On sums of primes

H. Riesel and R. C. Vaughan

1. Introduction

In this paper we prove the following

Theorem. *Every even natural number can be represented as a sum of at most eighteen primes.*

It follows at once that every natural number n with $n>1$ is a sum of at most nineteen primes. The previous best result of this kind is due to Deshouillers [2] who has twentysix in place of nineteen.

Let $N(x)$ denote the number of even numbers n not exceeding x for which n is the sum of at most two primes. Then it suffices to show that

$$N(x) > x/18 \quad (x \geqq 2), \tag{1.1}$$

for then the theorem will follow in the usual manner (for example as in § 6 of [7]).

The proof of (1.1) is divided into three cases according to the size of x. When $\log x \geqq 375$ we use the method described in § 7 of [7], but with an important modification that enables us to dispense altogether with the Brun–Titchmarsh theorem. When $\log x \leqq 27$ the inequality (1.1) is easy to establish. This leaves the intermediate region $27 < \log x < 375$. Here we develop a completely new argument, based partly on sieve estimates and partly on calculation.

2. Some constants

We give here a list of constants that arise in the proof together with estimates for their values. A detailed description of the more difficult calculations is given in § 10.

Let

$$\gamma_k = \lim_{n\to\infty} \left(\sum_{m=1}^{n} m^{-1}(\log m)^k - \frac{(\log n)^{k+1}}{k+1}\right). \tag{2.1}$$

Then it is well known that

(2.2) $$0.577215 < \gamma_0 < 0.577216, \quad -0.072816 < \gamma_1 < -0.072815.$$

In fact γ_0 and γ_1 are easily calculated by means of the Euler-Maclaurin summation formula.

Let

(2.3) $$C = 2\prod_{p>2} \frac{p(p-2)}{(p-1)^2},$$

the twin prime constant. Then

(2.4) $$1.320323 < C < 1.320324.$$

Define the multiplicative function g by taking

(2.5) $$g(p^k) = 0 \quad \text{when} \quad k > 3, \quad g(2) = 0, \quad g(4) = -3/4, \quad g(8) = 1/4$$

and

(2.6)

$$g(p) = \frac{4}{p(p-2)}, \quad g(p^2) = \frac{-3p-2}{p^2(p-2)}, \quad g(p^3) = \frac{2}{p^2(p-2)} \quad \text{when} \quad p > 2.$$

Let

(2.7) $$H(w) = \sum_{m=1}^{\infty} |g(m)| m^{-w}.$$

Then

(2.8) $$251.0127 < H\left(-\frac{1}{3}\right) < 251.0128.$$

Futher define

(2.9) $$A_0 = \sum_p \frac{\log p}{p(p-1)},$$

(2.10) $$A_1 = \frac{1}{2}\log 2 + 2A_0,$$

(2.11) $$A_2 = \sum_p \frac{8p^2-10p+4}{p^2(p-1)^2}(\log p)^2,$$

(2.12) $$A_3 = 4\gamma_0 + 2A_1,$$

(2.13) $$A_4 = \frac{1}{4}A_3^2 - 2\gamma_0^2 - 4\gamma_1 + \frac{1}{4}(\log 2)^2 - A_2.$$

Then

(2.14) $$6.023476 < A_3 < 6.023477, \quad 1.114073 < A_4 < 1.114074.$$

Let

$$A_5 = 3.282CH\left(-\frac{1}{3}\right), \tag{2.15}$$

$$A_6(\lambda) = 2A_3 - 4\log 2 - 2\log\lambda, \tag{2.16}$$

$$A_7(\lambda) = \frac{2}{3}\pi^2 + 4A_4 - 4A_3\log 2 + (\log\lambda)^2 - (2A_3 - 4\log 2)\log\lambda_0, \tag{2.17}$$

$$A_8(\lambda) = 8A_5\lambda^{1/6}, \tag{2.18}$$

$$A_9(\lambda) = 4A_3\lambda^{1/2}. \tag{2.19}$$

Then

$$8.463433 < A_6\left(\frac{3}{2}\right) < 8.463434, \tag{2.20}$$

$$-9.260623 < A_7\left(\frac{3}{2}\right) < -9.260622, \tag{2.21}$$

$$9310.076 < A_8\left(\frac{3}{2}\right) < 9310.077, \quad 29.50888 < A_9\left(\frac{3}{2}\right) < 29.50889. \tag{2.22}$$

Let

$$A_{10} = \Pi_{p>2}\left(1 + \frac{1}{p(p-1)}\right) \tag{2.23}$$

and

$$A_{11} = \Pi_{p>2}\left(1 + \frac{2p-1}{p(p-1)^2}\right). \tag{2.24}$$

Then $A_{10} = \frac{2\zeta(3)\zeta(2)}{3\zeta(6)} = \frac{105}{\pi^4}\zeta(3)$ and $\zeta(3)$ is readily estimated by means of the Euler–Maclaurin summation formula. Thus

$$1.295730 < A_{10} < 1.295731. \tag{2.25}$$

We also have

$$1.772431 < A_{11} < 1.772432. \tag{2.26}$$

Let

$$T(u) = \sum\sum_{3 \leq p_1 < p_2 \leq u} \Pi_{\substack{p \mid p_2 - p_1 \\ p>2}} \frac{p-1}{p-2}, \tag{2.27}$$

and define

$$s = \pi(u) - 1. \tag{2.28}$$

Then we have

$$T(u) < t \tag{2.29}$$

where $t=t(u)$ satisfies

(2.30) $$t(79) = 328.5614, \quad t(99989) = 80096031.$$

We also have

(2.31) $$s(79) = 21, \quad s(99989) = 9590.$$

3. The sieve estimate

The fundamental information concerning prime numbers that we use in the proof is embodied in Lemma 5 below. It is a refinement of Lemma 8 of Vaughan [7] and likewise follows from Corollary 1 of Montgomery and Vaughan [4]. The improved values for A in Lemma 5 are essential to our argument.

The principal term that arises from Corollary 1 of [4] is related to the sum

$$\sum_{q\leqq Q} \mu(q)^2 \prod_{p|q,\ p>2} \frac{2}{p-2}$$

and in turn this is related to the sum

$$\sum_{m\leqq x} \frac{d(m)}{m}.$$

The following lemma gives a good quantitative estimate for this latter sum.

Lemma 1. *When* $x>0$, *let*

(3.1) $$E(x) = \sum_{m\leqq x} \frac{d(m)}{m} - \frac{1}{2}(\log x)^2 - 2\gamma_0 \log x - \gamma_0^2 + 2\gamma_1.$$

Then

$$|E(x)| < 1.641 x^{-1/3}.$$

Proof. We have

$$\sum_{m\leqq x} \frac{d(m)}{m} = \sum_{m\leqq x^{1/2}} \frac{1}{m} \sum_{n\leqq x/m} \frac{2}{n} - \left(\sum_{m\leqq x^{1/2}} \frac{1}{m}\right)^2.$$

Let $B_1(y)=y-[y]-\frac{1}{2}$, $B_2(y)=\frac{1}{2}\left(y-[y]-\frac{1}{2}\right)^2$. Then the Euler–Maclaurin summation formula gives

$$\sum_{n\leqq y} \frac{1}{n} = \log y + \gamma_0 - \frac{1}{y} B_1(y) - y^{-2} B_2(y) + \int_y^\infty B_2(u) 2u^{-3}\, du$$

and

$$\sum_{m\leqq y} \frac{\log m}{m} = \frac{1}{2}\log^2 y + \gamma_1 - \frac{\log y}{y} B_1(y) + \frac{1-\log y}{y^2} B_2(y) - \int_y^\infty B_2(u) \frac{3-2\log u}{u^3}\, du.$$

Hence

$$E(x)=-\frac{2}{x}B_2(\sqrt{x})+\int_{\sqrt{x}}^{\infty}B_2(u)u^{-3}(6-4\log(ux^{-1/2}))\,du-D(\sqrt{x})^2-2\sum_{m\leqq\sqrt{x}}\frac{1}{m}D(x/m)$$

where

$$D(y)=\frac{1}{y}B_1(y)+y^{-2}B_2(y)-\int_y^{\infty}B_2(u)2u^{-3}\,du.$$

Clearly $-\frac{1}{2}\leqq B_1(u)<\frac{1}{2}$ and $0\leqq B_2(u)\leqq\frac{1}{8}$. Thus, for $x\geqq 1$,

$$E(x)\leqq\int_{\sqrt{x}}^{e^{\frac{3}{2}}\sqrt{x}}\frac{3-2\log(ux^{-1/2})}{4u^3}\,du+x^{-1/2}+\sum_{m\leqq\sqrt{x}}\frac{1}{m}\int_{x/m}^{\infty}\frac{du}{2u^3}$$

$$\leqq\left(\frac{1}{2}+\frac{1}{8}e^{-3}\right)x^{-1}+x^{-1/2},$$

and

$$E(x)\geqq-\frac{1}{4}x^{-1}+\int_{e^{\frac{3}{2}}\sqrt{x}}^{\infty}\frac{3-2\log(ux^{-1/2})}{4u^3}\,du-x^{-1/2}-\sum_{m\leqq\sqrt{x}}\frac{m}{4x^2}-\left(\frac{1}{2\sqrt{x}}+\frac{1}{8x}\right)^2$$

$$\geqq-x^{-1/2}-\frac{3}{4}x^{-1}+\left[\frac{3-2\log(ux^{-1/2})}{-8u^2}\right]_{e^{\frac{3}{2}}\sqrt{x}}^{\infty}-\int_{e^{\frac{3}{2}}\sqrt{x}}^{\infty}\frac{du}{4u^3}-\frac{1}{8x^{3/2}}-\frac{1}{64x^2}$$

$$=-x^{-1/2}-\left(\frac{3}{4}+\frac{1}{8}e^{-3}\right)x^{-1}-\frac{1}{8}x^{-3/2}-\frac{1}{64}x^{-2}.$$

Therefore, for $x\geqq 2$ we have

$$|E(x)|x^{1/3}\leqq 2^{-1/6}+\left(\frac{3}{4}+\frac{1}{8}e^{-3}\right)2^{-2/3}+\frac{1}{8}2^{-7/6}+\frac{1}{64}2^{-5/3}<1.5.$$

When $1\leqq x<2$ we have

$$E(x)=1-\frac{1}{2}(\log x)^2-2\gamma_0\log x-\gamma_0^2+2\gamma_1.$$

Moreover $E(x)$ is strictly decreasing on $(1, 2)$, $E(1)=1-\gamma_0^2+2\gamma_1<0.53$ and $E(2-)=1-\frac{1}{2}(\log 2)^2-2\gamma_0\log 2-\gamma_0^2+2\gamma_1>-0.52$. Hence

$$|E(x)|x^{1/3}<0.67.$$

When $0<x<1$ we have

$$E(x)=-\frac{1}{2}(\log x)^2-2\gamma_0\log x-\gamma_0^2+2\gamma_1.$$

Let

$$F(x)=-\left(\frac{1}{2}(\log x)^2+2\gamma_0\log x+\gamma_0^2-2\gamma_1\right)x^{1/3}.$$

Then $F(x)\to 0-$ as $x\to 0+$, $F(x)\to -\infty$ as $x\to +\infty$ and $F(x)$ has a local minimum at x_- and a local maximum at x_+ where $x_\pm$ is given by

$$\log x_\pm = -2\gamma_0 - 3 \pm (2\gamma_0^2 + 9 + 4\gamma_1)^{1/2}.$$

Moreover $0 < x_- < x_+ < 1$, $F(x_\pm) = 3(\log x_\pm + 2\gamma_0)x_\pm^{1/3}$, $F(1-) = 2\gamma_1 - \gamma_0^2 > -0.48$ and

$$-1.641 < F(x_-) < 0 < F(x_+) < 0.13.$$

Hence

$$|E(x)|x^{1/3} = |F(x)| < 1.641.$$

Lemma 2. *Let*

$$S(y) = \sum_{q\le y} \mu(q)^2 \prod_{\substack{p|q \\ p>2}} \frac{2}{p-2}. \tag{3.2}$$

Then for $y \ge 1$,

$$|2CS(y) - (\log y)^2 - A_3 \log y - A_4| < A_5 y^{-1/3}.$$

Proof. Let g be as in (2.5) and (2.6) and define for $w > -\frac{1}{2}$

$$G(w) = \sum_{n=1}^\infty g(n) n^{-w}. \tag{3.3}$$

Then it is easily verified that when $w > 0$

$$\sum_{q=1}^\infty \frac{\mu(q)^2}{q^w} \prod_{\substack{p|q \\ p>2}} \frac{2}{p-2} = \zeta(w+1)^2 G(w). \tag{3.4}$$

Thus, be the identity theorem for Dirichlet series,

$$S(y) = \sum_m g(m) \sum_{n\le y/m} \frac{d(n)}{n}. \tag{3.5}$$

Therefore, by (3.1),

$$S(y) = \sum_m g(m) \left(\frac{1}{2}\left(\log\frac{y}{m}\right)^2 + 2\gamma_0 \log\frac{y}{m} + \gamma_0^2 - 2\gamma_1 + E\left(\frac{y}{m}\right)\right). \tag{3.6}$$

By (2.7) and Lemma 1,

$$\left|\sum_m g(m) E\left(\frac{y}{m}\right)\right| < 1.641 y^{-1/3} H\left(-\frac{1}{3}\right). \tag{3.7}$$

The main term in (3.6) is

$$\left(\tfrac{1}{2}(\log y)^2 + 2\gamma_0 \log y + \gamma_0^2 - 2\gamma_1\right) G(0) + (\log y + 2\gamma_0) G'(0) + \tfrac{1}{2} G''(0). \tag{3.8}$$

By (3.3), (2.5), (2.6) and (2.3),

$$G(0) = \frac{1}{2} \prod_{p>2} \left(1 + \frac{4p-2-3p+2}{p^2(p-2)}\right) = C^{-1}. \tag{3.9}$$

By (3.4), when $w>0$

$$G(w) = \zeta(w+1)^{-2}(1+2^{-w}) \prod_{p>2}\left(1+\frac{2}{p^w(p-2)}\right).$$

Hence

$$\frac{G'}{G}(w) = J(w)$$

where

$$J(w) = \frac{2\log 2}{2^{w+1}-1} - \frac{\log 2}{2^w+1} + \sum_{p>2}\left(\frac{2\log p}{p^{w+1}-1} - \frac{2\log p}{p^w(p-2)+2}\right).$$

Letting $w\to 0+$ gives, by (2.10) and (2.9),

(3.10) $$G'(0) = A_1 C^{-1}.$$

We also have

$$G''(w) = (J'(w)+J(w)^2)G(w)$$

and

$$J'(w) = \frac{2^w\log^2 2}{(2^w+1)^2} + \sum_p\left(\frac{2p^w(p-2)\log^2 p}{(p^w(p-2)+2)^2} - \frac{2p^{w+1}\log^2 p}{(p^{w+1}-1)^2}\right).$$

Hence, by (2.11),

$$G''(0) = \left(\frac{1}{4}(\log 2)^2 - A_2 + A_1^2\right)C^{-1}.$$

Therefore the main term in (3.6) is

$$\frac{1}{2C}\left((\log y)^2 + (4\gamma_0+2A_1)(\log y) + 2\gamma_0^2 - 4\gamma_1 + 4\gamma_0 A_1 + \frac{1}{4}(\log 2)^2 - A_2 + A_1^2\right).$$

The lemma now follows from (2.12), (2.13), (2.15), (3.6) and (3.7).

Lemma 3. *When n is even, let*

(3.12) $$S_n(z) = \sum_{q\leq z}\frac{\mu(q)^2}{1+z^{-1}q}\left(\prod_{\substack{p\mid q\\ p\nmid n}}\frac{2}{p-2}\right)\prod_{p\mid(q,n)}\frac{1}{p-1}.$$

Then

(3.13) $$S_n(z) \geq S_2(z)\prod_{\substack{p\mid n\\ p>2}}\frac{p-2}{p-1}$$

and, for $z\geq 1$,

(3.14)

$$|2CS_2(z) - (\log z)^2 - \frac{1}{2}A_6(1)(\log z) - \frac{1}{4}A_7(1)| < \frac{1}{4}A_8(1)z^{-1/3} + \frac{1}{4}A_9(1)z^{-1}.$$

Proof. Let $s(q)=\Pi_{p|q}p$, the squarefree kernel of q. By considering the expansions

$$\frac{2}{p-2}=\sum_{h=1}^{\infty}\left(\frac{2}{p}\right)^h, \quad \frac{1}{p-1}=\sum_{h=1}^{\infty}\frac{1}{p^h}$$

it follows that

$$S_n(z)=\sum_{s(q)\le z}\frac{1}{q+z^{-1}qs(q)}\sum_{\substack{d|q\\(d,n)=1}}f(d)$$

where f is the multiplicative function with $f(p^m)=2^{m-1}$. Thus

$$S_n(z)\ge\sum_{\substack{s(dr)\le z\\(d,n)=1}}\frac{f(d)}{dr+z^{-1}drs(dr)}\left(\sum_{\substack{s(qdr)\le z\\s(q)|n,\,q\text{ odd}}}\frac{f(q)}{q}\right)\left(\sum_{\substack{s(q)|n\\q\text{ odd}}}\frac{f(q)}{q}\right)^{-1}$$

$$=\sum_{s(m)\le z}\frac{1}{m+z^{-1}ms(m)}\sum_{\substack{k|m\\k\text{ odd}}}f(k)\prod_{\substack{p|n\\p>2}}\left(1+\frac{1}{p}\left(1+\frac{2}{p}+\frac{2^2}{p^2}+\dots\right)\right)^{-1}$$

$$=\left(\prod_{\substack{p|n\\p>2}}\frac{p-2}{p-1}\right)S_2(z),$$

which gives (3.13).

By (3.12) and (3.2),

$$S_2(z)=\frac{1}{2}S(z)+\int_1^z\frac{zS(u)}{(z+u)^2}\,du.$$

Let

(3.15) $$M(y)=(\log y)^2+A_3\log y+A_4.$$

Then by Lemma 2,

$$\left|2CS_2(z)-\left[\frac{1}{2}M(z)-\int_1^z\frac{zM(u)}{(z+u)^2}\,du\right|<\frac{1}{2}A_5z^{-1/3}+\int_1^z\frac{zA_5u^{-1/3}}{(z+u)^2}\,du\right.$$

$$\le\tfrac{1}{2}A_5z^{-1/3}+z^{-1}A_5\int_1^z u^{-1/3}\,du.$$

Therefore

(3.16) $$\left|2CS_2(z)-\frac{1}{2}M(z)-\int_1^z\frac{zM(u)}{(z+u)^2}\,du\right|<2A_5z^{-1/3}.$$

By (3.15),

$$\int_1^z\frac{zM(u)}{(z+u)^2}\,du=\left[\frac{-zM(u)}{z+u}\right]_1^z+\int_1^z\frac{zM'(u)}{z+u}\,du.$$

The first term on the right contributes

$$\frac{zA_4}{z+1}-\frac{1}{2}(\log z)^2-\frac{1}{2}A_3\log z-\frac{1}{2}A_4$$

and the integral on the right contributes

$$\int_1^z \frac{2z}{u(z+u)}(\log u)\,du+\int_1^z \frac{zA_3}{u(z+u)}\,du.$$

On expanding $z(z+u)^{-1}$ as an infinite series in powers of u and interchanging the order of summation and integration (obviously justified by bounded convergence) the first integral becomes

$$\sum_{h=0}^{\infty}\int_1^z \frac{2}{u}\left(-\frac{u}{z}\right)^h (\log u)\,du = (\log z)^2+\sum_{h=1}^{\infty} 2\left(-\frac{1}{z}\right)^h\left(\frac{z^h}{h}\log z-\frac{z^h-1}{h^2}\right)$$

$$= (\log z)^2-2(\log 2)(\log z)+\frac{\pi^2}{6}-\sum_{h=1}^{\infty}\frac{2(-1)^{h-1}}{z^h h^2}.$$

Hence, by (3.15),

$$\frac{1}{2}M(z)+\int_1^z \frac{zM(u)}{(z+u)^2}\,du = (\log z)^2+(A_3-2\log 2)\log z+A_4$$

$$+\frac{\pi^2}{6}-A_3\log 2-\sum_{h=1}^{\infty}\frac{2(-1)^{h-1}}{z^h h^2}-\frac{A_4}{z+1}+A_3\log\left(1+\frac{1}{z}\right).$$

The terms in the series $\sum_{h=1}^{\infty} 2z^{-h}(-1)^{h-1}h^{-2}$ decrease in absolute value and oscillate in sign. Thus the series lies between 0 and $2/z$. Also, by (2.16) $A_3-2\log 2=\frac{1}{2}A_6(1)$, by (2.17) $A_4+\frac{\pi^2}{6}-A_3\log 2=\frac{1}{4}A_7(1)$, by (2.14) $0<A_4<A_3-2$, by (2.18) $2A_5=\frac{1}{4}A_8(1)$, and by (2.19) $A_3=\frac{1}{4}A_9(1)$. Hence, by (3.16) we have the lemma.

Lemma 4. *Suppose that* $x\geqq\lambda$ *and* $z=(x/\lambda)^{1/2}$. *Then*

$$|8CS_2(z)-(\log x)^2-A_6(\lambda)\log x-A_7(\lambda)|<A_8(\lambda)x^{-1/6}+A_9(\lambda)x^{-1/2}.$$

Proof. The lemma follows at once from (2.16), (2.17), (2.18), (2.19) and Lemma 3.

Lemma 5. *Let*

$$R(x, a, b) = \sup_I \sum_{\substack{p\in I\\ ap+b \text{ prime}}} 1 \tag{3.17}$$

where the supremum is taken over all intervals I of length x. Suppose that L and $A=A(L)$ *are related by the table below. Then, whenever* $x\geqq e^L$ *and* $ab\neq 0$ *we have*

$$R(x, a, b) < \left(\frac{8Cx}{(\log x)(A+\log x)}-100x^{1/2}\right)\prod_{\substack{p|ab\\ p>2}}\frac{p-1}{p-2}.$$

L	A	B	L	A	B
24	0	0.97	48	8.2	8.2054
25	1	2.31	60	8.3	8.302
26	2	3.40	82	8.35	8.3503
27	3	4.28	100	8.37	8.3708
28	4	5.00	127	8.39	8.3905
29	5	5.58	147	8.4	8.4004
31	6	6.45	174	8.41	8.4102
34	7	7.24	214	8.42	8.4201
36	7.5	7.56	278	8.43	8.4301
42	8	8.04	396	8.44	8.44004
44	8.1	8.11	690	8.45	8.45001

Proof. We may suppose that $(a, b)=1$ and ab is even, for otherwise $R(x, a, b)\leqq 2$ and the conclusion is trivial. Let $N=[x]$ and let I denote a typical interval of length x. For some integer M the integers h in I satisfy $M<h\leqq M+N+1$. Let

$$z = \left(\frac{2}{3}x\right)^{1/2}. \tag{3.18}$$

Then

$$\sum_{\substack{p\in I\\ ap+b \text{ prime}}} 1 \leqq \sum_{\substack{h=M+1\\ (h(ah+b),\, Q)=1}}^{M+N} 1+2\pi(z)+1$$

where $Q=\prod_{p\leqq z} p$. Therefore, by Corollary 1 of Montgomery and Vaughan [4],

$$R(x, a, b) \leqq \left(\sum_{q\leqq z} \frac{\mu(q)^2}{N+3/2qz}\left(\prod_{p|(q,ab)} \frac{1}{p-1}\right)\left(\prod_{\substack{p|q\\ p\nmid ab}} \frac{2}{p-2}\right)\right)^{-1}+2x^{1/2}.$$

Hence, by (3.12) and (3.18),

$$R(x, a, b) \leqq x(S_{ab}(z))^{-1}+2x^{1/2}.$$

Therefore, by (3.13),

$$R(x, a, b) \leqq \left(x(S_2(z))^{-1}+2x^{1/2}\right) \prod_{\substack{p|ab\\ p>2}} \frac{p-1}{p-2}. \tag{3.19}$$

By Lemma 4 with $\lambda=\frac{3}{2}$ we have

$$8CS_2(z) > (\log x)^2+F(x)\log x$$

where

$$F(x) = A_6\left(\frac{3}{2}\right)+\frac{A_7\left(\frac{3}{2}\right)}{\log x}-\frac{A_8\left(\frac{3}{2}\right)}{x^{1/6}\log x}-\frac{A_9\left(\frac{3}{2}\right)}{x^{1/2}\log x}.$$

By (2.20), (2.21) and (2.22), $F(x)$ is an increasing function of x for $x>1$ and

$$F(x) > B \quad (x \geqq e^L)$$

where B is given by the above table. Hence, by (3.19),

$$R(x, a, b) < \left(\frac{8Cx}{(\log x)(B+\log x)} + 2x^{1/2}\right) \prod_{\substack{p|ab \\ p>2}} \frac{p-1}{p-2}. \tag{3.20}$$

Since $(\log x)x^{-1/6}$ is a decreasing function for $x \geqq e^6$ and, by (2.4),

$$(\log x)(A+\log x)(B+\log x) < \frac{4C(B-A)}{51} x^{1/2} \tag{3.21}$$

when $x=e^L$ and A and B are given by the above table, it follows that (3.21) holds whenever $x \geqq e^L$. Moreover (3.21) is equivalent to

$$\frac{8Cx}{(\log x)(B+\log x)} + 102x^{1/2} < \frac{8Cx}{(\log x)(A+\log x)}.$$

The lemma now follows from (3.20).

4. An auxiliary lemma concerning prime numbers

In order to treat $N(x)$ we need to know that the prime numbers are fairly plentiful, and are reasonably well distributed. This information is provided by the following lemma.

Lemma 6. (i) *Suppose that* $\log x \geqq 17$. *Then*

$$\pi(x) > \frac{x}{\log x} + (0.9911)\frac{x}{(\log x)^2}. \tag{4.1}$$

(ii) *Suppose that* $\log x \geqq 300$. *Then*

$$\pi(x) < \frac{x}{\log x} + (1.0151)\frac{x}{(\log x)^2}. \tag{4.2}$$

Proof. We quote a number of results from Rosser and Schoenfeld [6]. Their Theorem 2 gives

$$|\theta(x) - x| < x\varepsilon(x) \quad (\log x \geqq 105)$$

where

$$\varepsilon(x) = 0.257634\left(1+\frac{0.96642}{X}\right) X^{3/4} e - X$$

with $X=(R^{-1}\log x)^{1/2}$ and $R=9.645908801$. Now $\varepsilon(x)\log x=\varepsilon(x)X^2R$ and $X^{11/4}e^{-X}$ is decreasing for $X>11/4$. Hence

(4.3) $$|\theta(x)-x| < (0.000154)\frac{x}{\log x} \quad (\log x \geqq 3000).$$

The table on page 267 of Rosser and Schoenfeld [6], the use of which is described at the beginning of their §4, gives values of ε and b such that

$$|\psi(x)-x| < \varepsilon x \quad (\log x \geqq b).$$

Inspection of this table shows that

(4.4) $$|\psi(x)-x| < (0.00822)\frac{x}{\log x} \quad (22 \leqq \log x \leqq 5000).$$

Theorem 6 of Rosser and Schoenfeld [5] gives

$$\theta(x) > \psi(x)-(1.001102)x^{1/2}-3x^{1/3} \quad (x>0).$$

Thus

$$\theta(x) > \psi(x)-(0.0003961)\frac{x}{\log x} \quad (\log x \geqq 22).$$

Hence, by (4.4) and (4.3),

(4.5) $$\theta(x) > x-(0.00862)\frac{x}{\log x} \quad (\log x \geqq 22).$$

Now

(4.6) $$\pi(x) = \frac{\theta(x)}{\log x}+\int_2^x \frac{\theta(u)}{(\log u)^2}\,du.$$

Writing $\delta=0.00862$, $y=e^{22}$ we obtain for $x\geqq y$

$$\pi(x) > \frac{x}{\log x}-\frac{\delta x}{(\log x)^2}+\int_y^x \frac{1}{(\log u)^2}-\frac{\delta}{(\log u)^3}\,du$$

$$= \frac{x}{\log x}-\frac{\delta x}{(\log x)^2}+\left[\frac{u}{(\log u)^2}-\frac{\delta u}{(\log u)^3}\right]_y^x+\int_y^x \frac{2}{(\log u)^3}-\frac{3\delta}{(\log u)^4}\,du$$

$$> \frac{x}{\log x}+\frac{x}{(\log x)^2}\left(1-\delta-\frac{\delta}{\log x}-\frac{y(\log x)^2}{x(\log y)^3}\right).$$

When $\log x\geqq 32$

$$\frac{\delta}{\log x}+\frac{y(\log x)^2}{x(\log y)^3}$$

is a decreasing function of x and so does not exceed 0.00028. This gives (4.1) when $\log x\geqq 32$.

Corollary 2 to Theorem 7 of Rosser and Schoenfeld [6] gives

$$\theta(x) > x - \frac{x}{40\log x} \quad (x \geqq 678{,}407).$$

Let $y=678{,}407$. Then, by (4.6), when $x \geqq y$ we have

$$\pi(x) > \frac{x}{\log x} - \frac{x}{40(\log x)^2} + \int_y^x \frac{1}{(\log u)^2} - \frac{1}{40(\log u)^3}\,du = \frac{x}{\log x} - \frac{x}{40(\log x)^2}$$

$$+\left[\frac{u}{(\log u)^2} - \frac{u}{40(\log u)^3} + \frac{2u}{(\log u)^3} - \frac{3u}{40(\log u)^4}\right]_y^x + \int_y^x \frac{6}{(\log u)^4} - \frac{12}{40(\log u)^5}\,du$$

$$> \frac{x}{\log x} + \frac{x}{(\log x)^2} + \frac{x}{(\log x)^3}\left(\frac{79-\log x}{40} - \frac{3}{40\log x} - \frac{y(\log x)^3}{x(\log y)^2}\left(1+\frac{2}{\log y}\right)\right).$$

Hence, for $17 \leqq \log x \leqq 35$ we have

$$\pi(x) > \frac{x}{\log x} + \frac{x}{(\log x)^2}$$

which is more than is required.

It remains to prove (4.2). We have $\theta(x) \leqq \psi(x)$. Hence, by (4.3) and (4.4)

$$\theta(x) < x + (0.00822)\frac{x}{\log x} \quad (\log x \geqq 22).$$

Let $y=e^{200}$ and $\delta=0.00822$. Then, by (4.6),

$$\pi(x) < \frac{x}{\log x}\left(1+\frac{\delta}{\log x}\right) + \int_y^x \frac{1}{(\log u)^2}\left(1+\frac{\delta}{\log u}\right)du + \pi(y). \tag{4.7}$$

Let

$$I = \int_y^x \frac{du}{(\log u)^2}.$$

Then

$$I = \left[\frac{u}{(\log u)^2} + \frac{2u}{(\log u)^3}\right]_y^x + \int_y^x \frac{6\,du}{(\log u)^4} < \frac{x}{(\log x)^2} + \frac{2x}{(\log x)^3} + \frac{6I}{(\log y)^2}.$$

Hence, when $\log x \geqq 300$,

$$I < 1.006818\frac{x}{(\log x)^2}.$$

Similarly

$$\int_y^x \frac{du}{(\log u)^3} < (0.003385)\frac{x}{(\log x)^2}$$

Therefore, by (4.7), we have (4.2) as desired.

5. The estimation of $N(x)$ when x is small

Lemma 7. *Suppose that* $2 \leqq x \leqq e^{27}$. *Then* $N(x) > x/18$.

Proof. Each of 2, 4, 6, 8 is the sum of at most two primes. Hence $N(x) > x/18$ when $2 \leqq x \leqq 67$.

By considering those numbers of the form $p+3$ and $p+5$ with $p \geqq 3$ it follows that

$$N(x) \geqq \pi(x-3)+\pi(x-5)-1-\sum_{\substack{3 \leqq p \leqq x-3 \\ p-2 \text{ prime}}} 1.$$

If $p>7$ and $p-2$ is prime, then $p-4$ is not prime whereas both p and $p-2$ are counted by $\pi(x-3)$. Hence

$$\sum_{\substack{3 \leqq p \leqq x-3 \\ p-2 \text{ prime}}} 2 \leqq \pi(x-3).$$

Thus, when $x \geqq 8$,

$$N(x) \geqq \frac{1}{2}\pi(x-3)+\pi(x-5)-1 \geqq \frac{3}{2}\pi(x)-4.$$

By (3.3) of Theorem 2 of Rosser and Schoenfeld [5], when $x \geqq 67$ we have

$$\pi(x) \geqq \frac{2x}{(2\log x)-1}.$$

We have $\dfrac{y}{\log y} > \left(\dfrac{32}{3\sqrt{e}}\right)^{1/2}$ whenever $y>1$. Thus, on writing $y=\left(\dfrac{x}{\sqrt{e}}\right)^{1/2}$ we have

$$-\frac{6}{x(2\log x-1)^2}+\frac{4}{x^2}<0 \quad \text{for} \quad x>2.$$

Thus

$$\frac{3}{(2\log x)-1}-\frac{4}{x}-\frac{1}{18}$$

is decreasing for $x \geqq 67$ and is positive when $x=e^{27}$.

6. The intermediate region

It is in the proof of the following lemma that the improved form of Lemma 5 plays a crucial rôle.

Lemma 8. *Suppose that* $24 \leqq \log x \leqq 424$. *Then*

$$N(x) > x/18.$$

Proof. Let

(6.1) $$R(n) = \sum\sum_{\substack{3 \leqq p_1 \leqq u,\, 3 \leqq p_2 \leqq u \\ p_1+p_2=n}} 1$$

where u is a parameter at our disposal with

(6.2) $$3 \leqq u \leqq 10^5.$$

Note that $R(n)=0$ when $n>x$. Hence, by Cauchy's inequality

$$(\sum_n R(n))^2 \leqq N(x) \sum_n R(n)^2.$$

We also have

$$\sum_n R(n)^2 = \sum_n R(n) + \sum\sum_{3 \leqq p_1 < p_2 \leqq u} \sum_{\substack{3+p_2-p_1 \leqq p_3 \leqq x-u \\ p_3-p_2+p_1 \text{ prime}}} 2.$$

Therefore, by Lemma 5 and (2.27),

(6.3) $$\sum_n R(n)(\sum_n R(n) - N(x)) \leqq \frac{16CxN(x)T(u)}{(\log x)(A+\log x)}.$$

By (6.1),

(6.4) $$\sum_n R(n) = (\pi(u)-1)(\pi(x-u)-1)$$

and, by (6.2) and Lemma 6,

$$\pi(x-u)-1 \geqq \pi(x)-u \geqq \alpha$$

where

(6.5) $$\alpha = \frac{x}{\log x} + \frac{Dx}{(\log x)^2}$$

and

(6.6) $$D = 0 \quad (24 \leqq \log x \leqq 42), \quad D = 0.99 \ (\log x > 42).$$

Therefore, by (2.28) and (6.4),

$$\alpha s(\alpha s - N(x)) \leqq \alpha s(\sum_n R(n) - N(x)) \leqq (\sum_n R(n))(\sum_n R(n) - N(x)).$$

Let

(6.7) $$\beta = \frac{16Cx}{(\log x)(A+\log x)}.$$

then, by (2.29) and (6.3)

$$\alpha s(\alpha s - N(x)) \leqq \beta t N(x).$$

Hence

$$N(x) \geqq \frac{\alpha^2 s^2}{\alpha s + \beta t}.$$

Therefore, it suffices to show that for suitable choices of u we have

$$\frac{\alpha^2 s^2}{\alpha s + \beta t} > \frac{x}{18}.$$

By (6.5) and (6.7) this is equivalent to

(6.8) $$s\left(\frac{A}{l}+1\right)\left(\frac{D}{l}+1\right)\left[18s\left(\frac{D}{l}+1\right)-l\right]-16Ct > 0$$

where

$$l = \log x.$$

For given A, D, u with $A \geqq 0$, $D \geqq 0$ the left hand side of (6.8) is a decreasing function of l. We choose our parameters as follows.

$A=0$,	$D=0$,	$u=79$	when	$24 \leqq l \leqq 42$.
$A=8$,	$D=0.99$,	$u=99989$	when	$42 < l \leqq 300$.
$A=8.43$,	$D=0.99$,	$u=99989$	when	$300 < l \leqq 400$.
$A=8.44$,	$D=0.99$,	$u=99989$	when	$400 < l \leqq 424$.

These choices are in conformity with Lemma 5 and (6.6). Then on inserting in (6.8) the corresponding values of s and t given by (2.29) and (2.30) and the upper bound for C given by (2.4) we see that the left hand side of (6.8) is positive when $l=42$, when $l=300$, when $l=400$ and when $l=424$ respectively. The lemma now follows.

7. Preliminaries to the estimation of $N(x)$ when x is large

Let

(7.1) $$K = 200, \quad y = x/(K+2),$$

(7.2) $$I_k = \left(\tfrac{1}{2}ky, \tfrac{1}{2}ky+y\right] \quad (k = 1, 2, ..., K)$$

and define

(7.3) $$R_k(n) = \sum_{\substack{p, p' \\ p+p'=n \\ p\in I_k,\, p'\in I_k}} 1,$$

(7.4) $$w(n) = \prod_{\substack{p|n \\ p>2}} \frac{p-2}{p-1}$$

and

(7.5) $$\Psi = \sum_{k=1}^{K} \sum_n R_k(n)\, w(n).$$

Lemma 9. *Suppose that* $\log y > 350$. *Then*

$$\Psi < (N(x)-N(y)) \frac{8Cy}{\left(\log \frac{y}{2}\right)\left(8.3+\log \frac{y}{2}\right)}.$$

Proof. By (7.3), $R_k(n)=0$ when $n \leqq ky$ or $n > ky+2y$,

$$R_k(n) = \sum_{\substack{\frac{1}{2}ky < p < n - \frac{1}{2}ky \\ n-p \text{ prime}}} 1 \quad \text{when} \quad ky < n \leqq ky+y.$$

$$R_k(n) = \sum_{\substack{n-\frac{1}{2}ky-y \leqq p \leqq \frac{1}{2}ky+y \\ n-p \text{ prime}}} 1 \quad \text{when} \quad ky+y < n \leqq ky+2y.$$

Hence, by (7.5),

$$\Psi = \sum_{y<n\leqq 2y} R_1(n)w(n)+\sum_{k=2}^{K}\sum_{ky<n\leqq ky+y}(R_k(n)+R_{k-1}(n))w(n) + \sum_{Ky+y<n\leqq Ky+2y} R_K(n)w(n), \tag{7.6}$$

and, by (7.4) and Lemma 5, when $ky+e^{60}<n\leqq ky+y-e^{60}$ we have

$$(R_k(n)+R_{k-1}(n))w(n) < 8C\left(\frac{u}{(\log u)(8.3+\log u)}+\frac{y-u}{(\log (y-u))(8.3+\log (y-u))}\right) \tag{7.7}$$

where $u=n-ky$. If instead $ky+y-e^{60}<n\leqq ky+y$, then

$$(R_k(n)+R_{k-1}(n))w(n) < \left(\frac{8Cu}{(\log u)(8.3+\log u)}-100u^{1/2}+e^{60}\right)$$

and since $u>y-e^{60}>e^{120}$ it follows that

$$(R_k(n)+R_{k-1}(n))w(n) < \frac{8Cy}{(\log y)(8.3+\log y)}.$$

A similar argument gives the same inequality when $ky<n\leqq ky+e^{60}$. Also, by Lemma 5, we have

$$R_k(n)w(n) < \frac{8Cy}{(\log y)(8.3+\log y)} \quad (k=1 \text{ or } K).$$

Therefore the lemma will follow from (7.6) and (7.7) provided that we can show that

$$\frac{u}{(\log u)(8.3+\log u)}+\frac{y-u}{(\log (y-u))(8.3+\log (y-u))} \leqq \frac{y}{\left(\log\frac{y}{2}\right)\left(8.3+\log\frac{y}{2}\right)} \tag{7.8}$$

whenever $e^{60}\leqq u\leqq\frac{1}{2}y$. Write $f(u)$ for the left hand side of (7.8) and consider it as a function of the continuous variable u. For brevity write $l=\log u$, $m=\log(y-u)$, so that $60\leqq l\leqq \log\frac{y}{2}\leqq m$. Then

$$f'(u)=\frac{l(8.3+l)-8.3-2l}{l^2(8.3+l)^2}-\frac{m(8.3+m)-8.3-2m}{m^2(8.3+m)^2}.$$

Now $(l(8.3+l)-8.3-2l)l^{-2}(8.3+l)^{-2}$ is strictly decreasing for $l\geqq 60$. Thus $f'(u)>0$ when $m>l\geqq 60$ and so (7.8) holds. This completes the proof of the lemma.

8. A lower bound for Ψ

By (7.4),

$$w(n) = \sum_{\substack{d|n \\ 2\nmid p}} \frac{\mu(d)}{\varphi(d)}.$$

Hence, by (7.5),

(8.1)
$$\Psi = \sum_{k=1}^{K} \Psi_k$$

where

(8.2)
$$\Psi_k = \sum_{\substack{p \\ 2\nmid d}} \frac{\mu(d)}{\varphi(d)} \sum_{\substack{n \\ d|n}} R_k(n).$$

By (7.3), when $y>4$ and d is odd,

(8.3)
$$\sum_{\substack{n \\ d|n}} R_k(n) = \sum_{\substack{p\in I_k \\ d|p+p'}} \sum_{\substack{p'\in I_k \\ d|p+p'}} 1.$$

Hence, by (7.1) and (7.2), this expression is zero when $d>\frac{1}{2}x$. Let

(8.4)
$$\Xi_k = \sum_{\substack{\frac{1}{2}y<d\leqq\frac{1}{2}x \\ 2\nmid d}} \frac{\mu(d)}{\varphi(d)} \sum_{\substack{n \\ d|n}} R_k(n).$$

Then, by (8.2),

(8.5)
$$\Psi_k = \Xi_k + \sum_{\substack{d\leqq\frac{y}{2} \\ 2\nmid d}} \frac{\mu(d)}{\varphi(d)} \sum_{\substack{n \\ d|n}} R_k(n).$$

Let

(8.6)
$$M_k = \left[\frac{1}{2}ky\right], \quad N_k = \left[\frac{1}{2}ky+y\right]-M_k, \quad z = z_k = \left(\frac{2}{3}N_k\right)^{1/2}, \quad w = w_k = \frac{z}{100}.$$

When $d\leqq\frac{1}{2}y$ it follows from (7.2) that every prime p in I_k satisfies $p\nmid d$. Hence, by (8.3),

(8.7)
$$\sum_{\substack{n \\ d|n}} R_k(n) = \frac{1}{\varphi(d)} \sum_{\chi \bmod d} \chi(-1)|S_k(\chi)|^2$$

where

(8.8)
$$S_k(\chi) = \sum_{p\in I_k} \chi(p).$$

Moreover each term in (8.7) is unaltered if we replace χ by the primitive character χ^* that induces it. Let d^* denote the conductor of χ. Then, by (8.5) and (8.7),

(8.9)
$$\Psi_k = \Phi_k - \Delta_k + \theta_k + \Xi_k$$

where

$$\Phi_k = \sum_{\substack{d \\ 2\nmid d}} \frac{\mu(d)}{\varphi(d)^2} \sum_{\substack{\chi \bmod d \\ d^* \leqq w}} \chi^*(-1)|S_k(\chi^*)|^2, \tag{8.10}$$

$$\Delta_k = \sum_{\substack{d > \frac{1}{2}y \\ 2\nmid d}} \frac{\mu(d)}{\varphi(d)^2} \sum_{\substack{\chi \bmod d \\ d^* \leqq w}} \chi^*(-1)|S_k(\chi^*)|^2, \tag{8.11}$$

$$\theta_k = \sum_{\substack{d \leqq \frac{1}{2}y \\ 2\nmid d}} \frac{\mu(d)}{\varphi(d)^2} \sum_{\substack{\chi \bmod d \\ d^* > w}} \chi(-1)|S_k(\chi)|^2. \tag{8.12}$$

Let

$$S_k = \sum_{p \in I_k} 1. \tag{8.13}$$

Lemma 10. *Suppose that* $\log y > 350$. *Then*

$$|\Xi_k| \leqq 6.31 A_{10} S_k.$$

Proof. The length of I_k is y and in (8.3) p' is determined by p modulo $2d$. Hence when $d > \frac{1}{2}y$, p' is uniquely determined. Therefore

$$\sum_{\substack{n \\ d|n}} R_k(n) \leqq S_k.$$

Hence, by (8.4),

$$|\Xi_k| \leqq \sum_{\substack{\frac{1}{2}y < d \leqq \frac{1}{2}x \\ 2\nmid d}} \frac{\mu(d)^2}{\varphi(d)} S_k.$$

We have

$$\frac{1}{\varphi(d)} = \frac{1}{d} \sum_{r|d} \frac{\mu(r)^2}{\varphi(r)}. \tag{8.14}$$

Hence

$$\sum_{\substack{\frac{1}{2}y < d \leqq \frac{1}{2}x \\ 2\nmid d}} \frac{\mu(d)^2}{\varphi(d)} \leqq \sum_{r \text{ odd}} \frac{\mu(r)^2}{r\varphi(r)} \sum_{\frac{1}{2r}y < m \leqq \frac{1}{2r}x} \frac{1}{m} < \sum_{r \text{ odd}} \frac{\mu(r)^2}{r\varphi(r)} \left(1 + \log \frac{x}{y}\right).$$

Therefore, by (7.1) and (2.23),

$$|\Xi_k| \leqq A_{10}(1 + \log(202)) S_k.$$

Lemma 11. *Suppose that* $\log y > 350$. *Then*

$$|\theta_k| \leqq 3A_{10} \frac{y}{w} S_k.$$

Proof. When $w < d \leqq \frac{1}{2}y$, it follows from (8.8) that

$$\sum_{\chi \bmod d} |S_k(\chi)|^2 = \varphi(d) \sum\sum_{\substack{p \in I_k,\ p' \in I_k \\ p \equiv p' \pmod d}} 1 \leqq \varphi(d) S_k \left(\frac{y}{d} + 1\right) \leqq \frac{3\varphi(d)}{2d} y S_k.$$

Therefore, by (8.12),

$$|\theta_k| \leqq \frac{3}{2} S_k y \sum_{\substack{d>w \\ 2\nmid d}} \frac{\mu(d)^2}{d\varphi(d)}.$$

By (8.14),

$$\sum_{\substack{d>w \\ 2\nmid d}} \frac{\mu(d)^2}{d\varphi(d)} \leqq \sum_{r\,\mathrm{odd}} \frac{\mu(r)^2}{r^2\varphi(r)} \sum_{m>w/r} \frac{1}{m^2} \leqq \frac{2}{w} \prod_{p>2}\left(1+\frac{1}{p(p-1)}\right).$$

The lemma now follows from (2.23).

Lemma 12. *Suppose that* $\log y > 350$. *Then*

$$|\Delta_k| \leqq 8A_{11} w S_k.$$

Proof. Clearly

$$\sum_{\substack{\chi \bmod d \\ d^* \leqq w}} |S_k(\chi^*)|^2 \leqq \sum_{\substack{r|d \\ r\leqq w}} \sum_{\chi \bmod r} |S_k(\chi)|^2 \leqq \sum_{r\leqq w} \varphi(r) \sum\sum_{\substack{p\in I_k,\ p'\in I_k \\ p\equiv p' \pmod r}} 1$$

$$\leqq \sum_{r\leqq w} \varphi(r) S_k\left(\frac{y}{r}+1\right).$$

Hence, by (8.6) and (8.11),

$$|\Delta_k| \leqq \sum_{\substack{d>\frac{1}{2}y \\ 2\nmid d}} \frac{\mu(d)^2}{\varphi(d)^2} 2wyS_k. \tag{8.15}$$

Define the multiplicative function g by

$$g(2) = 0, \quad g(p) = \frac{2p-1}{(p-1)^2} \quad (p > 2), \quad g(p^k) = 0 \quad (k > 1).$$

Then for odd squarefree d

$$\frac{1}{\varphi(d)^2} = \frac{1}{d^2} \sum_{r|d} g(r).$$

Hence

$$\sum_{\substack{d>\frac{1}{2}y \\ 2\nmid d}} \frac{\mu(d)^2}{\varphi(d)^2} \leqq \sum_r \frac{g(r)}{r^2} \sum_{m>\frac{y}{2r}} \frac{1}{m^2} \leqq \frac{4}{y} \sum_r \frac{g(r)}{r}.$$

Therefore, by (8.15) and (2.24), we have the lemma.

Lemma 13. *Suppose that* $\log y > 350$. *Then*

$$\Phi_k \geqq \frac{1}{2} C\left(\frac{16}{15} S_k^2 - \frac{2S_k(N_k - S_k \log 5)}{15(-2.9024 + \log N_k)}\right)$$

Proof. By (8.10),

$$\Phi_k = \sum_{q\leqq w} \sum\nolimits^*_{\chi \bmod q} \chi(-1) |S_k(\chi)|^2 \sum_{\substack{d \\ q|d \\ 2\nmid d}} \frac{\mu(d)}{\varphi(d)^2}$$

where $\sum^*$ means that we sun only over the primitive characters modulo q. Let

$$f(q) = \prod_{p|q} \frac{1}{p(p-2)}$$

when q is odd and squarefree, and let $f(q)=0$ otherwise. Then, by (2.3),

$$\sum_{\substack{d \\ q|d \\ 2\nmid d}} \frac{\mu(d)}{\varphi(d)^2} = \frac{1}{2} C\mu(q)f(q).$$

Hence

(8.16) $$\Phi_k = \tfrac{1}{2} C \sum_{q \leqq w} \mu(q)f(q) \sum^*_{\chi \bmod q} \chi(-1)|S_k(\chi)|^2.$$

Let

$$S(\alpha) = \sum_{p \in I_k} e(\alpha p)$$

where $e(\beta)=e^{2\pi i\beta}$. Then, by (2.6) of Montgomery and Vaughan [4], (7.2) and (8.6), we have

(8.17) $$\sum_{q \leqq z} \left(N_k + \frac{3}{2} qz\right)^{-1} \sum^q_{\substack{a=1 \\ (a,q)=1}} \left|S\left(\frac{a}{q}\right)\right|^2 \leqq S_k.$$

When χ is a character modulo q, let $\tau(\chi)$ denote the gaussian sum associated with χ,

$$\tau(\chi) = \sum^q_{r=1} \chi(r) e\left(\frac{r}{q}\right).$$

Then, for $q \leqq \dfrac{1}{2} y$,

$$S_k\left(\frac{a}{q}\right) = \frac{1}{\varphi(q)} \sum_{\chi \bmod q} \chi(a)\tau(\bar{\chi})S_k(\chi).$$

Hence

$$\sum^q_{\substack{a=1 \\ (a,q)=1}} \left|S_k\left(\frac{a}{q}\right)\right|^2 = \frac{1}{\varphi(q)} \sum_\chi |\tau(\chi)|^2 |S_k(\chi)|^2.$$

Let q^* denote the conductor of χ. It is easily shown (e.g. on page 67 of Davenport [1]) that $|\tau(\chi)|^2 = q^*$ when q/q^* is squarefree and $(q/q^*, q^*)=1$, and that $|\tau(\chi)|^2=0$ otherwise. Hence

$$\sum^q_{\substack{a=1 \\ (a,q)=1}} \left|S_k\left(\frac{a}{q}\right)\right|^2 = \sum_{\substack{r|q \\ (q/r,r)=1}} \frac{\mu(q/r)^2 r}{\varphi(q)} \sum^*_{\chi \bmod r} |S_k(\chi)|^2.$$

Therefore, by (8.6) and (8.17),

$$\sum_{r \leqq z} \frac{r}{\varphi(r)} \left(\sum_{\substack{m \leqq z/r \\ (m,r)=1}} \frac{\mu(m)^2}{\varphi(m)} \cdot \frac{1}{1+rmz^{-1}}\right) \sum^*_{\chi \bmod r} |S_k(\chi)|^2 \leqq S_k N_k.$$

By Lemmas 3 and 8 of Montgomery and Vaughan [4] and (8.6), whenever

$$r \leqq w$$

we have

$$\frac{r}{\varphi(r)} \sum_{\substack{m \leqq z/r \\ (m,r)=1}} \frac{\mu(m)^2}{\varphi(m)} \cdot \frac{1}{1+rmz^{-1}} > 0.361 + \log \frac{z}{r}.$$

Hence

$$\sum_{5 \leqq r \leqq w} \left(0.361 + \log \frac{z}{r}\right) \sum^{*}_{\chi \bmod r} |S_k(\chi)|^2 \leqq S_k N_k - S_k^2 (0.361 + \log z). \tag{8.18}$$

There is only one primitive character χ modulo 3, and for that character we have $\chi(-1) = -1$. Hence, by (8.16),

$$\Phi_k \geqq \frac{1}{2} C\left(S_k^2 - \sum_{5 \leqq q \leqq w} f(q) \sum^{*}_{\chi \bmod q} |S_k(\chi)|^2\right).$$

Therefore, by (8.18),

$$\Phi_k \geqq \tfrac{1}{2} C\left(S_k^2 - F S_k N_k + F S_k^2 (0.361 + \log z)\right)$$

where

$$F = \max_{5 \leqq q \leqq w} \frac{f(q)}{0.361 + \log \frac{z}{q}}.$$

The lemma will now follow from (8.6) if we show that the maximum occurs when $q=5$. Consider the function of α

$$\alpha\left(0.361 + \log \frac{z}{\alpha}\right) \quad (1 \leqq \alpha \leqq z).$$

This has its maximum when $\alpha = z \exp(-0.639)$, i.e., by (8.6), when $\alpha > w$. Hence it is strictly increasing when $5 \leqq \alpha \leqq w$. Therefore, when $7 \leqq q \leqq w$ and q is odd and and squarefree we have

$$\frac{f(q)}{0.361 + \log \frac{z}{q}} = \left(\prod_{p \mid q} \frac{1}{p-2}\right) \frac{1}{q\left(0.361 + \log \frac{z}{q}\right)} \leqq \frac{1}{3} \frac{1}{7\left(0.361 + \log \frac{z}{7}\right)}.$$

By (8.6) and the hypothesis $\log y > 350$ this is

$$< \frac{1}{15\left(0.361 + \log \frac{z}{5}\right)}.$$

This establishes that the maximum occurs when $q=5$ and completes the proof of the lemma.

Lemma 14. *Suppose that* $\log y > 350$. *Then*

$$\Psi_k \geqq \frac{C}{15}\left(\left(8+\frac{\log 5}{\log y}\right) S_k^2 - \frac{y}{\log y}\left(1+\frac{2.9267}{\log y}\right) S_k\right).$$

Proof. By (2.25),

$$6.31 A_{10} < e^{-300}\frac{y}{(\log y)^2}.$$

Hence, by Lemma 10 and (2.4),

$$|\Xi_k| \leqq \frac{C}{15} 10^{-50} \frac{y}{(\log y)^2} S_k. \tag{8.19}$$

By (8.6), $y-1 < N_k < y+1$, $z=\left(\frac{2}{3}N_k\right)^{1/2}$, $w=\frac{z}{100}$.

Therefore $w > \frac{1}{200} y^{1/2} > e^{125}(\log y)^2$. Therefore, by (2.25), Lemma 11 and (2.4),

$$|\theta_k| \leqq \frac{C}{15} 10^{-50} \frac{y}{(\log y)^2} S_k. \tag{8.20}$$

Similarly $w < \left(\frac{2}{3}N_k\right)^{1/2} < y^{1/2} < e^{-125} y (\log y)^{-2}$. Hence, by (2.26), Lemma 12 and (2.4),

$$|\Delta_k| \leqq \frac{C}{15} 10^{-50} \frac{y}{(\log y)^2} S_k. \tag{8.21}$$

We have

$$\frac{N_k}{-2.9024+\log N_k} < \frac{y}{\log y}\left(1+\frac{1}{y}\right)\left(1-\frac{2.9024}{\log y}\right)^{-1} = \frac{y}{\log y}$$

$$+\frac{y}{(\log y)^2}\left(\frac{\log y}{y}+(2.9024)\frac{\log y}{-2.9024+\log y}+\frac{(2.9024)\log y}{y(-2.9024+\log y)}\right).$$

Since $\log y \geqq 350$, this does not exceed

$$\frac{y}{\log y}+(2.92667)\frac{y}{(\log y)^2}.$$

We also have

$$-2.9024+\log N_k < -2.9024+\log(y+1) < \log y.$$

Hence, by Lemma 13,

$$\Phi_k \geqq \frac{C}{15}\left(\left(8+\frac{\log 5}{\log y}\right) S_k^2 - \frac{y}{\log y}\left(1+\frac{2.92667}{\log y}\right) S_k\right).$$

The lemma now follows from this and (8.9), (8.19) (8.20) and (8.21).

We now have to estimate S_k^2 from below, and the following lemma gives a suitable bound.

Lemma 15. *Suppose that* $\log y > 350$. *Then*

$$\sum_{k=1}^{K} S_k > \frac{Ky}{\log y}\left(1-\frac{3.6581}{\log y}\right).$$

Proof. By (7.2) and (8.13),

$$\sum_{k=1}^{K} S_k = \pi\left(\tfrac{1}{2}Ky+\tfrac{1}{2}y\right)+\pi\left(\tfrac{1}{2}Ky+y\right)-\pi\left(\tfrac{1}{2}y\right)-\pi(y). \tag{8.22}$$

By (4.1), when $\lambda \geqq 1$ we have

$$\pi(\lambda y) > \frac{\lambda y}{\log \lambda y}+(0.9911)\frac{\lambda y}{(\log \lambda y)^2}$$

$$= \frac{\lambda y}{\log y}\left(1-\frac{1}{\log y}\left((\log \lambda)\frac{\log y}{\log \lambda y}-(0.9911)\left(\frac{\log y}{\log \lambda y}\right)^2\right)\right).$$

Moreover, when $\lambda \geqq \exp(1.9822)$

$$(\log \lambda)z-(0.9911)z^2$$

is an increasing function of z for $z \leqq 1$ and $(\log y)/\log \lambda y$ is an increasing function of y bounded above by 1. Thus

$$\pi(\lambda y) > \frac{\lambda y}{\log y}\left(1-\frac{(\log \lambda)-0.9911}{\log y}\right). \tag{8.23}$$

By (4.2),

$$\pi(y) < \frac{y}{\log y}\left(1+\frac{1.0151}{\log y}\right)$$

and

$$\pi\left(\frac{1}{2}y\right) < \frac{y}{2\log y}\left(1+\frac{1}{\log y}\left((\log 2)\frac{\log y}{\log y/2}+(1.0151)\left(\frac{\log y}{\log y/2}\right)^2\right)\right)$$

$$< \frac{y}{\log y}\left(\frac{1}{2}+\frac{0.85683}{\log y}\right).$$

Therefore, by (8.22) and (8.23),

$$\sum_{k=1}^{K} S_k > \frac{Ky}{\log y}\left(1-\frac{A_{12}}{\log y}\right)$$

where

$$A_{12} = \frac{1}{K}\left(\frac{K+1}{2}\left(\log\frac{K+1}{2} - 0.9911\right)\right.$$

$$\left.+\frac{K+2}{2}\left(\log\frac{K+2}{2} - 0.9911\right) + 1.0151 + 0.85683\right).$$

The lemma now follows from (7.1).

9. Completion of the proof of (1.1)

In view of Lemmas 7 and 8 it suffices now to show that

$$N(x) - N(y) > \frac{x-y}{18} \tag{9.1}$$

where $\log x \geqq 375$ and y is given by (7.1) (so that $\log y > 350$).

By Cauchy's inequality and Lemma 15,

$$\sum_{k=1}^{K} S_k^2 \geqq \frac{1}{K}\left(\sum_{k=1}^{K} S_k\right)^2 > \frac{y}{\log y}\left(1 - \frac{3.6581}{\log y}\right)\sum_{k+1}^{K} S_k.$$

Therefore, by Lemma 14 and (8.1),

$$\Psi > \frac{C}{15}\left(\left(8 + \frac{\log 5}{\log y}\right)\left(1 - \frac{3.6581}{\log y}\right) - \left(1 + \frac{2.9267}{\log y}\right)\right)\frac{y}{\log y}\sum_{k=1}^{K} S_k$$

$$> \frac{C}{15}\left(7 - \frac{30.5989}{\log y}\right)\frac{y}{\log y}\sum_{k=1}^{K} S_k.$$

Thus, by Lemma 15 again,

$$\Psi > \frac{C}{15}\left(7 - \frac{56.2056}{\log y}\right)\frac{Ky^2}{(\log y)^2}. \tag{9.2}$$

Since $\log y \geqq 350$ we have

$$\frac{\left(\log\frac{y}{2}\right)\left(8.3 + \log\frac{y}{2}\right)}{(\log y)^2} = 1 + \frac{1}{\log y}\left(8.3 - 2\log 2 - \frac{(8.3 - \log 2)\log 2}{\log y}\right)$$

$$> 1 + \frac{6.8986}{\log y}.$$

Therefore, by (9.2) and Lemma 9,

$$N(x)-N(y) > \frac{Ky}{120}\left(7-\frac{56.2056}{\log y}\right)\left(1+\frac{6.8986}{\log y}\right)$$

$$> \frac{Ky}{120}\left(7-\frac{9.0233}{\log y}\right) > (6.974)\,\frac{Ky}{120}.$$

Hence, by (7.1),

$$N(x)-N(y) > (0.0578)(x-y),$$

which gives (9.1) and so completes the proof of (1.1).

10. The computations

The different products taken over all primes p with $p \geqq q$ were computed in the following manner. Consider

$$Q = \prod_{p \geqq q} f(p)$$

or equivalently

$$R = \log Q = \sum_{p \geqq q} \log f(p).$$

In each case it is possible to expand $\log f(p)$ in the form

$$\sum_{j=r+1}^{\infty} a_j p^{-j/r}.$$

Usually $r=1$, but in the case of (2.8) it is necessary to take $r=3$. Thus

$$R = \sum_{j=r+1}^{\infty} a_j P_q(j/r)$$

where

$$P_q(s) = \sum_{p \geqq q} p^{-s}.$$

The value of $P_q(s)$ can be easily deduced from the corresponding value of the prime zeta function

$$P(s) = P_2(s).$$

For some values of s this has been computed by Fröberg [3]. In general the value of $P(s)$ can be obtained from the relation

$$P(s) = \sum_{k=1}^{\infty} \mu(k) k^{-1} \log \zeta(ks).$$

Since $\log \zeta(ks) \sim \log(1+2^{-ks}) \sim 2^{-ks}$ this converges more rapidly than the geometric series

$$\sum_{k=1}^{\infty} (2^{-s})^k.$$

However the convergence is still quite slow when s is close to 1. For instance, to find $P\left(\frac{4}{3}\right)$ correct to 10 decimal places would already require about 25 terms. This difficulty was surmounted by using instead the relation

$$P_q(s) = \sum_{k=1}^{\infty} \mu(k)k^{-1} \log \zeta_q(ks)$$

where

$$\zeta_q(ks) = \prod_{p \geqq q} (1-p^{-s})^{-1} = \zeta(s) \prod_{2 \leqq p < q} (1-p^{-s}).$$

We chose $q=19$ and terminated the summation at $k=15$. Then the relative error is

$$P_{19}(s)^{-1} \sum_{k=17}^{\infty} \mu(k)k^{-1} \log \zeta_{19}(ks)$$

and since $n^{-s} \leqq \frac{1}{2} \int_{n-1}^{n+1} u^{-s}\, du$, so that $S_{19}(s) \leqq 1+19^{-s}+\frac{1}{2}\int_{22}^{\infty} u^{-s}\, du$, the relative error is majorized by

$$19^s \sum_{k=17}^{\infty} \mu(k)^2 k^{-1} 19^{-ks} \left(1+\frac{11}{16}\left(\frac{19}{22}\right)^{17}\right) < 19^{-16s} 15^{-1} < (2.4)10^{-22}.$$

Thus we have only a small error in $P_{19}(s)$ provided that $\log \zeta_{19}(ks)$ can be computed with a small error. Since $\zeta_{19}(ks)$ is close to 1 when ks is large it is necessary, in order to avoid loss of accuracy in this case, to write $x=\zeta_{19}(ks)-1$ and to calculate $\log(1+x)$ as

$$x-\frac{1}{2}x^2+\frac{1}{3}x^3-\ldots$$

This in turn requires an accurate estimate for $\zeta_{19}(s)-1$ when $s>1$. This was obtained, without much cancellation, from the relation

$$\zeta_{19}(s) = \zeta_3(s) \prod_{3 \leqq p \leqq 17} (1-p^{-s})$$

rewritten as

$$\zeta_{19}(s)-1 = T \prod_{3 \leqq p \leqq 17} (1-p^{-s})+U$$

where

$$T = \sum_{\substack{n \geqq 19 \\ n \text{ odd}}} n^{-s}$$

and U is a finite sum consisting of powers of primes not exceeding 17. The sum T was computed, as usual, *via* the Euler—Maclaurin summation formula and the rest of the calculation introduced only rounding errors.

As a check on the programme we also computed $\log \zeta_{19}(s)$ as

$$\sum_{k=1}^{20} P_{19}(ks)\, k^{-1}$$

for different values of s. In no case was the difference larger than $10^{-17} \log \zeta_{19}(s)$.

All the sums and products needed were computed by using the calculated values of $P_{19}(s)$. For example, in the case of (2.8), the general factor in $H\left(-\frac{1}{3}\right)$ when

$p\geqq 3$ is given by

$$1+\frac{4}{p^{2/3}(p-2)}+\frac{3p+2}{p^{1/3}(p-2)}+\frac{2}{p(p-2)}$$

and on writing $z=p^{-1/3}$ this becomes

$$1+\sum_{k=4}^{\infty} c_k z^k$$

with

$$c_4=3,\ c_5=4,\ c_k=2^{l-1}\quad (k=3l),\ c_k=2^{1+l}\quad (k=3l+1,\ k=3l+2)$$

when $l\geqq 2$. The logarithm of this has an expansion of the form

$$\sum_{k=4}^{\infty} b_k z^k,$$

where

$$b_k=\sum_{1\leqq j\leqq \frac{1}{4}k}(-1)^{j-1}j^{-1}\sum_{\substack{l_1,\dots,l_j\\ l_1+\dots+l_j=k}} c_{l_1}\dots c_{l_j}.$$

We calculated $H\left(-\frac{1}{3}\right)$ by truncating at $k=62$. This probably gives rise to a truncation error $<10^{-19}$. However this is quite difficult to prove. Instead the following crude argument suffices for our purposes.

Clearly, when $0\leqq z\leqq 9/16$.

$$\sum_{k=4}^{\infty}|b_k|z^k\leqq -\log\left(1-\sum_{k=4}^{\infty}c_k z^k\right).$$

Hence

$$\left|\sum_{k=63}^{\infty} b_k z^k\right|\leqq \left(\frac{16}{9}z\right)^{63} F\left(\frac{9}{16}\right)\quad \left(0\leqq z\leqq \frac{9}{16}\right)$$

where

$$F(z)=-\log\left(1-\frac{4z^5}{1-2z^3}-\frac{3z^4+2z^7}{1-2z^3}-\frac{2z^6}{1-2z^3}\right).$$

Thus

$$F\left(\frac{9}{16}\right)<3.5$$

and

$$\left|\sum_{k=63}^{\infty} b_k z^k\right|<3.5\left(\frac{16}{9}z\right)^{63}.$$

It follows that

$$\log\prod_{p\geqq 19}\left(1+\frac{4}{p^{2/3}(p-2)}+\frac{3p+2}{p^{4/3}(p-2)}+\frac{2}{p(p-2)}\right)=\sum_{k=4}^{62} b_k P_{19}(k/3)+E$$

where

$$|E|<(3.5)\left(\frac{16}{9}\right)^{63} P_{19}(21)<10^{-10}.$$

The coefficients b_k were evaluated exactly by a computer programme when $k \leqq 62$. Then the prior estimates for $P_{19}(s)$ give

$$\sum_{k=4}^{62} b_k P_{19}(k/3) = 0.8850635511946\ldots.$$

The estimate (2.8) now follows.

The series containing $\log p$ and $(\log p)^2$ were computed as derivatives, *via* the relation

$$\frac{d^k}{ds^k} p^{-s} f(p) = (-\log p)^k p^{-s} f(p),$$

by Richardson extrapolation with successive differences $h=0.08$, 0.04, 0.02, 0.01. An analysis of the errors arising shows that, using floating point, double precision arithmetic (61 bits = 18 decimals) throughout we obtained about 16 decimal places in function values, 14 in first derivatives, and 12 in second derivatives.

The very laborious computation of $T(u)$, given by (2.27), for all primes $u < 10^5$, was speeded up in the following manner. First of all the value of

$$\prod_{\substack{p|d \\ p>2}} \frac{p-1}{p-2}$$

was calculated for each even $d < 10^5$ and stored. Then for each prime u the value of $T(u)$ was updated from the value of T for the previous prime by adding on the contributions arising from each d with $d=u-p$ and $p<u$. This required about $\frac{1}{2}\pi(10^5)^2 \cong 46 \cdot 10^6$ accesses to the values stored at the beginning. Using double precision arithmetic we finally found

$$T(99989) = 80096030.30\ldots$$

correct to at least 10 significant figures.

11. Acknowledgement

The second named author is indebted to the Mittag—Leffler Institute for support and hospitality while the work culminating in this paper was in progress.

References

1. DAVENPORT, H., *Multiplicative number theory,* second edition, Springer, Berlin, 1980.
2. DESHOUILLERS, J.-M., Sur la constante de Šnirel'man, *Sém. Delange—Pisot—Poitou* 1975/76, Fasc. **2**, Exp. No. G **16**, Paris 1977.
3. FRÖBERG, C.-E., On the prime zeta funcion, *Nordisk Tidskr. Informationsbehandling* (BIT) **8** (1968), 187—202.
4. H. L. MONTGOMERY and VAUGHAN R. C., The large sieve, *Mathematika* **20** (1973), 119—134.
5. ROSSER J. B. and SCHOENFELD L., Approximate formulas for some functions of prime numbers, *Illinois J. Math.* **6** (1962), 64—94.
6. ROSSER J. B. and SCHOENFELD, Sharper bounds for the Chebyshev functions $\theta(x)$ and $\psi(x)$ *Math. Comp.* **29** (1975), 243—269.
7, VAUGHAN R. C. On the estimation of Schnirelman's constant, *J. reine ang. Math.* **290** (1977), 93—108.

Received October 8, 1981

Royal Institute of Technology.
S—10044 Stockholm
Imperial College,
London, SW7 2BZ