

LARGEST KNOWN TWIN PRIMES AND SOPHIE GERMAIN PRIMES

KARL-HEINZ INDLEKOFER AND ANTAL JÁRAI

ABSTRACT. The numbers $242206083 \cdot 2^{38880} \pm 1$ are twin primes. The number $p = 2375063906985 \cdot 2^{19380} - 1$ is a Sophie Germain prime, i.e. p and $2p + 1$ are both primes. For $p = 4610194180515 \cdot 2^{5056} - 1$, the numbers p , $p + 2$ and $2p + 1$ are all primes.

In the first days of October, 1995, Harvey Dubner [4] found the largest known twin primes with 5129 decimal digits. (Our earlier twin prime record was $697053813 \cdot 2^{16352} \pm 1$, with 4932 decimal digits; see [5]). Some days before he found the largest known Sophie Germain prime, with 5089 decimal digits (see [8], p. 330), beating our records $157324389 \cdot 2^{16352} - 1$ and $470943129 \cdot 2^{16352} - 1$ with 4931 and 4932 decimal digits, respectively, found in February 1995 (first appearing here in print). In the same days our twin prime program was already running to find large twins in the range with approximately 11700 decimal digits. Because the total expected running time was 1500–2000 CPU days on Sun[®] workstations with SuperSPARC[®] processors of Texas Instruments with clock frequency from 33 MHz to 60 MHz, we decided to stop the search for some time and to use our new combined test program which is able to search for twin primes and Sophie Germain primes simultaneously, with half the running time. The idea of combining the two tests originated from Dubner [4]. Since our estimate for the running time of this new test to find twin primes and Sophie Germain primes with ≈ 5850 decimal digits was somewhat below 100 CPU days (total), we thought it would be more realistic to do this search first and later to continue with the combined test in the range with ≈ 11700 decimal digits. To test the combined sieve program and to do the new, combined sieving for twin and Sophie Germain primes needed some days, and until this we let the “old” program run. In the last day before stopping the old program we found the twin primes $242206083 \cdot 2^{38880} \pm 1$ (11713 digits). After this we stopped the old program, and started to run the new combined test, which found the largest known Sophie Germain prime $p = 2375063906985 \cdot 2^{19380} - 1$ (5847 digits). Large Sophie Germain primes p congruent to 3 mod 4 lead to large composite Mersenne numbers. So $2^p - 1$ for $p = 2375063906985 \cdot 2^{19380} - 1$ is the largest known composite Mersenne number.

In an e-mail message Harvey Dubner suggested looking for primes p for which $p + 2$ and $2p + 1$ are primes, too. Our new combined test program was able to do this without any changes, so we also have done a search in this direction and found

Received by the editor April 7, 1997 and, in revised form, February 5, 1998.
1991 *Mathematics Subject Classification*. Primary 11-04; Secondary: 11A41.
[®]Sun is a registered trademark of Sun Microsystems Inc.
[®]SPARC is a registered trademark of SPARC International.

the prime $p = 4610194180515 \cdot 2^{5056} - 1$ (1535 digits) for which $p + 2$ and $2p + 1$ are primes, too.

To our best knowledge the earlier records for twin primes were $1692923232 \cdot 10^{4020} \pm 1$ (4030 digits), $4655478828 \cdot 10^{3429} \pm 1$ (3439 digits) found by Dubner in 1994 and $1706595 \cdot 2^{11235} \pm 1$ (3389 digits) found by Parady, Smith, Zarantonello in 1990 (see [10]). Recently Tony Forbes informed us about his new twins $6797727 \cdot 2^{15328} \pm 1$, which will be published in *Mathematics of Computation*. The largest known Sophie Germain primes known earlier are, to our best knowledge, $1803301 \cdot 3003 \cdot 10^{4526} - 1$, $488964 \cdot 3003 \cdot 10^{4003} - 1$ and $5199545 \cdot 3003 \cdot 10^{3529} - 1$ found by Harvey Dubner in 1994, $15655515 \cdot 3003 \cdot 10^{2999} - 1$, $581436 \cdot 3003 \cdot 10^{2591} - 1$ and $7014 \cdot 3003 \cdot 10^{2110} - 1$ found by Dubner in 1993 (see [4]).

In our searches first we sieved out candidates divisible by small primes up to $2^{35} \dots 2^{41}$. Second, we used the strong probabilistic primality test (Miller–Rabin test, see Knuth [6], pp. 379–380) to find good prime candidates. All numbers which passed one probabilistic test passed all the following ones, too. The last step was to do exact tests: for numbers having the form $m \cdot 2^n - 1$ using a Lucasian test (see Riesel [9]) and for numbers having the form $m \cdot 2^n + 1$ using the test of Brillhart, Lehmer and Selfridge (see Ribenboim [7], pp. 37–39).

Our searches for large prime pairs and large Sophie Germain primes were performed at the University of Paderborn, Germany, in the frame of a project for parallel computing in computational number theory supported by the Heinz Nixdorf Institute, Paderborn. For the searches treated here we used a large part of the SuperSPARC CPU's of the “mathematical” net of the Department of Mathematics and Computer Science of the University of Paderborn.

We used the arithmetical routines developed in this project for fast parallel and sequential computations with very large numbers. After our earlier twin prime record our arithmetical routines were considerably modified to gain speed.

We wrote a special routine with linear running time for division by numbers $m2^e + \delta$, where the odd multiplier m fits into some computer words and the signed number δ fits into one computer word (32 bits). We used an arithmetic sequence $m = h_0 + Ph$, $h = 0, 1, \dots$, with an appropriate positive integer h_0 , where P is the product of small primes. For such an arithmetic sequence, it is easy to sieve with small primes, and there is no need to sieve with the prime divisors of P . It is better to choose an even product P to keep $m = h_0 + Ph$ odd for each h .

The low-level routines of the sieving, a large part of the multiplication routines for the tests and the modular reduction were written in assembly language using the properties of the SuperSPARC processor. Other parts of the programs were written in C. In this range the multiplication routine using Fast Fourier Transform over the complex number field (see Knuth [6], pp. 290–295) was found to be the fastest. For massively parallel applications, using Fermat number transform seems to be much better because of the smaller amount of data for communication.

The distribution of the program source lines (including comments) for the SuperSPARC version used in these searches is shown in Table 1.

The search for the largest twin prime pair consisted of the following steps:

1. The search was planned among the numbers in the arithmetical sequences $(3 + 30h)2^{38880} \pm 1$. We started with the 2^{27} nonnegative values of h below 2^{27} . The exponent 38880 was fixed during the calculations. With these choices, the multipliers $3 + 30h$ fit into one computer word, there is no need for sieving with 2,

TABLE 1

Program	C lines	Assembly lines
Classical algorithms for arithmetic	933	852
Karatsuba multiplication	0	2348
Fermat number transform multiplication	1471	2168
Complex FFT multiplication	2078	4520
Arithmetic for short and special modulus	0	3138
General modular arithmetic	4901	0
Sieving	1403	0
Probabilistic tests	1131	0
Exact tests	512	0

3 and 5, and the numbers have length $\leq 19 \cdot 2^{11}$ bits. Hence complex FFT with 2^{11} dimensional vectors can be used for multiplication.

2. Both the case $+1$ and the case -1 were sieved with factors from 7 up to $44000 \cdot 2^{25}$. This needed 80–100 CPU days. After sieving, 594866 candidates remained.

3. Using the strong probabilistic primality test, the candidates were tested, first the $+1$ case and then the -1 case, until a “probable twin prime pair” was found. Altogether 55440 candidates were tested. The CPU time of one probabilistic test was ≈ 6.7 minutes with 60 MHz clock speed. The number of SuperSPARC CPU’s we used varied between 10 and 25. The test needed approximately 20–30 days idle time. We found 98 probable primes, too.

4. The “probable twin prime pair” was tested with exact tests: the -1 case using a Lucasian type test and the $+1$ case using the test of Brillhart, Lehmer and Selfridge.

The combined search for a large Sophie Germain prime pair **or** a large twin prime pair consisted of the following steps:

1. The search was planned among the numbers in the arithmetical sequence $p = (h_0 + 30030h)2^{19380} - 1$ with $h_0 = 5775$. We started with the 2^{27} nonnegative values of h below 2^{27} . The exponent was fixed during the calculations. For simplicity, an exponent $e \equiv 0 \pmod{60}$ was chosen; in this case $2^e \equiv 1 \pmod{3 \cdot 5 \cdot 7 \cdot 11 \cdot 13}$. Now if $h_0 \equiv 0 \pmod{3 \cdot 5 \cdot 7 \cdot 11}$ then p , $p + 2$ and $2p + 1$ are not divisible by 3, 5, 7 and 11. Moreover if $h_0 \equiv 3, 6$ or $8 \pmod{13}$, then p , $p + 2$ and $2p + 1$ are not divisible by 13 and the Lucasian type test of Riesel [9] can be used with our favorite discriminant 13 for p and $2p + 1$. The constant $h_0 = 5775$ is one of the possible choices. The length of the numbers is not larger than $19 \cdot 2^{10}$ bits, which allows to use complex FFT with 2^{10} dimensional vectors.

2. The numbers p for which p or $p + 2$ or $2p + 1$ is divisible by a small prime from 17 up to $1000 \cdot 2^{25}$ were sieved out. This needed ≈ 2 CPU days. After sieving, 223401 candidates remained.

3. Using the strong probabilistic primality test, the candidates p were tested. Periodically, all primes p were tested for whether $p + 2$ or $2p + 1$ is a prime, too. This was repeated until the first “probable Sophie Germain prime” was found. Altogether 182488 candidates were tested. We found 597 probable primes, too. The running time of one probabilistic test was ≈ 1.5 minutes with 60 MHz clock

speed. The number of SuperSPARC CPU's we used varied between 10 and 25. The test needed approximately 12 days idle time.

4. The "probable Sophie Germain prime" was tested with exact tests: the p and $2p+1$ cases using a Lucasian type test and the $p+2$ case using the test of Brillhart, Lehmer and Selfridge.

The combined search for a prime p for which $p+2$ and $2p+1$ are primes, too, i.e. which is a large Sophie Germain prime **and** a large twin prime too, consisted of the following steps:

1. The search was planned among numbers in different arithmetical sequences

$$p = (5775 + 30030h)2^{5040} - 1,$$

$$p = (5775 + 30030h)2^{4980} - 1,$$

$$p = (21945 + 30030h)2^{5056} - 1.$$

In each case, we started with the 2^{28} nonnegative values of h below 2^{28} . The exponent was fixed during the calculations. The considerations in choosing these sequences are similar to above. The maximal amount of available core memory limited the sieve table to 2^{28} bits; hence we had to use more than one sequence.

2. The numbers p for which p or $p+2$ or $2p+1$ is divisible by a small prime from 17 up to $1000 \cdot 2^{25}$ in the first and the second case and up to $8000 \cdot 2^{25}$ in the third case were sieved out. This needed ≈ 2 CPU days in the first two cases and ≈ 14 days in the last case. After sieving, 449152, 448181 and 349954 candidates, respectively, remained.

3. Using the strong probabilistic primality test, the candidates $p+2$ were tested (from the last sequence only 215000). In the three parts 5452, 5646 and 2819 probable primes respectively were found. All primes $p+2$ were tested for whether p is a prime, too. In the three parts, 68, 60 and 31 probable twin primes, respectively, were found. Finally, these numbers were tested for whether $2p+1$ is a prime, too. The running time of one probabilistic test was ≈ 4.4 seconds with 60 MHz clock speed. We used 4 SuperSPARC CPU's, and in the second half of the search one UltraSPARC CPU. The test needed approximately 30 days idle time.

4. The only p which passed the probabilistic tests were tested with exact tests: the p and $2p+1$ cases using a Lucasian type test and the $p+2$ case using the test of Brillhart, Lehmer and Selfridge.

To estimate the limit for the search we had to calculate the expected number of twin primes, Sophie Germain primes, etc. To do this, we used a conjecture of Bateman and Horn [1]:

Conjecture. *Let f_1, f_2, \dots, f_s be irreducible polynomials, with integral coefficients and positive leading coefficients. If $Q(N)$ denotes the number of integers $1 < n < N$ such that $f_1(n), \dots, f_s(n)$ are all primes, then*

$$Q(N) \sim C_{f_1, \dots, f_s} \frac{1}{\deg(f_1) \cdots \deg(f_s)} \sum_2^N \frac{1}{(\ln(N))^s},$$

where

$$C_{f_1, \dots, f_s} = \prod_p \left(1 - \frac{w(p)}{p}\right) \left(1 - \frac{1}{p}\right)^{-s};$$

here $w(p)$ denotes the number of solutions x of the congruence

$$f_1(x) \cdots f_s(x) \equiv 0 \pmod{p}.$$

It is proved in [1] that the infinite product is always convergent. The simple idea behind the conjecture is that, by the prime number theorem, the probability that a large number n is a prime is $1/\ln(n)$. Thus, the probability that the numbers $f_1(n), \dots, f_s(n)$ are simultaneously prime is, if these events are independent,

$$\frac{1}{\ln f_1(n) \cdots \ln f_s(n)}.$$

However, the s -tuples $(f_1(n), \dots, f_s(n))$ are not random. In the definition of C_{f_1, \dots, f_s} , the number $1 - w(p)/p$ is the chance that none of the integers $f_1(n), \dots, f_s(n)$ is divisible by p , and $(1 - 1/p)^s$ is the chance that none of the integers of a random s -tuple is divisible by p ; hence it is reasonable to state that the probability that $f_1(n), \dots, f_s(n)$ are simultaneously prime is

$$\frac{C_{f_1, \dots, f_s}}{\ln f_1(n) \cdots \ln f_s(n)}.$$

Hence the expected number $Q(a, b)$ of n 's in $[a, b]$ for which $f_1(n), \dots, f_s(n)$ are simultaneously prime is

$$Q(a, b) \sim C_{f_1, \dots, f_s} \int_a^b \frac{du}{\ln f_1(u) \cdots \ln f_s(u)}.$$

In our cases the polynomials are linear functions; hence C_{f_1, \dots, f_s} can be easily calculated from the constants

$$C_s = \prod_{p>s} \frac{1 - s/p}{(1 - 1/p)^s},$$

$s = 1, 2, 3$. Clearly $C_1 = 1$, $C_2 = 0.66016\dots$, the twin prime constant, and $C_3 \approx 0.635$.

In our cases, the values of the functions f_1, \dots, f_s are very large; hence the logarithms are almost constants. So we used Simpson's rule for the approximation of the integral above.

The above heuristic suggests that, if we use the sieve with primes $A \leq p < B$, then the density of the prime s -tuples is increased by the factor

$$D_{f_1, \dots, f_s}^{A, B} = \prod_{A \leq p < B} \frac{1}{1 - \frac{w(p)}{p}},$$

and the number of candidates is decreased by this factor. In our cases these products are reduced to the products

$$D_s^{A, B} = \prod_{A \leq p < B} \frac{1}{1 - \frac{s}{p}}.$$

These products were calculated in the following way: for $p < L = 1\,000\,000$ we did the multiplication, and for the remaining part of the product we used the approximation $(\ln(B)/\ln(L))^s$. This approximation is estimated to have relative error below 0.1%.

As an example let us consider the search for the largest known twin prime. Here $f_1(h) = (3 + 30h)2^{38880} - 1$ and $f_2(h) = (3 + 30h)2^{38880} + 1$. If we plan the search for the interval $[a, b) = [0, 2^{27})$, then we expect

$$\begin{aligned} Q(0, 2^{27}) &\sim C_{f_1, f_2} \int_0^{2^{27}} \frac{du}{\ln f_1(u) \cdot \ln f_2(u)} \\ &\approx C_{f_1, f_2} \frac{2^{27}}{6} (0.1376769251 + 4 \cdot 0.1374695060 + 0.1374624404) \cdot 10^{-8} \\ &\approx C_{f_1, f_2} \cdot 0.1845532660 \end{aligned}$$

twin primes. Here

$$C_{f_1, f_2} = \left(1 - \frac{1}{2}\right)^{-2} \cdot \left(1 - \frac{1}{3}\right)^{-2} \cdot \left(1 - \frac{1}{5}\right)^{-2} \prod_{p>5} \frac{1 - 2/p}{(1 - 1/p)^2} = 20C_2 \approx 13.2032,$$

hence $Q(0, 2^{27}) \approx 2.4367$. The “twin prime density” $\approx 2.4367/2^{27} \approx 1.815482974 \cdot 10^{-8}$ is increased by the factor

$$\begin{aligned} D_{f_1, f_2}^{7, 44\,000 \cdot 2^{25}} &= \prod_{7 \leq p < 44\,000 \cdot 2^{25}} \frac{1}{1 - \frac{2}{p}} = D_2^{7, 44\,000 \cdot 2^{25}} \\ &\approx D_2^{7, 1\,000\,000} \left(\frac{\ln 44\,000 \cdot 2^{25}}{\ln 1\,000\,000}\right)^2 \\ &\approx 45.86172510 \cdot 4.113596977 \approx 188.6566536, \end{aligned}$$

if we sieve with primes in the interval $[A, B) = [7, 44\,000 \cdot 2^{25})$. Hence after sieving we expect $\approx 2^{27}/188.6566536 \approx 711439.1432$ remaining numbers and an increased “twin prime density” $\approx 188.6566536 \cdot 1.815482974 \cdot 10^{-8} \approx 3.425029425 \cdot 10^{-6}$. Testing 55 440 numbers, we expect $55\,440 \cdot 3.425029425 \cdot 10^{-6} \approx 0.1899$ twin primes.

Tables 2–4 compare the results of the sieves with the expected results.

Since the deviation in the first line of Table 2 between the found and expected number of the remaining candidates is large, we repeated the sieve for several smaller limits. The results are shown in Table 3.

As you see, in these cases there was a good accordance. So we guess that there was an error, probably in the parametrization of the sieve program, which we cannot reconstruct. Of course, this does not affect the validity of the twin prime record.

In Table 4 we compare the number of known primes, twins, etc., with the expected number.

TABLE 2

sequence	range	sieve limit	after sieve	expected
$(3 + 30h)2^{38880} \pm 1$	$0 \leq h < 2^{27}$	$44000 \cdot 2^{25}$	594 866	711 439
$(5775 + 30030h)2^{19380+1} \pm 1$	$0 \leq h < 2^{27}$	$1000 \cdot 2^{25}$	223 401	223 641
$(5775 + 30030h)2^{5040+1} \pm 1$	$0 \leq h < 2^{28}$	$1000 \cdot 2^{25}$	449 119	447 601
$(5775 + 30030h)2^{4980+1} \pm 1$	$0 \leq h < 2^{28}$	$1000 \cdot 2^{25}$	448 181	447 601
$(21945 + 30030h)2^{5056+1} \pm 1$	$0 \leq h < 2^{28}$	$8000 \cdot 2^{25}$	349 954	349 641

TABLE 3

sequence	range	sieve limit	after sieve	expected
$(3 + 30h)2^{38880} \pm 1$	$0 \leq h < 2^{27}$	2^{25}	1 859 586	1 860 213
$(3 + 30h)2^{38880} \pm 1$	$0 \leq h < 2^{27}$	2^{26}	1 719 334	1 719 871
$(3 + 30h)2^{38880} \pm 1$	$0 \leq h < 2^{27}$	2^{27}	1 594 321	1 594 834
$(3 + 30h)2^{38880} \pm 1$	$0 \leq h < 2^{27}$	2^{28}	1 482 166	1 482 950
$(3 + 30h)2^{38880} \pm 1$	$0 \leq h < 2^{27}$	2^{29}	1 381 749	1 382 442
$(3 + 30h)2^{38880} \pm 1$	$0 \leq h < 2^{27}$	2^{30}	1 290 965	1 291 815
$(3 + 30h)2^{38880} \pm 1$	$0 \leq h < 2^{27}$	2^{31}	1 208 806	1 209 816
$(3 + 30h)2^{38880} \pm 1$	$0 \leq h < 2^{27}$	2^{32}	1 134 255	1 135 383
$(3 + 30h)2^{38880} \pm 1$	$0 \leq h < 2^{27}$	2^{33}	1 066 708	1 067 615
$(3 + 30h)2^{38880} \pm 1$	$0 \leq h < 2^{27}$	2^{34}	1 004 800	1 005 738
$(3 + 30h)2^{38880} \pm 1$	$0 \leq h < 2^{27}$	2^{35}	947 738	949 087
$(3 + 30h)2^{38880} \pm 1$	$0 \leq h < 2^{27}$	2^{36}	895 968	897 093

TABLE 4

sequence	tested	prime	exp.	twin	exp.	S. G.	exp.
$(3 + 30h)2^{38880} \pm 1$	55440	99	102.6	1	0.1899	–	–
$(5775 + 30030h)2^{19380+1} \pm 1$	182 488	598	585.3	0	1.878	1	1.877
$(5775 + 30030h)2^{5040+1} \pm 1$	449 119	5452	5510	68	67.6	0	0.829
$(5775 + 30030h)2^{4980+1} \pm 1$	448 181	5646	5564	60	69.1	0	0.857
$(21945 + 30030h)2^{5056+1} \pm 1$	215 000	2819	2855	31	37.9	1	0.526

ACKNOWLEDGMENTS

We thank very much Prof. Buchmann and Dr. Papanikolaou for confirming the validity of the new twin prime record. We thank Texas Instruments Inc. for important information about the SuperSPARC processor. We thank Mr. Heinz-Georg Wassing for running our probabilistic test program on several workstations. Mr. András Baligács wrote the C parts, binding together the different high-speed routines into a modular arithmetic package, and the new version of the strong probabilistic primality test used in the last two searches. Mr. Béla Almási took part in debugging the modular shift for the Fermat number transform and in the debugging of some parts of the implementation of the Karatsuba method. Other persons also took part in our project, but their work was not used in these searches.

REFERENCES

1. P. T. Bateman and R. A. Horn, *A heuristic asymptotic formula concerning the distribution of prime numbers.*, Math. Comp. **16** (1962), 363–367. MR **26**:6139
2. P. T. Bateman and R. A. Horn, *Primes represented by irreducible polynomials in one variable.*, Theory of Numbers, Proc. Symp. Pure Math. Vol. VIII, Amer. Math. Soc., Providence, R. I., 1965, pp. 119–132. MR **31**:1234
3. C. Caldwell, *The largest known primes. A regularly updated list available on request*, e-mail:caldwell@UTmartn.bitnet (1995).
4. H. Dubner, *Large Sophie Germain Primes*, Math. Comp. **65** (1996), 393–396. MR **96d**:11008

5. K.-H. Indlekofer, A. Jári, *Largest known twin primes*, Math. Comp. **65** (1996), 427–428. MR **96d**:11009
6. D. E. Knuth, *The Art of Computer Programming, Vol. 1–3. Second Edition*, Addison-Wesley, 1981. MR **83i**:68003
7. P. Ribenboim, *The Book of Prime Number Records*, Springer-Verlag, 1989. MR **90g**:11127
8. P. Ribenboim, *The New Book of Prime Number Records*, Springer-Verlag, 1996. MR **96k**:11112
9. H. Riesel, *Lucasian criteria for the primality of $N = h \cdot 2^n - 1$* , Math. Comp. **23** (1969), 869–875. MR **41**:6773
10. B. K. Parady, J. F. Smith, S. E. Zarantonello, *Largest known twin primes*, Math. Comp. **55** (1990), 381–382. MR **90j**:11013

UNIVERSITÄT GH PADERBORN, FB 17, D-33095 PADERBORN, GERMANY

E-mail address: k-heinz@uni-paderborn.de

E-mail address: jarai@uni-paderborn.de