Mathematics. — *On additive properties of squares of primes.* I. By P. ERDÖS. (Communicated by Prof. J. G. VAN DER CORPUT).

(Communicated at the meeting of November 27, 1937.)

Introduction.

In three papers, one of which is published [1]), and the other two of which will appear shortly, I proved the following results: The densities of the sets of integers $p^2 + q^2 - r^2$ (p, q, r, odd primes) and $p^2 - q^2 + 2^l$ are positive. In the introduction of I, I stated the following conjecture: The density of the integers of the form $p^2 + q^2 + r^2$ is positive. Now I have succeeded in proving this conjecture and the present paper will contain proofs of the following 4 theorems [2]):

1. The number of integers not exceeding n, of the form $p^2 + q^2$ is greater than $c_1 \frac{n}{(\log n)^2}$.

2. There exists an infinity of integers m such that the equation $m = p^2 + q^2$ has one and only one solution.

3. The density of the integers of the form $p^2 + q^2 + r^2$ is positive.

4. The density of the integers of the form $p^2 + q^2 + 2^{l_1} + 2^{l_2}$ is positive. It is clear that 1. follows from 4. Nevertheless we give an independent proof of 1, partly because it will help to make clear the method and partly because we shall be able to deduce 2. from it. From 3. and a well known result of SCHNIRELMANN we immediately deduce that a constant c_2 exists such that every integer is the sum of c_2 or less positive squares of primes. The chief importance of 4. lies in the fact that the number of integers of the form $p^2 + q^2 + 2^{l_1} + 2^{l_2}$ not exceeding n does not exceed $c_3\, n$.

Throughout this paper n denotes a sufficiently large integer, $c_1, c_2, \ldots$ and γ positive absolute constants, γ will be used only as an exponent of n.

§ 1.

We require the following

Lemma 1. *Let $e_1, e_2, \ldots, e_k$; $f_1, f_2, \ldots, f_k$ be integers with $|e_i|, |f_j| < n^{c_4}$, and for no i and j $e_i = e_j$, $f_i = f_j$ at the same time, then the number A of pairs of positive integers x and y, not exceeding n, for which*

$$e_1 x + f_1 y,\ e_2 x + f_2 y, \ldots, e_k x + f_k y \quad . \quad . \quad . \quad . \quad . \quad (1)$$

are all primes, is less than $c_5 \frac{n^2 (\log \log n)^{k+1}}{(\log n)^k}$.

1) On the easier WARING problem for powers of primes, Proc. Cambridge Phil. Soc. **33**, 6—12 (1937). I shall refer to this paper as I.

2) In connection with all these problems, see VINOGRADOV, Comptes Rendus de l'Acad. des Sciences de l'U. S. S. R., **16**, 131—132 (1937).

Proof. We first estimate for fixed y the number A_y of x not exceeding n, for which all the integers (1) are primes. We may evidently suppose $(e_i, f_i) = 1$, $i = 1, 2, \ldots k$. Let p be any prime not exceeding n^γ with $p \nmid e_1 e_2 \ldots e_k$ and $|e_i x + f_i y| > n^\gamma$ $(i = 1, 2, \ldots, k)$, then

$$x \not\equiv -\frac{f_1 y}{e_1}, -\frac{f_2 y}{e_2}, \ldots, -\frac{f_k y}{e_k} \pmod p,$$

since $e_i x + f_i y$ are all primes.

These k residues are all different if $p \nmid f(y)$, where

$$f(y) = y e_1 e_2 \ldots e_k \prod_{i<j\leq k} (e_i f_j - e_j f_i).$$

We may suppose $f(y) \neq 0$ for evidently $e_i f_j - e_j f_i \neq 0$).

Thus, by Lemma 2. of I and by the fact that there are at most $2kn^\gamma$, values of x for which one of the inequalities $|e_i x + f_i y| > n^\gamma$ $(i = 1, 2 \ldots k)$ is not true, we obtain

$$A_y < c_6 n \prod_{\substack{p<n^\gamma \\ p \nmid f(y)}} \left(1 - \frac{k}{p}\right) + 2kn^\gamma. \quad \ldots \ldots \quad (2)$$

By applying the inequality

$$1 < \left(1 - \frac{k}{p}\right)\left(1 + \frac{k+1}{p}\right), \qquad p > k(k+1)$$

for the primes $p > k(k+1)$ dividing $f(y)$ we obtain from (2)

$$A_y < c_6 n \prod_{k(k+1)<p<n^\gamma} \left(1 - \frac{k}{p}\right) \prod_{p|f(y)} \left(1 + \frac{k+1}{p}\right) + 2kn^\gamma < c_5 \frac{n(\log\log n)^{k+1}}{(\log n)^k} \quad (3)$$

since

$$\prod_{k(k+1)<p<n^\gamma} \left(1 - \frac{k}{p}\right) < \frac{c_7}{(\log n)^k}$$

and

$$\prod_{p|f(y)} \left(1 + \frac{k+1}{p}\right) < c_8 (\log\log f(y))^{k+1} < c_9 (\log\log n)^{k+1}.$$

Thus, from (3),

$$A = \sum_{y=1}^{n} A_y < c_5 \frac{n^2 (\log\log n)^{k+1}}{(\log n)^k},$$

which proves the lemma.

Theorem I. *The number B of integers m, not exceeding n, of the form $p^2 + q^2$ is greater then* $c_1 \dfrac{n}{(\log n)^2}$.

We prove a more general result containing at the same time Theorem II. namely:

The number B' of integers m not exceeding n for which the equation

$$m = p^2 + q^2 \text{ with } p \geqq q$$

has one and only one solution is greater than $c_{10} \dfrac{n}{(\log n)^2}$.

Proof. Denote by $\varphi(m)$ the number of solutions of the equation

$$m = p^2 + q^2 \text{ with } p \geqq q;$$

we then have

$$B' = \sum_{m=1}^{n} \varphi(m) - \sum_{\substack{m=1 \\ \varphi(m) \geqq 2}}^{n} \varphi(m) \geqq \sum_{m=1}^{n} \varphi(m) - \sum_{m=1}^{n} ([\varphi(m)]^2 - \varphi(m)). \quad (4)$$

Evidently

$$\sum_{m=1}^{n} \varphi(m) > \tfrac{1}{2} [\pi (1/2\, n^{1/2})]^2 > c_{11} \frac{n}{(\log n)^2}. \quad . \quad . \quad . \quad . \quad (5)$$

Next we prove

$$\sum_{m=1}^{n} ([\varphi(m)]^2 - \varphi(m)) < \frac{n}{(\log n)^{5/2}}. \quad . \quad . \quad . \quad . \quad . \quad (6)$$

Denote by C the number of solutions of

$$p^2 + q^2 = r^2 + s^2 \qquad (p^2 + q^2 \leqq n, p < r); \quad . \quad . \quad . \quad . \quad (7)$$

then

$$\sum_{m=1}^{n} ([\varphi(m)]^2 - \varphi(m)) = 2C,$$

so that it will suffice to estimate C.

We write (7) in the form

$$r^2 - p^2 = q^2 - s^2 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

and put

$$r - p = 2a, \quad q - s = 2b \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

Evidently $a, b < n^{1/2}$. We may suppose $a > b > 0$.

By (8) and (9),

$$ap - bs = b^2 - a^2,$$

or

$$p = \frac{b(s + b)}{a} - a.$$

Put $(a, b) = d$, $\dfrac{a}{d} = a'$, $\dfrac{b}{d} = b'$, then

$$p = \frac{b'(s + b'd)}{a'} - a'd \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

From (10) we obtain

$$s \equiv -b' d \pmod{a'},$$

which means

$$s = -b' d + a' x \quad \dots\dots\dots \quad (11)$$

From (11), (10) and (9) we have

$$\left.\begin{aligned} p &= -a' d + b' x \\ r &= a' d + b' x \\ q &= b' d + a' x \end{aligned}\right\} \quad \dots\dots\dots \quad (12)$$

Denote now by D the number of solutions of (11) and (12) with

$$1 \leqslant a' d \leqslant n^{1/2}, \quad 1 \leqslant a' x \leqslant n^{1/2} \quad (a' > b' > 0) \quad \dots \quad (13)$$

then evidently

$$D \geqslant C$$

(for if a', d, x do not satisfy (13) then at least one of the primes p, q, r, s is greater than $n^{1/2}$).

Hence it will be sufficient to estimate D.

Now we write

$$D = D' + D'',$$

where D' denotes the number of solutions of (11) and (12) with $d \leqslant n^{1/4}$, and D'' the number of solutions of (11) and (12) with $n^{1/4} < d \leqslant n^{1/2}$.

First we estimate D'. Denote by $D'_{d,x}$ the number of solutions in a', b' of (11), and (12) for fixed d and x.

From Lemma 1. by putting

$$e_1 = -d, \; e_2 = e_3 = x, \; e_4 = d, \; f_1 = f_4 = x, \; f_2 = -d, \; f_3 = d, \; b' = x, \; a' = y$$

and replacing n by $\dfrac{n^{1/2}}{\max(x, d)}$, we obtain

$$D'_{d,x} < \frac{c_{12}\, n \,(\log \log n)^5}{[\max(x, d)]^2 \,(\log n)^4} \quad \text{for } x \leqslant n^{3/8} \quad \dots \quad (14)$$

For $x > n^{3/8}$ we evidently have

$$D'_{d,x} < \frac{n}{x^2} \quad \dots\dots\dots \quad (15)$$

From (14) and (15) we obtain

$$\left.\begin{aligned} D'_d = \sum_{x=1}^{n^{1/2}} D'_{d,x} &= \sum_{x=1}^{d} D'_{d,x} + \sum_{x>d}^{n^{3/8}} D'_{d,x} + \sum_{x=n^{3/8}}^{n^{1/2}} D'_{d,x} \\ &< \frac{c_{12}\, n\,(\log\log n)^5}{d\,(\log n)^4} + \frac{c_{13}\, n\,(\log\log n)^5}{d\,(\log n)^4} + 2\,n^{5/8} \end{aligned}\right\} \quad . \; . \quad (16)$$

Finally from (16)

$$D' = \sum_{d=1}^{n^{1/4}} D'_d < \frac{c_{14}\, n\,(\log\log n)^5}{(\log n)^3} + 2\,n^{7/8} < c_{15}\,\frac{n\,(\log\log n)^5}{(\log n)^3} \quad . \; . \; . \; . \quad (17)$$

Now we estimate D''. Denote by $D''_{a',b',d}$ the number of solutions in x of (11) and (12) for fixed a', b', d.
By putting

$$-b'd = f_1 y, \quad -a'd = f_2 y, \quad a'd = f_3 y, \quad b'd = f_4 y, \quad e_1 = e_4 = a', \quad e_2 = e_3 = b',$$

and replacing n by $\dfrac{n^{1/2}}{a'}$ $\left(a' \leqslant \dfrac{n^{1/2}}{d} < n^{1/4}\right)$ we obtain from (3)

$$D''_{a',b',d} < c_{16}\,\frac{n^{1/2}\,(\log\log n)^5}{a'\,(\log n)^4}.$$

Now

$$D''_{a',d} = \sum_{b'=1}^{a'} D''_{a',b',d} < c_{16}\,\frac{n^{1/2}\,(\log\log n)^5}{(\log n)^4},$$

and

$$D''_d = \sum_{a'=1}^{\frac{n^{1/2}}{d}} D''_{a',d} < c_{16}\,\frac{n\,(\log\log n)^5}{d\,(\log n)^4}$$

Thus finally

$$D'' = \sum_{d>n^{1/4}}^{n^{1/2}} D''_d < c_{17}\,\frac{n\,(\log\log n)^5}{(\log n)^3} \quad . \; . \; . \; . \; . \; . \quad (18)$$

From (17) and (18) we obtain

$$D = D' + D'' < (c_{15} + c_{17})\,\frac{n\,(\log\log n)^5}{(\log n)^3} < \frac{n}{(\log n)^{5/2}},$$

which proves (6).
From (4), (5) and (6) we obtain

$$B' > c_{11}\,\frac{n}{(\log n)^2} - \frac{n}{(\log n)^{5/2}} > c_{10}\,\frac{n}{(\log n)^2}.$$

Hence the result.

(To be continued).