Pace University
DigitalCommons@Pace

CSIS Technical Reports

Ivan G. Seidenberg School of Computer Science and Information Systems

8-1-2002

On The Theory of Integers Factorization

Nelson A. Carella
Pace University

Follow this and additional works at: http://digitalcommons.pace.edu/csis_tech_reports

Recommended Citation
Carella, Nelson A., "On The Theory of Integers Factorization" (2002). *CSIS Technical Reports.* Paper 4.
http://digitalcommons.pace.edu/csis_tech_reports/4

This Article is brought to you for free and open access by the Ivan G. Seidenberg School of Computer Science and Information Systems at DigitalCommons@Pace. It has been accepted for inclusion in CSIS Technical Reports by an authorized administrator of DigitalCommons@Pace. For more information, please contact rracelis@pace.edu.

SCHOOL OF COMPUTER SCIENCE AND INFORMATION SYSTEMS

TECHNICAL REPORT

Number 180, August 2002

On the Theory of Integers Factorization

Nelson A. Carella

PACE UNIVERSITY

Nelson A. Carella, a mathematician who hails from the City University of New York, adjuncts in the Information Systems Department at Pace University in Manhattan.

On the Theory of Integers Factorization

Nelson A Carella
Email pobox5000@email com

Abstract: This paper establishes the preliminary theoretical frameworks for a new theory of integer factorization The corresponding algorithms are capable of factoring any composite integers $n = pq$ from certain large subset S of integers of exponential size $\#S = O(n^{1/2-\varepsilon})$, $\varepsilon > 0$, in polynomial time A generalization of this technique and the existence of a polynomial time factoring algorithm are also considered

Key Words Integers Factorization, Polynomial Time Algorithms
Mathematics Subjects Classification 11Y05

1 Introduction

In the last few decades very extensive research and resources has been devoted to the factorization of large integers The specific case $n = pq$ and the equivalent problem of determining the value of $\varphi(n)$ is of considerable interest These efforts have resulted in incremental improvements in both the factorization algorithms and the hardware/software A detailed account of the factorization of a 512-bit integer (155 digits) appears in [CV], see also [BR], [LA2], [OD], and [PA] for perspectives on current methods and future projections

The method introduced here relies on fundamental mathematical concepts It combines elements of additive number theory and elliptic curves to derive a new integer factorization theory These preliminary results have the potential to develop into a general purpose factorization method The recent developments in computational Diophantine analysis make all the algorithms probably practical, and utilize moderate amount of resources The technique is applicable to the factorization of composite integers $n = pq$ from certain subset of integers $S = \{\ n = pq \quad (p, q) \in P_n \times Q_n$, and $f(p,q) = 0\ \}$, where $P_n = \{\ p \le \alpha\sqrt{n})\ \}$, $Q_n = \{\ p \le \beta\sqrt{n})\ \}$, $\alpha\beta = 1$, and some equation $f(p,q) = 0$ It will be shown that the set S has exponential size $\#S = O(n^{1/2-\varepsilon})$, $\varepsilon > 0$

A few of the most effective integers factorization methods, the $p-1$ method [PO], the $p+1$ method [WI], the rho method, and the elliptic curve method [LH], rely on the specific structures of the prime factors of n, i e , smoothness of $p \pm 1$ etc The method herein does not rely any specific structures of $p \pm 1$, nor the nearby integers are required to have specific properties The factors of the integers n arise as the solutions of certain additive problems

The fastest algorithm herein can resolve an integer from the set S into its prime factors in time polynomial in double logarithm of the input size

Complexity Issues

A problem is in the complexity class **P** if it can be solved on a sequential machine by a deterministic algorithm in polynomially many steps $O(\log^c(n))$ in the input size $O(\log n)$, where $c > 0$ is a constant independent of the input n Similarly, a problem is in the complexity class **NP** if it can be solved by a nondeterministic algorithm in polynomially many steps $O(\log^u(n))$ in the input size $O(\log n)$, where $u > 0$ is a variable depending on the input n

The expression $L_n[\alpha,\beta] = e^{\beta \ln(n)^\alpha (\ln\ln(n))^{1-\alpha}}$, $0 \le \alpha \le 1$, $\beta > 0$, is widely used to measure the expected running time complexities of algorithms as $n \to \infty$ As α varies between 0 and 1, β constant, it interpolates between polynomial time and exponential time The three instances of the parameter $\alpha = 1$, $0 < \alpha < 1$, and $\alpha = 0$ describe three major types of algorithms

(1) $L_n[1,\beta] = n^\beta$, exponential, (1)

(2) $L_n[0 < \alpha < 0,\beta] = e^{\beta \ln(n)^\alpha (\ln\ln(n))^{1-\alpha}}$, subexponential, and

(3) $L_n[0,\beta] = (\ln n)^\beta$, polynomial running times respectively

An algorithm of running time complexity $L_n[\alpha,\beta]$ is defined as subexponential if

$$\lim_{n\to\infty} \frac{L_n[\alpha,\beta]}{n^\varepsilon} = 0 \text{ for all } \varepsilon > 0 \qquad (2)$$

Another class of interest is the nearly polynomial time class An algorithm from this class has running time complexity of the form $O(\log(n)^{O(\log\log n)})$, or similar expression equivalent to $O(n^{o(1)})$ In practice this is very effective For example, the fastest deterministic prime recognition algorithm runs in nearly polynomial time, see [AD]

All the currently known deterministic factoring algorithms run in exponential times The trial division algorithm, which is the slowest factorization algorithm, has an exponential running time complexity of $L_n[1,1/2+\varepsilon] = n^{1/2+\varepsilon}$, $\varepsilon > 0$, and the fastest, the class group method, runs in time $L_n[1,1/5+\varepsilon] = n^{1/5+\varepsilon}$ In contrast, all the currently known probabilistic factoring algorithms run in exponential times to subexponential times The $p \pm 1$ methods have exponential time complexities $L_n[1,1/4+\varepsilon] = O(n^{1/4}\log^2 n)$, but the elliptic curve method, the quadratic sieve, and the number fields sieve factorization algorithms have subexponential time complexities Currently the number fields sieve factorization algorithm has the best performance, it has a subexponential time complexity of $L_n[1/3,\beta] = e^{[1\ 9229+o(1)](\ln(n))^{1/3}(\ln\ln(n))^{2/3}]}$, see [LA1, p 12], or [PM]

The research efforts are focused in the development of algorithms of polynomial running time complexities $L_n[0,\beta] = e^{\beta \ln\ln(n)} = (\log n)^{\beta}$, some constant $\beta > 0$

2 Background Materials

This section is essentially a survey of various results in additive number theory This subject deals with the representations of integers as sums of other integers One of the earliest documented attempt is the binary additive problem The binary problem (otherwise known as Goldbach conjecture, circa 1742) claims that every even integer N is a sum of two primes $N = p + q$

For each even integer N there are up to $\rho(N) \le \pi(N-2) - \pi(N/2-1)$ representations $N = p + q$, where $\pi(x)$ is the primes counting function Trivially the maximal number of representations of N as a sum of two primes is approximately

$$\rho(N) \le \pi(N-2) - \pi(N/2-1) \le \frac{N}{\log N} \tag{3}$$

The equality on the left holds for $N = 210$ only, see [DE] Using sieving techniques, the number of representations of $N \ne 210$ possible can be estimated as

$$\rho(N) = \frac{c_0 N}{\log^2 N}\prod_{p|N}(1+1/p) = \frac{c_1 N \log\log N}{\log^2 N} \tag{4}$$

where $\rho(N) = \#\{\, p \le N \ \ \exists\, q \text{ such that } p+q = N\,\}$, and c_0, c_1 are absolute constants It is also known that there are infinitely many integers N with the estimated maximal

number of representations, see [PR] On the other hand, the minimal number is not known But the Numerical data suggests that it should be about $\rho(N) >> \log^2(N)\text{loglog}(N)$, see [RC] for a recent numerical experiment Thus in practice $\rho(N) >> 1$ Exempli gratia,

$100 = 3 + 97 = 11 + 89 = 17 + 83 = 29 + 71 = 41 + 59 = 47 + 53,$

$\rho(100) = 6$

$900 = 13 + 887 = 17 + 883 = 19 + 881 = \quad = 439 + 461 = 443 + 457 = 433 + 467,$

$\rho(900) = 48$

$2500 = 23 + 2477 = 41 + 2459 = 53 + 2447 =$
$= 1193 + 1307 = 1217 + 1283 = 1223 + 1277,$

$\rho(2500) = 47$

Almost Binary Additive Problem

The almost binary problem is a somewhat simpler version of the binary problem, but in the applications to be considered here it has an advantage which will appear later

Theorem 1 (Linnik 1953) All large even integers N have representations of the form

$$N = p_1 + p_2 + 2^{\nu_k} + \quad + 2^{\nu_2} + 2^{\nu_1}, \tag{5}$$

where $\nu_i \geq 0$, and $k < c$, c some absolute constant

The effective formula for the number of representations $r_k(N)$ for all $k \geq 2$, and even $N \geq N_k$ is

$$r_k(N) = 2\frac{N \log_2^k N}{\log^2 N}\left(1 + O\left(\frac{\log^2 k}{k}\right)\right) \tag{6}$$

The original formula for $r_k(N)$ given by Linnik was simplified to the present form in [GL] Note that the same also holds for odd integers since $N + 2^i + 2^j =$ odd for some $i, j = 0, 1, 2$

For the application envisioned here the more general equation

$$N = x + y \pm 2^{\nu_k} \pm \quad \pm 2^{\nu_2} \pm 2^{\nu_1}, \tag{7}$$

with integers arguments $x, y \in \mathbb{Z}$ is also of interest

Ternary Additive Problem

The quadratic equation $x^2 + y^2 + z^2 \equiv N \bmod p$ is solvable for all integers $N \neq 4^u(8v + 7)$, and these are precisely the integers that are expressable as sums of three squares (Legendere's Theorem) The same equation with primes arguments has local solutions for all integers N that satisfy the congruences $N \not\equiv 0 \bmod 5$, and $N \equiv 3 \bmod 24$

Let $\mathcal{B} = \{ N \geq 1 \quad N \equiv 3, 27, 51, 99 \bmod 120 \}$ The set $\mathcal{B}$ has a positive density $\delta(\mathcal{B}) = 1/30$ in the set of natural numbers $\mathbb{N}$

***Theorem* 2.** (Hua 1938) Almost every integer $N \in \mathcal{B}$ is a sum of three squares primes

$$N = p_1^2 + p_2^2 + p_3^2 \tag{8}$$

A similar result for $N = p_1^2 + p_2^2 + p_3^k$, all $k \geq 1$, is also given in the same paper

The exceptional set $E(X) = \{ N \leq X \quad N \in \mathcal{B},\ N \neq p_1^2 + p_2^2 + p_3^2 \}$ consists of all the integers from the set $\mathcal{B}$ that are not expressable as sums of three square primes Very recently it has been shown that this later set is a thin set in $\mathcal{B}$, more precisely $\#E(X) << X^{151/160+\varepsilon}$, $\varepsilon > 0$, see [BA]

The number of representations $r(N)$ of an integer N as a sum of three primes has exponential magnitude

$$r(N) >> \frac{\sqrt{N}}{\log^3 N} \tag{9}$$

for all sufficiently large N

The proofs of all the stated results consist of careful analysis of the singular integrals and the singular series of the corresponding problem For details on the circle method, the singular integral/series, and a few other types of representations of integers, see [GR], [VA], etc

There are many completely solved problems in additive number theory The first fully solved additive problem appears to be the representations of integers as sums of four

squares $N = x^2 + y^2 + z^2 + w^2$, $x, y, z, w \in \mathbb{Z}$, (Lagrange 1770) But the binary problem, also one of the earliest, still remains unsolved

The more general equation

$$N = x^2 + y^2 \pm 2^{v_k} \pm \quad \pm 2^{v_2} \pm 2^{v_1}, \tag{10}$$

with integers arguments $x, y \in \mathbb{Z}$ will be utilized later on

3 Nonsingular Curves

The algebraic curves appearing in this work are of the form $C \quad y^2 = f(x)$, where $f(x) = (x - e_1)(x - e_2) \quad (x - e_d)$, $e_i \in \mathbb{Q}(\xi)$, is a polynomial of degree $\deg(f(x)) = d$ The discriminant $D = (e_1 - e_2)^2(e_1 - e_3)^2 \quad (e_d - e_{d-1})^2$ of the curve is a function of the coefficients of $f(x)$ The discriminant is nonzero $D \neq 0$ if and only if $f(x)$ has no repeated roots If this is the case then the curve is nonsingular of genus $g = (d - 1)(d - 2)/2$ The genus $g \geq 0$ of the curve, a quantity introduced by Riemann, and other invariants of algebraic curves measure the structure and complexity of the sets of integral points

$C(\mathbb{Q}) = \{ P = (x, y) \quad f(x,y) = 0, \text{ and } x, y \in \mathbb{Q} \}$

on the curves The algebraic sets $C(\mathbb{Q})$ (loci of curves) of curves of genus $g = 0$ have infinitely many integral points, but the algebraic sets of curves of genus $g > 0$ have finitely many integral points, (Faltings 1983), and [EV] Thus if a curve of genus $g > 0$ has an integral point $P = (x, y)$ that can be used in integers factorization, then the search for this point is limited to a smaller set of integral points

One of the most important integers factorization equation $f_n(x,y) = x^2 - y^2 - n = 0$ has genus $g = 0$ Accordingly it has infinitely many integral points, but only a few of these points are of interest in integers factorizations The existence of a similar equation $g_n(x,y) = 0$, but of genus $g > 0$ could reduce the search to a very small set of integral points

***Lemma* 3.** Let $s \geq 1$, and let $n = pq$ Then there exists a fixed integer $m \in \mathbb{Z}$, $| m | \leq an^{s/4}$, such that $P = (x = p^s \pm q^s, y \neq 0)$ is an integral point on the nonsingular curve $E_{m\,s}$ $y^2 = (x + m)(x^2 \pm 4n^s)$, some small number $a > 0$

The determination of another squarefree polynomial $f(x)$ similar to $x^2 - 4n^s$ immediately leads to the nonsingular curve $E_s \quad y^2 = (x^2 - 4n^s)f(x)$ Since a squarefree polynomial $f(x)$ is a square $f(k)$ integer for infinite many integers k if and only if $deg(f(x)) \leq 2$, the curve

has genus $g = 1$ (assuming it exists) This curve together with an integral points search algorithm provide a general purpose factorization algorithm for all integers $n = pq$

The complex algorithms for computing all the integral points on several types of algebraic curves $C \quad f(x,y) = 0$ have been under extensive developments for last few decades There is a vast literature on this subject, see [PT2] for a survey, and [SM], [ZM] for comprehensive introduction to computational Diophantine analysis, and elliptic curves algorithms respectively A few of the integral points search methods are listed here

(I) A curve $E \quad y^2 = (x - a_1)(x - a_2)(x - a_3)$ with three rational roots $P = (a_1, 0)$, $P = (a_2, 0)$, $P = (a_3, 0)$ The set of integral points $P = (x, y)$ in a large range $| x |, | y | \le B$ can be determined in time polynomial in (double logarithmic) loglog B, by applying 2-descend techniques and continued fraction approximations, see [ZA, p 426] for the complexity analysis, and [KE] for related details

(II) A curve $E \quad y^2 = (x - a_1)(x^2 + ax + b)$ with a single rational root $P = (a_1, 0)$, and $x^2 + ax + b$ is irreducible over $\mathbb{Z}$ The points are determined by applying 2-descend technique and using a basis of the group of points

(III) Given a basis $P_1, P_2, \quad , P_r$ of the free group of rational points $E(\mathbb{Q}) / E(\mathbb{Q})_{\text{tors}}$ of an elliptic curve $E \quad y^2 = x^3 + ax^2 + bx + c$, the Lang–Zagier algorithm determines all the linear combinations

$$P = n_1P_1 + n_2P_2 + \quad + n_rP_r + T, \tag{11}$$

where $n_i \in \mathbb{Z}$, $T \in E(\mathbb{Q})_{\text{tors}}$ This method gives a complete solution to the integral points problem The complexity analysis of this method is not available yet, but for nonsingular elliptic curves of rank ≤ 6, the algorithm is effective on such curves, see [GB, p 172], [ST], [PT1] et cetera

Since some of these algorithms can work faster on certain type of curves, the choice of algorithm depends on the specific curve In the present problem the entire set of integral points on the curve $E_{m\,s} \quad y^2 = (x + m)(x^2 - 4n^s)$ is not required, only those integral points $P = (x, y)$ with x-coordinates in a short interval Specifically, if $n = pq$ and the ranges of the prime factors are as follows

$$\alpha\sqrt{n} < p < \sqrt{n} < q < \beta\sqrt{n}\,, \tag{12}$$

where α, $0 < \alpha < 1$, and $\alpha\beta = 1$ Then the short interval of the x-coordinate is $2n^{s/2} \le x = p^s + q^s \le (\alpha^s + \alpha^{-s})n^{s/2}$, $s \ge 1$ For example, at $s = 1, 2$, the intervals are

(1) $2\sqrt{n} \le x = p+q \le (\alpha+\alpha^{-1})\sqrt{n}$, for the curve $E_m \quad y^2 = (x+m)(x^2-4n)$, (13)

(2) $2n \le x = p^2+q^2 \le (\alpha^2+\alpha^{-2})n$, for the curve $E_m \quad y^2 = (x+m)(x^2-4n^2)$

4 Nonsingular Curves Factorization Algorithms

The works in this section combine the theory of integers representations and elliptic curve algorithms to derive a new theory of integers factorization and the related factorization algorithms These algorithms effectively resolve certain integers $n = pq$ into their prime factors

Curves With One Rational Point

Let $\mathfrak{C}(N, k) = \{ n = pq \quad N = p+q+2^{v_k} + \quad +2^{v_2}+2^{v_1} \}$ Since $\#\mathfrak{C}(N, k) = r_k(N) >> N\log_2(N) / (\log^2(N))$, this is a set of cardinality $\#\mathfrak{C}(N, k) \ge O(n^{1/2-\varepsilon})$, $\varepsilon > 0$ Any integer form this set can be factored in running time $O((\log n))^k)$ using the following algorithm

Algorithm 1
Input $n = pq$, N, and k
Output p, q

(1) For $x \in R_k = \{ x = N \pm 2^{v_k} \pm \quad \pm 2^{v_2} \pm 2^{v_1},\ 0 \le v_i \le \log_2 n \}$, compute $2\hat{p} = x - \sqrt{x^2-4n}$, $2\hat{q} = x + \sqrt{x^2-4n}$
(2) If $\hat{p}\hat{q} = n$, then return the triple $p = \hat{p}$, $q = \hat{q}$, $\varphi(n) = n + 1 - x$, and terminate

Now define $T(n)$ as the sieving time on the curve $C \quad y^2 = f(x)$ for all integral points $P = (x, y)$ with x-coordinates in the range $2\sqrt{n} \le x \le \tau\sqrt{n}$, where $\tau = \alpha + \alpha^{-1} > 2$ In addition let $N = p+q \pm 2^{v_k} \pm \quad \pm 2^{v_2} \pm 2^{v_1}, 0 \le v_i \le 5\log_2 n$

***Theorem* 4.** Let $n = pq$, and suppose that N is a square integer for some $k \le c$, where c is an absolute constant Then there exists an algorithm that can factors $n = pq$ in running time $O(T(n)(\log n)^{c+3})$

Proof· Consider the parametized curve $E_m \quad y^2 = (x+m)(x^2-4n)$, $m \in S_k = \{ m = \pm 2^{v_k} \pm \quad \pm 2^{v_2} \pm 2^{v_1},\ 0 \le v_i \le \log_2 n \}$ By hypothesis the factors of n maps to an almost binary representation of of a square N, which coincides with the integral point $P = (x = p+q, y \ne 0)$ on one of these curves To determine the curve and the point, one searches through all the curves (for each choice of m or vector $\boldsymbol{v} = (v_k, \quad , v_2, v_1)$, there is a different curve), it takes at most $\log_2^k N < \log_2^c N$ trials Thus the running time to determine the point $P = (x = p+q, y \ne 0)$ is at most

$\log_2^k N \;\; T(n) \le (\log n)^3 \log_2^c N \;\; T(n) = O(T(n)(\log n)^{c+3})$ ■

Algorithm 2
Input $n = pq$
Output p, q
(1) For $m \in R_k = \{ \; m = \pm 2^{v_k} \pm \quad \pm 2^{v_2} \pm 2^{v_1} \in \mathbb{Z} : 0 \le v_i \le \log_2 n \; \}$, compute the set
$V_m = \{ \; x \in \mathbb{N} \quad (x, y \neq 0)$ is $\mathbb{Z}$-point on $y^2 = (x+m)(x^2 - 4n)$, and $2\sqrt{n} \le x \le \tau\sqrt{n} \; \}$
(2) For $x \in V_m$, compute $2\hat{p} = x - \sqrt{x^2 - 4n}, \quad 2\hat{q} = x + \sqrt{x^2 - 4n}$
(3) If $\hat{p}\hat{q} = n$, then return the triple $p = \hat{p}, q = \hat{q}$, $\varphi(n) = n + 1 - x$, and terminate

***Example* 5.** To factor $n = 253$, let $\alpha = 5$, and put $m = m(\mathbf{v}) = \pm 2^{v_1}, \pm 2^{v_2} \pm 2^{v_1}, \quad , \; 0 \le v_i \le 5\log_2(39) < 3, k = 1, 2,$ Then the corresponding parametized curve is

$E_t \;\; y^2 = (x+m)(x^2 - 1012) = x^3 + mx^2 - 1012x - 1012m$

Apply a direct sieve in the interval $2\sqrt{253} \le x \le 2\,5\sqrt{253}$, ($\tau = \alpha + \alpha^{-1} = 2\,5$), for $m(\mathbf{v})$ $= -2, -1, 0, 1, 2, -4, -3, \quad$, to obtain the followings table (this procedure is algorithm 2)

$m(\mathbf{v})$	Integral Points (x, y)	Factors $\hat{p}, \quad \hat{q}$	$\varphi(n)$
−2	None		
−1	None		
0	None		
1	None		
2	(34, ±72)	$\hat{p} = 11, \quad \hat{q} = 23$	$\varphi(n) = 220$

Curves With Three Rational Points
The curve $E_1 \;\; y^2 = (x - mt)(x^2 - 4n)$ of the previous result has a single rational root, so the sieving method is probably not the faster one The next curve $E_2 \;\; y^2 = (x - m)(x^2 - 4n^2)$ in the collection $E_s \;\; y^2 = (x - m)(x^2 - 4n^s)$ has three rational roots, and sieving method is probably the fastest one

Let $\mathcal{L}(N) = \{ \; n = p_1 p_2 \quad N = p_1^2 + p_2^2 + p_3^2 < n^2$, and $(p_1, p_2) \in P_n \times Q_n \; \}$ be a subset of integers (other similarly parametized subsets can also be used) Note that omitting the permutations of a primes triple $p_{1,i}, p_{2,i}, p_{3,i}$, the transition from the ith representation

$N = p_{1,i}^2 + p_{2,i}^2 + p_{3,i}^2$ to the jth representation $N = p_{1,j}^2 + p_{2\,j}^2 + p_{3\,j}^2$ has at least one change in primes Thus each representation maps to one or more distinct integers $n = pq$, (up to three)

***Theorem* 6.** There exists an algorithm that can factor $n \in \mathcal{L}(N)$ in running time $O((\log n)^3(\text{loglog}(2n^2))^d)$, where $d > 0$ is a constant

Proof· Let $n = pq$, and $N = p^2 + q^2 + r^2$ Then the integral point $P = (x = p^2 + q^2, y \neq 0)$ is on the curve E_N $y^2 = (N - x)(x^2 - 4n^2)$ By the sieving method (I), the running time to determine all the integral points $P = (x, y \neq 0)$ such that $2n \leq x \leq \tau n$, is

$$T(n) = O((\text{loglog } B)^d) = O((\text{loglog } 2n^2)^d),$$

where $d \geq 1$ is a constant, and $B = \max \{ |x|, |y| \}$ ■

Algorithm 3
Input $n = pq$, N
Output p, q
(1) Compute $V_N = \{ x \in \mathbb{N} \quad (x, y \neq 0)$ is $\mathbb{Z}$-point on $y^2 = (N - x)(x^2 - 4n^2)$, and $2\sqrt{n} \leq x \leq \tau\sqrt{n}$ $\}$
(2) For $x \in V_N$, compute $2\hat{p} = \sqrt{x + 2n} - \sqrt{x - 2n}$, $2\hat{q} = \sqrt{x + 2n} + \sqrt{x - 2n}$
(3) If $\hat{p}\hat{q} = n$, then return the triple $p = \hat{p}$, $q = \hat{q}$, $\varphi(n) = n + 1 - p - q$, and terminate

Note The second line recovers the factors p and q from the integral point $P = (x = p^2 + q^2, y \neq 0)$ To accomplish this one can utilize the well known formulas

$$2p = n + 1 - \varphi(n) - \sqrt{[n + 1 - \varphi(n)]^2 - 4n}, \text{ and} \qquad (14)$$

$$2q = n + 1 - \varphi(n) + \sqrt{[n + 1 - \varphi(n)]^2 - 4n}$$

to derive the new formulas

$$\text{Compute } 2p = \sqrt{x + 2n} - \sqrt{x - 2n}, \quad 2q = \sqrt{x + 2n} + \sqrt{x - 2n}, \text{ and} \qquad (15)$$

$$\varphi(n) = n + 1 - \sqrt{x + 2n}$$

Remark: These algorithms were originally developed specifically for the two-prime composites $n = pq$ However the structures are richer, and can actually be used to discompose any composite integers, including even, perfect powers, and nonsquaresfree integers The length of the sieving interval $2n^{s/2} \le x \le \tau^s n^{s/2}$, $(\tau^s = \alpha^s + \alpha^{-s} \ge 2)$, determines the size of the factors, for example, $\tau \approx 2$ for nearly equal factors The same formulae (15) given above can be used to recover the factors, but the corresponding formula for $\varphi(n)$ works only for $n = pq$

The Special Case $E_0 : y^2 = x^3 - 4nx$

Each of the curve on the collection E_m $y^2 = (x + m)(x^2 - 4n)$ has a rational root, which makes the determination of the integral points some what easier The curve E_0 $y^2 = x(x^2 - 4n)$, the simplest case, provides an algorithm for factoring all integers $n = n_1 n_2$ such that $n_1 + n_2 = \square$ This includes an infinite sequence of integers $n = pq$ with the primes factors in the short interval $\alpha\sqrt{n} < p < \sqrt{n} < q < \beta\sqrt{n}$ as $n \to \infty$

Some of its characteristics are as follows

(1) The torsion group $E(\mathbb{Q})_{\text{tors}} = \{ O = (\infty, \infty), (0, 0) \} \cong \mathbb{Z}_2$

(2) The set of rational points $E(\mathbb{Q}) \cong \mathbb{Z}_2 \times \mathbb{Z}^r$, $1 \le r \le 5$

The upper limit of the rank $r \le 5$ is determined using a 2-descend method, the actual calculations appear in various places in the literature On the other hand, since there is an integral point $(x > 0, y \ne 0)$ in $E(\mathbb{Q})/2E(\mathbb{Q})$, which is not in $E(\mathbb{Q})_{\text{tors}}$, the rank $r \ge 1$ is nontrivial

Since only the integral points (x, y) in a large range, perhaps $2\sqrt{n} \le x \le \tau\sqrt{n}$, are required, there are several plausible strategies

(1) To sieve for integral points directly in the range $2\sqrt{n} \le x \le 2\,5\sqrt{n}$, (the algebraic system SIMATH is a possibility)

(2) To sieve for integral points in the compact component $E(\mathbb{Q})^c = \{ (x, y) \quad -2n^{1/2} \le x \le 0 \}$ and then use the duplication and addition formulas to take linear combinations of the form

$$P = n_1 P_1 + n_2 P_2 + T, \tag{16}$$

where $n_i \in \mathbb{Z}$, $P_1 = (x_1, y_1)$, $P_2, = (x_2, y_2) \in E(\mathbb{Q})^c$, and $T = (0, 0)$, to get integral points in the identity component $E(\mathbb{Q})^0 = \{ (x, y) \quad 2n^{1/2} \le x \le \infty \}$ The duplication and addition formulas are

$$2P = \left(\frac{x^2 - 4n}{4x}, -\left(\frac{3x^2 - 4n}{2y} \right)^2 \left(\frac{x^2 - 4n}{4x} \right) + \frac{x^3 + 4n}{2y} \right), \tag{17}$$

for $P = (x, y) \neq (0, 0)$ and for pair of distinct points $P_1 = (x_1, y_1) \neq \pm P_2 = (x_2, y_2)$

$$P_1 + P_2 = \left(\left(\frac{y_2 - y_1}{x_2 - x_1} \right)^2 - x_1 - x_2, -\left(\frac{y_2 - y_1}{x_2 - x_1} \right) \left(\left(\frac{y_2 - y_1}{x_2 - x_1} \right)^2 - x_1 - x_2 \right) - \left(\frac{x_2 y_1 - x_1 y_2}{x_2 - x_1} \right) \right)$$

(3) Compute a basis, and find all the linear combinations in the range of interest The authors in [ST] claims that the problem is tractable for curves of rank $r \leq 6$

***Example* 7.** For $n = 39$, the corresponding curve is $E \quad y^2 = x(x^2 - 156)$ The integral compact component is $E(\mathbb{Q})^c = \{ (x, y) \quad -2n^{1/2} \leq x \leq 0 \} = \{ (-12, \pm 12), (-3, \pm 21), (0, 0) \}$ Now adding points from this set gives
$(13, -13) = (-12, 12) + (0, 0), \qquad (16, -40) = (-12, 12) + (-3, 21)$, etc

Thus the identity component is

$E(\mathbb{Q})^0 = \{ O, (13, \pm 13), (16, \pm 40), (52, \pm 2548), \quad \}$

Applying algorithm 2, (here $V_m = E(\mathbb{Q})^0$), returns $p = 3$, $q = 13$, and $\varphi(n) = 24$

Additive and Multiplicative Types Factoring Algorithms

The integers factoring algorithms, (CFRAC, QS, NFS, etc) based on the equation $f_n(x,y) = x^2 + y^2 - n = 0$ construct the solutions of this equation by a process of completing the square through products These will be classified as *multiplicative* type factoring algorithms On the other hand, the integers factoring algorithms based on the equation $g_n(x,y) = 0$ that construct the solutions of this equation by a process of completing the square through sums will be classified as *additive* type factoring algorithms All the new algorithms introduced here are additive type

One of the marked difference between these two types of algorithms is the running times The running time of a multiplicative type factoring algorithm grows exponentially or subexponentially as $n \to \infty$ (conjecturally or heuristically speaking) In contrast, the running time of an additive type factoring algorithm grows polynomially or nearly polynomially as $n \to \infty$, (conjecturally speaking, see below)

This classification of the algorithms is motivated by the following simple observations

Consider the (signed binary) representations

$$N = n_1^2 + n_2^2 \pm 2^{v_k} \pm \quad \pm 2^{v_2} \pm 2^{v_1}, \tag{19}$$

and the corresponding curve

$$E_m \; y^2 = (x + m)(x^2 - 4n^2), \tag{20}$$

where the error term

$$m = \pm 2^{v_k} \pm \quad \pm 2^{v_2} \pm 2^{v_1}, \tag{21}$$

and $0 \le v_i \le c_0 \log_2(\tau n)$, $c_0 > 0$ The error term is used to complete the square

Trivially, any integer $N \ge 2n$ has a representation of the form (19) for some k in the range $2 \le k \le 1 + 5\log_2(\tau n)$, and $n \in \mathbb{N}$ It is quite plausible that every integer has one or more representations as in (19) with $k \le c$, where c is an absolute constant There is evidence to back this claim, for instance, the subset of integers

$$U_2 = \{ n = n_1 n_2 \quad n_1^2 + n_2^2 \pm 2^{v_2} \pm 2^{v_1} = \text{square, and } 0 \le v_i \le \log_2 n \} \tag{22}$$

(only two powers of 2 are needed to complete the squares) is a set of positive density in the set of integers $\mathbb{N}$ Some related analysis is given in [LI], for instance the number $A(X)$ of integers $N \le X$ representable as $N = p_1^2 + p_2^2 + 2^{v_2} + 2^{v_1}$ is $A(X) = c_1 X$, some absolute constant $c_1 > 0$

This mild assumption $k \le c$ on the parameter k leads to a very interesting result a polynomial time algorithm On the other hand, if k grows as $k(n) = c_2 \log_2 \log_2(n)$, $c_2 > 0$, (or a similar function perhaps $k(n) = c_3(\log_2 \log_2 n)^e$, $e > 1$,), then the corresponding algorithm runs in nearly polynomial time

***Theorem* 8.** Suppose that there exists an absolute constant $c > 2$ such that $k \le c$ in the representations of any square integers N Then the integer n can be factored in running time $O(\log n)^{c+3}(\log\log(2n^2))^d)$

Algorithm 4
Input $n = n_1 n_2$
Output n_1, n_2

(1) For $m \in R_k = \{ \pm 2^{v_k} \pm \quad \pm 2^{v_2} \pm 2^{v_1}, 0 \le v_i \le \log_2 n \}$, compute the set of points
$V_m = \{ x \in \mathbb{N} \quad (x, y \ne 0)$ is a $\mathbb{Z}$-point on $y^2 = (x + m)(x^2 - 4n^2)$, and $2\sqrt{n} \le x \le \tau\sqrt{n} \}$

(2) For $x \in V_m$, compute $2\hat{p} = \sqrt{x+2n} - \sqrt{x-2n}$, $2\hat{q} = \sqrt{x+2n} + \sqrt{x-2n}$

(3) If $\hat{p}\hat{q} = n$, then return the triple $p = \hat{p}, q = \hat{q}$, $\varphi(n) = n + 1 - p - q$, and terminate

Proof· The outer loop consisting of line 1 is repeated at most $(2(1 + \log_2 n))^k = O((\log n)^k)$ times In each repetition the set V_m is determined at a cost of $O((\text{loglog } 2n^2)^d)$, some $d > 0$, this follows from the analysis in [ZA] The inner loop consisting of lines 2, and 3 is repeated at most $3\ 7^8$ times, this estimate of the cardinality of the set V_m follows from a result of [EV] In each repetition a pair of integer square roots is computed and a few other arithmetic operations at a cost of about $O((\log n)^3)$ The overall complexity is

$$O((\log n)^k) \times O((\text{loglog } 2n^2)^d) \times O((\log n)^3) = O((\log n)^{k+3}(\text{loglog } 2n^2)^d)$$

Therefore if one assumes the existence of an absolute constant $c > 2$ such that $k \le c$, then the algorithm runs in polynomial time But if the integer k grows as $k(n) = c_2\log_2\log_2 n$, or $c_2(\log_2\log_2 n,)^e$, $e > 1$, then the algorithm runs in nearly polynomial time ■

***Corollary* 9.** There exists a polynomial time prime recognition algorithm

Proof· Since the factors n_1 and n_2 of any integer $n = n_1n_2$ satisfy the inequalities $2 \le n_1 \le \sqrt{n} \le n_2 \le n/2$, let $\tau n = 4 + n^2/4 = B$, and $2n \le x \le \tau n$ Clearly, if the algorithm 4 does not find a factor of n, then n is prime ■

The running time of the penultimate result has a resemblance to the running time $O(\log^2(n)\text{loglog}(n)\text{logloglog}(n))$ of the theoretical integers factorization on a quantum computer, see [SH], which is a probabilistic algorithm

The running time complexities of the fastest probabilistic factoring algorithms, ECM, QS, NFS, are all based on some reasonable conjectures and/or heuristic arguments, see [PM, p 265] The same is true of the quantum integers factorization algorithm Similarly the running time complexity $O(n^{1/5+\varepsilon})$ of the fastest deterministic factoring algorithm, the class group method, is based on the validity of the extended Riemann hypothesis In the same tradition, this paper using a reasonable assumption, (which is probably not difficult to overcome), has shown the existence of a polynomial time (or nearly polynomial time) integers factorization algorithm

***Example* 9.** To factor $n = 253$, and put $\alpha = 5$ The idea of the algorithm is to search for integer points on the sequence of curves

$y^2 = (x + 2^{v_2} + 2^{v_1})(x^2 - 4n^2)$, $y^2 = (x + 2^{v_3} + 2^{v_2} + 2^{v_1})(x^2 - 4n^2)$, ,

where $m(v) = 2^{v_k} + \quad + 2^{v_2} + 2^{v_1}$, $0 \le v_i \le 1 + 5\log_2(253) < 6$, $k = 2, 3,$ Apply a direct sieve in the interval $2\,253 \le x \le 2\,253$, ($\tau = \alpha^2 + \alpha^{-2} = 4\,25$), and the verification procedure to obtain the followings table

$m(v)$	Integral Points (x, y)	Factors $\hat{p}$, $\hat{q}$	$\varphi(n)$
14	(519, 2665), (966, 25760)	$\hat{p} = 14\,21$, $\hat{q} = 17\,81$ $\hat{p} = 8\,46$, $\hat{q} = 28\,71$	
16	(572, 6468)	$\hat{p} = 12\,35$, $\hat{q} = 20\,48$	
26	(650, 10608), (814, 18480)	$\hat{p} = 11$, $\hat{q} = 23$ $\hat{p} = 9\,39$, $\hat{q} = 26\,94$	$\varphi(n) = 220$

The curves for $m = 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 17, 18, 19, 20, 22, 24,$ and 25 do not have integral points in this range Note that the congruence $N \equiv 0, 1 \bmod 4$ for square integer $N = p^2 \pm q^2 + m$ is effective in reducing the number of curves to be sieved

References:

[AD] LM Adleman, C Pomerance, RS Rumely, *On Distinguishing Prime Numbers From Composites Numbers*, Ann Math Vol 117, 173-206, 1983
[BA] C Bauer et al, *On a sum of three primes squares*, J Number Theory 85 (2000), 336-339
[BR] R Brent, *Recent Progress And Prospects For Integers Factorisation Algorithms*, 2000, available at www
[CV] Stefania Cavallar, et al, *Factorization of a 512-Bit Modulus*, EUROCRYPT 2000, LNCS 1807, pp 1-18
[DE] JM Deshouillers et al, *An Upper Bound In Goldbach's Problem, Math Computation,* available at www
[EV] JH Evertse, *On Equations In S-Units And The Thue-Mahler Equation*, Inv Math 75, 561-584, 1984
[GB] J Gebel, A Petho, HG Zimmer, *Computing Integral Points On Elliptic Curves*, Acta Arithimeticae Vol 68 No 2, (1994), pp 171-190
[GL] PX Gallagher, *Primes and Power of* 2, Inventiones Mathematicae 29, 125-142 (1975)
[GR] E Grosswald, Representations of Integers as Sums of Squares, Springer-Verlag, NY, 1985
[HA] LK Hua, *Some Results In The Additive Prime Number Theory*, Q J Math 1938, pp 68-80
[KE] KS Kedlaya, *Solving Constrainted Pell Equations*, Math Computation, Vol 67 No 222, (1998) pp 833-842
[LA1] A Lenstra et al , Development of Numbers Field Sieve, LNM 1554, Springer-Velag, NY 1993
[LA2] A Lenstra, *Integers Factoring*, Designs, Codes, and Cryptography, 19, 101-128 (2000)
[LH] HW Lenstra, *Factoring Integers With Elliptic Curves*, Ann Math 126 (1987) 649-673
[OD] AM Odlyzko, *The Future Of Integer Factorization*, 1995, available at www
[PO] JM Pollard, *Theorems On Factorizations And Primality Testing*, Proc Cambridge Phil Soc 5 (1974) 521-528
[PA] AN Parshin, IR Shafarevich, Number Theory I, EMS Series Vol 49, Springer-Verlag, 1995
[PM] C Pomerance RE Crandall, Primes Numbers A Computational Perspective, Springer-Verlag, 2001
[PR] K Prachar, *On Integers Having Many Representations As A Sum Of 2 Primes*, J London Math Soc Vol 29, No 116, 1954, pp 347-350
[PT1] A Petho, et al, *Computing All S–Integral Points On Elliptic Curves*, Math Proc Camb Phil Soc (1999) 127, 383-402
[PT2] A Petho, HG Zimmer, *S-Integral Points On Elliptic Curve, Theory And Practice,* Algebraic Number Theory And Diophantine Analysis, pp 352–356, deGruyter, NY 2000

[RC] Jorg Richstein, *Verifying The Goldbach Conjecture Up To* $4\,10^{14}$, Math Computation, Vol 70 No 236, (2001) pp 1745-1749
[LI] J Liu et al, *Squares Of Primes And Powers Of 2*, Monatshefte Mathematik, 128, 283-313 (1999)
[SM] NP Smart, The Algorithmic Resolution of Diophantine Equations, ST Series Vol 41, London Math Soc , 1998
[SH] PW Shor, *Polynomial Time Algorithms For Prime Factorization And Discrete Logarithms On A Quantum Computer*, SIAM J Computing Vol 26, No 5 Pp 1484-1509, 1997
[ST] RJ Stroeker, N Tzanakis, *Solving Elliptic Diophantine Equations by Estimating Linear Form in Elliptic Logarithms*, Acta Arith Vol 67 No 2, (1994), pp 177-196
[VA] RC Vaughan, Hardy-Littlewood Method, Combridge U Press, 1997
[WI] HC Williams, *A p + 1 method of factoring*, Math Computation, Vol 39 No ?, (1982) pp 225-234
[ZA] D Zagier, *Large Integral Points On Elliptic Curves*, Math Computation, Vol 48 No 177, (1987) pp 425-436
[ZM] HG Zimmer, *Basic Elliptic Curves Algorithms*, available at www

School of Computer Science and Information Systems
Pace University
Technical Report Series

EDITORIAL BOARD

Editor
Allen Stix, Computer Science, Pace--Westchester

Associate Editors:
Connie Knapp, Information Systems, Pace--New York
Susan M Merritt, Dean, SCSIS--Pace

Members
Howard S Blum, Computer Science, Pace--New York
Donald M Booker, Information Systems, Pace--New York
M Judith Caouette, Office Information Systems, Pace--Westchester
Nicholas J DeLillo, Mathematics and Computer Science, Manhattan College
Fred Grossman, Information Systems, Pace--New York
Fran Goertzel Gustavson, Information Systems, Pace--Westchester
Joseph F Malerba, Computer Science, Pace--Westchester
John S Mallozzi, Computer Information Sciences, Iona College
John C Molluzzo, Information Systems, Pace--New York
Narayan S Murthy, Computer Science, Pace--New York
Catherine Ricardo, Computer Information Sciences, Iona College
Sylvester Tuohy, Computer Science, Pace--Westchester
C T Zahn, Computer Science, Pace--Westchester

The School of Computer Science and Information Systems, through the Technical Report Series, provides members of the community an opportunity to disseminate the results of their research by publishing monographs, working papers, and tutorials *Technical Reports* is a place where scholarly striving is respected

All preprints and recent reprints are requested and accepted New manuscripts are read by two members of the editorial board, the editor decides upon publication Authors, please note that production is Xerographic from the pages you have submitted Statements of policy and mission may be found in issues #29 (April 1990) and #34 (September 1990)

Please direct submissions as well as requests for single copies to

Allen Stix
School of CS & IS - Suite 412 Graduate Center
Pace University
1 Martine Avenue
White Plains, NY 10606-1932