SUPERCOMPUTERS AND THE RIEMANN ZETA FUNCTION

A. M. Odlyzko

AT&T Bell Laboratories
Murray Hill, New Jersey 07974

ABSTRACT

The Riemann Hypothesis, which specifies the location of zeros of the Riemann zeta function, and thus describes the behavior of primes, is one of the most famous unsolved problems in mathematics, and extensive efforts have been made over more than a century to check it numerically for large sets of cases. Recently a new algorithm, invented by the speaker and A. Schönhage, has been implemented, and used to compute over 175 million zeros near zero number 10^{20}. The new algorithm turned out to be over 5 orders of magnitude faster than older methods. The crucial ingredients in it are a rational function evaluation method similar to the Greengard-Rokhlin gravitational potential evaluation algorithm, the FFT, and band-limited function interpolation. While the only present implementation is on a Cray, the algorithm can easily be parallelized.

1. Introduction and History

The Riemann zeta function is defined for complex values of s with $\mathrm{Re}(s) > 1$ by

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} , \qquad (1.1)$$

and it can be continued analytically to the entire complex plane with the exception of $s=1$, where it has a first order pole. The zeta function was actually first defined by Euler in the first half of the eighteenth century [2, 26]. Euler's work was motivated by the problem of evaluating

$$\sum_{n=1}^{\infty} \frac{1}{n^2} ,$$

(which is $\zeta(2)$ in our notation), which was posed in the seventeenth century by Mengoli. Euler eventually showed that $\zeta(2) = \pi^2/6$, but some of his initial efforts went into numerically evaluating $\zeta(2)$ and involved development of what is now called the Euler-Maclaurin formula. This was the first of many connections between numerical analysis and the zeta function.

Euler found some important relations between the zeta function and primes, but it was only Riemann in the 1850's who showed the full extent of these connections. In particular, Riemann showed that the distribution of primes is determined by the *nontrivial zeros* of $\zeta(s)$; i.e., those zeros of $\zeta(s)$ that lie in the strip $0 < \mathrm{Re}(s) < 1$ (called the critical strip). Riemann further conjectured that all of the nontrivial zeros lie on the *critical line*, $\mathrm{Re}(s) = 1/2$. This conjecture, known as the *Riemann Hypothesis* (RH), is probably the most important unsolved problem in mathematics. (For general background and history, see [9].)

Given the importance of the RH, it is not surprising that many attempts have been made to check it numerically for various sets of zeros. Since $\overline{\rho}$ is a zero of $\zeta(s)$ whenever ρ is, we consider only the nontrivial zeros ρ with $\mathrm{Im}(\rho) > 0$, and number them ρ_1, ρ_2, ..., so that

$$0 < \mathrm{Im}(\rho_1) \leq \mathrm{Im}(\rho_2) \leq \cdots .$$

(In case of multiple zeros, they have to be counted according to their multiplicity.) The RH is the equivalent to the claim that each $\rho_n = 1/2 + i\,\gamma_n$, where γ_n is real. A partial list of the most important numerical verifications of the RH is given in Table 1. The number in parentheses refers to the date of publication, and the n entry denotes that the RH has been checked for the first n zeros, so that $\gamma_1, \gamma_2, \ldots, \gamma_n$ are real. This means not just that the γ_n are real to within some bound, like 10^{-20}, but that they are actually real. It is possible to establish this rigorously, assuming of course that the programs and hardware are correct, due to some special properties of the zeta function. For a fuller history of these computations, see [9, 21].

There were two main innovations that were introduced in the computations listed in Table 1. One was a change in technology; starting with the work of A. Turing, all computations were performed on electronic digital computers, as opposed to paper and pencil or the electromechanical devices of earlier workers, with the latest result, the verification of the RH for the first 1.5×10^9 zeros by van de Lune, te Riele, and Winter [17] taking about 1500 hours on a Cyber 205. The other innovation was in algorithms, and was at least as important. The computations of Gram, Backlund, and Hutchinson used a method for computing the zeta function that is based on the Euler-Maclaurin formula. It is effective, but not very efficient, since to compute a single value of $\zeta(1/2 + it)$ with this method takes on the order of t steps. (The n-th zero $1/2 + i\,\gamma_n$ has $\gamma_n \sim 2\pi n/(\log n)$, with the 1.5×10^9-th zero having $\gamma_n \approx 5 \times 10^8$.) Starting with the work of Titchmarsh, a much more efficient algorithm has been employed. As it turns out, this was not a new invention, but rather a discovery by C. L. Siegel [25] in Riemann's unpublished notes of a method now known as the Riemann-Siegel formula. This method was already used by Riemann to compute at least the first few zeros of the zeta function, and the results of these computations may have been a crucial factor in stimulating him to make the RH.

The Euler-Maclaurin formula is a very general technique that applies to many summations. In contrast, the Riemann-Siegel formula relies on special properties of the zeta function. It is much more efficient in that only about $\sqrt{t}$ operations are needed to evaluate $\zeta(1/2 + it)$, in contrast to the roughly t operations needed by the Euler-Maclaurin method. For computations near the 10^9-th zero, the gain in efficiency is about 4 orders of magnitude.

Recently A. Schönhage and the author [19, 24] have invented an even faster method for computing large sets of zeros of the zeta function. In time roughly $\sqrt{T}$, it enables one to compute up to B zeros, where $B \lesssim \sqrt{T}$ is approximately the storage capacity on moderately fast devices (such as magnetic disks) that is available. This algorithm has been implemented on a Cray X-MP, and used to compute the zeros listed in Table 2. (The $N = 10^{12}$ entry, for example, means that 1,592,196 zeros were computed, starting with zero number $10^{12} - 6032$ and ending with zero number $10^{12} + 1586163$.) An earlier implementation of the standard Riemann-Siegel formula method [20] calculated 10^5 zeros near zero number 10^{12} in about 15 hours on the Cray X-MP. Since $\gamma_n \approx 2.6\times10^{11}$ for $n = 10^{12}$, and $\gamma_n \approx 1.5\times10^{19}$ for $n = 10^{20}$, this indicates that computing 6.5×10^7 zeros near zero number 10^{20} would have required about

$$\frac{6.5\times10^7}{10^5} \cdot \left[\frac{1.5\times10^{19}}{2.6\times10^{11}}\right]^{1/2} \approx 5\times10^6$$

as long, or about 7.5×10^7 hours, if one were to use the Riemann-Siegel method. On the other hand, one run of the new algorithm determined those 6.5×10^7 zeros in about 250 hours (of otherwise idle time), which is about 3×10^5 times faster.

A full description of the implementation of the new algorithm and the results that have been obtained with it are presented in [22]. The next section sketches some of the main features of the new algorithm. Now we will mention some of the main results from [22]. First of all, no counterexamples to the RH were found! Further, extensive statistics about the zeros were collected. The main purpose for this (and, in fact, for the whole project) was to obtain more insight into the behavior of the zeros of the zeta function. The fact that the RH holds for the first 1.5×10^9 zeros, as well as for over 3×10^8 zeros at heights as far up as the 2×10^{20}-th zero, is not entirely convincing evidence in favor of the RH, since there are many examples of number theoretic conjectures that have been shown to be false, but where the smallest known counterexamples are very high up (up to $\exp(\exp(65))$ in one case [23]). Thus it is important to examine more sophisticated heuristics as to whether the RH is likely to be true, and computational results are very valuable in this regard. (See [21] for a full discussion.) In particular, there is a set of conjectures, going back to Hilbert and Pólya, which imply the RH, and, when combined imaginatively with some observations of Dyson and H. Montgomery might lead one to expect that the zeros of the zeta function behave like eigenvalues of random Hermitian matrices. More precisely, if we let

$$\delta_n = (\gamma_{n+1} - \gamma_n) \frac{\log(\gamma_n/(2\pi))}{2\pi}$$

be the normalized gap between consecutive zeros of the zeta function, then the conjecture is that the δ_n behave like the gaps between successive normalized eigenvalues in the Gaussian unitary ensemble (*GUE*), which has been investigated very extensively by physicists [5, 6, 18]. In the case of the GUE, it is known that the δ_n have a particular (complicated) distribution. Fig. 1 compares the distribution of the δ_n for the first 10^6 zeros of the zeta function (scatterplot) to the GUE distribution (solid line). As can be seen, the agreement is only moderate. On the other hand, Fig. 2 shows a similar graph based on about 10^6 zeros near zero number 2×10^{20}, and here the agreement as almost perfect. This might be taken as evidence that the highly speculative Hilbert and Pólya conjectures are indeed correct, and that the RH is true.

The papers [21, 22] contain many more statistical studies of the zeros and discussions of what they might mean for the truth of the RH. At this point we will only mention that there are some interesting conjectural connections between zeta function zeros and quantum chaos [4].

2. Algorithms and Implementations

The new algorithm is based on the Riemann-Siegel formula. In the standard implementation, almost all of the computing time is devoted to evaluating sums of the form

$$F(t) \;=\; 2 \sum_{k=2}^{k_1} k^{-1/2} \; \exp(it \log k) \;, \tag{2.1}$$

where

$$k_1 \;=\; \left\lfloor (t/(2\pi))^{1/2} \right\rfloor . \tag{2.2}$$

Once $F(t)$ (or, actually, the $F(t)\exp(-i\theta(t))$, where $\theta(t)$ is a certain easy to evaluate function) is computed, $\zeta(1/2 + it)$ is very easy to obtain.

For values of t near zero number 10^{20}, $k_1 \approx 1.5\times10^9$, so the sum in (2.1) is very long, and it is desirable to avoid evaluating that sum term-by-term for each value of t that one might wish to investigate. The first crucial ingredient in the new algorithm is to observe that if one can precompute $F(t)$ at an evenly spaced grid of points, $t = T$, $T+\delta$, $T + 2\delta$, ..., $T + (R-1)\delta$, which is dense enough, then one can compute $F(t)$ at any point in that interval from the precomputed values. The initial method that was proposed for doing this involved Taylor series expansions [24]. However, a much more efficient method can be obtained by using band-limited function interpolation, and it was used in the implementation [22]. It relies on the fact, well known to communication engineers and complex analysts, that a band-limited function $G(t)$,

$$G(t) \;=\; \int_{-\tau}^{\tau} g(x) e^{ixt} \, dx \;, \tag{2.3}$$

is determined by its samples at the points $n\pi/\tau$, n running through the integers, provided only that it satisfies some mild growth conditions [7, 14, 15]. In fact, one has the classical ‘‘cardinal series’’

$$G(t) \;=\; \sum_{n=-\infty}^{\infty} G\left(\frac{n\pi}{\tau}\right) \; \frac{\sin(\tau t - n\pi)}{\tau t - n\pi} \;. \tag{2.4}$$

This series is not suitable for zeta function computations because of slow convergence, but it is possible to obtain a similar series with the $(\sin u)/u$ kernel replaced by one that drops off much more rapidly, provided one is willing to compute samples $G(n\pi/\beta)$ for some $\beta > \tau$. As is shown in [22], this yields a very efficient method for computing $G(t)$ from the $G(n\pi/\beta)$.

We next consider the problem of evaluating $F(t)$, $F(T+\delta), \ldots, F(T+(R-1)\delta)$. We need to find a method that requires substantially fewer than Rk_1 operations to compute all these values. The first step is to apply the discrete Fourier transform. We let

$$h_k = \sum_{j=0}^{R-1} F(T+j\delta) \exp(2\pi ij/R), \quad 0 \le k < R . \qquad (2.5)$$

If R is chosen appropriately (say $R = 2^r$), then the $F(T+j\delta)$ can be computed efficiently from the h_k using the FFT. (In the current implementation, the FFT routines that were used were those of Bailey [3]. They consume a negligible fraction of the total computing time.) Thus we reduce to the problem of computing all the h_k efficiently.

To compute the h_k, which consumes most of the time, we note that when we put in the definition (2.1) of $F(t)$ into (2.5), we obtain $h_k = h(\exp(2\pi ik/R))$, where

$$h(z) = \sum_{k=2}^{k_1} \frac{a_k}{z-b_k} ,$$
$$b_k = \exp(i\delta \log k) , \qquad (2.6)$$

and the a_k are certain complex constants. The straightforward term-by-term evaluation of the series (2.6) takes $\approx k_1$ steps for each k, which takes $\approx Rk_1$ steps in all, and offers no saving over the standard evaluation of the Riemann-Siegel formula. However, there is a way to compute all of h_k simultaneously in $O(k_1(\log k_1)^2)$ operations (for $R \lesssim k_1$) [24]. This method relies on Taylor series expansions of partial sums of the sum in (2.1). It is explained in detail in [22, 24].

The method of [24] for evaluating the rational function $h(z)$ of (2.6) simultaneously is very similar in spirit (although not in form) to the Greengard-Rokhlin algorithm for evaluating gravitational and Coulomb potentials [11] (see also [1,8,12,13]), and it can be adapted to handle such problems. For gravitational potential problems, it appears that the algorithms of [11] and [24] are of comparable efficiences. However, as is discussed in [22], it seems likely that for computing zeros of the zeta function, the Greengard-Rokhlin algorithm could turn out to be faster, due to the special structure of the zeta function problem and space limitations.

Although the current implementation of the new algorithm is over five orders of magnitude faster than the standard Riemann-Siegel method, substantial further improvements are possible. One could come from using the Greengard-Rokhlin algorithm. Others could come from more careful coding, including the use of assembly language. (Right now all the programs are written in Fortran, and most of the important loops are vectorized by the Cray Fortran compiler, with the most important loop operating at 100 Mflops/sec, but given the mix of arithmetic operations, this could probably be increased by using assembly language and rearranging the computation.) Perhaps the largest improvement that can be obtained would come from use of larger disk storage. Main memory is not a major constraint (the basic program uses under 7 Mbytes), but lack of temporary disk storage is. By using larger devices (which, if one were willing to do extra work, could even be magnetic tapes or optical disks) one could obtain substantially higher speeds.

While the current program has only been implemented on a Cray X-MP (and its has also been tested on a Cray-2), it can easily be parallelized. There are already some parallel implementations of the Greengard-Rokhlin algorithm [10, 16, 27, 28], and so similar ones could be carried out for the zeta function algorithm.

One interesting feature of the new computations is the problem of accuracy. The current computations are not completely rigorous due to incomplete control over roundoff errors. This problem is due to the extremely large heights at which the computations are being carried out and to the 64-bit word size on the Cray. Basically, double precision arithmetic is being used, but even that is not sufficient. Larger word sizes (72 or 80 bits) would be very helpful here!

REFERENCES

(1) J. Ambrosiano, L. Greengard, and V. Rokhlin, The fast multipole method for gridless particle simulations, *Computer Physics Comm. 48* (1988), 117-125.

(2) R. Ayoub, Euler and the zeta function, *Amer. Math. Monthly 81* (1974), 1067-1086.

(3) D. H. Bailey, A high-performance FFT algorithm for vector supercomputers, *Intern. J. Supercomputer Appl. 2, no. 1,* (1988), 82-87.

(4) M. V. Berry, Quantum chaology, *Proc. Royal Soc. London A 413* (1987), 183-198.

(5) O. Bohigas and M.-J. Giannoni, Chaotic motion and random matrix theories, pp. 1-99 in *Mathematical and Computational Methods in Nuclear Physics*, J. S. Dehesa, J. M. G. Gomez, and A. Polls, eds., Lecture Notes in Physics #209, Springer 1984.

(6) T. A. Brody, J. Flores, J. P. French, P. A. Mello, A. Pandey, and S. S. M. Wong, Random-matrix physics: spectrum and strength fluctuations, *Rev. Modern Physics 53* (1981), 385-479.

(7) P. L. Butzer, W. Splettstösser, and R. L. Stens, The sampling theorem and linear prediction in signal analysis, *Jber. d. Deutschen Math.-Verein. 90* (1988), 1-70.

(8) J. Carrier, L. Greengard, and V. Rokhlin, A fast adaptive algorithm for particle simulations, *SIAM J. Stat. Sci. Comp. 9* (1988), 669-686.

(9) H. M. Edwards, *Riemann's Zeta Function*, Academic Press, 1974.

(10) L. Greengard and W. D. Gropp, A parallel version of the fast multipole algorithm, to be published.

(11) L. Greengard and V. Rokhlin, A fast algorithm for particle simulations, *J. Computational Phys. 73* (1987), 325-348.

(12) L. Greengard and V. Rokhlin, Rapid evaluation of potential fields in three dimensions, to be published.

(13) L. Greengard and V. Rokhlin, On the efficient implementation of the fast multipole algorithm, to be published.

(14) J. R. Higgins, Five short stories about the cardinal series, *Bull. Amer. Math. Soc. 12* (1985), 45-89.

(15) A. J. Jerri, The Shannon sampling theorem - its various extensions and applications: a tutorial review, *Proc. IEEE 65* (1977), 1565-1596.

(16) J. Katzenelson, Computational structure of the N-body problem, *SIAM J. Stat. Sci. Comp.*, to appear.

(17) J. van de Lune, H. J. J. te Riele, and D. T. Winter, On the zeros of the Riemann zeta function in the critical strip. IV., *Math. Comp. 46* (1986), 667-681.

(18) M. L. Mehta, *Random Matrices*, Academic Press, 1967.

(19) A. M. Odlyzko, New analytic algorithms in number theory, pp. 466-475 in *Proc. Intern. Congress Math. 1986*, Amer. Math. Soc. 1987.

(20) A. M. Odlyzko, On the distribution of spacings between zeros of the zeta function, *Math. Comp. 48* (1987), 273-308.

(21) A. M. Odlyzko, Zeros of the Riemann zeta function: Conjectures and computations, manuscript in preparation.

(22) A. M. Odlyzko, The 10^{20}-th zero of the Riemann zeta function and 175 million of its neighbors, manuscript in preparation.

(23) A. M. Odlyzko and H. J. J. te Riele, Disproof of the Mertens conjecture, *J. reine angew. Math. 357* (1985), 138-160.

(24) A. M. Odlyzko and A. Schönhage, Fast algorithms for multiple evaluations of the Riemann zeta function, *Trans. Amer. Math. Soc. 309* (1988), 797-809.

(25) C. L. Siegel, Über Riemanns Nachlass zur analytischen Zahlentheorie, *Quellen und Studien zur Geschichte der Math. Astr. Phys. 2* (1932), 45-80. Reprinted in C. L. Siegel, *Gesammelte Abhandlungen*, Springer, 1966, Vol. 1, pp. 275-310.

(26) A. Weil, Prehistory of the zeta-function, in *Number Theory, Trace Formulas, and Discrete Groups,* K. E. Aubert, E. Bombieri, and D. M. Goldfeld, eds., Proc. 1987 Selberg Symposium, Academic Press, 1988, pp. 1-9.

(27) Feng Zhao, An $O(N)$ algorithm for three-dimensional N-body simulations, MIT AI Lab. report # 995, October 1987.

(28) Feng Zhao and L. Johnsson, The parallel multipole method on the Connection Machine, paper in preparation.

Table 1. Numerical verifications of the Riemann Hypothesis for the first n zeros.

Investigator	n
Gram (1903)	10
Backlund (1914)	79
Hutchinson (1925)	138
Titchmarsh et al. (1936)	1,041
Turing (1953)	1,104
Lehmer (1956)	25,000
Meller (1958)	35,337
Lehman (1966)	250,000
Rosser et al. (1969)	3,500,000
Brent (1979)	81,000,001
van de Lune et al. (1986)	1,500,000,000

Table 2. Large computed sets of zeros of the Riemann zeta function.

N	number of zeros	index of first zero in set
10^{12}	1,592,196	$N - 6,032$
10^{14}	1,685,452	$N - 736$
10^{16}	16,480,973	$N - 5,946$
10^{18}	16,671,047	$N - 8,839$
10^{19}	16,749,725	$N - 13,607$
10^{20}	175,587,726	$N - 30,769,710$
2×10^{20}	101,305,325	$N - 633,984$

FIGURE CAPTIONS

Fig. 1. Probability density of the normalized spacings δ_n. Solid line: GUE prediction. Scatterplot: empirical data based on initial 1,000,000 zeros.

Fig. 2. Probability density of the normalized spacings δ_n. Solid line: GUE prediction. Scatterplot: empirical data based on 1,041,600 zeros near zero number 2×10^{20}.