001$	-706	3.866041
2	-706	$2 \cdot 4471$	3013	5.705002
3	3013	$2 \cdot 4568$	-415	7.820150
4	-415	$2 \cdot 4562$	3145	9.913212
5	3145	$2 \cdot 1728$	-6083	10.298666
6	-6083	$2 \cdot 4355$	518	11.928441
7	518	$2 \cdot 4451$	-4451	13.724944
8	-4451	$2 \cdot 4451$	518	

TABLE 2. Step 4 (Cycle in reverse) for $N = 22117019$.

again by $\hat{Q}$, P, Q. Once more $\hat{Q}$ and Q in the algorithm are the absolute values of the numbers shown in Table 2.

Notice that the infrastructure distance covered in Step 4 (13.724944) is exactly half that covered in Step 2 (27.449888).

The algorithm does not actually compute the entire last form. As soon as it finds the middle coefficient P' of that form and notices that $P' = P$, the factor of N is at hand. Since $Q = 4451$ (the absolute value of $(-1)^i S_i = -4451$ in line 7 of Table 2) is odd, it is a factor of N and we find that $N = 4451 \cdot 4969$.

3.5. **Sufficient List.** Some implementations of SQUFOF do not use the previously described queue structure. Instead, when a form $(*,*,c)$ is discovered with $|c| < L$ when c is even, or with $|c| < L/2$ when c is odd, then $|c|$ is put into a list. Then any square form $\left(*,*,c^2\right)$ is ignored if $|c|$ is found to be in the list. This "sufficient list" is simpler, though potentially slower because some proper square forms may be skipped. For the running time analysis we will assume that the queue, and not the list, is used.

3.6. **Binary Quadratic Forms Description.** In [3], Cohen presents a different version of SQUFOF entirely in the language of binary quadratic forms. It reduces to the continued fraction version of SQUFOF whenever $N \equiv 2$ or $3 \bmod 4$. (However, the middle coefficients have their factors of 2 and the end coefficients have their proper signs.) Whenever $N \equiv 1 \bmod 4$ the algorithm defines $\Delta = N$ and works with this fundamental discriminant of binary quadratic forms. Although it is slower than the previous algorithm, because each iteration of ρ requires several divisions, the methods we use to analyze the complexity apply to it as well.

The binary quadratic forms version of SQUFOF follows. For simplicity we use the sufficient list instead of the queue. The description is shorter than that of the continued fraction version given above because we use the ρ function defined earlier.

```
1. Initialize:
      Read the odd positive integer $N$ to be factored. If $N$ is the square of an integer, output the square root and stop. If $N \equiv 1 \bmod 4$, then set $D \leftarrow N$, $m \leftarrow 1$, $d \leftarrow \left\lfloor \sqrt{D} \right\rfloor$, and $b \leftarrow 2\lfloor (d-1)/2 \rfloor + 1$. Otherwise ($N \equiv 2$ or $3 \bmod 4$), set $D \leftarrow 4 \cdot N$, $m \leftarrow 2$, $d \leftarrow \left\lfloor \sqrt{D} \right\rfloor$, and $b \leftarrow 2\lfloor d/2 \rfloor$. Let $F \leftarrow (1, b, (b^2 - D)/4)$, $i \leftarrow 2$, $L \leftarrow \left\lfloor \sqrt{d} \right\rfloor$, and $Bound \leftarrow 4 \cdot L$. Create an empty list. Let $g \leftarrow |(b^2 - D)/4| / \gcd(|(b^2 - D)/4|, m)$. If $g \le L$, add $g$ to the list.
2. Cycle forward to find a proper square form:
   2a: Set $F = (A, B, C) \leftarrow \rho(F)$, where $\rho$ was defined in 2.1.2.
   2b: If $i$ is even, go to Step 2d. If $C$ is the square of an integer, let $c > 0$ be a square root. If $c$ is not in the list, go to Step 3. If $c = 1$, stop because the algorithm has gone through the entire principal period without finding a proper square form.
   2c: Let $g \leftarrow |C| / \gcd(|C|, m)$. If $g \le L$, add $g$ to the list.
   2d: Let $i \leftarrow i + 1$. If $i > Bound$, give up. Otherwise, go to Step 2a.
3. Compute an inverse square root of the square form:
      Set $G = (a, b, c) \leftarrow \rho((cA, -B, -C))$.
4. Cycle in the reverse direction to find a factor of $N$:
   4a: Set $b' \leftarrow b$ and $G = (a, b, c) \leftarrow \rho(G)$.
   4b: If $b = b'$, then go to Step 5, else go to Step 4a.
5. Print the factor of $N$:
      If $c$ is even, let $c \leftarrow c/2$. Output $|c|$ as a non-trivial factor of $N$.
```

We give an example to illustrate the binary quadratic forms version of SQUFOF and factor an $N \equiv 1 \bmod 4$. We factor $N = 633003781$, with $D = N$, $\sqrt{D} \approx 25159.566391$, $d = 25159$, $b = 25159$, $(b^2 - D)/4 = -7125$, and $L = 158$.

The first two square forms we find in the forward cycle, $F_{154} = (-11923, 1509, 115^2)$ and $F_{184} = (-4009, 21359, 105^2)$, had square roots c already in the list, so we continued cycling forward. We eventually find the square form $F_{242} = (-18765, 23479, 33^2)$, which is proper because 33 is not in the list. The inverse square root of F_{242} is the form $(-33, -23479, 619245)$, which reduces to $G_0 = (-33, 25129, 11645)$. The form $G_{124} = (-15735, 8821, 8821)$ is ambiguous and gives the factor 8821 of N. The infrastructure distance covered in Step 4 (152.770486) is exactly half that traversed in Step 2 (305.540972).

4. SQUFOF Time and Space Complexity

To preserve the continuity of the complexity analysis of SQUFOF we collect in the next subsection some general lemmas we will need later.

4.1. Helpful Lemmas.

The following lemmas will aid in computing the average numbers of reduced and square forms on the principal cycle. We will find the asymptotic behavior of many quantities, all of which depend on N, the number we are trying to factor, or on Δ, which is either N or $4N$. We will write $f(N) \sim g(N)$ if $g(N) \neq 0$ for $N > 0$ and $\lim_{N\to\infty} f(N)/g(N) = 1$.

Lemmas 4.2 to 4.4 below are easily proved by mathematical induction, beginning with an identity of the form

$$\sum_{c<x, a\nmid c} h(c) = \sum_{c<x} h(c) - \sum_{c<x/a} h(ac)\ ,$$

and estimating the difference with the following simple lemma.

Lemma 4.1. *Let $a > b > 0$. Suppose $f(\Delta) \sim ah(\Delta)$ and $g(\Delta) \sim bh(\Delta)$, as $\Delta \to \infty$, where $h(\Delta) \neq 0$ for all Δ. Then $f(\Delta) - g(\Delta) \sim (a-b)h(\Delta)$, as $\Delta \to \infty$.*

Lemma 4.2. *Let Δ be a positive integer and suppose $p_1, \ldots, p_n$, for $n \geq 0$, are distinct small primes. Then, as $\Delta \to \infty$,*

$$\sum_{\substack{c=\sqrt{\Delta}/2 \\ p_i^2\nmid c,\ i=1,\ldots,n}}^{\sqrt{\Delta}} \frac{1}{c} \quad\sim\quad \log 2 \prod_{i=1}^{n} \frac{p_i^2-1}{p_i^2}\ .$$

Lemma 4.3. *Let Δ be a positive integer and suppose $p_1, \ldots, p_n$, for $n \geq 0$, are distinct small primes. Then, as $\Delta \to \infty$,*

$$\sum_{\substack{c=\sqrt[4]{\Delta}/\sqrt{2} \\ p_i\nmid c,\ i=1,\ldots,n}}^{\sqrt[4]{\Delta}} \frac{1}{c^2} \quad\sim\quad \frac{\sqrt{2}-1}{\sqrt[4]{\Delta}} \prod_{i=1}^{n} \frac{p_i-1}{p_i}\ .$$

Lemma 4.4. *Let Δ be a positive integer and suppose $p_1, \ldots, p_n$, for $n \geq 0$, are distinct small primes. Then, as $\Delta \to \infty$,*

$$\sum_{\substack{c=1 \\ p_i\nmid c,\ i=1,\ldots,n}}^{\sqrt[4]{\Delta}/\sqrt{2}} 1 \;-\; \sum_{\substack{c=\sqrt[4]{\Delta}/\sqrt{2} \\ p_i\nmid c,\ i=1,\ldots,n}}^{\sqrt[4]{\Delta}} 1 \quad\sim\quad \sqrt[4]{\Delta}\left(\sqrt{2}-1\right) \prod_{i=1}^{n} \frac{p_i-1}{p_i}\ .$$

4.2. **Outline of the Complexity Analysis.** As we have seen, once SQUFOF finds a proper square form F_n, it will find an ambiguous form (and factor N) at a distance of about $n/2$ forms away from $F_n^{-1/2}$. So we take the number of forms examined before finding a proper square form to be a fair measure of the running time. We omit from our analysis those N for which there is no proper square form in the principal cycle. SQUFOF cannot factor such N.

We have seen in Section 3 that SQUFOF generates several sequences depending on N. It looks for numbers with certain properties (proper squares). The complexity analysis is a heuristic argument based on several assumptions. Most of these assumptions say that these sequences of integers behave like random sequences of numbers of the same approximate size. Our first assumption, however, is not of this type. It simplifies the analysis by permitting the use of theorems about fundamental discriminants. It almost certainly holds in the most common uses of SQUFOF.

Assumption 4.5. *We assume that N is a square-free positive integer with k large odd prime divisors.*

Assuming that N is square-free (that is, using Assumption 4.5) implies that

$$\Delta = \begin{cases} N & \text{if } N \equiv 1 \bmod 4, \\ 4N & \text{if } N \equiv 2 \text{ or } 3 \bmod 4, \end{cases} \tag{4.1}$$

is a fundamental discriminant. This allows us to use many results from the theory of binary quadratic forms. Recall that if we use the continued fraction version of SQUFOF described in Sections 3.1 and 3.3, then any $N \equiv 1 \bmod 4$ will be multiplied by 2 at once. The $N \equiv 1 \bmod 4$ case here implies that we are using the binary quadratic forms version of Section 3.6.

In any case, SQUFOF is used mainly as an auxiliary algorithm in larger factorization algorithms and hence SQUFOF will typically be used to factor integers of modest size with no small prime factors. Such integers are typically the product of a small number of distinct primes.

4.3. **Counting Reduced Forms.** There is an obvious correspondence between forms of discriminant Δ and solutions to the congruence

$$b^2 \equiv \Delta \bmod 4c\,. \tag{4.2}$$

When $0 < y - x < 4c$, we will use the notation

$$N_{\Delta,c}(x, y) = |\{b \bmod 4c \,:\, (4.2) \text{ holds and } x < b < y\}|$$

later. Given an integer c, we will need to know the average number of reduced forms $(*, *, c)$. It is clear that there will be no such forms if c is divisible by any inert prime, or if c is divisible by the square of a ramified prime. So we may assume that c is divisible by no inert primes and by ramified primes to at most the first power. Under these restrictions, the following three lemmas calculate the number of solutions to (4.2) from which the number of reduced forms will follow.

Lemma 4.6. *Let $0 < c < \sqrt{\Delta}/2$ and suppose c is divisible by no inert primes, by ramified primes to at most the first power, and by exactly l distinct split primes. Then there are 2^l reduced forms $(*, *, c)$ of discriminant Δ.*

Proof. Suppose $c = q_1^{e_1} \cdots q_l^{e_l} r_1 \cdots r_t$, where the r_i are ramified primes and the q_j are split primes. For each odd r_i, the congruence $b^2 \equiv \Delta \bmod r_i$ has only the trivial solution. For each odd q_j, the congruence $b^2 \equiv \Delta \bmod q_j$ has exactly two solutions. Since these two solutions are both nonsingular, they each lift to a unique solution of $b^2 \equiv \Delta \bmod q_j^{e_j}$.

If c is odd, then we must count the number of solutions to $b^2 \equiv \Delta \bmod 4$. Since $\Delta \equiv 0$ or $1 \bmod 4$, in either case we have two solutions. Finally, the Chinese Remainder Theorem gives 2^{l+1} solutions to (4.2).

Now suppose c is even. Either 2 is ramified (so that 2 exactly divides c) or 2 is split. If 2 is ramified, then we must count the number of solutions to $b^2 \equiv \Delta \bmod 8$. Since $N \equiv 2$ or $3 \bmod 4$, we see that $\Delta \equiv 0$ or $4 \bmod 8$. There are two solutions to $b^2 \equiv 0 \bmod 8$ and two solutions to $b^2 \equiv 4 \bmod 8$, so once again the Chinese Remainder Theorem gives 2^{l+1} solutions to (4.2).

Finally, suppose 2 is a split prime and 2^e exactly divides c, where $e \geq 1$. We must count the number of solutions to $b^2 \equiv \Delta \bmod 2^{e+2}$. In this case $N \equiv 1 \bmod 8$. The congruence $b^2 \equiv 1 \bmod 8$ has four solutions. It is not hard to show that these four solutions lift to exactly four solutions of $b^2 \equiv \Delta \bmod 2^{e+2}$ for any $e \geq 1$ (c.f Theorem 2.24 of [10].) For any of the other $l-1$ odd split primes, we will have two solutions to $b^2 \equiv \Delta \bmod q_j^{e_j}$ as before. Again, the Chinese Remainder Theorem gives $4 \cdot 2^{l-1} = 2^{l+1}$ solutions to (4.2).

Recall that a form (a, b, c) is reduced if and only if $\left|\sqrt{\Delta} - 2|c|\right| < b < \sqrt{\Delta}$. By hypothesis $0 < c < \sqrt{\Delta}/2$; hence $\left|\sqrt{\Delta} - 2|c|\right| = \sqrt{\Delta} - 2c$. The condition $\sqrt{\Delta} - 2c < b < \sqrt{\Delta}$ defines an interval of length $2c$. Now suppose $0 < b_1, b_2, \ldots, b_{2^{l+1}} < 4c$ are the 2^{l+1} solutions of (4.2) in the interval $(0, 4c)$. Note that half of these solutions must be in $(0, 2c)$ and half must be in $(2c, 4c)$. By translating these solutions to the interval $(\sqrt{\Delta} - 4c, \sqrt{\Delta})$, we see that the 2^l solutions in $(\sqrt{\Delta} - 2c, \sqrt{\Delta})$ lead to 2^l reduced forms $(*, *, c)$ of discriminant Δ. Finally, if (a, b, c) is a reduced form of discriminant Δ, then clearly b must be one of the translated b_i. This finishes the proof of the lemma. □

Lemma 4.7. *Let $\sqrt{\Delta} < c$. There are no reduced forms $(*, *, c)$ of discriminant Δ.*

Proof. A form (a, b, c) is reduced if and only if $\left|\sqrt{\Delta} - 2|c|\right| < b < \sqrt{\Delta}$. No b can satisfy this condition since $\sqrt{\Delta} < c$ implies that $\sqrt{\Delta} < 2c - \sqrt{\Delta} = \left|\sqrt{\Delta} - 2|c|\right|$. □

The previous two lemmas give us the exact number of reduced forms $(*, *, c)$ of discriminant Δ whenever $0 < c < \sqrt{\Delta}/2$ or $\sqrt{\Delta} < c$. We must settle for an "average number" whenever $\sqrt{\Delta}/2 < c < \sqrt{\Delta}$. We make this notion precise as follows. Let Δ be a fundamental discriminant in the interval (c^2, ∞) and let $f(\Delta)$ denote a function of Δ. We say $f(\Delta)$ has average value $e(\Delta)$ if, as $c \to \infty$,

$$\sum_{c^2 < \Delta' \leq \Delta} f(\Delta') \sim \sum_{c^2 < \Delta' \leq \Delta} e(\Delta')\,.$$

We will write $A[f]$ for an average value e of f. Note that $A[\cdot]$ is asymptotically linear: If k is constant and f and g are two functions of Δ, then $A[kf + g] \sim kA[f] + A[g]$.

We make the following assumption regarding the distribution of quadratic residues in a complete system of residues modulo $4c$.

Assumption 4.8. *Let Δ be a fundamental discriminant in the interval $(c^2, 4c^2)$. Then the average value of $N_{\Delta,c}(x,y)$, as $c \to \infty$, is asymptotically*

$$\frac{y-x}{2c} N_{\Delta,c}(0,2c) \text{ when } 0 < x < y < 2c$$

and

$$\frac{y-x}{2c} N_{\Delta,c}(2c,4c) \text{ when } 2c < x < y < 4c.$$

In the following, when we write "$\Delta \to \infty$" we mean that $\Delta \to \infty$ through fundamental discriminants. Often there will be other restrictions on Δ, such as that it lie in a certain residue class modulo 4. Remember that Δ and N are always related by (4.1), so that we may write $N \to \infty$ instead of $\Delta \to \infty$.

Lemma 4.9. *Let $\sqrt{\Delta}/2 < c < \sqrt{\Delta}$ and suppose c is divisible by no inert primes, by ramified primes to at most the first power, and by exactly l distinct split primes. Then, as $\Delta \to \infty$, the average number of reduced forms $(*,*,c)$ of discriminant Δ is asymptotically*

$$\frac{2^l\left(\sqrt{\Delta}-c\right)}{c}.$$

Proof. Since $\sqrt{\Delta}/2 < c < \sqrt{\Delta}$, the condition $\left|\sqrt{\Delta} - 2|c|\right| < b < \sqrt{\Delta}$ is equivalent to $2c - \sqrt{\Delta} < b < \sqrt{\Delta}$. This defines the interval $(2c-\sqrt{\Delta}, \sqrt{\Delta})$ of length $2\left(\sqrt{\Delta}-c\right)$. We translate the 2^{l+1} solutions of the congruence (4.2) in $(0,4c)$ to the interval $(2c-\sqrt{\Delta}, 6c-\sqrt{\Delta})$. Half of these solutions will be in the interval $(2c-\sqrt{\Delta}, 4c-\sqrt{\Delta})$. We apply Assumption 4.8 with $x = 2c - \sqrt{\Delta} > 0$ and $y = \sqrt{\Delta} < 2c$. Then $(y-x)/2c = (\sqrt{\Delta}-c)/c$ and $N_{\Delta,c}(0,2c) = 2^l$. The number of reduced forms we are counting equals the number of solutions to (4.2) with $2c - \sqrt{\Delta} < b < \sqrt{\Delta}$, that is, $N_{\Delta,c}(2c-\sqrt{\Delta}, \sqrt{\Delta})$. By Assumption 4.8,

$$\sum_{c^2 < \Delta' \leq \Delta} N_{\Delta',c}(2c-\sqrt{\Delta'}, \sqrt{\Delta'}) \sim \sum_{c^2 < \Delta' \leq \Delta} \frac{\left(\sqrt{\Delta'}-c\right)}{c} 2^l$$

as $\Delta \to \infty$. This shows that the average number of reduced forms $(*,*,c)$ of discriminant Δ is asymptotically $2^l(\sqrt{\Delta}-c)/c$, as $\Delta \to \infty$, as claimed. □

Note that if (a,b,c) is a reduced form of discriminant Δ, then so is $(-a,b,-c)$, so the previous three lemmas tell us the average number of forms $(*,*,-c)$ for $c > 0$. For $c > 0$ we let $Y_c = Y_c(\Delta)$ be the number of reduced forms (a',b',c') of discriminant Δ with $|c'| = c$. We will not compute this quantity for every possible value of c. Instead we compute the average value $A[Y_c]$ of Y_c. The previous three lemmas can be used to compute this quantity.

The fraction of non-ramified primes $p < \sqrt{\Delta}$ that split is asymptotically $1/2$, as $\Delta \to \infty$, by the Chebotarev density theorem.

Proposition 4.10. *Suppose $c > 0$ is an integer divisible by ramified primes to at most the first power, and let Y_c be the number of reduced forms $(*,*,c')$ of* *discriminant* Δ *with* $|c'| = c$. *Then, as* $\Delta \to \infty$,

$$A[Y_c] \sim \begin{cases} 2 & \text{if } 0 < c < \sqrt{\Delta}/2 \ , \\ \frac{2(\sqrt{\Delta}-c)}{c} & \text{if } \sqrt{\Delta}/2 < c < \sqrt{\Delta} \ , \\ 0 & \text{if } \sqrt{\Delta} < c \ . \end{cases}$$

Proof. First suppose $0 < c < \sqrt{\Delta}/2$ and that c is divisible by l non-ramified primes. By the remark above, the fraction of c divisible by no inert prime is 2^{-l}. Lemma 4.6 says that if c is divisible by no inert primes, by ramified primes to at most the first power, and by exactly l split primes, then there will be 2^l reduced forms $(*, *, c)$ of discriminant Δ. So we have, as $\Delta \to \infty$,

$$A[Y_c] \sim 2\left(2^{-l} \cdot 2^l + (1 - 2^{-l}) \cdot 0\right) = 2 \ ,$$

where we multiply by two since (a, b, c) is a reduced form of discriminant Δ if and only if $(-a, b, -c)$ is a reduced form of discriminant Δ.

Now suppose that $\sqrt{\Delta}/2 < c < \sqrt{\Delta}$, and that c is divisible by l non-ramified primes. Again, the fraction of c divisible by no inert prime is 2^{-l}. Lemma 4.9 implies that if c is divisible by no inert primes, by ramified primes to at most the first power, and by exactly l split primes, then we expect $2^l(\sqrt{\Delta} - c)/c$ reduced forms $(*, *, c)$ of discriminant Δ. So we have, as $\Delta \to \infty$,

$$A[Y_c] \sim 2\left(2^{-l} \cdot \frac{2^l\left(\sqrt{\Delta} - c\right)}{c} + (1 - 2^{-l}) \cdot 0\right) = \frac{2\left(\sqrt{\Delta} - c\right)}{c} \ .$$

Finally, suppose $\sqrt{\Delta} < c$. Lemma 4.7 implies that there are no reduced forms $(*, *, c)$ of discriminant Δ; hence $A[Y_c] = 0$. □

4.4. **Successive Square Forms.**

We now use the results of the previous subsection to compute the average index-difference between successive square forms. Using similar techniques, we will count both the average number of reduced forms and the average number of reduced square forms on the principal cycle. Then the average number of steps between successive square forms will be the ratio of these two average numbers.

Let C be the group of equivalence classes of binary quadratic forms of discriminant Δ. Recall that this group is isomorphic to the narrow class group of $\mathbb{Q}(\sqrt{\Delta})$; hence $|C| = h^+$, the narrow class number of $\mathbb{Q}(\sqrt{\Delta})$. Let G be the group of genera of forms of discriminant Δ. There is a surjective group homomorphism $\phi : C \to G$ taking an equivalence class to its genus, which we identify with its corresponding assigned value. The kernel of this homomorphism is the set of classes in the principal genus. The first group isomorphism theorem implies that $C/\ker\phi \cong G$; hence $|\ker\phi| = h^+/|G|$. It remains to compute the value of $|G|$.

Let $\kappa = k$ when $N \equiv 1 \bmod 4$, $\Delta = N$ and $\kappa = k + 1$ when $N \equiv 2$ or $3 \bmod 4$, $\Delta = 4N$. Then κ is the number of generic characters of Δ, as defined in Section 2.1.2. Since the number of genera is equal to one half the possible assigned values, we see that $|G| = 2^{\kappa-1}$. Finally we see that the number of classes in the principal genus is $h^+/2^{\kappa-1}$.

4.4.1. *Number of Reduced Forms on the Principal Cycle.*

Let $c > 0$, $X_c = X_c(\Delta)$ be the number of reduced forms $(*, *, c')$ of discriminant Δ with $|c'| = c$ on the principal cycle, and $X = X(\Delta)$ be the total number of reduced forms with discriminant Δ

on the principal cycle. Then $X = \sum_{0<c} X_c$. We have seen (Lemma 4.7) that if $(*,*,c)$ is a reduced form, then $0 < |c| < \sqrt{\Delta}$, so

$$X = \sum_{c=1}^{\sqrt{\Delta}} X_c \; .$$

We will compute $A[X]$, the average number of reduced forms on the principal cycle, and we have, as $\Delta \to \infty$,

$$A[X] \sim \sum_{c=1}^{\sqrt{\Delta}} A[X_c] \; .$$

We now make a few observations about the distribution of forms among the the h^+ cycles. First observe that since the principal cycle is ambiguous, a non-ambiguous reduced form (a, b, c) will be on the principal cycle if and only if its associate (c, b, a) is on the principal cycle. But this means that (a, b, c) is on the principal cycle if and only if $\rho^{-1}(c, b, a) = (a', b', c)$ is on the principal cycle. Existence of the reduced forms (a, b, c), (a', b', c) implies the existence of the reduced forms $(-a, b, -c)$, $(-a', b', -c)$. These four forms will be collectively referred to as the *quartet of forms* associated with the form (a, b, c).

Suppose that one of the κ generic characters ξ associated to Δ satisfies $\xi(-1) = -1$. Then the forms (a, b, c), (a', b', c) are on the principal cycle if and only if the forms $(-a, b, -c)$, $(-a', b', -c)$ are *not* on the principal cycle. (Here we are using Assumption 4.5.) Therefore, for a given $c > 0$, at most two forms from each quartet can be on the principal cycle. This leads us to make the following assumption.

Assumption 4.11. *Assume that, as $\Delta \to \infty$ through values for which there is some generic character ξ associated with Δ such that $\xi(-1) = -1$, we have*

$$A[X_c] \sim A[Y_c]/2h^+ \; . \tag{4.3}$$

Assumption 4.11 is reasonable because for each quartet of forms $(*,*,c')$ with $|c'| = c$, at most two forms (that is, half of the forms in the quartet) may be on the principal cycle, and the principal cycle is one of the h^+ cycles. Since Proposition 4.10 gives us an expression for $A[Y_c]$ when $0 < c < \sqrt{\Delta}$, Assumption 4.11 enables us to compute $A[X]$.

Now suppose that $\xi(-1) = 1$ for all generic characters ξ associated to Δ. Then (a, b, c) is on the principal cycle if and only if its entire quartet is on the principal cycle. (Here we are using Assumption 4.5 again.) We now make the following assumption.

Assumption 4.12. *Assume that, as $\Delta \to \infty$ through values for which $\xi(-1) = 1$ for all generic characters ξ associated with Δ, we have*

$$A[X_c] \sim A[Y_c]/h^+ \; .$$

Assumption 4.12 is reasonable because either the entire quartet associated with (a, b, c) is on the principal cycle or not, and the principal cycle is one of the h^+ cycles. In summary, we have assumed that $A[X_c] \sim \nu A[Y_c]/h^+$, as $\Delta \to \infty$, where

$$\nu = \nu(\Delta) = \begin{cases} \frac{1}{2} & \text{if } \xi(-1) = -1 \text{ for some generic character } \xi \text{ of } \Delta \; , \\ 1 & \text{if } \xi(-1) = 1 \text{ for all generic characters } \xi \text{ of } \Delta \; . \end{cases}$$

If $N \equiv 3 \bmod 4$, then some prime dividing N must be congruent to 3 modulo 4. In this case ν must equal $1/2$ and so (4.3) holds. If $N \equiv 1$ or $2 \bmod 4$, then we cannot know which value to use for ν, so we make the following assumption about the average value of ν.

Assumption 4.13. *Assume that when $N \equiv 1$ or $2 \bmod 4$ we have, as $\Delta \to \infty$,*

$$A[\nu] \sim \left(1 - 2^{-\kappa}\right) \cdot \frac{1}{2} + 2^{-\kappa} \cdot 1 = \frac{2^\kappa + 1}{2^{\kappa+1}} \ ,$$

Assumption 4.13 is reasonable because the κ generic characters associated with Δ should independently each take the value 1 at -1 about half of the time, so that all take the value 1 at -1 in about one case out of 2^κ. Assumptions 4.11, 4.12 and 4.13 together imply that when $N \equiv 1$ or $2 \bmod 4$,

$$A[X_c] \sim A[\nu]A[Y_c]/h^+ \sim \frac{(2^\kappa + 1)A[Y_c]}{2^{\kappa+1}h^+} \ . \tag{4.4}$$

In any case, we can now calculate $A[X]$.

Proposition 4.14. *As $\Delta \to \infty$, the asymptotic average number of reduced forms of discriminant Δ on the principal cycle is*

$$A[X] \sim \begin{cases} \dfrac{\left(2^k+1\right)\sqrt{N}\log 2}{2^k h^+} & \textit{if } N \equiv 1 \bmod 4 \ , \\[2ex] \dfrac{3\left(2^{k+1}+1\right)\sqrt{N}\log 2}{2^{k+2} h^+} & \textit{if } N \equiv 2 \bmod 4 \ , \\[2ex] \dfrac{3\sqrt{N}\log 2}{2h^+} & \textit{if } N \equiv 3 \bmod 4 \ . \end{cases}$$

Proof. We assume that the odd prime divisors p_i of N (all of which are ramified) are so large that the chance that c is divisible by p_i^2 is negligibly small. This means that we shall use the results of Proposition 4.10 for all values of c, except when 2 is ramified ($N \equiv 2$ or $3 \bmod 4$.) When 2 is ramified, we will use $A[X_c] = 0$ for any c divisible by 4.

Note that to get the result in the case of $N \equiv 2 \bmod 4$, we may multiply the result in the case of $N \equiv 3 \bmod 4$ by $(2^\kappa + 1)/2^\kappa = \left(2^{k+1}+1\right)/2^{k+1}$, since the only difference is that we replace $1/2$ with $A[\nu]$.

Case 1: ($N \equiv 1 \bmod 4$) In this case $\Delta = N$. We have

$$\begin{aligned} A[X] &\sim \sum_{c=1}^{\sqrt{\Delta}} A[X_c] \sim \sum_{c=1}^{\sqrt{N}} \frac{\left(2^k+1\right)A[Y_c]}{2^{k+1}h^+} \quad \text{(using Equation (4.4))} \\ &\sim \frac{2^k+1}{2^{k+1}h^+} \sum_{c=1}^{\sqrt{N}} A[Y_c] \\ &\sim \frac{2^k+1}{2^{k+1}h^+} \left[\sum_{c=1}^{\sqrt{N}/2} 2 \ + \sum_{c=\sqrt{N}/2}^{\sqrt{N}} \frac{2\left(\sqrt{N}-c\right)}{c} \right] \quad \text{(by Proposition 4.10)} \\ &\sim \frac{\left(2^k+1\right)\sqrt{N}}{2^k h^+} \sum_{c=\sqrt{N}/2}^{\sqrt{N}} \frac{1}{c} \end{aligned}$$

$$\sim \frac{\left(2^k+1\right)\sqrt{N}\log 2}{2^k h^+} \quad \text{(using Lemma 4.2).}$$

Case 2: ($N \equiv 3 \bmod 4$) In this case $\Delta = 4N$, and 2 is a ramified prime. We have

$$\begin{aligned}
A[X] &\sim \sum_{c=1}^{\sqrt{\Delta}} A[X_c] \sim \sum_{c=1}^{2\sqrt{N}} A[Y_c]/2h^+ \; - \sum_{c=1}^{\sqrt{N}/2} A[Y_{4c}]/2h^+ \\
&\qquad\qquad\qquad \text{(using Equation (4.3))} \\
&\sim \frac{1}{2h^+}\left[\sum_{c=1}^{2\sqrt{N}} A[Y_c] \; - \sum_{c=1}^{\sqrt{N}/2} A[Y_{4c}]\right] \\
&\sim \frac{1}{2h^+}\left[\sum_{c=1}^{\sqrt{N}} 2 \; + \sum_{c=\sqrt{N}}^{2\sqrt{N}} \frac{2\left(2\sqrt{N}-c\right)}{c}\right. \\
&\qquad\qquad \left. - \sum_{c=1}^{\sqrt{N}/4} 2 \; - \sum_{c=\sqrt{N}/4}^{\sqrt{N}/2} \frac{2\left(2\sqrt{N}-4c\right)}{4c}\right] \quad \text{(by Proposition 4.10)} \\
&\sim \frac{1}{2h^+}\left[4\sqrt{N}\sum_{c=\sqrt{N}}^{2\sqrt{N}} \frac{1}{c} \; - \sqrt{N}\sum_{c=\sqrt{N}/4}^{\sqrt{N}/2} \frac{1}{c}\right] \\
&\sim \frac{3\sqrt{N}\log 2}{2h^+} \quad \text{(using Lemma 4.2).}
\end{aligned}$$

□

4.4.2. *Number of Square Forms on the Principal Cycle.* We can use the same methods used in the previous subsection to count $X_{sq} = X_{sq}(\Delta)$, the number of reduced square forms $(*, *, c^2)$ on the principal cycle. As before, we will actually compute $A[X_{sq}]$, the average number of reduced square forms on the principal cycle. Here we begin with $X_{sq} = \sum X_{c^2}/2$, where we divide by two since square forms must have a positive right-end coefficient and exactly half of the X_{c^2} forms will satisfy this condition. Lemma 4.7 implies that

$$X_{sq} = \sum_{c=1}^{\sqrt[4]{\Delta}} X_{c^2}/2 \; .$$

Hence

$$A[X_{sq}] \sim \sum_{c=1}^{\sqrt[4]{\Delta}} A[X_{c^2}]/2 \; .$$

As before, for a given $c > 0$ there are Y_{c^2} reduced forms (a, b, c') of discriminant Δ with $|c'| = c^2$. Also, for each non-ambiguous form (a, b, c^2) we have the associated quartet of forms: (a, b, c^2), $\rho^{-1}(c^2, b, a) = (a', b', c^2)$, $(-a, b, -c^2)$, and $(-a', b', -c^2)$. We make the following assumption.

Assumption 4.15.

$$A[X_{c^2}] \sim \begin{cases} \dfrac{2^\kappa+1}{4h^+}\, A[Y_{c^2}] & \textit{if } N \equiv 1 \textit{ or } 2 \bmod 4, \\[2ex] \dfrac{2^{\kappa-2}}{h^+}\, A[Y_{c^2}] & \textit{if } N \equiv 3 \bmod 4. \end{cases} \tag{4.5}$$

We justify Assumption 4.15 as follows. When $N \equiv 3 \bmod 4$, reason as for Assumption 4.11, noting that a square form has to lie on a cycle in the principal genus, and the principal cycle is one of the $h^+/2^{\kappa-1}$ cycles in the principal genus. When $N \equiv 1$ or $2 \bmod 4$, reason as for (4.4), with the same change noted.

Since Proposition 4.10 gives us an expression for $A[Y_{c^2}]$ when $0 < c < \sqrt[4]{\Delta}$, we can now compute $A[X_{sq}]$.

Proposition 4.16. *As $\Delta \to \infty$, the asymptotic average number of reduced square forms of discriminant Δ on the principal cycle is*

$$A[X_{sq}] \sim \begin{cases} \dfrac{\left(2^k+1\right)\left(\sqrt{2}-1\right)\sqrt[4]{N}}{2h^+} & \textit{if } N \equiv 1 \bmod 4\ , \\[2ex] \dfrac{\left(2^{k+1}+1\right)\left(2-\sqrt{2}\right)\sqrt[4]{N}}{4h^+} & \textit{if } N \equiv 2 \bmod 4\ , \\[2ex] \dfrac{2^k\left(2-\sqrt{2}\right)\sqrt[4]{N}}{2h^+} & \textit{if } N \equiv 3 \bmod 4\ . \end{cases}$$

Proof. As in the proof of Proposition 4.14, we assume that the odd prime divisors p_i of N are so large that the fraction of c^2 that are divisible by p_i^2 is negligibly small. So we shall once again use the results of Proposition 4.10 for all values of c^2, except when 2 is ramified. When 2 is ramified, $2|c^2$ implies that $4|c^2$, and hence $A[X_{c^2}] = 0$. Also we can easily obtain the result for $N \equiv 2 \bmod 4$ once we have the result for $N \equiv 3 \bmod 4$ as in Proposition 4.14.

Case 1: ($N \equiv 1 \bmod 4$) In this case $\Delta = N$. We have

$$\begin{aligned} A[X_{sq}] &\sim \sum_{c=1}^{\sqrt[4]{\Delta}} A[X_{c^2}]/2 \sim \sum_{c=1}^{\sqrt[4]{N}} \frac{\left(2^\kappa+1\right)A[Y_{c^2}]}{8h^+} \quad \text{(by Equation (4.5))} \\ &\sim \frac{2^\kappa+1}{8h^+}\sum_{c=1}^{\sqrt[4]{N}} A[Y_{c^2}] \\ &\sim \frac{2^\kappa+1}{8h^+}\left[\sum_{c=1}^{\sqrt[4]{N}/\sqrt{2}} 2 \;+\; \sum_{c=\sqrt[4]{N}/\sqrt{2}}^{\sqrt[4]{N}} \frac{2\left(\sqrt{N}-c^2\right)}{c^2}\right] \quad \text{(by Proposition 4.10)} \\ &\sim \frac{2^\kappa+1}{4h^+}\left[\left(\sqrt{2}-1\right)\sqrt[4]{N} \;+\; \sqrt{N}\sum_{c=\sqrt[4]{N}/\sqrt{2}}^{\sqrt[4]{N}} \frac{1}{c^2}\right] \\ &\sim \frac{\left(2^\kappa+1\right)\left(\sqrt{2}-1\right)\sqrt[4]{N}}{2h^+} \quad \text{(by Lemma 4.3).} \end{aligned}$$

When $N \equiv 1 \bmod 4$, we have $\kappa = k$ generic characters, one for each prime divisor of N. Thus

$$A[X_{sq}] \sim \frac{\left(2^k + 1\right)\left(\sqrt{2} - 1\right)\sqrt[4]{N}}{2h^+}\,.$$

Case 2: ($N \equiv 3 \bmod 4$) In this case $\Delta = 4N$.

$$\begin{aligned}
A[X_{sq}] &\sim \sum_{c=1}^{\sqrt[4]{\Delta}} A[X_{c^2}]/2 \\
&\sim \sum_{c=1}^{\sqrt{2}\sqrt[4]{N}} \frac{2^{\kappa-2}A[Y_{c^2}]}{2h^+} - \sum_{c=1}^{\sqrt[4]{N}/\sqrt{2}} \frac{2^{\kappa-2}A[Y_{4c^2}]}{2h^+} \\
&\qquad \text{(by Equation (4.5))} \\
&\sim \frac{2^\kappa}{8h^+}\left[\sum_{c=1}^{\sqrt{2}\sqrt[4]{N}} A[Y_{c^2}] \;-\; \sum_{c=1}^{\sqrt[4]{N}/\sqrt{2}} A[Y_{4c^2}]\right] \\
&\sim \frac{2^\kappa}{8h^+}\left[\sum_{c=1}^{\sqrt[4]{N}} 2 \;+\; \sum_{c=\sqrt[4]{N}}^{\sqrt{2}\sqrt[4]{N}} \frac{2\left(2\sqrt{N} - c^2\right)}{c^2}\right. \\
&\qquad\left. - \sum_{c=1}^{\sqrt[4]{N}/2} 2 \;-\; \sum_{c=\sqrt[4]{N}/2}^{\sqrt[4]{N}/\sqrt{2}} \frac{2\left(2\sqrt{N} - 4c^2\right)}{4c^2}\right] \\
&\qquad \text{(by Proposition 4.10)} \\
&\sim \frac{2^\kappa}{4h^+}\left[\left(2 - \sqrt{2}\right)\sqrt[4]{N} \;+\; 2\sqrt{N}\sum_{c=\sqrt[4]{N}}^{\sqrt{2}\sqrt[4]{N}} \frac{1}{c^2}\right. \\
&\qquad\left. - \frac{2-\sqrt{2}}{2}\sqrt[4]{N} \;-\; \frac{\sqrt{N}}{2}\sum_{c=\sqrt[4]{N}/2}^{\sqrt[4]{N}/\sqrt{2}} \frac{1}{c^2}\right] \\
&\sim \frac{2^\kappa\left(2 - \sqrt{2}\right)\sqrt[4]{N}}{4h^+} \quad \text{(by Lemma 4.3).}
\end{aligned}$$

In this case $\kappa = k + 1$, thus

$$A[X_{sq}] \sim \frac{2^k\left(2 - \sqrt{2}\right)\sqrt[4]{N}}{2h^+}\,.$$

□

The Brauer-Siegel theorem (see pages 216 and 297 of [3]) says that $\log(R(\Delta)h(\Delta)) \sim \log(\sqrt{\Delta})$ as $\Delta \to \infty$, where $R(\Delta)$ is the regulator of $\mathbb{Q}(\sqrt{N})$. It is conjectured that $R(\Delta)$ usually has size about $\sqrt{\Delta}$, so that h and therefore h^+ are typically very small. In this "usual" case, we have $X \approx \text{constant} \cdot \sqrt{N}/h^+ \approx \text{constant} \cdot \sqrt{N}$ and $X_{sq} \approx \text{constant} \cdot \sqrt[4]{N}/h^+ \approx \text{constant} \cdot \sqrt[4]{N}$.

4.4.3. *Average Index-Difference between Successive Square Forms.* SQUFOF begins with the first reduced form $(1, 2P, -Q)$ following a (trivial) reduced square form $(-Q, 2P, 1^2)$ in the principal cycle (using the continued fraction description).

It steps through the principal cycle until it finds the next reduced square form. Our measure of the time complexity of SQUFOF to factor N is the number of reduced forms it examines in the principal cycle. In our heuristic analysis we will approximate this number by the average number of reduced forms between successive square forms in the principal cycle. Let $D = D(\Delta) = X/X_{sq} = X(\Delta)/X_{sq}(\Delta)$ be this average number. The following assumption allows us to use Propositions 4.14 and 4.16 to find the average index-difference $A[D]$ between successive square forms.

Assumption 4.17. *Assume that, as $\Delta \to \infty$, we have $A[D] \sim A[X]/A[X_{sq}]$,*

Assumption 4.17 is plausible because we are using averages, $A[X]$ is roughly of size $\sqrt{\Delta}$ and $A[X_{sq}]$ is roughly of size $\sqrt[4]{\Delta}$, so that $A[D]$ is roughly of size $\sqrt[4]{\Delta}$.

The following corollary allows us to compute $A[D]$.

Corollary 4.18. *As $N \to \infty$, we have*

$$A[D] \sim \begin{cases} \dfrac{\left(\sqrt{2}+1\right)\sqrt[4]{N}\log 2}{2^{k-1}} & \textit{if } N \equiv 1 \bmod 4\ , \\[2ex] \dfrac{3\left(\sqrt{2}+2\right)\sqrt[4]{N}\log 2}{2^{k+1}} & \textit{if } N \equiv 2 \textit{ or } 3 \bmod 4\ . \end{cases} \tag{4.6}$$

Proof. We prove the case $N \equiv 1 \bmod 4$. The cases $N \equiv 2$ and $3 \bmod 4$ are proved in the same way.

Case 1: ($N \equiv 1 \bmod 4$) Proposition 4.14 implies that

$$A[X] \sim \frac{\left(2^k+1\right)\sqrt{N}\log 2}{2^k h^+}\ ,$$

and Proposition 4.16 implies that

$$A[X_{sq}] \sim \frac{\left(2^k+1\right)\left(\sqrt{2}-1\right)\sqrt[4]{N}}{2h^+}\ .$$

Thus, by Assumption 4.17,

$$A[D] \sim \frac{\left(\left(2^k+1\right)\sqrt{N}\log 2\right)\Big/2^k h^+}{\left(\left(2^k+1\right)\left(\sqrt{2}-1\right)\sqrt[4]{N}\right)\Big/2h^+} = \frac{\left(\sqrt{2}+1\right)\sqrt[4]{N}\log 2}{2^{k-1}}\ .$$

□

4.5. **Proper Square Forms.** In this subsection we will derive the average number of square forms we must examine to successfully factor N. Recall that when SQUFOF finds a square form, it forms an inverse square root and follows its cycle to an ambiguous form, where there is a factor of N. Also, a proper square form is one that leads to an ambiguous form with a nontrivial divisor of N. As before, κ is the number of generic characters for Δ. We assume that any square form in the principal cycle is equally likely to lead to any ambiguous form.

Assumption 4.19. *Let f be one of the 2^κ reduced ambiguous forms of discriminant Δ. If one begins with a square form $(*, *, c^2)$ on the principal cycle of reduced forms of discriminant Δ, computes its inverse square root as SQUFOF does, and follows its cycle to the first ambiguous form, then there is one chance in 2^κ that this ambiguous form will be f.*

Assumption 4.19 is reasonable because the square root of a square form on the principal cycle must lie in an ambiguous cycle, so we will reach one of the 2^κ ambiguous forms, and f is one of them. Recall that we are modeling SQUFOF as a random walk. We are assuming that when we compute the inverse square root, we jump to a random ambiguous cycle.

We always assume that N is the product of k distinct primes. To prepare for multipliers in Section 5 we prove the next proposition in the case when Δ contains not only the prime factors of N, but also those of an odd multiplier. The latter primes are small and known before SQUFOF begins. Furthermore, we will show later how to tell whether a square form on the principal cycle will lead to a divisor of twice the multiplier, so that SQUFOF with a multiplier can avoid finding a trivial factor of N.

Proposition 4.20. *The average asymptotic fraction of square forms that are proper is*

$$\frac{2^k - 2}{2^k} \; . \tag{4.7}$$

Proof. As we have seen, a square form leads to an ambiguous form f, hence to a factor of Δ. There are as many ambiguous classes as there are genera, and this latter quantity is known to be $2^{\kappa-1}$. There are two ambiguous forms per ambiguous class; hence there are 2^κ ambiguous forms. These forms are in bijective correspondence with the square-free divisors d of Δ with $|d| < \sqrt{\Delta}$.

Now suppose Δ has n small ramified primes (known prime factors of a multiplier, or 2 when $N \equiv 2$ or 3 mod 4) and k large ramified primes (the factors of N). Then $\kappa = k + n$ and there will be 2^{n+1} improper squares (one for each of the possible 2^{n+1} square-free divisors d of Δ with $|d| < \sqrt{\Delta}$ and divisible only by the small ramified primes). Thus, by Assumption 4.19, the fraction of square forms that are proper is

$$\frac{2^\kappa - 2^{n+1}}{2^\kappa} = \frac{2^{n+k} - 2^{n+1}}{2^{n+k}} = \frac{2^k - 2}{2^k} \; .$$

□

Corollary 4.21. *The asymptotic average number of square forms that SQUFOF must examine before finding a proper square form is*

$$\frac{2^k}{2^k - 2} \; . \tag{4.8}$$

4.6. **The Time Complexity of SQUFOF.** We now have everything we need to compute the asymptotic behavior of the average number of forms SQUFOF must examine to find a proper square form.

Theorem 4.22. *Let $W = W(N)$ be the number of quadratic forms that SQUFOF, when factoring a square-free integer N having k prime factors, must examine before finding a proper square form. Then, as $N \to \infty$, the asymptotic average value of W is*

$$A[W] \sim \begin{cases} \dfrac{2\left(\sqrt{2}+1\right)\sqrt[4]{N}\log 2}{2^k - 2} & \text{if } N \equiv 1 \bmod 4\;, \\[2ex] \dfrac{3\left(\sqrt{2}+2\right)\sqrt[4]{N}\log 2}{2\left(2^k - 2\right)} & \text{if } N \equiv 2 \text{ or } 3 \bmod 4\;. \end{cases}$$

k	$A[W]/\sqrt[4]{N},\ N \equiv 1 \bmod 4$	$A[W]/\sqrt[4]{N},\ N \equiv 2 \text{ or } 3 \bmod 4$
2	1.6734	1.7749
3	0.5578	0.5916
4	0.2391	0.2536

TABLE 3. Estimates of $A[W]/\sqrt[4]{N}$ for $k = 2, 3, 4$.

Proof. This is simply the product of (4.6) and (4.8). □

Table 3 lists the predicted values for $A[W]/\sqrt[4]{N}$ when N is a product of two, three, and four primes.

4.7. **Average Queue Size.** Now that we have the average number of forms that SQUFOF will examine before finding a proper square form, it is a simple matter to calculate the average queue size. If $N \equiv 1 \bmod 4$, then $(*, *, c)$ will be enqueued if $|c| < \sqrt[4]{\Delta}$. There are $2\sqrt[4]{\Delta}$ integers c such that $|c| < \sqrt[4]{\Delta}$, of which only $3\sqrt[4]{\Delta}/2$ satisfy $4 \nmid c$. There are $3\sqrt{\Delta}/2$ integers c such that $|c| < \sqrt{\Delta}$ and $4 \nmid c$. If we chose random integers c not divisible by 4 with uniform distribution from the interval $(-\sqrt{\Delta}, \sqrt{\Delta})$, about one in every $\sqrt[4]{\Delta}$ integers c would satisfy $|c| < \sqrt[4]{\Delta}$. We will assume that the same fraction $1/\sqrt[4]{\Delta}$ of actual numbers c that arise when SQUFOF is used to factor N satisfy $|c| < \sqrt[4]{\Delta}$.

Now consider the case $N \equiv 2$ or $3 \bmod 4$. If a form $(*, *, c)$ is such that $|c| < \sqrt[4]{\Delta}$ when c is odd, or $|c/2| < \sqrt[4]{\Delta}$ when c is even, then SQUFOF will enqueue this form. There are $2\sqrt[4]{\Delta}$ integers c such that $|c| < \sqrt[4]{\Delta}$, and only $3\sqrt[4]{\Delta}/2$ such that $4 \nmid c$ as well. Of this latter quantity, $\sqrt[4]{\Delta}$ of these c are odd, and so $2c$ satisfies $|c| = |2c/2| < \sqrt[4]{\Delta}$. So there are $5\sqrt[4]{\Delta}/2$ integers c such that $|c| < \sqrt[4]{\Delta}$ when c is odd, and $|c/2| < \sqrt[4]{\Delta}$ when c is even. Finally, there are $3\sqrt{\Delta}/2$ integers c with $|c| < \sqrt{\Delta}$ and $4 \nmid c$. If we chose random integers c not divisible by 4 with uniform distribution from the interval $(-\sqrt{\Delta}, \sqrt{\Delta})$, then the fraction of them such that c or $c/2$ is in $(-\sqrt[4]{\Delta}, \sqrt[4]{\Delta})$ is $(5\sqrt[4]{\Delta}/2)/(3\sqrt{\Delta}/2) = 5/(3\sqrt[4]{\Delta})$. We will assume that the same fraction of actual numbers c that arise when SQUFOF is used to factor N satisfy this inequality. We summarize our assumptions this way.

Assumption 4.23. *Let W be the number of forms examined and $Q = Q(N)$ be the number of forms enqueued during the factorization of N. Then $A[Q]/A[W]$, the average fraction of the examined forms that are enqueued, is either $1/\sqrt[4]{\Delta}$ or $5/(3\sqrt[4]{\Delta})$, according as $N \equiv 1 \bmod 4$ or $N \equiv 2$ or $3 \bmod 4$.*

Theorem 4.24. *As $N \to \infty$, the asymptotic average value of Q is*

$$A[Q] \sim \begin{cases} \dfrac{2\left(\sqrt{2}+1\right)\log 2}{2^k - 2} & \text{if } N \equiv 1 \bmod 4\ , \\[2ex] \dfrac{5\left(\sqrt{2}+1\right)\log 2}{2\left(2^k - 2\right)} & \text{if } N \equiv 2 \text{ or } 3 \bmod 4\ . \end{cases}$$

Proof. **Case 1:** ($N \equiv 1 \bmod 4$.) Since $\Delta = N$, we have by Assumption 4.23,

$$A[Q] = A[W]/\sqrt[4]{N} = \left(\frac{2\left(\sqrt{2}+1\right)\sqrt[4]{N}\log 2}{2^k - 2}\right) \Big/ \sqrt[4]{N} = \frac{2\left(\sqrt{2}+1\right)\log 2}{2^k - 2}\ .$$

k	$A[Q],\ N \equiv 1 \bmod 4$	$A[Q],\ N \equiv 2 \text{ or } 3 \bmod 4$
2	1.6734	2.0918
3	0.5578	0.6973
4	0.2391	0.2988

TABLE 4. Estimates of $A[Q]$ for $k = 2, 3, 4$.

Case 2: ($N \equiv 2$ or $3 \bmod 4$.) Since $\Delta = 4N$, we have by Assumption 4.23,

$$A[Q] = (5A[W]) \,/\, (3\sqrt[4]{4N}) = \left(\frac{5 \cdot 3\left(\sqrt{2}+2\right)\sqrt[4]{N}\log 2}{2\left(2^k-2\right)}\right) \Big/ \left(3\sqrt[4]{4N}\right)$$

$$= \frac{5\left(\sqrt{2}+1\right)\log 2}{2\left(2^k-2\right)}\ .$$

□

Table 4 lists the predicted values for $A[Q]$ when N is a product of two, three, and four primes.

5. The Effect of Multipliers

We now consider how multiplying N by small odd primes changes the running time of SQUFOF and the queue length. Our strategy will be similar to that of Section 4 in that we will compute $A[X]$ and $A[X_{sq}]$ for $p_1 p_2 \cdots p_n N$ for distinct small odd primes p_i.

5.1. The Time Complexity with Multipliers.

Proposition 5.1. *Let N be a square-free positive integer with k distinct large odd prime divisors and let $p_1, \ldots, p_n$ be n distinct small odd primes ($n \geq 0$) with $p_i \nmid N$ for all i. Define Δ by*

$$\Delta = \begin{cases} p_1 \cdots p_n N & \text{if} \quad p_1 \cdots p_n N \equiv 1 \bmod 4\ , \\ 4p_1 \cdots p_n N & \text{if} \quad p_1 \cdots p_n N \equiv 2 \text{ or } 3 \bmod 4\ . \end{cases}$$

If X is the number of reduced forms on the principal cycle of discriminant Δ then, as $N \to \infty$, the asymptotic average value of X is

$$A[X] \sim \begin{cases} \dfrac{\left(2^{k+n}+1\right)\sqrt{N}\log 2}{2^{k+n}h^+} \displaystyle\prod_{i=1}^{n} \frac{p_i^2-1}{p_i^{3/2}} & \text{if} \quad \Delta \equiv 1 \bmod 4 \text{ and} \\ & \qquad p_i \equiv 1 \bmod 4\ \forall i\ , \\ \dfrac{\sqrt{N}\log 2}{h^+} \displaystyle\prod_{i=1}^{n} \frac{p_i^2-1}{p_i^{3/2}} & \text{if} \quad \Delta \equiv 1 \bmod 4 \text{ and} \\ & \qquad \exists p_i \equiv 3 \bmod 4\ , \\ \dfrac{3\left(2^{k+n+1}+1\right)\sqrt{N}\log 2}{2^{k+n+2}h^+} \displaystyle\prod_{i=1}^{n} \frac{p_i^2-1}{p_i^{3/2}} & \text{if} \quad N \equiv 2 \bmod 4 \text{ and} \\ & \qquad p_i \equiv 1 \bmod 4\ \forall i\ , \\ \dfrac{3\sqrt{N}\log 2}{2h^+} \displaystyle\prod_{i=1}^{n} \frac{p_i^2-1}{p_i^{3/2}} & \text{otherwise.} \end{cases}$$

Proof. The proof of each case is similar to the corresponding proof in Proposition 4.14. The main difference is that we will need Lemma 4.2 to handle several small ramified primes. □

Proposition 5.2. *Let N be a square-free positive integer with k distinct large odd prime divisors and let $p_1, \ldots, p_n$ be n distinct small odd primes ($n \geq 0$) with $p_i \nmid N$ for all i. Define Δ as in Proposition 5.1. If X_{sq} is the number of reduced square forms on the principal cycle of discriminant Δ then, as $N \to \infty$, the asymptotic average value of X_{sq} is*

$$A[X_{sq}] \sim \begin{cases} \dfrac{\left(2^{k+n}+1\right)\left(\sqrt{2}-1\right)\sqrt[4]{N}}{2h^+}\displaystyle\prod_{i=1}^{n}\frac{p_i-1}{p_i^{3/4}} & \text{if} \quad \Delta \equiv 1 \bmod 4 \text{ and} \\ & \quad p_i \equiv 1 \bmod 4 \ \forall i\ , \\ \dfrac{2^{k+n}\left(\sqrt{2}-1\right)\sqrt[4]{N}}{2h^+}\displaystyle\prod_{i=1}^{n}\frac{p_i-1}{p_i^{3/4}} & \text{if} \quad \Delta \equiv 1 \bmod 4 \text{ and} \\ & \quad \exists p_i \equiv 3 \bmod 4\ , \\ \dfrac{\left(2^{k+n+1}+1\right)\left(2-\sqrt{2}\right)\sqrt[4]{N}}{4h^+}\displaystyle\prod_{i=1}^{n}\frac{p_i-1}{p_i^{3/4}} & \text{if} \quad N \equiv 2 \bmod 4 \text{ and} \\ & \quad p_i \equiv 1 \bmod 4 \ \forall i\ , \\ \dfrac{2^{k+n}\left(2-\sqrt{2}\right)\sqrt[4]{N}}{2h^+}\displaystyle\prod_{i=1}^{n}\frac{p_i-1}{p_i^{3/4}} & \text{otherwise.} \end{cases}$$

Proof. The proof of each case is similar to the corresponding proof in Proposition 4.16. The main difference is that we will need Lemmas 4.3 and 4.4 to handle several small ramified primes. □

Corollary 5.3. *Let N be a square-free positive integer with k distinct large odd prime divisors and let $p_1, \ldots, p_n$ be n distinct small odd primes ($n \geq 0$) with $p_i \nmid N$ for all i. Define Δ as in Proposition 5.1. If D is the index-difference between successive square forms on the principal cycle, then, as $N \to \infty$, the asymptotic average value of D is*

$$A[D] \sim \begin{cases} \dfrac{\left(\sqrt{2}+1\right)\sqrt[4]{N}\log 2}{2^{k-1}}\displaystyle\prod_{i=1}^{n}\frac{p_i+1}{2p_i^{3/4}} & \text{if} \quad \Delta \equiv 1 \bmod 4\ , \\ \dfrac{3\left(\sqrt{2}+2\right)\sqrt[4]{N}\log 2}{2^{k+1}}\displaystyle\prod_{i=1}^{n}\frac{p_i+1}{2p_i^{3/4}} & \text{if} \quad \Delta \equiv 0 \bmod 4\ . \end{cases}$$

Proof. Just as in Corollary 4.18, we obtain the result by computing $A[X]/A[X_{sq}]$. □

Theorem 5.4. *Let N be a square-free positive integer with k distinct large odd prime divisors and let $p_1, \ldots, p_n$ be n distinct small odd primes ($n \geq 0$) with $p_i \nmid N$ for all i. Define Δ as in Proposition 5.1. If W is the number of forms that SQUFOF must examine before finding a proper square form, then, as $N \to \infty$, the asymptotic* *average value of* W *is*

$$A[W] \sim \begin{cases} \dfrac{2\left(\sqrt{2}+1\right)\sqrt[4]{N}\log 2}{2^k-2}\displaystyle\prod_{i=1}^{n}\frac{p_i+1}{2p_i^{3/4}} & \text{if} \quad \Delta \equiv 1 \bmod 4\ , \\ \\ \dfrac{3\left(\sqrt{2}+2\right)\sqrt[4]{N}\log 2}{2\left(2^k-2\right)}\displaystyle\prod_{i=1}^{n}\frac{p_i+1}{2p_i^{3/4}} & \text{if} \quad \Delta \equiv 0 \bmod 4\ . \end{cases}$$

Proof. As in Theorem 4.22, this is the product of the results from Corollaries 4.21 (which allows multipliers) and 5.3. □

5.2. **Using the Queue with Multipliers.** We begin with propositions analogous to Propositions 3.1 and 3.2.

Proposition 5.5. *Let* N *be a square-free positive integer with* k *distinct large odd prime divisors and let* $p_1, \ldots, p_n$ *be* n *distinct small odd primes* ($n \geq 0$) *with* $p_i \nmid N$ *for all* i*. Define* Δ *as in Proposition 5.1. Suppose that* a *is a positive odd integer,* b *is a positive integer,* $\gcd(a,b) = 1$*, and that* $\left(a^2, b, -c\right)$ *is a square form on the principal cycle of discriminant* Δ *with* $c > 0$*. Then* $(-a, b, ac)^2 \sim \left(a^2, b, -c\right)$*.*

Proof. This follows directly from the definition of composition. □

There are 2^κ reduced ambiguous forms of discriminant Δ. The forms $(\pm d, *, *)$, where d is a square-free divisor of Δ relatively prime to N with $|d| < \sqrt{\Delta}$, lead to trivial factorizations of N. Let $\pm\mathbf{d}$ denote the reduced ambiguous form $(\pm d, *, *)$.

Proposition 5.6. *Let* N *be a square-free positive integer with* k *distinct large odd prime divisors and let* $p_1, \ldots, p_n$ *be* n *distinct small odd primes* ($n \geq 0$) *with* $p_i \nmid N$ *for all* i*. Define* Δ *as in Proposition 5.1, and define* $\mu = p_1 \cdots p_n$ *if* $\Delta \equiv 1 \bmod 4$ *and* $\mu = 2p_1 \cdots p_n$ *if* $\Delta \equiv 0 \bmod 4$*. Assume that the* p_i *are chosen so that* $\mu^{3/4} < \sqrt[4]{N}$ *when* $\Delta \equiv 1 \bmod 4$*, or* $\sqrt{2}\mu^{3/4} < \sqrt[4]{N}$ *when* $\Delta \equiv 0 \bmod 4$*. Suppose that* a *is a positive odd integer,* b *is a positive integer,* $\gcd(a,b) = 1$*, and that* $F_n = \left(a^2, b, -c\right)$ *is a square form on the principal cycle of discriminant* Δ*, with* $c > 0$*. Some form* $(\alpha, \beta, *)$ *appears on the principal cycle at position* $m < n$ *with* $\alpha \in \{\pm da\}$*, where* d *is a square-free divisor of* Δ *that is relatively prime to* N *with* $|d| < \sqrt{\Delta}$*, and* $\beta \equiv b \bmod a$ *if and only if* $(-a, b, ac)$ *is equivalent to one of the ambiguous forms* $\pm\mathbf{d}$*.*

Proof. The largest element in the set $\{\pm da\}$ is μa. The inequality on μ in the hypothesis insures that each element of $\{\pm da\}$ actually appears as an end coefficient for some reduced form. To see this, suppose that $a^2 < \sqrt{\Delta}$ and we want $\mu a < \sqrt{\Delta}$, too, so that the square root can appear as a reduced form in some ambiguous cycle. Since $a < \sqrt[4]{\Delta}$, we can insure that $\mu a < \sqrt{\Delta}$ by taking $\mu < \sqrt[4]{\Delta}$. Using the definitions of μ and Δ, we obtain the inequalities on μ in the hypotheses.

Suppose now that some form $(\alpha, \beta, *)$ appears on the principal cycle at position $m < n$ with $\alpha = da$, where d is a square-free divisor of Δ that is relatively prime to N with $|d| < \sqrt{\Delta}$, and $\beta \equiv b \bmod a$. Then we can write the form as $(\alpha, \beta, *) \sim (da, \beta, *)$. It is easy to see that $-\mathbf{d} \sim (da, \beta, *) \circ -\mathbf{d} \sim (-a, \beta, *) \sim (-a, b, ac)$.

Conversely, suppose there is no form $(\alpha, \beta, *)$ appearing on the principal cycle before F_n with $\alpha \in \{\pm da\}$ and $\beta \equiv b \bmod a$. Let $f = (-a, b, ac)$. If f is equivalent to some $g \in \{\pm\mathbf{d}\}$, then $f \circ g \sim \mathbf{1}$ and $f \circ g$ is equivalent to some form $(\alpha, \beta', *)$

with $\alpha \in \{\pm da\}$ and $\beta' \equiv b \bmod a$. But this square root is equivalent to a reduced square root $(\alpha, \beta, *)$, with $\alpha \in \{\pm da\}$ and $\beta \equiv \beta' \equiv b \bmod a$, that must be on the principal cycle. But then this reduced square root must appear before the form F_n, a contradiction. Therefore, f is not equivalent to any of the ambiguous forms $\pm\mathbf{d}$. □

Proposition 5.6 says that when we have several small ramified primes, the test for whether a form is enqueued or not is the following. First suppose $N \equiv 1 \bmod 4$. If a form $(*, b, c)$ is found such that $|c'| < \sqrt[4]{\Delta}$, where $c' = c/\gcd(c, p_1 \cdots p_n)$, then SQUFOF will enqueue the pair $(c', b \bmod c')$. If $N \equiv 2$ or $3 \bmod 4$, then the additional ramified prime 2 means that we should take $c' = c/\gcd(c, 2p_1 \cdots p_n)$.

We now describe the changes in the algorithm descriptions in Subsections 3.3 and 3.6 needed if a multiplier is used. First, the multiplier m should be a square-free product of small odd primes, certainly smaller than any prime factor of N. The multiplier should also be small enough to imply the inequalities on μ in Proposition 5.6.

Change Step 1 of the binary quadratic forms version to this:

Read the odd positive integer N to be factored. If N is the square of an integer, output the square root and stop. If $mN \equiv 1 \bmod 4$, then set $D \leftarrow mN$, $m' \leftarrow m$, $d \leftarrow \left\lfloor \sqrt{D} \right\rfloor$, and $b \leftarrow 2\lfloor (d-1)/2 \rfloor + 1$. Otherwise ($N \equiv 2$ or $3 \bmod 4$), set $D \leftarrow 4mN$, $m' \leftarrow 2m$, $d \leftarrow \left\lfloor \sqrt{D} \right\rfloor$, and $b \leftarrow 2\lfloor d/2 \rfloor$. Let $F \leftarrow (1, b, (b^2 - D)/4)$, $i \leftarrow 2$, $L \leftarrow \left\lfloor \sqrt{d} \right\rfloor$, and $Bound \leftarrow 4 \cdot L$. Create an empty list. Let $g \leftarrow |(b^2 - D)/4|/\gcd(|(b^2 - D)/4|, m')$. If $g \le L$, add g to the list. (To use a queue, put $(g, b \bmod g)$ onto the QUEUE here.)

In Step 2c, change m to m'. (For a queue, put $(g, B \bmod g)$ onto the QUEUE instead.)

Change Step 1 of the continued fraction version to this:

Read the odd positive integer N to be factored. If N is the square of an integer, output the square root and stop. If $mN \equiv 1 \bmod 4$, then set $D \leftarrow 2mN$; otherwise, set $D \leftarrow mN$. In any case, set $S \leftarrow \left\lfloor \sqrt{D} \right\rfloor$, $\hat{Q} \leftarrow 1$, $P \leftarrow S$, $Q \leftarrow D - P \cdot P$, $L \leftarrow \left\lfloor \sqrt{2\sqrt{D}} \right\rfloor$, $B \leftarrow 2 \cdot L$, and $i \leftarrow 0$.

In the continued fraction version of SQUFOF, change Step 2b to:

Let $g \leftarrow Q/\gcd(Q, 2m)$. If $g \le L$, put $(g, B \bmod g)$ onto the QUEUE.

In Step 5 of the continued fraction version, replace Q by $Q/\gcd(Q, 2m)$ before the factor Q is output. Likewise, any common factor of $|c|$ and m' is divided out from $|c|$ before the factor $|c|$ is written in the binary quadratic forms version.

We now turn to the task of computing the average number of forms enqueued in terms of N and the p_i. The numbers $|c'|$ in the next proposition are the numbers g enqueued in Step 2b of the continued fraction version of SQUFOF.

Proposition 5.7. *Let N, Δ, $p_1, \ldots, p_n$, and μ as in Proposition 5.6. Suppose that $\mu^{3/4} < \sqrt[4]{N}$ when $\Delta \equiv 1 \bmod 4$, or $\sqrt{2}\mu^{3/4} < \sqrt[4]{N}$ when $\Delta \equiv 0 \bmod 4$. Let $\mathcal{S}$ be the set of integers c with*

(1) *$p_i^2 \nmid c$ for $i = 1, 2, \ldots, n$, and either*

(2) *$|c| < \sqrt[4]{\Delta}$, or*

(3) *if* $|c| > \sqrt[4]{\Delta}$, *then* $|c'| < \sqrt[4]{\Delta}$, *where* $c' = c/\gcd(p_1 \cdots p_n, c)$ *when* $\Delta \equiv 1 \bmod 4$, *and* $c' = c/\gcd(2p_1 \cdots p_n, c)$ *when* $\Delta \equiv 0 \bmod 4$.

Define $K_{\mathcal{S}} = K_{\mathcal{T}} = 1$ *when* $\Delta \equiv 1 \bmod 4$, *and* $K_{\mathcal{S}} = 5/3$ *and* $K_{\mathcal{T}} = 3/4$ *when* $\Delta \equiv 0 \bmod 4$. *Then, as* $N \to \infty$,

$$|\mathcal{S}| \sim |\mathcal{T}|\, K_{\mathcal{S}} \prod_{i=1}^{n} \frac{2p_i + 1}{p_i + 1} \,,$$

where

$$|\mathcal{T}| \sim 2\sqrt[4]{\Delta}\, K_{\mathcal{T}} \prod_{i=1}^{n} \frac{p_i^2 - 1}{p_i^2} \,, \tag{5.1}$$

and $\mathcal{T}$ *is the set of integers satisfying only the first two conditions above.*

Proof. We prove the theorem just for the case $\Delta \equiv 1 \bmod 4$. The only difference in the other case, when $\Delta \equiv 0 \bmod 4$, is that the prime 2 is ramified and behaves like the odd primes p_i.

The asymptotic behavior of the cardinality of the set $\mathcal{T}$ is clear, as is the behavior of the cardinality of the set $\mathcal{S}$ when $n = 0$. We prove the asymptotic behavior of the cardinality of the set $\mathcal{S}$ for $n > 0$ by induction on n.

Suppose $n = 1$. Of the integers in $\mathcal{T}$, the subset of integers c divisible by p_1 (given that c is not divisible by p_1^2) has size $|\mathcal{T}|/(p_1 + 1)$. None of these numbers can also appear as a c' for some $c > \sqrt[4]{\Delta}$. The rest of the integers in the set $\mathcal{T}$ are not divisible by p_1, so if we multiply each of these by p_1, we get a new set of integers that must satisfy Conditions (1) and (3). Thus

$$|\mathcal{S}| \sim |\mathcal{T}| \frac{1}{p_1 + 1} + 2\,|\mathcal{T}| \frac{p_1}{p_1 + 1} = |\mathcal{T}| \frac{2p_1 + 1}{p_1 + 1} \,.$$

Now suppose that the claim holds for any choice of $n - 1$ primes and consider the case of n primes $p_1, \ldots, p_n$. Again, it is easy to see that $|\mathcal{T}|$ is given by Formula (5.1). The subset of these numbers divisible by p_n has cardinality $|\mathcal{T}|/(p_n + 1)$. Each of these can be multiplied by some square-free product (perhaps trivial) of only the $p_1, \ldots, p_{n-1}$. By the induction hypothesis, this subset leads to

$$|\mathcal{T}| \frac{1}{p_n + 1} \prod_{i=1}^{n-1} \frac{2p_i + 1}{p_i + 1}$$

integers that satisfy either (1) and (2), or (1) and (3). The rest of the integers in $\mathcal{T}$ are not divisible by p_n. Again, by the induction hypothesis, this subset leads to

$$2\,|\mathcal{T}| \frac{p_n}{p_n + 1} \prod_{i=1}^{n-1} \frac{2p_i + 1}{p_i + 1} \,,$$

where we first count those numbers that we get by multiplying by some square-free product (perhaps trivial) of the $p_1, \ldots, p_{n-1}$, then we double our count since for each of these we may multiply by either 1 or p_n. Finally, $|\mathcal{S}|$, the total number of integers that satisfy either (1) and (2), or (1) and (3) is

$$|\mathcal{S}| \sim |\mathcal{T}| \frac{1}{p_n + 1} \prod_{i=1}^{n-1} \frac{2p_i + 1}{p_i + 1} + 2\,|\mathcal{T}| \frac{p_n}{p_n + 1} \prod_{i=1}^{n-1} \frac{2p_i + 1}{p_i + 1} = |\mathcal{T}| \prod_{i=1}^{n} \frac{2p_i + 1}{p_i + 1} \,.$$

Thus the claim holds for all $n \geq 0$. □

Theorem 5.8. *Let N be a square-free positive integer with k distinct large odd prime divisors and let $p_1, \ldots, p_n$ be n distinct small odd primes ($n \geq 0$) with $p_i \nmid N$ for all i. Define Δ as in Proposition 5.1 and μ as in Proposition 5.6. Assume that the p_i are chosen so that $\mu^{3/4} < \sqrt[4]{N}$ when $\Delta \equiv 1 \bmod 4$, or $\sqrt{2}\mu^{3/4} < \sqrt[4]{N}$ when $\Delta \equiv 0 \bmod 4$. If Q is the number of forms that SQUFOF enqueues before finding a proper square form, then, as $N \to \infty$, the asymptotic average value of Q is*

$$A[Q] \sim \begin{cases} \dfrac{2\left(\sqrt{2}+1\right)\log 2}{2^k-2} \displaystyle\prod_{i=1}^{n} \frac{2p_i+1}{2p_i} & \textit{if} \quad \Delta \equiv 1 \bmod 4\ , \\[2ex] \dfrac{5\left(\sqrt{2}+1\right)\log 2}{2\left(2^k-2\right)} \displaystyle\prod_{i=1}^{n} \frac{2p_i+1}{2p_i} & \textit{if} \quad \Delta \equiv 0 \bmod 4\ . \end{cases}$$

Proof. The size $|\mathcal{S}|$ in Proposition 5.7 is the number of end coefficients that will lead to a form being enqueued. Let Condition (2)$'$ denote Condition (2) of Proposition 5.7 with $\sqrt[4]{\Delta}$ replaced with $\sqrt{\Delta}$, and let $\mathcal{T}'$ be the set of integers c satisfying Conditions (1) and (2)$'$. Then, as in the proof of Proposition 5.7,

$$|\mathcal{T}'| \ \sim\ 2\sqrt{\Delta}\, K_{\mathcal{T}} \prod_{i=1}^{n} \frac{p_i^2-1}{p_i^2}\ .$$

The number of end coefficients c with $|c| < \sqrt{\Delta}$ and c not divisible by the square of any ramified prime is given by $|\mathcal{T}'|$. We take the ratio $|\mathcal{S}|/|\mathcal{T}'|$ to be the fraction of forms enqueued. Finally, we take the product of this number with $A[W]$ to get $A[Q]$, as in the proof of Theorem 4.24. □

5.3. **Optimal Multipliers for SQUFOF.** Recall that the continued fraction version of SQUFOF always works with $N \equiv 2$ or $3 \bmod 4$, multiplying N by 2 whenever $mN \equiv 1 \bmod 4$, where m is the odd multiplier. Notice then that if the multiplier m is the product of the odd primes p_1, …, p_n, the integer $p_1 \cdots p_n N \equiv 2$ or $3 \bmod 4$, where N has a factor of 2 if and only if $p_1 \cdots p_n N \equiv 1 \bmod 4$ before the factor of 2 was put in. In this situation the asymptotic average number of forms examined to find a proper square form, as $N \to \infty$, is

$$A[W] = \frac{3\left(\sqrt{2}+2\right)\sqrt[4]{N}\log 2}{2\left(2^k-2\right)} \prod_{i=1}^{n} \frac{p_i+1}{2p_i^{3/4}}\ .$$

We seek p_i that minimize this quantity, the last product being the factor by which SQUFOF factorization of $p_1 \cdots p_n N$ is faster or slower than that of N.

Theorem 5.9. *Let Ω be the set of all finite sets of distinct odd primes and define the mapping $F : \Omega \to \mathbb{Z}$ by $F(\emptyset) = 1$ and*

$$F\left(\{p_1, \ldots, p_n\}\right) = \prod_{i=1}^{n} \frac{p_i+1}{2p_i^{3/4}}\ .$$

Then F is minimized at the set $\{3,5,7,11\}$ and

$$F\left(\{3,5,7,11\}\right) \approx 0.7268\ .$$

Proof. It is easy to check that $F\left(\{3,5,7,11\}\right) \approx 0.7268$. We will show that for any other finite set of odd primes $\{p_1, \ldots, p_n\}$, we will have

$$F\left(\{p_1, \ldots, p_n\}\right) > F\left(\{3,5,7,11\}\right)\ ,$$

$p_1 \cdots p_n$	$F(\{p_1,\ldots,p_n\})$	$G(\{p_1,\ldots,p_n\})$
3	0.8774	1.1667
5	0.8972	1.1000
7	0.9295	1.0714
11	0.9934	1.0455
$3 \cdot 5$	0.7872	1.2833
$3 \cdot 7$	0.8155	1.2500
$3 \cdot 11$	0.8716	1.2197
$5 \cdot 7$	0.8339	1.1786
$5 \cdot 11$	0.8913	1.1500
$7 \cdot 11$	0.9233	1.1201
$3 \cdot 5 \cdot 7$	0.7317	1.3750
$3 \cdot 5 \cdot 11$	0.7820	1.3417
$3 \cdot 7 \cdot 11$	0.8101	1.3068
$5 \cdot 7 \cdot 11$	0.8284	1.2321
$3 \cdot 5 \cdot 7 \cdot 11$	0.7268	1.4375

Table 5. Good candidate multipliers for $N \equiv 2$ or 3 mod 4.

which will prove the claim. So suppose by way of contradiction that there exists a finite set of odd primes $\{p_1,\ldots,p_n\}$ such that $F(\{p_1,\ldots,p_n\}) < F(\{3,5,7,11\})$. Since $F(\emptyset) = 1$, F is not minimized at $\emptyset$ and so $n > 0$.

It is easy to check that the function $f(x) = (x+1)/2x^{3/4}$ is strictly increasing on $[3,\infty)$ and so for a given n, among all sets of n primes, F is minimized at $\{3,5,7,\ldots,p_n\}$, where p_n is the n^{th} odd prime. Straightforward computation shows that for sets of n primes with $n = 1,2,3,4$, F is minimized at $\{3,5,7,11\}$. Finally, one easily sees that $(x+1)/2x^{3/4} > 1$ for $x \geq 13$. This means that adding any additional primes to the set $\{3,5,7,11\}$ will increase the value of F at this new set. Therefore, F is minimized at the set $\{3,5,7,11\}$. □

Theorem 5.9 shows that the optimal multiplier is $3 \cdot 5 \cdot 7 \cdot 11 = 1155$, and that in fact we can expect SQUFOF to find a non-trivial factor of N using $1155N$ in about 73% of the time that it would take using N. However, for practical reasons associated with the size of single precision integers, SQUFOF may actually run faster for smaller multipliers.

Let F be defined as in Theorem 5.9 and let $G(\{p_1,\ldots,p_n\}) = \prod_{i=1}^{n} \frac{2p_i+1}{2p_i}$ be the factor by which the number of forms enqueued is larger when factoring $p_1 \cdots p_n N$ than when factoring N. Table 5 lists some good candidate multipliers, along with the associated values of F and G. Note that for the values of $p_1 \cdots p_n$ considered in Table 5, the value of G is no larger than 1.5. In other words, at worst we can expect a 50% increase in the number of forms enqueued when using one of these multipliers. However, the number of forms enqueued without using a multiplier is very small—about 2.1 forms. So even though the rules for enqueuing a form are more complicated (hence more time consuming) when using multipliers, this average time cost is negligible compared with the average time savings.

5.4. **Racing SQUFOF with Multipliers.** The original reason for using multipliers was to exploit the great variation in $W/\sqrt[4]{N}$, where W is the actual number of forms that SQUFOF must examine before finding a proper square form. Racing several multipliers succeeds when the first proper square form is found, which may appear early for at least one of the multiples of N. The results of the previous subsection suggest that if we choose the multipliers wisely, we can expect the proper square form to come quickly for one multiplier, so that the total work is less.

Our experiments suggest that $A[W]/\sqrt[4]{N}$ behaves like a random variable with an exponential distribution, since we find its mean and variance to be approximately the same, and the exponential distribution is the only common one with this property. Furthermore, the chance that any form in the principal period is a proper square form seems to be independent of whether any forms seen earlier in the period are proper square forms. This lack of memory is another characteristic of the exponential distribution. We conjecture that $A[W]/\sqrt[4]{N}$ does have an exponential distribution.

To illustrate an implication of this conjecture, let N be a product of two primes. Let $m_1, \dots, m_s$ be distinct multipliers. Let $A[W_i]$ be the average number of forms that SQUFOF on m_iN must examine to find the first proper square form. Let $A[W_1, \dots, W_s] = s \cdot A[\min(W_1, \dots, W_s)]$ be the average total number of forms that SQUFOF must examine to find the first proper square form when racing the multipliers $m_1N, \dots, m_sN$. Let

$$H(x_1, \dots, x_s) = \left(\frac{1}{s} \sum_{i=1}^{s} \frac{1}{x_i} \right)^{-1}$$

be the harmonic mean of $x_1, \dots, x_s$. Then an exponential distribution for $A[W]/\sqrt[4]{N}$ would imply that

$$A[W_1, \dots, W_s]/\sqrt[4]{N} = H(A[W_1]/\sqrt[4]{N}, \dots, A[W_s]/\sqrt[4]{N}).$$

Our experimental data supports this formula. The minimum of several exponential random variables is again an exponential random variable, which is probably why we see this behavior when racing multipliers.

For example, our experiments for racing the multipliers $m_1 = 11$ and $m_2 = 3\cdot5\cdot7$ give the value 1.4687 for $A[W_1, W_2]/\sqrt[4]{N}$ averaged over thousands of different N. The theory predicts that $A[W_1]/\sqrt[4]{N} = 0.9934 \cdot A[W]/\sqrt[4]{N}$ and $A[W_2]/\sqrt[4]{N} = 0.7317 \cdot A[W]/\sqrt[4]{N}$. Hence, the conjecture would predict that

$$\begin{aligned} A[W_1, W_2]/\sqrt[4]{N} &= H(0.9934 \cdot A[W]/\sqrt[4]{N}, 0.7317 \cdot A[W]/\sqrt[4]{N}) \\ &= H(0.9934, 0.7317) \cdot A[W]/\sqrt[4]{N} \\ &= 0.8427 \cdot A[W]/\sqrt[4]{N}. \end{aligned}$$

The theory predicts that $A[W]/\sqrt[4]{N} = 1.7749$, so we multiply 0.8427 by 1.7749 to get 1.4957, which is close to the observed value 1.4687 from the experiment. Many other examples from [5] give similar approximate confirmation of the conjecture.

6. Experimental Results

To test the conclusions of the heuristic arguments in the preceding two sections, we factored hundreds of thousands of integers N with SQUFOF. These numbers were all square-free with two, three or four prime factors. About one-third of the numbers had each number of prime factors. The size of N in our experiments ranged from about 10^9 to about 10^{15}, with a few larger N.

We used both the continued fraction and the binary quadratic forms versions of SQUFOF. Half of the numbers were $\equiv 1 \bmod 4$ and half were $\equiv 3 \bmod 4$. We used each one of the 16 divisors of $1155 = 3 \cdot 5 \cdot 7 \cdot 11$ as a multiplier m. Whenever $mN \equiv 1 \bmod 4$ and we were using the continued fraction version of SQUFOF, we factored $2mN$, as specified in the heuristic argument. But we also tried to factor mN directly in these cases, even though the results of this paper do not apply there.

For each pair of multipliers $m_1 \neq m_2$, we raced SQUFOF on m_1N and m_2N until the first proper square form was found.

The queue in these experiments had space for 50 entries. It never overflowed.

In a small number of cases, SQUFOF could not factor N because the principal period contained no proper square forms. This happened less than 1% of the time for N near 10^9 or 10^{10} and even less frequently for N near 10^{15}.

For each number N successfully factored, we noted the number W of forms examined before finding the first proper square form and the total number Q of entries into the queue. For each case (number of prime factors of N, whether $N \equiv 1$ or $3 \bmod 4$, version of SQUFOF used, multiplier m or pair m_1, m_2 or racing multipliers), we computed the mean and standard deviation of $W/\sqrt[4]{N}$ and Q.

In every case the mean of $W/\sqrt[4]{N}$ was close to the value predicted by Theorems 4.22 and 5.4. Also, the mean of Q was close to the prediction given in Theorems 4.24 and 5.8. In general, the experimental average values were closer to the theoretical predictions for larger N than for smaller N.

In all of our experiments, the standard deviation of $W/\sqrt[4]{N}$ was close to the mean for that statistic, which supports the hypothesis of an exponential distribution for the random variable.

Table 6 gives a tiny sample of the extensive tables we generated. Each line in it gives the results of factoring 40,000 values of N, each the product of two primes near 1,000,000. Each N was factored using each multiplier in Table 5.

Let FWRD be the number of forms of discriminant Δ that SQUFOF examines before finding a proper square form, divided by the fourth root of N. Let QUEUE be the total number of forms that SQUFOF enqueues during the search for a proper square form. We computed FWRD and QUEUE for each successful factorization. We then computed the average values $\overline{\text{FWRD}}$ and $\overline{\text{QUEUE}}$, along with the standard deviations $\sigma\,(\text{FWRD})$ and $\sigma\,(\text{QUEUE})$. We also the computed the maximum and minimum value for FWRD and QUEUE, which gives the inequalities: $0.0008 \leq \text{FWRD} \leq 34.4793$ and $0 \leq \text{QUEUE} \leq 49$. Table 6 compares the predicted and calculated values for FWRD and QUEUE.

7. Future Work

We conclude with some questions for further study.

1.) Non-fundamental discriminants: SQUFOF appears to work for non-fundamental discriminants Δ. We have factored millions of $N \equiv 1 \bmod 4$ without multiplying by 2. The success rate was more than 99%, just as for $N \equiv 3 \bmod 4$. In addition, we have factored tens of thousands of N having an odd square factor > 8 with a similar success rate. We believe that an analysis similar to that in this paper will yield the same time and average number of forms enqueued for such Δ. In future work we will re-examine the points where we assume Δ to be a fundamental discriminant.

m	$\frac{A[W]}{\sqrt[4]{N}}$	$\overline{\text{FWRD}}$	$\sigma\,(\text{FWRD})$	$A[Q]$	$\overline{\text{QUEUE}}$	$\sigma\,(\text{QUEUE})$	failures
1	1.7749	1.7587	1.7570	2.0918	2.4009	2.9398	111
3	1.5573	1.5294	1.5144	2.4404	2.3750	2.8554	103
5	1.5925	1.5822	1.5799	2.3009	2.3970	2.9107	74
7	1.6497	1.6193	1.6254	2.2412	2.3850	2.9285	102
11	1.7631	1.7460	1.7536	2.1868	2.3848	2.9158	49
15	1.3972	1.3744	1.3829	2.6844	2.3602	2.8420	68
21	1.4474	1.4273	1.4186	2.6147	2.3733	2.8655	78
33	1.5469	1.5421	1.5326	2.5513	2.3996	2.8646	84
35	1.4802	1.4664	1.4652	2.4653	2.4034	2.8916	51
55	1.5819	1.5616	1.5556	2.4055	2.3789	2.8545	52
77	1.6388	1.6304	1.6282	2.3430	2.4089	2.9155	55
105	1.2987	1.2714	1.2747	2.8762	2.3666	2.8318	43
165	1.3879	1.3773	1.3791	2.8064	2.3800	2.8723	55
231	1.4378	1.4243	1.4332	2.7335	2.3767	2.8659	47
385	1.4703	1.4565	1.4542	2.5773	2.4006	2.8811	50
1155	1.2900	1.2770	1.2766	3.0069	2.3897	2.8553	55

Table 6. Two-prime statistics for FWRD and QUEUE.

2.) Distributions of $A[W]$ and $A[Q]$: Our experiments suggest that $A[W]/\sqrt[4]{N}$ may be a random variable with an exponential distribution, since we observe its mean and variance to be approximately the same. In future work, we would like to prove this, and investigate the implications it holds for the distribution of $A[Q]$ and for racing several multipliers.

3.) Racing Multipliers: First, we would like to prove our experimental results for $A[W]$ for the case of racing multipliers. If we can do this, then we will be able to give a good estimate for $A[W_r]$, the average number of forms examined during a race between several multiples of N. We also hope to discover the distribution of $A[Q_r]$, the average number of forms enqueued during a race between several multiples of N. Also, given that there are several multipliers m such that we can expect to factor mN faster than we can expect to factor N, it may be worthwhile to race several multiples of N.

4.) Fast Return: Once we have found a proper square form and switched to the cycle of its inverse square root, we know approximately how many forms we must traverse to reach an ambiguous form with the factor of N. Namely, it is close to half the number of forms considered before we found the proper square form. Using a fast exponentiation algorithm with composition of forms, we can swiftly compute a reduced form in this neighborhood. If it is not ambiguous, then we can cycle through forms adjacent to this one in both directions until we find an ambiguous form. This device greatly reduces the time for Step 4 of the algorithm.

5.) 64-bit Architecture: Modern workstations which perform arithmetic on 64-bit integers allow SQUFOF to factor integers as large as about 36 digits using only single precision operations. SQUFOF would take a few seconds to a minute to factor a typical 36-digit integer on such a machine.

The elliptic curve algorithm would be a strong competitor to SQUFOF for a number of that size, and faster for factoring larger integers. The cross-over point should be investigated.

References

[1] J. Brillhart and M. A. Morrison. A Method of Factoring and the Factorization of F_7. *Mathematics of Computation*, 29:183–205, 1975.

[2] Duncan A. Buell. *Binary Quadratic Forms: Classical Theory and Modern Computations*. Springer-Verlag, 1989.

[3] Henri Cohen. *A Course in Computational Algebraic Number Theory*. Springer-Verlag, 1996.

[4] Harvey Cohn. *A Second Course in Number Theory*. John Wiley & Sons, Inc., 1962.

[5] Jason E. Gower. *Square form factorization*. PhD thesis, Purdue University, December 2004.

[6] David Joyner and Stephen McMath. `http://cadigweb.ew.usna.edu/~wdj/mcmath/`.

[7] A. Y. Khinchin. *Continued Fractions*. The University of Chicago Press, third edition, 1964.

[8] H. W. Lenstra, Jr. On the Calculation of Regulators and Class Numbers of Quadratic Fields. In J. V. Armitage, editor, *Journées Arithmétiques, 1980*, volume 56 of *Lecture Notes Series*, pages 123–150. London Mathematical Society, 1982.

[9] J. S. Milne. Algebraic Number Theory, 1998. available at `http://www.math.lsa.umich.edu/~jmilne`.

[10] Ivan Niven, Herbert S. Zuckerman, and Hugh L. Montgomery. *An Introduction to the Theory of Numbers*. John Wiley & Sons, Inc., fifth edition, 1991.

[11] Hans Riesel. *Prime Numbers and Computer Methods for Factorization*. Birkhaüser, second edition, 1994.

[12] D. Shanks. Analysis and Improvement of the Continued Fraction Method of Factorization. Abstract 720-10-43. *Amer. Math. Soc. Notices*, 22:A–68, 1975.

[13] Daniel Shanks. An Attempt to Factor $N = 1002742628021$. Manuscript, 3 pages, available at [6].

[14] Daniel Shanks. Notes for *Analysis and Improvement of the Continued Fraction Method of Factorization*. Manuscript, 15 pages, available at [6].

[15] Daniel Shanks. SQUFOF Notes. Manuscript, 30 pages, available at [6].

[16] Daniel Shanks. Class Number, a Theory of Factorization, and Genera. In *Proceedings of Symposia in Pure Mathematics*, volume 20, pages 415–440. American Mathematical Society, 1971.

[17] Daniel Shanks. The Infrastructure of a Real Quadratic Field and its Applications. In *Proceedings of the 1972 Number Theory Conference*, pages 217–224, Boulder, Colorado, 1972.

[18] Daniel Shanks. Five Number-Theoretic Algorithms. In *Proceedings of the Second Manitoba Conference on Numerical Mathematics*, pages 51–70. Utilitas Mathematica, 1973. Number VII in Congressus Numerantium.

[19] Samuel S. Wagstaff, Jr. *Cryptanalysis of Number Theoretic Ciphers*. CRC Press, 2003.

Institute for Mathematics and its Applications, University of Minnesota, 424 Lind Hall, 207 Church St. S.E., Minneapolis, MN 55455-0436

E-mail address: `gower@ima.umn.edu`

Center for Education and Research in Information Assurance and Security, and Department of Computer Science, Purdue University, West Lafayette, IN 47907

E-mail address: `ssw@cerias.purdue.edu`