Integer Factorization Algorithms Illustrated by the Factorization of Fermat Numbers[*]

Richard P. Brent
Computing Laboratory
University of Oxford
`rpb@comlab.ox.ac.uk`

[*] Presented at Number Theory Day, CWI, Amsterdam, 8 May 1998. Copyright ©1998, R. P. Brent. CWI98t

Abstract

We first compare several integer factorization algorithms, including ECM, MPQS and NFS, for the application of factoring "typical" or "random" large integers. We then illustrate some of the conclusions by giving a brief historical summary of attempts to factor Fermat numbers.

Outline

- Notation and definitions.
- The elliptic curve method (ECM).
- Comparison of ECM and MPQS.
- Comparison of ECM and NFS.
- Factorization of Fermat numbers.
- Factorization of F_{10}: some details.

Notation

n and N always denote positive integers.

p_n denotes a prime number with n decimal digits, e.g. $p_3 = 163$. Similarly, c_n denotes a composite number with n decimal digits, e.g. $c_4 = 1729$.

Almost Always and Almost Never

If $P(n)$ is a predicate, we say that $P(n)$ holds *almost always* if

$$\lim_{N\to\infty} \frac{|\{n \le N : P(n)\}|}{N} = 1$$

and we say that $P(n)$ holds *almost never* if

$$\lim_{N\to\infty} \frac{|\{n \le N : P(n)\}|}{N} = 0 \; .$$

Example (Erdős–Kac): For any $\varepsilon > 0$, n almost always has between $(1-\varepsilon)\log\log n$ and $(1+\varepsilon)\log\log n$ prime factors.

A Brief Description of ECM

The *elliptic curve method* (ECM) was discovered by H. W. Lenstra, Jr. in 1985. Various practical refinements were suggested by Montgomery, Suyama, and others. References can be found in my report [4].

Lenstra's key idea was to apply Pollard's "$p-1$" method but to work over a different group G. If the method fails, another group can be tried. This is not possible for the $p-1$ method, because it uses a fixed group.

ECM uses groups defined by pseudo-random elliptic curves over F_p, where $p > 3$ is the prime factor we hope to find. (Fortunately, we don't need to know p in advance.) By a theorem of Hasse (1934), the group order g for an elliptic curve over F_p satisfies

$$|g - p - 1| < 2\sqrt{p}\ .$$

By a result of Deuring, all g satisfying this inequality are possible.

Lenstra's Analysis of ECM

Consider applying ECM to a composite integer N with smallest prime factor p. Modulo an unproved but plausible assumption regarding the distribution of prime factors of random integers in "short" intervals, Lenstra showed that ECM will find p in an expected number

$$W(p) = \exp\left(\sqrt{(2+o(1))\log p \log\log p}\right)$$

of multiplications (mod N), where the "$o(1)$" term tends to zero as $p \to \infty$.

ECM can routinely find factors p of size about 30 decimal digits, and it has successfully found factors as large as 49 decimal digits. Details can be found in [3].

Choice of Parameters

ECM has several parameters. The most important is the first-phase limit B_1. The optimal choice of the parameters depends on the size of the factor p. Since p is unknown, we have to guess or use some sort of adaptive strategy. Some suggestions are given in my report [4]. Fortunately, the expected performance of ECM is not very sensitive to the choice of parameters.

Expected Performance of ECM

In Table 1 we give a small table of $\log_{10} W$ for factors of D decimal digits. The precise figures depend on assumptions about the implementation, see [4].

Table 1: Expected work for ECM

digits D	$\log_{10} W$
20	7.35
30	9.57
40	11.49
50	13.22
60	14.80

ECM, MPQS and NFS

We assume that the *multiple polynomial quadratic sieve* (MPQS), and the (general) *number field sieve* (NFS) are familiar.

If N is a (large) integer with prime factors $p_1 \geq p_2 \geq \ldots$, we *assume* that the expected time to factor N by these three methods is $T_{ECM}(N), T_{MPQS}(N), T_{NFS}(N)$ respectively, where

$$\log T_{ECM} = \sqrt{(2+o(1))\log p_2 \log\log p_2}$$

$$\log T_{MPQS} = \sqrt{(1+o(1))\log N \log\log N}$$

$$\log T_{NFS} = \sqrt[3]{(c+o(1))\log N\,(\log\log N)^2}$$

Here c is some positive constant, and the $o(1)$ terms are as $p_2 \to \infty$ or $N \to \infty$.

ECM and MPQS

Theorem
$T_{MPQS}(N) > e^{\sqrt{\log N}} T_{ECM}(N)$ almost always.

Idea of Proof

$$\left(\frac{\log T_{MPQS}}{\log T_{ECM}}\right)^2 \geq \frac{\log N}{(2+o(1))\log p_2}$$

but from the known distribution of $\log p_2 / \log N$ this is at least $1+\varepsilon$ with probability at least $1 - O(\varepsilon^2)$. Thus, the Theorem holds if $e^{\sqrt{\log N}}$ is replaced by $f(N)$, where

$$\log f(N) = o\left(\sqrt{\log N \log\log N}\right) .$$

Corollary
For all $\varepsilon > 0$, $T_{ECM} < \varepsilon T_{MPQS}$ holds almost always.

ECM and NFS

Theorem
For all $\varepsilon > 0$, $T_{NFS} < \varepsilon T_{ECM}$ holds almost always.

However, *this is not the full story*, because ECM can find small factors quickly, and after dividing them out NFS can finish the factorization more quickly than if ECM had not been used.

Let $T_{ECM}^{(\lambda)}(N)$ be the expected time for ECM to find at least λk prime factors of N, where k is the total number of prime factors of N. (It does not matter how we count multiple factors.)

Theorem
Let K be any positive constant, $0 \leq \lambda \leq 1$.
If $\lambda < 2/3$ then $T_{ECM}^{(\lambda)} < KT_{NFS}$ almost always, and if $\lambda > 2/3$ then $T_{ECM}^{(\lambda)} > KT_{NFS}$ almost always.

Thus, it is better to use a combination of ECM and NFS than either alone, and with a sensible strategy we expect to find about two thirds of the prime factors by ECM and the remaining one third by NFS.

Some history of Fermat numbers

For a nonnegative integer n, the n-th *Fermat number* is $F_n = 2^{2^n} + 1$. It is known that F_n is prime for $0 \le n \le 4$, and composite for $5 \le n \le 23$. Also, for $n \ge 2$, the factors of F_n are of the form

$$k2^{n+2} + 1 \, .$$

In 1732 Euler found that $641 = 5 \cdot 2^7 + 1$ is a factor of F_5, thus disproving Fermat's belief that all F_n are prime. Euler apparently used trial division by primes of the form $64k + 1$ (not just $128k + 1$).

The complete factorization of the Fermat numbers $F_6, F_7, \ldots$ has been a challenge since Euler's time. Because the F_n grow rapidly in size, a method which factors F_n may be inadequate for F_{n+1}.

No Fermat primes larger than F_4 are known, and a probabilistic argument makes it plausible that only a finite number of F_n (perhaps only $F_0, \ldots, F_4$) are prime. It is known that F_n is composite for $5 \le n \le 23$.

F_6

In 1880, Landry factored $F_6 = 274177 \cdot p_{14}$. Landry's method was never published in full, but Williams has attempted to reconstruct it.

Hand Computations

In the period 1877–1970, several small factors of F_n for various $n \ge 9$ were found by taking advantage of the special form of these factors. For example, in 1903 Western found the factor $p_7 = 2424833 = 37 \cdot 2^{16} + 1$ of F_9.

Significant further progress was only possible with the development of the digital computer and more efficient algorithms.

F_7

In 1970, Morrison and Brillhart factored

$$F_7 = 59649589127497217 \cdot p_{22}$$

by the continued fraction method. This method has now been superseded by MPQS which, perhaps surprisingly, has never been the first to factor a Fermat number.

F_8

In 1980, Brent and Pollard factored

$$F_8 = 1238926361552897 \cdot p_{62}$$

by a modification of Pollard's "rho" method. The "rho" method is now largely superseded by ECM.

The larger factor p_{62} of F_8 was first proved prime by Williams using the method of Williams and Judd. Later, I provided a simpler proof by factoring $p_{62} - 1$.

Nowadays, F_7 and F_8 are "easy" to factor by ECM or MPQS.

F_9

Logically, the next step after the factorization of F_8 was the factorization of F_9. It was known that

$$F_9 = 2424833 \cdot c_{148}$$

The 148-digit composite number resisted attack by methods such as Pollard rho, Pollard $p \pm 1$, and the elliptic curve method (ECM), which would have found "small" factors. It was too large to factor by the continued fraction method or even by MPQS.

The difficulty was finally overcome by the invention of the (special) number field sieve (SNFS), based on a new idea of Pollard.
In 1990, Lenstra, Lenstra, Manasse and Pollard, with the assistance of many collaborators and approximately 700 workstations scattered around the world completely factored F_9 by SNFS.

$$\begin{aligned} F_9 &= 2424833 \cdot p_{49} \cdot p_{99} \,, \\ p_{49} &= 7455602825647884208337395736200454918783366342657 \end{aligned}$$

Later, SNFS was generalised to GNFS (what we called simply NFS above).

F_{10}

After the factorization of F_9 in 1990, F_{10} was the "most wanted" number in various lists of composite numbers.

F_{10} was proved composite in 1952 by Robinson, using Pépin's test on the SWAC. A small factor, 45592577, was found by Selfridge [18] in 1953 (also on the SWAC). Another small factor, 6487031809, was found by Brillhart in 1962 on an IBM 704. Brillhart later found that the cofactor was a 291-digit composite.

Using ECM we found a 40-digit factor of F_{10} on October 20, 1995. The 252-digit cofactor c_{291}/p_{40} passed a probabilistic primality test and was soon proved to be prime using the method of Atkin and Morain (based, appropriately, on elliptic curves). Thus, the complete factorization of F_{10} is

$$\begin{aligned} F_{10} &= 45592577 \cdot 6487031809 \cdot p_{40} \cdot p_{252} \,, \\ p_{40} &= 4659775785220018543264560743076778192897 \end{aligned}$$

F_{11}

F_{11} was completely factored in 1988, *before* the factorization of F_9 and F_{10}. In fact,

$$\begin{aligned} F_{11} = \; & 319489 \cdot 974849 \cdot \\ & 167988556341760475137 \cdot \\ & 3560841906445833920513 \cdot p_{564} \end{aligned}$$

The two 6-digit factors were found by Cunningham in 1899, and I found the remaining factors in May 1988, using ECM on a Fujitsu VP100. The 564-digit factor passed a probabilistic primality test, and a rigorous proof of primality was provided by Morain.

The reason why F_{11} could be completely factored before F_9 and F_{10} is that the difficulty of completely factoring numbers by ECM is determined mainly by the size of the *second-largest* prime factor of the number.

The second-largest prime factor of F_{11} has 22 digits and is much easier to find by ECM than the 40-digit factor of F_{10} or the 49-digit factor of F_9.

Summary

A brief summary of the history of factorization of $F_5, \ldots, F_{11}$ is given in the Table.

Table 2: Complete factorization of F_n, $n = 5, \ldots, 11$

n	Factorization	Date	Comments
5	$p_3 \cdot p_7$	1732	Euler
6	$p_6 \cdot p_{14}$	1880	Landry
7	$p_{17} \cdot p_{22}$	1970	Morrison and Brillhart
8	$p_{16} \cdot p_{62}$	1980	Brent and Pollard (p_{16}, p_{62})
		1980	Williams (primality of p_{62})
9	$p_7 \cdot p_{49} \cdot p_{99}$	1903	Western (p_7)
		1990	Lenstra *et al* (p_{49}, p_{99})
10	$p_8 \cdot p_{10} \cdot p_{40} \cdot p_{252}$	1953	Selfridge (p_8)
		1962	Brillhart (p_{10})
		1995	Brent (p_{40}, p_{252})
11	$p_6 \cdot p_6' \cdot p_{21} \cdot p_{22} \cdot p_{564}$	1899	Cunningham (p_6, p_6')
		1988	Brent (p_{21}, p_{22}, p_{564})
		1988	Morain (primality of p_{564})

F_{12}

The smallest Fermat number which is not yet completely factored is F_{12}. It is known that

$$\begin{aligned} F_{12} \quad &= \quad 114689 \cdot 26017793 \cdot \\ & \quad\quad 63766529 \cdot 190274191361 \cdot \\ & \quad\quad 1256132134125569 \cdot c_{1187} \ , \end{aligned}$$

where the 16-digit factor was found by Baillie in 1986, using the Pollard $p-1$ method (and rediscovered in 1988 using ECM).

F_{12} has at least seven prime factors, spoiling a "conjecture" based on the observation that F_n has exactly $n-6$ prime factors for $8 \le n \le 11$.

F_{13}

It is known that

$$\begin{aligned} F_{13} \quad &= \quad 2710954639361 \cdot \\ & \quad\quad 2663848877152141313 \cdot \\ & \quad\quad 3603109844542291969 \cdot \\ & \quad\quad 319546020820551643220672513 \cdot c_{2391} \ , \end{aligned}$$

where the 13-digit factor was found by Hallyburton and Brillhart (1975), and the two 19-digit factors were found by Crandall (1991).

I found the 27-digit factor in June 1995, using ECM on an IBM PC equipped with a Dubner Cruncher board.

F_{14}

$F_{14} = c_{4933}$ is composite, but no nontrivial factors are known. The smallest prime factor probably has at least 30 decimal digits.

A new factor of F_{16}

$$F_{16} = 825753601{\cdot}188981757975021318420037633{\cdot}c_{19694}$$

where the 9-digit factor was found by Selfridge (1953), and the 27-digit factor was found in December 1996 by Brent, Crandall, Dilcher and Van Halewyn (BCDH) using ECM. For details, see our report [5].

A new factor of F_{15}

$$F_{15} = 1214251009 \cdot 2327042503868417 \cdot c_{9840},$$

where the 13- and 16-digit prime factors were found by Kraitchik (1925) and Gostin (1987). On July 3, 1997 BCDH found a 33-digit factor

$$p_{33} = 168768817029516972383024127016961$$

using ECM. The quotient is c_{9808}.

Factorization of F_{10}: some details

ECM was implemented on a Fujitsu VP100 in March 1988. The program was soon successful in completing the factorization of F_{11}, but had no success with other Fermat numbers, apart from rediscovering known factors. The VP100 was upgraded to a VP2200 in 1991.

In September 1994 we started running a similar program on one or two 60 Mhz SuperSparc processors. In July 1995 six more 60 Mhz SuperSparc processors became available for a limited period. We attempted to factor F_{10} on all eight SuperSparcs.

The p_{40} factor of F_{10} was found by a run which started on Oct 14 and finished on Oct 20, 1995. The run tried 10 curves with $B_1 = 2000000$ in about 114 hours of CPU time.

All the computations were performed at the Australian National University, Canberra.

Summary of F_{10} runs

In Table 3, F is an estimate of the expected number of times that the factor p_{40} should be found with the given B_1 and number of curves.
E is an estimate of the efficiency compared to the optimal choice of $B_1 \simeq 3400000$.
The last row of the table gives totals (for number of curves and F) and weighted means (for B_1 and E).

Table 3: ECM runs on F_{10}

B_1	curves	F	E	machine(s) and dates
6×10^4	2000	0.0010	0.14	VP100, Mar 1988 – Nov 1990
2×10^5	17360	0.0910	0.42	VP2200, Aug 1991 – Aug 1995
5×10^5	700	0.0152	0.69	Sparc × 2, Sep 1994 – Jul 1995
10^6	480	0.0262	0.87	Sparc × 8, Jul 1995 – Aug 1995
2×10^6	900	0.1100	0.98	Sparc × 8, Aug 1995 – Oct 1995
2.9×10^5	21440	0.2434	0.63	

The Computational Work

Each curve on a 60 Mhz SuperSparc takes about $5.7 \times 10^{-6} B_1$ hours of CPU time. If a 60 Mhz SuperSparc is counted as a 60-Mips machine, then our computation took about 240 Mips-years. This is comparable to the 340 Mips-years estimated for sieving to factor F_9 by SNFS. (SNFS has since been improved, so the 340 Mips-years could now be reduced by an order of magnitude.) A 130-digit number, RSA130, took 500 Mips-years by GNFS (May 1996).

Since the inner loops of our programs use floating-point arithmetic, Mflops are a more appropriate measure than Mips. The VP2200/10 is rated at 1250 Mflop (peak performance). If our factorization of F_{10} had been performed entirely on the VP2200, it would have taken about 6 weeks of machine time, or 140 Mflop-years. Cryptographers should note that this amounts to about 75 minutes on a 1 Teraflop machine.

Multiplications

The number of multiplications (mod N) is a machine-independent measure of the work to factor N. Each curve takes about $22.9\, B_1$ such multiplications.

Overall, our factorization of F_{10} took 1.4×10^{11} multiplications (mod N), where $N = c_{291}$. Table 1 predicts 3.3×10^{11} with the optimal choice of parameters.

Numbers mod c_{291} were represented with 38 digits and base 2^{26} (on the VP100/VP2200) or with 41 digits and base 2^{24} (on the Sparc), so each multiplication (mod N) took more than 10^4 floating-point operations.

The Group Order

The successful elliptic curve leading to the factorization of F_{10} had order

$$\begin{aligned} g &= p_{40} + 1 - 3674872259129499038 \\ &= 2^2 \cdot 3^2 \cdot 5 \cdot 149 \cdot 163 \cdot 197 \cdot 7187 \cdot \\ &\quad\; 18311 \cdot 123677 \cdot 226133 \cdot 314263 \cdot 4677853\,. \end{aligned}$$

The probability that a random integer near $g/12$ has largest prime factor at most 4677853 and second-largest prime factor at most 314263 is about 5.8×10^{-6}. The phase 1 limit for the successful run was $B_1 = 2 \times 10^6$, but our program finds p_{40} with B_1 as small as 314263 if the same curve and starting point are used. (The largest factor 4677853 of g is caught by "phase 2" of ECM.)

The final word – F_9 again

In April 1997, using ECM on a 250 Mhz DEC alpha, we “rediscovered” the 49-digit factor of F_9:

7455602825647884208337395736200454918783366342657

We used the equivalent of about 73,000 curves with $B_1 = 10^7$; the number of curves predicted is about 90,000. The group order for the lucky curve is

$2^2 \cdot 3^2 \cdot 5^2 \cdot 7 \cdot 331 \cdot 1231 \cdot 1289 \cdot 6277 \cdot 68147 \cdot 1296877 \cdot 9304783 \cdot 9859051 \cdot 44275577$

Of course, the 49-digit factor was already known, but it is interesting to see that it *could* have been found by ECM. Excluding this example, the current record for ECM is a 48-digit factor

662926550178509475639682769961460088456141816377

of $24^{121} + 1$, which I found on 8 October 1997.

How much computation was required ?

The factorization of F_9 by ECM required about 1.7×10^{13} multiplications mod N_9, where $N_9 = F_9/(\text{small factor})$.

Recall that the factorization of F_{10} required about 1.4×10^{11} multiplications mod N_{10}, where $N_{10} = F_{10}/(\text{known factors})$. Allowing for the sizes of N_9 and N_{10}, we see that the factorization of F_9 required about thirty times as much computation as the factorization of F_{10}.

However, *we saved a factor of about ten* by working mod p_{49} rather than mod N_9.

> “This is cheating”
>
> Arjen Lenstra, 4 Dec 1997

Acknowledgements

Thanks are due to the Australian National University Supercomputer Facility and to many people, including: John Brillhart, Richard Crandall, Bruce Dodson, Harvey Dubner, Wilfrid Keller, Donald Knuth, Arjen Lenstra, Hendrik Lenstra, Jr., Peter Montgomery, François Morain, John Pollard, John Selfridge, Robert Silverman, Hiromi Suyama, and Sam Wagstaff, Jr.

A preliminary version of this talk was presented at CANT’97, Sydney, 5 December 1997.

References

[1] A. O. L. Atkin and F. Morain, *Elliptic curves and primality proving,* Math. Comp. **61** (1993), 29–68. Programs available from `ftp://ftp.inria.fr/INRIA/ecpp.V3.4.1.tar.Z` .

[2] R. P. Brent, *Factorization of the eleventh Fermat number (preliminary report),* AMS Abstracts **10** (1989), 89T-11-73.

[3] R. P. Brent, *Large factors found by ECM,* Computer Sciences Laboratory, Australian National University, Dec. 1995 (and more recent updates). `ftp://nimbus.anu.edu.au/pub/Brent/champs.ecm` .

[4] R. P. Brent, *Factorization of the tenth and eleventh Fermat numbers,* Report TR-CS-96-02, Computer Sciences Laboratory, Australian National University, Feb. 1996, 25 pp. `ftp://nimbus.anu.edu.au:/pub/Brent/rpb161tr.dvi.gz` . Revision to appear in Math. Comp., 1998. See `ftp://nimbus.anu.edu.au:/pub/Brent/rpb161.dvi.gz` .

[5] R. P. Brent, R. E. Crandall and K. Dilcher, *Two new factors of Fermat numbers,* Report TR-CS-97-11, CSL, ANU, May 1997. `ftp://nimbus.anu.edu.au:/pub/Brent/rpb175tr.dvi.gz` . Revision (with C. Van Halewyn and one more factor) submitted for publication.

[6] R. P. Brent and J. M. Pollard, *Factorization of the eighth Fermat number,* Math. Comp. **36** (1981), 627-630.

[7] J. Brillhart, D. H. Lehmer, J. L. Selfridge, B. Tuckerman, and S. S. Wagstaff, Jr., *Factorizations of* $b^n \pm 1$, $b = 2, 3, 5, 6, 7, 10, 11, 12$ *up to high powers*, 2nd ed., Amer. Math. Soc., Providence, RI, 1988.

[8] C. Caldwell, *The Dubner PC Cruncher – a microcomputer coprocessor card for doing integer arithmetic*, review in J. Rec. Math. **25**(1), 1993.

[9] D. V. and G. V. Chudnovsky, *Sequences of numbers generated by addition in formal groups and new primality and factorization tests,* Adv. in Appl. Math. **7** (1986), 187–237.

[10] L. Euler, *Observationes de theoremate quodam Fermatiano aliisque ad numeros primos spectantibus,* Comm. Acad. Sci. Petropol. **6**, ad annos 1732–33 (1738), 103–107; Leonhardi Euleri Opera Omnia, Ser. I, vol. II, Teubner, Leipzig, 1915, 1–5.

[11] F. Landry, *Note sur la décomposition du nombre* $2^{64} + 1$ (Extrait), C. R. Acad. Sci. Paris **91** (1880), 138.

[12] A. K. Lenstra and H. W. Lenstra, Jr. (editors), *The development of the number field sieve,* Lecture Notes in Mathematics **1554**, Springer-Verlag, Berlin, 1993.

[13] A. K. Lenstra, H. W. Lenstra, Jr., M. S. Manasse, and J. M. Pollard, *The factorization of the ninth Fermat number,* Math. Comp. **61** (1993), 319–349.

[14] H. W. Lenstra, Jr., *Factoring integers with elliptic curves,* Annals of Mathematics (2) **126** (1987), 649–673.

[15] P. L. Montgomery, *Speeding the Pollard and elliptic curve methods of factorization,* Math. Comp. **48** (1987), 243–264.

[16] P. L. Montgomery, *An FFT extension of the elliptic curve method of factorization,* Ph. D. dissertation, Mathematics, University of California at Los Angeles, 1992.

[17] M. A. Morrison and J. Brillhart, *A method of factorization and the factorization of* F_7*,* Math. Comp. **29** (1975), 183–205.

[18] J. L. Selfridge, *Factors of Fermat numbers,* MTAC **7** (1953), 274–275.

[19] J. L. Selfridge and A. Hurwitz, *Fermat numbers and Mersenne numbers,* Math. Comp. **18** (1964), 146–148.

[20] A. M. Vershik, *The asymptotic distribution of factorizations of natural numbers into prime divisors,* Dokl. Akad. Nauk SSSR **289** (1986), 269–272; English transl. in Soviet Math. Dokl. **34** (1987), 57–61.

[21] H. C. Williams, *How was* F_6 *factored?,* Math. Comp. **61** (1993), 463–474.