ON THE REPRESENTATION OF LARGE EVEN INTEGERS AS SUMS OF TWO ALMOST PRIMES. I

By

Akiyo TOGASHI and Saburô UCHIYAMA

The classical Goldbach problem, which still survives unsolved, is to prove that every even integer $\geqq 6$ is a sum of two prime numbers. In 1948 A. Rényi [4] succeeded, by making use of his refinement of the large sieve of Yu. V. Linnik, in proving that every even integer $\geqq 6$ is a sum of a prime and of an almost prime. Here an almost prime is a positive integer (>1) the total number of prime factors of which is bounded by a certain constant. Recently this result was sharpened in part by Ch.-D. Pan [3], who showed that every sufficiently large even integer can be represented as a sum of a prime and of an almost prime possessing at most five prime factors.

On the other hand, A. A. Buhštab [1] has proved that every large even integer can be written as a sum of two almost primes, each of which is composed of at most four prime factors. The purpose of the present paper is to improve this result of Buhštab. Indeed, we shall prove the following

Theorem. *Every sufficiently large even integer is representable as a sum of two integers, each of which has not more than three prime factors.*

We know that this result is originally due to A. I. Vinogradov [7]. However, as has been reviewed by H. Davenport [2], the exposition of Vinogradov in [7] does not seem to be quite clear. Thus it will be worth while, we believe, to give another proof for the theorem. Our proof of the above theorem is based on a combination of the sieve methods of Viggo Brun and of A. Selberg: in fact, it is substantially a deduction from an intermediate result obtained by Buhštab [1].

It should be noted that the following result can also be proved by the same argument *mutatis mutandis*: for every fixed integral value of $k \geqq 1$ there are infinitely many pairs of integers m, $m+2k$ with $V(m) \leqq 3$, $V(m+2k) \leqq 3$, where $V(n)$ denotes the total number of prime factors of n.

1. Throughout in this paper the letters d, k, m, n, r are used to denote positive integers, p, q to denote prime numbers, and x to denote a positive real number. $\varepsilon\left(0<\varepsilon<\frac{1}{30}\right)$ is a sufficiently small positive real number. c will represent positive constants not necessarily the same in each occurrence; the constants implied in the symbol O may depend only upon the parameter ε.

Let $N\geqq N_0$ be a fixed even integer, where $N_0=N_0(\varepsilon)>3$. We consider the $\varphi(N)$ integers $a_n=n(N-n)$ $(1\leqq n\leqq N,\ (n, N)=1)$. Denote by S_d the number of those $a_n(1\leqq n\leqq N,\ (n, N)=1)$ which are divisible by d.

Lemma 1. *We have*

$$S_d=\frac{\omega(d)}{d}\varphi(N)+R(d),$$

where

$$\omega(d)=\begin{cases} 2^{v(d)} & \text{if } (d, N)=1, \\ 0 & \text{if } (d, N)>1, \end{cases}$$

and

$$|R(d)|\leqq B_N\omega(d) \quad \textit{with} \quad B_N=c2^{v(N)}.$$

Here $v(m)$ denotes the number of distinct prime divisors of m. Lemma 1 is essentially the same as [5, Lemma 1].

We now put $f(m)=m/\omega(m)$ and define $f_1(m)$ by

$$f_1(m)=\sum_{d|m}\mu(d)f\left(\frac{m}{d}\right).$$

The functions $f(m)$ and $f_1(m)$ are multiplicative functions of m, and we find easily that $1<f(m)\leqq\infty$ for $m>1$ and if $\mu^2(m)=1$ then

$$f_1(m)=f(m)\prod_{p|m}\left(1-\frac{1}{f(p)}\right).$$

For the sake of convenience let us set

$$C_N=\frac{1}{8}\prod_{p>2}\frac{(p-1)^2}{p(p-2)}\prod_{\substack{p|N\\ p>2}}\frac{p-2}{p}.$$

It is easy to see that

$$c>C_N>c(\log\log N)^{-2}.$$

Lemma 2. *For $x\geqq 2$ we have*

$$\prod_{p\leqq x}\left(1-\frac{1}{f(p)}\right)=\frac{A_N}{\log^2 x}+O\left(\frac{(\log\log N)^2}{\log^3 x}\right)+O\left(\frac{\log N(\log\log N)^2}{x\log^2 x}\right),$$

where $A_N=(2e^{2C}C_N)^{-1}$, *C being the Euler constant.*

Proof. We have

$$\prod_{p\leq x}\left(1-\frac{1}{f(p)}\right)=\prod_{\substack{p\leq x\\ p\nmid N}}\left(1-\frac{2}{p}\right)=\prod_{2<p\leq x}\left(1-\frac{2}{p}\right)\prod_{\substack{2<p\leq x\\ p|N}}\left(1-\frac{2}{p}\right)^{-1},$$

where

$$\prod_{2<p\leq x}\left(1-\frac{2}{p}\right)=4\prod_{2<p\leq x}\frac{p(p-2)}{(p-1)^2}\prod_{p\leq x}\left(1-\frac{1}{p}\right)^2.$$

Now

$$\prod_{\substack{2<p\leq x\\ p|N}}\left(1-\frac{2}{p}\right)^{-1}=\prod_{\substack{p>2\\ p|N}}\left(1-\frac{2}{p}\right)^{-1}\cdot\left(1+O\left(\frac{\log N}{x}\right)\right)$$
$$=\prod_{\substack{p|N\\ p>2}}\frac{p}{p-2}+O\left(\frac{\log N(\log\log N)^2}{x}\right);$$

$$\prod_{2<p\leq x}\frac{p(p-2)}{(p-1)^2}=\prod_{p>2}\frac{p(p-2)}{(p-1)^2}\cdot\left(1+O\left(\frac{1}{x}\right)\right)$$
$$=\prod_{p>2}\frac{p(p-2)}{(p-1)^2}+O\left(\frac{1}{x}\right);$$

$$\prod_{p\leq x}\left(1-\frac{1}{p}\right)^2=\frac{e^{-2C}}{\log^2 x}+O\left(\frac{1}{\log^3 x}\right).$$

Gathering up these results, we obtain Lemma 2.

Lemma 3. *For* $x\leq N$ *we have*

$$\sum_{\substack{m\leq x\\ (m,k)=1}}\frac{\mu^2(m)}{f_1(m)}=C_{kN}\log^2 x+O(\log kN\log\log kN).$$

This is [5, Lemma 3].

2. Let $P(N^{\frac{1}{u}})$ $(u\geq 2)$ denote the number of those integers $a_n=n(N-n)$ $(1\leq n\leq N,\ (n,N)=1)$ which are not divisible by any prime $p\leq N^{\frac{1}{u}}$. We evaluate $P(N^{\frac{1}{u}})$ for some values of u by Brun's method, just as in Buhštab [1].

Let $\varepsilon\left(0<\varepsilon<\dfrac{1}{30}\right)$ be small enough and put $h=(5-30\varepsilon)/(4-30\varepsilon)$. For $x\geq x_0=x_0(\varepsilon)$ we have

$$0.4463>\tau\overset{\text{def}}{=}2\log(h+\varepsilon)>\sum_{x<p\leq x^h}\frac{1}{f(p)},$$

$$1.5630 > \lambda \overset{\text{def}}{=} (h+\varepsilon)^2 > \prod_{x<p\leqq x^h}\left(1-\frac{1}{f(p)}\right)^{-1}.$$

Then we get

$$P(N^{\frac{1}{15}}) > \left(1-\sum_{m=0}^{\infty}\frac{\lambda^{m+1}((m+1)\tau)^{2m+8}}{(2m+8)!}\right)\prod_{p\leqq N^{\frac{1}{15}}}\left(1-\frac{1}{f(p)}\right)\cdot\varphi(N)+O(N^{1-\epsilon});$$

$$> 224.9997\,A_N\frac{\varphi(N)}{\log^2 N}+O\left(\frac{\varphi(N)\,(\log\log N)^2}{\log^3 N}\right),$$

on taking account of Lemmas 1 and 2 and noticing that $B_N = O(N^{\epsilon})$ and

$$\max_{1\leqq d\leqq cN^2} R(d) = cB_N \max_{1\leqq d\leqq cN^2} 2^{v(d)} = O(N^{3\epsilon}).$$

In like manner we can show that

$$P(N^{\frac{1}{14}}) < 196.0022\,A_N\frac{\varphi(N)}{\log^2 N}+O\left(\frac{\varphi(N)\,(\log\log N)^2}{\log^3 N}\right),$$

$$P(N^{\frac{1}{12}}) < 144.1328\,A_N\frac{\varphi(N)}{\log^2 N}+O\left(\frac{\varphi(N)\,(\log\log N)^2}{\log^3 N}\right),$$

$$P(N^{\frac{1}{10}}) > 98.0\,A_N\frac{\varphi(N)}{\log^2 N}+O\left(\frac{\varphi(N)\,(\log\log N)^2}{\log^3 N}\right),$$

$$P(N^{\frac{1}{10}}) < 101.6\,A_N\frac{\varphi(N)}{\log^2 N}+O\left(\frac{\varphi(N)\,(\log\log N)^2}{\log^3 N}\right).$$

Now we can proceed along with Buhštab [1], obtaining finally the following result:

Lemma 4. *We have*

$$S_1 \overset{\text{def}}{=} P(N^{\frac{1}{6}}) > 26.4612\,A_N\frac{\varphi(N)}{\log^2 N}+O\left(\frac{\varphi(N)\,(\log\log N)^2}{\log^3 N}\right).$$

3. Let q be a fixed prime number in the interval $z<q\leqq z_1$ with $(q, N)=1$, where

$$z = N^{\frac{1}{6}}, \qquad z_1 = N^{\frac{1}{3}}.$$

We wish to evaluate from above the number $S(q)$ of those integers $a_n = n(N-n)$ $(1\leqq n\leqq N,\ (n, N)=1)$ which are not divisible by any prime $p\leqq z$ and are divisible by the prime q. Applying the generalized 'upper' sieve of A. Selberg (cf. [6, Appendix]), we find on account of Lemma 1 that

$$S(q) \leqq \frac{2\varphi(N)}{qW_q}+R_q,$$

where

$$W_q = \sum_{\substack{m \leqq z^a \\ g(m) \leqq z}} \frac{\mu^2(m)}{f_1(m)} \quad \text{with} \quad a = 3\left(1 - 2\varepsilon - \frac{\log q}{\log N}\right)$$

and

$$R_q = O(B_N z^{2a} (\log\log z)^2) = O\left(\frac{B_N N^{1-2\varepsilon} (\log\log N)^2}{q}\right).$$

In the expression for W_q the function $g(m)$ is defined as follows: $g(1)=1$ and for $m>1$ $g(m)=$ the greatest prime divisor of m.

If $0<a\leqq 1$ then we have, by Lemma 3 (with $k=1$),

$$\begin{aligned} W_q &= \sum_{m \leqq z^a} \frac{\mu^2(m)}{f_1(m)} = C_N \log^2 z^a + O(\log N \log\log N) \\ &= \frac{a^2}{36} C_N \log^2 N + O(\log N \log\log N), \end{aligned}$$

while if $a>1$ then we have, by Lemma 3 again,

$$\begin{aligned} W_q &\geqq \sum_{m \leqq z^a} \frac{\mu^2(m)}{f_1(m)} - \sum_{z<p\leqq z^a} \sum_{\substack{m \leqq z^a \\ m \equiv 0(p)}} \frac{\mu^2(m)}{f_1(m)} \\ &= \sum_{m \leqq z^a} \frac{\mu^2(m)}{f_1(m)} - \sum_{z<p\leqq z^a} \frac{1}{f_1(p)} \sum_{\substack{m \leqq z^a/p \\ (m,p)=1}} \frac{\mu^2(m)}{f_1(m)} \\ &= C_N \log^2 z^a + O(\log N \log\log N) \\ &\quad - \sum_{\substack{z<p\leqq z^a \\ (p,N)=1}} \frac{2}{p-2}\left(C_{pN} \log^2 \frac{z^a}{p} + O(\log pN \log\log pN)\right) \\ &= \frac{(2a-1)^2 - 2a^2 \log a}{36} C_N \log^2 N + O(\log N \log\log N), \end{aligned}$$

since $\dfrac{2}{p-2} C_{pN} = \dfrac{2}{p} C_N$ for $(p, N)=1$, where we have used the well-known inequalities

$$\sum_{p \leqq x} \frac{1}{p} = \log\log x + c + O\left(\frac{1}{\log x}\right),$$

$$\sum_{p \leqq x} \frac{\log p}{p} = \log x + O(1),$$

and

$$\sum_{p \leqq x} \frac{\log^2 p}{p} = \frac{1}{2} \log^2 x + O(\log x).$$

We define for $t>(1-2\varepsilon)^{-1}$ the function

$$C_\varepsilon(t)=\begin{cases}\dfrac{144e^{2C}}{a^2} & (0<a\leqq 1)\\[2ex] \dfrac{144e^{2C}}{(2a-1)^2-2a^2\log a} & (1<a\leqq 3)\end{cases}$$

where

$$a=3\left(1-2\varepsilon-\frac{1}{t}\right):$$

the function $C_\varepsilon(t)$ is positive and continuous for $t>(1-2\varepsilon)^{-1}$, since the denominator in the definition of $C_\varepsilon(t)$ is positive and continuous for $0<a\leqq 3$ (moreover, it is not difficult to see that $C_\varepsilon(t)$ is monotone decreasing for $t>(1-2\varepsilon)^{-1}$). Then it follows from the above results that

$$S(q)\leqq A_N\frac{C_\varepsilon(t_q)}{q}\,\frac{\varphi(N)}{\log^2 N}+O\left(\frac{\varphi(N)\,(\log\log N)^5}{q\log^3 N}\right)+O\left(\frac{B_N N^{1-2\varepsilon}(\log\log N)^2}{q}\right)$$

where $t_q=(\log N)/\log q$.

Now, let $r\geqq 1$ be a fixed integer and let S_2 denote the number of those integers $a_n(1\leqq n\leqq N,\ (n,\ N)=1)$ which are not divisible by any prime $p\leqq z$ and are divisible by at least $r+1$ distinct primes q in the interval $z<q\leqq z_1$, $(q,\ N)=1$. Clearly S_2 is not greater than

$$\frac{1}{r+1}\sum_{\substack{z<q\leqq z_1\\(q,N)=1}}S(q)\,.$$

Here we can argue just as in [6, §5], obtaining the evaluation

$$\sum_{\substack{z<q\leqq z_1\\(q,N)=1}}\frac{C_\varepsilon(t_q)}{q}\leqq\int_3^6\frac{C_\varepsilon(u)}{u}du+O\left(\frac{1}{\log^{1/2}N}\right).$$

We thus have proved the following lemma, since $B_N=O(N^\varepsilon)$:

Lemma 5. *We have*

$$S_2\leqq\frac{1}{r+1}\int_3^6\frac{C_\varepsilon(u)}{u}du\,A_N\frac{\varphi(N)}{\log^2 N}+O\left(\frac{\varphi(N)\,(\log\log N)^2}{\log^{5/2}N}\right).$$

4. We shall prove that for some sufficiently small $\varepsilon\left(0<\varepsilon<\dfrac{1}{30}\right)$ we have

$$I(\varepsilon)\overset{\text{def}}{=}\int_3^6\frac{C_\varepsilon(u)}{u}du\leqq 79.3026\,.$$

To accomplish this it will suffice to show that

$$I(0) < 79.3025\,,$$

since the integral $I(\varepsilon)$ is itself a continuous function of ε.

We have

$$I(0) = \int_3^6 \frac{C_0(u)}{u}\,du = 144e^{2C}\int_2^{2.5} \frac{ds}{(3-s)\,((2s-1)^2-2s^2\log s)}\,,$$

on substituting $s=3\left(1-\dfrac{1}{u}\right)$.

Lemma 6. *The function*

$$h(s) = \frac{1}{(3-s)\,((2s-1)^2-2s^2\log s)}$$

is positive and convex for $2\leqq s\leqq 2.5$.

Proof. Putting $h_1(s)=(h(s))^{-1}$, we find that

$$\begin{aligned}
h_1(s) &= -4s^3+16s^2-13s+3+2s^3\log s-6s^2\log s\,,\\
h_1'(s) &= -10s^2+26s-13+6s^2\log s-12s\log s\,,\\
h_1''(s) &= -14s+14+12s\log s-12\log s\,,\\
h_1'''(s) &= -2+12\log s-\frac{12}{s}\,.
\end{aligned}$$

Thus, $h_1'''(s)$ is monotone increasing for $s>0$ and $h_1'''(2)=12\log 2-8>0$, which implies that $h_1'''(s)>0$ for $s\geqq 2$. Hence $h_1''(s)$ is monotone increasing for $s\geqq 2$. But $h_1''(2.5)=18\log 2.5-21<0$ and, therefore, $h_1''(s)<0$ for $2\leqq s\leqq 2.5$. This implies in turn that $h_1'(s)$ is monotone decreasing for $2\leqq s\leqq 2.5$. But $h_1'(2)=-1$, so that $h_1'(s)<0$ for $2\leqq s\leqq 2.5$. Hence $h_1(s)$ is monotone decaeasing for $2\leqq s\leqq 2.5$ and, as will be seen a moment later, $h_1(2.5)>0$, which means that $h_1(s)>0$ for $2\leqq s\leqq 2.5$. Therefore

$$h''(s) = \frac{2(h_1'(s))^2-h_1(s)h_1''(s)}{h_1^3(s)} > 0$$

for $2\leqq s\leqq 2.5$. This completes the proof of the lemma.

Now we have

$$h(2) < 0.2895\,, \qquad h(2.3) < 0.3445\,,$$
$$0.4399 < h(2.5) < 0.4400\,.$$

By virtue of Lemma 6, we thus find that

$$\int_2^{2.5} h(s)ds \leqq 0.15(h(2)+h(2.3))+0.1(h(2.3)+h(2.5))$$

$$< 0.09510 + 0.07845 < 0.1736 .$$

Since $e^{2C} < 3.1723$, we finally obtain

$$I(0) = 144e^{2C}\int_2^{2.5} h(s)\,ds < 144e^{2C}\cdot 0.1736 < 79.3025 ,$$

which is the desired result.

5. We are now going to conclude our proof of the theorem. Let us fix $\varepsilon\left(0<\varepsilon<\frac{1}{30}\right)$ so small that Lemma 4 holds and that we have

$$I(\varepsilon) = \int_3^6 \frac{C_\varepsilon(u)}{u}\,du < 79.3026 ,$$

and suppose that $N \geqq N_0 = N_0(\varepsilon)$ be a sufficiently large even integer. We put $z = N^{\frac{1}{6}}$ and $z_1 = N^{\frac{1}{3}}$ as before. Then, by Lemma 5 with $r=2$, the number S_2 of those integers $a_n = n(N-n)$ $(1 \leqq n \leqq N,\ (n, N)=1)$ which are not divisible by any prime $p \leqq z$ and are divisible by at least three distinct primes q in the interval $z < q \leqq z_1$ with $(q, N)=1$ is not greater than

$$\frac{I(\varepsilon)}{3} A_N \frac{\varphi(N)}{\log^2 N} + O\left(\frac{\varphi(N)(\log\log N)^2}{\log^{5/2} N}\right)$$
$$< 26.4342\, A_N \frac{\varphi(N)}{\log^2 N} + O\left(\frac{\varphi(N)(\log\log N)^2}{\log^{5/2} N}\right).$$

Now, the number S_3 of those integers $a_n (1 \leqq n \leqq N,\ (n, N)=1)$ which are not divisible by any prime $p \leqq z$ and are divisible by some integer q^2 with q in $z < q \leqq z_1$ does not exceed

$$\sum_{z<q\leqq z_1}\left(\frac{2\varphi(N)}{q^2} + O(N^\varepsilon)\right) = O(N^{\frac{5}{6}}) + O(N^{\frac{1}{3}+\varepsilon}) = O(N^{\frac{5}{6}}) .$$

Therefore, if we denote by S the number of those integers $a_n (1 \leqq n \leqq N,$ $(n, N)=1)$ which are divisible by no primes $p \leqq z$, by at most two primes q in $z < q \leqq z_1$, and by no integers of the form q^2 with q in $z < q \leqq z_1$, then, by Lemma 4,

$$\begin{aligned} S &\geqq S_1 - S_2 - S_3 \\ &= (26.4612 - 26.4342) A_N \frac{\varphi(N)}{\log^2 N} + O\left(\frac{\varphi(N)(\log\log N)^2}{\log^{5/2} N}\right) \\ &> 0.0269\, A_N \frac{\varphi(N)}{\log^2 N} > 2 . \end{aligned}$$

Hence there exists at least one integer n with $1 < n < N-1$, $(n, N)=1$, such that $V(n) \leqq 3$, $V(N-n) \leqq 3$. Since

$$N = n + (N-n),$$

our proof of the theorem is now complete.

References

[1] A. A. BUHŠTAB: On the representation of even numbers as the sum of two numbers with a bounded number of prime factors, Dokl. Akad. Nauk SSSR, vol. 29 (1940), pp. 544–548 (in Russian). Here some results are referred to of the paper by A. A. Buhštab: New improvements in the sieve method of Eratosthenes, Matem. Sbornik, vol. 4 (46) (1938), pp. 375–387 (in Russian).

[2] Review by H. DAVENPORT to the paper [7] below of A. I. Vinogradov, Math. Reviews, vol. 20 (1959), #3836, p. 634.

[3] CH.-D. PAN: Representation of even integers as sums of a prime and an almost prime, Acta Math. Sinica, vol. 12 (1962), pp. 95–106 (in Chinese).

[4] A. RÉNYI: On the representation of an even number as the sum of a prime and of an almost prime, Izv. Akad. Nauk SSSR, Ser. Mat., vol. 12 (1948), pp. 57–78 (in Russian). An English version of this paper can be found in the Amer. Math. Soc. Translations, Ser. 2, vol. 19, Amer. Math. Soc., Providence, R.I., U.S.A., 1962, pp. 299–321.

[5] H. N. SHAPIRO and J. WARGA: On the representation of large integers as sums of primes, Part I, Comm. on Pure and Appl. Math., vol. 3 (1950), pp. 153–176.

[6] S. UCHIYAMA: On the distribution of almost primes in an arithmetic progression, J. Fac. Sci., Hokkaidô Univ., Ser. I, vol. 18 (1964), pp. 1–22.

[7] A. I. VINOGRADOV: Application of $\zeta(s)$ to the sieve of Eratosthenes, Matem. Sbornik, vol. 41 (83) (1957), pp. 49–80; Correction, ibid., pp. 415–416 (in Russian).

Department of Mathematics,
Hokkaidô University

(Received February 12, 1964)