Det Kgl. Danske Videnskabernes Selskab.
Mathematisk-fysiske Meddelelser. **VIII,** 1.

ON THE LOGARITHMIC DERIVATIVES OF THE GAMMA FUNCTION

BY

EINAR HILLE

KØBENHAVN
HOVEDKOMMISSIONÆR: ANDR. FRED. HØST & SØN, KGL. HOF-BOGHANDEL
BIANCO LUNOS BOGTRYKKERI
1927

Pris: Kr. 2,40.

Det Kgl. Danske Videnskabernes Selskabs videnskabelige Meddelelser udkommer fra 1917 indtil videre i følgende Rækker:

Historisk-filologiske Meddelelser,
Filosofiske Meddelelser,
Mathematisk-fysiske Meddelelser,
Biologiske Meddelelser.

Hele Bind af disse Rækker sælges 25 pCt. billigere end Summen af Bogladepriserne for de enkelte Hefter.

Selskabets Hovedkommissionær er *Andr. Fred. Høst & Søn*, Kgl. Hof-Boghandel, København.

Det Kgl. Danske Videnskabernes Selskab.

Mathematisk-fysiske Meddelelser. **VIII,** 1.

ON THE LOGARITHMIC DERIVATIVES OF THE GAMMA FUNCTION

BY

EINAR HILLE

KØBENHAVN

HOVEDKOMMISSIONÆR: ANDR. FRED. HØST & SØN, KGL. HOF-BOGHANDEL

BIANCO LUNOS BOGTRYKKERI

1927

1. Introduction. The difference calculus has led to the introduction into analysis of new classes of functions defined as solutions of equations of the type

$$\underset{\omega}{\Delta} F(z) = \varphi(z)$$

or of difference equations of higher order. Among the simplest and most important of the functions defined in this manner is $\psi(z)$, the logarithmic derivative of the gamma function.

The central role played by $\psi(z)$ in the difference calculus, as well as its importance for analysis in general, would seem to justify a detailed study of the properties of this function. Most of these have been known a long time, but there are still some problems outstanding. In the present paper we undertake an investigation of the distribution of the values taken on by $\psi(z)$ and of the corresponding conformal mapping. This problem requires a detailed study of the properties of $\psi'(z)$ and in particular of the zeros of this function. In Part I of the paper we are chiefly concerned with a determination of regions in the plane where the real part of $\psi(z)$ is positive. The study of $\psi'(z)$ follows in Part II; the main problem is attacked in Part III.[1]

[1] The present investigation was undertaken at the suggestion of Professor N. E. Nörlund. I should like to use this opportunity to express my gratitude to Professor Nörlund and to all the Copenhagen mathematicians for their friendly interest and for the cordial reception which they have given me.

Part I.

A Preliminary Study of $\psi(z)$.

2. Formal properties of $\psi(z)$. The function $\psi(z)$ is defined as that principal solution of the equation

$$\Delta F(z) = \frac{1}{z}$$

which assumes the value $-C$ for $z = +1$, where $C = 0.5772156649\ldots$ is Euler's constant. We have

$$\psi(z) = -C + \sum_{n=0}^{\infty}\left[\frac{1}{n+1} - \frac{1}{n+z}\right]. \tag{1}$$

Of the many relations satisfied by $\psi(z)$ we notice the following

$$\psi(z+1) = \psi(z) + \frac{1}{z}, \tag{2}$$

$$\psi(1-z) = \psi(z) + \pi \cot \pi z, \tag{3}$$

$$m\,\psi(mz) = \sum_{n=0}^{m-1} \psi\left(z + \frac{n}{m}\right) + m \log m, \tag{4}$$

$$\lim_{\varrho \to \infty}\left[\psi(z) - \log z\right] = 0. \tag{5}$$

Here m is a positive integer and $\log z$ denotes the principal determination of the logarithm; ϱ is the least distance of z from the negative real axis. Let us write

$$\psi(x+iy) = R(x,y) + i\,I(x,y), \tag{6}$$

where

$$R(x,y) = -C + \sum_{n=0}^{\infty}\left[\frac{1}{n+1} - \frac{n+x}{(n+x)^2+y^2}\right], \tag{7}$$

(8) $$I(x, y) = y \sum_{n=0}^{\infty} \frac{1}{(n+x)^2 + y^2}.$$

In view of formulas (2)—(4) these functions satisfy the following relations:

(9) $$R(x+1, y) = R(x, y) + \frac{x}{x^2+y^2},$$

(10) $$I(x+1, y) = I(x, y) - \frac{y}{x^2+y^2},$$

(11) $$\begin{cases} R(1-x, -y) = R(1-x, y) = \\ = R(x, y) + \pi \cot \pi x \dfrac{\coth^2 \pi y - 1}{\cot^2 \pi x + \coth^2 \pi y}, \end{cases}$$

(12) $$\begin{cases} I(1-x, -y) = -I(1-x, y) = \\ = I(x, y) - \pi \coth \pi y \dfrac{\cot^2 \pi x + 1}{\cot^2 \pi x + \coth^2 \pi y}, \end{cases}$$

(13) $$m R(mx, my) = \sum_{n=0}^{m-1} R\left(x + \frac{n}{m}, y\right) + m \log m,$$

(14) $$m I(mx, my) = \sum_{n=0}^{m-1} I\left(x + \frac{n}{m}, y\right),$$

(15) $$\lim_{\varrho \to \infty} \left[R(x, y) - \log |z|\right] = 0,$$

(16) $$\lim_{\varrho \to \infty} \left[I(x, y) - \arg z\right] = 0.$$

For particular values of x we can express $I(x, y)$ in terms of elementary functions. Thus

(17) $$I(0, y) = \frac{\pi}{2} \coth \pi y + \frac{1}{2y},$$

(18) $$I\left(\frac{1}{2}, y\right) = \frac{\pi}{2} \operatorname{th} \pi y.$$

The former relation is obtainable from (10) and (12) by letting $x \to 0$, the latter from (12) by putting $x = \frac{1}{2}$.[1]

For purposes of numerical calculation we shall use the following relation[2]

$$(19)\quad \psi(z) = \log z - \frac{1}{2z} - \sum_{\nu=1}^{m} \frac{B_{2\nu}}{2\nu z^{2\nu}} + \int_0^\infty \frac{\bar{B}_{2m}(t)\,dt}{(t+z)^{2m+1}}.$$

Here $B_2, B_4, \ldots$ are the Bernoullian numbers; $\bar{B}_{2m}(t)$ is that periodic function of period unity which on the interval (0, 1) coincides with $B_{2m}(t)$, the Bernoullian polynomial of order $2m$. We shall use this formula for purely imaginary values of z. Setting $z = iy$ we get

$$(20)\quad R(0, y) = \log|y| + \sum_{\nu=1}^{m} \frac{|B_{2\nu}|}{2\nu y^{2\nu}} + \Re \int_0^\infty \frac{\bar{B}_{2m}(t)\,dt}{(t+iy)^{2m+1}},$$

where the absolute value of the remainder is less than

$$(21)\quad \operatorname{Max}|B_{2m}(t)| \int_0^\infty \frac{dt}{|t+iy|^{2m+1}} = \frac{2\cdot 4\cdot 6\ldots(2m-2)}{1\cdot 3\cdot 5\ldots(2m-1)} \cdot \frac{|B_{2m}|}{y^{2m}}.$$

Finally we shall have some use for the following factorial series

$$(22)\quad \psi(z+h) - \psi(z) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n+1} \frac{h(h-1)\ldots(h-n)}{z(z+1)\ldots(z+n)},$$

which converges when $\Re(z) > 0$ and $\Re(z+h) > 0$.[3]

[1] I am indebted to Professor N. E. Nörlund for formula (18) which will be found useful below.

[2] See N. E. Nörlund: Vorlesungen über Differenzenrechnung, Berlin, J. Springer, 1924, p. 106. All the fundamental formulas for $\psi(z)$ which we use in the present paper are to be found in this book, chiefly in Chapter Five.

[3] See Nörlund, l. c. p. 251.

3. Properties of $R(x, y)$ and $I(x, y)$. It follows from (7) that

(23) $$I(x, y) \gtreqless 0 \text{ according as } y \gtreqless 0.$$

Hence all the zeros of $\psi(z)$ are real. As

$$\psi'(x) = \sum_{n=0}^{\infty} \frac{1}{(n+x)^2} > 0$$

for all real values of x, we conclude that $\psi(z)$ vanishes once and only once on each of the intervals $(-n-1, -n)$, $n = 0, 1, 2, \ldots$ and in addition once on the positive real axis. The positive zero x_0 lies between 1 and 2; it was computed by Gauss and Legendre who found $x_0 = 1.46163 \ldots$

Substituting $z = -n - \frac{1}{2}$ in (3) we find that

(24) $$\psi\left(-n-\frac{1}{2}\right) = \psi\left(n+\frac{3}{2}\right) > 0 \quad \text{for } n = 0, 1, 2, \ldots$$

It follows that the zero x_n of $\psi(z)$ on the interval $(-n, -n+1)$ lies on the left half of this interval. With the aid of (3) in conjunction with (5) we conclude that

(25) $$x_n \sim -n + \frac{1}{\log n}.$$

All these facts are of course well known. We shall now take up a detailed discussion of $R(x, y)$. It follows from (7) that

(26) $$R(x, y_1) > R(x, y_2) \text{ when } x \geqq 0 \text{ and } |y_1| > |y_2|.$$

Hence in particular

(27) $$R(x, y) > 0 \quad \text{when} \quad x \geqq x_0, \ z \neq x_0.$$

Using formula (9) we conclude that

(28) $$R(x, y) \lesseqgtr R(x+1, y) \text{ according as } x \gtreqless 0.$$

Further, formula (11) implies that

(29) $$R(x, y) \begin{cases} > R(1-x, y) & \text{when } n+\frac{1}{2} < x < n+1, \\ = R(1-x, y) & \text{when } x = n+\frac{1}{2} \text{ or } n+1, \\ < R(1-x, y) & \text{when } n < x < n+\frac{1}{2}. \end{cases}$$

Here n is an arbitrary integer including zero. Suppose that $n \leqq -2$, then (29) together with (27) implies that

(30) $$R(x, y) > 0 \text{ when } -n-\frac{1}{2} \leqq x \leqq -n, \; n = 1, 2, 3, \ldots$$

If we set $n = -1$ in (29) we merely get that $R(x, y) > 0$ when $-\frac{1}{2} \leqq x \leqq 1 - x_0$.

The result stated in formula (30) can be improved upon; in fact we have

(31) $$R(x, y) > 0 \text{ when } x_{n+1} \leqq x \leqq -n, \; n = 1, 2, 3, \ldots,$$

where, as above, x_{n+1} denotes the zero of $\psi(z)$ on the interval $(-n-1, -n)$. It is evidently sufficient to prove that $R(x, y)$ is positive for $x_{n+1} \leqq x \leqq -n-\frac{1}{2}$, as the remainder of the interval is already taken care of. But this follows from formula (11). We have

$$R(x, y) = R(1-x, y) - \pi \cot \pi x \frac{\coth^2 \pi y - 1}{\coth^2 \pi y + \cot^2 \pi x}.$$

Let x be fixed on the interval $\left(-n-1, -n-\frac{1}{2}\right)$. The first term on the right hand side is always positive and increases with $|y|$. The second term is also positive, but decreases when $|y|$ increases. Consequently

$$(32)\quad \begin{cases} R(x, y_1) > R(x, y_2) \text{ when } -n-1 \leqq x \leqq -n-\dfrac{1}{2} \\ \qquad\qquad \text{and } |y_1| > |y_2|. \end{cases}$$

If we set $y_2 = 0$ in (32) and assume $x_{n+1} \leqq x \leqq -n-\frac{1}{2}$, then the right hand side is positive; this suffices to prove (31).

Next we proceed to prove that

$$(33)\qquad R(x, y) > 0 \text{ when } x \leqq -\frac{3}{4},\ |y| \geqq \frac{1}{2}.$$

For this purpose we again use formula (11). Let us give y a fixed positive value and vary x, then

$$(34)\quad -\frac{\pi}{sh\, 2\pi y} \leqq \pi \cot \pi x \frac{\coth^2 \pi y - 1}{\coth^2 \pi y + \cot^2 \pi x} \leqq \frac{\pi}{sh\, 2\pi y}.$$

If $|y| \geqq \frac{1}{2}$, (34) implies that $R(x, y)$ differs from $R(1-x, y)$ by at most $\frac{\pi}{sh\,\pi} = 0.2704$. But if $x \leqq -\frac{3}{4}$ and $|y| \geqq \frac{1}{2}$, $R(1-x, y) \geqq R\left(\frac{7}{4}, \frac{1}{2}\right)$. In fact, the least value of $R(1-x, y)$ in the region in question must be reached on the boundary. In view of (26) the least value on the vertical boundary is to be found at the lowest point. The horizontal boundary remains. Consider formula (7) with $x \geqq \frac{7}{4}$ and $y = \frac{1}{2}$. All the terms

$$\frac{n+x}{(n+y)^2 + \frac{1}{4}} \qquad (n = 0, 1, 2, \ldots)$$

will then be decreasing functions of x when x increases. Hence the least value of $R\left(1-x, \frac{1}{2}\right)$ for $x \leqq -\frac{3}{4}$ will be reached at $x = -\frac{3}{4}$. It is difficult to estimate the size of $R\left(\frac{7}{4}, \frac{1}{2}\right)$ without computation so we use the computed value 0.3136 (>0.2704), to be found in Table I on p. 53. Hence (33) is true. The same type of argument can be used in order to show that $R(x, y) > 0$ when $x \leqq 0$, $|y| > 1$. There is some doubt whether or not $R(x, y)$ will take on negative values on the line segment from $-\frac{3}{4} + \frac{i}{2}$ to $-\frac{1}{2} + \frac{i}{2}$.

Now let us assume that $x \leqq -n - \frac{1}{2}$ where n is an integer $\geqq 3$, and that $|y| \geqq y_0 > 0$. In view of (11) and (34) we have that

$$\begin{aligned} R(x, y) &\geqq R\left(n + \frac{3}{2},\, y_0\right) - \frac{\pi}{sh\, 2\pi y_0} \\ &> R\left(n + \frac{3}{2},\, 0\right) - \frac{\pi}{sh\, 2\pi y_0} \\ &= -C - 2\log 2 + 2\left[1 + \frac{1}{3} + \dots + \frac{1}{2n+1}\right] - \frac{\pi}{sh\, 2\pi y_0} \\ &> -C - \log 2 + \log\left(n + \frac{3}{2}\right) - \frac{\pi}{sh\, 2\pi y_0}. \end{aligned}$$

Thus, in order that $R(x, y)$ be positive when $x \leqq -n - \frac{1}{2}$, $|y| \geqq y_0$, it is sufficient that

$$sh\, 2\pi y_0 \geqq \pi \left[\log\left(n + \frac{3}{2}\right) - C - \log 2\right]^{-1}.$$

Hence, a fortiori,

$$(35)\quad \begin{cases} R(x, y) > 0 \text{ when} \\ x \leqq -n - \frac{1}{2} \text{ and } |y| \geqq \frac{1}{2}\left[\log\left(n + \frac{3}{2}\right) - C - \log 2\right]^{-1}. \end{cases}$$

Formula (35) gives a better estimate than (34) when $n \geqq 9$. Thus we see that the region in the neighborhood of $z = -n$ where $R(x, y) < 0$, contracts indefinitely when $n \to \infty$. Its maximum diameter is

$$O\left(\frac{1}{\log n}\right).$$

The arcs on which $R = 0$ contract steadily to zero in the following sense. Consider the arc of $R = 0$ on which $-n \leq x \leq x_n \, (n \geq 1)$ which arc we denote by R_n. Let us imagine that R_n be moved parallel to the real axis a distance of one unit to the left. The transferred curve will then completely enclose R_{n+1}, the two curves having only the point $z = -n-1$ in common. This follows from (9). In fact, if $z+1$ is on R_n then $R(x+1, y) = 0$ and

$$R(x, y) = -\frac{x}{x^2+y^2} > 0,$$

i. e. the point z lies outside of R_{n+1} provided $z \neq -n-1$.

Part II.

Investigation of $\psi'(z)$.

4. Formal properties of $\psi'(z)$. In order to continue the discussion profitably we shall need to investigate the derivative of $\psi(z)$ in some detail and especially the location of the points where $\psi'(z) = 0$, i. e. the non-singular points where the mapping ceases to be conformal. We have

$$\psi'(z) = \sum_{n=0}^{\infty} \frac{1}{(z+n)^2}. \tag{36}$$

The most important relations satisfied by $\psi'(z)$ are the following:

$$\psi'(z+1) - \psi'(z) = -\frac{1}{z^2}, \tag{37}$$

$$\psi'(z) + \psi'(1-z) = \frac{\pi^2}{\sin^2 \pi z}, \tag{38}$$

$$\lim_{\varrho \to \infty} \left[\psi'(z) - \frac{1}{z}\right] = 0. \tag{39}$$

$$(40)\qquad m^2\,\psi'(mz) = \sum_{n=0}^{m-1} \psi'\left(z + \frac{n}{m}\right).$$

We set

$$\psi'(z) = r(x, y) + ij(x, y),$$

with

$$(41)\qquad r(x, y) = \sum_{n=0}^{\infty} \frac{(x+n)^2 - y^2}{\left[(x+n)^2 + y^2\right]^2},$$

$$(42)\qquad j(x, y) = -2y \sum_{n=0}^{\infty} \frac{x+n}{\left[(x+n)^2 + y^2\right]^2}.$$

Of the relations satisfied by $r(x, y)$ and $j(x, y)$ which are a consequence of (37)—(40) we notice the following:

$$(43)\qquad r(x+1, y) = r(x, y) - \frac{x^2 - y^2}{(x^2 + y^2)^2},$$

$$(44)\qquad j(x+1, y) = j(x, y) + \frac{2xy}{(x^2 + y^2)^2},$$

$$(45)\quad r(x, y) = -r(1-x, y) + \pi^2 \frac{\sin^2 \pi x\, ch^2 \pi y - \cos^2 \pi x\, sh^2 \pi y}{[\sin^2 \pi x + sh^2 \pi y]^2},$$

$$(46)\quad j(x, y) = -j(1-x, y) - \frac{\pi^2}{2} \frac{\sin 2\pi x\, sh\, 2\pi y}{[\sin^2 \pi x + sh^2 \pi y]^2},$$

$$(47)\qquad m^2\, r(mx, my) = \sum_{n=0}^{m-1} r\left(x + \frac{n}{m}, y\right),$$

$$(48)\qquad m^2\, j(mx, my) = \sum_{n=0}^{m-1} j\left(x + \frac{n}{m}, y\right).$$

For certain special purposes we shall need the factorial series

$$(49)\qquad \psi'(z) = \sum_{n=0}^{\infty} \frac{n!}{(n+1)\, z(z+1)\, \ldots\, (z+n)},$$

which converges when $\Re(z) > 0$. This series is easily obtainable from the corresponding series for $\psi(z+h) - \psi(z)$ in formula (22) by dividing by h and then letting h tend to zero.[1]

It is trivial to notice but useful to remember that

$$(50) \qquad r(x, y) = \frac{\partial}{\partial x} R(x, y) = \frac{\partial}{\partial y} I(x, y)$$

$$(51) \qquad j(x, y) = -\frac{\partial}{\partial y} R(x, y) = \frac{\partial}{\partial x} I(x, y).$$

5. $\psi'(z)$ **in the right half-plane.** It is obvious that

$$(52) \qquad \operatorname{sgn} j(x, y) = -\operatorname{sgn} y \quad \text{when} \quad x \geqq 0.$$

It is further clear that $\psi'(x)$ is real positive when x is real. From these two observations we conclude that $\psi'(z) \neq 0$ when $\Re(z) \geqq 0$. In the expression

$$\frac{\pi^2}{\sin^2 \pi z} = \sum_{n=-\infty}^{+\infty} \frac{1}{(z+n)^2}$$

we set $z = iy$. The result can be written in the form

$$-\frac{\pi^2}{sh^2 \pi y} = -\frac{1}{y^2} + 2 \sum_{n=0}^{\infty} \frac{n^2 - y^2}{(n^2 + y^2)^2}.$$

Hence we have

$$(53) \qquad r(0, y) = -\frac{1}{2y^2} - \frac{\pi^2}{2 sh^2 \pi y} < 0.$$

Thus $w = \psi'(z)$ maps the line $x = 0$ in the z-plane upon a curve J in the w-plane

$$u = r(0, y), \qquad v = j(0, y),$$

[1] See also Nörlund, l. c. p. 243.

which curve lies entirely in the half-plane $u < 0$ except for the point $(0, 0)$, where J is tangent to the v-axis. J does not intersect itself for $r(0, y)$ increases steadily with $|y|$; it consists of two branches symmetric with respect to the negative u-axis, which is the asymptote of both. Let the region outside of J be denoted by Δ. It will be proved in § 11 that $\psi''(z) \neq 0$ in $\Delta + J$. Thus $w = \psi'(z)$ maps the half-plane $\Re(z) > 0$ conformally upon Δ. Thus every value in Δ is taken on once and only once by $\psi'(z)$ in the right half-plane. A simple calculation shows that

$$|v| < \frac{\sqrt{3}}{16}\pi^2 = 1.069$$

on J; hence the values not taken on in $\Re(z) > 0$ have negative real part and a numerically small imaginary part.[1]

In the half-plane $\Re(z) \geqq 1$, $r(x, y) > 0$. To see this we notice first that

$$r(1, y) = r(0, y) + \frac{1}{y^2} = \frac{1}{2y^2} - \frac{\pi^2}{2\, sh^2 \pi y} > 0$$

in view of formulas (43) and (53). Thus the curve $r(x, y) = 0$ does not intersect the line $x = +1$. On the other hand, there are two branches of this curve in the right half-plane which pass through the origin, where they have the slopes $+1$ and -1 respectively, and which admit of the imaginary axis as their asymptote. Hence the branches of $r(x, y) = 0$ which lie in the right half-plane must be enclosed in the strip $0 \leqq x < +1$. It follows from formula (39) that there are no other branches of the curve $r(x, y) = 0$ in the right half-plane. Hence $r(x, y) > 0$ when $x \geqq +1$.

[1] To obtain the estimate given for $|v|$ we replace each term in the series (42) by its maximum value for $x = 0$ and sum these maximum values. The estimate is rather crude; $|v|$ probably does not exceed 0.8.

6. $\psi'(z)$ in the left half-plane. We now turn our attention to the left half-plane. Let k be a positive integer; then

$$(54)\qquad r(-k, y) = \sum_{n=1}^{k} \frac{n^2-y^2}{(n^2+y^2)^2} - \frac{1}{y^2} + \sum_{n=1}^{\infty} \frac{n^2-y^2}{(n^2+y^2)^2},$$

or

$$(55)\qquad r(-k, y) = \sum_{n=1}^{k} \frac{n^2-y^2}{(n^2+y^2)^2} + r(0, y).$$

In view of (53) we can conclude that $r(-k, y) < 0$ when $|y| \geq k$. When $|y| < k$ we have

$$r(-k, y) < \sum_{n=1}^{\infty} \frac{n^2-y^2}{(n^2+y^2)^2} + r(0, y)$$

$$= \frac{1}{y^2} + 2r(0, y) = -\frac{\pi^2}{sh^2\,\pi y} < 0.$$

Hence

$$(56)\qquad r(-k, y) < 0, \qquad k = 0, 1, 2, \ldots$$

for all values of y. Further

$$(57)\qquad j(-k, y) = -2y \sum_{n=k+1}^{\infty} \frac{n}{(n^2+y^2)^2},$$

$$(58)\qquad j\left(-k-\frac{1}{2}, y\right) = -2y \sum_{n=k+1}^{\infty} \frac{n+\frac{1}{2}}{\left[\left(n+\frac{1}{2}\right)^2+y^2\right]^2}.$$

Consequently $\psi'(z) \neq 0$ on all the lines $x = -\frac{n}{2}$ $(n = 0, 1, 2, \ldots)$ and

$$\operatorname{sgn} \Re\left[\psi'(-n+iy)\right] = -1,$$

$$(59)\qquad \operatorname{sgn} \Im\left[\psi'\left(-\frac{n}{2}+iy\right)\right] = -\operatorname{sgn} y.$$

7. Introduction of the cells. The lines $x = -\frac{n}{2}(n = 0, 1, 2, \ldots)$ and $y = 0$ divide the left half-plane into an infinite number of cells

$$C_n : -\frac{n}{2} < x < -\frac{n-1}{2}, \quad y > 0,$$

and

$$\bar{C}_n : -\frac{n}{2} < x < -\frac{n-1}{2}, \quad y < 0.$$

Theorem: Each of the cells C_{2k-1} and $\bar{C}_{2k-1}$ contains one and only one complex zero of $\psi'(z)$. The cells C_{2k} and $\bar{C}_{2k}$ do not contain any zeros $(k = 1, 2, 3, \ldots)$.

In order to prove this theorem we trace the image of the boundary of a cell C_n by the transformation $w = \psi'(z)$ avoiding the vertices of the cell at the singular points in the usual manner. For the following discussion consult Fig. 1 which gives a schematic representation of the situation. The line drawn in full corresponds to the case when n is odd and the dotted line to the case when n is even.

Let the image of the line segment $x = -\frac{n}{2}$, $0 \leqq y$ be denoted by J_n. In view of (57) and (58) the curve J_{2k} lies entirely in the third quadrant of the w-plane; it is asymptotic to the negative real axis and tangent to the v-axis at the origin. According to (59) J_{2k-1} lies in the lower half-plane; starting from a point on the positive real axis, it ends in the third quadrant at the origin and tangent to the v-axis. J_{2k-2} and J_{2k-1} intersect at least once in the third quadrant forming a loop together; it is probable that J_{2k-1} and J_{2k} do not intersect each other, also that the curves J_n do not intersect themselves, but this is immaterial for our present purpose.

The lower boundary of the cell is mapped upon a segment of the positive real axis which is in parts covered twice when n is odd. Finally a small circular arc $|z+k+1| = \varrho e^{i\theta}$, $0 \leqq \theta \leqq \frac{\pi}{2}$, is mapped upon a large contour in the

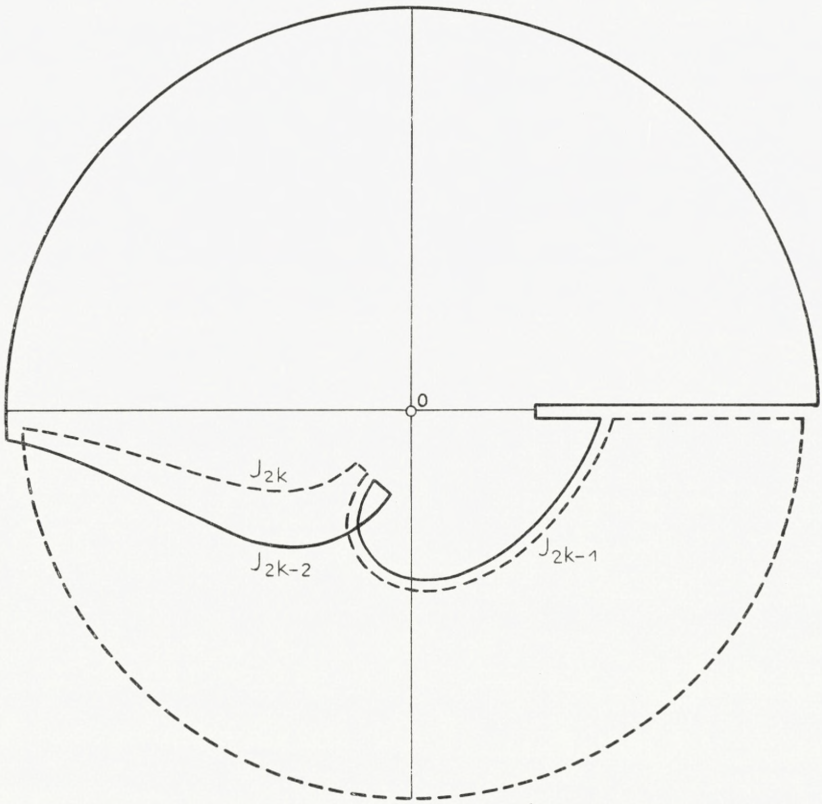


Fig. 1.

lower half-plane, and an arc $|z+k| = \varrho e^{i\theta}$, $\frac{\pi}{2} \leqq \theta \leqq \pi$, is mapped upon a contour in the upper half-plane. Keeping these facts in mind or consulting the figure the reader will see that the argument of $\psi'(z)$ remains unchanged when we trace the boundary of C_{2k} but increases by 2π along the boundary of C_{2k-1}, a result which suffices to prove our theorem.

We shall prove in § 11 that $\psi''(z) \neq 0$ in C_{2k} for all values of k. It follows that $\psi''(z)$ maps the interior of C_{2k} conformally upon a region in the lower half of the w-plane a region which, however, may partly overlap itself. The map of C_{2k-1} is neither conformal in the interior nor on the boundary. Since $\psi''(-k-1+\varepsilon) < 0$, $\psi''\left(-k-\frac{1}{2}\right) < 0$, $\psi''(-k-\varepsilon) > 0$ and $\psi'''(x) > 0$, where k is a positive integer or zero, $\varepsilon > 0$ and x is real, we conclude that $\psi''(z)$ vanishes once and only once in the interval $(-k-1, -k)$ and, in fact, on the right half of this interval. We have also noticed that the curves J_{2k-1} and J_{2k-2} intersect in the third quadrant where they form a loop. This indicates that $\psi''(z)$ vanishes at least once in the interior of C_{2k-1}. Thus we have at least 3 zeros of $\psi''(z)$ in the strip $-k-\frac{1}{2} < x < -k$ for every integral $k \geqq 0$. We shall see later that there are exactly 3 zeros of $\psi''(z)$ in this strip.

8. The curves $r = 0$ and $j = 0$. In order to gain additional information regarding the map corresponding to $w = \psi'(z)$ we consider the curves $r(x, y) = 0$ and $j(x, y) = 0$. The points $z = -n$ $(n \geqq 0)$ are double poles of $\psi'(z)$; hence they are double points of the curves $r = 0$ and $j = 0$. The r-curves have the slopes $+1$ and -1 at $z = -n$, the j-curves have the slopes 0 and ∞ at this point.

One of the j-curves through $z = -n$ is the real axis. Let the other j-curve through this point be denoted by j_n. We have already seen that $j(x, y) < 0$ in C_{2n} and > 0 in $\overline{C}_{2n}$ $(n = 1, 2, 3, \ldots)$. This follows also directly from formula (46) which shows that $j(x, y) < 0$ if $-n \leqq x \leqq -n + \frac{1}{2}$, $y > 0$. Consequently j_n lies entirely in $C_{2n-1} + \overline{C}_{2n-1}$. It is a closed curve which intersects the real axis at $z = -n + 1$ and at the point where $\psi''(x) = 0$. The curve j_n grows steadily with n in the following sense. Let us imagine that j_n be moved parallel to the real axis a distance of one unit to the left. It will then have a contact with j_{n+1} at $z = -n$; with the exception of this point, the transferred curve lies entirely within j_{n+1}. This follows from formula (44); indeed, if $z+1$ lies on j_n then $j(x+1, y) = 0$ and

$$j(x, y) = -\frac{2xy}{(x^2+y^2)^2} > 0, \qquad (y > 0),$$

i. e., the point z lies inside of j_{n+1}.

It is possible to find upper limits for $|y|$ on j_n with the aid of formulas (44), (46) and (48). If y is fixed positive

$$\frac{\pi^2}{2}\frac{\sin 2\pi x \operatorname{sh} 2\pi y}{(\sin^2 \pi x + \operatorname{sh}^2 \pi y)^2} < \pi^2 \frac{\operatorname{ch} \pi y}{\operatorname{sh}^3 \pi y}.$$

The latter expression is less than 0.075 when $y \geqq 1$. On the other hand, we can show by a simple but tedious calculation that $j(3.5, 1) < -0.075$. Further,

$$\frac{\partial j}{\partial x} = 2y \sum_{n=0}^{\infty} \frac{3(x+n)^2 - y^2}{\left[(x+n)^2 + y^2\right]^3}.$$

This expression is certainly positive when $0 \leqq y \leqq \sqrt{3}\,x$. Hence $j(x, 1)$ increases with x when $x \geqq \frac{1}{\sqrt{3}}$. Thus $j(x, 1) \leqq j(3.5, 1)$ when $1 \leqq x \leqq 3.5$. We conclude, with the aid of (46), that

$$j(x, 1) < j(3.5, 1) + \pi^2 \frac{\operatorname{ch} \pi}{\operatorname{sh}^3 \pi} < 0$$

when $-2.5 \leqq x \leqq 0$, and we can obviously draw the same conclusion for the larger interval $-3 \leqq x \leqq 0$. Hence, $|y| < 1$ on j_1, j_2 and j_3. We now use (48) with $m = 2$, viz.

$$4j(2x, 2y) = j(x, y) + j\left(x + \frac{1}{2}, y\right)$$

and set $-3 \leq x \leq 0$ and $y = 1$. It follows that $|y| < 2$ on j_4, j_5 and j_6. Repeating the argument we conclude successively that $|y| < 4$ on j_7—j_{12}, $|y| < 8$ on j_{13}—j_{24}, and so on. These limits for $|y|$ on j_n are probably not very good for large values of n; they could be improved upon, but the task is rather laborious.

We now turn our attention to the curves $r = 0$. We have already discussed in § 5 the branches of this curve in the right half-plane. Two arcs of $r = 0$ start at $z = -n$ in the interior of j_n. These arcs cannot remain inside of j_n; if they did so, they would have to intersect on the real axis forming a closed curve which, however, is impossible since $\psi'(z)$ does not have any real zeros. Hence these two arcs have to intersect j_n and obviously at $z = z_n$ and $\bar{z}_n$ where $\psi'(z) = 0$. These two arcs must pass through $z = -n-1$ since there is no other place where they can intersect the line $x = -n-1$, in view of (56), and they cannot wander off to the point at infinity.

We refer the reader to Fig. 2, which gives a schematic representation of the curves $r = 0$ (drawn in full lines) and $j = 0$ (drawn in dotted lines).

It is possible to find limits for the curves $r = 0$ which are somewhat more satisfactory than those found for the j-curves. Let r_n denote that arc of $r = 0$ on which $-n \leq x \leq -n+1$ $(n \geq 1)$ and $y > 0$. We notice first that r_n expands with n just as j_n does. If $z+1$ lies on r_n, then $r(x+1, y) = 0$ and

$$r(x, y) = \frac{x^2 - y^2}{(x^2+y^2)^2} > 0$$

provided $|y| < -x$. That this proviso is verified follows from formulas (62) below. Hence, we are justified in concluding that the point z lies inside of r_{n+1}, if z does not coincide with one of the poles.

It follows from (41) that $r(x, y) > 0$ if all the following inequalities are simultaneously fulfilled:

(60) $$(x+n)^2 - y^2 \geqq 0, \quad n = 0, 1, 2, \ldots$$

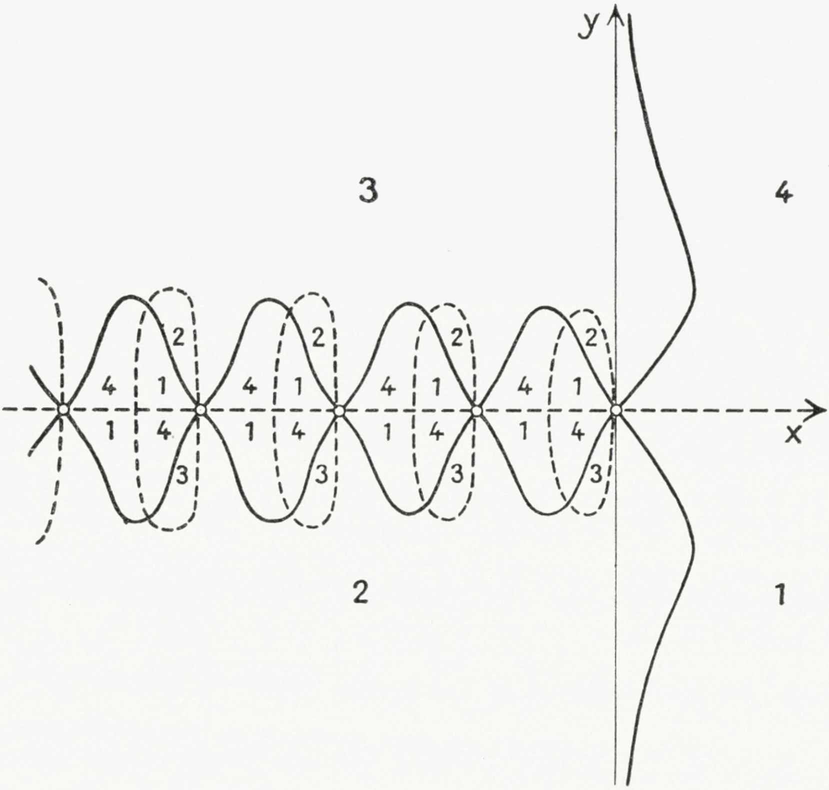


Fig. 2.

These inequalities determine a sector of opening $\frac{\pi}{2}$ in the right half-plane, and, in addition, a set of squares in the left half-plane each square having a line segment $(-n-1, -n)$ as one of its diagonals. Thus r_n lies above the polygonal line joining $z = -n$, $-n + \frac{1}{2} + \frac{i}{2}$ and $-n+1$.

A partial limitation of r_n from above can be found with the aid of (45). If $x \leqq 0$, $r(1-x, y) > 0$. Hence,

$r(x, y)$ will be negative when $x \leq 0$ and $\sin^2 \pi x \operatorname{ch}^2 \pi y - \cos^2 \pi x \operatorname{sh}^2 \pi y \leq 0$, or

(61) $r(x, y) < 0$ when $x \leq 0$ and $\tan^2 \pi x \leq \operatorname{th}^2 \pi y$.

This inequality implies that r_n lies below the corresponding arcs of the curve

$$\tan^2 \pi x = \operatorname{th}^2 \pi y \quad \text{or} \quad y = \frac{1}{2\pi} \log \tan \pi \left(\frac{1}{4} \pm x\right).$$

This curve consists of infinitely many arcs, passing in pairs through the points $z = -n$ where they have slopes equal to ± 1, and having the lines $x = -n \pm \frac{1}{4}$ as asymptotes. This method of course does not give any upper bound for r_n in the interval $-n + \frac{1}{4} \leq x \leq -n + \frac{3}{4}$.

In order to fill this gap we use the same method as above for j_n. We have

$$-\frac{\pi^2}{\operatorname{sh}^2 \pi y} \leq \pi^2 \frac{\sin^2 \pi x \operatorname{ch}^2 \pi y - \cos^2 \pi x \operatorname{sh}^2 \pi y}{[\sin^2 \pi x + \operatorname{sh}^2 \pi y]^2} \leq \frac{\pi^2}{\operatorname{ch}^2 \pi y},$$

when y is fixed. Let us set $y = +1$ and vary x on the interval $(-k, -1)$ where k is a positive integer which will be chosen below. Then

$$r(x, 1) \leq -\min r(1-x, 1) + \frac{\pi^2}{\operatorname{ch}^2 \pi}.$$

We have

$$\frac{\partial r}{\partial x} = 2 \sum_{n=0}^{\infty} (x+n) \frac{3y^2 - (x+n)^2}{\left[y^2 + (x+n)^2\right]^3} < 0$$

if $0 \leq \sqrt{3}\, y \leq x$. Thus $r(1-x, 1) \geq r(1+k, 1)$ when $-k \leq x \leq -1$. Now

$$r(1+k, 1) = \frac{1}{2} - \frac{\pi^2}{2 \operatorname{sh}^2 \pi} - \frac{3}{25} - \frac{8}{100} - \cdots\cdots - \frac{k^2-1}{(k^2+1)^2},$$

whence

$$r(x, 1) \leqq \frac{\pi^2}{2\,\mathrm{sh}^2\,\pi} + \frac{\pi^2}{\mathrm{ch}^2\,\pi} + \frac{3}{25} + \frac{8}{100} + \cdots\cdots + \frac{k^2-1}{(k^2+1)^2} - \frac{1}{2}.$$

The expression on the right hand side is negative for $k \leq 12$. Thus

(62 a) $\qquad r(x, 1) < 0 \quad \text{for} \quad -12 \leqq x \leqq 0.$

Using (47) with $m = 2$ we conclude that

(62 b) $\qquad r(x, 2) < 0 \quad \text{for} \quad -24 \leqq x < -12,$

(62 c) $\qquad r(x, 4) < 0 \quad \text{for} \quad -48 \leqq x < -24$

and so on. These estimates are probably rather crude, but they seem to justify the conclusion that the maximum ordinate on r_n grows considerably slower with n than the maximum ordinate on j_n.

The curves $r = 0$ and $j = 0$ divide the z-plane into an infinity of regions. Four of these are infinite in extent, all the others are finite. All the finite regions and the infinite ones in the right half-plane are mapped conformally and without overlapping upon a complete quadrant of the w-plane by the transformation $w = \psi'(z)$. The numbers plotted in the different regions of the figure indicate which quadrant corresponds to the region in question. The other infinite regions are mapped, the upper one upon the third and the lower one upon the second quadrant, but the map is not conformal and overlaps itself infinitely often since the regions under consideration contain all the complex zeros of $\psi''(z)$.

In order to build up the corresponding Riemann surface we can proceed as in § 17 below. To carry through the discussion properly would, however, require rather elaborate considerations so we restrict ourselves to these indications.

9. Lower limitation of the zeros of $\psi'(z)$. We shall now proceed to a further delimitation of the zeros of $\psi'(z)$. The inequalities obtained for $r(x, y)$ and $j(x, y)$ in the preceding paragraph give upper limits for y_n, the ordinate of the zero z_n of $\psi'(z)$ in the cell C_n. In particular, formulas (62) imply that

$$y_n < \left[\frac{n-1}{12}\right] + 1, \quad n = 1, 2, 3, \ldots \tag{63}$$

where as usual $[u]$ denotes the largest integer less than or equal to u. This estimate is of course rather unsatisfactory for large values of n, but shows nevertheless that y_n grows rather slowly.

A lower limit for y_n can be obtained with the aid of formulas (38) and (49). It follows from (49) that, when $\Re(z) > 1$,

$$|z\,\psi'(z)| \leqq \sum_{n=0}^{\infty} \frac{n!}{(n+1)\,|z+1| \ldots |z+n|} < \sum_{n=0}^{\infty} \frac{1}{(n+1)^2} = \frac{\pi^2}{6},$$

or

$$|(1-z)\,\psi'(1-z)| < \frac{\pi^2}{6} \quad \text{when } \Re(z) < 0. \tag{64}$$

Similarly

$$\left|\psi'(1-z) - \frac{1}{1-z}\right| < \frac{\frac{\pi^2}{3} - 2}{|1-z|\,|2-z|}. \tag{65}$$

In virtue of (38) we have that $\psi'(z) \neq 0$ if

$$\left|\frac{\pi^2}{\sin^2 \pi z}\right| > |\psi'(1-z)|,$$

and, using (64), we see that this is a fortiori the case when

$$\left|\frac{\pi^2}{\sin^2 \pi z}\right| \geqq \frac{\pi^2}{6\,|1-z|},$$

or

(66) $$|\sin^2 \pi z| \leq 6\,|1-z|.$$

Thus the two branches of the curve C

(67) $$\sin^2 \pi x + \mathrm{sh}^2 \pi y = 6 \sqrt{(x-1)^2 + y^2}$$

for $x \leq 0$, together with a connecting segment on the imaginary axis, bound a simply-connected region R such that $\psi'(z) \neq 0$ on $R + C$. A fairly simple reckoning shows that $\frac{dy}{dx} < 0$ on the upper half of C, i. e., y decreases when x increases.

We can now obtain a lower limit for y_n as follows. Evidently y_n exceeds the ordinate of the point on C whose abscissa is $-n+1$; this ordinate is determined as the real positive root of the equation

$$\mathrm{sh}\, \pi y = \sqrt{6}\, \sqrt[4]{n^2 + y^2}.$$

This equation implies that

$$\mathrm{sh}\, \pi y > \sqrt{6n},$$

or

$$y > \frac{1}{\pi} \log \left[\sqrt{6n} + \sqrt{6n+1}\right] > \frac{1}{\pi} \log 2\sqrt{6n}.$$

Hence

(68) $$y_n > \frac{1}{\pi} \log 2\sqrt{6n}.$$

In particular, $y_1 > 0.5$. For small values of n, formulas (63) and (68) give comparatively narrow limits for y_n.

10. The asymptotic distribution of the zeros of $\psi'(z)$. We shall now take up the asymptotic distribution of the zeros. We introduce the function

(69) $$\Phi(z) = \frac{\pi^2}{\sin^2 \pi z} - \frac{1}{1-z},$$

and proceed to prove the following

Theorem: $\Phi(z)$ has exactly one zero in each of the cells C_{2n-1} and $\overline{C}_{2n-1}$ and no zeros in the cells C_{2n} and $\overline{C}_{2n}$. If we denote the zero in C_{2n-1} by ζ_n and set $\zeta_n = \xi_n + i\eta_n$, then $-n+\frac{1}{2} < \zeta_n < -n + \frac{3}{4}$, and

$$(70)\left\{\begin{aligned} \zeta_n = -n + \frac{1}{2} + \frac{\log 2\pi\sqrt{n+\frac{1}{2}}}{2\pi^2\left(n+\frac{1}{2}\right)} + \\ + \frac{i}{\pi}\left[\log 2\pi\sqrt{n+\frac{1}{2}} - \frac{1}{4\pi^2\left(n+\frac{1}{2}\right)}\right] + O\left(\frac{\log^2 n}{n^2}\right). \end{aligned}\right.$$

Let ζ_n be the center of a circle Γ_n of radius $\frac{3}{n+1}$. Then each circle Γ_n with $n \geqq 11$ contains one and only one zero of $\psi'(z)$.

We postpone the proof of formula (70) until the rest of the theorem has been proved. We readily verify that

$$\pi < \arg \Phi(-n+iy) < \frac{3\pi}{2},$$

$$\Im\, \Phi\left(-n-\frac{1}{2}+iy\right) < 0,$$

when $y > 0$ and $n = 0, 1, 2, \ldots$ Further $\Phi(x) > 0$ for x real and negative. These relations are exactly the same as those satisfied by $\psi'(z)$ on the lines in question; they permit us to repeat the proof given in § 7 with $\psi'(z)$ replaced by $\Phi(z)$; this suffices to prove the statement about the cells.

To verify that $-n+\frac{1}{2} < \xi_n < -n+\frac{3}{4}$ we notice that

$$\Re\left[\sin^2 \pi(x+iy)\right] = \frac{1}{2}(1-\cos 2\pi x \operatorname{ch} 2\pi y) \leqq \frac{1}{2}$$

if $\cos 2\pi x \geqq 0$, i. e., if $-n+\frac{3}{4} \leqq x \leqq -n+\frac{5}{4}$. On the other hand,

$$\Re\left[\pi^2(1-z)\right] = \pi^2(1-x) \geqq \pi^2$$

when $x \leqq 0$. Hence ξ_n must be limited in the way just mentioned.

Let $\zeta = \xi + i\eta$ $(\xi < 0,\ \eta > 0)$ be an arbitrary zero of $\Phi(z)$. We shall study $\Phi(\zeta+\omega) = \underset{\omega}{\Delta}\Phi(\zeta)$ when $|\omega| = r$, a fixed number. We have

$$\Phi(\zeta+\omega) = \pi^2 \frac{\sin^2 \pi\zeta - \sin^2 \pi(\zeta+\omega)}{\sin^2 \pi\zeta \sin^2 \pi(\zeta+\omega)} + \frac{\omega}{(1-\zeta)(1-\zeta-\omega)}$$

$$= -\frac{\omega}{1-\zeta}\left[\frac{\sin \pi\omega}{\omega}\,\frac{\sin \pi(2\zeta+\omega)}{\sin^2 \pi(\zeta+\omega)} - \frac{1}{1-\zeta-\omega}\right].$$

We now assume $r \leqq \frac{1}{4}$. Then

$$|\Phi(\zeta+\omega)| \geqq \left|\frac{\omega}{1-\zeta}\right| \left[2\sqrt{2}\,\frac{\operatorname{sh} \pi(2\eta-r)}{\operatorname{ch}^2 \pi(\eta-r)} - \frac{1}{|1-\zeta|-r}\right].$$

The fraction involving the hyperbolic functions increases steadily with η when r is fixed, and decreases when r increases if η is fixed $> \frac{r}{2}$. Thus the fraction will be made as small as possible if we give η its least value and take $r = \frac{1}{4}$.

In order to obtain a suitable lower limit for the bracket we set $\zeta = \zeta_n$ with $n \geqq 11$. This implies $|1-\zeta|-r > 11$, and, since $\eta_n > \frac{1}{\pi}\log 2\pi\sqrt{n}$,

$$\eta > \eta_{11} > \frac{1}{\pi}\log 2\pi\sqrt{11} > 0.96.^{1}$$

[1] In order to obtain this estimate, use the same type of argument as in the proof of formula (68).

With these restrictions upon ζ and ω we find that

$$(71) \qquad |\Phi(\zeta+\omega)| \geqq \frac{K|\omega|}{|1-\zeta|} \quad \text{where} \quad K > 0.445.$$ [1]

In view of formulas (38) and (65) we have

$$(72)\ \psi'(z) = \Phi(z) + P(z) \ \text{where} \ |P(z)| < \frac{\frac{\pi^2}{3}-2}{|(1-z)(2-z)|},$$

when $\Re(z) < 0$. With each of the points ζ_n, $n \geqq 11$, as center we lay a circle Γ_n of radius r_n. We shall determine r_n in such a manner that

$$(73) \qquad |\Phi(z)| > |P(z)| \quad \text{on} \quad \Gamma_n,$$

and impose in advance the condition $r_n \leqq \frac{1}{4}$. Setting $z = \zeta_n + \omega_n$ $(|\omega_n| = r_n)$ on Γ_n we have from (65) and (71)

$$|\Phi(\zeta_n+\omega_n)| \geqq \frac{K r_n}{|1-\zeta_n|},$$

$$|P(\zeta_n+\omega_n)| < \frac{\frac{\pi^2}{3}-2}{|(1-\zeta_n-\omega_n)(2-\zeta_n-\omega_n)|}.$$

Thus (73) will be fulfilled if

$$r_n \leqq \frac{\frac{\pi^2}{3}-2}{K} \, \frac{|1-\zeta_n|}{|1-\zeta_n|-r_n} \, \frac{1}{|2-\zeta_n|-r_n}.$$

We now use the assumptions $n \geqq 11$, $r_n \leqq \frac{1}{4}$ together with the fact that $K > 0.445$. These premises imply that $|1-\zeta_n| > 11.25$, $|2-\zeta_n| - r_n > n+1$, and

[1] By imposing more severe restrictions upon ζ and ω we can get K as near to 2π as we please.

$$\frac{\frac{\pi^2}{3}-2}{K}\;\frac{|1-\zeta_n|}{|1-\zeta_n|-r_n}<2.97<3.$$

Consequently (73) will hold for $n \geqq 11$ when $r_n = \dfrac{3}{n+1}$. But then it follows from the theorem of Rouché that each circle Γ_n contains one and only one zero of $\psi'(z)$.

It remains to prove formula (70). We begin by determining a set of numbers $\zeta_{m,n}$ satisfying the following conditions

$$(74)\quad \begin{cases} \sin \pi \zeta_{m+1,n} = \pi \sqrt{1-\zeta_{m,n}}, \quad \zeta_{1,n} = -n+\dfrac{1}{2}, \\ -n < \Re(\zeta_{m,n}) < -n+1, \quad \Im(\zeta_{m,n}) > 0. \end{cases}$$

Here $m, n = 1, 2, 3, \ldots$ and $\sqrt{1-z}$ means that determination of the square root which equals to $+1$ when $z=0$. We have

$$\zeta_{2,n} = -n+\frac{1}{2}+\frac{i}{\pi}\log\left[\pi\sqrt{n+\frac{1}{2}}+\pi\sqrt{n+\frac{1}{2}-\frac{1}{\pi^2}}\right],$$

$$(75)\quad \begin{cases} \zeta_{3,n} = -n+\dfrac{1}{2}+\dfrac{\log 2\pi\sqrt{n+\frac{1}{2}}}{2\pi^2\left(n+\frac{1}{2}\right)}+ \\ +\dfrac{i}{\pi}\left[\log 2\pi\sqrt{n+\dfrac{1}{2}}-\dfrac{1}{4\pi^2\left(n+\frac{1}{2}\right)}\right]+O\left(\dfrac{\log^2 n}{n^2}\right), \end{cases}$$

. . . .

We easily verify that

$$\sin\frac{\pi}{2}(\zeta_n-\zeta_{m,n}) = -\frac{\pi\,[\zeta_n-\zeta_{m-1,n}]}{2\cos\frac{\pi}{2}(\zeta_n+\zeta_{m,n})\left[\sqrt{1-\zeta_n}+\sqrt{1-\zeta_{m-1,n}}\right]}.$$

For $m=2$ we have

$$|\zeta_n - \zeta_{1,n}| < c_1 \log n, \quad \left|\cos\frac{\pi}{2}(\zeta_n - \zeta_{2,n})\right| > c_2\sqrt{n},$$

$$\left|\sqrt{1-\zeta_n} + \sqrt{1-\zeta_{1,n}}\right| > c_3\sqrt{n},$$

where the c's are positive constants independent of n. Hence

$$\left|\sin\frac{\pi}{2}(\zeta_n - \zeta_{2,n})\right| < \frac{c_1}{2\,c_2\,c_3} \cdot \frac{\log n}{n}$$

and

$$|\zeta_n - \zeta_{2,n}| < C_1 \frac{\log n}{n}.$$

Repeating the argument with $m = 3$ we see that

$$|\zeta_n - \zeta_{3,n}| < C_2 \frac{\log n}{n^2}. \tag{76}$$

Combining formulas (75) and (76) we get formula (70).

11. The zeros of $\psi''(z)$. In §§ 5 and 7 we made certain statements regarding the zeros of $\psi''(z)$. We shall now prove the following

Theorem: $\psi''(z)$ has exactly three zeros in each of the strips $-n+\frac{1}{2} \leq x \leq -n+1$ $(n = 1, 2, 3, \ldots)$ of which one and only one is real. There are no other zeros.

We begin by proving that

$$\operatorname{sgn} \Im\left[\psi''(-n+iy)\right] = -\operatorname{sgn} y \tag{77}$$

for $n = 0, 1, 2, \ldots$ We have from (53) and (55)

$$\Im\left[\psi''(-n+iy)\right] = -\frac{\partial}{\partial y} r(-n, y)$$

$$= 2y \sum_{m=1}^{n} \frac{3m^2 - y^2}{(m^2+y^2)^3} - \frac{1}{y^3} - \pi^3 \frac{\operatorname{ch} \pi y}{\operatorname{sh}^3 \pi y},$$

where the finite sum is to be suppressed when $n = 0$. This expression clearly has opposite sign to that of y when $|y| \geq n\sqrt{3}$. If $|y| < n\sqrt{3}$ we have

$$\frac{1}{y}\Im\left[\psi''(-n+iy)\right] < 2\sum_{m=1}^{\infty}\frac{3m^2-y^2}{(m^2+y^2)^3}-\frac{1}{y}\left[\frac{1}{y^3}+\pi^3\frac{\operatorname{ch}\pi y}{\operatorname{sh}^3\pi y}\right]$$

$$= -\frac{2\pi^3}{y}\frac{\operatorname{ch}\pi y}{\operatorname{sh}^3\pi y} < 0.$$

This completes the proof.

We shall now prove that the variation of the argument of $\psi''(z)$ is zero when z describes the perimeter of a large square with vertices at the points $n(\pm 1 \pm i)$ avoiding the point $z = -n$ by a small semi-circle to the right of this point. This contour contains n triple poles of $\psi''(z)$, further at least $3n$ zeros, namely, at least three in each of the strips $-m+\frac{1}{2} < x < -m+1$, $m = 1, 2, 3, \ldots, n$, as we have seen in § 7. If we can prove the statement about the variation of the argument then the theorem follows immediately.

In the neighborhood of $z = \infty$ in the sector $|\arg z| \leq \pi - \varepsilon$ we have

$$\psi''(z) = -\frac{2}{z^2} + O\left(\frac{1}{z^3}\right).$$

Now let us start with z at $+n$ and describe the contour in the positive sense. Then $w = \psi''(z)$ starts with a small negative value, and its argument decreases from π to approximately $-\frac{\pi}{2}$ when z goes from $+n$ to $n(-1+i)$. When z goes from $n(-1+i)$ to $-n+\varepsilon i$, w remains in the lower half-plane in view of (77), and when $z = -n+i\varepsilon$ $|w|$ is large and arg w is nearly $-\frac{\pi}{2}$ since

$$\psi''(z) = -\frac{2}{(z+n)^3} + \mathfrak{P}(z+n).$$

When z describes the circular arc $|z+n| = \varepsilon$, $0 \leq \arg (z+n) \leq \frac{\pi}{2}$, $|w|$ remains large and $\arg w$ increases from $-\frac{\pi}{2}$ to $+\pi$. Consequently $\arg w$ is back to its initial value after we have described the upper half of the contour, and by reasons of symmetry, $\arg w$ will return to the initial value after we have described the lower half of the contour. This completes the proof of the theorem.

Part III.

The conformal correspondence $w = \psi(z)$.

12. The R, I**-net.** We shall now return to the psi-function itself, and consider the question of how its values are distributed in the plane. We shall attack this problem from two different angles. First, we have obtained in Parts I and II of the present paper a variety of results which permit us to give a rather detailed discussion of the curves

$$\Re\left[\psi(z)\right] = \text{const., and } \Im\left[\psi(z)\right] = \text{const.}$$

We shall give this discussion in §§ 12—15. Secondly, we try to complete the information so obtained by numerical computation of the psi-function for some values of z. Finally, in § 17 we discuss the Riemann surface corresponding to $w = \psi(z)$ in the light of the results obtained in §§ 12—16.

For the whole discussion the reader should consult Fig. 3, which gives a representation of the curves in question. In the upper half-plane the curves

$R(x, y) = c$, with $c = -2, -1.5, -1, \ldots, +1.5$ and $+2$, are traced; in the lower half-plane we have marked the curves

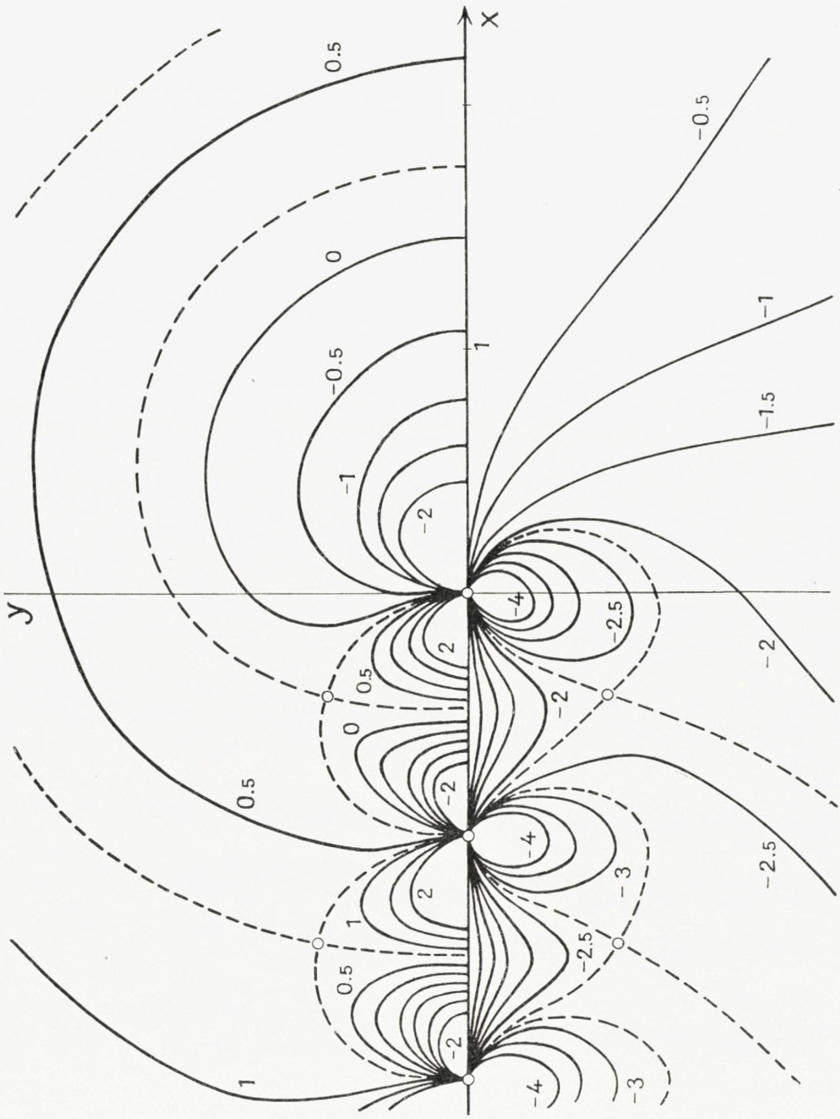


Fig. 3.

$I(x, y) = \gamma$ with $\gamma = -4, -3.5, -3, \ldots, -0.5$ and 0. In addition we have plotted in dotted lines the curves of the two systems which pass through the four zeros of $\psi'(z)$

which are nearest to the origin. The diagram is based upon the results of §§ 12—16, and is believed to give a fairly accurate picture of the situation, but, naturally, it must not be trusted too far.

In the sector $|\arg z| \leq \pi - \varepsilon$ we have

$$\psi(z) = \log z + O\left(\frac{1}{z}\right).$$

It follows that within this sector and sufficiently far from the origin, the curves $R = c$ correspond to large positive values of c and each curve lies between two circles $|z| = e^c - \delta$ and $|z| = e^c + \delta$. The curves $I = \gamma$ on the other hand, correspond to values of γ between $-\pi + \varepsilon$ and $\pi - \varepsilon$ and are asymptotic to the lines $\arg z = \gamma$.

In the remaining sector we have

$$\psi(z) = \log(1 - z) - \pi \cot \pi z + O\left(\frac{1}{z}\right).$$

Here we have evidently quite a complicated situation; the net corresponding to $\log(1 - z)$ is distorted by the superimposed net due to $-\pi \cot \pi z$.

The points $z = -n$ $(n \geq 0)$ are simple poles of residue -1 for $\psi(z)$. Let N_ε be a small neighborhood of $z = -n$. Any R-curve in N_ε will pass through $z = -n$ where it will have a vertical tangent. If c is sufficiently large positive (negative) the curve $R = c$ will be closed in N_ε and located to the left (right) of the vertical tangent; further it will be almost circular in shape. Any I-curve in N_ε will pass through $z = -n$ and be tangent to the x-axis. If γ is sufficiently large numerically, the curve $I = \gamma$ will be closed in N_ε and almost circular; it will be above or below the x-axis according as γ is positive or negative. The curves of the two nets have a perfectly definite order in N_ε. Thus, for example, if we describe the upper half of the curve $R = -M$ (M large positive) in N_ε starting from $z = -n + \delta$ (δ real positive) and ending at $z = -n$, then $v = I(x, y)$ will be steadily growing along the curve from the initial value 0 to the final limit $+\infty$, every intermediate value being taken on once and only once. Similarly with the R-curves.

Any curve $R = c$ will consist of an infinity of separate branches, beginning and ending at $z = -n$, one branch for each pole. Any curve $I = \gamma$ will consist of an infinity of branches, which, however, may and as a rule do have end-points in common. Such a branch will join a pole either with itself or with another pole or with the point at infinity.

Through the points z_n where $\psi'(z) = 0$ will pass two and only two branches of each system. If we set

$$\psi(z_n) = w_n = u_n + i\,v_n, \tag{78}$$

it is two branches of the curve $R = u_n$ and two branches of $I = v_n$ which pass through z_n. These curves are of fundamental importance for the whole discussion and will be considered at length in §§ 14 and 15. No other curve of either system can intersect itself or have a non-singular point in common with any other curve belonging to the same system.

We have discussed the curve $R = 0$ in some detail in § 3. This curve was found to consist of infinitely many separate ovals R_n, one for each pole $z = -n$, $n \geqq 0$, all being outside each other in accordance with the inequalities (30) and (31). Indeed, these inequalities prove the existence in every strip $-n-1 < x < -n$ ($n \geqq 0$) of a substrip where $R(x, y) > 0$. Further, formulas (25) and (35) prove that the oval R_n contracts indefinitely to zero when $n \to +\infty$, and we have shown at the end of § 3 that this contraction process is monotone in a perfectly definite sense.

Let us now turn to a curve $R = c < 0$. This curve clearly consists of separate ovals $R_n(c)$, namely, one and only one oval inside each oval R_n $(n = 0, 1, 2, \ldots)$. Thus the ovals $R_n(c)$ are outside of each other when $c \leq 0$. They will contract indefinitely when $n \to +\infty$ and the process can be shown to be steady or monotone in the sense above mentioned. The same conclusions will hold for sufficiently small positive values of c, but will cease to hold when c is large. Let $R_n(c)$ still denote that branch of $R = c$ which goes through $z = -n$. If c is large we can no longer affirm that the $R_n(c)$ are all outside of each other, but they will have this property for sufficiently large values of n, i. e., if we disregard a finite number of the branches the remaining ones will be outside of each other and of the disregarded branches[1]. Our previous conclusions are valid for the residual infinite set.

13. Differential properties of the net. Now we turn our attention to questions of increase and direction. We have

$$r(x, y) = \frac{\partial}{\partial x} R(x, y) = \frac{\partial}{\partial y} I(x, y),$$

$$j(x, y) = -\frac{\partial}{\partial y} R(x, y) = \frac{\partial}{\partial x} I(x, y).$$

In a region where $r(x, y) > 0$, $R(x, y)$ increases with x and $I(x, y)$ increases with y. In a region where $j(x, y) > 0$,

[1] In order that the $R_n(c)$ be outside of each other for $n \geq m$ it is necessary and sufficient that $c \leq u_m$. This follows from the results stated at the end of § 14.

$R(x, y)$ decreases when y increases and $I(x, y)$ increases with y. At a point $z = z_0 = x_0 + iy_0$ where $r(x, y)$ and $j(x, y)$ have the same (opposite) sign, the slope of $R(x, y) = c$ is positive (negative) and the slope of $I(x, y) = \gamma$ is the negative reciprocal of the slope of the R-curve. Let us define as *positive direction of the tangent* of $R(x, y) = c$ at z_0 *that direction in which* $I(x, y)$ *increases*, with a similar definition for the I-curve. This direction is uniquely defined unless z_0 happens to be a zero or a pole of $\psi'(z)$. Let $\varphi_1(z_0)$ be the angle which the positive direction of the tangent of $\Re[\psi(z)] = \Re[\psi(z_0)]$ at $z = z_0$ makes with the positive direction of the real axis, the angle being measured from the axis to the tangent; and let $\varphi_2(z_0)$ be the corresponding angle for the I-curve $\Im[\psi(z)] = \Im[\psi(z_0)]$. Then we have

$$\text{(79)} \qquad \varphi_1(z_0) \equiv \frac{\pi}{2} - \arg \psi'(z_0) \qquad (\text{mod. } 2\pi),$$

$$\text{(80)} \qquad \varphi_2(z_0) \equiv \quad - \arg \psi'(z_0) \qquad (\text{mod. } 2\pi).$$

Fig. 2 suffices to give us a general notion of the mode of variation of $\arg \psi'(z)$. This figure, it will be remembered, is based upon the discussion of $\psi'(z)$ in §§ 5—8. It will perhaps be useful to collect at this point some of the consequences of this discussion.

It has been noticed that $j(x, y)$ is negative in all the cells C_{2n} and in the first quadrant and positive in the symmetric regions below the x-axis. This implies that $R(x, y)$ grows with $|y|$ in these regions when x is kept fixed. In particular, this will be the case on the boundary of any one of the cells, hence[1]

[1] In order to obtain the lower limits in (81) and (82) use formula (29).

$$R(-n, y) > \psi(n+1), \tag{81}$$

$$R\left(-n-\frac{1}{2}, y\right) \geq \psi\left(n+\frac{3}{2}\right). \qquad (n = 0, 1, 2, \ldots) \tag{82}$$

In that part of C_{2n} which lies above r_n, we have $\pi < \varphi_1(z_0) < \frac{3\pi}{2}$ and $\frac{\pi}{2} < \varphi_2(z_0) < \pi$; below r_n we have $\frac{\pi}{2} < \varphi_1(z_0) < \pi$ and $0 < \varphi_2(z_0) < \frac{\pi}{2}$.

The R-curves have vertical tangents on j_n, horizontal ones on r_n and $\bar{r}_n$[1]. For the I-curves the situation is of course reversed. Finally, we notice that any vertical line which does not intersect any of the curves j_n, will either intersect an arbitrary curve $R = c$ in two points symmetric to the x-axis or not at all.

14. Qualitative description of the net. We shall now take up the properties of the net in the gross. We aim at a qualitative description of the net which will tell us how the separate branches of the different curves go, what singular points they join, how they separate the plane into regions, and so on. We shall see that the solution of this problem depends essentially upon a special case of the same problem, namely how the critical curves through the zeros of $\psi'(z)$ behave in this or that respect.

We begin by considering the I-curves. Let us inspect the branches of the I-curves which radiate from $z = -n$ $(n \geq 1)$. One of these curves is the real axis. Now give γ a small positive value. We conclude by reasons of continuity that there is a branch of the curve $I = \gamma$ which joins $z = -n$ with $z = -n+1$ and which lies entirely within a rectangle $-n \leq x \leq -n+1$, $0 \leq y \leq \delta(\gamma)$ where

[1] We denote the arc of $r = 0$ which lies in the first quadrant by r_0 and let $\bar{r}_n$ mean the curve symmetric to r_n in the lower half-plane $(n = 0, 1, 2, \ldots)$.

$\delta(\gamma) \to 0$ with γ. There is also a branch of the same curve which joins $z = -n$ with $z = -n-1$, but we disregard this arc for the present. Let γ_n be the largest value of γ such that for $\gamma \leqq \gamma_n$ the curve $I = \gamma$ has a branch $I_n(\gamma)$ joining $z = -n$ with $z = -n+1$ without passing through any other singular point. I claim that $I_n(\gamma_n)$ goes through a zero of $\psi'(z)$; to be more specific, I assert that $I_n(\gamma_n)$ goes through $z = z_n$, i. e. $\gamma_n = v_n$. Suppose this were not so and consider that arc of the curve $I = \gamma_n + \delta$ (δ small positive) which starts at $z = -n$ and on which $x+n$ is small positive when y is small positive. Since $\psi'(z) \neq 0$ on $I_n(\gamma_n)$, this arc will be uniformly near to $I_n(\gamma_n)$, i. e., we can find an $\varepsilon = \varepsilon(\delta)$ which tends to zero with δ, such that the distance between the two curves nowhere exceeds ε[1]. But then this branch of $I = \gamma_n + \delta$ must end at $z = -n+1$, which is contrary to the definition of γ_n.

Thus $I_n(\gamma_n)$ goes through a zero of $\psi'(z)$. Suppose that this zero were not z_n. Then $I_n(\gamma_n)$ which joins $z = -n$ with $z = -n+1$, must intersect either the line $x = -n$ or the line $x = -n+1$ in two distinct points with positive ordinates. This, however, is impossible since $I(-m, y)$ is steadily decreasing when y increases, m being a positive integer or 0, in accordance with (56). Hence $I_n(\gamma_n)$ passes

[1] That this is actually the case follows from the following consideration. Leaving out two small arcs at the end-points of $I_n(\gamma_n)$ we can cover the residual arc by a finite number of circles such that: (i) every point on the arc is interior to at least one of the circles, and (ii) the interior of any one of the circles is mapped conformally and without overlapping upon a region in the w-plane by the transformation $w = \psi(z)$. The image of the set of points which belong to at least one of these circles is simply-connected and contains a segment of the line $v = \gamma_n$, hence also a segment of $v = \gamma_n + \delta$ if δ is sufficiently small. This proves the assertion except near the end-points of $I_n(\gamma_n)$. But these do not cause any difficulties since the curves under consideration are tangent to each other at these points. This completes the proof.

through z_n and does not go through any other zero of $\psi'(z)$. Further, $\gamma_n = v_n$. Incidentally we notice that $I_n(v_n)$ does not intersect the lines $x = -n$ or $x = -n+1$ except at the end-points.

Now let $z+1$ trace the arc $I_n(v_n)$ from $-n$ to $-n+1$. Then z traces an arc I_n^* from $-n-1$ to $-n$. On this latter arc $I(x, y) > v_n$ except at the end-points where equality holds, as we see from formula (10). Let D_n^* be the region bounded by I_n^* and the real axis between $-n-1$ and $-n$. The point z_{n+1} may be located (i) within D_n^*, or (ii) on I_n^*, or (iii) outside of D_n^*. Whichever be the actual case, we shall prove that $v_n < v_{n+1}$.

Suppose case (i) be realized, and consider the four arcs of the curve $I = v_{n+1}$ which start at $z = z_{n+1}$. We know that two of these arcs form the branch $I_{n+1}(v_{n+1})$ with end-points at $z = -n-1$ and $-n$. The other two arcs cannot lie completely within D_n^*. If they did, we should have two distinct arcs $I = v_{n+1}$ the ends of which would belong to a small sector $|z+n| < \delta$, $0 < \arg(z+n) < \varepsilon$; this is clearly impossible in view of the order relations between the I-curves in the neighborhood of a pole. Hence these two arcs must intersect I_n^* at a point where $I(x, y) > v_n$, and thus $v_{n+1} > v_n$. In cases (ii) and (iii) we see almost directly that the same conclusion is valid.

Let us now study the I-curves which emanate from $z = 0$ and of which the initial arcs belong to the first quadrant. Such a branch of the curve $I = \gamma$ will be designated by $I_0(\gamma)$. As long as $0 \leq \gamma \leq \frac{\pi}{2}$, $I_0(\gamma)$ will remain in the first quadrant and go from $z = 0$ to $z = \infty$, having the line $\arg z = \gamma$ as its asymptote. That $I_0(\gamma)$ cannot intersect the positive imaginary axis follows from formulas (16) and (56), which imply that

(83) $\quad I(-n, y) > \frac{\pi}{2}$ when $y > 0, \; n = 0, 1, 2, \ldots$

When $\gamma > \frac{\pi}{2}$ the arc $I_0(\gamma)$ intersects the imaginary axis and proceeds to the point at infinity as long as $\gamma - \frac{\pi}{2}$ is sufficiently small. There exists a largest value, Γ_0 say, such that $I_0(\gamma)$ ends at infinity for every $\gamma \leqq \Gamma_0$. Just as above we prove that the curve $I_0(\Gamma_0)$ must pass through a zero of $\psi'(z)$, and this zero must be z_1. Suppose contrariwise that it would be z_2 instead. Then $\Gamma_0 = v_2$ and there exists an arc of an I-curve joining $z = 0$ with $z = -1$ on which $I = v_2$. This arc together with the segment of the real axis from 0 to -1 bounds a region D_1 which evidently contains the point z_1 in its interior. Moreover, the four arcs of the curve $I = v_1$ which meet at $z = z_1$ must be enclosed in D_1. But this is impossible since $v_2 > v_1$; indeed, if $\Gamma_0 = v_2 > v_1$ then $I_0(v_1)$ goes from $z = 0$ to $z = \infty$ entirely outside of D_1 in view of the definition of Γ_0. But there are only two arcs of $I = v_1$ which begin or end at $z = 0$ and only two such arcs which begin or end at $z = -1$; if one of the former arcs is outside of D_1, then there is at least one of the four arcs of $I = v_1$ starting at z_1 which does not end in the interior of D_1. We are thus led to a contradiction by assuming that $\Gamma_0 = v_2$; in exactly the same manner we disprove the assumption that $\Gamma_0 = v_n$, $n \neq 1$.

Hence $\Gamma_0 = v_1$. We can now account for all the I-curves which emanate from $z = 0$. As long as $\gamma < v_1$, $I_0(\gamma)$ goes from $z = 0$ to $z = \infty$; when $\gamma > v_1$, $I_0(\gamma)$ is a closed curve beginning and ending at the origin. The two types of curves are separated by arcs of $I = v_1$, namely, $I_0(v_1)$, which goes from $z = 0$ over z_1 to ∞, plus $I_1(v_1)$ which goes from $z = -1$ over z_1 to 0.

We now pass to the second pole at $z = -1$. We designate by $I_1(\gamma)$ that arc of the curve $I = \gamma$ which starts at $z = -1$ and on which $\arg(z+1)$ is small positive when $|z+1|$ is small. As long as $0 \leq \gamma \leq v_1$, $I_1(\gamma)$ goes from -1 to 0. Thus if we return for a moment to $z = 0$ we see that the I-curves in a small neighborhood of the origin, $|z| < \delta$, $y \geq 0$, are either of the type $I_0(\gamma)$, $0 \leq \gamma$, or of the type $I_1(\gamma)$, $0 \leq \gamma \leq v_1$. The former curves begin at $z = 0$, the latter ones end at this point according to our present convention, which is in agreement with our previous way of orienting the curves with the aid of the positive direction of the tangent.

When $0 < \gamma - v_1 < \varepsilon$, $I_1(\gamma)$ joins $z = -1$ with $z = \infty$. There exists a largest value, Γ_1 say, such that $I_1(\gamma)$ has this property for every γ, $v_1 < \gamma \leq \Gamma_1$. As above we show that $\Gamma_1 = v_2$. Thus the branches $I_1(\gamma)$ join $z = -1$ with $z = \infty$ when $v_1 < \gamma \leq v_2$, and when $\gamma > v_2$ they are closed curves beginning and ending at $z = -1$. The remaining I-curves which belong to the upper half of an ε-neighborhood of $z = -1$ are curves of the type $I_2(\gamma)$ with $0 \leq \gamma \leq v_2$ which start at $z = -2$ and end at $z = -1$. In this manner we can proceed step by step. The situation in the lower half-plane is symmetric to the situation just described.

We notice that

$$(84) \qquad \frac{\pi}{2} < \Im\left[\psi(z_n)\right] < \Im\left[\psi(z_{n+1})\right] < \pi.$$

Here the lower limit $\frac{\pi}{2}$ could be raised somewhat; v_1 is certainly greater than 2, — on the other hand

(85) $v_1 < \operatorname{Max} I\left(-\frac{1}{2}, y\right) = \operatorname{Max}\left[\frac{\pi}{2} \operatorname{th} \pi y + \frac{4y}{1+4y^2}\right] < \frac{\pi}{2} + 1,$

and in general

$$(86)\begin{cases} v_n < \operatorname{Max} I\left(-n+\frac{1}{2}, y\right) \\ = \operatorname{Max}\left[\frac{\pi}{2} \operatorname{th} \pi y + 4y \sum_{m=1}^{n} \frac{1}{(2m-1)^2 + y^2}\right]. \end{cases}$$

These limits are unfortunately not well suited for numerical estimates. That π is the true upper limit in (84) follows from the following consideration. Let ε be arbitrarily small positive and let $\pi - \varepsilon < v < \pi$. There is a unique I-curve which admits of the line $\arg z = v$ as its asymptote, this curve is a branch of $I = v$. We know that any such branch when traced in the negative sense will ultimately lead us to a pole. Suppose that our branch leads to $z = -m$. Then we are dealing with $I_m(v)$ according to the nomenclature adopted above. But if $I_m(\gamma)$ joins $z = -m$ with $z = \infty$ then $v_m < \gamma \leq v_{m+1}$. Hence the same inequality has to be satisfied by v, i. e. $v_{m+1} > \pi - \varepsilon$.

We now proceed to discuss the fate of the R-curves which emanate from the different poles, and start with $z = 0$. The corresponding arcs of the R-system have been designated by $R_0(c)$ in § 12. As long as $c < -C$, $R_0(c)$ remains in the right half-plane and intersects the positive real axis between 0 and $+1$. There exists a largest value of c, c_1 say, such that all the ovals $R_0(c)$ intersect the positive real axis as long as $c \leq c_1$. As above, we prove that $R_0(c_1)$ passes through a zero of $\psi'(z)$, and owing to symmetry it will have to pass through two conjugate imaginary zeros. The zero in the upper half-plane must be z_1, i. e. $c_1 = u_1$. Indeed, if $R_0(c_1)$ passed through any other zero but z_1, it would have to intersect the line $x = -\frac{1}{2}$ twice; this is impossible since $R(x, y)$ increases steadily with $|y|$ along this line.

Let the point of intersection of $R_0(u_1)$ with the positive real axis be denoted by P_1. We can find a point $z = p_1$ on the interval $(-1, 0)$ where $\psi(z) = u_1$. Through the latter point passes a branch of $R = u_1$. There is also a branch of the same curve which goes through $z = -1$. These branches must pass through $z = z_1$. In order to see that this is really true, we notice that there are four arcs of the curve $R = u_1$ which meet at $z = z_1$. Two of these have already been accounted for; one joins z_1 with P_1, the other joins z_1 with the origin. Let us follow the remaining two arcs away from $z < z_1$. None of these arcs can intersect the imaginary axis as there is already one arc of the curve $R = u_1$ which does so and $R(0, y)$ increases steadily with $|y|$. Further, none of the arcs in question can wander off to infinity or end at the origin. We are thus sure that one of these arcs will intersect the real axis between $-\frac{1}{2}$ and 0 and the other will intersect the line $x = -\frac{1}{2}$, $y > 0$. As $R(x, y)$ is monotone on both lines there cannot be more than one intersection on each. The arc which intersects the real axis clearly joins $z = z_1$ with $z = p_1$. It follows that

$$u_1 = \psi(p_1) > \psi\left(-\frac{1}{2}\right) = \psi\left(\frac{3}{2}\right) > 0. \tag{87}$$

The arc which intersects $x = -\frac{1}{2}$ remains. This arc will pass through $z = -1$ if we can prove that it cannot intersect the line $x = -1$ at a point of ordinate different from 0. It clearly cannot intersect the real axis between -1 and $-\frac{1}{2}$. In view of (81) it is sufficient for our purpose to prove that

$$(88) \qquad u_1 < \psi(2).$$

Let us consider the rectangle whose vertices are $A = 0$, $B = i\eta$, $C = -\frac{1}{2} + i\eta$, $D = -\frac{1}{2}$, where $\eta\,(> 0)$ is to be suitably chosen. It is clear that the curve $R = u_1$ intersects the polygonal line $ABCD$ at least once. Hence $u_1 \leqq \text{Max}\, R(x, y)$ on $ABCD$. Now $R(x, y)$ is monotone increasing on AB and on DC. Hence $u_1 \leqq \text{Max}\, R(x, y)$ on CB. The latter maximum can be estimated with the aid of the methods which we have used in the latter half of § 3. In view of (11) and (34) we have

$$R(x, \eta) \leqq \text{Max}\; R(1-x, \eta) + \frac{\pi}{\text{sh}\, 2\pi\eta},$$

when z lies on CB. Hence

$$(89) \qquad u_1 \leqq R\left(\frac{3}{2}, \eta\right) + \frac{\pi}{\text{sh}\, 2\pi\eta},$$

no matter how η be chosen > 0. Now it will be proved in § 15 that it is always possible to find an η such that

$$(90) \quad R\left(n + \frac{1}{2}, \eta\right) + \frac{\pi}{\text{sh}\, 2\pi\eta} < R(n+1, 0) = \psi(n+1).$$

Hence (88) is actually true[1].

We can now account for all the arcs $R_0(c)$. When $c < u_1$, $R_0(c)$ intersects the positive real axis between 0 and P_1 where $\frac{3}{2} < P_1 < 2$. When $c > u_1$, $R_0(c)$ intersects the negative real axis between p_1 and 0 where $-\frac{1}{2} > p_1$. The two different types of curves are separated and enclosed by lobes of $R_0(u_1)$.

[1] It follows from Table II in § 16 and the corresponding Fig. 4 that u_1 lies between 0.1 and 0.2.

At $z = -1$ we have a similar situation. As long as $c < u_1$, $R_1(c)$ intersects the negative real axis between -1 and p_1 and all these curves are enclosed by a lobe of $R_0(u_1) = R_1(u_1)$. When c is somewhat larger than u_1, $R_1(c)$ intersects the positive real axis beyond P_1. There is a largest value of c, c_2 say, for which this is the case, and we prove in the same manner as above that $c_2 = u_2$. Thus all the curves $R_1(c)$ with $u_1 < c < u_2$ intersect the positive real axis between P_1 and P_2 where $\psi(P_2) = u_2$. When $c > u_2$, $R_1(c)$ intersects the negative real axis to the left of $z = -1$, namely, between p_2 and -1 where $p_2 > -\frac{3}{2}$ and $\psi(p_2) = u_2$.

Finally, in the general case the curves $R_n(c)$ fall into three classes: (i) Curves corresponding to $c < u_n$; these curves intersect the negative real axis to the right of $z = -n$ between $-n$ and p_n, where $-n + \frac{1}{2} < p_n < -n + 1$ and where $\psi(p_n) = u_n$. (ii) Curves corresponding to $u_n < c < u_{n+1}$; these curves intersect the positive real axis between P_n and P_{n+1} where $n + \frac{1}{2} < P_n < n + 1$ and where $\psi(P_n) = u_n$. (iii) Curves corresponding to $c > u_{n+1}$; these curves intersect the negative real axis to the left of $z = -n$, between p_{n+1} and $-n$. The three different types of curves are separated by lobes of the critical curves $R_n(u_n)$ and $R_n(u_{n+1})$.

15. Inequalities for the critical values. We have thus completed the qualitative description of the R, I-net. It remains to prove formula (90). For this purpose we resort to formula (22), which has not been used in the earlier part of the paper, namely,

$$\psi(z+h)-\psi(z) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n+1} \frac{h(h-1)\dots(h-n)}{z(z+1)\dots(z+n)}.$$

Consequently, if $z = x$ is real and positive

$$\left|\psi(x+h)-\psi(x)-\frac{h}{x}\right|$$

$$\leq \frac{|h|}{x}\sum_{n=1}^{\infty}\frac{1}{n+1}\frac{(|h|+1)(|h|+2)\dots(|h|+n)}{(x+1)(x+2)\dots(x+n)}$$

$$\leq \frac{|h|}{2x}\sum_{n=1}^{\infty}\frac{(|h|+1)\dots(|h|+n)}{(x+1)\dots(x+n)} = \frac{|h|(|h|+1)}{2x(x-|h|-1)}$$

provided $x > |h|+1$. Hence

(91) $$\psi(x+h)-\psi(x) = \frac{h}{x}\left[1+\varrho(x,h)\right],$$

where

(92) $$|\varrho(x,h)| < \frac{|h|+1}{2(x-|h|-1)}\ \text{when}\ x > |h|+1.$$

We now choose $x > (|h|+1)(2|h|+1)$ and set $h = k+il$. Then

$$|\varrho(x,h)| < \frac{1}{4|h|}$$

and

(93) $R(x+k,l)-R(x,0) = \dfrac{k}{x}+P(x,h)$ where $|P(x,h)| < \dfrac{1}{4x}$.

In (93) we put $x = n+1$, $k = -\frac{1}{2}$, $l = \eta$ and obtain

(94) $$R(n+1,0)-R\left(n+\frac{1}{2},\eta\right) > \frac{1}{4(n+1)}$$

provided $n \geq 2\left(\eta^2+\frac{1}{4}\right)+3\sqrt{\eta^2+\frac{1}{4}}$, and, a fortiori, if $n \geq 5\left(\eta^2+\frac{1}{4}\right)$ when $\eta \geq \frac{\sqrt{3}}{2}$. It is possible to replace (94) by

$$R(n+1,0) - R\left(n+\frac{1}{2},\eta\right) > \frac{\pi}{\operatorname{sh} 2\pi\eta}$$

as long as

$$\frac{\pi}{\operatorname{sh} 2\pi\eta} < \frac{1}{4(n+1)}.$$

Suppose that we choose n and η subject to the following double inequality

$$(95) \qquad 5\left(\eta^2+\frac{1}{4}\right) \leqq n < \frac{1}{4\pi} \operatorname{sh} 2\pi\eta - 1, \quad \eta \geqq \frac{\sqrt{3}}{2};$$

then (94) implies that (90) is fulfilled for such values of n and η. It is now obvious that when we give ourselves an $n \geqq 5$, we can find an η satisfying (95). Thus to every $n \geqq 5$ there exists an η for which (90) holds. We can verify by numerical calculation that (90) holds for $\eta = 1$ when $n = 1, 2, 3$ and 4. We obtain from Table I that

n	1	2	3	4
$R\left(n+\frac{1}{2}, 1\right)$	0.3480	0.8096	1.1544	1.4386
$\psi(n+1)$	0.4228	0.9228	1.2561	1.5061

Since $\dfrac{\pi}{\operatorname{sh} 2\pi} = 0.0117$ we have verified our statement.

We can obtain an asymptotic expression for $w_n = \psi(z_n)$ for large values of n with the aid of (19) and (70). The result is rather complicated and will not be given here; it permits us to conclude, however, that

$$\psi(P_n) - \psi\left(n+\frac{1}{2}\right) = O\left(\frac{\log^2 n}{n^2}\right);$$

whence

$$(96) \qquad P_n - n - \frac{1}{2} = O\left(\frac{\log^2 n}{n}\right),$$

and similarly

$$(97) \qquad p_n + n - \frac{1}{2} = O\left(\frac{\log^2 n}{n}\right).$$

The inequalities obtained for the critical values w_n can be summarized as follows:

Theorem: Let z_n and $\bar{z}_n$ $(n = 1, 2, 3, \ldots)$ denote the zeros of $\psi'(z)$ where $-n+\frac{1}{2} < \Re(z_n) < -n+1$ and set $w_n = \psi(z_n) = u_n + iv_n$. Then

$$\psi\left(n+\frac{1}{2}\right) < u_n < \psi(n+1), \tag{98}$$

$$\frac{\pi}{2} < v_n < v_{n+1} < \pi, \tag{99}$$

where $n = 1, 2, 3, \ldots$ Further,

$$u_n = \psi\left(n+\frac{1}{2}\right) + O\left(\frac{\log^2 n}{n}\right), \tag{100}$$

$$v_n = \pi - O\left(\frac{\log n}{n}\right). \tag{101}$$

16. Numerical computation of $\psi(z)$. We can also attack the question of how the values of $\psi(z)$ are distributed with the aid of numerical calculation. Such computations are fairly easy to carry out on the imaginary axis; with the aid of the formulas in § 2 we can afterwards obtain values of the function on other vertical lines.

Formulas (17) and (18), namely,

$$I(0, y) = \frac{\pi}{2}\coth \pi y + \frac{1}{2y}, \quad I\left(\frac{1}{2}, y\right) = \frac{\pi}{2}\operatorname{th} \pi y,$$

enable us to calculate the imaginary part of $\psi(z)$ on the lines $\Re(z) = 0$ and $\frac{1}{2}$. The values on the lines $\Re(z) = n$ and $n+\frac{1}{2}$ are then obtainable with the aid of (10). It does not seem to be possible to get the values of $I(x, y)$ on any other vertical lines with the aid of the formulas in § 2 without the use of (19).

The situation with regard to $R(x, y)$ is rather different. We have

$$R(0, y) = -C + y^2 \sum_{n=1}^{\infty} \frac{1}{n(n^2 + y^2)}. \tag{102}$$

This series is not well suited for numerical work, nor does its sum seem to be expressible in terms of elementary functions. To obtain more rapidly convergent series we use transformations of Kummer's type. Writing

$$S_p = \sum_{n=1}^{\infty} \frac{1}{n^p} \qquad (p \geqq 2)$$

we easily see that

$$\left\{\begin{aligned} R(0, y) &= -C + y^2 S_3 - y^4 S_5 + \dots \\ &+ (-1)^{k-1} y^{2k} S_{2k+1} + (-1)^k y^{2k+2} \sum_{n=1}^{\infty} \frac{1}{n^{2k+1}(n^2 + y^2)}. \end{aligned}\right. \tag{103}$$

If $|y| \leqq 1$ and k equals 4 or 5, this expression is quite suited for computations. When $|y| > 1$ we can still apply the same method if we let the transformations apply to the remainder after a suitably chosen term of the original series. For certain values of y the series

$$S_m(y) = \sum_{n=m}^{\infty} \frac{1}{n(n^2 - y^2)}$$

has a known value. Thus

$$S_1\left(\frac{1}{2}\right) = 8 \log 2 - 4, \; S_2(1) = \frac{1}{4}, \; S_3(2) = \frac{11}{96} \text{ etc.}$$

For such values of y we can obtain a rapidly convergent series in fewer steps, e. g.,

$$R\left(0, \frac{1}{2}\right) = -C + 2\log 2 - 1 - \frac{1}{8} S_5 - \frac{1}{256} S_9$$
$$- \frac{1}{2048} \sum_{n=1}^{\infty} \frac{1}{n^9\left(n^4 - \frac{1}{16}\right)}.$$

The remainder after the first term of the infinite series contributes less than 10^{-7} to the value of $R\left(0, \frac{1}{2}\right)$. The sums S_p which are needed for the computations can be taken from Stieltjes' table in Acta Mathematica, vol. 10.

Formula (103) becomes unmanageable when $|y|$ is larger than about 3. For such values we have to resort to formula (19), which is very convenient for numerical work. Using formulas (20) and (21) with $m = 5$ and substituting the values of the Bernoullian numbers we obtain

(104) $$R(0, y) = \log|y| + \frac{1}{12y^2} + \frac{1}{120y^4} + \frac{1}{252y^6} + \frac{1}{240y^8} + \frac{1}{132y^{10}} + R_{11},$$

(104 a) $$|R_{11}| < \frac{64}{2079\, y^{10}}.$$

Using one or the other of these series we have computed the following values

$$R\left(0, \frac{1}{4}\right) = -0.505907, \quad R\left(0, \frac{1}{3}\right) = -0.455210, \quad R\left(0, \frac{1}{2}\right) = -0.328886,$$
$$R\left(0, \frac{2}{3}\right) = -0.186352, \quad R\left(0, \frac{3}{4}\right) = -0.113901, \quad R(0, 1) = 0.094650,$$
$$R\left(0, \frac{3}{2}\right) = 0.444698, \quad R(0, 2) = 0.714592, \quad R\left(0, \frac{9}{4}\right) = 0.827758,$$
$$R(0, 3) = 1.108907, \quad R(0, 4) = 1.391537.$$

The error in these values, barring unfortunate accidents, amounts to less than one unit in the last decimal place.

Knowing $R(0, y)$ and $R(0, 2y)$ we can compute $R\left(\frac{n}{2}, y\right)$ with the aid of (9) and (13). If we know $R(0, 3y)$ in addition, we can obtain $R\left(\frac{n}{3}, y\right)$ with the aid of (9), (11) and (13). Finally, if we know $R(0, y)$, $R(0, 2y)$ and $R(0, 4y)$ we can get $R\left(\frac{n}{4}, y\right)$ with the aid of the same formulas which supply the necessary number of equations.

In the adjoining Table I we have listed the values of $\psi(x+iy)$ for some values of x and y. The sign » in any place of the table indicates that the corresponding value has not been calculated; thus the imaginary part is given for only half of the entries. A last digit set in heavier type indicates that the decimal in question has been raised. The values listed above permit extending the table considerably. In Table II we have listed the real part of $\psi(x+iy)$ at 40 different points in the square $-1 \leqq x \leqq 0$, $0 \leqq y \leqq 1$. This table illustrates the run of $R(x, y)$ in the neighborhood of the critical point $z = z_1$. The adjoining Figure 4 is based upon this table; it shows the interpolated curves $R = c$ for $c = -0.5, -0.4, \ldots, 0.5$ and 0.6. In order to avoid crowding the figure we have left out most of the arcs of these curves in the lower half of the diagram. The dots in the figure mark the points where the values of $R(x, y)$ have been calculated. The table and the figure together would seem to suggest that z_1 is near to the point $z = -\frac{1}{3} + \frac{2i}{3}$ and that u_1 is about 0.16.

17. The Riemann surface of $w = \psi(z)$. We can now form a fairly good idea of the structure of the Riemann surface corresponding to $w = \psi(z)$ and its inverse. The singularities of the inverse function $z = \psi^{-1}(w)$ are $w = \infty$, which is a *transcendental critical point*, together with all the points $w = w_n$ and $\overline{w}_n$. The latter points are

Table I.

Values of $\psi(x+iy)$.

$x \backslash y$	0	0.25	0.50	0.75	1
— 2	∞	0.9276 + 4.6859 i	0.9417 + 3.2305 i	0.9644 + 2.7903 i	0.9947 + 2.7767 i
— 1.75	— 2.3028	— 0.4273 + »	0.5709 + »	0.8099 + »	0.8940 + »
— 1.50	0.7032	0.7106 + 1.9720 i	0.7319 + 2.6406 i	0.7667 + 2.7362 i	0.8096 + 2.6025 i
— 1.25	3.7142	1.8340 + »	0.8802 + »	0.7052 + »	0.7181 + »
— 1	∞	0.4353 + 4.6244 i	0.4711 + 3.1128 i	0.5261 + 2.6259 i	0.5947 + 2.5767 i
— 0.75	— 2.8942	0.9873 + »	0.0426 + »	0.3272 + »	0.4633 + »
— 0.50	0.0365	0.0619 + 1.8365 i	0.1319 + 2.4406 i	0.2324 + 2.4695 i	0.3480 + 2.3649 i
— 0.25	2.9142	1.0647 + »	0.1905 + »	0.1169 + »	0.2303 + »
0	∞	— 0.5059 + 4.3891 i	— 0.3289 + 2.7128 i	— 0.1139 + 2.2659 i	0.0947 + 2.0767 i
0.25	— 4.2275	— 2.1873 + »	— 0.8803 + »	— 0.3395 + »	— 0.0167 + »
0.50	— 1.9635	— 1.5381 + 1.0365 i	— 0.8681 + 1.4406 i	— 0.3830 + 1.5424 i	— 0.0520 + 1.5649 i
0.75	— 1.0858	— 0.9353 + »	— 0.6095 + »	— 0.2831 + »	— 0.0050 + »
1	— 0.5772	— 0.5059 + 0.3891 i	— 0.3289 + 0.7128 i	— 0.1139 + 0.9326 i	0.0947 + 1.0767 i
1.25	— 0.2275	— 0.1873 + »	— 0.0803 + »	0.0605 + »	0.2186 + »
1.50	0.0365	0.0619 + 0.2365 i	0.1319 + 0.4406 i	0.2324 + 0.6193 i	0.3480 + 0.7649 i
1.75	0.2475	0.2647 + »	0.3136 + »	0.3836 + »	0.4750 + »
2	0.4228	0.4353 + 0.1538 i	0.4711 + 0.3128 i	0.5261 + 0.4526 i	0.5947 + 0.5767 i

Table II.

Values of $R(x, y)$.

$y \backslash x$	-1	$-\frac{3}{4}$	$-\frac{2}{3}$	$-\frac{1}{2}$	$-\frac{1}{3}$	$-\frac{1}{4}$	0
1	0.5947	0.4633	0.4234	0.3480	0.2720	0.2303	0.0947
$\frac{3}{4}$	0.5261	0.3272	0.2895	0.2324	0.1686	0.1169	−0.1139
$\frac{2}{3}$	0.5060	»	0.2265	»	0.1578	»	−0.1864
$\frac{1}{2}$	0.4711	0.0426	0.0311	0.1319	0.2192	0.1905	−0.3289
$\frac{1}{3}$	0.4448	»	−0.3727	0.0808	0.5160	»	−0.4552
$\frac{1}{4}$	0.4353	−0.9873	»	0.0619	»	1.0647	−0.5059
0	∞	−2.8942	−1.0548	0.0365	1.2590	2.9142	∞

algebraic branch-points in the neighborhood of which two determinations of z are interchanged.

In order to build up the Riemann surface we consider the map of the z-plane corresponding to the transformation $w = \psi(z)$. It is clear that this map will cover itself infinitely often. Thus we have to cut up the z-plane into regions such that each region has a smooth non-overlapping image and then we must piece these different images together. It then becomes a question of how these regions should be chosen. Our previous study of the R, I-net shows that the critical curves through the points z_n and $\bar{z}_n$ give a natural division of the plane into suitable regions. We can choose either the curves $R_n(u_n)$ or the curves $I_n(v_n)$ for this purpose; we select the former curves. We then imagine the plane cut up along those arcs of $R_n(u_n)$ which join z_n and $\bar{z}_n$ with $-n$ and $-n+1$. We do not, however, cut the plane along the remaining arcs of $R_n(u_n)$ which join z_n

with $\bar{z}_n$ over p_n and P_n, respectively. The region outside of all the cuts we denote by D_0. The region inside of the cuts from $z = -n$ over z_n, $-n+1$ and $\bar{z}_n$ back to $-n$ will be denoted by D_n $(n = 1, 2, 3, \ldots)$.

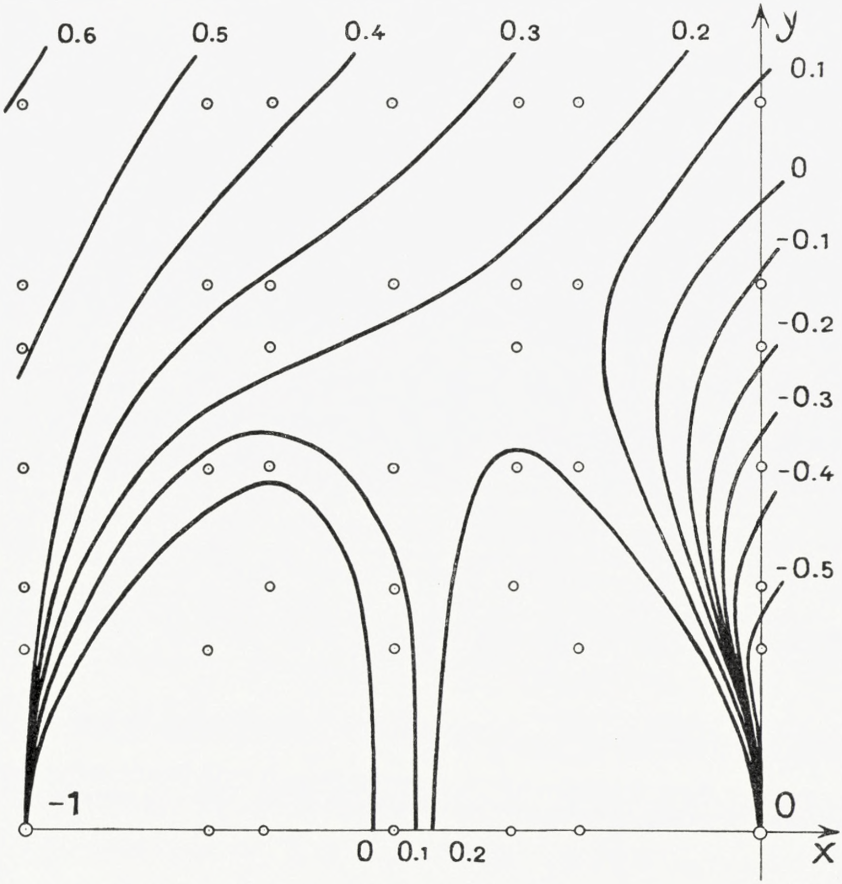


Fig. 4.

We begin by considering D_0. This is a simply-connected region, if we leave out the points $z = -n$ $(n \geqq 1)$, in the interior of which $\psi(z)$ is holomorphic and $\psi'(z) \neq 0$. We shall prove that the image of D_0 by the transformation $w = \psi(z)$ is a full plane slit up along the lines $u = u_n$, $v \geqq v_n$ and $u = u_n$, $v \leqq -v_n$ $(n = 1, 2, 3, \ldots)$.

In order to see this we shall consider the equation

$$\psi(z) = u + iv.$$

It is not difficult to see that this equation has one and only one solution in the interior of D_0 if $u+iv$ is not on the slits just mentioned, and if $u+iv$ is located on one of the slits there are two solutions on the boundary of D_0. In fact, suppose that $u_m < u < u_{m+1}$.[1] We can then locate $R_m(u)$ in D_0; this curve goes from $z = -m$ back to this point, intersecting the positive real axis between P_m and P_{m+1}. It lies entirely in D_0 and it is the only branch of the curve $R = u$ in D_0, all the other branches are in the excluded regions. If we trace $R_m(u)$ once from $-m$ back to this point going in the positive sense, $I(x, y)$ increases steadily from $-\infty$ to $+\infty$. Thus there is one and only one point on the curve where $I(x, y) = v$ and this point gives the desired solution, which is obviously unique. The case in which $u = u_m$ is easily disposed and will not be considered here. We designate the image of D_0 by Π_0.

In the interior of D_1, $\psi(z)$ takes on every value once and only once with the exception of the values $u = u_1$, $v \geqq v_1$ and $u = u_1$, $v \leqq -v_1$ which are not taken on at all in the interior but twice on the boundary instead. Thus we find that D_1 is mapped upon a full plane slit along the lines $u = u_1$, $v \geqq v_1$ and $u = u_1$, $v \leqq -v_1$. Let this slit plane be denoted by Π_1; Π_0 and Π_1 are evidently connected along the common cuts. In general, the region D_n $(n \geqq 1)$ is mapped upon a full plane Π_n slit along the lines $u = u_n$, $v \geqq v_n$ and $u = u_n$, $v \leqq -v_n$,

[1] We set $u_0 = -\infty$.

and this plane is connected with Π_0 along the common cuts. There is obviously no direct connection between Π_m and Π_n if $mn \neq 0$. The totality of these sheets Π_m constitutes the Riemann surface of $\psi(z)$.

18. Generalizations. In concluding we shall raise the question of the extent to which the results obtained in the present paper may be considered typical for the class of functions defined as principal solutions of equations of the form

$$\underset{\omega}{\Delta} F(z) = \varphi(z). \tag{105}$$

Without pretending to answer this question we shall call attention to a few facts which have an obvious bearing on the situation.

There are many details in the preceeding discussion which are of a highly special nature and which cannot be carried over to a more general case. But the fundamental results of the investigation have been derived either directly from the defining difference equation (2) or from the complementary theorem (3), the multiplication theorem (4) and the asymptotic formula (19). The latter three theorems are all immediate consequences of the difference equation and are not dependent upon the special analytic form of the solution. Now the principal solution of (105) does satisfy a complementary theorem, a multiplication theorem and an asymptotic relation all of a fairly simple nature under very general assumptions on $\varphi(z)$. Further, if $\varphi(z)$ is single-valued the nature and distribution of the singularities of $F(z)$ shows considerable resemblance to the corresponding situation for $\psi(z)$. There is consequently some ground for expecting that also the finer structure of the distribution of the values taken on by the several functions shall show striking resemblances in the special case here treated and the general case mentioned above.

Færdig fra Trykkeriet den 15. August 1927.

MATHEMATISK-FYSISKE MEDDELELSER

UDGIVNE AF

DET KGL. DANSKE VIDENSKABERNES SELSKAB

6. BIND (Kr. 17,00):

7. BIND (Kr. 17,70):

8. BIND (under Pressen):

Det Kgl. Danske Videnskabernes Selskab.
Mathematisk-fysiske Meddelelser. **VIII,** 2.

BEITRAG ZUR THEORIE DER UNVOLLSTÄNDIGEN GAMMAFUNKTIONEN

NACH HINTERLASSENEN PAPIEREN
VON J. L. W. V. JENSEN

VON

G. RASCH

KØBENHAVN
HOVEDKOMMISSIONÆR: ANDR. FRED. HØST & SØN, KGL. HOF-BOGHANDEL
BIANCO LUNOS BOGTRYKKERI
1927

1. Einleitung.

Die unvollständigen Gammafunktionen $P(z,\varrho)$ und $Q(z,\varrho)$ können definiert werden durch die Integrale

$$(1) \qquad P(z,\varrho) = \int_0^{\varrho} e^{-t}\, t^{z-1}\, dt, \qquad \Re z > 0$$

$$(2) \qquad Q(z,\varrho) = \int_{\varrho}^{\infty} e^{-t}\, t^{z-1}\, dt,$$

wobei ϱ nicht negativ reell ist. Der Integrationsweg soll die negative reelle Achse nicht schneiden. Die Summe der beiden Funktionen ist das zweite Eulersche Integral

$$(3) \qquad \Gamma(z) = \int_0^{\infty} e^{-t}\, t^{z-1}\, dt, \qquad \Re z > 0.$$

Für verschiedene Eigenschaften der beiden Funktionen, die wir im folgenden brauchen, sei verwiesen auf N. E. Nörlund, Differenzenrechnung, Berlin 1924, p. 388—406, N. Nielsen, Theorie der Gammafunktion, Leipzig 1906, p. 25—36, 209—219 und N. Nielsen, Theorie des Integrallogarithmus, Leipzig 1906.

Über die unvollständigen Gammafunktionen finden sich in Jensens nachgelassenen Papieren gewisse Untersuchungen, welche die wesentliche Grundlage der folgenden Arbeit bilden. Und zwar handelt es sich einmal um zwei Sätze über die Nullstellen der P-Funktion, die mit naheliegenden Verallgemeinerungen in § 2 mitgeteilt werden. Im nächsten Paragrafen wird der asymptotische Ausdruck für die Nullstellen der Q-Funktion hergeleitet, den Jensen 1924 der

Dänischen Gesellschaft der Wissenschaften vorlegte und übrigens schon 1893 in der Kopenhagener Mathematischen Vereinigung vorgetragen hatte. Dieser Ausdruck wird auch etwas verschärft. § 4 enthält einen von Jensen gegebenen Beweis für eine von Schlömilch herrührende Fakultätenreihendarstellung von $Q(z,\varrho)$ und § 5 die neuen Entwicklungen (53) und (54) für die unvollständigen Gammafunktionen, die auch von Jensen herrühren; für diese Entwicklungen werden einige Anwendungen gegeben.

2. Die Nullstellen der *P*-Funktion.

Gronwall[1] hat einen sehr einfachen Beweis dafür gegeben, dass die imaginären Nullstellen von $P(z) = P(z,1)$ in der Halbebene $\Re z \leqq -\frac{3}{2}$ liegen. Seine Beweisführung kann vertieft werden, so dass man einen Satz über die Lage der imaginären Nullstellen von $P(z,\varrho)$ für $0 < \varrho \leqq 1$ erhält.

Bei $0 < \varrho \leqq 1$ liegt jede imaginäre Nullstelle von $P(z,\varrho)$ in dem Gebiete, das gemeinsam überdeckt wird von zwei Kreissystemen C_s und C'_s $(s = 0, 1, 2, \ldots)$, die auf folgende Weise definiert sind: C_s ist der Kreis mit dem Mittelpunkte

$$c_s = -2s - \frac{1}{1 - \frac{\varrho}{2s+1}} \tag{4}$$

und dem Radius

$$\varrho_s = \frac{\sqrt{\frac{\varrho}{2s+1}}}{1 - \frac{\varrho}{2s+1}}, \tag{5}$$

während C'_s den Mittelpunkt $c'_s = c_s - 1$ und den Radius ϱ_s hat.

[1] Ann. Éc. Norm. (3) 33 (1916), p. 381—393.

Beweis: Aus der Partialbruchreihe

$$P(z,\varrho) = \sum_{s=0}^{\infty} \frac{(-1)^s}{s!} \cdot \frac{\varrho^{z+s}}{z+s} \tag{6}$$

gewinnt man mit $z = x + iy$

$$\left\{\begin{aligned} -\frac{1}{y}\Im\left(\varrho^{-z}P(z,\varrho)\right) &= \sum_{s=0}^{\infty} \frac{(-1)^s}{s!} \cdot \frac{\varrho^s}{|z+s|^2} \\ &= \sum_{s=0}^{\infty} \frac{\varrho^{2s}}{(2s)!} \cdot \left[\frac{1}{|z+2s|^2} - \frac{\varrho}{(2s+1)\,|z+2s+1|^2}\right]. \end{aligned}\right. \tag{7}$$

Die rechte Seite ist offenbar positiv, wofern

$$|z+2s+1|^2 \geqq \frac{\varrho}{2s+1} |z+2s|^2$$

für $s = 0, 1, 2, \ldots$. D. h.: $P(z,\varrho) \neq 0$, wenn z ausserhalb aller Kreise C_s oder auf einem von ihnen liegt.

Mit Benützung der Differenzengleichung

$$P(z+1,\varrho) = z\,P(z,\varrho) - e^{-\varrho}\varrho^z \tag{8}$$

folgt

$$\Im\left(z\varrho^{-z}P(z,\varrho)\right) = \Im\left(\varrho^{-z}P(z+1,\varrho)\right), \tag{9}$$

und hieraus ersieht man, dass $P(z,\varrho)$ auch $\neq 0$, wenn $z+1$ ausserhalb der Kreise C_s liegt. Hiermit ist der Satz bewiesen.

Für grosse s stellt C_s einen sehr kleinen Kreis um $-2s-1$ dar, während C'_s ein sehr kleiner Kreis um $-2s-2$ ist. Hieraus geht hervor, dass das beiden Kreissystemen C_s und C'_s gemeinsame Gebiet in einer endlichen Anzahl dieser Kreise enthalten sein muss. Wir wollen demgemäss zeigen, dass die imaginären Nullstellen innerhalb der Kreise C_0, C_1, C_2, C_3 und C'_0, C'_1, C'_2 liegen. Zunächst merken wir an, dass die Mittelpunkte c_1, c'_1, c_2, c'_2, ... eine monotone Folge bilden. Man hat ja immer $c_s > c'_s$; die Ungleichheit $c'_s > c_s + 1$ kann umgeschrieben werden zu

$$1 - \frac{2\varrho}{2s+1} + \frac{\varrho^2}{(2s+1)(2s+3)} > 0,$$

und man sieht sogleich ein, dass sie bei $s \geqq 1$ für alle positiven $\varrho \leqq 1$ erfüllt ist. Weiter wollen wir beweisen, dass die Kreise C_3, C'_3, C_4, C'_4, ... ganz getrennt liegen. Die Schnittpunkte zwischen dem Kreise C_s und der negativen reellen Achse befinden sich in

$$c_s \pm \varrho_s = -2s - \frac{1}{1 \pm \sqrt{\frac{\varrho}{2s+1}}}, \tag{10}$$

während für C'_s die entsprechenden Punkte $c_s - 1 \pm \varrho_s$ sind. Die Bedingung dafür, dass C_s ganz ausserhalb C'_s liegt, lautet

$$\varrho_s < \frac{1}{2}.$$

Diese Ungleichheit ist identisch mit

$$2s + 1 > \varrho\left(3 + 3\sqrt{8}\right) = 5{,}83\,\varrho,$$

was bei $s \geqq 3$ offenbar für alle positiven $\varrho \leqq 1$ zutrifft. Die Bedingung dafür, dass C'_s keinen Punkt mit C_{s+1} gemein hat, heisst

$$c'_s - \varrho_s > c_{s+1} + \varrho_{s+1}$$

und kann mit Hilfe von (10) umgeschrieben werden zu

$$1 - 2\sqrt{\frac{\varrho}{2s+1}} - \frac{\varrho}{\sqrt{(2s+1)(2s+3)}} > 0.$$

Eine einfache Rechnung lehrt, dass auch diese Ungleichheit bei $s \geqq 3$ für alle positiven $\varrho \leqq 1$ stattfindet.

Das soeben Bewiesene schliesst natürlich nicht aus, dass z. B. C_0' einen oder mehrere der Kreise C_s und C_s' mit $s \geqq 3$ umfasst. Für $\varrho = 1$ ist C_0' sogar die ganze Halbebene $\Re z \leqq -\frac{3}{2}$ und enthält demnach alle diese Kreise.

Der Beweis kann nicht ohne weiteres auf den Fall $\varrho > 1$ übertragen werden; jedoch kann man mit Hilfe der Fakultätenreihe

$$(11) \qquad P(z,\varrho) = e^{-\varrho}\varrho^z \sum_{s=0}^{\infty} \frac{\varrho^s}{z(z+1)\cdots(z+s)}$$

den folgenden Satz erhalten, der für alle (auch komplexen) ϱ gilt, aber für positives $\varrho \leqq 1$ nicht so weit geht wie der eben bewiesene.

$P(z,\varrho)$ hat keine Nullstellen in dem ausserhalb des Kreises $|z+1| = 2|\varrho|$ gelegenen Teile der Halbebene $\Re z \geqq -\frac{3}{2}$.

Beweis: In der betrachteten Halbebene ist

$$|z+s+1| \geqq |z+1|$$

für alle positiven ganzen s. Aus (11) bekommt man daher

$$|ze^{\varrho}\varrho^{-z}P(z,\varrho) - 1| < \sum_{s=1}^{\infty} \left|\frac{\varrho}{s+1}\right|^s = \frac{|\varrho|}{|z+1| - |\varrho|},$$

wofern $|z+1| > |\varrho|$. Befindet sich z sogar ausserhalb des Kreises $|z+1| > 2|\varrho|$, so fällt

$$1 > \frac{|\varrho|}{|z+1| - |\varrho|},$$

also $P(z,\varrho) \neq 0$ aus.

3. Die Nullstellen der Q-Funktion.

Nach Lindhagen und Nielsen[1] liegen die Nullstellen der ganzen Funktion $Q(z,\varrho)$ bei $\varrho \geqq 0$ in der Halbebene $\Re z > \varrho$. Stieltjes[2] bewies in schöner Anwendung des Picardschen Satzes über ganze Funktionen, dass $Q(z,\varrho)$ wirklich unendlich viele Nullstellen hat. Dies folgt übrigens auch aus den Anfangsgründen der Hadamardschen Theorie, weil $Q(z,\varrho)$ offenbar eine ganze Funktion vom Geschlecht 1 ist, die nicht von der Form $p(z)e^{kz}$ mit ganzem rationalem $p(z)$ und konstantem k sein kann; denn wäre sie es, so könnte sie nicht längs der positiven reellen Achse dieselbe Grössenordnung wie $\Gamma(z)$ besitzen.

Aus der Integraldarstellung (2) für $Q(z,\varrho)$ entnimmt man weiterhin $Q(z,\varrho) > 0$ für reelles z, sodass alle Nullstellen imaginär ausfallen. Wir wollen ihre asymptotische Lage bestimmen. Die Gleichung

$$Q(z,\varrho) = 0 \tag{12}$$

ist identisch mit

$$\Gamma(z) = P(z,\varrho), \tag{12a}$$

so dass die Nullstellen die Schnittpunkte zwischen der Kurve γ

$$\log|\Gamma(z)| = \log|P(z,\varrho)| \tag{13}$$

und den Kurven α_n

$$\operatorname{arc} \Gamma(z) = \operatorname{arc} P(z,\varrho) + 2n\pi \tag{14}$$

sind, wobei n alle positiven und negativen ganzen Zahlen einschliesslich 0 durchläuft. In der Halbebene $\Re z > 0$ ist $\operatorname{arc} \Gamma(z)$ eindeutig bestimmt als stetige Fortsetzung von

[1] Vgl. z. B. N. Nielsen, Gammafunktion, p. 211—212.

[2] Correspondance d'Hermite et de Stieltjes, Paris 1905, lettre no. 219.

arc $\Gamma(z) = 0$ für positives z. Um arc $P(z,\varrho)$ festzulegen, müssen wir einen Schnitt führen, der z am Umlaufen einer Nullstelle von $P(z,\varrho)$ verhindert; nach dem Schlusse von § 2 können wir als solchen Schnitt den Kreis $|z+1| = 2\varrho$ wählen.

Zunächst soll gezeigt werden, dass α_n bei hinreichend grossem $|n|$ die Kurve γ ein- und nur einmal schneidet. γ besteht aus zwei Zweigen, die zur positiven reellen Achse symmetrisch liegen, und α_n ist in bezug auf diese symmetrisch zu α_{-n}. Wie aus dem folgenden hervorgehen wird, befindet sich α_n für genügend grosse $n > 0$ ganz oberhalb der positiven reellen Achse, und es genügt deshalb, den Fall $n > 0$, $\Im z \geqq 0$ zu betrachten.

Wird

$$(15) \qquad \omega(z) = \log \Gamma(z) - \left(z - \frac{1}{2}\right) \log z - z - \log \sqrt{2\pi}$$

und

$$(16) \qquad \omega(z,\varrho) = \log\left(1 + \sum_{s=1}^{\infty} \frac{\varrho^s}{(z+1)\cdots(z+s)}\right)$$

sowie $z = x + iy$ gesetzt, so kann die Gleichung (13) auch

$$(17) \quad \begin{cases} \left(x + \frac{1}{2}\right) \log|z| - y \operatorname{arctg}\frac{y}{x} - x \log \varrho e + \varrho + \log\sqrt{2\pi} \\ \qquad = \Re\left(\omega(z,\varrho) - \omega(z)\right) = O\left(\frac{1}{z}\right) \end{cases}$$

geschrieben werden. Wenn z in der Halbebene $\Re z > 0$ gegen Unendlich geht und (17) befriedigen soll, so kann x offenbar nicht beschränkt bleiben. Dividiert man die Gleichung durch x, so erhält man

$$\left(1 + \frac{1}{2x}\right) \log|z| - \frac{y}{x} \operatorname{arctg} \frac{y}{x} = O(1)$$

und sieht hieraus, dass auch

$$\frac{y}{x} \to \infty. \tag{18}$$

Die Tangentenrichtung in einem Punkte z auf γ ist bestimmt durch

$$\frac{dy}{dx} = \frac{\log|z| + \dfrac{y}{2|z|^2} - \log\varrho + \Re\left(\omega'(z) - \omega'(z,\varrho)\right)}{\operatorname{arctg}\dfrac{y}{x} - \dfrac{y}{2|z|^2} + \Im\left(\omega'(z) - \omega'(z,\varrho)\right)}; \tag{19}$$

da $\omega'(z)$ und $\omega'(z,\varrho)$ für $z \to \infty$ nach 0 streben, wird auf Grund von (18)

$$\frac{dy}{dx} \simeq \frac{2}{\pi}\log|z| \qquad \text{für } |z| \to \infty. \tag{19a}$$

Hieraus folgt, dass die Kurve γ für grosse $|z|$ monoton wächst.

Jetzt gehen wir an die Untersuchung der Kurven α_n. Nach (15) und (16) können ihre Gleichungen in der Gestalt

$$\left\{\begin{aligned} &\left(x + \frac{1}{2}\right)\operatorname{arctg}\frac{y}{x} + y\log|z| - y\log\varrho e \\ &= 2n\pi + \Im\left(\omega(z,\varrho) - \omega(z)\right) = 2n\pi + O\left(\frac{1}{z}\right) \end{aligned}\right. \tag{20}$$

geschrieben werden. Es ist unmittelbar zu sehen, dass für jedes feste n die positive Achse eine Asymptote für α_n bildet; im entgegengesetzten Falle würde nämlich $y\log|z|$ für grosse $|z|$ das überwiegende Glied in der Gleichung (20) werden, die deshalb nicht erfüllt sein könnte.

In einem gegebenen Punkte z auf α_n ist die Tangente bestimmt durch

$$(21)\begin{cases} \dfrac{dy}{dx} = \dfrac{\operatorname{arctg}\dfrac{y}{x} - \dfrac{y}{2|z|^2} + \Im\left(\omega'(z) - \omega'(z,\varrho)\right)}{\log|z| + \dfrac{y}{2|z|^2} - \log\varrho + \Re\left(\omega'(z) - \omega'(z,\varrho)\right)} \\ \qquad \underline{\simeq} - \dfrac{\operatorname{arctg}\dfrac{y}{x}}{\log|z|}, \end{cases}$$

woraus hervorgeht, dass α_n für grosse $|z|$ monoton fällt. Für $n \to \infty$ rückt der Punkt auf der Kurve α_n, der dem Nullpunkte am nächsten liegt, ins Unendliche; dies hat zur Folge, dass für hinreichend grosse n jede Kurve α_n monoton fällt. Ausserdem schneidet sie die imaginäre Achse; denn für $x = 0$ gewinnt man aus (20)

$$y \log y - y \log \varrho e \; - 2 n \pi \; - \Im\left(\omega(iy,\varrho) - \omega(iy)\right) = 0,$$

worin die linke Seite für grosse y offenbar positiv ist und für $y = 0$ und hinreichend grosses n negativ.

Die gefundenen Eigenschaften für die Kurven γ und α_n lehren, dass α_n bei hinreichend grossem n die Kurve γ in genau einem Punkte schneidet. Nur eine endliche Anzahl Kurven α_n schneiden γ möglicherweise nicht oder in mehreren Punkten. Es sei m die kleinste positive ganze Zahl derart, dass α_n die Kurve γ für jedes $n \geqq m$ in genau einem Punkte schneidet, dann hat man für die n-te Nullstelle (gerechnet in der Lage auf γ)

$$(22) \qquad \log \Gamma(z) = \log P(z,\varrho) + 2(n+p)\,\pi i \quad (n \geqq m);$$

p ist eine gewisse von n unabhängige ganze Zahl. $\log \Gamma(z)$ und $\log P(z,\varrho)$ sind in Übereinstimmung mit den obigen Definitionen von $\operatorname{arc} \Gamma(z)$ und $\operatorname{arc} P(z,\varrho)$ gewählt.

Um jetzt die asymptotische Lage der n-ten Nullstelle z_n zu bestimmen, schreiben wir (22) in der Form

$$
(23)\begin{cases}
\left(z_n+\frac{1}{2}\right)\log z_n - z_n \log \varrho e \\
\quad = 2n\pi i + 2p\pi i - \varrho - \log\sqrt{2\pi} - \omega(z_n) + \omega(z_n,\varrho) \\
\quad = 2n\pi i + O(1).
\end{cases}
$$

In erster Annäherung ist

$$
(24)\qquad z_n \log z_n = 2n\pi i + O(z).
$$

Man sieht unmittelbar $\frac{z}{n} \to 0$ für $n \to \infty$; daraus folgt

$$
(25)\quad \log z_n + \log_2 z_n = \log(2n\pi i) + O\left(\frac{z_n}{n}\right) = \log(2n\pi i) + o(1)
$$

und weiter

$$
(26)\begin{cases}
\log_2 z_n = \log_2(2n\pi i) + O\left(\frac{\log_2 z_n}{\log(2n\pi i)}\right) \\
\quad = \log_2(2n\pi i) + o(1).
\end{cases}
$$

Wird dies in (25) eingesetzt, so erhält man

$$
(27)\qquad \log z_n = \log(2n\pi i) - \log_2(2n\pi i) + o(1),
$$

was auch

$$
(28)\qquad z_n = \frac{2n\pi i}{\log(2n\pi i)}\bigl(1 + o(1)\bigr)
$$

geschrieben werden kann. Dieses Ergebnis kann sogleich etwas verschärft werden; trägt man es nämlich in (25) und (26) ein, so findet man anstelle von (27)

$$
(27\,\mathrm{a})\quad \log z_n = \log(2n\pi i) - \log_2(2n\pi i) + O\left(\frac{\log_2 n}{\log n}\right)
$$

und daraus

$$
(28\,\mathrm{a})\qquad z_n = \frac{2n\pi i}{\log(2n\pi i)}\left(1 + O\left(\frac{\log_2 n}{\log n}\right)\right).
$$

Unter Berücksichtigung von $\log(2n\pi i) = \log(2n\pi) + \frac{\pi i}{2}$ bekommt man durch Trennung von Reellem und Imaginärem auf der rechten Seite von (28 a)

$$(29) \qquad z_n = \frac{\pi^2 n}{\log^2 n}(1+\varepsilon) + \frac{2n\pi i}{\log n}(1+\varepsilon'),$$

worin ε und ε' beide $O\left(\frac{\log_2 n}{\log n}\right)$ sind. Diese Formel — bis auf die Grössenordnung von ε und ε' — ist es gerade, die Jensen 1893 in der Kopenhagener Mathematischen Vereinigung und 1924 in der Dänischen Gesellschaft der Wissenschaften mitteilte.

Es ist nicht schwer einen beträchtlich genaueren asymptotischen Ausdruck für den Nullstellen anzugeben. Aus (27) geht hervor, dass (23) geschrieben werden kann

$$(30) \begin{cases} z_n \log \dfrac{z_n}{\varrho e} = 2n\pi i - \dfrac{1}{2}\log \dfrac{2n\pi i}{\log(2n\pi i)} + O(1) \\ \qquad = 2n\pi i - \dfrac{1}{2}\log \dfrac{n}{\log n} + O(1). \end{cases}$$

Wird

$$(31) \qquad \zeta_n = \frac{z_n}{\varrho e}, \; \xi_n = \frac{2n\pi i}{\varrho e} - \frac{1}{2\varrho e}\log \frac{n}{\log n}$$

gesetzt, so nimmt (30) die Gestalt

$$(30\text{ a}) \qquad \zeta_n \log \zeta_n = \xi_n + O(1)$$

an. Wir führen nun eine Funktion $\lambda(z)$ ein, die durch die Funktionalgleichung

$$(32) \qquad \lambda(z)\log\lambda(z) = z$$

definiert ist. Da $\lambda(z) \neq 0$ ausser für $z = 0$, hat $\lambda(z)$ dieselben Singularitäten wie die Funktion $\log \lambda(z)$, die in dieser Hinsicht von Painlevé[1] genau untersucht worden ist. Indem man Einzelheiten und Beweise aus seiner Arbeit entnehmen möge, soll hier nur das erwähnt werden, was wir im folgendem brauchen. $\lambda(z)$ ist eine unendlich viel-

[1] P. Boutroux, Équations différentielles du premier ordre, Note de P. Painlevé. Paris 1908, pp. 143—145, 157—158.

deutige Funktion; sie hat einen Zweig, der sich im Nullpunkte regulär verhält und dort den Wert 1 hat, während alle anderen Zweige daselbst einen Verzweigungspunkt unendlich hoher Ordnung aufweisen und den Wert 0 annehmen. Jeder Zweig hat einen Verzweigungspunkt zweiter Ordnung in $-\frac{1}{e}$. Durch Zerschneidung der Ebene längs der negativen reellen Achse werden also alle Zweige eindeutig gemacht.

Wir betrachten jetzt in (22) n als eine stetige Veränderliche, welche die positive Achse durchläuft. Dann sind z_n und damit auch ζ_n für hinreichend grosses n reguläre Funktionen von n. Die Untersuchung der Kurven γ und α_n ändert sich nämlich nicht wesentlich, wenn man n auch komplexe Werte gibt, wofern diese nur hinreichend nahe an der positiven rellen Achse liegen. Hieraus folgt, dass z_n eindeutig ist und daher keine algebraischen Singularitäten haben kann; unendlich kann die Funktion auch nicht werden, da α_n die Kurven γ immer in einem im Endlichem gelegenen Punkte schneidet.

Nach (30 a) entspricht jedem n ein gewisser Zweig von $\lambda(z)$ mit

$$\zeta_n = \lambda\left(\xi_n + O(1)\right); \tag{33}$$

da aber ζ_n regulär ist und $\xi_n + O(1)$ keinen singulären Punkt von $\lambda(z)$ umkreisen kann, wenn n hinreichend gross ist, so ist dieser Zweig von $\lambda(z)$ für alle hinreichend grossen n derselbe.

Indem $\lambda(z)$ von jetzt an diesen wohlbestimmten Zweig der mehrdeutigen Funktion bedeutet, wollen wir den Ausdruck auf der rechten Seite in (33) vereinfachen.

Nach dem Mittelwertsatze von Darboux ist

$$(34)\begin{cases} \lambda(z+x) = \lambda(z) + \dot{x}\lambda'(z) + \frac{\eta}{2}x^2\lambda''(z+\vartheta x), \\ |\eta| \leqq 1,\ 0 \leqq \vartheta \leqq 1. \end{cases}$$

Aus der Funktionalgleichung rechnet man

$$(35) \qquad \lambda'(z) = \frac{1}{\log\lambda(z)+1}$$

und

$$(36) \quad \lambda''(z) = -\frac{\lambda'(z)}{\lambda(z)}\cdot\frac{1}{(\log\lambda(z)+1)^2} = \frac{-1}{\lambda(z)\,(\log\lambda(z)+1)^3}$$

aus. Denkt man an die Herleitung von (27), so überzeugt man sich leicht, dass diese Formel in Wirklichkeit

$$(37) \qquad \log\lambda(z) = \log z - \log_2 z + o(1)$$

aussagt, während (28) auch als

$$(38) \qquad \lambda(z) = \frac{z}{\log z}\big(1+o(1)\big)$$

gefasst werden kann. Folglich ist

$$\lambda'(z) = \frac{1}{\log\dfrac{z}{\log z}} + O\left(\frac{1}{\log^2 z}\right)$$

und

$$\lambda''(z) = \frac{-1}{z\log^4 z}\big(1+o(1)\big).$$

Nehmen wir in (34)

$$(39) \qquad z = \frac{2n\pi i}{\varrho e},\ x = -\frac{1}{2\varrho e}\log\frac{n}{\log n} + O(1),$$

so ergibt sich also

$$\zeta_n = \lambda\big(\xi_n + O(1)\big) = \lambda\left(\frac{2n\pi i}{\varrho e}\right) - \frac{1}{2\varrho e} + O\left(\frac{1}{\log n}\right),$$

d. h.

$$(40) \qquad z_n = \varrho e\,\lambda\left(\frac{2n\pi i}{\varrho e}\right) - \frac{1}{2} + O\left(\frac{1}{\log n}\right).$$

Hiermit ist die n-te Nullstelle von $Q(z,\varrho)$ bis auf eine für $n \to \infty$ nullstrebige Grösse bestimmt, während sie in (29) nur bis auf ein Restglied festgelegt ist, das wie $\frac{n \log_2 n}{\log^2 n}$ gegen Unendlich konvergiert. Andererseit tritt jedoch in (40) die nicht ganz elementare Funktion $\lambda(z)$ auf.

In (38) steht, wie sich $\lambda(z)$ in erster Annäherung für grosse z verhält; dies kann nach (28 a) ersetzt werden durch

$$\lambda(z) = \frac{z}{\log z}\left(1 + O\left(\frac{\log_2 z}{\log z}\right)\right), \tag{41}$$

woraus man erhält

$$\begin{aligned} \log \lambda(z) &= \log \frac{z}{\log z} + O\left(\frac{\log_2 z}{\log z}\right) \\ &= \log \frac{z}{\log z}\left(1 + O\left(\frac{\log^2 z}{\log_2 z}\right)\right). \end{aligned} \tag{42}$$

Wird dies in die Funktionalgleichung (32) eingetragen, so entsteht

$$\lambda(z) = \frac{z}{\log \frac{z}{\log z}}\left(1 + O\left(\frac{\log_2 z}{\log^2 z}\right)\right). \tag{43}$$

Hieraus kann man für $\log \lambda(z)$ einen schärferen Ausdruck als (43) gewinnen, diesen dann in (32) einsetzen u. s. w. Schreiben wir

$$\begin{cases} \lambda_1(z) = \frac{z}{\log z}, \quad \lambda_2(z) = \frac{z}{\log \frac{z}{\log z}} = \frac{z}{\log \lambda_1(z)}, \; \ldots \\ \lambda_{n+1}(z) = \frac{z}{\log \lambda_n(z)}, \; \ldots, \end{cases} \tag{44}$$

so findet man durch Induktion leicht die allgemeine Formel

$$\lambda(z) = \lambda_\nu(z)\left(1 + O\left(\frac{\log_2 z}{\log^\nu z}\right)\right). \tag{45}$$

Durch Einsetzen dieses Ausdruckes in (40) erkennen wir, dass auch dann, wenn man sich an elementare Funktionen halten will, (40) mehr als (28 a) liefert.

4. Die Fakultätenreihe von Schlömilch für $Q(z, \varrho)$.

Aus der von Schlömilch[1] gegebenen Integraldarstellung

$$(46)\qquad Q(1-z, \varrho) = \frac{e^{-\varrho}\varrho^{1-z}}{\Gamma(z)} \int_0^\infty \frac{e^{-t}t^{z-1}}{\varrho+t}\,dt, \qquad \Re z > 0,$$

in der für ϱ lediglich negative reelle Werte ausgeschlossen sind, findet man auf bekannte Weise

$$(47)\qquad \Gamma(z)e^{\varrho}\varrho^{z-1}Q(1-z, \varrho) = \sum_{\nu=0}^{n} \frac{c_\nu}{\varrho(\varrho+1)\cdots(\varrho+\nu)} + R_n$$

mit

$$(48)\qquad c_\nu = (-1)^\nu \int_0^\infty e^{-t}t^z(t-1)\cdots(t-\nu+1)\,dt$$

und

$$(49)\qquad R_n = \frac{(-1)^{n+1}n!}{\varrho(\varrho+1)\cdots(\varrho+n)} \int_0^\infty \frac{e^{-t}t^z}{\varrho+t}\binom{t-1}{n}\,dt.$$

Bei $n \to \infty$ strebt R_n nach 0 für $\Re z > 0$, $\Re \varrho > 0$, und man erhält damit die von Schlömilch in seinem Kompendium der höheren Analysis II, 2. Aufl., p. 263—265 mit unvollständigem Beweis gegebene Fakultätenreihe; später haben andere und auch er selbst befriedigende Beweise gegeben, aber keiner von diesen ist so einfach wie der folgende.

Aus der Cauchyschen Ungleichung für die Koeffizienten einer Potenzreihe folgt für $t \geqq 1$

$$\left|\binom{t-1}{n}\right| \leqq \frac{(1+r)^{t-1}}{r^n},$$

[1] Vgl. z. B. Komp. d. höheren Analysis II, 2. Aufl., Braunschweig 1866, p. 263.

wenn r zwischen 0 und 1 liegt; bei $r \to 1$ bekommt man

$$\left|\binom{t-1}{n}\right| \leqq 2^{t-1} < 2^t. \tag{50}$$

Für $0 \leqq t \leqq 1$ gilt offenbar

$$(-1)^n \binom{t-1}{n} = \left(1-\frac{t}{1}\right)\left(1-\frac{t}{2}\right)\cdots\left(1-\frac{t}{n}\right) \leqq 1 \leqq 2^t.$$

Mit $\Re z = x$ besteht also folgende Ungleichung

$$|R_n| \leqq \left|\frac{\Gamma(\varrho)\,\Gamma(n+1)}{\Gamma(\varrho+n+1)}\right| \cdot \int_0^\infty \frac{e^{-t}\, t^x\, 2^t}{|\varrho+t|}\, dt < C \cdot n^{-\Re \varrho},$$

wobei C eine von n unabhängige Konstante ist. Hieraus entfliesst jedoch unmittelbar $R_n \to 0$ für $n \to \infty$, wofern $\Re z > 0$, $\Re \varrho > 0$.

5. Reihen für die unvollständigen Gammafunktionen.

Differenziert man die Funktion $e^\sigma P(z, \varrho+\sigma)$ nach σ, so findet man mit Hilfe der Differenzengleichung der P-Funktion

$$\left\{\begin{aligned} \frac{\partial\left(e^\sigma P(z,\varrho+\sigma)\right)}{\partial\sigma} &= e^\sigma P(z,\varrho+\sigma) + e^{-\varrho}(\varrho+\sigma)^{z-1} \\ &= (z-1)\, e^\sigma P(z-1,\, \varrho+\sigma) \end{aligned}\right. \tag{51}$$

und darnach durch Induktion

$$\frac{\partial^n\left(e^\sigma P(z,\varrho+\sigma)\right)}{n!\, \partial\sigma^n} = \binom{z-1}{n} e^\sigma P(z-n,\, \varrho+\sigma). \tag{52}$$

Ist z keine positive oder negative ganze Zahl einschliesslich 0, so haben wir also

$$(53)\qquad P(z,\varrho+\sigma) = e^{-\sigma}\sum_{n=0}^{\infty}\binom{z-1}{n}P(z-n,\varrho)\,\sigma^n$$

für $|\sigma| < |\varrho|$, indem $\sigma = \varrho$ der einzige singuläre Punkt von $P(z,\varrho+\sigma)$ als Funktion von σ im Endlichen ist. Für positiv ganzzahliges z gilt die Reihe sogar für alle σ.

Eine ähnliche Betrachtung liefert für die Q-Funktion

$$(54)\qquad Q(z,\varrho+\sigma) = e^{-\sigma}\sum_{n=0}^{\infty}\binom{z-1}{n}Q(z-n,\varrho)\,\sigma^n,$$

und zwar bei $|\sigma|<|\varrho|$ für jedes z; ist z ganzzahlig (positiv, negativ oder 0), so konvergiert die Reihe sogar für alle σ.

Führt man in (54) die Identität

$$(55)\qquad \frac{Q(z-n,\varrho)}{\Gamma(z-n)} = \frac{Q(z,\varrho)}{\Gamma(z)} - e^{-\varrho}\sum_{s=0}^{n-1}\frac{\varrho^{z+s-n}}{\Gamma(z+s-n-1)}$$

ein, so gewinnt man nach einer einfachen Umrechnung

$$(56)\qquad Q(z,\varrho+\sigma) = Q(z,\varrho) - e^{-\varrho-\sigma}\varrho^z\sum_{n=1}^{\infty}\frac{\sigma^n}{n!}\sum_{s=0}^{n-1}\frac{(z-1)\cdots(z-s)}{\varrho^{s+1}},$$

woraus man durch Umkehrung der Summationsordnung die Formel

$$(57)\qquad Q(z,\varrho+\sigma) = Q(z,\varrho) - e^{-\varrho}\varrho^z\sum_{s=0}^{\infty}\binom{z-1}{s}\frac{P(s+1,\sigma)}{\varrho^{s+1}}$$

herleiten kann, die man N. Nielsen[1] verdankt.

Für $z = 0$ stossen wir nach (54) und (56) auf die folgenden bemerkenswerten Entwicklungen für den Integrallogarithmus $li\,e^{-x} = -Q(0,x)$

[1] Integrallogarithmus, p. 84.

$$
(58)\quad \begin{cases} li\, e^{-x-y} = e^{-y} \sum\limits_{n=0}^{\infty} (-1)^{n-1} Q(-n,x)\, y^{n} \\ \qquad\quad = e^{-x-y} \sum\limits_{n=0}^{\infty} \dfrac{y^{n}}{n!} \sum\limits_{s=0}^{n-1} \dfrac{(-1)^{s} s!}{x^{s+1}}, \end{cases}
$$

während (57) die von Bessel[1] herrührende Reihe

$$
(59)\quad li\, e^{-x-y} - li\, e^{-x} = e^{-x} \sum_{s=0}^{\infty} \frac{(-1)^{s}}{x^{s+1}} P(s+1, y)
$$

liefert.

[1] N. Nielsen l. c. p. 85.

Færdig fra Trykkeriet d. 30. Juni 1927.

Det Kgl. Danske Videnskabernes Selskab.
Mathematisk-fysiske Meddelelser. **VIII,** 3.

THERMAL MOLECULAR PRESSURE IN TUBES

BY

MARTIN KNUDSEN

KØBENHAVN
HOVEDKOMMISSIONÆR: ANDR. FRED. HØST & SØN, KGL. HOF-BOGHANDEL
BIANCO LUNOS BOGTRYKKERI
1927

Pris: Kr. 1,50.

Det Kgl. Danske Videnskabernes Selskabs videnskabelige Meddelelser udkommer fra 1917 indtil videre i følgende Rækker:

Historisk-filologiske Meddelelser,
Filosofiske Meddelelser,
Mathematisk-fysiske Meddelelser,
Biologiske Meddelelser.

Hele Bind af disse Rækker sælges 25 pCt. billigere end Summen af Bogladepriserne for de enkelte Hefter.

Selskabets Hovedkommissionær er *Andr. Fred. Høst & Søn*, Kgl. Hof-Boghandel, København.

Det Kgl. Danske Videnskabernes Selskab.
Mathematisk-fysiske Meddelelser. **VIII,** 3.

THERMAL MOLECULAR PRESSURE IN TUBES

BY

MARTIN KNUDSEN

KØBENHAVN
HOVEDKOMMISSIONÆR: ANDR. FRED. HØST & SØN, KGL. HOF-BOGHANDEL
BIANCO LUNOS BOGTRYKKERI
1927

Introduction.

If a gas is enclosed in two vessels communicating with each other by a tube, the gas will, as known, be in equilibrium if its pressure is of the same magnitude throughout. True, this condition of equilibrium only holds good if we disregard the differences in pressure produced by the effect of gravity. This we will do in the following. If the two vessels are given a different temperature, the condition of equilibrium will still hold good in many cases, even if the temperature is varied through the communication tube, and this fact is made use of, e. g. in the gas thermometer, it being a well-known fact that the gas in the manometer may have a temperature quite different from the gas in the thermometer bulb.

If, however, we employ the word equilibrium as a term for the state in which the amounts of the masses of gas found in the two vessels do not change any more if the temperature of the vessels remains unchanged, the condition of equilibrium mentioned may in certain cases become quite wrong. This was already shown by O. Reynolds[1], who by means of the kinetic theory derives the equation $\frac{p_1}{p_2} = \left(\frac{T_1}{T_2}\right)^{\frac{1}{2}}$ as valid in the case of p_1 and T_1 being pressure and temperature respectively on one side of a porous plate, while p_2 and T_2 are the corresponding

[1] O. Reynolds, Phil. Trans. p. 727, London 1879.

quantities on the other side of the plate. Putting $p_1 = p_2$ and T_1 different from T_2 the gas will not be in equilibrium, there will be a flow of gas through the plate from the cold to the warm side. This flow of gas which Reynolds called "thermical transpiration" was demonstrated experimentally by Reynolds himself in experiments with plates of gypsum and meerschaum. The temperatures T_1 and T_2 were not, however, measured directly, so that Reynolds did not obtain a numerical confirmation of the equation given above. Such a confirmation I have achieved by means of a glass tube in which a magnesia plug had been firmly fixed, and the thermical transpiration was demonstrated in the following way.

I used a vessel holding from $^1/_2$ to 1 litre, and which was made of porous porcelain (a filtration bulb). The neck was closed with a rubber-stopper through which was passed a glass tube ending under a water surface. The gas in the bulb was heated by an electric current sent through a coil placed inside the bulb. The walls of the bulb being thus heated from the inside, and continually cooled on the outer side, a fall of temperature will take place in the porous wall, and this will cause gas to be sucked through the wall into the bulb. Gas bubbles will then rise through the surface of the water so that, in the course of a few minutes, more gas can be collected than the porous vessel holds. It will be noted too that the flow of gas will continue with constant velocity as long as the temperatures on the inner and outer sides of the walls are kept constant, and that the velocity increases when the heating current is increased or the cooling velocity of the bulb is augmented by blowing cold air on to it.

The condition of equilibrium stated by Reynolds is, however, only valid when the effect of the collision of the molecules with each other as compared with the number of the impacts with the tube walls may be disregarded, or, in other words, when the cross-section dimensions of the tube are negligible compared with the mean free path λ of the gas molecules. For a cylindrical tube with the radius r the quantity $\frac{r}{\lambda}$ must thus be negligible compared with 1 if Reynolds' formula is to hold good.

If λ is negligible compared with r, the condition of equilibrium will, as known, be that the pressure is the same throughout the whole system whatever is the distribution of the temperature.

The case when λ is small but not negligible compared with r has been theoretically dealt with by MAXWELL[1] who made use of the results of KUNDT and WARBURG's experiments on the slipping of the gases. By a consideration which I have formerly[2] explained I have arrived at a relation which formally agrees perfectly with Maxwell's. The constants found by me deviate somewhat from those found by Maxwell. A series of experiments previously made by me shows that the formulas in question are formally right, but that the constants found by experiments are again somewhat different from the theoretical ones, which is easily explained.

The theoretically found formula corroborated by experiment may be written as follows:

$$p_1^2 - p_2^2 = c\,(T_1^2 - T_2^2),$$

where p_1 and T_1 are the pressure and absolute temperature in one vessel, p_2 and T_2 the corresponding quantities in

[1] J. Clerk Maxwell, Phil. Trans. p. 231, London 1879.

[2] Martin Knudsen, Ann. d. Phys. Bd. 31, p. 214, 1910.

the other, while c is dependent on the radius of the tube and the mean free path λ_1 of the gas at a pressure of 1 bar. It may be expected too that c will in some degree be dependent on the temperature. For the cases when r is either negligible or very large compared with λ we have neither theoretical nor experimental investigations of the relation between p and T, yet a knowledge of this relation may be of great importance e. g. when the gas thermometer is to be used to measure the lowest temperatures that can now be produced.

In the following I shall give an account of a series of experiments performed by me for the purpose of learning more of this relation.

Experiments with the Gas Thermometer.

To solve the problem mentioned above I have tried using a gas thermometer and carrying out measurements at constant pressure and varying volume. The glass apparatus is sketched in fig. 1. The volume V of the thermometer tube was measured, and likewise the radius r of the communication tube to the hot-wire manometer M,

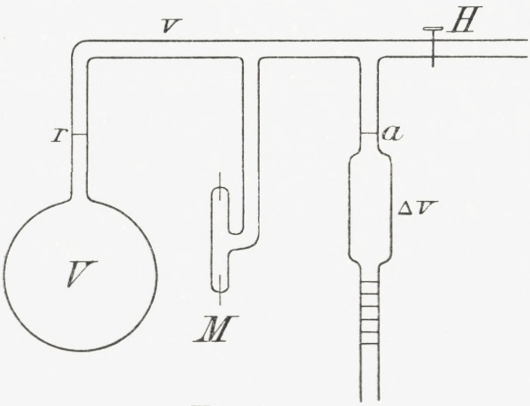


Fig 1

and the volume $\varDelta v$ of a pipette continued at the bottom in a graduated tube closed with mercury. H is a stop-cock through which hydrogen is introduced into the apparatus until the desired pressure is obtained.

Let us suppose that the whole apparatus has the absolute temperature T_1 and that the pipette is filled with mercury up to the mark a. There will then be the same pressure p_1 throughout the apparatus. Now the vessel V is heated to the temperature T_2 while the remaining part of the apparatus is kept at the temperature T_1. The mercury is made to sink in the pipette until the manometer again shows the initial pressure p_1. If the volume of the manometer and communication tube be designated v, and the new pressure in the heated vessel p_2, the expression for the constancy of the mass of gas gives that

$$\frac{p_1 V}{T_1} + \frac{p_1 v}{T_1} = \frac{p_2 V}{T_2} + \frac{p_1 v}{T_1} + \frac{p_1 \Delta v}{T_1}$$

from which we get that

$$\frac{p_2}{p_1} = \frac{T_2}{T_1}\left(1 - \frac{\Delta v}{V}\right).$$

Here p_2 and p_1 designate the pressures in the communication tube with the radius r in the state of equilibrium in those places where the temperatures are T_2 and T_1 respectively, and p_1 and all quantities on the right side of the sign of equation being measured, the equation gives the sought relation between p_2 and p_1.

We know that for large values of the pressure, that is to say abt. 1 cm. mercury pressure, we ought to find $p_2 = p_1$, and that for small values of the pressure, that is to say small values of $\frac{r}{\lambda}$, we ought to find small deviations from the equation $\frac{p_2}{p_1} = \sqrt{\frac{T_2}{T_1}}$. The first of these requirements was fairly well satisfied, which showed that the errors of observation were small. But from the second equation such great deviations appeared that the method must be considered unsuitable.

The reason for this is that in each measurement great accuracy in the determination both of temperatures and volumes must be demanded, but of decisive importance are the adsorption phenomena that manifest themselves vigorously at lower pressures. This source of error I have not been able to eliminate, and my experiments seem to me to have shown that the gas thermometer is not suited for temperature measurements when the gas pressure in the vessel must necessarily be very low. Heating of the thermometer bulb and the use of Geräte glass somewhat reduced the error, though far from sufficiently.

Plan of the Experimental Investigation.

The method used for the final series of measurements was the following.

M (fig. 2) is a hot-wire manometer as previously described[1]. By a series of glass tubes of unequal widths the

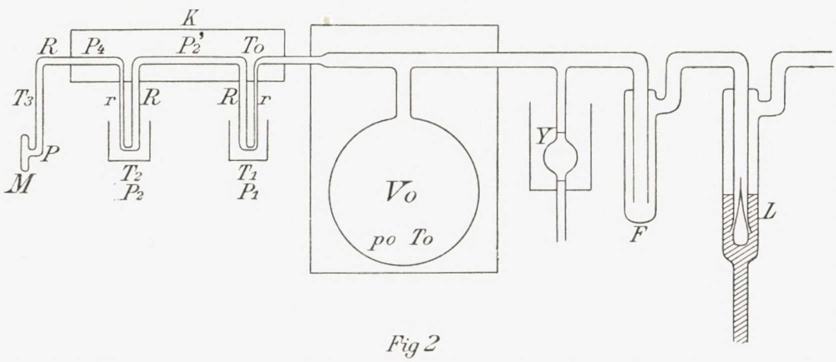


Fig 2

manometer is connected with a large glass vessel the volume of which has been measured to be V_0. The widest tubes

[1] Martin Knudsen: Det Kgl. Danske Videnskabernes Selskab. Mathematisk-fysiske Meddelelser VII, 15, 1927.

in the system have a radius of abt. 1 mm. The two pieces of tubing marked R in the figure have equal radii:

$$R = 0.099655 \text{ cm.}$$

and the two pieces of tubing marked r have equal radii, viz.

$$r = 0.026845 \text{ cm.}$$

These four pieces of tubing were as circularly cylindrical as they could be found in a large collection of tubes. Their upper joints, which during the experiments should have the same temperature, were surrounded by a rather large water bath, K, contained in a rectangular copper box. The vessel with the volume V_0, abt. 1.2 litres, was likewise placed in a water bath the walls of which were well insulating for heat. Y is a pipette, the gauge vessel, placed in a water bath. It may be filled with and emptied of mercury through the tube at the bottom, and it serves to calibrate the hot-wire manometer, its volume between the marks being measured to 8.8746 cm^3. F is a trap which is kept in liquid air during the experiments, and L is a mercury seal with a ground glass float which can shut off the apparatus from the pipette system serving to introduce a gas of known pressure into the apparatus. The glass float very effectually prevents the penetration of mercury vapours into the apparatus, and the tube below the small gauge vessel is very narrow, so that by this way too the entrance of mercury vapours will be negligible.

During the first measurement on the potentiometer the two joints between the tubes R and r are placed each in a separate bath, the temperatures of which are measured or known. Thereupon the two baths are interchanged and a second measurement is made. From the two measurements the ratio between the two pressures may be found.

Let fig. 3 be a circularly cylindrical tube closed at both ends. Let the radius of the tube be r and let one end of it have the absolute temperature T_1, the other T_2. In the state of equilibrium the pressures of the gas contained in the tube will be different at the two ends. If they be designated p_1 and p_2, as indicated in the figure, $p_1 - p_2$, i. e. the quantity which I term the thermal molecular pressure will differ from 0 when $T_1 - T_2$ does so.

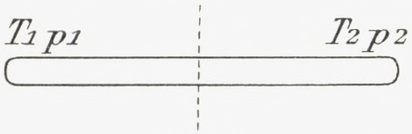


Fig 3

We may know beforehand that the thermal molecular pressure must be a very complicated function of the temperature, the pressure, and the radius of the tube, therefore I have considered it advisable to make a series of measurements at such small temperature differences that $p_1 - p_2$ may with sufficient approximation be put proportional to $T_1 - T_2$ at all pressures. Hence for the tube with the radius r we put $p_1 - p_2 = f(T_1 - T_2)$, and for the tube with the radius R we put $p_1 - p_2 = F(T_1 - T_2)$, on the assumption that the mean pressures in the two tubes may with approximation be put equal.

With the designations of pressures and temperatures given in fig. 2 we may then note the following set of equations:

$$
\begin{aligned}
P_1 - p_0 &= f(T_1 - T_0) \\
P_2' - P_1 &= F(T_0 - T_1) \\
P_2 - P_2' &= F(T_2 - T_0) \\
P_4 - P_2 &= f(T_0 - T_2) \\
P - P_4 &= F(T_3 - T_0)
\end{aligned}
$$

If now we interchange the two baths whose temperatures were designated T_1 and T_2, the pressures will change throughout except in the large vessel V_0, for compared to its volume the rest of the volumes are negligible. If the new pressures are designated p with the same indices which P had before, we shall get the following set of equations:

$$\begin{aligned} p_1 - p_0 &= f(T_2 - T_0) \\ p'_2 - p_1 &= F(T_0 - T_2) \\ p_2 - p'_2 &= F(T_1 - T_0) \\ p_4 - p_2 &= f(T_0 - T_1) \\ p - p_4 &= F(T_3 - T_0) \end{aligned}$$

If the two sets of equations are added separately, we get:

$$\begin{aligned} P - p_0 &= F(T_2 - T_1) + f(T_1 - T_2) + F(T_3 - T_0) \\ p - p_0 &= F(T_1 - T_2) + f(T_2 - T_1) + F(T_3 - T_0), \end{aligned}$$

from which by subtraction

$$\frac{1}{2}(P-p) = (f-F)(T_1 - T_2)$$

and hence, $\frac{1}{2}(P-p)$ being designated by Δp and $T_1 - T_2$ by ΔT,

$$f - F = \frac{\Delta p}{\Delta T}.$$

The quantity f may be characterised as being equal to $\frac{dp}{dT}$ in the tube with the radius r, while F is equal to $\frac{dp}{dT}$ in the tube with radius R, in which the pressure is very nearly equal to that in the first-mentioned tube.

In the series of experiments in question in which F and f were determined, the temperatures $(T_0 - 273^\circ)$ of the water baths were kept at 20° centigrade. T_1 was kept at the temperature of melting ice, the bath consisting of scraped ice in a Dewar vessel. T_2 was kept 40° higher by a water bath which was continually stirred by a current of air, while the loss of heat to the surroundings was compensated by electric heating. The temperature of this bath was read on a mercury thermometer graded $^1/_{10}$ and bent in an angle so that it could be placed under the copper box *K*. The two baths were both placed on a revolving stand, so that they could be easily and quickly interchanged.

The temperature of the manometer was throughout that of melting ice, but as will be seen, neither this temperature nor T_0, that of the water baths, enters into the equation by which $f-F$ is determined. Besides this the method further presents the advantage that the volumes changing temperatures are not altered by the interchange, and thus changes in pressure caused by Gay-Lussac's expansion are avoided. Finally, what is most essential is that the areas of the glass surfaces subjected to the changes in temperature is here reduced to a minimum of abt. 8 cm^2. For it is only in the comparatively narrow and short pieces of tubing that the temperature changes cause the harmful adsorption phenomena.

Example of a Measurement.

All measurements were, in all essentials, made in the same manner, so we shall here only describe one chosen at random from the entire material.

The whole apparatus with pipette system and mercury manometer was exhausted by means of a mercury diffusion pump. Between this and the apparatus was inserted a trap cooled in liquid air. When after pumping for a short time the apparatus was almost devoid of air, electric heating coils were placed round each of the bends of the tubes under the copper box. Thus these tubes were baked out, being kept heated to a temperature of abt. 330° centigrade for 10—12 hours while the pump was at work now and then. This heating was intended to diminish the harmful adsorption effects, and was to some extent successful, though unfortunately not absolutely. The tubes used were of good Thüring glass, an experiment with Geräte glass and Pyrex glass having shown that these kinds did not present such advantages that it would be profitable to use them.

When the heating was finished and also the exhaustion, the result of which was followed on the hot wire manometer and in some cases on an absolute manometer, hydrogen was brought into the apparatus. This hydrogen was taken from a steel receptacle and had been tested and proved sufficiently pure. It was dried by passing through a trap cooled in liquid air. Its influx could be so accurately regulated that the desired pressure in the mercury manometer could be produced with great approximation. In the experiment to be described here the reading on the mercury manometer was 21.514 cm. mercury pressure at 20°. By means of the pipette system a fraction hereof was introduced into the apparatus, liquid air being placed round the trap F (fig. 2). From the known volumes of the pipette system and the apparatus the pressure in the apparatus was found to be 718.6 bar. The mercury seal L (fig. 2) was closed, and a water bath of 20° centigrade placed round both bends of the tubes, r and R.

The resistance of the manometer wire was 743.3 ohms at 0°. Wheatstone's bridge was adjusted so that there would be no current in its galvanometer when the wire was heated so much that its resistance would be 900 ohms. The mean temperature of the wire is then abt. 60° centigrade. The potentiometer whose total resistance was 11000 ohms was shunted with a resistance of 310 ohms in order to obtain a suitably large reading (between 9000 and 10000 ohms). This is of importance for the interpolation which is made by measuring the deflection of the galvanometer mirror caused by a change of 1 ohm in the potentiometer.

The table below shows a column indicating the hour when the measurements were made. The next two columns show the temperatures t_1 and t_2 in degrees centigrade of the baths surrounding the bends of the tubes, and under Comp., finally, are given the readings on the potentiometer.

Time	t_1	t_2	Comp.
11^h45	20°	20°	9656,97
57	40°,63	0°	9745,47
12^h10	0°	40°,40	9568,88
23	40°,58	0°	9745,10
35	20°	20°	9656,32
47	0°	40°,54	9568,60
59	40°,40	0°	9744,32
1^h11	0°	40°,20	9568,92
24	20°	20°	9655,14

From this series it will be seen that the measurements were made at very nearly equal intervals of time. This was done on account of the adsorption effects. Between the measurements are interposed some with both bends of the tubes at the same temperature 20° so as to follow the slow changes of pressure in the apparatus. It was as far as possible attempted to do away with these by grouping the measurements three and three together as shown in the table.

By taking means the table is reduced to the following two

t_1	t_2	Comp.	Δ Comp. = Comp.$_{t_1 t_2}$ — Comp.$_{20, 20}$	$\frac{\Delta \text{ Comp.}}{t_1 - t_2}$	Mean values $\frac{\Delta \text{ Comp.}}{t_1 - t_2}$
20°	20°	9656,65			
40°,61	0°	9745,29	+ 88,64	2,1827	2,1776
0°	40°,40	9568,88	— 87,77	2,1725	
20°	20°	9655,73			
40°,40	0°	9744,32	+ 88,59	2,1928	2,1736
0°	40°,37	9568,76	— 86,97	2,1543	

The two measurements made independently of each other, and which were almost independent of the measurements at 20°, 20°, have thus given the following values for $\frac{\Delta \text{ Comp.}}{t_1 - t_2}$, 2.1776 and 2.1736, hence the mean value, $t_1 - t_2$ being designated by ΔT.

$$\frac{\Delta \text{ Comp.}}{\Delta T} = 2.1756 = \text{Measurement.}$$

In connection with these measurements gaugings were made, as a rule two, one before and the other after the measurements at different temperatures had been made. During the gauging measurements the bends of the tubes were kept at the same temperature abt. 20°, both bends being placed in a single large water bath. In connection with the above-mentioned measurements the following gauge-measurements were made, partly with the gauge vessel Y empty (fig. 2) partly when it was filled with mercury

	Comp.	Comp. (empty-full)	Mean d Comp.
Y empty	9655,14		
Y full	9622,79	32,28	
Y empty	9655,00		
			32,30 = gauge
Y empty	9650,49		
Y full	9618,04	32,32	
Y empty	9650,22		

The compensation change thus produced, viz. d Comp. = 32.30 = gauge is due to the fact that the volume $V = 1219.04 + 8.8746$ cm^3 of the gas content of the whole apparatus has been diminished by the volume $\Delta v = 8.8746$ cm^3 of the gauge vessel. This reduction of the volume produces an increase of the pressure dp which, when the pressure in the apparatus is designated p, is determined by

$$\frac{dp}{p} = \frac{8.8746}{1219.04} = 0.007280.$$

Since here, where the relative changes are small, we can put the changes in pressure proportional to the changes in compensation we get

$$\frac{\Delta p}{dp} = \frac{\Delta \text{Comp.}}{d\,\text{Comp.}} = \Delta T \cdot \frac{\text{measurement}}{\text{gauge}}$$

and consequently

$$10^4 \frac{\Delta p}{p \Delta T} = 72{,}80 \cdot \frac{\text{measurement}}{\text{gauge}}.$$

In the example here considered, where measurement = 2.1756 and gauge = 32.30 we thus get

$$10^4 \frac{\Delta p}{p \Delta T} = 4.904.$$

Such a determination was made at eleven different pressures p, which very nearly formed a geometrical progression, 19082 bar being the highest pressure and 27.15 the lowest, while the quotient was equal to the square root of the ratio between the radii of the two tubes used in the apparatus. Judging from repetitions following immediately after one another the uncertainty of the values found only amounts to a few per mille, but as we shall see later, the real uncertainty is much greater, especially at low pressures.

Results of the Measurements and their provisional Treatment.

The measurements made with the temperatures 0 and 40 degrees centigrade at various pressures p' gave the following values for $10^4 \frac{\Delta p'}{p' \Delta T}$

number n =	0	1	2	3	4	5
p' Bar =	19082	9920	5142	2705	1388,9	718,6
$10^4 \frac{\Delta p'}{p' \Delta T}$ =	0,0790	0,2528	0,7456	1,648	3,314	4,904

number n =	6	7	8	9	10
p' Bar =	373,5	193,2	100,76	52,22	27,15
$10^4 \frac{\Delta p'}{p' \Delta T}$ =	5,424	4,719	3,481	2,542	1,786

The measured values of p' agree very closely with the pressures calculated from the formula $p = p_0 \left(\frac{r}{R}\right)^{\frac{n}{2}}$ so that interpolations may be made with great certainty which give $10^4 \frac{\Delta p}{p \Delta T}$ for the pressures p given in the geometrical progression. The interpolated values are given in the following table, together with some other quantities to be mentioned later on.

p Bar	$10^4 \frac{\Delta p}{p \Delta T}$	$10^4 \frac{f}{p}$	$\frac{1}{u} = \frac{2Tf}{p}$
19082	0,0790	0,0857	0,00502
9904	0,2537	0,2774	0,01626
5140	0,7464	0,8321	0,04878
2668	1,678	1,955	0,1146
1385	3,323	4,155	0,2436
718,7	4,904	6,859	0,4021
373,0	5,422	9,577	0,5614
193,6	4,724	11,583	0,6790
100,5	3,474	13,051	0,7650
52,15	2,542	14,125	0,8280
27,07	1,782	14,833	0,8695

For the further treatment of the observation series we remind the reader that for a tube with radius r, in which there is the pressure gradient dp originating from the temperature gradient dT, we have put $\frac{dp}{dT} = f$, while for the wide tube with radius R we put $\frac{dp}{dT} = F$.

Further it was proved that

$$f - F = \frac{\Delta p}{\Delta T}$$

or

$$10^4 \frac{f}{p} - 10^4 \frac{F}{p} = 10^4 \frac{\Delta p}{p \Delta T}$$

where Δp and ΔT are just the quantities that with the same designation enter into the tabulated values for $10^4 \frac{\Delta p}{p \Delta T}$. Hence we may consider $f - F$ as the quantity observed at different pressures. Our problem is now to find both f and F from the differences observed, which may be done by theoretical considerations that have been found tenable by previous experiments.

For a tube with the radius r I have previously[1] found an expression for f or $\frac{dp}{dT}$ which by inserting the constants gives

$$f = \frac{0{,}02996}{r + 0{,}01191\, r^2 p}.$$

If r in this expression be replaced by R, we get F, and from the values thus found we calculate $10^4 \frac{f}{p}$ and $10^4 \frac{F}{p}$ and from these again $10^4 \frac{f}{p} - 10^4 \frac{F}{p}$, which is a provisionally found value for $10^4 \frac{\varDelta p}{p \varDelta T}$. If we compare the values of $10^4 \frac{\varDelta p}{p \varDelta T}$ thus calculated with those observed, we find, in the case of the three greatest pressures, deviations not amounting to more than 5 p. c. of the values. For a pressure of 19082 bar $10^4 \frac{F}{p}$ is calculated to be 0.0067 and for a pressure of 9904 bar $10^4 \frac{F}{p}$ is calculated to be 0.0237. Having $10^4 \frac{f}{p} - 10^4 \frac{F}{p} = 10^4 \frac{\varDelta p}{p \varDelta T}$ we find by adding the calculated values for $10^4 \frac{F}{p}$ to those observed for $10^4 \frac{\varDelta p}{p \varDelta T}$ that for the pressures

$$19082 \text{ bar we get } 10^4 \frac{f}{p} = 0.0857$$

and for

$$9904 \text{ bar we get } 10^4 \frac{f}{p} = 0.2774.$$

These values have been tabulated and are employed to calculate the rest of the values given for $10^4 \frac{f}{p}$.

In this calculation we avail ourselves of the fact that $\frac{f}{p}$ at the pressure 19082 is equal to $\frac{F}{p}$ at the pressure $19082 \cdot \frac{r}{R}$, that is to say, at the pressure 5140, for which we have an observation of $10^4 \frac{f}{p} - 10^4 \frac{F}{p}$. Hence to this

[1] Martin Knudsen, Ann. d. Phys. Bd. 33, p. 1444, 1910.

observation, the tabulated 0.7464, we need only add 0.0857 to find $10^4 \frac{f}{p}$ valid for the pressure 5140. This proceeding is continued throughout the table so that for each pressure we have the corresponding value for $10^4 \frac{f}{p}$. In view of the following calculations and considerations the next column in the table is formed, $\frac{1}{u} = 2\,T \cdot \frac{f}{p}$ where T is the mean temperature 293.1° at which the measurements were made. The quantities $\frac{1}{u}$ thus calculated indicate how great is the effect of the thermal molecular pressure in a tube of the given radius r.

The formula $f = \frac{0,02996}{r + 0,01191\,r^2 p}$ warrants this proceeding in the case of the large pressures. Hence for a tube with radius r and pressure p and another with radius R and pressure P we have $\frac{f}{p} = \frac{0,02996}{rp + 0,01191\,(rp)^2}$ and $\frac{F}{P} = \frac{0,02996}{RP + 0,01191\,(RP)^2}$. From this we see that $\frac{f}{p} = \frac{F}{P}$ when $rp = RP$ which was just what was made use of in the calculation.

That this proceeding holds good for all pressures may be seen by the following consideration. We will take it for granted that when a closed circularly cylindrical tube, containing a gas at the pressure p, has different temperatures at the two ends, the difference in pressure found between the two ends in the state of equilibrium will be independent of the way in which the temperature varies from end to end. From this it follows that when an increase of temperature dT is found on the length dl of the tube, this will involve an increase of pressure dp, which is independent of dl, but determined by other quantities. What these are may be determined by considering the quantity $\frac{\frac{dp}{p}}{\frac{dT}{T}}$. This quantity is dependent both on the dimensions of the tube and on the physical properties of the gas. Since the dimensions of the tube only comprise its length and radius and since the length, as we just stated, does not influence $\frac{dp}{dT}$, the radius r will be the only dimension of the tube of which $\frac{dp}{dT}$ may be a function. Hence we may put $\frac{\frac{dp}{p}}{\frac{dT}{T}} = \Phi$ (r, the physical properties of the gas). Since $\frac{\frac{dp}{p}}{\frac{dT}{T}}$ is a pure number the physical properties of the gas that can be taken into account can only be a length L which enters into the equation so that we can put $\frac{\frac{dp}{p}}{\frac{dT}{T}} = \Phi\left(\frac{r}{L}\right)$ or $2\,T\frac{f}{p} = 2\,\Phi\left(\frac{r}{L}\right)$. If we have a series of tubes of different radii and assume that they all of them have the same temperature and all contain hydrogen, their hydrogen content will be determined entirely by the dimensions of the tubes and the mean free path λ of the hydrogen. The quantity L in the above-mentioned formula may therefore be put identical with λ. If now we remember that we have $p\lambda = \lambda_1$, we get $\frac{r}{L} = \frac{rp}{\lambda_1}$ and hence

$$\frac{f}{p} = \frac{1}{T}\Phi\left(\frac{rp}{\lambda_1}\right).$$

Here λ_1 is the mean free path of the hydrogen at the pressure 1 bar and the temperature T, that is to say, independent of r and p, so the expression shows that when

r and p vary in such a way that their product is kept constant $\frac{f}{p}$ will remain unaltered at constant temperature. As has been mentioned, the calculations made have been based on this rule, and it is important in giving expressions for the thermal molecular pressure to keep to the formula

$$\frac{dp}{dT} = \frac{p}{T}\Phi\left(\frac{r}{\lambda}\right).$$

The tabulated quantity $\frac{1}{u} = 2T\frac{f}{p} = 2T\frac{1}{p}\frac{dp}{dT}$ is thus a function of $\frac{r}{\lambda}$ alone, and our next problem will be to find an expression for this functional dependency. For the solution of this problem it must be remembered that the values found in the table for $\frac{1}{u}$ were calculated successively, so that any inaccuracy in one of the measurements at high pressures will make its influence felt at all the lower pressures. For these, therefore, all the errors will be added up. This unfortunate circumstance may, however, be entirely avoided, as will be shown in the following.

Theoretical Considerations in the Formation of a Formula for the Thermal Molecular Pressure.

Reynolds' formula, which was given in the introduction, may be arrived at by the following simple kinetic consideration. Let N be the number of gas molecules in each cm.^{3}, m the mass of each molecule, and c the molecular velocity. $\bar{c}$ denotes the mean value of the molecular velocities, and $\overline{c^2}$ the mean value of the squares of the velocities. n denotes the number of impacts that is to say, the number of molecules which in each second passes through a cm.^{2} coming from one side of it. If the pressure of the gas be p, its absolute temperature T, and its molecular weight M, the kinetic theory in conjunction with the simple equation of state will give the following fundamental expressions

$$p = \frac{1}{3} Nm\,\overline{c^2} \quad \text{and} \quad n = \frac{1}{4} N\,\overline{c}$$

and it follows from Maxwell's law of the distribution of velocities that

$$\frac{1}{3}\overline{c^2} = \frac{\pi}{8}\left(\overline{c}\right)^2 \qquad \overline{c} = 14550\sqrt{\frac{T}{M}}.$$

If we can disregard the effect of the mutual impacts of the molecules in the places where the temperature varies from place to place, we have, when the state is to be a state of equilibrium, that the number of impacts n must have the same value everywhere. For let us suppose that the two vessels have the absolute temperatures T_1 and T_2, and that the total temperature difference $T_1 - T_2$ is found in a single definite cross-section of the communication tube. Then, for the molecules coming from one side towards this cross-section we have the number of impacts $n_1 = \frac{1}{4} N_1\,\overline{c_1}$ and for those coming from the other side we have $n_2 = \frac{1}{4} N_2\,\overline{c_2}$. As it is presupposed that no more molecules pass through the cross-section in one direction than in the opposite direction, we must have $n_1 = n_2$ and hence $N_1\overline{c_1} = N_2\overline{c_2}$. From the fundamental equation we see that $\frac{p_1}{p_2} = \frac{N_1\overline{c_1^2}}{N_2\overline{c_2^2}} = \frac{N_1\left(\overline{c_1}\right)^2}{N_2\left(\overline{c_2}\right)^2}$ and consequently $\frac{p_1}{p_2} = \frac{\overline{c_1}}{\overline{c_2}} = \left(\frac{T_1}{T_2}\right)^{\frac{1}{2}}$. If the difference in temperature $T_1 - T_2$ is infinitely small and equal to dT, we get the expression

$$\frac{dp}{p} = \frac{1}{2}\frac{dT}{T} \quad \text{or} \quad \frac{dp}{dT} = \frac{p}{T}\cdot\frac{1}{2}.$$

This expression can, however, only be expected to be valid when we can disregard the number of the mutual impacts of the molecules as compared with the number of impacts against the walls of the tubes, or, in other words, when the cross-section dimensions of the tube are negligible compared with the mean free path λ of the gas molecules. If this requirement is not satisfied, $\frac{dp}{p}$ may be expected to be less than $\frac{1}{2}\frac{dT}{T}$, and for a cylindrical tube with radius r it may be expected that $\frac{dp}{p}$ will decrease when the ratio $\frac{2r}{\lambda}$ increases. For the case when $\frac{2r}{\lambda}$ is small compared with 1, I have previously[1] given the following condition of equilibrium

$$\frac{dp}{p} = \frac{1}{1 + \frac{2r}{\lambda}} \frac{d\bar{c}}{\bar{c}}.$$

Since $\frac{d\bar{c}}{\bar{c}} = \frac{dT}{2T}$ the expression is transformed into

$$\frac{dp}{p} = \frac{1}{1 + 2\frac{r}{\lambda}} \frac{dT}{2T}.$$

For the correctness of this expression I have previously, l. c., endeavoured to give reasons. These do not now seem to me to be conclusive. Hence I will for the time being substitute an unknown factor for the factor 2 in the denominator.

In all cases in which r is not negligible compared with λ there will be currents in the tube in the state of equilibrium. A current along the wall of the tube from the cold to the warm end will cause the pressure at the warm end to be greater than that at the cold end, and

[1] Martin Knudsen, Ann. d. Phys. Bd. 31, p. 223, 1910.

this gradient of pressure will cause a current to flow along the axis of the tube from the warm to the cold end.

For the momentum M received by each surface unit of the tube owing to the molecular velocity $\bar{c}$ varying through the tube, I have previously[1] given an expression which with a slight transcription gives

$$M = -\frac{3\pi}{128} p \varrho_1 \bar{c} \lambda \frac{d\bar{c}}{dl} k_1$$

where ϱ_1 denotes the density of the gas at a pressure of 1 bar and the temperature T, while dl denotes an element of the length of the tube. k_1 is a quantity which for very small values of $\frac{r}{\lambda}$ may be put equal to 1 and increases with increasing values of $\frac{r}{\lambda}$ to a limit which according to previous measurements lies between 2 and 3.

Since

$$\bar{c} = \sqrt{\frac{8}{\pi}} \sqrt{\frac{1}{\varrho_1}} \text{ and } \frac{d\bar{c}}{\bar{c}} = \frac{dT}{2T} \text{ we get } M = -\frac{3}{32} p\lambda \frac{1}{T} \frac{dT}{dl} \cdot k_1.$$

For the momentum B received by each surface unit in each second when the pressure gradient $\frac{dp}{dl}$ produces a current at constant temperature, calculation and experiments give

$$B = \frac{3 \cdot 0{,}81}{32} \sqrt{\frac{\pi}{8}} \frac{\sqrt{\varrho_1}}{\eta} r^2 p \frac{dp}{dl}$$

where η denotes the coefficient of viscosity which is connected with λ by the equation

$$\lambda = \frac{1}{0{,}49} \sqrt{\frac{\pi}{8}} \frac{\eta}{p\sqrt{\varrho_1}} \quad \text{so that} \quad B = \frac{3\pi}{256} \frac{0{,}81}{0{,}49} \frac{r^2}{\lambda} \frac{dp}{dl}.$$

[1] Martin Knudsen, Ann. d. Phys. Bd. 31, p. 214, 1910.

If temperature gradient and current are found at the same time, each length unit of the tube will receive the momentum $2\pi r (M + B)$ so that the condition of equilibrium will be $2\pi r (M + B) + \pi r^2 \frac{dp}{dl} = 0$. Hence, by the insertion of M and B we get

$$\frac{dp}{dT} = \frac{1}{\frac{8}{3}\frac{1}{k_1}\frac{r}{\lambda} + \frac{\pi}{16}\frac{0{,}81}{0{,}49}\frac{1}{k_1}\frac{r^2}{\lambda^2}} \cdot \frac{p}{2T}.$$

By a previous[1] series of experiments I have shown that this expression may be assumed to be correct when $\frac{r}{\lambda}$ is large.

An expression of this form and the expression given for small values of $\frac{r}{\lambda}$ may be embodied in the following

$$\frac{dp}{dT} = \frac{1}{\left(1 + a' \frac{r}{\lambda}\right)^2} \frac{p}{2T}$$

and it may then be expected that a' for small values of $\frac{r}{\lambda}$ will be of the same order of size as 1. For if we put $a' = 1$, the expression will be identical with the above-mentioned

$$\frac{dp}{dT} = \frac{1}{1 + 2\frac{r}{\lambda}} \frac{p}{2T}.$$

If we compare the equation containing a' with that into which k_1 enters, we should expect a' to decrease with increasing values of $\frac{r}{\lambda}$. The following expression satisfies this requirement

$$a' = a \frac{1 + b\frac{r}{\lambda}}{1 + c\frac{r}{\lambda}} \qquad \text{where} \quad b < c.$$

[1] Martin Knudsen, Ann. d. Phys. Bd. 33, p. 1435, 1910.

If we insert this value for a', we get the following general formula for the thermal molecular pressure

$$\frac{dp}{dT} = \frac{1}{\left(1 + a\frac{r}{\lambda}\frac{1 + b\frac{r}{\lambda}}{1 + c\frac{r}{\lambda}}\right)^2}\frac{p}{2T}. \tag{1}$$

The mean free path λ is, however, a function of p and T. We have $\lambda p = \lambda_1$ where λ_1 is the mean free path at a pressure of 1 bar and the temperature T. For the temperature interval at which my measurements were made the temperature dependency of the viscosity of hydrogen is given by the formula

$$\eta = \eta_0 \left(\frac{T}{273}\right)^{0,682}$$

and since

$$\lambda p = \frac{1}{0,49}\sqrt{\frac{\pi}{8}}\frac{\eta}{\sqrt{\varrho_1}} = \frac{1}{0,49}\sqrt{\frac{\pi}{8}}\frac{\eta}{\sqrt{\varrho_0}}\left(\frac{T}{273}\right)^{\frac{1}{2}}$$

where ϱ_0 is the density of the hydrogen at the pressure 1 bar and the temperature of melting ice ($T = 273°$), we get

$$\frac{1}{\lambda} = \frac{p}{\frac{1}{0,49}\sqrt{\frac{\pi}{8}}\frac{\eta_0}{\sqrt{\varrho_0}}\left(\frac{T}{273}\right)^{1,182}}.$$

Putting for hydrogen

$$\frac{\eta_0}{\sqrt{\varrho_0}} = 8,933$$

we get

$$\frac{1}{\lambda} = 0,08753\, p\left(\frac{273}{T}\right)^{1,182}. \tag{2}$$

Determination of the Constants in the General Formula for the Thermal Molecular Pressure.

Having tried various formulas I have chosen the one given above as that which with the fewest constants agrees best with the experimental material at the mean temperature 20.1° centigrade. If we remember that, when describing the results of our measurements, we put

$$2\,T\frac{f}{p} = 2\,T\frac{dp}{pdT} = \frac{1}{u}$$

we should according to formula (1) be able to put

$$\frac{dp}{dT} = \frac{1}{u}\cdot\frac{p}{2\,T}, \text{ where } u = \left(1 + a\frac{r}{\lambda}\frac{1 + b\frac{r}{\lambda}}{1 + c\frac{r}{\lambda}}\right)^2$$

and $$\frac{r}{\lambda} = 0{,}08753\cdot p\cdot r\left(\frac{273}{T}\right)^{1{,}182}.$$

By formation of differences in the table containing $\frac{1}{u}$ as a function of p and thence of $\frac{r}{\lambda}$ were found the following provisional values for the constants $a = 2.212$, $b = 2.85$, $c = 20.0$. These constants which must be expected to be influenced by the errors due to summation and to the use of earlier and uncertain observations, must now be improved, these sources of error being avoided. This is done by returning to the directly observed values $10^4\frac{\varDelta p}{p\varDelta T}$ given in the table. From this series a new series is formed by multiplication with $10^{-4}\cdot 2\,T$, where T is the mean temperature 293.1°. In this way we get a value for $2\,T\frac{\varDelta p}{p\varDelta T}$ which is regarded as the quantity observed at each single pressure.

Forming U from u by replacing r by R we get

$$2\,T\frac{\varDelta p}{p\varDelta T} = \frac{1}{u} - \frac{1}{U}$$

that is to say, an equation for each of the observations of the table. From these equations a, b and c are determined by the method of least squares, and thus we find

$$a = 2.46 \qquad b = 3.15 \qquad c = 24.6.$$

In this calculation it is supposed that all the measurements of $2\,T\dfrac{\Delta p}{p\,\Delta T}$ have been made with equal accuracy, so that occurring systematical errors will influence the constants.

Hence the result of the investigation is that the thermal molecular pressure in a circularly cylindrical tube with the radius r may be expressed as follows

$$\frac{dp}{dT} = \frac{1}{\left(1 + 2{,}46\,\dfrac{r}{\lambda}\cdot\dfrac{1 + 3{,}15\,\dfrac{r}{\lambda}}{1 + 24{,}6\,\dfrac{r}{\lambda}}\right)^2}\cdot\frac{p}{2\,T}$$

when putting for hydrogen

$$\frac{1}{\lambda} = 0{,}08753\,p\left(\frac{273}{T}\right)^{1{,}182}.$$

In the case of such great temperature differences that the differential formula cannot be directly applied, an integration may be undertaken, the last equation giving

$$\frac{dp}{p} = \frac{1{,}182\,dT}{T} - \frac{d\lambda}{\lambda}$$

which, inserted in the last but one, gives

$$\frac{dT}{2\,T} = \frac{d\lambda}{\lambda}\,\frac{1}{2{,}364 - \dfrac{1}{u}}.$$

If in this we insert the value found for u the result will be an equation which can easily be integrated. Such an integration between the limits $293° \pm 20°$ showed that we are entirely warranted in applying the differential formula without integration within these limits.

In order to investigate how the experimental results are rendered by the differential formula with the constants found the following table was calculated.

p Bar	$\frac{r}{\lambda}$ at 20° C.	$2T\frac{\Delta p}{p\Delta T}$ observed	$2T\frac{\Delta p}{p\Delta T}$ calculated	observed—calculated
19082	41,22	0,00463	0,00464	— 0,00001
9904	21,39	0,01486	0,01486	0,00000
5140	11,103	0,0437	0,0425	+ 0,0012
2668	5,763	0,0986	0,1024	— 0,0038
1385	2,991	0,195	0,196	— 0,001
718,7	1,552	0,287	0,285	+ 0,002
373,0	0,8057	0,318	0,317	+ 0,001
193,6	0,4182	0,276	0,278	— 0,002
100,5	0,2170	0,204	0,207	— 0,003
52,15	0,1126	0,149	0,146	+ 0,003
27,07	0,0585	0,105	0,107	— 0,002

From the above table it will be seen that the general formula gives a very good representation of the experimental results within the range investigated. The differences between the observed and the calculated values is of the order of 1 p. c. of the observed values except for the observation made at the pressure of 2668 bar, where the difference amounts to almost 4 p. c. This may possibly be due to an incorrect determination of the pressure which is confirmed by the following.

As it would be of interest to investigate other temperatures, two other series of experiments were made, simultaneously with those between 0° and 40°, at the same pressures. In one of these the one bath was scraped ice, the other a mixture of carbonic acid and ether. The temperature of this mixture was put at —78.5 C. In the second series of experiments liquid air was used in one bath, while the other was again a mixture of carbonic acid and ether. The temperature of the liquid air was determined by means of the usual small floats. In these two series of measurements the quantity $2T\frac{\Delta p}{p\Delta T}$ given in the following tables was again determined. For comparison with the general formula $\frac{1}{u}-\frac{1}{U}$ was calculated by integration and the result subtracted from the observed values for $2T\frac{\Delta p}{\Delta T}$. The mean temperature T (absolute) of the baths used is also given in the tables, as well as the values $\frac{r}{\lambda}$ calculated from p and T.

p	$T = 233.75°$ Ice and Carbonic Acid			$T = 138.3°$ Carbonic Acid and Liquid Air		
	$\frac{r}{\lambda}$	$2T\frac{\Delta p}{p\Delta T}$	observed-calculated	$\frac{r}{\lambda}$	$2T\frac{\Delta p}{p\Delta T}$	observed—calculated
19082	53,87	0,0031	+ 0,0003	100,18	0,0010	+ 0,0002
9920	28,00	0,0093	0,0000	52,08	0,0031	+ 0,0001
5142	14,52	0,0288	+ 0,0007	27,00	0,0098	— 0,0001
2705	7,636	0,067	— 0,005	14,20	0,025	— 0,003
1389	3,921	0,153	— 0,001	7,292	0,071	— 0,004
718,6	2,029	0,255	+ 0,004	3,773	0,151	— 0,007
373,5	1,054	0,313	+ 0,004	1,961	0,254	+ 0,001
193,2	0,5454	0,293	+ 0,001	1,014	0,301	0,000
100,8	0,2844	0,227	— 0,001	0,529	0,274	— 0,002
52,22	0,1474	0,170	+ 0,009	0,274	0,217	— 0,005
27,15	0,0766	0,117	+ 0,001	0,143	0,157	+ 0,008

The differences between the observed and the calculated values are not any greater than might reasonably be expected. A systematic course in the differences only appears in the series with carbonic acid and liquid air and is not marked enough to give reasons for a change of the constants in the general formula. The greatest percentage difference between the observed and the calculated values appears in both series at a pressure of 2705 bar, that is to say, at the same pressure at which the greatest deviation in the series 40°-ice was found. This would seem to indicate that an error has crept in in the determination of the pressure, which is not, however, so great that we should feel justified in leaving the observations at this pressure out of consideration.

In order to investigate the thermal molecular pressure at higher temperatures a series of experiments were conducted at a mean temperature of abt. 260°. The values observed for $2\,T\frac{\Delta p}{p\,\Delta T}$ here proved to be abt. 10 p. c. lower than those calculated by the formula given above. For pressures higher than 1000 bar this discrepancy is chiefly due to the fact that the temperature dependency given in the formula $\eta = \eta_0\left(\frac{T}{273}\right)^{0.682}$ is not valid at high temperatures. Breitenbach's[1] exponent 0.5832 instead of 0.682 would give a considerably better correspondence. At lower pressures such an alteration of the exponent will not, however, greatly alter the calculated values, and the explanation may be perhaps that the harmful adsorption phenomena make their influence more felt at high than at low temperatures because the adsorption processes take place more rapidly in the first case.

[1] Breitenbach, Ann. d. Phys. Bd. 67, 1899, p. 817.

In the above I have mentioned that as a guide in the formation of the general empirical differential formula I used the theoretical expression

$$\frac{dp}{dT} = \frac{1}{1 + 2\frac{r}{\lambda}} \frac{p}{2T}$$

which I thought must hold good when r is very small but not negligible compared with λ. For this case the empirical formula gives

$$\frac{dp}{dT} = \frac{1}{1 + 2a\frac{r}{\lambda}} \frac{p}{2T}.$$

If the theoretical formula were correct, the measurements should thus have given $a = 1$. They have, however, given $a = 2.46$, which is a considerable discrepancy. Whether this is due to an incorrect determination of the quantity a on account of adsorption phenomena or whether some error attaches to the theoretical formula I dare not say.

In order to elucidate this latter question, a theoretical derivation of the formula ought to be made, based solely on the kinetic theory of gases. Such a derivation would presumably be very difficult in the general case, but would seem feasible here where we are considering the case of r being small compared with λ. In case this calculation were made and in case it would in future be possible to avoid the adsorption phenomena so that a could be determined with sufficient accuracy, this method presents a direct measurement of λ, this quantity, the mean free path, of which it is now hardly possible to give an exact definition, would in that case be directly compared with the radius r of the tube, and we should have another means of obtaining information of the direction in which the molecules moved after the so-called mutual impacts.

With regard to the influence of the adsorption the following remark may be made. In the way the apparatus was arranged we can hardly suppose that it contained mercury in other places than in the trap which was cooled in liquid air. It may be supposed, however, that there will be adsorbed water all over the walls of the glass, which will pass at an extremely slow rate towards the trap with liquid air. This passage will presumably be so slow and regular when the temperature of the apparatus is kept constant that the hydrogen pressure will practically be the same throughout. Otherwise when the joint of a tube is heated or cooled. In the former case water is liberated from the walls, in the latter case water is adsorbed. Both processes will produce currents which are different in the wide tube from those in the narrow tube, and these currents may be expected to produce differences in pressure which will become sources of error in the measurements.

Finally, it cannot be precluded that the hydrogen itself may to some extent be adsorbed to the glass wall. Even if such an adsorption is not appreciably altered at the small differences of temperature employed in the experiments, it will, however, cause λ to be smaller close to the wall of the tube than it is at the axis of the tube, and thus explain that the constant a has been found greater than 1.

It might perhaps be supposed that the harmful effect of the adsorption phenomena would appear less at high than at low temperatures. My measurements do not indicate this, however. The adsorption phenomena cause currents in the tubes lasting at least twenty-four hours and probably several days and nights, so that a stationary condition is not obtained within a reasonable time. Hence it is a reasonable supposition that even if the adsorbed masses are much smaller at high than at low temperatures, the liberation of adsorbed substance at increased temperature will take place at a much quicker rate with a high than with a low temperature, and produce just as strong or perhaps stronger currents in the tubes.

In the execution of the above-described measurements and calculations Mr. K. Thiesen and Mr. G. Nørgaard have rendered much valuable assistance for which I offer sincere thanks. My grateful acknowledgments are also due to the directors of the Carlsberg Fund who have granted financial aid without which it would have been impossible for me to carry through these investigations.

Forelagt paa Mødet den 6. Maj 1927.
Færdig fra Trykkeriet den 30. Juni 1927.

MATHEMATISK-FYSISKE MEDDELELSER

UDGIVNE AF

DET KGL. DANSKE VIDENSKABERNES SELSKAB

6. BIND (Kr. 17,00):

7. BIND (Kr. 17,70):

Kr. Ø.

9. Bjerrum, Niels: Untersuchungen über Ionenassoziation. I. Der Einfluss der Ionenassoziation auf die Aktivität der Ionen bei mittleren Assoziationsgraden. 1926 .... 2.00
10. Nielsen, Niels: Recherches sur les fonctions cylindriques et sur certaines fonctions analogues. 1926 .... 1.35
11. Hevesy, G.: On the Missing Element 87. 1926 .... 0.60
12. Kuhn, W.: Die Stärke der anomalen Dispersion in nicht leuchtendem Dampfe von Thallium und Cadmium. 1926 .... 3.75
13. Sugiura, Y. and Urey, H. C.: On the Quantum Theory explanation of the Anomalies in the 6th and 7th Periods of the Periodic Table. 1926 .... 1.15
14. Burrau, Øyvind: Berechnung des Energiewertes des Wasserstoffmolekel-Ions (H_2^+) im Normalzustand. 1927 .... 0.90
15. Knudsen, Martin: The Hot-wire Manometer. 1927 .... 0.80
16. Juel, C.: Om v. Staudt's Definitioner. 1927 .... 0.85
17. Pólya, G.: Über die algebraisch-funktionentheoretischen Untersuchungen von J. L. W. V. Jensen. 1927 .... 1.40

8. BIND (under Pressen):

1. Hille, Einar: On the Logarithmic Derivatives of the Gamma Function. (Under Pressen).
2. Rasch, G.: Beitrag zur Theorie der unvollständigen Gammafunktionen. Nach hinterlassenen Papieren von J. L. W. V. Jensen. 1927 .... 0.85
3. Knudsen, Martin: Thermal Molecular Pressure in Tubes. 1927 1.50

Det Kgl. Danske Videnskabernes Selskab.
Mathematisk-fysiske Meddelelser. **VIII**, 4.

DEN HØJERE ATMOSFÆRES SAMMENSÆTNING, TRYK, TEMPERATUR OG ELEKTRISKE LEDNINGSEVNE I BELYSNING AF RADIOBØLGERNES UDBREDELSESFORHOLD

AF

P. O. PEDERSEN

KØBENHAVN
HOVEDKOMMISSIONÆR: ANDR. FRED. HØST & SØN, KGL. HOF-BOGHANDEL
BIANCO LUNOS BOGTRYKKERI
1927

Pris: Kr. 0,60.

Det Kgl. Danske Videnskabernes Selskabs videnskabelige Meddelelser udkommer fra 1917 indtil videre i følgende Rækker:

Historisk-filologiske Meddelelser,
Filosofiske Meddelelser,
Mathematisk-fysiske Meddelelser,
Biologiske Meddelelser.

Hele Bind af disse Rækker sælges 25 pCt. billigere end Summen af Bogladepriserne for de enkelte Hefter.

Selskabets Hovedkommissionær er *Andr. Fred. Høst & Søn*, Kgl. Hof-Boghandel, København.

Det Kgl. Danske Videnskabernes Selskab.
Mathematisk-fysiske Meddelelser. **VIII,** 4.

DEN HØJERE ATMOSFÆRES SAMMENSÆTNING, TRYK, TEMPERATUR OG ELEKTRISKE LEDNINGSEVNE I BELYSNING AF RADIOBØLGERNES UDBREDELSESFORHOLD

AF

P. O. PEDERSEN

KØBENHAVN
HOVEDKOMMISSIONÆR: ANDR. FRED. HØST & SØN, KGL. HOF-BOGHANDEL
BIANCO LUNOS BOGTRYKKERI
1927

I Løbet af det sidste Par Aar har Forfatteren foretaget nogle teoretiske Undersøgelser vedrørende Radiobølgers Udbredelse, og Resultaterne af disse Undersøgelser er i Mellemtiden samlede i en Bog, samt i nogle Tidsskriftsartikler, der nærmere uddyber nogle af de radiotekniske Konsekvenser[1]. Enkelte af de vundne Resultater kan formentlig have Interesse ud over Radioteknikernes Kreds, men da Bogen og de nævnte Tidsskriftartikler i Hovedsagen er skrevet ud fra et radioteknisk Synspunkt, saa findes de Strejflys, som Undersøgelsen kaster over Problemer fra Meteorologien og Geofysikken, spredt over de nævnte Publikationer. Det er saaledes ikke helt let alene paa Grundlag af disse at danne sig et Overblik over Undersøgelsens mulige Betydning for de nævnte Videnskaber. Det kan derfor formentlig være rimeligt ganske kort at gøre Rede for disse Forhold samtidig med, at der fremføres enkelte nye Momenter.

Erfaringen har vist, at korte Bølger, omkring 15 til 20 m, kan gaa Jorden een, ja endog to eller endnu flere Gange rundt og dog modtages med betydelig Intensitet[2]. Det

[1] »The Propagation of Radio Waves«, København 1927, citeret som »P. R. W.«
»Ingeniøren« Nr. 9, p. 101—104. 1927.
»Radio Posten« Nr. 1, p. 5—7, Nr. 2, p. 5—6. 1927.
»Radio Magasinet« Nr. 3, p. 155—159, Nr. 4, p. 227—232. 1927.
»Radiofoni Aarbogen«. 1927.

[2] E. Quäck: »E. N. T.« Bd. 4, p. 308—312. 1927. Meddelelsen om, at det er lykkedes at paavise Radiobølger, der har gaaet to Gange rundt

drejer sig følgelig om Transmissionsafstande paa 40 000 til 80 000 km og derover.

Det er let at paavise, at denne Udbredelse af Radiobølgerne hverken kan ske gennem Jorden eller direkte langs dennes Overflade [1]. Tilbage staar kun den Mulighed, at Bølgeudbredelsen sker ved Hjælp af og igennem Atmosfæren, og her maa det i særlig Grad være den højere Atmosfære, der spiller ind. Denne maa nødvendigvis virke som en Art »Skal«, der i hvert Fald til en vis Grad holder sammen paa Radiobølgerne og hindrer disses fuldstændige Udstraaling til Verdensrummet.

Denne Skals Højde over Jordens Overflade ligger mellem 100 og 160 km [2]; men vi kommer senere tilbage til Bestemmelsen af denne Højde.

Hovedspørgsmaalet er nu:

1) Foregaar Bølgeudbredelsen paa den Maade, at Bølgestraalerne fra Jordens Overflade gaar op til »Skallen« og enten reflekteres fra eller brydes i denne, saaledes at de atter vender tilbage til Jordens Overflade, fra hvilken de igen reflekteres; hvorefter den samme Proces gentages en Række Gange, idet hvert enkelt af disse »Skridt«, som Bølgestraalen saaledes kan tage, kun vilde have en Længde paa omkring 1000 til 2000 km? I dette Tilfælde vil kun en mindre Del af Straalens Bane ligge i den højere Atmosfære.

2) Eller tager Bølgestraalerne kun et enkelt eller nogle faa Skridt? I dette Tilfælde vil den allerstørste Del af

om Jorden, er først fremkommen efter Afslutningen af »P. R. W.«, men den bekræfter og understreger kun yderligere Berettigelsen af de der dragne Slutninger med Hensyn til Bølgernes maksimale Dæmpningskonstant.

[1] »P. R. W.« p. 20 (Fig. 4) og p. 189; »Radio Posten« Nr. 1.

[2] De følgende Betragtninger gælder derfor i Hovedsagen ogsaa kun Atmosfærens Tilstand op til ca. 200 km. Højde.

Straalens Bane komme til at ligge i den højere Atmosfære, idet Bølgestraalen paa den allerstørste Del af Vejen følger »Skallen«.

Det kan nu paavises, at det gennemsnitlige Energitab[1] ved en enkelt Reflektion af de korte Bølger fra Jordens Overflade ikke kan sættes til mindre end 75 %, og i mange Tilfælde endda vil overstige 90 %, saaledes at den reflekterede Straale kun har indtil 10 til 25 % af den indfaldende Straales Energi. Men naar dette er Tilfældet, ser man let, at Bølgernes Udbredelse over de meget lange Afstande ikke kan ske gennem mange, forholdsvis korte Skridt[2]. Jeg skal belyse dette ved et Taleksempel: Hvis Bølgen tog 2000 km lange Skridt og gik Jorden to Gange rundt, saa medgik dertil ialt 40 Skridt. Ved disse 40 Reflektioner vilde Bølgernes Energi formindskes til $1 \cdot 10^{-24}$ af den oprindelige Energi, naar Tabet ved hver Reflektion var 75 %, og til $1 \cdot 10^{-40}$, hvis Tabet var 90 %. Men selv den første af disse Svækkelser er mere end 10^{15} Gange for stor. De korte Bølger maa nødvendigvis tage de store Afstande i et enkelt eller højst i nogle faa Skridt.

Radiobølgerne maa derfor den allerstørste Del af Vejen bevæge sig i den højere Atmosfære i Skallens Højde og praktisk talt parallelt med Jordens Overflade.

Naar dette er slaaet fast, kan man uden Vanskelighed angive en højere Grænse for Bølgernes Dæmpning under deres Udbredelse i denne Del af Atmosfæren. Denne

[1] Hvortil ogsaa regnes det Energitab, der fremkommer ved, at en Del af den reflekterede Straaleenergi har en saadan Retning, at den gaar tabt til Verdensrummet.

[2] »P. R. W.« p. 194—200. »Radio Magasinet« Nr. 3 og 4.

Dæmpning kan skrives paa Formen

$$\varepsilon^{-\gamma_0 x}, \tag{1}$$

hvor x er den af Bølgen gennemløbne Vej, og hvor Dæmpningskonstanten γ_0 for de her betragtede korte Bølger og for Højder over 100 km med tilstrækkelig Tilnærmelse er bestemt ved

$$\gamma_0 = k_0 \nu, \tag{2}$$

hvor k_0 er en Konstant, der bl. a. paa kendt Maade[1] afhænger af Tætheden af de fri Elektroner og af Bølgelængden, medens ν er det Antal Stød, som en Elektron i Gennemsnit lider per Sekund.

Her kan man, som nævnt, fastslaa en Maksimumsværdi for γ_0 [2] og en Minimumsværdi for k_0 [3]. Ligning (2) giver saaledes ogsaa en Maksimumsværdi for Stødtallet ν.

Men her møder man en Vanskelighed, som vi maa betragte lidt nærmere.

Regner man, at der i den nederste Del af Atmosfæren foregaar en fuldstændig Blanding af de enkelte Luftarter op til 12 km, medens en saadan Blanding ikke finder Sted over denne Højde, og regner man samtidig, at Temperaturen her er gaaet ned til —55° C. = ca. 220° K. og holder sig paa dette Punkt i hele Resten af Atmosfæren, saa kan man for enhver Højde over Jorden let udregne Partialtrykket for hver enkelt af de Luftarter, der indgaar i Atmosfæren.

Af disse Luftarter har kun Kvælstof, Ilt, Helium, Brint

[1] »P. R. W.«, p. 214.

[2] »P. R. W.«, p. 205.

[3] Ifølge »P. R. W.«, p. 214, Formel (48) er $k_0 = N \cdot \frac{2\pi e^2}{m c n_0 \omega^2}$, hvor e, m, c, n_0, ω er kendte Konstanter, medens N er Antallet af fri Elektroner per cm³. Som vist »P. R. W.« p. 179—182 kan man paa Grundlag af de korte Bølgers Udbredelsesforhold angive en lavere Grænse for den maksimale Elektrontæthed N.

og Vanddamp Mulighed for at spille nogen større Rolle i Højder over 100 km. Vi vil imidlertid se bort baade fra Vanddamp og fra Brint. For den førstes Vedkommende fordi det kan forudses, at der ved den nævnte, lave Temperatur kun kan findes ubetydelige Mængder af Vanddamp i den højere Atmosfære.

For Brintens Vedkommende stiller Sagen sig noget anderledes, idet denne Luftart paa Grund af sin store Lethed vilde blive den dominerende Konstituent i den højeste Atmosfære, hvis der er et kendeligt Indhold af fri Brint ved Jordens Overflade, og der ikke skete et Tab af fri Brint i Atmosfæren, f. Eks. ved at den forenede sig med Ilt til Vanddamp (under Paavirkning af Lyset eller af elektriske Udladninger eller paa anden Maade), eller ved at den forsvandt ud til Verdensrummet. Brintindholdet ved Jordens Overflade angives sædvanligvis til 0.01 % ned til 0.001 % Vol.[1]; A. Krogh[2] har dog vist, at Indholdet af fri Brint i den nedre Atmosfære i hvert Fald ikke er over 0.0005 % og sandsynligvis mindre end 0.0002 %. Selv med det mindste af disse Tal vilde Brinten blive den væsentligste Bestanddel af Luften i 140 km Højde og derover. Vi vil indtil videre gaa ud fra, at Brintindholdet i den højere Atmosfære er lig med Nul. Berettigelsen af denne Antagelse vil fremgaa af det følgende.

Tilbage bliver da kun Kvælstof, Ilt og Helium, hvis Partialtryk og samlede Tryk vi under de angivne Forudsætninger let kan bestemme.

[1] A. Wegener: Thermodynamik der Atmosphäre, anfører 0,003 % Brint, (p. 46). Leipzig 1911.
W. J. Humphreys: Physics of the Air, anfører 0,01 % Brint, (p. 69). Philadelphia 1920.
Hann u. Süring: Lehrbuch der Meteorologie (4. Aufl.), anfører 0.001 %, (p. 5). Leipzig 1926.

[2] A. Krogh: Vid. Selsk. Math.-fys. Medd. I, Nr. 12. (1919).

Til Bestemmelse af Stødtallet ν for Elektronerne maa man kende dels Elektronernes gennemsnitlige Hastighed, dels den gennemsnitlige Længde af deres fri Vejlængde, idet Stødtallet ν er bestemt ved

$$\nu = \frac{\text{Hastighed}}{\text{fri Vejlængde}}. \tag{3}$$

Her kan Elektronernes gennemsnitlige Hastighed i hvert Fald ikke være mindre end den til —55° C. svarende termiske Hastighed. Indsætter vi denne termiske Hastighed i (3) og indsætter vi ligeledes Elektronernes fri Vejlængde bestemt paa Grundlag af Lufttrykket og under Hensyntagen til de ved de foreliggende Forsøg bestemte Værdier af Elektronernes fri Vejlængde i de tre nævnte Luftarter, saa viser det sig, at det ved (3) bestemte Stødtal er for stort, større end det er tilladeligt efter Formel (2). Med andre Ord, Bølgernes Dæmpning bliver herefter alt for stor.

Hvorledes kommer man udenom denne Vanskelighed?

Da den fri Vejlængde er omvendt proportional med Lufttrykket, og Stødtallet derfor ifølge (3) er ligefrem proportionalt med dette, saa vilde Vanskeligheden forsvinde, hvis Trykket paa en eller anden Maade var ca. 10 Gange mindre end antaget.

Denne Udvej støder formentlig paa uovervindelige Vanskeligheder: For det første vilde vi ikke komme saa langt ned med Trykket, selv om vi antog, at Blandingshøjden gik helt op til 20 km, og at Temperaturen i Stratosfæren gik helt ned til —100° C. = c. 175° K., og hertil kommer endda, at forskellige Ejendommeligheder ved Lydbølgers Udbredelse over store Afstande i Atmosfæren tyder paa, at der atter finder en ikke ubetydelig Temperaturstigning Sted i en Højde omkring 40 til 60 km, saaledes at den lave Temperatur af ca. —55° C. = ca. 220° K., som findes i

Tropopausen, maa vige for en højere Temperatur i de større Højder. Nyere Arbejder, bl. a. af Helge Petersen, tyder i samme Retning.[1]

Med Hensyn til Temperaturen i den højere Atmosfære synes Fantasien hidtil at have haft ret frit Spil, da selv smaa Ændringer i Forudsætningerne angaaende de fysiske Forhold fører til ret store Ændringer i den beregnede Temperatur. Saaledes er Lindemann og Dobson kommen til en Temperatur paa omkring 300° K. for Højder over 60 km; L. Vegard har tidligere sluttet sig hertil, men er senere gaaet over til den Opfattelse, at Temperaturen kun er ca. 35° K.

I »Propagation of Radio Waves« er der over 12 km regnet med ca. 220° K. Ændredes denne Forudsætning til, at Temperaturen over 60 km var 330° K., saa vilde Følgen blive, at »Skallens« Højde ved Dag vilde hæves fra omkring 130 til omkring 165 km. Denne Højde er, saa vidt Erfaringerne fra Radiobølgeomraadet rækker, noget for stor. Ogsaa Bølgernes Dæmpning vilde blive noget for stor. Alt i alt lader denne Forudsætning sig dog ikke helt afvise, og den vilde passe ret godt med Meteorerfaringerne. Derimod vilde Radiobølgeerfaringerne komme i afgjort Strid med Antagelsen af en endnu højere Temperatur i Atmosfæren, i hvert Fald for Højder under 150—200 km.

Der er iøvrigt næppe Tvivl om, at man i Løbet af kort Tid vil blive i Stand til ad Radiovejen at bestemme de ledende Lags Højde saa nøje, at Usikkerheden med Hensyn til de højere Luftlags Temperatur for Højder op til 150—200 km vil blive meget mindre.

Den nævnte Ændring, nemlig fra 220° til 330° K., af den antagne Temperatur i de højeste Luftlag, vilde ikke

[1] Helge Petersen: Phys. Zeitschrift Bd. 28, p. 510—513. 1927.

ændre de følgende Slutningers Rigtighed, men kun enkelte af de anførte Højdeangivelser for »Skallen«.

Og sidst, men ikke mindst, vilde et væsentlig mindre Lufttryk i den højere Atmosfære (omkring 150 km) end her antaget komme i afgjort Modstrid med de foreliggende Kendsgerninger vedrørende Meteorers Hastighed, Lysstyrke m. m.[1]

Disse forskellige Forhold tyder saaledes bestemt paa, at Lufttrykket ikke kan være lavere, men snarere maa antages at være lidt højere end forudsat i »P. R. W.«

Den foran forsøgte Udvej er saaledes ikke farbar. Vi skal derfor undersøge en anden Mulighed.

Vi har hidtil antaget, at »Skallen« laa i en Højde af omkring 130 km. Tænker vi os, at den rykkede op til 250 km, saa vilde Trykket der være saa meget lavere, at det ifølge (3) beregnede Stødtal ikke blev ret meget for stort. Ogsaa denne Udvej er imidlertid spærret. Dels kan man eksperimentelt bestemme »Skallens« Højde, og selv om disse Bestemmelser ikke er meget sikre, er de i hvert Fald gode nok til at vise, at »Skallens« Højde ikke kan være saa stor.

Men hertil kommer, at i saa store Højder (eller udtrykt paa anden Maade: ved saa lave Ilt- og Kvælstoftryk) vil Elektrontætheden forandre sig saa langsomt, at de meget

[1] Ganske vist kræver den af Lindemann og Dobson (Proc. Roy. Soc. (A). Vol. 102, p. 411—437, 1922. Vol. 103, p. 339—442, 1923) opstillede Meteorteori en alt for høj Massetæthed og selv efter den af Sparrow (Astrophys. J. Vol. 63, p. 90—110. 1925) foretagne, meget væsentlige Korrektion af Teorien er dennes Krav til Massetæthed i den højere Atmosfære formentlig for stort. Men det vil paa den anden Side næppe være muligt at forklare Meteorfænomenerne med en mindre Massetæthed end den, hvormed der er regnet i »P. R. W.« Ja det vilde vistnok være ønskeligt, set alene fra Meteorsynspunktet, at have en 2 à 3 Gange saa stor Massetæthed.

udprægede Forskelligheder ved Bølgernes Udbredelse ved Dag og ved Nat, slet ikke vilde kunne komme frem.

Endelig er i saa stor Højde Ilt- og Kvælstoftrykket saa ringe, at den nødvendige Ionisation ikke alene kan stamme fra de der tilstedeværende Ilt- og Kvælstofmolekuler, men for den allerstørste Dels Vedkommende maatte skyldes det tilstedeværende Helium. Men til at fremkalde en saadan Ionisation i Helium er Intensiteten af det ultraviolette Sollys omkring 10^7 til 10^8 Gange for svagt.

Hertil kommer en Række andre Forhold, der viser, at »Skallen« ikke kan ligge saa højt; men de anførte Grunde er formentlig tilstrækkelige.

Da man saaledes hverken kan formindske Elektronernes Hastighed eller Lufttrykket, synes man at staa overfor en uovervindelig Vanskelighed. Der er dog ikke desto mindre en Udvej. C. Ramsauer og andre har nemlig vist [1], at det »Tværsnitsareal«, som Atomerne i de ædle Luftarter Neon [2], Argon, Krypton og Xenon frembyder overfor Sammenstød med Elektroner konvergerer mod Nul, naar den relative Hastighed ved Stødet antager meget smaa Værdier, og netop Værdier af den Størrelsesorden, som det drejer sig om ved de her betragtede Stød.

Med Hensyn til Helium foreligger der endnu ikke afgørende Bevis for, at det samme er Tilfældet, men Forsøgene tyder vel nærmest i den Retning, lige saa vel som det i og for sig vel nok er rimeligt at antage, at denne Luftart i saa Henseende følger de nævnte, ædle Luftarter. [3]

Vi vil derfor forudsætte, at den fri Vejlængde for lang-

[1] Se f. Eks.: J. Franck und P. Jordan: Anregung von Quantensprüngen durch Stösse, p. 20. Berlin 1926.

[2] E. Brücke: Ann. d. Physik, Bd. 84, p. 279—291. 1927.

[3] Direkte Forsøg med saa langsomme Elektroner, som det her drejer sig om (nemlig svarende til ca. $^1/_{86}$ Volt), foreligger formentlig ikke

somme Elektroner i Helium er mellem 18 og 20 Gange saa stor som i Kvælstof ved samme Tryk. Er dette Tilfældet, saa faar Stødtallet netop en saadan Værdi, at den paa Grundlag deraf beregnede Dæmpning for Bølgerne bliver af den rigtige Størrelsesorden.

Vi har saaledes paa naturlig Maade overvundet den nævnte Vanskelighed; men man ser tillige, at den benyttede Udvej ikke vilde føre til Maalet, hvis der i den højere Atmosfære foruden de omhandlede tre Luftarter, Kvælstof, Ilt og Helium, tillige var Brint [1] tilstede i betydelig Mængde. Vi føres derfor til den før nævnte Slutning, at Brintindholdet i disse Højder er lig med Nul. Mod denne Antagelse er der, saa vidt vides, intet der strider.

Man spørger uvilkaarlig: Hvorledes har de mange andre Forfattere [2], der har beskæftiget sig med Spørgsmaalet, stillet sig til den ovenfor behandlede Vanskelighed? Svaret

hverken for Helium eller for andre Luftarter. Det vil vistnok ogsaa være vanskeligt at udføre saadanne i Laboratoriet. Dog er der en Mulighed for, at man ved Hjælp af Laboratorieforsøg med meget korte Radiobølger ($\lambda < 4$ m) kan opnaa betydningsfulde Resultater angaaende langsomme Elektroners Bevægelse i stærkt fortyndet Luft. Der kan i saa Henseende henvises til nogle af H. GUTTON og J. CLÉMENT udførte, interessante Forsøg (L'Onde électrique, Vol. 6, p. 137—151. 1927). De nævnte Forfatteres teoretiske Behandling af Forsøgsresultaterne er dog formentlig urigtig, se nærmere »P. R. W.« p. 92—94. Hvis denne Metode ikke skulde vise sig anvendelig, frembyder den højere Atmosfære maaske den eneste Mulighed for saadanne Forsøgs Udførelse. Hvis Slutningerne i »P. R. W.« er rigtige, saa viser de korte Bølgers Udbredelsesforhold ikke alene, at Heliumatomets "Tværsnitsareal" nærmer sig til Nul overfor meget langsomme Elektroner, men tillige, at Elektronernes "Ophold" i Heliumatomet eller i dets umiddelbare Nærhed er mindre end $5 \cdot 10^{-8}$ Sekund. I Laboratoriet har det, saa vidt vides, ikke hidtil været muligt at skaffe Klarhed over dette Forhold.

[1] Eller en anden Luftart, der ikke viser RAMSAUER-Effekten.

[2] I Løbet af det sidste Par Aar er der skrevet omkring 300 videnskabelige Afhandlinger om det foreliggende Emne.

er for saa vidt meget let. Ingen af disse Forfattere er, saa vidt vides, gaaet saa dybt ind paa Spørgsmaalets fysiske og meteorologiske Side, at de overhovedet er stødt paa denne Vanskelighed. De har hver for sig behandlet en mindre Del af Problemet og derved gjort de Forudsætninger, som gav Resultater, der stemmede med Erfaringerne indenfor det behandlede, begrænsede Omraade, men uden Hensyn til om disse Forudsætninger overhovedet var forenelige med hele det foreliggende Erfaringskompleks, ikke alene vedrørende Radiobølgernes Udbredelse, men ogsaa vedrørende den spektrale Fordeling af Solstraalingen og Forholdene i den højere Atmosfære. [1]

Den eneste mig bekendte Undtagelse herfra danner de af H. Lassen [2] i Løbet af det sidste Aarstid publicerede Afhandlinger, hvor han gaar noget ind paa alle disse Forhold. Men denne Forfatter ser helt bort fra de frie Elektroners Indflydelse, idet han forudsætter, at der i den højere Atmosfære findes rigeligt af fri Brint, og at de frigjorte Elektroner straks fanges ind af de neutrale Brintmolekuler under Dannelse af H_2^--Ioner. Ogsaa han undgaar derved at komme ind paa Spørgsmaalet; men den omtalte Iondannelse er meget lidet sandsynlig og spiller sikkert en meget ringe Rolle. Dette fremgaar bl. a. af Arbejder af W. B. Haines (1915). Ja selv i Kanalstraalerør, i hvilke man iøvrigt finder næsten alle mulige Brintioner rigelig repræsenterede (H^+, H^-, H_2^+, H_3^+) har man først i den allersidste Tid været i Stand til at paavise H_2^--Ioner [3].

[1] Som Eksempel herpaa skal nævnes E. O. Hulburt's Artikel i Phys. Rev. (II), Vol. 29, p. 712, Maj 1927, hvori det siges: Elektrontætheden stiger med Højden og naar en Værdi af omkring 10^5 Elektroner pr. cm^3 i en Højde af 100 miles. Men den nævnte Forfatter kommer iøvrigt ikke nærmere ind paa Spørgsmaalet. Se ogsaa »P. R. W.« p. 237.

[2] »P. R. W.« p. 46 og 184.

[3] Handb. d. Phys. Bd. XXIV, p. 78. 1927.

LASSEN's Opfattelse frembyder iøvrigt en Række andre Vanskeligheder, som det dog vilde føre for langt at komme ind paa ved denne Lejlighed, og den kan næppe siges at give Mulighed for en tilfredsstillende Løsning af de foreliggende Problemer.

Vi skal derefter gaa over til at betragte den omhandlede Undersøgelses Relation til et andet Spørgsmaal indenfor Geofysikfens Omraade, nemlig til Teorien for de regelmæssige daglige Variationer i Jordens magnetiske Felt, hvilke Variationer staar i Forhold til Solens og Maanens og Jordens indbyrdes Stilling. A. SCHUSTER [1] har udviklet en Teori for disse Variationer, ifølge hvilken de skyldes Atmosfærens elektriske Ledningsevne, og S. CHAPMAN har i en lang Række Afhandlinger uddybet denne Teori yderligere [2]. Han kommer derved bl. a. til det Resultat, at Atmosfærens totale Ledningsevne skal være omkring $25 \cdot 10^{-6}$ (e.m.e.; cm) for at give en tilfredsstillende Forklaring af de nævnte Variationer.

SCHUSTER-CHAPMAN's Teori giver ikke sikre Holdepunkter for, hvor højt oppe i Atmosfæren denne Ledningsevne skal være lokaliseret, men det antages almindelig, at det ledende Lag ligger i omkring 50 km Højde [3]. Denne Opfattelse er formentlig uholdbar; til Opretholdelse af den omtalte Ledningsevne, der forudsættes at være jævnt fordelt mellem 45 og 55 km Højde, vilde der kræves, at den ultraviolette Straaling fra Solen var omkring 10^9 Gange stærkere, end den kan antages at være [4]. Den omtalte Led-

[1] A. SCHUSTER: Trans. Roy. Soc. (A). Vol. 208, p. 163—204; 1908.

[2] S. CHAPMAN: Trans. Roy. Soc. (A). Vol. 213, p. 279—321; 1913; Vol. 214, p. 295—317, 1914; Vol. 215, p. 161—179, 1915; Vol. 218, p. 1—118, 1919; Vol. 225, p. 45—91, 1925.

[3] Se saaledes: G. ANGENHEISTER: Handb. d. Phys. Bd. XV, p. 274. 1927.

[4] Eller udtrykt paa anden Maade: Til Opretholdelse af den nævnte

ningsevne maa i Hovedsagen søges højere oppe, mellem 130 og 200 m Højde.

Men nu melder der sig et nyt Spørgsmaal. SCHUSTER, CHAPMAN og de andre Forskere, der har beskæftiget sig med dette Spørgsmaal, taler ret og slet om elektrisk Ledningsevne. Nu er Sagen imidlertid den, at Atmosfærens elektriske Ledningsevne i en Retning parallel med Kraftlinierne i Jordens magnetiske Felt i Højderne 120, 140 og 160 km er omkring henholdsvis 300, 4000 og 9000 Gange større end Ledningsevnen vinkelret paa Feltets Retning.

Den totale Ledningsevne af Atmosfæren ved Dag, hvormed der er regnet i »P. R.W.«, er $450 \cdot 10^{-6}$ (e. m. e.; cm), naar Maalingen foregaar parallelt med de magnetiske Kraftlinier, men er kun $0.5 \cdot 10^{-6}$ (e. m. e.; cm) maalt vinkelret paa disse[1]. Den af CHAPMAN beregnede Ledningsevne, $25 \cdot 10^{-6}$ (e. m. e.; cm) ligger for saa vidt meget pænt indenfor disse Grænser. Men det vilde være ønskeligt, om SCHUSTER's og CHAPMAN's Teori gjordes til Genstand for en Revison under Hensyntagen til Atmosfærens vidt forskellige Ledningsevne i forskellige Retninger.

Det er formentlig ikke urimeligt at antage, at denne Ejendommelighed hos Atmosfærens elektriske Ledningsevne vil kunne bidrage til at forklare nogle af de magnetiske Ejendommeligheder, som man møder i de polare Egne.

Af astrofysiske Problemer, til hvis Løsning Radiobølgernes Udbredelsesforhold maaske vil blive i Stand til at

Ledningsevne i den nævnte Højde vilde der kræves en Energitilførsel, der var flere Gange Solens totale Udstraaling, naar Solen regnes som et sort Legeme med en Temperatur af 6000° K. En Opretholdelse af den nævnte Ledningsevne gennem en korpuskulær Straaling fra Solen maa derfor formentlig ogsaa anses for udelukket.

[1] $450 \cdot 10^{-6}$ (e. m. e.; cm) svarer til Ledningsevnen i en 7 mm tyk Kobberskal.

yde et Bidrag, skal nævnes Størrelsen af den af den meget gennemtrængende, kosmiske Straaling og af Stjernernes ultraviolette Straaling fremkaldte Ionisation. Den første Art af Ionisation er nemlig afgørende for Dæmpningen af meget lange Bølger. Den sidste er i Hovedsagen Aarsagen til, at lange Radiobølger er mere dæmpede ved Vinterdag end ved Sommerdag.

Der kan næppe være Tvivl om, at Kendskab til Radiobølgernes Udbredelsesforhold og Teorien herfor vil blive et af de mest effektive Hjælpemidler i Studiet af den højere Atmosfære, dens Sammensætning, Tryk, Temperatur og elektriske Ledningsevne, selv om det maaske nok er muligt, at man fra radioteknisk Side i nogen Grad er tilbøjelig til at overvurdere Sikkerheden af de dragne Slutninger og derfor ogsaa Betydningen af dette Hjælpemiddel.

Forelagt paa Mødet den 4. November 1927.
Færdig fra Trykkeriet den 19. December 1927.

MATHEMATISK-FYSISKE MEDDELELSER

UDGIVNE AF

DET KGL. DANSKE VIDENSKABERNES SELSKAB

6. BIND (Kr. 17,00):

Kr. Ø.

1. Nielsen, Niels: Sur l'opération itérative des Équations de Lagrange. 1924 .... 3.10
2. Urey, H. C.: On the Effect of perturbing Electric Fields on the Zeeman Effect of the Hydrogen Spectrum. 1924 .... 0.65
3. Bøggild, O. B.: On the Labradorization of the Feldspars. With one plate. 1924 .... 3.00
4. Pedersen, P. O.: Om elektriske Gnister. II. Eksperimentelle Undersøgelser over Gnistforsinkelse og Gnistdannelse. Med 7 Tavler. 1924 .... 4.30
5. Juel, C.: Über Flächen von Maximalindex. 1924 .... 1.25
6. Nielsen, Niels: Sur une Équation de Lagrange. 1924 .... 1.25
7. Hevesy, G. de: Recherches sur les propriétés du Hafnium. Avec 2 planches. 1925 .... 6.25
8. Bohr, Harald: Neuer Beweis eines allgemeinen Kronecker'schen Approximationssatzes. 1924 .... 0.50
9. Bjerrum, Niels and Ebert, Ludwig: On some recent Investigations concerning Mixtures of Strong Electrolytes (Transference Numbers and Amalgam Equilibria). 1925 .... 0.75
10. Landau, Edm.: Die Ungleichungen für zweimal differentiierbare Funktionen. 1925 .... 1.60

7. BIND (Kr. 17,70):

1. Bohr, Harald: Unendlich viele lineare Kongruenzen mit unendlich vielen Unbekannten. 1925 .... 1.40
2. Hartmann, Jul., and Trolle, Birgit: On Beat-phenomena in Cylindrical Tubes exposed to Sound-waves. With three plates. 1925 .... 2.85
3. Pauli, W. jr.: Ueber die Intensitäten der im elektrischen Feld erscheinenden Kombinationslinien. 1925 .... 0.65
4. Hardy, G. H. and Littlewood, J. E.: A theorem concerning series of positive terms, with applications to the theory of functions. 1925 .... 0.90
5. Steffensen, J. F.: On a Generalization of Nörlund's Polynomials. 1926 .... 1.00
6. Hartmann, Jul., and Trolle, Birgit: New investigation on the air jet generator for acoustic wawes. 1926 .... 2.40
7. Mollerup, Johs.: Sur l'approximation d'un nombre irrationnel par des carrés rationnels. 1926 .... 0.80
8. Nielsen, Niels: Sur certains développements d'une fonction holomorphe. 1926 .... 0.75

Kr. Ø.

9. Bjerrum, Niels: Untersuchungen über Ionenassoziation. I. Der Einfluss der Ionenassoziation auf die Aktivität der Ionen bei mittleren Assoziationsgraden. 1926 .... 2.00
10. Nielsen, Niels: Recherches sur les fonctions cylindriques et sur certaines fonctions analogues. 1926 .... 1.35
11. Hevesy, G.: On the Missing Element 87. 1926 .... 0.60
12. Kuhn, W.: Die Stärke der anomalen Dispersion in nicht leuchtendem Dampfe von Thallium und Cadmium. 1926 .... 3.75
13. Sugiura, Y. and Urey, H. C.: On the Quantum Theory explanation of the Anomalies in the 6th and 7th Periods of the Periodic Table. 1926 .... 1.15
14. Burrau, Øyvind: Berechnung des Energiewertes des Wasserstoffmolekel-Ions (H_2^+) im Normalzustand. 1927 .... 0.90
15. Knudsen, Martin: The Hot-wire Manometer. 1927 .... 0.80
16. Juel, C.: Om v. Staudt's Definitioner. 1927 .... 0.85
17. Pólya, G.: Über die algebraisch-funktionentheoretischen Untersuchungen von J. L. W. V. Jensen. 1927 .... 1.40

8. BIND (under Pressen):

1. Hille, Einar: On the Logarithmic Derivatives of the Gamma Function. 1927 .... 2.40
2. Rasch, G.: Beitrag zur Theorie der unvollständigen Gammafunktionen. Nach hinterlassenen Papieren von J. L. W. V. Jensen. 1927 .... 0.85
3. Knudsen, Martin: Thermal Molecular Pressure in Tubes. 1927 1.50
4. Pedersen, P. O.: Den højere Atmosfæres Sammensætning, Tryk, Temperatur og elektriske Ledningsevne i Belysning af Radiobølgernes Udbredelsesforhold. 1927 .... 0.60

Det Kgl. Danske Videnskabernes Selskab.
Mathematisk-fysiske Meddelelser. **VIII**, 5.

SOME REMARKS ON GENERALISATIONS OF ALMOST PERIODIC FUNCTIONS

BY

A. BESICOVITCH AND H. BOHR
(CAMBRIDGE) (COPENHAGEN)

KØBENHAVN
HOVEDKOMMISSIONÆR: ANDR. FRED. HØST & SØN, KGL. HOF-BOGHANDEL
BIANCO LUNOS BOGTRYKKERI
1927

Pris: Kr. 1,50.

Det Kgl. Danske Videnskabernes Selskabs videnskabelige Meddelelser udkommer fra 1917 indtil videre i følgende Rækker:

Historisk-filologiske Meddelelser,
Filosofiske Meddelelser,
Mathematisk-fysiske Meddelelser,
Biologiske Meddelelser.

Hele Bind af disse Rækker sælges 25 pCt. billigere end Summen af Bogladepriserne for de enkelte Hefter.

Selskabets Hovedkommissionær er *Andr. Fred. Høst & Søn*, Kgl. Hof-Boghandel, København.

Det Kgl. Danske Videnskabernes Selskab.
Mathematisk-fysiske Meddelelser. **VIII,** 5.

SOME REMARKS ON GENERALISATIONS OF ALMOST PERIODIC FUNCTIONS

BY

A. BESICOVITCH AND H. BOHR
(CAMBRIDGE) (COPENHAGEN)

KØBENHAVN
HOVEDKOMMISSIONÆR: ANDR. FRED. HØST & SØN, KGL. HOF-BOGHANDEL
BIANCO LUNOS BOGTRYKKERI
1927

Introduction.

The theory of almost periodic functions had its origin in the problem, what functions $f(x) = u(x) + iv(x)$ can be decomposed, in the interval $-\infty < x < \infty$, into pure oscillations, i. e. into oscillations of the form $e^{i\lambda x}$. The problem stated in these general terms has evidently no definite meaning until the notion of "decomposition" has been strictly defined and this of course can be done in many different ways.

The first and most primitive way of interpreting the term would perhaps be to regard as decomposable only those functions which can be represented as the sum of a finite number of oscillations:

$$(1) \qquad s(x) = \sum_{\nu=1}^{N} a_\nu e^{i\lambda_\nu x}.$$

We shall denote by A the class of such functions $s(x)$. But at the first attempt to develop the theory of unctions of this class we see that the definition is too narrow. Indeed the class A is not "closed" to limit processes, so that when working only with functions of the class A we should have to exclude from the start those operations which involve the idea of continuity in "functional space".

We must therefore close the class A, that is we must extend it to a larger class C (A) consisting of all functions $f(x)$ (including the functions of A itself) which are the limits of sequences of functions $s(x)$ of A. But here again the content of the class $C(\mathrm{A})$ depends on the kind of the limit process employed.

The simplest limit process is that of ordinary convergence: $f(x)$ is a limit-function of the class A if there is a sequence

$$s_1(x), \quad s_2(x), \ldots, \quad s_n(x), \ldots$$

of functions of A such that for every x

$$f(x) = \lim_{n \to \infty} s_n(x). \tag{2}$$

But the class $C(\mathrm{A})$ to which this limit process leads is found to be too wide, in the sense that practically none of the characteristic properties of the functions $s(x)$ (those relating to oscillations) are conserved. In fact, as Besicovitch [1][1] has shown, the above class $C(\mathrm{A})$ includes all bounded continuous functions.

The same is true even when we demand that the convergence shall be uniform in every finite interval.

We are therefore led to consider only limit processes which involve some sort of uniformity in the whole interval $-\infty < x < \infty$.

In the theory of a. p. (almost periodic) functions developed by Bohr in his papers in Acta Mathematica [1, 2, 3] the limit process employed was that of ordinary uniform convergence in the whole infinite interval $-\infty < x < \infty$. The class $C(\mathrm{A})$ corresponding to this limit process is the narrowest possible closure of the class A. But, as we shall see, the theory of larger closures $C(\mathrm{A})$ derived from more general processes of uniform convergence can be treated simply as generalisation of the theory of the above

[1] The list of papers quoted is given as an appendix.

closure C (A) in the sense that we can extend very many results directly, without repeating the arguments of their proof.

In § 1 of this paper we give a short outline of a part of the theory of a. p. functions. § 2 contains some general remarks on the generalisation of the theory to larger classes. Finally in § 3 we give in full detail an application of the principles lied down in § 2 to a particular generalisation of a. p. functions; for this purpose we choose the class of summable functions discussed by STEPANOFF in his interesting paper in Math. Ann. [1], where such generalisations were studied for the first time.

§ 1.

For convenience and also in order to bring out as clearly as possible the similarity between the definitions and proofs of this paragraph and these of § 3 we introduce the following notation:

By

$$U\text{-}\lim f_n(x) = f(x) \tag{3}$$

we mean that $f_n(x)$ tends to $f(x)$ uniformly in the whole interval $-\infty < x < \infty$. We shall call the upper bound of the difference $|f(x) - g(x)|$ for $-\infty < x < \infty$ the U-distance between the functions $f(x)$ and $g(x)$ and shall denote it by $D_U[f(x), g(x)]$; thus (3) can be written in the equivalent form

$$D_U[f_n(x), f(x)] \to 0, \quad \text{as } n \to \infty.$$

By $D_U[f(x)]$ we mean $D_U[f(x), 0]$, i. e. the upper bound of $|f(x)|$ in the interval $-\infty < x < \infty$.

A being as before the class of all finite sums

$$s(x) = \sum_{\nu=1}^{N} a_\nu e^{i\lambda_\nu x}$$

where the λ_ν are real and different, and the a_ν arbitrary complex numbers, we denote by $C_U(\text{A})$ the class of all functions $f(x)$ for each of which there is a sequence $s_n(x)$ of functions of A such that $f(x) = U\text{-lim}\, s_n(x)$.

We proceed to establish some properties of functions of the class $C_U(\text{A})$, which follow directly from this definition.

1°. Every function $f(x)$ of $C_U(\text{A})$ is bounded for $-\infty < x < \infty$.

For, given $f(x)$ we can choose an $s(x)$ so that

$$D_U[f(x),\ s(x)] < 1;$$

since $s(x)$ is plainly bounded, the result follows from the inequality

$$D_U[f(x)] \leqq D_U[f(x),\ s(x)] + D_U[s(x)].$$

2°. Every function $f(x)$ of $C_U(\text{A})$ is uniformly continuous in the whole interval $-\infty < x < \infty$.

For, given ε we can choose an $s(x)$ such that

$$D_U[f(x),\ s(x)] < \frac{\varepsilon}{3}.$$

$s(x)$ is evidently uniformly continuous; we can therefore choose δ so that

$$D_U[s(x+h),\ s(x)] < \frac{\varepsilon}{3} \quad \text{for } |h| < \delta.$$

From the inequality

$$D_U[f(x+h),\ f(x)] \leqq D_U[s(x+h),\ s(x)] + 2D_U[f(x),\ s(x)]$$

it now follows that for $|h| < \delta$

$$D_U[f(x+h),\ f(x)] < \varepsilon.$$

3^0. The sum and the product of two functions of $C_U(\mathrm{A})$ are again functions of $C_U(\mathrm{A})$.

This follows at once from the fact that the sum and the product of two functions $s(x)$ are functions $s(x)$.

In particular we observe that if $f(x)$ belongs to $C_U(\mathrm{A})$, then so does $f(x)\,e^{i\lambda x}$ for every real value of λ.

4^0. Every function $f(x)$ of $C_U(\mathrm{A})$ has a mean value $M\{f(x)\}$, i. e.

$$\lim_{T\to\infty} \frac{1}{T}\int_{\gamma}^{\gamma+T} f(x)\,dx$$

exists; this is even true uniformly in γ.

The property is obvious for any function $s(x)$ (the mean value in this case being the constant term in $s(x)$). Choosing $s(x)$ so that $D_U[f(x),\ s(x)]$ is "small" the result follows in the usual way from the inequality

$$\left|\frac{1}{T}\int_{\gamma}^{\gamma+T} f(x)\,dx - \frac{1}{T}\int_{\gamma}^{\gamma+T} s(x)\,dx\right| \leqq D_U[f(x),\ s(x)].$$

From the same inequality we see that if

(4) $$U\text{-lim}\, s_n(x) = f(x)$$

then

$$M\{s_n(x)\} \to M\{f(x)\}, \quad \text{as } n\to\infty,$$

and further, for any real value of λ,

(5) $$M\{s_n(x)\,e^{-i\lambda x}\} \to M\{f(x)\,e^{-i\lambda x}\}, \quad \text{as } n\to\infty.$$

For (4) implies

$$U\text{-lim}\, s_n(x)\,e^{-i\lambda x} = f(x)\,e^{-i\lambda x}.$$

5^0. For any function $f(x)$ of $C_U(\mathrm{A})$ the mean value

$$M\{f(x)\,e^{-i\lambda x}\} = a(\lambda)$$

differs from zero for at most an enumerable set of values of λ.

Let $s_1(x)$, $s_2(x)$, ... be a sequence of functions of A such that

$$U\text{-}\lim s_n(x) = f(x).$$

For each function $s_n(x)$ the mean value $M\{s_n(x)\,e^{-i\lambda x}\}$ differs from zero only for a finite set of values of λ, namely those occurring as the exponents in the polynomial expression for $s_n(x)$. It follows that, except in at most an enumerable set of λ's, $M\{s_n(x)\,e^{-i\lambda x}\}$ is zero for every n, and so by (5) that $a(\lambda) = M\{f(x)\,e^{-i\lambda x}\}$ is zero except for at most an enumerable set of values of λ.

We can now denote these λ's by

$$\Lambda_1,\ \Lambda_2,\ \ldots$$

and the corresponding $a(\lambda)$ by

$$A_1,\ A_2,\ \ldots.$$

We express this symbolically by writing

$$f(x) \sim \sum A_\nu e^{i\Lambda_\nu x} \tag{6}$$

and we call the series (finite or infinite) on the right the Fourier series of the function $f(x)$.

The Fourier series of a fonction $s(x)$ of A evidently coincides with its polynomial expression (1). From (5) we see that, as $n \to \infty$, the polynomial expression of $s_n(x)$ goes over by a "formal" limit passage into the Fourier series of $f(x)$; this already shows that Fourier series are likely to play an important part in the study of functions of $C_U(A)$.

We next consider a property of functions of $C_U(A)$ of a different kind. We call a real number $\tau = \tau(\varepsilon)$ a translation number of the function $f(x)$ belonging to ε if

$$D_U\left[f(x+\tau),\ f(x)\right] \leqq \varepsilon .$$

Corresponding to any given function $s(x) = \sum_1^N a_\nu e^{i\lambda_\nu x}$ of the class A there exist for any given $\varepsilon > 0$ an infinite number of translation numbers, and the set of these numbers $\tau = \tau(\varepsilon)$ is even "relatively dense" in the sense that any interval of a certain length $l = l(\varepsilon)$ contains at least one such number $\tau(\varepsilon)$. This is an immediate consequence of the BOHL-WENNBERG theorem on diophantine approximations (see f. inst. Bohr [1], p. 120), which states that for an arbitrarily small δ the N diophantine inequalities[1]

$$|\lambda_\nu \tau| < \delta \pmod{2\pi} \qquad (\nu = 1, 2, \ldots, N)$$

have relatively dense solutions with respect to τ. In the ordinary way (i. e. by approaching the function $f(x)$ by a function $s(x)$ such that $D_U\left[f(x),\ s(x)\right]$ is small) we see that the functions $f(x)$ of $C_U(\mathrm{A})$ also possess the above property, namely that for every $\varepsilon > 0$ the ε-translation numbers exist and are relatively dense.

Functions which are continuous in $-\infty < x < \infty$ and possess this property are said to be almost periodic; we have just seen that every function of A, and even every function of $C_U(\mathrm{A})$, is almost periodic.

The main result of the theory of a. p. (almost periodic) functions is that the converse of the last statement is also true; every a. p. function is a function of $C_U(\mathrm{A})$, so that

The class $C_U(A)$ is identical with the class of a. p. functions.

[1] By $|a| < b \pmod{c}$, where a, b, c are real and b and c positive, we mean that there exists an integral n such that $|a - nc| < b$.

Naturally we shall not enter the proof of this theorem which would involve the development of almost the whole theory of a. p. functions. In proving the identity of the class $C_U(\mathrm{A})$ with the class of a. p. functions Bohr had first to show that the latter class possesses all the above properties 1°, . . . , 5° of the class $C_U(\mathrm{A})$; and though this was not the main difficulty of the investigation, it nevertheless involved considerations of a different character from the immediate deductions employed above in establishing these properties for the class $C_U(\mathrm{A})$. The main difficulty to be overcome was the proof of the "fundamental theorem" (Parseval's theorem) namely

$$M\{|f(x)|^2\} = \sum |A_\nu|^2,$$

from which follows as an immediate corollary the "uniqueness theorem":

An a. p. function is uniquely determined by its Fourier series, i. e. two different a. p. functions cannot have the same Fourier series.

When the identity of the class $C_U(\mathrm{A})$ and the class of a. p. functions has been established the question naturally arises: given an a. p. function $f(x)$, actually to find a sequence of functions $s_n(x)$ such that

$$U\text{-lim}\, s_n(x) = f(x).$$

A method of obtaining such a sequence of functions was given by Bohr in his second paper in Acta Math.; his sums $s_n(x)$ contained as exponents only exponents from the Fourier series of $f(x)$, a fact of importance in the extension of the theory to functions of a complex variable. An essentially simpler method of obtaining such approximation functions $s_n(x)$ was given by Bochner [1], who succeded in extending the Fejér summation method of classical Fourier theory to the class of a. p. functions. Like Bohr, he started from the representation of the "Fourier exponents" Λ_ν with the help of a "base" $\alpha_1, \alpha_2, \ldots$. By a base we mean a sequence of linearly independent[1] positive numbers $\alpha_1, \alpha_2, \ldots$ (which generally is enumerable but in particular cases may be finite) such thal every the α's Λ_ν may be expressed as a finite linear form in exponent with rational coefficients,

$$\Lambda_\nu = r_{\nu,1}\,\alpha_1 + r_{\nu,2}\,\alpha_2 + \ldots + r_{\nu,q_\nu}\,\alpha_{q_\nu}.$$

Fejér in his summation of Fourier series of pure periodic functions $f(x)$, with period 2π, used as approximation sums the expressions

$$s_n(x) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x+t)\,\Pi_n(t)\,dt = M\{f(x+t)\,\Pi_n(t)\},$$

where the "kernel" $\Pi_n(t)$ was given by

$$\Pi_n(t) = \sum_{\nu=-n}^{n}\left(1-\frac{|\nu|}{n}\right)e^{-i\nu t} = \frac{1}{n}\left(\frac{\sin n\frac{t}{2}}{\sin\frac{t}{2}}\right)^2.$$

Bochner replaced Fejér's simple kernel by a finite product of such kernels

$$\Pi(t) = \Pi_{n_1}(\beta_1 t)\ldots\Pi_{n_p}(\beta_p t) =$$

$$\sum_{\substack{-n_1\leq\nu_1\leq n_1\\ \cdots\cdots\\ -n_p\leq\nu_p\leq n_p}}\left(1-\frac{|\nu_1|}{n_1}\right)\ldots\left(1-\frac{|\nu_p|}{n_p}\right)e^{-i(\nu_1\beta_1+\ldots+\nu_p\beta_p)t}$$

[1] $\alpha_1, \alpha_2, \ldots$ are said to be linearly independent if no equation of the form

$$r_1\alpha_1 + r_2\alpha_2 + \ldots + r_N\alpha_N = 0$$

holds, where $N \geq 1$ and the r's are rational, not all naught.

where the β's are linearly independent numbers. This composite kernel has the same characteristic properties as the Fejér kernel — it is always positive and its mean-value $M\{\Pi(t)\}$ is equal to 1 (the constant term in the polynomial expansion of $\Pi(t)$ being 1 on account of the linear independence of the β's). Bochner considers an expression of the form

$$M\left\{f(x+t)\,\Pi_{n_1}\left(\frac{\alpha_1}{N_1!}\,t\right)\,\ldots\,\Pi_{n_p}\left(\frac{\alpha_p}{N_p!}\,t\right)\right\}$$

which, since

$$f(x+t) \sim \sum A_\nu e^{i\Lambda_\nu x}\cdot e^{i\Lambda_\nu t},$$

is equal to the finite sum

$$s(x) = \sum_{\substack{-n_1 \leq \nu_1 \leq n_1 \\ \cdots\cdots\cdots \\ -n_p \leq \nu_p \leq n_p}} \left(1-\frac{|\nu_1|}{n_1}\right)\,\ldots\,\left(1-\frac{|\nu_p|}{n_p}\right) A_\nu e^{i\Lambda_\nu x}$$

where

(7) $$\Lambda_\nu = \frac{\nu_1}{N_1!}\alpha_1 + \ldots + \frac{\nu_p}{N_p!}\alpha_p$$

(and A_ν is to be interpreted as zero when the linear combination (7) of α's is not an exponent in the Fourier series of $f(x)$). Bochners result is:

The sum $s(x)$ *tends uniformly to* $f(x)$, *as* $p\to\infty$, $N_1\to\infty$, $N_2\to\infty$, ... *and* $\frac{n_1}{N_1!}\to\infty$, $\frac{n_2}{N_2!}\to\infty$, ..., in other words, provided the limit process is carried out in such a way that every exponent Λ_ν occurs sooner or later in $s(x)$.

As we shall have to prove in § 3 an analogous theorem for a more general class of functions, we next give a proof of Bochners theorem which, though perhaps not so elegant as Bochner's own, is better adapted to generalisation. Like

Bochner's, our proof depends on the fundamental theorem. We start from the following property of a. p. functions, which (see Bohr, II p. 110) is a fairly direct consequence of the fundamental theorem:

"To any ε corresponds an integer M and a positive number η such that every solution t of the system of diophantine inequalities

$$|\Lambda_1 t| < \eta \pmod{2\pi}, \ \ldots, \ |\Lambda_M t| < \eta \pmod{2\pi}$$

is a translation number $\tau(\varepsilon)$ of the given a. p. function".

Since each exponent Λ_ν can be represented as a finite linear form in the α's with rational coefficients, we have the immediate corollary:

To any ε correspond integers P and Q and a positive number $\delta < \pi$ such that every solution t of the system of diophantine inequalities

$$(8) \qquad \left|\frac{\alpha_1}{Q} t\right| < \delta \pmod{2\pi}, \ \ldots, \ \left|\frac{\alpha_P}{Q} t\right| < \delta \pmod{2\pi}$$

is a translation number $\tau\left(\frac{\varepsilon}{2}\right)$ of the given a. p. function.

Let I_1 denote the set of all values of t satisfying (8) and I_2 the set of all other values of t. We can write the kernel

$$\Pi(t) = \Pi_{n_1}\left(\frac{\alpha_1}{N_1!} t\right) \ldots \Pi_{n_p}\left(\frac{\alpha_p}{N_p!} t\right)$$

as the sum of two kernels

$$\Pi(t) = \Pi'(t) + \Pi''(t)$$

where

$$\Pi'(t) = \Pi(t), \quad \Pi''(t) = 0 \qquad \text{for } t \subset I_1,$$
$$\Pi'(t) = 0, \quad \Pi''(t) = \Pi(t) \qquad \text{for } t \subset I_2.$$

We may further assume that $p > P$ and $N_1, \ldots, N_P$ all $> Q$, where P and Q are the integers (depending on ε) which occur in (8).

From

$$s(x)-f(x) = M\{(f(x+t)-f(x))\,\Pi(t)\}$$

we have

$$|s(x)-f(x)| \leq \overline{M}\{|f(x+t)-f(x)|\,\Pi'(t)\} + \overline{M}\{|f(x+t)-f(x)|\,\Pi''(t)\}$$

where $\overline{M}\{g(t)\}$ for a positive function $g(t)$ (which need not necessarily possess a mean value) denotes $\limsup \frac{1}{2T}\int_{-T}^{T} g(t)\,dt$ as $T \to \infty$. Thus

$$(9)\quad \begin{cases} D_U[s(x),\, f(x)] \leq \overline{M}\{D_U[f(x+t),\, f(x)]\,\Pi'(t)\} \\ \qquad\qquad + \overline{M}\{D_U[f(x+t),\, f(x)]\,\Pi''(t)\} \end{cases}$$

where the distance D_U on the right hand side is taken with respect to x, t remaining fixed.

Since for all values of t belonging to I_1

$$D_U[f(x+t),\, f(x)] \leq \frac{\varepsilon}{2}$$

while for all other values of t

$$\Pi'(t) = 0$$

we have at once

$$(10)\quad \begin{cases} \overline{M}\{D_U[f(x+t),\, f(x)]\,\Pi'(t)\} \\ < \frac{\varepsilon}{2}\overline{M}\{\Pi'(t)\} \leq \frac{\varepsilon}{2}M\{\Pi(t)\} = \frac{\varepsilon}{2}. \end{cases}$$

Writing $K = D_U[f(x)]$, we have for the second term on the right hand side of (9)

$$(11)\quad \overline{M}\{D_U[f(x+t),\, f(x)]\,\Pi''(t)\} \leq 2K\overline{M}\{\Pi''(t)\}.$$

Let $I_{2,q}$ $(q = 1, 2, \ldots, P)$ denote the set of values t, at which the q-th of the inequalities (8) is n o t satisfied.

This set $I_{2,q}$ consists of the whole t-axis with the exception of small intervals of length $2\,l_q = 2\,\frac{\delta Q}{\alpha_q}$ with their centres at the zeros of $\sin\frac{\alpha_q t}{2\,Q}$. Since $N_q > Q$ and $N_q!$ is thus a multiple of Q, we observe that the zeros of $\sin\frac{\alpha_q t}{2\,N_q!}$ are included among these of $\sin\frac{\alpha_q t}{2\,Q}$. Putting

$$\overset{*}{\Pi}_{n_q}\left(\frac{\alpha_q}{N_q!}\,t\right) = \Pi_{n_q}\left(\frac{\alpha_q}{N_q!}\,t\right) \text{ for } t \subset I_{2,q}, \text{ and} = 0 \text{ for } t \text{ outside } I_{2,q}$$

we evidently have for $-\infty < t < \infty$

$$(12)\quad \Pi''(t) \leq \sum_{q=1}^{P} \overset{*}{\Pi}_{n_q} \cdot \Pi_{n_1}\Pi_{n_2}\ldots\Pi_{n_{q-1}}\Pi_{n_{q+}}\ \ldots\Pi_{n_p}$$

since for any t in I_2 at least one of the inequalities (8) is not satisfied and thus at least one of the P term of this sum is equal to $\Pi(t)$. Therefore

$$\overline{M}\{\Pi''(t)\} \leq \sum_{q=1}^{P} M\{\overset{*}{\Pi}_{n_q} \cdot \Pi_{n_1}\Pi_{n_2}\ldots\Pi_{n_{q-1}}\Pi_{n_{q+1}}\ldots\Pi_{n_p}\}.$$

As the product $\Pi_{n_1}\ldots\Pi_{n_{q-1}}\Pi_{n_{q+1}}\ldots\Pi_{n_p}$ is a finite sum $s(x) = \sum a_\nu e^{i\lambda_\nu x}$ with the constant term 1 whose other exponents do not satisfy any relation of the form $\lambda_\nu + r\,\alpha_q = 0$ with rational r, the mean value under the sign of summation is simply equal to the mean value $M\{\overset{*}{\Pi}_{n_q}\}$, both being evidently equal to the constant term in the ordinary Fourier series for the pure periodic function $\overset{*}{\Pi}_{n_q}$ $\left(\text{with period } 2\,\omega_q = 2\,\pi\frac{N_q!}{\alpha_q}\right)$. Thus

$$(13)\qquad \overline{M}\{\Pi''(t)\} \leq \sum_{q=1}^{P} M\left\{\overset{*}{\Pi}_{n_q}\left(\frac{\alpha_q}{N_q!}\,t\right)\right\}.$$

Further

$$M\left\{\overset{*}{\Pi}_{n_q}\left(\frac{\alpha_q}{N_q!}t\right)\right\} \leq \frac{1}{\omega_q}\int_{l_q}^{\omega_q} \Pi_{n_q}\left(\frac{\alpha_q}{N_q!}t\right)dt$$

$$= \frac{1}{\omega_q}\int_{l_q}^{\omega_q}\frac{1}{n_q}\frac{\sin^2\frac{n_q\alpha_q}{2N_q!}t}{\sin^2\frac{\alpha_q}{2N_q!}t}dt \leq \frac{1}{n_q\omega_q}\int_{l_q}^{\omega_q}\frac{1}{\left(\frac{1}{2}\cdot\frac{\alpha_q t}{2N_q!}\right)^2}dt$$

$$< \frac{16(N_q!)^2}{n_q\omega_q\alpha_q^2}\int_{l_q}^{\infty}\frac{dt}{t^2} = \frac{16(N_q!)^2}{n_q\omega_q\alpha_q^2 l_q} = \frac{16}{\pi\delta Q}\cdot\frac{N_q!}{n_q} < \frac{6}{\delta Q}\,\frac{N_q!}{n_q}.$$

Thus by (13)

$$\overline{M}\{\Pi''(t)\} < \frac{6}{\delta Q}\sum_{q=1}^{P}\frac{N_q!}{n_q}.$$

On putting

$$n_q > \frac{24\,KP}{\delta Q\,\varepsilon}N_q! \qquad (q = 1, \ldots, P)$$

we have therefore

$$2K\overline{M}\{\Pi''(t)\} < \frac{\varepsilon}{2}$$

and thus finally by (9), (10), (11)

$$D_U[s(x), f(x)] < \varepsilon$$

for $p > P$, $N_q > Q$, $n_q > CN_q!$ $(q = 1, \ldots, P)$, where P, Q, C depend only on ε

Q. E. D.

§ 2.

A natural way of generalising the theory of a. p. functions is to use, in closing the class A, a limit process more general than the simple U-lim employed in § 1. Let G denote such a limit process. Then the first and main problem which arises is to determine the generalised almost periodic properties which characterise the class $C_G(\mathrm{A})$, the closure of A by the limit process G.

But once this problem has been solved in the original theory its solution for generalised classes may be reduced to considerations concerning only the limit process G and its effect on almost-periodicity. Stepanoff himself proceeded along these lines in the paper already quoted and thus escaped entering once more into the difficulties involved in establishing the fundamental theorem or the uniqueness theorem. It is however possible to go further in this way. We have only to take into account and use to the full the fact that the U-lim process employed in the original theory, while general enough to bring out the main properties of almost periodicity, is at the same time the "narrowest" of all limit processes of the kind described in the introduction. It is the narrowest in the sense that the closure of the class A by any limit process G coincides with the closure by G of the class $C_U(\mathrm{A})$ already closed by the U-process, i. e. in symbols

$$C_G(\mathrm{A}) = C_G\big(C_U(\mathrm{A})\big).$$

Thus

$$C_G(\mathrm{A}) = C_G(\text{a. p. functions})$$

which shows that the problem of characterising the class $C_G(\mathrm{A})$ by almost periodic properties is equivalent to the investigation of the effect of the limit process G on ordinary almost periodicity.

Once the character of the almost periodicity corresponding to a given process G has been determined, the next main question is to find an "algorithm" which, applied to a function $f(x)$ possessing this type of almost periodicity, will lead to a sequence $s_n(x)$ of functions of A which approach the function $f(x)$ in the sense of the given limit process G, i. e. for which

$$G\text{-lim}\, s_n(x) = f(x), \qquad \text{as } n \to \infty.$$

The natural method of treating this question will generally be by establishing the existence of the Fourier series of $f(x)$ and applying to it a suitable method of summation.

We now carry out an investigation of the type just outlined, obtaining finally what is perhaps the most natural generalisation of the class of a. p. functions, namely the class of integrable (L) functions considered by Stepanoff. The very fact that this class of functions — defined by Stepanoff himself through generalised almost periodic properties — can also be characterised as the closure of the class of (ordinary) a. p. functions by a certain limit process (the S-process given below) has already been pointed out by Bochner [2].

§ 3.

The functions which we have to consider in this section are assumed to be defined almost everywhere in the whole interval $-\infty < x < \infty$ and to be integrable (L) over any finite interval. We begin by introducing a notation analogous to that of § 1. We say that the function $f(x)$ is the S-limit (Stepanoff limit) of the sequence $f_n(x)$ $(n = 1, 2, \ldots)$ and write

(14) $$f(x) = S\text{-lim}\, f_n(x)$$

if

$$\underset{-\infty < x < \infty}{\text{Upper bound}} \int_x^{x+1} |f(\xi) - f_n(\xi)|\, d\xi \to 0, \quad \text{as } n \to \infty^{1}.$$

By the S-distance $D_S[f(x), g(x)]$ between the functions $f(x)$ and $g(x)$ we mean

[1] Evidently the definition of the S-limit will not be altered if we replace $\int_x^{x+1} |\ |\, d\xi$ by $\int_x^{x+k} |\ |\, d\xi$, where k is an arbitrary positive constant.

$$\underset{-\infty < x < \infty}{\text{Upper bound}} \int_x^{x+1} |f(\xi) - g(\xi)|\, d\xi;$$

thus the equation (14) is equivalent to

$$D_S[f(x),\ f_n(x)] \to 0, \qquad \text{as } n \to \infty.$$

Further by $D_S[f(x)]$ we mean $D_S[f(x),\ 0]$, i. e. the upper bound of $\int_x^{x+1} |f(\xi)|\, d\xi$ in the interval $-\infty < x < \infty$.

This section is devoted to the study of the class $C_S(\mathrm{A})$, i. e. the closure of the class A by the S-limit process.

As in § 1 we begin by establishing a number of properties of the functions of $C_S(\mathrm{A})$ which follow directly from their definition as S-limits of functions $s(x)$ of A. The deduction of these properties is on the same lines as before, the only difference being that the U-processes of § 1 are here replaced by S-processes. Our notation is essentially that of Bochner [2].

1^0. Every function $f(x)$ of $C_S(\mathrm{A})$ is "S-bounded" in $-\infty < x < \infty$, i. e. $D_S[f(x)]$ is finite.

The property follows from the inequality

$$D_S[f(x)] \leqq D_S[f(x),\ s(x)] + D_S[s(x)]$$

in the same way as in § 1 ($s(x)$, being bounded, is a fortiori S-bounded).

2^0. Every function $f(x)$ of $C_S(A)$ is "S-uniformly continuous" in $-\infty < x < \infty$, i. e.

$$D_S[f(x+h),\ f(x)] < \varepsilon \qquad \text{for } |h| < \delta.$$

In the same way as before the property follows from the inequality

$$D_S[f(x+h),\ f(x)] \leqq D_S[s(x+h),\ s(x)] + 2D_S[f(x),\ s(x)].$$

3,1°. Evidently the sum of two functions of $C_S(A)$ is again a function of $C_S(A)$.

3,2°. If $f(x)$ belongs to $C_S(A)$, then the product $f(x)e^{i\lambda x}$ belongs to $C_S(A)$ for every real λ.

For

$$D_S\left[f(x)e^{i\lambda x},\ s(x)e^{i\lambda x}\right] = D_S\left[f(x),\ s(x)\right].$$

(The property does not hold for the product of any two functions of $C_S(A)$, since such a product may not be integrable).

4°. Every function $f(x)$ of $C_S(A)$ has a mean value

$$M\{f(x)\} = \lim \frac{1}{T}\int_{\gamma}^{\gamma+T} f(x)\,dx, \qquad \text{as } T \to \infty$$

(and the limit even exists uniformly in γ).

This follows from the inequality

$$\left|\frac{1}{T}\int_{\gamma}^{\gamma+T} f(x)\,dx - \frac{1}{T}\int_{\gamma}^{\gamma+T} s(x)\,dx\right| \leqq \frac{1}{T}\int_{\gamma}^{\gamma+[T]+1} |f(x)-s(x)|\,dx$$

$$\leqq \frac{[T]+1}{T} D_S\left[f(x),\ s(x)\right] \leqq 2D_S\left[f(x),\ s(x)\right] \qquad \text{for } T \geqq 1.$$

From the same inequality we see that if

$$S\text{-}\lim s_n(x) = f(x)$$

then

$$M\{s_n(x)\} \to M\{f(x)\}$$

and, more generally,

$$M\{s_n(x)e^{i\lambda x}\} \to M\{f(x)e^{i\lambda x}\}$$

as $n \to \infty$.

5°. Repeating word for word the argument of § 1 we can now establish the existence of a Fourier series, write

$$f(x) \sim \sum A_\nu e^{i\Lambda_\nu x},$$

in the same sense as before, and assert that when $S\text{-lim}\, s_n(x) = f(x)$ the Fourier series of $f(x)$ may be obtained from the expressions of $s_n(x)$ by a formal limit passage.

We now pass to the main part of the theory of the class $C_S(\mathrm{A})$, the investigation of the almost periodic properties characteristic of the class.

We shall in the future refer to the translation numbers defined in § 1 as U-translation numbers. We now introduce another kind of translation numbers, defined as follows: A number τ is said to be an S-translation number of the function $f(x)$ belonging to ε, if

$$D_S\left[f(x+\tau),\ f(x)\right] \leqq \varepsilon.$$

We call $f(x)$ an S. a. p. function if for every positive ε there exists a relatively dense set (in the sense of § 1) of S-translation numbers $\tau(\varepsilon)$ of $f(x)$.

Theorem. *The class $C_S(\mathrm{A})$ is identical with the class of S. a. p. functions.*

In proving the identity of the class $C_S(\mathrm{A})$ with the class of S. a. p. functions we shall, so far as the latter class is concerned, use nothing but its definition; we do not even need to begin, as Stepanoff did in his development of the theory, by establishing the elementary properties, just deduced for $C_S(\mathrm{A})$ as immediate consequences of its definition. In accordance with § 2 we base our proof on the fact that the class $C_S(\mathrm{A})$ is identical with the class

$$C_S\ \text{(a. p. functions)}.$$

For convenience we shall in future denote an a. p. function by $\sigma(x)$.

1. That every function $f(x)$ of $C_S(\mathrm{A})$ is an S. a. p. function, is obvious. We have only to choose $\sigma(x)$ in the inequality

$$D_S\left[f(x+\tau),\ f(x)\right] \leq D_S\left[\sigma(x+\tau),\ \sigma(x)\right] + 2D_S\left[f(x),\ \sigma(x)\right]$$

so that $D_S\left[f(x),\ \sigma(x)\right] < \frac{\varepsilon}{3}$, to ensure that every U-translation number of $\sigma(x)$ (which is a fortiori an S-translation number of $\sigma(x)$) belonging to $\frac{\varepsilon}{3}$ shall be an S-translation number of $f(x)$ belonging to ε.

2. We now proceed to the proof of the converse result, namely that every S. a. p. function is the S-limit of a sequence of functions $\sigma(x)$. For this purpose we consider the functions

$$\varphi_\delta(x) = \frac{1}{\delta}\int_x^{x+\delta} f(\xi)\, d\xi$$

already studied by Stepanoff. We shall prove first that $\varphi_\delta(x)$ (which is evidently continuous) is an a. p. function $\sigma(x)$, and secondly that

$$f(x) = S\text{-}\lim \varphi_\delta(x), \qquad \text{as } \delta \to 0.$$

We may suppose $\delta < 1$.

To prove the first statement we observe that every S-translation number of $f(x)$ which belongs to $\varepsilon\delta$ is also a U-translation number of $\varphi_\delta(x)$ belonging to ε. Given such an S-translation number $\tau = \tau(\varepsilon\delta)$ we have in fact for every x

$$|\varphi_\delta(x+\tau) - \varphi_\delta(x)| = \frac{1}{\delta}\left|\int_x^{x+\delta}\left(f(\xi+\tau) - f(\xi)\right) d\xi\right|$$

$$\leq \frac{1}{\delta}\int_x^{x+1} |f(\xi+\tau) - f(\xi)|\, d\xi \leq \varepsilon.$$

To prove the second statement we first observe that any S-translation number τ of $f(x)$ belonging to ε is also an S-translation number of $\varphi_\delta(x)$ belonging to 2ε. For we have for every x_0

$$\int_{x_0}^{x_0+1} |\varphi_\delta(x+\tau)-\varphi_\delta(x)|\,dx \leq \frac{1}{\delta}\int_{x_0}^{x_0+1} dx \int_{x}^{x+\delta} |f(\xi+\tau)-f(\xi)|\,d\xi$$

$$\leq \frac{1}{\delta}\int_{x_0}^{x_0+1+\delta} |f(\xi+\tau)-f(\xi)|\,d\xi \int_{\xi-\delta}^{\xi} dx \leq \int_{x_0}^{x_0+2} |f(\xi+\tau)-f(\xi)|\,d\xi \leq 2\varepsilon.$$

What we have to prove is that $D_S[f(x),\ \varphi_\delta(x)] < \varepsilon$ for $\delta < \delta_0(\varepsilon)$, i. e. that for every x_0

$$(15)\qquad \int_{x_0}^{x_0+1} |f(x)-\varphi_\delta(x)|\,dx < \varepsilon \qquad \text{for } \delta < \delta_0(\varepsilon).$$

Let $l = l\left(\frac{\varepsilon}{4}\right)$ be a number such that every interval of length l contains an S-translation number $\tau = \tau\left(\frac{\varepsilon}{4}\right)$ of $f(x)$ which (as we have just seen) is also an S-translation number, belonging to $\frac{\varepsilon}{2}$, of every $\varphi_\delta(x)$. Corresponding to the value of x_0 in (15) we select a translation number $\tau\left(\frac{\varepsilon}{4}\right)$ so that the point $x_0+\tau$ lies in the interval $(0, l)$. Then

$$\int_{x_0}^{x_0+1} |f(x)-\varphi_\delta(x)|\,dx \leq \int_{x_0}^{x_0+1} |f(x)-f(x+\tau)|\,dx$$

$$+\int_{x_0}^{x_0+1} |f(x+\tau)-\varphi_\delta(x+\tau)|\,dx + \int_{x_0}^{x_0+1} |\varphi_\delta(x+\tau)-\varphi_\delta(x)|\,dx$$

$$\leq \frac{\varepsilon}{4} + \int_{x_0+\tau}^{x_0+\tau+1} |f(x)-\varphi_\delta(x)|\,dx + \frac{\varepsilon}{2} \leq \frac{3}{4}\varepsilon + \int_0^{l+1} |f(x)-\varphi_\delta(x)|\,dx.$$

Thus the proof will be complete when we have established the following simple proposition, which we state as an independent lemma since it may be of use in other problems involving the smoothing of an integrable function.

Lemma. *Let $f(x)$ be a function integrable (L) in a finite interval (a, b), and let*

$$\varphi_\delta(x) = \frac{1}{\delta}\int_x^{x+\delta} f(\xi)\,d\xi \qquad (a < x < b-\delta).$$

Then, for any $\beta < b$,

$$\lim_{\delta \to 0}\int_a^\beta |f(x) - \varphi_\delta(x)|\,dx = 0.$$

We denote the number $b-\beta$ by d. Then $\varphi_\delta(x)$ is defined in the interval (a, β) for every $\delta \leqq d$. By the theorem on the differentiation of a Lebesgue integral we have that $\varphi_\delta(x) \to f(x)$ as $\delta \to 0$ almost everywhere in (a, β).

We first prove that to any $\varepsilon > 0$ corresponds an $\eta > 0$ such that

$$\int_{\mathrm{E}} |\varphi_\delta(x)|\,dx < \varepsilon \tag{16}$$

for all values of $\delta\,(\leqq d)$ and for every set $\mathrm{E} \subset (a, \beta)$ such that $m\mathrm{E} < \eta$. We write

$$\int_{\mathrm{E}} |\varphi_\delta(x)|\,dx \leqq \frac{1}{\delta}\int_{\mathrm{E}} dx \int_x^{x+\delta} |f(\xi)|\,d\xi = \frac{1}{\delta}\iint_G |f(\xi)|\,dx\,d\xi$$

where G denotes the two-dimensional set of points in the x, ξ plane

$$x \subset \mathrm{E}, \quad x < \xi < x+\delta$$

whose measure

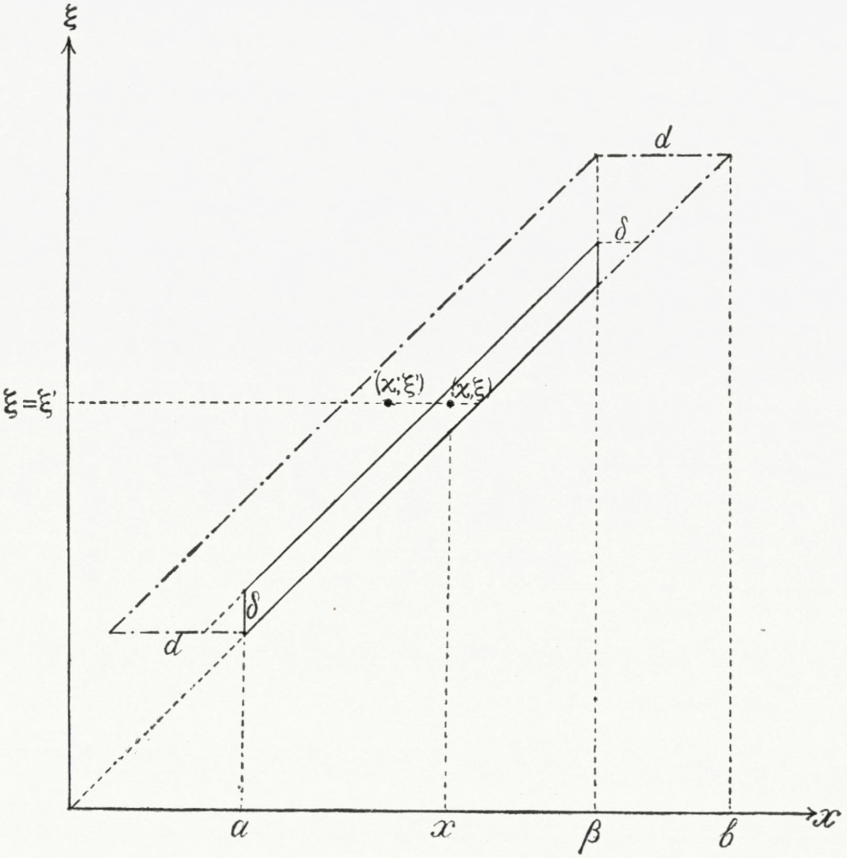
$$m\,G = \delta \cdot m\,\mathrm{E}.$$

On making the simple transformation

$$\xi = \xi', \quad x' - \xi' = \frac{d}{\delta}(x-\xi)$$

the set G goes over to a set G' in the x', ξ' plane, of measure

(17) $$mG' = \frac{d}{\delta} \cdot mG = d \cdot m\mathrm{E},$$



which evidently lies inside the constant (independent of δ) parallelogram P whose vertices are the points $(a-d,\ a)$, $(a,\ a)$, $(b,\ b)$, $(b-d,\ b)$. And

(18) $$\frac{1}{\delta}\iint\limits_{G} |f(\xi)|\,dx\,d\xi = \frac{1}{d}\iint\limits_{G'} |f(\xi')|\,dx'\,d\xi'.$$

Since the function $f(\xi')$, regarded as a function of the two variables x', ξ', is integrable in the parallelogram P, the integral on the right can be made less than ε by taking $m\mathrm{E}$ sufficiently small ($< \eta$), for by (17) this makes mG' "small".

(16) being established, the proof of the lemma may be completed in the usual way: We choose $\eta' \leqq \eta$ so that $\int\limits_{\mathrm{E}} |f(x)|\,dx < \varepsilon$ for $m\mathrm{E} < \eta'$, and then chose δ_0 so small that for $\delta < \delta_0$ the set $F = F_\delta$ of points of (α, β) at which $|f - \varphi_\delta| < \dfrac{\varepsilon}{\beta - \alpha}$ is of measure $> \beta - \alpha - \eta'$, and consequently the complementary set $C(F)$ of measure $< \eta'$. Then for $\delta < \delta_0$

$$\int_\alpha^\beta |f(x) - \varphi_\delta(x)|\,dx \leqq \int\limits_F |f(x) - \varphi_\delta(x)|\,dx + \int\limits_{C(F)} |f(x)|\,dx$$

$$+ \int\limits_{C(F)} |\varphi_\delta(x)|\,dx < \frac{\varepsilon}{\beta - \alpha}(\beta - \alpha) + \varepsilon + \varepsilon = 3\,\varepsilon.$$

When the identity of the class $C_S(\mathrm{A})$ with the class of S. a. p. functions has been established the next problem which arises is to find a simple algorithm which, applied to an arbitrary S. a. p. function $f(x)$, gives a sequence of finite sums $s_n(x)$ tending to $f(x)$ in the sense of the limit process characteristic of the class, i. e. such that

$$S\text{-lim}\, s_n(x) = f(x), \qquad \text{as } n \to \infty.$$

The above proof evidently provides a mean of constructing such a sequence. For the above functions $\varphi_\delta(x) = \dfrac{1}{\delta}\displaystyle\int_x^{x+\delta} f(\xi)\,d\xi$, being ordinary a. p. unctions, can in accordance with § 1 be approched within an arbitrarily small U-distance (and a fortiori within an arbitrarily small S-

distance) by Fejér sums of their Fourier series. Therefore by first taking δ sufficiently small and then approximating sufficiently closely to $\varphi_\delta(x)$ by a Fejér sum, we shall obtain finite sums whose S-distance from $f(x)$ can be made arbitrarily small. This process can be regarded as an algorithm on the Fourier series of $f(x)$ itself, since the Fourier series of $\varphi_\delta(x)$ involved can be obtained at once from this series by a formal integration:

$$\varphi_\delta(x) \sim \sum A_\nu \frac{e^{i\Lambda_\nu\delta}-1}{i\Lambda_\nu\delta} e^{i\Lambda_\nu x} \tag{19}$$

$\left(\frac{e^{i\Lambda_\nu\delta}-1}{i\Lambda_\nu\delta}\right.$ stands for 1 when $\left.\Lambda_\nu = 0\right)$. This follows from the obvious equation

$$M\left\{\frac{1}{\delta}\int_x^{x+\delta} f(\xi)\,d\xi\cdot e^{-i\lambda x}\right\} = M\{f(x)\,e^{-i\lambda x}\}\,\frac{e^{i\lambda\delta}-1}{i\lambda\delta}$$

already used by Bohr ([1], p. 62) for a. p. functions and by Stepanoff for S. a. p. functions.

The above algorithm for the construction of a sequence $s_n(x)$ is complicated by the presence of the parameter δ. This complication can be avoided; we shall show that *we may use as approximating sums the Fejér sums of $f(x)$ itself,* instead of the Fejér sums of the functions $\varphi_\delta(x)$.

For this purpose we first prove that the relation given in § 1 between the Fourier exponents and the translation numbers of an a. p. function holds also for an S. a. p. function; in other words that:

"If $\alpha_1, \alpha_2, \ldots$ is a base of the Fourier exponents Λ_ν of an S. a. p. function $f(x)$, then to any $\varepsilon > 0$ correspond integers P, Q and a positive number $\delta < \pi$ such that every solution of the system of diophantine inequalities

$$\left|\frac{\alpha_1}{Q} t\right| < \delta \pmod{2\pi}, \;\ldots,\; \left|\frac{\alpha_p}{Q} t\right| < \delta \pmod{2\pi}$$

is an S-translation number $\tau\left(\frac{\varepsilon}{2}\right)$ of $f(x)$".

We take a $\varphi_\delta(x)$ such that $D_S\left[f(x), \varphi_\delta(x)\right] < \frac{\varepsilon}{6}$. Then every S-translation number of $\varphi_\delta(x)$ (and a fortiori every U-translation number of $\varphi_\delta(x)$) belonging to $\frac{\varepsilon}{6}$ is also an S-translation number of $f(x)$ belonging to $\frac{\varepsilon}{2}$. We have now only to apply the theorem of § 1 $\left(\text{with } \frac{\varepsilon}{6} \text{ in place of } \frac{\varepsilon}{2}\right)$ to the (ordinary) a. p. function $\varphi_\delta(x)$, which by (19) also has the sequence $\alpha_1, \alpha_2, \ldots$ as a base for its Fourier exponents.

Using the notation of § 1 we write

$$s(x) = M\left\{f(x+t)\, \Pi_{n_1}\left(\frac{\alpha_1}{N_1!} t\right) \ldots \Pi_{n_p}\left(\frac{\alpha_p}{N_p!} t\right)\right\}$$

$$= \sum_{\substack{-n_1 \leq \nu_1 \leq n_1 \\ \cdots\cdots\cdots \\ -n_p \leq \nu_p \leq n_p}} \left(1 - \frac{|\nu_1|}{n_1}\right) \ldots \left(1 - \frac{|\nu_p|}{n_p}\right) A_\nu\, e^{i \Lambda_\nu x}$$

where

$$\Lambda_\nu = \frac{\nu_1}{N_1!}\alpha_1 + \ldots + \frac{\nu_p}{N_p!}\alpha_p$$

(and A_ν, as before, denotes zero when $\frac{\nu_1}{N_1!}\alpha_1 + \ldots + \frac{\nu_p}{N_p!}\alpha_p$ is not one of the Fourier exponents of $f(x)$). We shall prove, just as in § 1, that

$$S\text{-lim}\, s(x) = f(x)$$

provided only $p, N_1, N_2, \ldots, n_1, n_2, \ldots$ *tend to* ∞ *together in such a way that* $\frac{n_1}{N_1!} \to \infty, \frac{n_2}{N_2!} \to \infty, \ldots$.

For any x at which $f(x)$ is defined (and so for almost all values of x) we have

$$s(x) - f(x) = M\{(f(x+t) - f(x))\,\Pi(t)\}$$

and therefore

$$|s(x) - f(x)| \leq M\{|f(x+t) - f(x)|\,\Pi(t)\};$$

the mean value on the right exists, since $|f(x+t) - f(x)|$ is plainly an S. a. p. function of t. Thus for every x

$$(20) \quad \int_x^{x+1} |s(\xi) - f(\xi)|\,d\xi \leq \int_x^{x+1} M\{|f(\xi+t) - f(\xi)|\,\Pi(t)\}\,d\xi.$$

We shall show that the last integral is less than or equal to

$$(21) \qquad M\left\{\Pi(t)\int_x^{x+1} |f(\xi+t) - f(\xi)|\,d\xi\right\};$$

(this mean value also exists since $\int_x^{x+1} |f(\xi+t) - f(\xi)|\,d\xi$ is even an ordinary a. p. function of t). We have

$$(22) \quad \left\{ \begin{aligned} &\frac{1}{2T}\int_{-T}^{T} \Pi(t)\,dt \int_x^{x+1} |f(\xi+t) - f(\xi)|\,d\xi \\ &= \int_x^{x+1} d\xi\, \frac{1}{2T}\int_{-T}^{T} |f(\xi+t) - f(\xi)|\,\Pi(t)\,dt. \end{aligned} \right.$$

As $T \to \infty$ the left side of (22) and therefore also the right side tends to (21). But the limit of the right hand side is, by a theorem of Fatou[1] greater than or equal to

$$\int_x^{x+1} d\xi \lim \frac{1}{2T}\int_{-T}^{T} |f(\xi+t) - f(\xi)|\,\Pi(t)\,dt$$
$$= \int_x^{x+1} d\xi\, M\{|f(\xi+t) - f(\xi)|\,\Pi(t)\}.$$

[1] If $f_n(x) \geq 0$ in (a, b) and $f_n(x) \to f(x)$ then

$$\liminf \int_a^b f_n(x)\,dx \geq \int_a^b f(x)\,dx.$$

Thus by (20)

$$\int_x^{x+1} |s(\xi) - f(\xi)|\, d\xi \leq M\{\Pi(t) \int_x^{x+1} |f(\xi + t) - f(\xi)|\, d\xi\}.$$

Dividing the t-axis, as in § 1, into two sets I_1, I_2 and writing $\Pi(t) = \Pi'(t) + \Pi''(t)$ as before we have

$$D_S[s(x), f(x)] \leq \overline{M}\{D_S[f(x+t), f(x)]\,\Pi'(t)\} \\ + \overline{M}\{D_S[f(x+t), f(x)]\,\Pi''(t)\}.$$

This inequality differs from the inequality (9) of § 1 only in having D_U replaced by D_S. Since further the distance

$$D_S[f(x+t), f(x)]$$

is $\leq \frac{\varepsilon}{2}$ in I_1 and $\leq 2K$ in I_2 $\left(K \text{ is } D_S[f(x)]\right)$, the rest of the proof is word for word the same as that of § 1.

In conclusion we may remark that the above "summation theorem" implies the "uniqueness theorem" (see Stepanoff) which states that *an S. a. p. function is uniquely determined by its Fourier series.* For if two functions are both S-limits of the same sequence (Fejér sums of the given Fourier series) their S-distance must be zero and consequently they are equivalent, i. e. equal almost everywhere.

MEMOIRS REFERRED TO.

A. Besicovitch:

[1]. "On generalized almost periodic functions", Proc. London Math. Soc. (2), 25 (1926), p. 495—512.

S. Bochner:

[1]. "Beiträge zur Theorie der fastperiodischen Funktionen", Math. Ann. 96 (1926), p. 119—147.

[2]. "Properties of Fourier series of almost periodic functions", Proc. London Math. Soc. (2), 26 (1927), p. 433—452.

H. Bohr:

[1], [2], [3]. "Zur Theorie der fastperiodischen Funktionen" I, II, III, Acta Math. 45 (1924), p. 29—127, 46 (1925), p. 101—214, 47 (1926), p. 237—281.

W. Stepanoff:

[1]. "Über einige Verallgemeinerungen der fastperiodischen Funktionen", Math. Ann. 95 (1926), p. 473—498.

Forelagt paa Mødet den 16. December 1927.
Færdig fra Trykkeriet den 31. December 1927.

MATHEMATISK-FYSISKE MEDDELELSER

UDGIVNE AF

DET KGL. DANSKE VIDENSKABERNES SELSKAB

6. BIND (Kr. 17,00):

		Kr. Ø.
1.	Nielsen, Niels: Sur l'opération itérative des Équations de Lagrange. 1924	3.10
2.	Urey, H. C.: On the Effect of perturbing Electric Fields on the Zeeman Effect of the Hydrogen Spectrum. 1924	0.65
3.	Bøggild, O. B.: On the Labradorization of the Feldspars. With one plate. 1924	3.00
4.	Pedersen, P. O.: Om elektriske Gnister. II. Eksperimentelle Undersøgelser over Gnistforsinkelse og Gnistdannelse. Med 7 Tavler. 1924	4.30
5.	Juel, C.: Über Flächen von Maximalindex. 1924	1.25
6.	Nielsen, Niels: Sur une Équation de Lagrange. 1924	1.25
7.	Hevesy, G. de: Recherches sur les propriétés du Hafnium. Avec 2 planches. 1925	6.25
8.	Bohr, Harald: Neuer Beweis eines allgemeinen Kronecker'schen Approximationssatzes. 1924	0.50
9.	Bjerrum, Niels and Ebert, Ludwig: On some recent Investigations concerning Mixtures of Strong Electrolytes (Transference Numbers and Amalgam Equilibria). 1925	0.75
10.	Landau, Edm.: Die Ungleichungen für zweimal differentiierbare Funktionen. 1925	1.60

7. BIND (Kr. 17,70):

1.	Bohr, Harald: Unendlich viele lineare Kongruenzen mit unendlich vielen Unbekannten. 1925	1.40
2.	Hartmann, Jul., and Trolle, Birgit: On Beat-phenomena in Cylindrical Tubes exposed to Sound-waves. With three plates. 1925	2.85
3.	Pauli, W. jr.: Ueber die Intensitäten der im elektrischen Feld erscheinenden Kombinationslinien. 1925	0.65
4.	Hardy, G. H. and Littlewood, J. E.: A theorem concerning series of positive terms, with applications to the theory of functions. 1925	0.90
5.	Steffensen, J. F.: On a Generalization of Nörlund's Polynomials. 1926	1.00
6.	Hartmann, Jul., and Trolle, Birgit: New investigation on the air jet generator for acoustic wawes. 1926	2.40
7.	Mollerup, Johs.: Sur l'approximation d'un nombre irrationnel par des carrés rationnels. 1926	0.80
8.	Nielsen, Niels: Sur certains développements d'une fonction holomorphe. 1926	0.75

Kr. Ø.

9. Bjerrum, Niels: Untersuchungen über Ionenassoziation. I. Der Einfluss der Ionenassoziation auf die Aktivität der Ionen bei mittleren Assoziationsgraden. 1926 .... 2.00
10. Nielsen, Niels: Recherches sur les fonctions cylindriques et sur certaines fonctions analogues. 1926 .... 1.35
11. Hevesy, G.: On the Missing Element 87. 1926 .... 0.60
12. Kuhn, W.: Die Stärke der anomalen Dispersion in nicht leuchtendem Dampfe von Thallium und Cadmium. 1926 .... 3.75
13. Sugiura, Y. and Urey, H. C.: On the Quantum Theory explanation of the Anomalies in the 6th and 7th Periods of the Periodic Table. 1926 .... 1.15
14. Burrau, Øyvind: Berechnung des Energiewertes des Wasserstoffmolekel-Ions (H_2^+) im Normalzustand. 1927 .... 0.90
15. Knudsen, Martin: The Hot-wire Manometer. 1927 .... 0.80
16. Juel, C.: Om v. Staudt's Definitioner. 1927 .... 0.85
17. Pólya, G.: Über die algebraisch-funktionentheoretischen Untersuchungen von J. L. W. V. Jensen. 1927 .... 1.40

8. BIND (under Pressen):

1. Hille, Einar: On the Logarithmic Derivatives of the Gamma Function. 1927 .... 2.40
2. Rasch, G.: Beitrag zur Theorie der unvollständigen Gammafunktionen. Nach hinterlassenen Papieren von J. L. W. V. Jensen. 1927 .... 0.85
3. Knudsen, Martin: Thermal Molecular Pressure in Tubes. 1927 1.50
4. Pedersen, P. O.: Den højere Atmosfæres Sammensætning, Tryk, Temperatur og elektriske Ledningsevne i Belysning af Radiobølgernes Udbredelsesforhold. 1927 .... 0.60
5. Besicovitch, A. and Bohr, H.: Some remarks on Generalisations of almost Periodic Functions. 1927 .... 1.50

Det Kgl. Danske Videnskabernes Selskab.
Mathematisk-fysiske Meddelelser. **VIII,** 6.

ÜBER EINIGE BESONDERE KLASSEN VON FASTPERIODISCHEN FUNKTIONEN

VON

ELLEN PEDERSEN

KØBENHAVN
HOVEDKOMMISSIONÆR: ANDR. FRED. HØST & SØN, KGL. HOF-BOGHANDEL
BIANCO LUNOS BOGTRYKKERI
1928

Det Kgl. Danske Videnskabernes Selskabs videnskabelige Meddelelser udkommer fra 1917 indtil videre i følgende Rækker:

Historisk-filologiske Meddelelser,
Filosofiske Meddelelser,
Mathematisk-fysiske Meddelelser,
Biologiske Meddelelser.

Hele Bind af disse Rækker sælges 25 pCt. billigere end Summen af Bogladepriserne for de enkelte Hefter.

Selskabets Hovedkommissionær er *Andr. Fred. Høst & Søn*, Kgl. Hof-Boghandel, København.

Det Kgl. Danske Videnskabernes Selskab.
Mathematisk-fysiske Meddelelser. **VIII,** 6.

ÜBER EINIGE BESONDERE KLASSEN VON FASTPERIODISCHEN FUNKTIONEN

VON

ELLEN PEDERSEN

KØBENHAVN
HOVEDKOMMISSIONÆR: ANDR. FRED. HØST & SØN, KGL. HOF-BOGHANDEL
BIANCO LUNOS BOGTRYKKERI
1928

HARALD BOHR hat in den Abhandlungen in Acta Mathematica, in welchen er die Theorie der fastperiodischen Funktionen begründet hat, folgende Sätze, welche Verallgemeinerungen des klassischen Weierstrassschen Satzes über rein periodische Funktionen sind, bewiesen:

I. (Approximationssatz): Damit die für $-\infty < x < \infty$ definierte Funktion $f(x)$ durch ein trigonometrisches Polynom

$$c_1 e^{i\lambda_1 x} + c_2 e^{i\lambda_2 x} + \ldots + c_n e^{i\lambda_n x} \tag{1}$$

(mit willkürlich wählbaren reellen Exponenten λ_i) gleichmässig für alle x mit vorgeschriebener Genauigkeit ε approximierbar sei, ist notwendig und hinreichend, dass $f(x)$ fastperiodisch ist.

II. Damit die geforderte Approximation möglich sei mit Exponenten λ_i, die einer im voraus gegebenen Menge $\{\lambda\}$ von reellen Zahlen angehören, ist notwendig und hinreichend, dass (die Funktion fastperiodisch sei und) alle Exponenten in der Fourierreihe der Funktion Zahlen aus der Menge $\{\lambda\}$ sein sollen.

Von besonderem Interesse sind die Fälle, in denen $\{\lambda\}$ ein abzählbarer Modul ist. Für diesen Fall leitet man aus Satz II den folgenden Satz her:

IIa. Gegeben sei der abzählbare Modul

$$(2) \qquad \{\lambda\} = \{\lambda_1, \lambda_2, \ldots, \lambda_n, \ldots\}.$$

Damit eine für $-\infty < x < \infty$ definierte Funktion $f(x)$ durch eine Summe von der Form (1) mit Exponenten, die (2) angehören, beliebig genau approximierbar sei, ist notwendig und hinreichend, dass zu einem ε ein $\delta = \delta(\varepsilon)$ und ein $N = N(\varepsilon)$ existieren, so dass jedes die Ungleichungen

$$|\tau \lambda_i| < \delta \pmod{2\pi} \qquad (i = 1, 2, \ldots, N)$$

erfüllende τ eine zu ε gehörige Verschiebungszahl von $f(x)$ sein soll.

Das in Satz II a genannte Kriterium erhält eine einfachere und übersichtlichere Form, wenn über $\{\lambda\}$ weitere spezielle Annahmen gemacht werden.

Bohl[1] hat in einer grundlegenden Abhandlung den Fall betrachtet, in welchem die λ ganzzahlige Kombinationen einer endlichen Anzahl von linear unabhängigen α sind, während in den folgenden Überlegungen der Fall behandelt wird, wo der Modul $\{\lambda\}$ eine endliche oder unendliche Basis $\{\alpha\}$ hat, und wo die λ alle lineare Kombinationen der α mit entweder ganzzahligen oder rationalen Koeffizienten sind.

Beim Beweise der Sätze I—IIa benutzt Bohr die vollständige Theorie der fastperiodischen Funktionen. Im folgenden sollen die für die genannten Spezialfälle geltenden Sätze direkt und mit verhältnismässig einfachen Hülfsmitteln begründet werden, indem dieselben Beweismethoden angewendet werden, welche Bohl in seiner Abhandlung benutzt hat.

[1] P. Bohl, Über die Darstellung von Funktionen einer Variabeln durch trigonometrische Reihen mit mehreren einer Variabeln proportionalen Argumenten (Dorpat, 1893).

Was die im folgenden benutzten Begriffe der grenzperiodischen Funktionen von endlich vielen Variabeln sowie der rein- und grenzperiodischen Funktionen von unendlich vielen Variabeln betrifft, vergleiche man die zweite Arbeit von BOHR in Acta Mathematica.

I.

Gegeben ist eine abzählbar unendliche Folge $\alpha_1, \alpha_2, \ldots$ von linear unabhängigen Zahlen α; dieses heisst, dass $n_1\alpha_1 + n_2\alpha_2 + \ldots + n_\nu\alpha_\nu = 0$ für ganze n_i nur erfüllt wird, wenn $n_1 = n_2 = \ldots = n_\nu = 0$ ist. Es sei $f(x)$ eine für $-\infty < x < \infty$ definierte stetige Funktion, welche durch ein trigonometrisches Polynom

$$(1) \qquad \sum c_{n_1 \ldots n_m} e^{i(n_1\alpha_1 + \ldots + n_m\alpha_m)x}$$

für alle x gleichmässig approximierbar ist. Die Menge dieser Funktionen bezeichnen wir mit A.

Satz I. Die notwendige und hinreichende Bedingung dafür, dass $f(x)$ der Menge A angehört, ist folgende: Zu einem beliebigen $\varepsilon > 0$ existiert ein $\delta = \delta(\varepsilon)$ und $N = N(\varepsilon)$, so dass jedes τ, welches die N diophantischen Ungleichungen

$$(2) \qquad \begin{cases} |\tau\alpha_1| < \delta \quad (\mathrm{mod}\ 2\pi) \\ \ldots\ldots\ldots\ldots\ldots \\ |\tau\alpha_N| < \delta \quad (\mathrm{mod}\ 2\pi) \end{cases}$$

erfüllt, eine zu ε passende Verschiebungszahl von $f(x)$ ist, das heisst die Ungleichung

$$|f(x+\tau) - f(x)| < \varepsilon$$

für alle x erfüllt.

Beweis: Die Menge der Funktionen $f(x)$, welche der genannten (ε, δ, N)-Bedingung genügen, bezeichnen wir mit B. Zunächst beweisen wir, dass A eine Teilmenge von B ist. Es sei nämlich $f(x)$ eine Funktion aus A und $S(x)$ ein trigonometrisches Polynom von der Form (1), welches $f(x)$ mit der Genauigkeit $\frac{\varepsilon}{3}$ approximiert. $S(x+\tau)$ entsteht aus $S(x)$, indem jeder Faktor $e^{ik\alpha x}$ in jedem der Glieder einen Zusatzfaktor $e^{ik\alpha\tau}$ erhält. Daher wird $|S(x+\tau) - S(x)| < \frac{\varepsilon}{3}$ ausfallen, wenn jeder dieser Zusatzfaktoren hinreichend nahe an 1 liegt, das heisst: die Bedingung $|k\alpha\tau| < \delta^*$ (mod 2π) für hinreichend kleines δ^* erfüllt. Wir erreichen also $|S(x+\tau) - S(x)| < \frac{\varepsilon}{3}$ und damit $|f(x+\tau) - f(x)| < \varepsilon$, wenn wir N so gross wählen, dass alle in $S(x)$ auftretenden α mitkommen, und $\delta = \frac{\delta^*}{M}$ setzen, wo M das grösste aller in Betracht kommenden k ist.

Die Identität der Mengen A und B wird nun dadurch bewiesen, dass wir zeigen, dass sowohl A wie B identisch mit einer Funktionenmenge C sind, die wie folgt definiert wird: Eine für $-\infty < x < \infty$ definierte stetige Funktion soll dann und nur dann C angehören, wenn

$$f(x) = P(x, x, \ldots) \tag{3}$$

ist, wo $P(x_1, x_2, \ldots.)$ eine im Raume von abzählbar unendlich vielen Dimensionen definierte stetige periodische Funktion mit dem Periodensystem

$$(p_1, p_2, \ldots.) = \left(\frac{2\pi}{\alpha_1}, \frac{2\pi}{\alpha_2}, \ldots.\right) \tag{4}$$

ist.

Die Menge C ist in der Menge A enthalten. Auf Grund des Weierstrassschen Approximationssatzes für Funktionen von m Variabeln lässt sich nämlich ohne weiteres zeigen, dass $P(x_1, x_2, \ldots.)$ im ganzen Raume durch ein Polynom $\sum c_{n_1 \ldots n_m} e^{i(n_1 \alpha_1 x_1 + \ldots + n_m \alpha_m x_m)}$ mit ε-Genauigkeit approximierbar ist. Ist nun $f(x)$ eine Funktion aus C und $= P(x, x, \ldots)$, erhält man, indem in der Ungleichung

$$\left| P(x_1, x_2, \ldots) - \sum c_{n_1 \ldots n_m} e^{i(n_1 \alpha_1 x_1 + \ldots + n_m \alpha_m x_m)} \right| < \varepsilon$$

x_i durch x entsetzt wird,

$$\left| f(x) - \sum c_{n_1 \ldots n_m} e^{i(n_1 \alpha_1 + \ldots + n_m \alpha_m) x} \right| < \varepsilon,$$

das heisst $f(x)$ ist eine Funktion aus A.

Es erübrigt noch zu zeigen, dass die Menge B eine Teilmenge von C ist. Wir nehmen eine willkürliche Funktion $f(x)$ aus B und definieren zunächst $P(x_1, x_2, \ldots)$ auf der Hauptdiagonalen $x_1 = x_2 = \ldots$ durch die Gleichung

$$P(x, x, \ldots.) = f(x).$$

Hiernach definieren wir die Punktmenge Ξ wie folgt: Der Punkt $(\xi_1, \xi_2, \ldots.)$ soll zu Ξ gehören, wenn es ein x giebt, so dass alle Kongruenzen

$$(5) \qquad \xi_\nu \equiv x \pmod{p_\nu} \qquad (\nu = 1, 2, \ldots)$$

erfüllt sind. Auf Ξ definieren wir $P(x_1, x_2, \ldots)$ durch

$$(6) \qquad P(\xi_1, \xi_2, \ldots) = P(x, x, \ldots) = f(x),$$

wo $\xi_1, \xi_2, \ldots$ und x durch (5) verbunden sind. Die lineare Unabhängigkeit der Zahlen α führt mit sich, dass ein Punkt $(\xi_1, \xi_2, \ldots)$ von Ξ nicht mit zwei Punkten $(x', x', \ldots)$ und $(x'', x'', \ldots)$ der Hauptdiagonalen äquivalent sein kann, sonst müsste nämlich $x' - x''$ ein gemeinsames Vielfaches aller p sein, was nicht möglich ist, da das Verhältniss von p_i und p_k irrational ist.

Die Punkte von Ξ liegen im Raume überall dicht, da man nach Angabe eines Punktes $(a_1, a_2, \ldots)$, eines N und beliebig kleiner $\delta_1, \delta_2, \ldots \delta_N$ Punkte $(\xi_1, \xi_2, \ldots \xi_N, \ldots)$ finden kann, welche die N Ungleichungen

$$|\xi_\nu - a_\nu| < \delta_\nu \qquad (\nu = 1, 2, \ldots N)$$

oder die N diophantischen Ungleichungen

$$|x - a_\nu| < \delta_\nu \qquad (\text{mod } p_\nu)$$

erfüllen, welche letztere auch in der Form

$$|\alpha_\nu x - \alpha_\nu a_\nu| < \delta_\nu |\alpha_\nu| \qquad (\text{mod } 2\pi)$$

geschrieben werden können; und diese haben immer eine Lösung, da die α linear unabhängig sind. Es soll nun gezeigt werden, dass P in Ξ gleichmässig stetig ist. Hierbei benutzen wir die Voraussetzung, dass $f(x)$ der Menge B angehört. Wir zeigen, dass wir zu einem willkürlichen $\varepsilon > 0$ ein N und Zahlen $\delta_1, \delta_2, \ldots, \delta_N$ so finden können, dass die Ungleichung

$$|P(\xi_1', \xi_2', \ldots) - P(\xi_1'', \xi_2'', \ldots)| < \varepsilon$$

für zwei willkürliche Punkte $(\xi_1', \xi_2', \ldots)$ und $(\xi_1'', \xi_2'', \ldots)$ von Ξ erfüllt ist, sobald für die ersten N Koordinaten die Ungleichungen

$$|\xi_\nu' - \xi_\nu''| < \delta_\nu \qquad (\nu = 1, 2, \ldots, N)$$

gelten.

Wir bezeichnen mit x' und $x'' = x' + \tau$ die zwei Werte von x, von welchen aus die Punkte $(\xi_1', \xi_2', \ldots)$ und $(\xi_1'', \xi_2'', \ldots)$ hergeleitet wurden. Zu beweisen ist, dass

$$|f(x') - f(x'')| = |f(x') - f(x' + \tau)| < \varepsilon,$$

sobald die N Bedingungen

$$|x' - x''| = |\tau| < \delta_\nu \qquad (\text{mod } p_\nu)$$

oder

$$|\alpha_\nu \tau| < \delta_\nu |\alpha_\nu| \qquad (\text{mod } 2\pi)$$

erfüllt sind, und dieses trifft zu, wenn man die Zahlen $N, \delta_1, \delta_2, \ldots \delta_N$ so wählt, dass in den durch $f(x)$ erfüllten Bedingungen (2) $N = N(\varepsilon)$ und $\delta_\nu |\alpha_\nu| = \delta(\varepsilon)$ ist. Da $P(x_1, x_2, \ldots)$ in Ξ gleichmässig stetig ist, und Ξ im Raume überall dicht liegt, gibt es eine und nur eine im ganzen Raume definierte stetige Funktion, die auf Ξ mit der daselbst schon definierten Funktion übereinstimmt. Damit ist gezeigt, dass B eine Teilmenge von C ist; die gefundene Funktion $P(x_1, x_2, \ldots.)$ erfüllt nämlich alle gestellten Bedingungen: Sie ist periodisch mit dem Periodensystem $(p_1, p_2, \ldots)$, sie erfüllt die Gleichung $P(x, x, \ldots) = f(x)$ und sie ist stetig im ganzen Raume; denn wählen wir $\delta = \delta(\varepsilon)$ so, dass

$$|P(\xi_1', \xi_2', \ldots) - P(\xi_1'', \xi_2'', \ldots)| < \varepsilon$$

erfüllt wird für zwei beliebige Punkte von Ξ, für welche $|\xi_\nu' - \xi_\nu''| < \delta$ $(\nu = 1, 2, \ldots, N)$ gilt, wird für dasselbe δ und N auch die Ungleichung

$$|P(x_1', x_2', \ldots) - P(x_1'', x_2'', \ldots)| \leq \varepsilon$$

erfüllt sein, sobald $|x_\nu' - x_\nu''| < \delta$ $(\nu = 1, 2, \ldots N)$.

II.

Gegeben seien m reelle, linear unabhängige Zahlen: $\alpha_1, \alpha_2, \ldots, \alpha_m$. Wir betrachten die Menge aller Schwingungen $e^{i\alpha x}$, wo α die Form

$$\alpha = r_1\alpha_1 + r_2\alpha_2 + \ldots + r_m\alpha_m$$

hat, und wo $r_1, r_2, \ldots, r_m$ rationale Zahlen bedeuten. Es sei $f(x)$ eine für $-\infty < x < \infty$ definierte stetige Funktion, welche gleichmässig für alle x durch eine endliche Summe von der Form

(8) $$\sum c_{r_1 r_2 \ldots r_m} e^{i(r_1\alpha_1 + r_2\alpha_2 + \ldots + r_m\alpha_m)x}$$

approximiert werden kann. Die Menge aller dieser Funktionen bezeichnen wir mit A.

Die notwendige und hinreichende Bedingung dafür, dass eine Funktion $f(x)$ der Menge A angehört, ist folgende: Zu einem gegebenen $\varepsilon > 0$ gibt es ein $\delta = \delta(\varepsilon)$ und ein ganzzahliges $M = M(\varepsilon)$ so, dass jede Zahl τ, welche die m diophantischen Ungleichungen

(9) $$|\tau\alpha_\nu| < \delta \quad (\text{mod } 2\pi M) \quad (\nu = 1, 2, \ldots m)$$

erfüllt, eine zu ε passende Verschiebungszahl von $f(x)$ ist, das heisst der Ungleichung

$$|f(x+\tau) - f(x)| < \varepsilon \qquad (-\infty < x < \infty)$$

genügt.

Die Menge aller Funktionen mit der genannten (ε, δ, M)-Eigenschaft wollen wir wieder mit B bezeichnen. A ist eine Teilmenge von B; ist nämlich $f(x)$ eine Funktion aus A, kann diese mit $\frac{\varepsilon}{3}$ Genauigkeit durch eine Summe $S(x)$ von der Form (8) angenähert werden. Ist M ein Hauptnenner der in $S(x)$ auftretenden r, kann man $\delta > 0$ so klein wählen, dass für jedes τ, welches die Ungleichungen

$$|\alpha_\nu \tau| < \delta \quad (\text{mod } 2\pi M) \quad (\nu = 1, 2, \ldots, m)$$

erfüllt, gilt, dass

$$|S(x+\tau) - S(x)| < \frac{\varepsilon}{3}, \qquad (-\infty < x < \infty)$$

und für ein solches τ hat man also

$$|f(x+\tau) - f(x)| < \varepsilon.$$

Wir wollen nun zeigen, dass die Mengen A und B identisch sind, indem wir beweisen, dass beide mit einer

Funktionenmenge C zusammenfallen, welche wie folgt definiert ist: Die Funktion $f(x)$ gehört zu C, wenn sie in der Form

$$(10) \qquad f(x) = G(x, x, \ldots, x)$$

geschrieben werden kann, wo $G(x_1, x_2, \ldots, x_m)$ eine stetige grenzperiodische Funktion mit dem Periodensystem

$$(11) \qquad (p_1, p_2, \ldots, p_m) = \left(\frac{2\pi}{\alpha_1}, \frac{2\pi}{\alpha_2}, \ldots, \frac{2\pi}{\alpha_m}\right)$$

ist.

C ist eine Teilmenge von A; denn eine Funktion $f(x) = G(x, x, \ldots, x)$ aus C kann ja durch eine Summe von der Form (8) mit ε-Genauigkeit approximiert werden; die Menge der grenzperiodischen Funktionen $G(x_1, x_2, \ldots, x_m)$ mit dem Periodensystem (11) besteht nämlich eben aus denjenigen Funktionen, welche durch endliche Summen der Form $\sum c_{r_1 r_2 \ldots r_m} e^{i(r_1 \alpha_1 x_1 + \ldots + r_m \alpha_m x_m)}$ gleichmässig approximiert werden können; also braucht man nur die Funktion $G(x_1, x_2, \ldots, x_m)$ im ganzen m-dimensionalen Raume mit ε-Genauigkeit durch eine endliche Summe von der Form $\sum c_{r_1 \ldots r_m} e^{(r_1 \alpha_1 x_1 + \ldots + r_m \alpha_m x_m)}$ zu approximieren und danach in der Ungleichung

$$| G(x_1, \ldots, x_m) - \sum c_{r_1 \ldots r_m} e^{i(r_1 \alpha_1 x_1 + \ldots + r_m \alpha_m x_m)} | < \varepsilon$$

$x_i = x$ zu setzen.

Um die Identität von A und B zu beweisen, erübrigt es noch zu zeigen, dass B eine Teilmenge von C ist. Ist also $f(x)$ eine willkürliche Funktion aus B, soll gezeigt werden, dass es eine grenzperiodische Funktion $G(x_1, x_2, \ldots, x_m)$ mit dem Periodensystem (11) gibt, für welche $f(x) = G(x, x, \ldots, x)$.

Für ein willkürlich gegebenes x und ein gegebenes

s (= 1, 2, . . .) bestimmen wir alle die Punkte $\xi^{(s)} = (\xi_1^{(s)}, \xi_2^{(s)}, \ldots, \xi_m^{(s)})$, deren Koordinaten die Bedingungen

$$(12)\quad \begin{cases} \xi_1^{(s)} \equiv x \pmod{s!\, p_1} \\ \dots\dots\dots\dots\dots \\ \xi_m^{(s)} \equiv x \pmod{s!\, p_m} \end{cases}$$

erfüllen. Indem nun x alle Werte durchläuft, entsteht eine Menge Ξ_s von Punkten $\xi^{(s)}$; in Ξ_s definieren wir eine Funktion $g_s = g_s(\xi_1^{(s)}, \xi_2^{(s)}, \ldots, \xi_m^{(s)})$ durch die Gleichung

$$(13)\qquad g_s(\xi_1^{(s)}, \xi_2^{(s)}, \ldots, \xi_m^{(s)}) = f(x)$$

Da einem $\xi^{(s)}$ nur ein x entspricht, da die α linear unabhängig sind, ist die Funktion g_s eindeutig bestimmt.

Die Funktion g_s ist in Ξ_s unstetig, doch in der Weise, dass für $s \to \infty$ die Funktion g_s »mehr und mehr stetig wird«, womit folgendes gemeint ist: $|g_s(\xi') - g_s(\xi'')|$ wird für $s > M = M(\varepsilon)$ kleiner als ε ausfallen, sobald die Punkte ξ' und ξ'' von Ξ_s so nahe an einander liegen, dass $|\xi'_\nu - \xi''_\nu| < \delta_\nu = \dfrac{\delta(\varepsilon)}{|\alpha_\nu|}$ $(\nu = 1, 2, \ldots, m)$ ist.

Jede der Mengen $\Xi_1, \Xi_2, \ldots$ liegt überall dicht im Raume; Ξ_{n+1} ist eine Teilmenge von Ξ_n,

$$\Xi_1 > \Xi_2 > \ldots,$$

und nur Punkte der Hauptdiagonalen sind in allen Ξ_s enthalten. Ist also $p > q$, wird der Punkt $(\xi_1^{(p)}, \ldots, \xi_m^{(p)})$ von Ξ_p auch in Ξ_q enthalten sein, und es gilt

$$(14)\qquad g_p(\xi_1^{(p)}, \xi_2^{(p)}, \ldots, \xi_m^{(p)}) = g_q(\xi_1^{(p)}, \xi_2^{(p)}, \ldots, \xi_m^{(p)}).$$

Die Funktion $G(x_1, x_2, \ldots, x_m)$ wird in einem willkürlichen Punkte $(x_1, x_2, \ldots, x_m)$ wie folgt definiert: Wir wählen eine Punktfolge $(\xi_1^{(1)}, \xi_2^{(1)}, \ldots, \xi_m^{(1)}), (\xi_1^{(2)}, \xi_2^{(2)}, \ldots, \xi_m^{(2)}), \ldots,$

welche gegen $(x_1, x_2, \ldots, x_m)$ konvergiert, und wo der s-te Punkt $(\xi_1^{(s)}, \xi_2^{(s)}, \ldots, \xi_m^{(s)})$ der Menge Ξ_s angehört, und setzen

$$(15) \qquad G(x_1, x_2, \ldots, x_m) = \lim_{n \to \infty} g_n(\xi_1^{(n)}, \xi_2^{(n)}, \ldots, \xi_m^{(n)}).$$

Dass zu einem willkürlich gegebenen Punkt $(x_1, x_2, \ldots, x_m)$ eine solche Folge $(\xi_1^{(1)}, \xi_2^{(1)}, \ldots, \xi_m^{(1)}), (\xi_1^{(2)}, \xi_2^{(2)}, \ldots, \xi_m^{(2)}), \ldots$ existiert, folgt daraus, dass jede der Mengen Ξ_n im ganzen Raume überall dicht liegt. Um die Darstellung zu vereinfachen, wollen wir zur Approximation des Punktes $(x_1, x_2, \ldots, x_m)$ solche spezielle Folgen benutzen, für welche die m Ungleichungen

$$(16) \qquad |x_\nu - \xi_\nu^{(s)}| < \frac{1}{s\,|\alpha_\nu|} \qquad (\nu = 1, 2, \ldots, m)$$

erfüllt sind.

Um die Konvergenz der Folge $g_1(\xi^{(1)}), g_2(\xi^{(2)}), \ldots$ einzusehen, betrachten wir die Differenz zwischen dem q-ten und dem p-ten Glied in der Folge, für welche, wenn $p > q$, gilt, dass

$$|g_p(\xi^{(p)}) - g_q(\xi^{(q)})| = |g_q(\xi^{(p)}) - g_q(\xi^{(q)})|.$$

Diese letztere Differenz ist aber kleiner als die gegebene Zahl $\varepsilon > 0$, sobald q so gross gewählt wird, dass $\frac{2}{q} < \delta(\varepsilon)$ und gleichzeitig $q > M(\varepsilon)$ ist, wobei $\delta(\varepsilon)$ und $M(\varepsilon)$ durch die Funktion $f(x)$, welche B angehört, bestimmt sind. Demzufolge ist

$$|g_p(\xi^{(p)}) - g_q(\xi^{(q)})| = |f(x') - f(x'')| < \varepsilon,$$

denn x' und x'' erfüllen die m diophantischen Ungleichungen

$$|x' - x''| < \frac{\delta(\varepsilon)}{|\alpha_\nu|} \pmod{M!\,p_\nu} \qquad (\nu = 1, 2, \ldots, m)$$

oder

$$|(x' - x'')\alpha_\nu| < \delta(\varepsilon) \pmod{2\pi M!} \quad (\nu = 1, 2, \ldots, m).$$

Da $\lim g_s(\xi^{(s)})$ für jede Folge mit $\xi^{(s)} \to (x_1, x_2, \ldots\ldots, x_m)$ existiert, ist der Grenzwert $G(x_1, x_2, \ldots\ldots, x_m)$ unabhängig von der gewählten Folge, wenn nur diese die genannten Bedingungen erfüllt.

Die auf diese Weise definierte Funktion $G(x_1, x_2, \ldots, x_m)$ ist stetig, denn für zwei Punkte $(x'_1, x'_2, \ldots\ldots, x'_m)$ und $(x''_1, x''_2, \ldots, x''_m)$ mit

$$|x'_\nu - x''_\nu| < \frac{\delta(\varepsilon)}{3|\alpha_\nu|} \qquad (\nu = 1, 2, \ldots, m)$$

gilt

$$|\xi_\nu'^{(s)} - \xi_\nu''^{(s)}| < \frac{\delta(\varepsilon)}{|\alpha_\nu|}$$

für jedes $s \geq s_0$, sobald $\frac{1}{s_0} < \frac{\delta(\varepsilon)}{3}$, und ist ausserdem $s_0 > M(\varepsilon)$, erhalten wir für alle $s \geq s_0$

$$|g_s(\xi'^{(s)}) - g_s(\xi''^{(s)})| = |f(x'^{(s)}) - f(x''^{(s)})| < \varepsilon,$$

denn wir haben

$$|\alpha_\nu(x'^{(s)} - x''^{(s)})| < \delta(\varepsilon) \quad \left(\bmod\ M(\varepsilon)!\,2\pi\right), \quad \nu = 1, 2, \ldots, m,$$

woraus zuletzt folgt

$$|G(x'_1, x'_2, \ldots, x'_m) - G(x''_1, x''_2, \ldots, x''_m)| \leq \varepsilon.$$

Die auf diese Weise im ganzen Raume definierte Funktion $G(x_1, x_2, \ldots, x_m)$ ist von der gewünschten Art: sie ist stetig, grenzperiodisch mit dem Periodensystem $\left(\frac{2\pi}{\alpha_1}, \frac{2\pi}{\alpha_2}, \ldots, \frac{2\pi}{\alpha_m}\right)$ und erfüllt auf der Hauptdiagonalen die Gleichung $G(x, x, \ldots, x) = f(x)$. Dies letztere ist unmittelbar ersichtlich, da man für einen beliebigen Punkt $(x, x, \ldots, x)$ auf der Hauptdiagonalen als Punkt $(\xi_1^{(s)}, \xi_2^{(s)}, \ldots, \xi_m^{(s)})$ den Punkt $(x, x, \ldots, x)$ wählen kann, so dass $G(x, x, \ldots, x)$ der Grenzwert einer Folge ist, in der jedes Glied $f(x)$ ist. Dass die Funktion grenzperiodisch mit dem gegebenen

Periodensystem ist, das heisst mit ε-Genauigkeit durch eine stetige reinperiodische Funktion mit dem Periodensystem $\left(n!\,\frac{2\pi}{\alpha_1}, n!\,\frac{2\pi}{\alpha_2}, \ldots, n!\,\frac{2\pi}{\alpha_m}\right)$ approximierbar ist, zeigen wir wie folgt: Wir definieren zunächst eine reinperiodische, aber unstetige Funktion Π_n mit dem Periodensystem $\left(n!\,\frac{2\pi}{\alpha_1}, n!\,\frac{2\pi}{\alpha_2}, \ldots, n!\,\frac{2\pi}{\alpha_m}\right)$, welche im Bereich $0 \leq x_\nu < n!\,|p_\nu|$ $(\nu = 1, 2, \ldots, m)$ mit $G(x_1, x_2, \ldots, x_m)$ zusammenfällt. Um nun eine stetige periodische Funktion P_n zu erhalten, nehmen wir eine Ausglättung von Π_n vor, indem wir P_n in jedem Punkte des Raumes definieren als den Mittelwert der Funktion Π_n im Inneren einer Kugel mit dem Mittelpunkt in $(x_1, x_2, \ldots, x_m)$ und Radius δ. Ist n so gross, dass im ganzen Raume gilt

$$| G(x_1, x_2, \ldots, x_m) - g_n(\xi_1^{(n)}, \xi_2^{(n)}, \ldots, \xi_m^{(n)}) | < \frac{\varepsilon}{4},$$

ist gleichzeitig $n > M\left(\frac{\varepsilon}{2}\right)$, und wählen wir $\delta < \frac{\delta\left(\frac{\varepsilon}{2}\right)}{2\alpha}$, wo α die grösste unter den Zahlen $|\alpha_1|, \ldots, |\alpha_m|$ ist wird in jedem Punkte

$$| P_n - \Pi_n | \leq \frac{\varepsilon}{2}$$

gelten; denn die Differenz $|\Pi_n(x_1, x_2, \ldots x_m) - \Pi_n(y_1, y_2, \ldots y_m)|$, wo $(y_1, y_2, \ldots, y_m)$ ein Punkt der genannten Kugel ist, wird $\leq \frac{\varepsilon}{2}$ sein, denn wir haben

$$| \Pi_n(x_1, x_2, \ldots, x_m) - \Pi_n(y_1, y_2, \ldots, y_m) |$$
$$= | G(x_1', x_2', \ldots, x_m') - G(y_1', y_2', \ldots, y_m') |,$$

wo

$$0 \leq x_\nu' < n!\,p_\nu \quad \text{und} \quad 0 \leq y_\nu' < n!\,p_\nu \quad (\nu = 1, 2, \ldots, m),$$

und wo ausserdem gilt

$$| x_\nu' - y_\nu' | < \frac{\delta\left(\frac{\varepsilon}{2}\right)}{2\,|\alpha_\nu|} \qquad (\operatorname{mod}\ n!\,p_\nu).$$

Wenn nun die Punkte $\xi'^{(1)}, \xi'^{(2)}, \ldots$ und $\eta'^{(1)}, \eta'^{(2)}, \ldots$ gegen $(x'_1, x'_2, \ldots, x'_m)$, bez. $(y'_1, y'_2, \ldots, y'_m)$ konvergieren, wird für alle s von einer gewissen Stelle an die Ungleichung

$$|\xi'^{(s)}_\nu - \eta'^{(s)}_\nu| < \frac{\delta\left(\frac{\varepsilon}{2}\right)}{|\alpha_\nu|} \quad \left(\text{mod } M\left(\frac{\varepsilon}{2}\right)!\, p_\nu\right)$$

gelten, was zur Folge hat, dass

$$|g_s(\xi'^{(s)}) - g_s(\eta'^{(s)})| = |f(x^{(s)}) - f(y^{(s)})| < \frac{\varepsilon}{2},$$

indem

$$|x^{(s)} - y^{(s)}| < \frac{\delta\left(\frac{\varepsilon}{2}\right)}{|\alpha_\nu|} \quad \left(\text{mod } M\left(\frac{\varepsilon}{2}\right)!\, p_\nu\right) \quad (\nu = 1, 2, \ldots, m),$$

woraus zuletzt folgt, dass

$$|\Pi_n(x_1, x_2, \ldots, x_m) - \Pi_n(y_1, y_2, \ldots, y_m)| \leq \frac{\varepsilon}{2}.$$

Wir wollen nun zeigen, dass P_n eine Funktion der gewünschten Art ist, das heisst sie approximiert G im ganzen Raume mit ε-Genauigkeit. Ist nämlich $(x_1, x_2, \ldots, x_m)$ ein willkürlicher Punkt im Raume, $(\xi_1^{(n)}, \xi_2^{(n)}, \ldots, \xi_m^{(n)})$ ein Punkt in Ξ_n, der die m Ungleichungen $|x_\nu - \xi_\nu^{(n)}| < \dfrac{1}{n|\alpha_\nu|}$ $(\nu = 1, 2, \ldots, m)$ erfüllt, und ist $(x'_1, x'_2, \ldots, x'_m)$ und $(\xi_1'^{(n)}, \xi_2'^{(n)}, \ldots, \xi_m'^{(n)})$ die mit ihnen äquivalenten Punkte im Bereich $0 \leq x_\nu < n!\,|p_\nu|$ $(\nu = 1, 2, \ldots, m)$, hat man

$$\begin{aligned}|G(x_1, x_2, \ldots, x_m) - P_n(x_1, x_2, \ldots, x_m)| &= |G(x_1, x_2, \ldots, x_m) \\ - g_n(\xi_1^{(n)}, \xi_2^{(n)}, \ldots, \xi_m^{(n)}) + g_n(\xi_1'^{(n)}, \xi_2'^{(n)}, \ldots, \xi_m'^{(n)}) &- G(x'_1, x'_2, \ldots, x'_m) \\ + \Pi_n(x_1, x_2, \ldots, x_m) - P_n(x_1, x_2, \ldots, x_m)| &< \frac{\varepsilon}{4} + \frac{\varepsilon}{4} + \frac{\varepsilon}{2} = \varepsilon.\end{aligned}$$

Damit ist gezeigt, dass B eine Teilmenge von C ist, und auf Grund des vorhergehenden, dass A und B identisch sind.

III.

Gegeben ist eine abzählbar unendliche Folge von linear unabhängigen Zahlen $\alpha_1, \alpha_2, \ldots$ Wir wollen die Menge aller Funktionen bestimmen, welche gleichmässig für alle x durch eine endliche Summe $\sum c_\alpha e^{i\alpha x}$ approximiert werden können. Hierbei bedeutet α eine Zahl von der Form $r_1\alpha_1 + r_2\alpha_2 + \ldots + r_m\alpha_m$ mit rationalen Koeffizienten r. Die zu untersuchende Funktionenmenge bezeichnen wir wieder mit A. Das dem in II entwickelten analoge Kriterium ist hier das folgende: **Damit $f(x)$ zu A gehöre, ist notwendig und hinreichend, dass zu einem $\varepsilon > 0$ ein $\delta = \delta(\varepsilon)$, $N = N(\varepsilon)$ und $M = M(\varepsilon)$ existieren, so dass jedes τ, welches die N diophantischen Ungleichungen**

$$(17)\quad \begin{cases} |\tau\alpha_1| < \delta \pmod{2\pi M} \\ |\tau\alpha_2| < \delta \pmod{2\pi M} \\ \ldots\ldots\ldots\ldots\ldots \\ |\tau\alpha_N| < \delta \pmod{2\pi M} \end{cases}$$

erfüllt, eine zu ε passende Verschiebungszahl von $f(x)$ ist, das heisst die Ungleichung

$$|f(x+\tau) - f(x)| < \varepsilon \qquad -\infty < x < \infty$$

erfüllt.

Wie im vorhergehenden wollen wir die Menge aller Funktionen, welche die genannte ε-δ-M-N-Bedingung erfüllen, mit B bezeichnen. Es ist leicht zu zeigen, dass A eine Teilmenge von B ist; es sei nämlich $f(x)$ eine Funktion aus A; diese kann mit $\frac{\varepsilon}{3}$ Genauigkeit durch eine Summe $S(x) = \sum c_{r_1 r_2 \ldots r_m} e^{i(r_1\alpha_1 + r_2\alpha_2 + \ldots + r_m\alpha_m)x}$ approximiert werden. Hiernach wähle man N so gross, dass alle die α-Zahlen, die in $S(x)$ vorkommen, in (17) mitberüchsichtigt werden, und als M wähle man den Hauptnenner aller in S vorkommenden r-Zahlen; endlich wähle man δ so klein, dass für jedes τ, welches die Ungleichungen (17) erfüllt, gelten soll, dass

$$|S(x+\tau) - S(x)| < \frac{\varepsilon}{3};$$

für ein solches τ hat man

$$|f(x+\tau) - f(x)| < \varepsilon.$$

Wir gehen nun dazu über die Identität von A und B nachzuweisen, indem wir analog mit dem vorhergehenden zeigen, dass beide identisch sind mit der Menge C bestehend aus den Funktionen

$$f(x) = G(x, x, \ldots),$$

wo $G(x_1, x_2, \ldots)$ eine stetige grenzperiodische Funktion im Raume von abzählbar unendlich vielen Dimensionen mit dem Periodensystem

$$(18) \qquad (p_1, p_2, \ldots) = \left(\frac{2\pi}{\alpha_1}, \frac{2\pi}{\alpha_2}, \ldots\right)$$

ist.

C ist auch hier eine Teilmenge von A, denn ist $f(x)$ eine Funktion aus C, also $f(x) = G(x, x, \ldots.)$, kann sie gewiss mit einer endlichen Summe von der Form $\sum c_{r_1 r_2 \ldots r_m} e^{i(r_1\alpha_1 + r_2\alpha_2 + \ldots + r_m\alpha_m)x}$ approximiert werden. Hierzu braucht man nur die Funktion $G(x_1, x_2, \ldots)$ gleichmässig im ganzen Raume durch eine Summe von der Form $S(x_1, x_2, \ldots, x_m) = \sum c_{r_1 r_2 \ldots r_m} e^{i(r_1\alpha_1 x_1 + r_2\alpha_2 x_2 + \ldots + r_m\alpha_m x_m)}$ anzunähern, was zufolge der Annahme, dass $G(x_1, x_2, \ldots)$ grenzperiodisch mit dem Periodensystem (18) ist, gewiss geschehen kann. $S(x, x, \ldots, x)$ ist dann eine approximierende Summe der behaupteten Art. Es erübrigt also nur zu zeigen, dass B eine Teilmenge von C ist.

Wir bestimmen für ein willkürlich gegebenes x und ein gegebenes s ($s = 1, 2, \ldots$) alle Punkte $(\xi_1^{(s)}, \xi_2^{(s)}, \ldots)$ des Raumes, deren Koordinaten für jedes $\nu = 1, 2, \ldots$ die Kongruenzen

(19) $$\xi_\nu^{(s)} \equiv x \pmod{s!\, p_\nu}$$

erfüllen. Die Punktmenge, welche wir erhalten, indem x alle Werte durchläuft, bezeichnen wir mit Ξ_s. Danach definieren wir die Funktion $g_s(\xi_1^{(s)}, \xi_2^{(s)}, \ldots)$ in der Menge Ξ_s durch die Gleichung

$$g_s(\xi_1^{(s)}, \xi_2^{(s)}, \ldots) = f(x),$$

wo x der dem Punkte $(\xi_1^{(s)}, \xi_2^{(s)}, \ldots)$ im Sinne von (19) entsprechende Wert ist. Wie im vorhergehenden Fall ist auch jetzt Ξ_s eine in unserem jetzigen Raume überall dicht liegende Punktmenge, und wiederum gilt sowohl, dass

$$\Xi_1 > \Xi_2 > \ldots,$$

wie auch, wenn $p > q$, dass

$$g_p(\xi_1^{(p)}, \xi_2^{(p)}, \ldots) = g_q(\xi_1^{(p)}, \xi_2^{(p)}, \ldots).$$

Die Funktion $G(x_1, x_2, \ldots)$ wird dann in einem beliebigen Punkte $(x_1, x_2, \ldots)$ wie folgt definiert: Wir wählen eine Folge von Punkten $(\xi_1^{(1)}, \xi_2^{(1)}, \ldots)$, $(\xi_1^{(2)}, \xi_2^{(2)}, \ldots)$, ..., welche gegen $(x_1, x_2, \ldots)$ konvergiert, und wo der s-te Punkt Ξ_s angehört. $G(x_1, x_2, \ldots)$ soll dann der Grenzwert von $g_1(\xi^{(1)})$, $g_2(\xi^{(2)})$, ... sein:

$$G(x_1, x_2, \ldots) = \lim_{s \to \infty} g_s(\xi_1^{(s)}, \xi_2^{(s)}, \ldots).$$

Im folgenden werden nicht willkürliche Folgen $\xi^{(1)}, \xi^{(2)}, \ldots$, sondern solche spezielle benutzt, bei welchen

$$|x_\nu - \xi_\nu^{(s)}| < \frac{1}{s|\alpha_\nu|} \qquad (\nu = 1, 2, \ldots, s)$$

ist; man erreicht dadurch, dass alle Berechnungen wesentlich einfacher werden.

Um die Konvergenz einer Folge $g_1(\xi^{(1)}), g_2(\xi^{(2)}), \ldots$ zu zeigen, betrachten wir die Differenz zwischen dem p-ten und dem q-ten Glied der Folge, für welche wir, wenn $p > q$, haben, dass

$$|g_p(\xi^{(p)}) - g_q(\xi^{(q)})| = |g_q(\xi^{(p)}) - g_q(\xi^{(q)})| = |f(x^{(p)}) - f(x^{(q)})|.$$

Hierbei gilt

$$|\alpha_\nu(x^{(q)} - x^{(p)})| < \frac{2}{q} \pmod{2\pi q!} \qquad (\nu = 1, 2, \ldots, q)$$

und hat man q so gross gewählt, dass $\frac{2}{q} < \delta(\varepsilon)$, $q > M(\varepsilon)$ und $q > N(\varepsilon)$, gilt $|g_p(\xi^{(p)}) - g_q(\xi^{(q)})| < \varepsilon$. Ähnlich wie zuvor zeigt man, dass der die Funktion G definierende Grenzwert unabhängig von der gewählten Folge $\xi^{(1)}, \xi^{(2)}, \ldots$ ist.

Dass $G(x_1, x_2, \ldots)$ stetig ist, ersieht man wie folgt: Hat man zwei Punkte $(x_1', x_2', \ldots)$ und $(x_1'', x_2'', \ldots)$ mit den Koordinatendifferenzen

$$|x_\nu' - x_\nu''| < \delta, \qquad (\nu = 1, 2, \ldots N),$$

hat man auch

(20) $$|G(x_1', x_2', \ldots) - G(x_1'', x_2'', \ldots)| \leq \varepsilon,$$

wenn nur $\delta < \frac{1}{3}\frac{\delta(\varepsilon)}{|\alpha_\nu|}$ $(\nu = 1, 2, \ldots, N)$ und $N \geq N(\varepsilon)$, denn wir können in diesem Falle ein $s_0 > N$ so bestimmen, dass für alle $s > s_0$ sowohl

$$|\xi_\nu'^{(s)} - \xi_\nu''^{(s)}| < \frac{\delta(\varepsilon)}{|\alpha_\nu|} \qquad (\nu = 1, 2, \ldots, N)$$

wie $s > M(\varepsilon)$ erfüllt ist. Dann gilt

$$|g_s(\xi'^{(s)}) - g_s(\xi''^{(s)})| = |f(x'^{(s)}) - f(x''^{(s)})| < \varepsilon,$$

denn für $s > s_0$ hat man

$$|\alpha_\nu(x'^{(s)} - x''^{(s)})| < \delta(\varepsilon) \pmod{2\pi\, M(\varepsilon)!} \quad (\nu = 1, 2, \ldots, N(\varepsilon)),$$

woraus (20) folgt.

Um nun zu zeigen, dass die so definierte Funktion auch grenzperiodisch ist, benutzen wir das folgende Kriterium: Damit eine stetige Funktion $G(x_1, x_2, \ldots)$ grenzperiodisch mit dem Periodensystem $(p_1, p_2, \ldots)$ sei, ist (notwendig und) hinreichend, dass es zu einem $\varepsilon > 0$ eine Zahl N und eine rationale Zahl $r \neq 0$ geben soll, so dass in zwei Punkten $(x'_1, x'_2, \ldots)$ und $(x''_1, x''_2, \ldots)$, deren Koordinaten die N Kongruenzen

$$x'_\nu \equiv x''_\nu \pmod{r\, p_\nu} \qquad (\nu = 1, 2, \ldots, N)$$

erfüllen, gilt

$$|G(x'_1, x'_2, \ldots) - G(x''_1, x''_2, \ldots)| < \varepsilon.$$

Durch die Benutzung dieses Satzes kann die Konstruktion der reinperiodischen Annäherungsfunktion wie auch die hierbei nötige Glättung gespart werden, da diese im Laufe des Beweises des genannten Satzes vorgenommen werden.

Betrachten wir zwei willkürliche Punkte $(x'_1, x'_2, \ldots)$ und $(x''_1, x''_2, \ldots)$, für welche die T Kongruenzen

$$x'_\nu \equiv x''_\nu \pmod{T!\, p_\nu} \qquad (\nu = 1, 2, \ldots, T) \tag{21}$$

gelten, können in den Folgen $\xi'^{(1)}, \xi'^{(2)}, \ldots$ und $\xi''^{(1)}, \xi''^{(2)}, \ldots$, welche zur Bestimmung der Funktionswerte verwendet werden, die T-ten Punkte $\xi'^{(T)}$ und $\xi''^{(T)}$ so gewählt werden, dass auch

$$\xi'^{(T)}_\nu \equiv \xi''^{(T)}_\nu \pmod{T!\, p_\nu} \qquad (\nu = 1, 2, \ldots, T)$$

gilt, indem die Forderung, dass ein Punkt der T-te Annäherungspunkt für einen Punkt $(x_1, x_2, \ldots)$ sein soll, nur die T ersten Koordinaten betrifft. Die Punkte $\xi'^{(T)}$ und $\xi''^{(T)}$ sind dann äquivalent mit demselben Punkt $(x, x, \ldots)$ der Hauptdiagonalen, und man hat

$$g_T(\xi'^{(T)}) = g_T(\xi''^{(T)}).$$

Wurde also T so gross gewählt, dass gleichmässig im ganzen Raume gilt

$$|G(x_1, x_2, \ldots) - g_T(\xi_1^{(T)}, \xi_2^{(T)}, \ldots)| < \frac{\varepsilon}{2},$$

wird für zwei Punkte, welche (21) erfüllen, auch

$$|G(x'_1, x'_2, \ldots) - G(x''_1, x''_2, \ldots)| < \varepsilon$$

richtig sein. Hiermit ist gezeigt, dass $G(x_1, x_2, \ldots)$ eine Funktion der gewünschten Art ist und damit, dass B eine Teilmenge von C ist, woraus die Identität von A und B folgt.

Færdig fra Trykkeriet den 12. Marts 1928.

MATHEMATISK-FYSISKE MEDDELELSER

UDGIVNE AF

DET KGL. DANSKE VIDENSKABERNES SELSKAB

6. BIND (Kr. 17,00):

		Kr. Ø.
1.	NIELSEN, NIELS: Sur l'opération itérative des Équations de Lagrange. 1924	3.10
2.	UREY, H. C.: On the Effect of perturbing Electric Fields on the Zeeman Effect of the Hydrogen Spectrum. 1924	0.65
3.	BØGGILD, O. B.: On the Labradorization of the Feldspars. With one plate. 1924	3.00
4.	PEDERSEN, P. O.: Om elektriske Gnister. II. Eksperimentelle Undersøgelser over Gnistforsinkelse og Gnistdannelse. Med 7 Tavler. 1924	4.30
5.	JUEL, C.: Über Flächen von Maximalindex. 1924	1.25
6.	NIELSEN, NIELS: Sur une Équation de Lagrange. 1924	1.25
7.	HEVESY, G. DE: Recherches sur les propriétés du Hafnium. Avec 2 planches. 1925	6.25
8.	BOHR, HARALD: Neuer Beweis eines allgemeinen Kronecker'schen Approximationssatzes. 1924	0.50
9.	BJERRUM NIELS and EBERT, LUDWIG: On some recent Investigations concerning Mixtures of Strong Electrolytes (Transference Numbers and Amalgam Equilibria). 1925	0.75
10.	LANDAU, EDM.: Die Ungleichungen für zweimal differentiierbare Funktionen. 1925	1.60

7. BIND (Kr. 17,70):

1.	BOHR, HARALD: Unendlich viele lineare Kongruenzen mit unendlich vielen Unbekannten. 1925	1.40
2.	HARTMANN, JUL., and TROLLE, BIRGIT: On Beat-phenomena in Cylindrical Tubes exposed to Sound-waves. With three plates. 1925	2.85
3.	PAULI, W. jr.: Ueber die Intensitäten der im elektrischen Feld erscheinenden Kombinationslinien. 1925	0.65
4.	HARDY, G. H. and LITTLEWOOD, J. E.: A theorem concerning series of positive terms, with applications to the theory of functions. 1925	0.90
5.	STEFFENSEN, J. F.: On a Generalization of Nörlund's Polynomials. 1926	1.00
6.	HARTMANN, JUL., and TROLLE, BIRGIT: New investigation on the air jet generator for acoustic wawes. 1926	2.40
7.	MOLLERUP, JOHS.: Sur l'approximation d'un nombre irrationnel par des carrés rationnels. 1926	0.80
8.	NIELSEN, NIELS: Sur certains développements d'une fonction holomorphe. 1926	0.75

8. BIND (under Pressen):

Det Kgl. Danske Videnskabernes Selskab.
Mathematisk-fysiske Meddelelser. **VIII,** 7.

A GENERAL SUMMATION FORMULA

BY

J. F. STEFFENSEN

KØBENHAVN
HOVEDKOMMISSIONÆR: ANDR. FRED. HØST & SØN, KGL. HOF-BOGHANDEL
BIANCO LUNOS BOGTRYKKERI
1928

Pris: Kr. 1,25.

Det Kgl. Danske Videnskabernes Selskabs videnskabelige Meddelelser udkommer fra 1917 indtil videre i følgende Rækker:

Historisk-filologiske Meddelelser,
Filosofiske Meddelelser,
Mathematisk-fysiske Meddelelser,
Biologiske Meddelelser.

Hele Bind af disse Rækker sælges 25 pCt. billigere end Summen af Bogladepriserne for de enkelte Hefter.

Selskabets Hovedkommissionær er *Andr. Fred. Høst & Søn*, Kgl. Hof-Boghandel, København.

Det Kgl. Danske Videnskabernes Selskab.
Mathematisk-fysiske Meddelelser. **VIII**, 7.

A GENERAL SUMMATION FORMULA

BY

J. F. STEFFENSEN

KØBENHAVN
HOVEDKOMMISSIONÆR: ANDR. FRED. HØST & SØN, KGL. HOF-BOGHANDEL
BIANCO LUNOS BOGTRYKKERI
1928

1. In a recent paper[1] I have introduced certain polynomials $\overset{\nu}{x}_{\omega n}$, defined by having to satisfy the equations

$$\Delta \overset{\nu}{x}_{\omega n} = \nu \overset{\nu-1}{x}_{\omega, n-1} \tag{1}$$

$$\underset{\omega}{\Delta} \overset{\nu}{x}_{\omega n} = \nu \overset{\nu-1}{x}_{\omega n}, \tag{2}$$

besides the initial conditions

$$\overset{0}{x}_{\omega n} = 1, \qquad \overset{\nu}{x}_{\omega 0} = x(x-\omega) \ldots (x-\nu\omega+\omega). \tag{3}$$

These polynomials are the natural instrument for dealing with some of the most important problems of the theory of interpolation, such as expressing a difference of a certain order and with a given interval in terms of differences with another given interval, or expressing a sum of a certain order and with a given interval in terms of sums with another given interval. In the present paper we intend to occupy ourselves with the latter problem.

2. Proceeding in a way similar to that followed in deriving the generalized Euler—Maclaurin summation-formula[2], we begin by defining a function $\overline{x}^{\nu}_{\omega r}$, distinguished by a bar above the argument, r being a positive integer, which in the semi-closed interval $0 \leq x < r$ is identical with $\overset{\nu}{x}_{\omega r}$, that is

[1] On a Generalization of Nörlunds Polynomials. Det Kgl. Danske Videnskabernes Selskab, Mathematisk-fysiske Meddelelser, VII, 5 (1926). Quoted below as "G. N. P."

[2] Nörlund: Differenzenrechnung, pp. 154—161; Transactions of the American Mathematical Society, Vol. 25, No. 1, pp. 36—46.

$$\overline{x}^{\nu}_{\omega r} = x^{\nu}_{\omega r} \qquad (0 \leq x < r), \tag{4}$$

while for all x (positive, negative or zero)

$$\Delta^{r} \overline{x}^{\nu}_{\omega r} = 0. \tag{5}$$

As the latter relation is a linear relation between $\overline{x}^{\nu}_{\omega r}$, $\overline{x+1}^{\nu}_{\omega r}, \ldots \overline{x+r}^{\nu}_{\omega r}$, it is seen that $\overline{x}^{\nu}_{\omega r}$ is completely determined by (4) and (5).

It is easy to form an explicit expression for the calculation of $\overline{x}^{\nu}_{\omega r}$. Let k be an integer (positive, negative or zero), and let $0 \leq \theta < 1$. We may, then, always write $x = k + \theta$, and it may be proved that, putting $\binom{\nu}{s} = 0$ for $s > \nu$,

$$\overline{k+\theta}^{\nu}_{\omega r} = \sum_{s=0}^{r-1} \binom{\nu}{s} k^{(s)} \theta^{\nu-s}_{\omega, r-s} \qquad (0 \leq \theta < 1). \tag{6}$$

We only have to show that this expression satisfies (4) and (5). Now, performing Δ^{r} on both sides of (6) with respect to k, all the terms on the right vanish, so that (5) is satisfied. As regards (4), we assume for a moment $0 \leq k + \theta < r$, so that $0 \leq k \leq r-1$. As $k^{(s)}$ then vanishes for $s > k$, (6) becomes

$$\overline{k+\theta}^{\nu}_{\omega r} = \sum_{s=0}^{k} \binom{\nu}{s} k^{(s)} \theta^{\nu-s}_{\omega, r-s}$$

or, as $\binom{\nu}{s}$ vanishes for $s > \nu$, and $k^{(s)}$ for $s > k$,

$$\overline{k+\theta}^{\nu}_{\omega r} = \sum_{s=0}^{\nu} \binom{\nu}{s} k^{(s)} \theta^{\nu-s}_{\omega, r-s},$$

or, by G. N. P. (5),

$$\overline{k+\theta}^{\nu}_{\omega r} = (k+\theta)^{\nu}_{\omega r},$$

so that (4) is also satisfied.

It should be noted that if $\nu < r$, (4) is valid for all x, as in that case $\Delta^r \overset{\nu}{x}_{\omega r} = 0$ for all x, so that (5) is satisfied by $\overset{\nu}{x}_{\omega r}$ itself. The formula (6), though valid in all cases, need therefore only be applied if $\nu \geq r$.

3. We are now able to prove that

$$\Delta \overset{-\nu}{x}_{\omega r} = \nu \overset{-\nu-1}{x}_{\omega, r-1}; \tag{7}$$

as $\overset{-\nu}{x}_{\omega 0}$ has not yet been defined, we may put

$$\overset{-\nu}{x}_{\omega 0} = 0, \tag{8}$$

so that (7), owing to (5), also holds for $r = 1$.

We need only difference (6) with respect to k, the result being

$$\begin{aligned}\Delta \overline{k+\theta}^{\nu}_{\omega r} &= \sum_{s=0}^{r-1} \binom{\nu}{s} s k^{(s-1)} \overset{\nu-s}{\theta}_{\omega, r-s} \\ &= \nu \sum_{s=1}^{r-1} \binom{\nu-1}{s-1} k^{(s-1)} \overset{\nu-s}{\theta}_{\omega, r-s} \\ &= \nu \sum_{s=0}^{r-2} \binom{\nu-1}{s} k^{(s)} \overset{\nu-s-1}{\theta}_{\omega, r-s-1} \\ &= \nu \overline{k+\theta}^{\nu-1}_{\omega, r-1},\end{aligned}$$

or (7).

4. It follows evidently from (4) and (6) that the function $\overset{-\nu}{x}_{\omega r}$ is continuous in the interval $0 \leq x < r$ and also in the interval between any two consecutive integers. If $\nu < r$, $\overset{-\nu}{x}_{\omega r}$ is identical with $\overset{\nu}{x}_{\omega r}$ and, therefore, continuous for all x. If $\nu > r$, it can be proved that $\overset{-\nu}{x}_{\omega r}$ is still continuous for all x; but in the case of $\nu = r$ we shall arrive at the result that $\overset{-r}{x}_{\omega r}$ possesses points of discontinuity when x is an integer $\geq r$ or ≤ 0; other points of discontinuity do not exist.

Let us first assume $\nu > r$. The relation (5), written in full, is

$$\overline{x+r}^{\nu}_{\omega r} - \binom{r}{1}\overline{x+r-1}^{\nu}_{\omega r} + \binom{r}{2}\overline{x+r-2}^{\nu}_{\omega r} - \cdots + (-1)^r \overline{x}^{\nu}_{\omega r} = 0. \qquad (9)$$

This is valid for all x. Putting $x = 0$, and making use of (4), we have

$$\overline{r}^{\nu}_{\omega r} - \binom{r}{1}(r-1)^{\nu}_{\omega r} + \binom{r}{2}(r-2)^{\nu}_{\omega r} - \cdots + (-1)^r\, 0^{\nu}_{\omega r} = 0$$

or

$$\varDelta^r 0^{\nu}_{\omega r} + \overline{r}^{\nu}_{\omega r} - r^{\nu}_{\omega r} = 0.$$

But if $\nu > r$, $\varDelta^r 0^{\nu}_{\omega r} = \nu^{(r)} 0^{\nu - r}_{\omega 0}$ vanishes, and we have, therefore,

$$\overline{r}^{\nu}_{\omega r} = r^{\nu}_{\omega r} \qquad (\nu > r).$$

It follows that $\overline{x}^{\nu}_{\omega r}$ is, for $\nu > r$, continuous in the closed interval $0 \leq x \leq r$, and (9) shows clearly that $\overline{x}^{\nu}_{\omega r}$ must then be continuous for all x, as was to be proved.

It remains to investigate the case $\nu = r$. We obtain by (6) for $\nu = r$

$$\overline{k+\theta}^{r}_{\omega r} = \sum_{s=0}^{r-1} \binom{r}{s} k^{(s)} \theta^{r-s}_{\omega, r-s}$$

$$= \sum_{s=0}^{r} \binom{r}{s} k^{(s)} \theta^{r-s}_{\omega, r-s} - k^{(r)}$$

or, by G. N. P. (5),

$$\overline{k+\theta}^{r}_{\omega r} = (k+\theta)^{r}_{\omega r} - k^{(r)} \qquad (0 \leq \theta < 1). \qquad (10)$$

Hence we have, for $\theta = 0$,

$$\bar{k}^{r}_{\omega r} = k^{r}_{\omega r} - k^{(r)} \qquad (11)$$

and for $\theta \to 1$

$$\overline{k+1-0}^{r}_{\omega r} = (k+1)^{r}_{\omega r} - k^{(r)}$$

or, replacing k by $k-1$,

$$\overline{k-0}^{r}_{\omega r} = k^{r}_{\omega r} - (k-1)^{(r)}. \qquad (12)$$

Subtracting (12) from (11), we have

$$\bar{k}^{r}_{\omega r} - \overline{k-0}^{r}_{\omega r} = (k-1)^{(r)} - k^{(r)}$$
$$= -\Delta (k-1)^{(r)}$$

or finally

$$\bar{k}^{r}_{\omega r} - \overline{k-0}^{r}_{\omega r} = -r(k-1)^{(r-1)}. \qquad (13)$$

This expression shows that $\bar{x}^{r}_{\omega r}$ has discontinuities at all the points $x = r,\ r+1,\ r+2, \ldots$ and $x = 0, -1, -2, \ldots$; other discontinuities do not exist, as the expression on the right of (13) vanishes, if k has one of the values $1, 2, \ldots, r-1$ $(r > 1)$.

It is worth noting that the amount of the discontinuity, or the height of the "jump", is independent of ω, as appears from (13).

5. The relation G. N. P. (34), or

$$x^{\nu}_{\omega n} = (-1)^{\nu} (n-x)^{\nu}_{-\omega, n} \qquad (14)$$

also holds, with an obvious reservation, for the functions $\bar{x}^{\nu}_{\omega r}$. We begin by noting that instead of (6) we may use the following relation for the calculation of $\bar{x}^{\nu}_{\omega r}$

$$\overline{r-k-1+\theta}^{\nu}_{\omega r} = (-1)^{\nu} \sum_{s=0}^{r-1} \binom{\nu}{s} k^{(s)} (1-\theta)^{\nu-s}_{-\omega, r-s} \qquad (15)$$

where k and θ have the same meanings as before. For, putting $x = r-k-1+\theta$, it may be proved that the expression (15) satisfies (4) and (5), as we proceed to show.

As regards (5), it is seen at once that, differencing r times on both sides with respect to $-k$, all the terms on the right vanish, so that (5) is satisfied.

Next, we assume $0 \leq x < r$, that is $0 \leq r-k-1+\theta < r$, so that $0 \leq r-k-1 \leq r-1$, or $0 \leq k \leq r-1$. If k is comprised between these limits we may, as above, replace the upper limit of summation in (15) by ν, so that

$$\overline{r-k-1+\theta}^{\nu}_{\omega r} = (-1)^{\nu} \sum_{s=0}^{\nu} \binom{\nu}{s} k^{(s)} (1-\theta)^{\nu-s}_{-\omega, r-s}$$

or, by G. N. P. (5),

$$\overline{r-k-1+\theta}^{\nu}_{\omega r} = (-1)^{\nu} (k+1-\theta)^{\nu}_{-\omega, r}$$

or finally, by (14),

$$\overline{r-k-1+\theta}^{\nu}_{\omega r} = (r-k-1+\theta)^{\nu}_{\omega r},$$

so that also (4) is satisfied.

Having thus established (15), we may, if we exclude the value $\theta = 0$, replace θ by $1-\theta$ in (15). Changing the sign of ω, we thus obtain for $0 < \theta < 1$

$$\overline{r-k-\theta}^{\nu}_{-\omega, r} = (-1)^{\nu} \sum_{s=0}^{r-1} \binom{\nu}{s} k^{(s)} \theta^{\nu-s}_{\omega, r-s}. \tag{16}$$

But comparison of this relation and (6) shows that

$$\overline{x}^{\nu}_{\omega r} = (-1)^{\nu} \overline{r-x}^{\nu}_{-\omega, r}, \tag{17}$$

provided that x is not an integer. If x is an integer, (17) is still valid for $\nu \neq r$, as in that case $\overline{x}^{\nu}_{\omega r}$ is continuous for all x. But the case $\nu = r$ must be treated by putting

$\theta = 0$ in (6) and letting $\theta \to 0$ in (16), the result being the relation

$$\overline{k}^{r}_{\omega r} = (-1)^r \overline{r-k-0}^{r}_{-\omega, r}. \tag{18}$$

6. We shall now assume that

$$0 < \omega < 1, \qquad 0 \leq \theta < 1 - \omega, \tag{19}$$

so that $0 < \theta + \omega < 1$. It follows that (6) remains valid, if θ is replaced by $\theta + \omega$, and we therefore obtain

$$\begin{aligned} \underset{\omega}{\Delta} \overline{k+\theta}^{\nu}_{\omega r} &= \sum_{s=0}^{r-1} \binom{\nu}{s} k^{(s)} (\nu - s)\, \theta^{\nu-s-1}_{\omega, r-s} \\ &= \nu \sum_{s=0}^{r-1} \binom{\nu-1}{s} k^{(s)}\, \theta^{\nu-s-1}_{\omega, r-s} \\ &= \nu\, \overline{k+\theta}^{\nu-1}_{\omega r} \end{aligned}$$

or

$$\underset{\omega}{\Delta} \overline{x}^{\nu}_{\omega r} = \nu\, \overline{x}^{\nu-1}_{\omega r} \qquad \begin{pmatrix} 0 < \omega < 1 \\ k-1 \leq x < k-\omega \end{pmatrix} \tag{20}$$

where k has one of the values $0, \pm 1, \pm 2, \ldots$

If, in this formula, we let $\omega \to 0$, the symbol $\underset{0}{\Delta}$ may be replaced by the symbol of Differentiation D at every point where the derivative exists; at points where it does not exists $\underset{0}{\Delta}$ means the differential coefficient to the right.

For $x \to k - \omega$ we find from (20)

$$\frac{1}{\omega}\left(\overline{k-0}^{\nu}_{\omega r} - \overline{k-\omega}^{\nu}_{\omega r}\right) = \nu\, \overline{k-\omega}^{\nu-1}_{\omega r}$$

whence

$$\underset{\omega}{\Delta} \overline{k-\omega}^{\nu}_{\omega r} = \nu\, \overline{k-\omega}^{\nu-1}_{\omega r} + \frac{1}{\omega}\left(\overline{k}^{\nu}_{\omega r} - \overline{k-0}^{\nu}_{\omega r}\right). \tag{21}$$

It follows that, if $\nu \neq r$, (20) is still valid for $x = k - \omega$, while in the case $\nu = r$ we obtain, by (13),

$$\underset{\omega}{\Delta} \overline{k - \omega}_{\omega r}^{\,r} = r\,\overline{k - \omega}_{\omega r}^{\,r-1} - \frac{r}{\omega}(k-1)^{(r-1)}. \tag{22}$$

The supplementary term in (22), representing the discontinuity, vanishes for $k = 1, 2, \ldots, r-1$, $(r > 1)$, so that (20) is still valid for $\nu = r$, $x = k - \omega$, if k has one of the values $1, 2, \ldots, r-1$, $(r > 1)$.

7. After these preliminaries, the problem of summation may be attacked. We assume henceforth that $\frac{1}{\omega}$ is a positive integer > 1 whence follows, in particular, that the condition $0 < \omega < 1$, implied in (20), is satisfied.

Let h be a parameter, positive, negative or zero, of which we may dispose afterwards, and let us consider the expression

$$V_s^{(\nu)} = -\omega \sum_{\mu=0}^{\frac{1}{\omega}-1} \frac{\overline{\mu\omega + h}_{\omega s}^{\,\nu+s}}{(\nu+s)!} \underset{\omega}{\Delta}^{\nu+1} f(x+1-\omega-\mu\omega). \tag{23}$$

This expression may be transformed in the following way which is equivalent with partial summation. As $\underset{\omega}{\Delta}^{\nu+1} = \underset{\omega}{\Delta}^{\nu} \underset{\omega}{\Delta}$, we have

$$V_s^{(\nu)} = -\sum_{\mu=0}^{\frac{1}{\omega}-1} \frac{\overline{\mu\omega + h}_{\omega s}^{\,\nu+s}}{(\nu+s)!} \underset{\omega}{\Delta}^{\nu} f(x+1-\mu\omega)$$
$$+ \sum_{\mu=0}^{\frac{1}{\omega}-1} \frac{\overline{\mu\omega + h}_{\omega s}^{\,\nu+s}}{(\nu+s)!} \underset{\omega}{\Delta}^{\nu} f(x+1-\omega-\mu\omega)$$

or, if in the first sum we replace μ by $\mu + 1$,

$$V_s^{(\nu)} = -\sum_{\mu=-1}^{\frac{1}{\omega}-2} \frac{\overline{\mu\omega+\omega+h}_{\omega s}^{\nu+s}}{(\nu+s)!} \mathop{\Delta}_{\omega}^{\nu} f(x+1-\omega-\mu\omega)$$

$$+\sum_{\mu=0}^{\frac{1}{\omega}-1} \frac{\overline{\mu\omega+h}_{\omega s}^{\nu+s}}{(\nu+s)!} \mathop{\Delta}_{\omega}^{\nu} f(x+1-\omega-\mu\omega)$$

which may be reduced to

$$V_s^{(\nu)} = -\omega \sum_{\mu=0}^{\frac{1}{\omega}-1} \frac{\mathop{\Delta}_{\omega} \overline{\mu\omega+h}_{\omega s}^{\nu+s}}{(\nu+s)!} \mathop{\Delta}_{\omega}^{\nu} f(x+1-\omega-\mu\omega)$$

$$+\frac{\overline{1+h}_{\omega s}^{\nu+s}}{(\nu+s)!} \mathop{\Delta}_{\omega}^{\nu} f(x) - \frac{\overline{h}_{\omega s}^{\nu+s}}{(\nu+s)!} \mathop{\Delta}_{\omega}^{\nu} f(x+1).$$

Assuming $s \geq 1$, we have, by (7),

$$\overline{1+h}_{\omega s}^{\nu+s} = \overline{h}_{\omega s}^{\nu+s} + (\nu+s)\, \overline{h}_{\omega,\, s-1}^{\nu+s-1},$$

so that $V_s^{(\nu)}$ may be written

$$\left.\begin{aligned} V_s^{(\nu)} &= -\omega \sum_{\mu=0}^{\frac{1}{\omega}-1} \frac{\mathop{\Delta}_{\omega} \overline{\mu\omega+h}_{\omega s}^{\nu+s}}{(\nu+s)!} \mathop{\Delta}_{\omega}^{\nu} f(x+1-\omega-\mu\omega) \\ &+ \frac{\overline{h}_{\omega,\, s-1}^{\nu+s-1}}{(\nu+s-1)!} \mathop{\Delta}_{\omega}^{\nu} f(x) - \frac{\overline{h}_{\omega s}^{\nu+s}}{(\nu+s)!} \Delta \mathop{\Delta}_{\omega}^{\nu} f(x). \end{aligned}\right\} \quad (24)$$

We now assume that $\nu \geq 1$ and that h is a multiple of ω, that is, $h = p\omega$ where p denotes an integer (positive, negative or zero). In that case we have, according to No. 6, as $\nu + s \mp s$,

$$\mathop{\Delta}_{\omega} \overline{\mu\omega+h}_{\omega s}^{\nu+s} = (\nu+s)\, \overline{\mu\omega+h}_{\omega s}^{\nu+s-1},$$

so that we obtain from (24), by (23),

$$V_s^{(\nu)} = V_s^{(\nu-1)} + \frac{\overline{h}_{\omega, s-1}^{\nu+s-1}}{(\nu+s-1)!} \underset{\omega}{\Delta}^{\nu} f(x) - \frac{\overline{h}_{\omega s}^{\nu+s}}{(\nu+s)!} \Delta \underset{\omega}{\Delta}^{\nu} f(x). \quad (25)$$

Performing the operation Δ^{s-1} on both sides of (25) and summing from $s = 1$ to $s = r$, we find, putting

$$R_\nu = \sum_{s=1}^{r} \Delta^{s-1} V_s^{(\nu)} \quad (26)$$

and taking account of (8),

$$R_\nu = R_{\nu-1} - \frac{\overline{h}_{\omega r}^{\nu+r}}{(\nu+r)!} \Delta^r \underset{\omega}{\Delta}^{\nu} f(x). \quad (27)$$

Summing on both sides of this equation from $\nu = 1$ to $\nu = m$, we obtain

$$R_m = R_0 - \sum_{\nu=1}^{m} \frac{\overline{h}_{\omega r}^{\nu+r}}{(\nu+r)!} \Delta^r \underset{\omega}{\Delta}^{\nu} f(x). \quad (28)$$

It remains to investigate R_0. According to (24)

$$\left.\begin{aligned} V_s^{(0)} &= -\omega \sum_{\mu=0}^{\frac{1}{\omega}-1} \frac{\underset{\omega}{\Delta} \overline{\mu\omega + h_{\omega s}}^{s}}{s!} f(x+1-\omega-\mu\omega) \\ &+ \frac{\overline{h}_{\omega, s-1}^{s-1}}{(s-1)!} f(x) - \frac{\overline{h}_{\omega s}^{s}}{s!} \Delta f(x). \end{aligned}\right\} \quad (29)$$

Now it follows from No. 6 that we have generally

$$\underset{\omega}{\Delta} \overline{\mu\omega + h_{\omega s}}^{s} = s \, \overline{\mu\omega + h_{\omega s}}^{s-1},$$

exception being made at the point

$$\mu\omega + h = k - \omega \tag{30}$$

where the term

$$-\frac{s}{\omega}(k-1)^{(s-1)}$$

must be added to the right-hand side, producing a term

$$+\binom{k-1}{s-1} f(x+1-k+h)$$

in $V_s^{(0)}$. We therefore obtain from (29), performing the operation Δ^{s-1} on both sides and summing from $s = 1$ to $s = r$

$$\left.\begin{aligned} R_0 = &-\omega \sum_{s=1}^{r} \sum_{\mu=0}^{\frac{1}{\omega}-1} \frac{\overline{\mu\omega + h}_{\omega s}^{s-1}}{(s-1)!} \Delta^{s-1} f(x+1-\omega-\mu\omega) \\ &+ \sum_{s=1}^{r} \binom{k-1}{s-1} \Delta^{s-1} f(x+1-k+h) - \frac{\bar{h}_{\omega r}^{r}}{r!} \Delta^{r} f(x). \end{aligned}\right\} \tag{31}$$

This expression may be simplified, if we assume $0 \leq h < r$. In that case $\bar{h}_{\omega r}^{r}$ may be replaced by $h_{\omega r}^{r}$, and it may be concluded from (30) that $r \geq k \geq 1$, so that

$$\sum_{s=1}^{r} \binom{k-1}{s-1} \Delta^{s-1} f(x+1-k+h)$$

$$= \sum_{s=1}^{k} \binom{k-1}{s-1} \Delta^{s-1} f(x+1-k+h) = f(x+h),$$

as follows from the identity

$$\begin{aligned} f(x+h) &= (1+\Delta)^{k-1} E^{-k+1} f(x+h) \\ &= (1+\Delta)^{k-1} f(x+1-k+h). \end{aligned}$$

We thus obtain from (31)

$$\left.\begin{aligned} R_0 = -\omega \sum_{s=1}^{r} \sum_{\mu=0}^{\frac{1}{\omega}-1} \frac{\overline{\mu\omega + h}_{\omega s}^{s-1}}{(s-1)!} \Delta^{s-1} f(x+1-\omega-\mu\omega) \\ + f(x+h) - \frac{h_{\omega r}^{r}}{r!} \Delta^{r} f(x). \end{aligned}\right\} \quad (32)$$

Finally, we insert this expression in (28) where $\overline{h}_{\omega r}^{\nu+r}$ may now be replaced by $h_{\omega r}^{\nu+r}$. Noting that $\overline{\mu\omega+h}_{\omega s}^{s-1}$ $= (\mu\omega + h)_{\omega s}^{s-1}$ according to No. 2, and writing $\mu\omega = 1 - \omega$ $-\nu\omega$, we find

$$\left.\begin{aligned} f(x+h) = \omega \sum_{s=1}^{r} \sum_{\nu=0}^{\frac{1}{\omega}-1} \frac{(h+1-\omega-\nu\omega)_{\omega s}^{s-1}}{(s-1)!} \Delta^{s-1} f(x+\nu\omega) \\ + \sum_{\nu=0}^{m} \frac{h_{\omega r}^{\nu+r}}{(\nu+r)!} \Delta^{r} \underset{\omega}{\Delta^{\nu}} f(x) + R_m \end{aligned}\right\} \quad (33)$$

where

$$0 \leq h = p\omega < r \tag{34}$$

and, by (26) and (23),

$$R_m = -\omega \sum_{s=1}^{r} \sum_{\nu=0}^{\frac{1}{\omega}-1} \frac{\overline{h+1-\omega-\nu\omega}_{\omega s}^{m+s}}{(m+s)!} \Delta^{s-1} \underset{\omega}{\Delta^{m+1}} f(x+\nu\omega). \tag{35}$$

8. The formula (33) is a generalization of Euler-Maclaurin's formula which is obtained for $r = 1$, $\omega \to 0$.

Our formula may be transformed in several ways. Thus, by G. N. P. (38), it may be written

$$\left.\begin{aligned} f(x+h) &= \omega \sum_{s=0}^{r-1} \sum_{\nu=0}^{\frac{1}{\omega}-1} \binom{h-\nu\omega}{s} \Delta^s f(x+\nu\omega) \\ &+ \sum_{\nu=0}^{m} \frac{h_{\omega r}^{\nu+r}}{(\nu+r)!} \Delta^r \underset{\omega}{\Delta}^\nu f(x) + R_m . \end{aligned}\right\} \quad (36)$$

The formula is in reality an identity between the $m + \frac{r}{\omega} + 1$ equidistant values of $f(t)$

$$f(x),\ f(x+\omega),\ f(x+2\omega),\ \ldots\ f(x+m\omega+r). \quad (37)$$

If, for $f(t)$, we take a polynomial of degree not exceeding m, we have $R_m = 0$, and comparison with G. N. P. (59) shows that we have

$$\omega \sum_{s=0}^{r-1} \sum_{\nu=0}^{\frac{1}{\omega}-1} \binom{h-\nu\omega}{s} \Delta^s f(x+\nu\omega) = \sum_{\nu=0}^{r-1} \frac{h_{\omega r}^{\nu}}{\nu!} \Delta^r \underset{\omega}{\Delta}^{\nu-r} f(x). \quad (38)$$

This relation has thus been proved for a polynomial. In order to extend this formula to other functions than polynomials, we note that $\Delta^r \underset{\omega}{\Delta}^{-r} f(x)$ has a definite meaning whether $f(x)$ is a polynomial or not, as the various meanings of $\underset{\omega}{\Delta}^{-r} f(x)$ only differ by a function which is cancelled by the subsequent application of Δ^r, $\frac{1}{\omega}$ being an integer. We have, in fact

$$\Delta^r \underset{\omega}{\Delta}^{-r} = \left[\left(1 + \omega \underset{\omega}{\Delta}\right)^{\frac{1}{\omega}} - 1 \right]^r \underset{\omega}{\Delta}^{-r} \quad (39)$$

or, on comparison with G. N. P. (17),

$$\Delta^r \underset{\omega}{\Delta}^{-r} = \sum_{s=0}^{\left(\frac{1}{\omega}-1\right)r} \frac{0_{\omega,-r}^{s}}{s!} \underset{\omega}{\Delta}^{s} \quad (40)$$

which may also, by G. N. P. (39), be written

$$\Delta^r \underset{\omega}{\Delta}^{-r} = \sum_{s=0}^{\left(\frac{1}{\omega}-1\right)r} \frac{\Delta^r 0_{\omega 0}^{s+r}}{(s+r)!} \underset{\omega}{\Delta}^s, \tag{41}$$

so that the operation $\Delta^r \underset{\omega}{\Delta}^{-r}$ has a well defined meaning whether it is applied to a polynomial or to any other function.

It is now seen that (38) is an identity between the $\frac{r}{\omega}$ values of $f(t)$

$$f(x),\ f(x+\omega),\ f(x+2\omega),\ \ldots\ f(x+r-\omega),$$

both sides being linear functions of these values with coefficients that are independent of $f(t)$. It follows that although (38) was only proved for polynomials, it is valid for any function $f(t)$.

We may therefore write (33) or (36) in the form

$$f(x+h) = \sum_{\nu=0}^{r+m} \frac{h_{\omega r}^{\nu}}{\nu!} \Delta^r \underset{\omega}{\Delta}^{\nu-r} f(x) + R_m. \tag{42}$$

For $\omega \to 0$ we obtain from this a formula due to Nörlund (Differenzenrechnung, p. 160).

9. If we impose certain restrictions on $f(t)$, the remainder R_m may be put into the convenient form

$$R_m = \omega \sum_{\nu=0}^{\infty} \frac{\overline{h-\omega-\nu\omega}_{\omega r}^{\,m+r}}{(m+r)!} \Delta^r \underset{\omega}{\Delta}^{m+1} f(x+\nu\omega). \tag{43}$$

The assumption we make about $f(t)$ is that the expression

$$\sum_{\nu=0}^{\infty} \overline{-\nu\omega}_{\omega r}^{\,m+r} f(z+\nu\omega) \tag{44}$$

must be convergent for $z \geq x$. This condition is, for instance, satisfied, if

$$\lim_{t \to \infty} t^{r+\varepsilon} f(t) = 0 \qquad (\varepsilon > 0); \tag{45}$$

for it follows from (6) that $\overline{x}_{\omega r}^{\nu}$ does not increase more rapidly than $|x|^{r-1}$. The condition (44) does not imply any analytical property of $f(t)$ but only concerns the rapidity with which the function must decrease for $t \to \infty$. It is clear that if (44) is satisfied for a given value of r, it is also satisfied for any smaller value of r.

We may now prove (43) by induction, writing $R_m^{(r)}$ instead of R_m in order to indicate that R_m depends on r. Let us first prove the formula for $r = 1$. We have by (35)

$$R_m^{(1)} = -\omega \sum_{\nu=0}^{\frac{1}{\omega}-1} \frac{\overline{h+1-\omega-\nu\omega}_{\omega 1}^{\,m+1}}{(m+1)!} \mathop{\Delta}_{\omega}^{m+1} f(x+\nu\omega).$$

If to the right-hand side we add the expression

$$\omega \sum_{\nu=0}^{\infty} \frac{\overline{h+1-\omega-\nu\omega}_{\omega 1}^{\,m+1} - \overline{h-\omega-\nu\omega}_{\omega 1}^{\,m+1}}{(m+1)!} \mathop{\Delta}_{\omega}^{m+1} f(x+\nu\omega)$$

which is convergent according to hypothesis, and vanishes identically, as $\overline{x}_{\omega 1}^{\,m+1}$ is periodical with the period 1, we obtain after an obvious reduction

$$R_m^{(1)} = \omega \sum_{\nu=\frac{1}{\omega}}^{\infty} \frac{\overline{h+1-\omega-\nu\omega}_{\omega 1}^{\,m+1}}{(m+1)!} \mathop{\Delta}_{\omega}^{m+1} f(x+\nu\omega)$$

$$- \omega \sum_{\nu=0}^{\infty} \frac{\overline{h-\omega-\nu\omega}_{\omega 1}^{\,m+1}}{(m+1)!} \mathop{\Delta}_{\omega}^{m+1} f(x+\nu\omega)$$

or, writing $\nu + \frac{1}{\omega}$ instead of ν in the first sum and reducing,

$$R_m^{(1)} = \omega \sum_{\nu=0}^{\infty} \frac{\overline{h-\omega-\nu\omega}_{\omega 1}^{\,m+1}}{(m+1)!} \Delta \mathop{\Delta}_{\omega}^{m+1} f(x+\nu\omega),$$

so that (43) is valid for $r = 1$.

It remains to show that if (43) is valid for $r = s-1$, it is also valid for $r = s$. Now, by (35),

$$R_m^{(s)} = R_m^{(s-1)} - \omega \sum_{\nu=0}^{\frac{1}{\omega}-1} \frac{\overline{h+1-\omega-\nu\omega}_{\omega s}^{\,m+s}}{(m+s)!} \Delta^{s-1} \mathop{\Delta}_{\omega}^{m+1} f(x+\nu\omega);$$

hence, if (43) is valid for $r = s-1$,

$$R_m^{(s)} = \omega \sum_{\nu=0}^{\infty} \frac{\overline{h-\omega-\nu\omega}_{\omega,\, s-1}^{\,m+s-1}}{(m+s-1)!} \Delta^{s-1} \mathop{\Delta}_{\omega}^{m+1} f(x+\nu\omega)$$

$$- \omega \sum_{\nu=0}^{\frac{1}{\omega}-1} \frac{\overline{h+1-\omega-\nu\omega}_{\omega s}^{\,m+s}}{(m+s)!} \Delta^{s-1} \mathop{\Delta}_{\omega}^{m+1} f(x+\nu\omega).$$

But, as

$$\overline{h-\omega-\nu\omega}_{\omega,\, s-1}^{\,m+s-1} = \frac{\overline{h+1-\omega-\nu\omega}_{\omega s}^{\,m+s} - \overline{h-\omega-\nu\omega}_{\omega s}^{\,m+s}}{m+s},$$

we find immediately, on reduction,

$$R_m^{(s)} = \omega \sum_{\nu=0}^{\infty} \frac{\overline{h - \omega - \nu\omega}\,^{m+s}_{\omega s}}{(m+s)!} \Delta^{s} \mathop{\Delta}_{\omega}^{m+1} f(x + \nu\omega),$$

and the proof is completed.

10. Before proceeding to establish the desired general summation-formula, we shall make a few remarks about *repeated summation*. The symbol $\mathop{\Delta}\limits_{\omega}^{-1}$ is generally defined in such a way that

$$\mathop{\Delta}_{\omega}^{-1} f(x) = \varphi(x) + \psi_{\omega}(x),$$

$\varphi(x)$ being any particular solution of the difference equation $\mathop{\Delta}\limits_{\omega} \varphi(x) = f(x)$, and $\psi_{\omega}(x)$ being an arbitrary periodic function with the period ω. It will now be advantageous to fix the meaning of $\mathop{\Delta}\limits_{\omega}^{-1}$. We put[1], assuming the convergence,

$$\mathop{\Delta}_{\omega}^{-1} f(x) = -\omega \sum_{\nu=0}^{\infty} f(x + \nu\omega), \tag{46}$$

and it is obvious that, with this definition, $\mathop{\Delta}\limits_{\omega}\mathop{\Delta}\limits_{\omega}^{-1} f(x) = f(x)$, as it should be. For the applications of the operation $\mathop{\Delta}\limits_{\omega}^{-1}$, thus defined, to summation between finite limits, the condition that (46) must be convergent is not a restriction of real importance; for, as we do not assume that $f(x)$ is an analytical function, the summation-process (46) may be applied to any table of *finite extent*, provided we put $f(t) = 0$ for values of t outside the range of the table.

The symbol $\mathop{\Delta}\limits_{\omega}^{-1}$, defined in this particular way, is *commutative* with $\mathop{\Delta}\limits_{\omega}$; for we have

[1] Compare Nörlund: Differenzenrechnung, p. 41.

$$\underset{\omega}{\Delta}^{-1}\underset{\omega}{\Delta}f(x) = -\omega\sum_{\nu=0}^{\infty}\frac{f(x+\nu\omega+\omega)-f(x+\nu\omega)}{\omega}$$

$$= -\sum_{\nu=0}^{\infty}f(x+\nu\omega+\omega)+\sum_{\nu=0}^{\infty}f(x+\nu\omega)$$

$$= f(x) = \underset{\omega}{\Delta}\underset{\omega}{\Delta}^{-1}f(x).$$

From (46) follows, for $\omega = 1$,

$$\Delta^{-1}f(x) = -\sum_{\nu=0}^{\infty}f(x+\nu), \tag{47}$$

and it is easily proved that any two of the symbols Δ, Δ^{-1}, $\underset{\omega}{\Delta}$, $\underset{\omega}{\Delta}^{-1}$ are commutative if, in exchanging the order of two symbols of summation, we assume the absolute convergence of the double sum.

The operation $\underset{\omega}{\Delta}^{-1}$ may be repeated, always assuming the convergence; and we find in the case of absolute convergence

$$\underset{\omega}{\Delta}^{-r}f(x) = (-\omega)^r\sum_{\nu=0}^{\infty}\binom{\nu+r-1}{r-1}f(x+\nu\omega). \tag{48}$$

For this formula is valid for $r = 1$; but being valid for any particular value of r, it is also valid for the following one, as

$$\underset{\omega}{\Delta}^{-1}\underset{\omega}{\Delta}^{-r}f(x) = (-\omega)^{r+1}\sum_{\mu=0}^{\infty}\sum_{\nu=0}^{\infty}\binom{\nu+r-1}{r-1}f(x+\mu\omega+\nu\omega)$$

$$= (-\omega)^{r+1}\sum_{s=0}^{\infty}\sum_{\mu=0}^{s}\binom{s-\mu+r-1}{r-1}f(x+s\omega)$$

$$= (-\omega)^{r+1}\sum_{s=0}^{\infty}\binom{s+r}{r}f(x+s\omega).$$

From (48) we find, putting $\omega = 1$,

$$\Delta^{-r} f(x) = (-1)^r \sum_{\nu=0}^{\infty} \binom{\nu+r-1}{r-1} f(x+\nu). \tag{49}$$

11. In order to derive the desired summation-formula, we will for a moment assume that $f(t)$ is a function that vanishes beyond a certain range, say, for $t > N$. We may then, in (42) and (43), perform the operation Δ^{-r} on both sides, and thus obtain

$$\Delta^{-r} f(x+h) = \sum_{\nu=0}^{r+m} \frac{h_{\omega r}^{\nu}}{\nu!} \Delta_{\omega}^{\nu-r} f(x) + R, \tag{50}$$

$$R = \omega \sum_{\nu=0}^{\infty} \frac{\overline{h-\omega-\nu\omega}_{\omega r}^{\,m+r}}{(m+r)!} \Delta_{\omega}^{m+1} f(x+\nu\omega). \tag{51}$$

But this formula evidently remains valid for $N \to \infty$, if all the sums are convergent.

A sufficient condition for the validity of (50) and (51) is, therefore, that the condition (45) is satisfied in which case all the sums are absolutely convergent. In particular, (50) and (51) may be applied to summation between finite limits, if we put $f(t) = 0$ for values of t outside the range of the table. The parameter h must satisfy the condition (34), and $\frac{1}{\omega}$ is a positive integer > 1.

By keeping the first term on the right of (50) apart, it is seen that the formula may be used for the approximate calculation of $\Delta_{\omega}^{-r} f(x)$ or $\Delta^{-r} f(x+h)$ if, besides one of these sums, we know the sums of lower order $\Delta_{\omega}^{1-r} f(x)$, $\Delta_{\omega}^{2-r} f(x)$, ...

12. The simplest and most important case of (50) is obtained for $r = 1$. The result may be written

$$\omega \sum_{\nu=0}^{\infty} f(x+\nu\omega) = \sum_{\nu=0}^{\infty} f(x+h+\nu) + \sum_{\nu=1}^{m+1} \frac{h_{\omega 1}^{\nu}}{\nu!} \underset{\omega}{\Delta}^{\nu-1} f(x) + R \qquad (52)$$

where $0 \leq h = p\omega < 1$, and

$$R = \omega \sum_{\nu=0}^{\infty} \frac{\overline{h-\omega-\nu\omega}_{\omega 1}^{m+1}}{(m+1)!} \underset{\omega}{\Delta}^{m+1} f(x+\nu\omega). \qquad (53)$$

The explicit expression of $x_{\omega 1}^{\nu}$ is, according to G. N. P. (46),

$$x_{\omega 1}^{\nu} = \sum_{s=0}^{\nu} \binom{\nu}{s} 0_{\omega 1}^{\nu-s} x_{\omega 0}^{s}, \qquad (54)$$

and $\overline{x}_{\omega 1}^{\nu}$ is a function, periodical with the period 1, which in the interval $0 \leq x < 1$ is identical with $x_{\omega 1}^{\nu}$.

From (52) and (53), Euler's summation formula is obtained by letting $\omega \to 0$; we need not go into details.

13. It is not always practical to use (50) for summation between finite limits, but another formula may be derived from (42) as follows.

Let x be an integer (this restriction being of no real consequence), and let γ be another integer, supposed to be constant. We put[1], for $x < \gamma$,

$$S' f(x) = \sum_{x}^{\gamma-1} f(x) = \sum_{\nu=0}^{\gamma-1-x} f(x+\nu) \qquad (55)$$

while $S' f(x) = 0$ for $x \geq \gamma$. Hence, on repeating the operation S' r times,

[1] Compare Steffensen: Interpolation (Baltimore 1927), art. 111 (where β is written for $\gamma - 1$).

$$S^{(r)} f(x) = \sum_{\nu=x}^{\gamma-1} \binom{\nu - x + r - 1}{r-1} f(\nu) = \sum_{\nu=0}^{\gamma-1-x} \binom{\nu + r - 1}{r-1} f(x+\nu), \quad (56)$$

as may be proved by induction, or be concluded from (49) (putting $f(t) = 0$ for $t \geq \gamma$).

It is now easy to prove that for $s \leq r$

$$S^{(s)} \Delta^r f(x) = (-1)^s \left[\Delta^{r-s} f(x) - \sum_{\nu=0}^{s-1} (-1)^\nu \binom{\gamma - 1 - x + \nu}{\nu} \Delta^{r-s+\nu} f(\gamma) \right]. \quad (57)$$

For this formula is valid for $s = 1$, as, by (55),

$$\begin{aligned} S' \Delta^r f(x) &= S' \Delta^{r-1} \Delta f(x) \\ &= S' \Delta^{r-1} f(x+1) - S' \Delta^{r-1} f(x) \\ &= \Delta^{r-1} f(\gamma) - \Delta^{r-1} f(x); \end{aligned}$$

and being valid for any particular value of s, (57) is proved to be valid also for the following one, on performing the operation S' on both sides and noting that

$$S' \binom{\gamma - 1 - x + \nu}{\nu} = \binom{\gamma - x + \nu}{\nu + 1}.$$

Similarly, we put

$$\underset{\omega}{S'} f(x) = \omega \left[f(x) + f(x+\omega) + \ldots + f(\gamma - \omega) \right] \quad (58)$$

besides $\underset{\omega}{S'} f(x) = 0$ for $x \geq \gamma$; whence, by induction or by (48),

$$\underset{\omega}{S^{(r)}} f(x) = \omega^r \sum_{\nu=0}^{\frac{\gamma - x}{\omega} - 1} \binom{\nu + r - 1}{r - 1} f(x + \nu\omega). \quad (59)$$

If now, in (42), we interpret $\underset{\omega}{\Delta}^{\nu-r} f(x)$ for $\nu < r$ as $(-1)^{\nu-r} \underset{\omega}{S}^{(r-\nu)} f(x)$, this formula may be written

$$\left.\begin{aligned} f(x+h) = \sum_{\nu=0}^{r-1} (-1)^{\nu+r} \frac{h_{\omega r}^{\nu}}{\nu!} \Delta^r \underset{\omega}{S}^{(r-\nu)} f(x) \\ + \sum_{\nu=0}^{m} \frac{h_{\omega r}^{\nu+r}}{(\nu+r)!} \Delta^r \underset{\omega}{\Delta}^{\nu} f(x) + R_m \end{aligned}\right\} \quad (60)$$

where R_m has the meaning (43).

Finally, performing the operation $S^{(r)}$ on both sides of (60), and taking into account that, according to (57),

$$S^{(r)} \Delta^r f(x) = (-1)^r \left[f(x) - \sum_{\nu=0}^{r-1} (-1)^{\nu} \binom{\gamma-1-x+\nu}{\nu} \Delta^{\nu} f(\gamma) \right], \quad (61)$$

we find, as $\underset{\omega}{S}^{(r-\nu)} f(\gamma) = 0$ for $\nu < r$,

$$\left.\begin{aligned} S^{(r)} f(x+h) = \sum_{\nu=0}^{r-1} (-1)^{\nu} \frac{h_{\omega r}^{\nu}}{\nu!} \underset{\omega}{S}^{(r-\nu)} f(x) \\ + (-1)^r \sum_{\nu=0}^{m} \frac{h_{\omega r}^{\nu+r}}{(\nu+r)!} \Bigg[\underset{\omega}{\Delta}^{\nu} f(x) \\ - \sum_{\mu=0}^{r-1} (-1)^{\mu} \binom{\gamma-1-x+\mu}{\mu} \Delta^{\mu} \underset{\omega}{\Delta}^{\nu} f(\gamma) \Bigg] + R \end{aligned}\right\} \quad (62)$$

where

$$\left.\begin{aligned} R = (-1)^r \omega \sum_{\nu=0}^{\infty} \frac{\overline{h-\omega-\nu\omega}_{\omega r}^{m+r}}{(m+r)!} \Bigg[\underset{\omega}{\Delta}^{m+1} f(x+\nu\omega) \\ - \sum_{\mu=0}^{r-1} (-1)^{\mu} \binom{\gamma-1-x+\mu}{\mu} \Delta^{\mu} \underset{\omega}{\Delta}^{m+1} f(\gamma+\nu\omega) \Bigg]. \end{aligned}\right\} \quad (63)$$

This formula is more cumbrous in appearance than (50), but has the advantage over the latter that the remainder may tend to a limit for $\omega \to 0$ which is not the case if (50) is applied to summation between finite limits by assuming that $f(t)$ vanishes beyond a certain range.

14. In the particular case where $r = 1$ we obtain from (62) and (63)

$$\underset{\omega}{S'} f(x) = S' f(x+h) + \sum_{\nu=0}^{m} \frac{h_{\omega 1}^{\nu+1}}{(\nu+1)!}\left[\underset{\omega}{\Delta}^{\nu} f(x) - \underset{\omega}{\Delta}^{\nu} f(\gamma)\right] + R \qquad (64)$$

where

$$R = \omega \sum_{\nu=0}^{\infty} \frac{\overline{h-\omega-\nu\omega}_{\omega 1}^{\,m+1}}{(m+1)!}\left[\underset{\omega}{\Delta}^{m+1} f(x+\nu\omega) - \underset{\omega}{\Delta}^{m+1} f(\gamma+\nu\omega)\right]. \qquad (65)$$

$S' f(x+h)$ has the value

$$S' f(x+h) = f(x+h) + f(x+h+1) + \ldots + f(\gamma - 1 + h)$$

while

$$\underset{\omega}{S'} f(x) = \omega\left[f(x) + f(x+\omega) + \ldots + f(\gamma - \omega)\right].$$

The parameter h must satisfy the condition

$$0 \leq h = p\omega < 1.$$

By letting $\omega \to 0$ we may, from (64) and (65), derive Euler's summation-formula.

(64) and (65) may also be obtained directly from (50) and (51) by writing γ for x and deducting.

Færdig fra Trykkeriet den 12. Marts 1928.

MATHEMATISK-FYSISKE MEDDELELSER

UDGIVNE AF

DET KGL. DANSKE VIDENSKABERNES SELSKAB

6. BIND (Kr. 17,00):

7. BIND (Kr. 17,70):

Kr. Ø.

9. Bjerrum, Niels: Untersuchungen über Ionenassoziation. I. Der Einfluss der Ionenassoziation auf die Aktivität der Ionen bei mittleren Assoziationsgraden. 1926 .... 2.00
10. Nielsen, Niels: Recherches sur les fonctions cylindriques et sur certaines fonctions analogues. 1926 .... 1.35
11. Hevesy, G.: On the Missing Element 87. 1926 .... 0.60
12. Kuhn, W.: Die Stärke der anomalen Dispersion in nicht leuchtendem Dampfe von Thallium und Cadmium. 1926 .... 3.75
13. Sugiura, Y. and Urey, H. C.: On the Quantum Theory explanation of the Anomalies in the 6th and 7th Periods of the Periodic Table. 1926 .... 1.15
14. Burrau, Øyvind: Berechnung des Energiewertes des Wasserstoffmolekel-Ions (H_2^+) im Normalzustand. 1927 .... 0.90
15. Knudsen, Martin: The Hot-wire Manometer. 1927 .... 0.80
16. Juel, C.: Om v. Staudt's Definitioner. 1927 .... 0.85
17. Pólya, G.: Über die algebraisch-funktionentheoretischen Untersuchungen von J. L. W. V. Jensen. 1927 .... 1.40

8. BIND (under Pressen):

1. Hille, Einar: On the Logarithmic Derivatives of the Gamma Function. 1927 .... 2.40
2. Rasch, G.: Beitrag zur Theorie der unvollständigen Gammafunktionen. Nach hinterlassenen Papieren von J. L. W. V. Jensen. 1927 .... 0.85
3. Knudsen, Martin: Thermal Molecular Pressure in Tubes. 1927 .... 1.50
4. Pedersen, P. O.: Den højere Atmosfæres Sammensætning, Tryk, Temperatur og elektriske Ledningsevne i Belysning af Radiobølgernes Udbredelsesforhold. 1927 .... 0.60
5. Besicovitch, A. and Bohr, H.: Some remarks on Generalisations of almost Periodic Functions. 1927 .... 1.50
6. Pedersen, Ellen: Über einige besondere Klassen von fastperiodischen Funktionen. 1928 .... 1.00
7. Steffensen, J. F.: A general Summation Formula. 1928 .... 1.25

Det Kgl. Danske Videnskabernes Selskab.
Mathematisk-fysiske Meddelelser. **VIII,** 8.

OBSERVATIONS SUR DES RECHERCHES ALGÉBRIQUES PLUS ANCIENNES QUE LE THÉORÈME D'ABEL

PAR

NIELS NIELSEN

KØBENHAVN
HOVEDKOMMISSIONÆR: ANDR. FRED. HØST & SØN, KGL. HOF-BOGHANDEL
BIANCO LUNOS BOGTRYKKERI
1928

Pris: Kr. 1,40.

Det Kgl. Danske Videnskabernes Selskabs videnskabelige Meddelelser udkommer fra 1917 indtil videre i følgende Rækker:

Historisk-filologiske Meddelelser,
Filosofiske Meddelelser,
Mathematisk-fysiske Meddelelser,
Biologiske Meddelelser.

Hele Bind af disse Rækker sælges 25 pCt. billigere end Summen af Bogladepriserne for de enkelte Hefter.

Selskabets Hovedkommissionær er *Andr. Fred. Høst & Søn*, Kgl. Hof-Boghandel, København.

Det Kgl. Danske Videnskabernes Selskab.
Mathematisk-fysiske Meddelelser. **VIII,** 8.

OBSERVATIONS SUR DES RECHERCHES ALGÉBRIQUES PLUS ANCIENNES QUE LE THÉORÈME D'ABEL

PAR

NIELS NIELSEN

KØBENHAVN
HOVEDKOMMISSIONÆR: ANDR. FRED. HØST & SØN, KGL. HOF-BOGHANDEL
BIANCO LUNOS BOGTRYKKERI
1928

Le dix-huitième siècle est remarquable dans l'histoire des mathématiques à cause de ses progrès étonnants.

Il suffit d'en citer deux exemples seulement: le développement des méthodes du calcul infinitésimal et ses applications à la géométrie analytique et à la mécanique; l'établissement de la géométrie descriptive.

Le grand siècle n'a pas donné des progrès semblables, en ce qui concerne l'algèbre, parce qu'il fut reservé au génie d'Abel d'éclaircir le grand énigme de tant de siècles: l'équation du cinquième degré.

Néanmoins on doit au dix-huitième siècle de très belles découvertes algébriques, par exemple les théorèmes de d'Alembert, de Bezout et de Fourier; les recherches de Lagrange, de Vandermonde, travaux qui sont très peu connus aujourd'hui, tant ils sont ombragés par les admirables recherches d'Abel et leurs conséquences.

Le but principal du présent mémoire, en grande partie extrait de mon cours universitaire, intitulé *Géomètres français sous la Révolution* et professé dans le second sémestre de 1927, c'est de tirer de l'oubli les recherches susdites, de faire justice aux auteurs, en leur procurant la reconnaissance qu'ils méritent.

[1] Un extrait de ce cours, 114 pages in-8°, vient de paraître en danois, dans la dissertation-programme de l'Université de Copenhague, novembre 1927. Le cours complet paraîtra en français le plus tôt possible.

I. Le cas irréductible du troisième degré.

On sait que des géomètres italiens, dans le deuxième quart du seizième siècle, ont brisé les bornes antiques de l'algèbre, en résolvant l'équation du troisième degré.

La méthode appliquée par ces géomètres n'a pas été transmise à la postérité, et la plus ancienne résolution connue est, je crois, celle de Hudde, publiée par exemple en 1694, dans un appendice à une edition de la Géométrie de Descartes. La méthode de Hudde étant indispensable, dans ce qui suit, il est nécessaire de la mentionner ici en peu de mots.

Soit posé, dans l'équation

$$x^3 - px + q = 0, \tag{1}$$

$x = u + iv$, cette équation est évidemment satisfaite, pourvu que

$$u^3 + v^3 = -q, \quad uv = \frac{p}{3}, \tag{2}$$

conditions qui déterminent immédiatement u et v, et l'on aura

$$x = \sqrt[3]{-\frac{q}{2} + \sqrt{\left(\frac{q}{2}\right)^2 - \left(\frac{p}{3}\right)^3}} + \sqrt[3]{-\frac{q}{2} - \sqrt{\left(\frac{q}{2}\right)^2 - \left(\frac{p}{3}\right)^3}}, \tag{3}$$

où les deux racines cubiques, u et v, sont à choisir, conformément à la dernière des équations (2), de sorte que $3uv = p$.

Désignons donc par u et v un tel couple de valeurs, par α une quelconque des deux racines imaginaires de l'équation binome $x^3 = 1$, les trois racines de l'équation proposée deviennent

$$u + v, \quad \alpha u + \alpha^2 v, \quad \alpha^2 u + \alpha v, \tag{4}$$

formule qui est souvent, mais injustement, attribuée à Cardan.

Supposons maintenant p et q réels, de sorte que

$$(5) \qquad \left(\frac{p}{3}\right)^3 > \left(\frac{q}{2}\right)^2,$$

les racines cubiques qui figurent au second membre de (3) sont toutes deux imaginaires, mais leur produit étant positif, elles sont des nombres conjugués, de sorte que les trois racines de l'équation proposée sont toutes réelles.

C'est le célèbre cas irréductible du troisième degré.

La formule (3) qui représente, sous forme imaginaire, un nombre réel, a beaucoup surpris les géomètres qui se sont certainement efforcés d'en chasser les imaginaires, mais vainement.

François Nicole, connu par ses belles méthodes pour sommer certaines séries, a étudié le premier, je crois, plus amplement le cas irréductible[1], et il écrit, dans un autre mémoire, qu'il avait toujours été un »sujet de scandale« qu'un nombre réel se présente sous une forme apparemment imaginaire.[2]

Or, remarquant que les racines (3) sont de la forme

$$(6) \qquad \sqrt[3]{a+ib} + \sqrt[3]{a-ib},$$

a et b étant réels, Nicole développe, d'après la formule binomiale, les deux expressions

$$\left(\frac{a}{b}+i\right)^n, \quad \left(\frac{a}{b}-i\right)^n,$$

où l'exposant n est réel, ce qui met en évidence que la somme (6) est réelle, parce que les termes imaginaires s'évanouissent.

Mais les expressions ainsi obtenues pour les racines se

[1] Mémoires de l'Académie des sciences 1738, 97—102.

[2] Ibid. 1741, 25.

présentent sous forme d'une série infinie, et Nicole avoue qu'il ne possède aucun moyen pour sommer cette série.

Le but de Nicole a évidemment été de représenter les racines réelles par des expressions algébriques proprement dites, savoir à l'aide des radicaux réels, ce qui est impossible, nous le savons grâce au beau théorème de M. Hölder.[1]

Dionis du Séjour[2], connu par sa résolution du triangle sphéroïdique, introduit dans (1)

$$x = a + ib, \quad b \neq 0,$$

ce qui donnera

$$3a^2 = p + b^2, \quad a^3 - 2ab^2 - pa + q = 0,$$

d'où, en éliminant a,

$$4(b^2 + p)(4b^2 + p)^2 = 27q^2,$$

ce qui est impossible, parce que p et b^2 sont des nombres positifs.

Or, d'Alembert[3] ayant démontré que toutes les racines non-réelles d'une équation algébrique sont toujours de la forme $a + ib$, b n'étant pas zéro, Dionis du Séjour dit, avec raison, qu'il a démontré par l'analyse, que, dans le cas irréductible, les racines de (1) sont toutes réelles.

Janot de Stainville[4] part de l'équation trinome

$$x^{2n+1} - px + q = 0 \tag{7}$$

et démontre que cette équation, p et q étant réels, a toujours trois, et seulement trois racines réelles, pourvu que

$$\left(\frac{p}{2n+1}\right)^{2n+1} > \left(\frac{q}{2n}\right)^{2n}; \tag{8}$$

[1] Mathematische Annalen XXXVIII, 307—312; 1891.
[2] Mémoires de l'Académie des sciences 1768, 207—208.
[3] Mémoires de l'Académie de Berlin 1746.
[4] Correspondance sur l'École polytechnique III, 58—60; janvier 1814.

soit $n = 1$, on aura précisément le cas irréductible du troisième degré.

Ayant, dans ce qui suit, à regarder d'autres méthodes pour la résolution de l'équation du troisième degré, nous remarquons ici que Stainville, dans une autre note[1], a résolu l'équation (1), en posant

$$x = y + z + t,$$

résolution qui est analogue à la seconde des méthodes appliquées par Euler pour la résolution de l'équation du quatrième degré, nous le verrons dans ce qui suit.

II. L'équation du quatrième degré.

Louis Ferrari, peu d'années après la résolution de l'équation du troisième degré, a résolu aussi celle du quatrième degré

$$(1) \qquad x^4 + ax^3 + bx^2 + cx + d = 0,$$

en écrivant cette équation sous la forme

$$\left(x^2 + \frac{a}{2}x + \frac{y}{2}\right)^2 = \left(y + \frac{a^2}{4} - b\right)x^2 + \left(\frac{ay}{2} - c\right)x + \frac{y^2}{4} - d,$$

puis déterminant le paramètre y, de sorte que le second membre devient un carré exact, ce qui donnera la réduite

$$(2) \qquad y^3 - by^2 + (ac - 4d)\,y + d\,(4b - a^2) - c^2 = 0.$$

Descartes a résolu l'équation du quatrième degré, ne contenant pas le second terme, savoir

$$(3) \qquad x^4 + px^2 + qy + r = 0,$$

en écrivant son premier membre sous la forme

$$(x^2 - \alpha x + \beta)(x^2 + \alpha x + \gamma),$$

ce qui conduira, pour $\alpha^2 = y$, à la réduite

[1] Annales de mathématiques pures et appliquées IX, 197—203; 1818.

(4) $$y^3+2py^2+(p^2-4r)y-q^2 = 0.$$

Appliquons cette méthode à l'équation complète (1), nous aurons une réduite du sixième degré contenant tous les termes, équation qui est par conséquent résoluble à l'aide des radicaux.

Ayant à regarder, dans ce qui suit, d'autres résolutions de l'équation du quatrième degré, nous remarquons ici que Pierre Pilatte[1], ancien élève de l'École polytechnique et ancien capitaine d'artillerie, puis professeur de mathématiques spéciales au lycée d'Angers, a résolu l'équation (3) en y posant $x = y+z$, ce qui conduira immédiatement à la réduite de Descartes. On voit que cette résolution est analogue à celle appliquée par Hudde pour la résolution de l'équation du troisième degré.

III. Méthodes „propres à résoudre toutes les équations".

Walther de Tschirnhausen[2] a donné le premier, je crois, une méthode qui permet de résoudre les équations et du troisième et du quatrième degré, méthode qu'il croyait propre à donner la résolution, à l'aide des radicaux, d'une équation algébrique quelconque.

Le fondement de la méthode de Tschirnhausen est l'élimination de x entre l'équation proposée

(1) $$x^n+a_1x^{n-1}+a_2x^{n-2}+\cdots+a_{n-1}x+a_n = 0$$

et celle-ci

(2) $$y = \alpha_0+\alpha_1x+\alpha_2x^2+\cdots+\alpha_{n-1}x^{n-1},$$

où les α_μ sont des paramètres quelconques.

[1] Annales de mathématiques pures et appliquées II, 152—154; 1811.
[2] Acta eruditorum 1683, 204—207.

A cet effet, désignons par $x_1 x_2 \ldots x_n$ les racines de (1), puis posons

$$y_r = \alpha_0 + \alpha_1 x_r + \alpha_2 x_r^2 + \cdots + \alpha_{n-1} x_r^{n-1},$$

la réduite susdite se présente sous la forme

$$(y - y_1)(y - y_2) \ldots (y - y_n) = 0$$

ou bien, ordonnée d'après des puissances déscendantes de y,

(3) $$y^n + A_1 y^{n-1} + A_2 y^{n-2} + \cdots + A_{n-1} y + A_n = 0,$$

et il est évident que les coefficients A_p sont des fonctions entières et rationnelles des coefficients donnés a_μ. De plus, les A_p sont des fonctions entières, rationnelles et homogènes des inconnues α_μ, de sorte que A_p est précisement du degré p.

Tschirnhausen détermine ensuite les inconnues α_μ de sorte que

(4) $$A_1 = 0, \quad A_2 = 0, \quad \ldots, \quad A_{n-1} = 0,$$

ce qui réduira l'équation (3) à une équation binome.

Dans le cas le plus simple $n = 3$, la réduite ainsi obtenue est donc du deuxième degré, ce qui s'accorde bien avec la méthode de Hudde.

Soit ensuite $n = 4$, la réduite est du sixième degré, comme dans la résolution de Descartes; mais nous savons que cette équation est résoluble par des radicaux.

Pour $n = 5$, la réduite est du degré 24; mais la réduite du sixième degré qui correspond à $n = 4$ étant résoluble à l'aide des radicaux, il n'était pas, en 1683 et beaucoup plus tard encore, une conclusion trop hardie que la réduite du degré 24 ait la même propriété.

Nous savons que la méthode de Tschirnhausen ne peut pas donner la résolution algébrique des équations du'n degré quelconque; cette méthode est néanmoins d'une haute valeur. Elle montrera par exemple facilement que l'équation générale du cinquième degré peut toujours être réduite à la forme

$$x^5 + x + a = 0,$$

dont les racines sont par conséquent des fonctions algébriques d'une seule variable, propriété qui a plus tard conduis à des résultats remarquables.

Cette réduction de l'équation générale du cinquième degré est due à Bring, plus tard professeur d'histoire à l'Université de Lund, qui l'a donnée dans une dissertation[1], évidemment sans connaître la publication de Tschirnhausen.

Euler[2], partant des expressions des racines de l'équation du troisième degré, essaya de résoudre l'équation algébrique générale, sans le second terme,

$$(5) \qquad x^n + a_2 x^{n-2} + a_3 x^{n-3} + \cdots + a_{n-1} x + a_n = 0,$$

en supposant

$$(6) \qquad x = \sqrt[n]{y_1} + \sqrt[n]{y_2} + \cdots + \sqrt[n]{y_{n-1}},$$

ce qui réduit la résolution de (5) à la détermination des $n-1$ nombres y_μ, savoir à la résolution d'une équation du degré $n-1$, ce qui est évident; mais comment déterminer les coefficients de cette nouvelle équation du degré $n-1$?

Soit $n = 3$, la méthode d'Euler n'est autre chose que celle de Hudde.

Pour $n = 4$, Euler suppose, au lieu de (6),

$$(7) \qquad x = \sqrt{y_1} + \sqrt{y_2} + \sqrt{y_3},$$

ce qui conduira à une nouvelle résolution de l'équation du

[1] publiée à Lund en 1786.

[2] Commentarii Academiæ Petropolitanæ VI, 216—231; 1738.

quatrième degré; la réduite qui détermine les trois inconnues $y_1\ y_2\ y_3$ est celle de Descartes, retrouvée par Pilatte.

Or, la méthode d'Euler étant applicable pour $n = 3$ et $n = 4$, la conclusion qu'elle soit généralement applicable n'était point hardie dans ce temps-là, où l'idée de l'élimination était très peu claire, nous le verrons dans ce qui suit.

Ce mémoire d'Euler semble très peu connu.

Lagrange, dans ses Réflexions sur la résolution des équations, ne le mentionne pas; dans ses leçons données aux Écoles normales en 1795, il résout l'équation du quatrième degré en posant

$$x = y + z + t,$$

ce qui n'est autre chose que la méthode d'Euler, mais cette méthode est tirée de l'Algèbre d'Euler, que Lagrange a traduite en français vers cette époque, en l'enrichissant par des suppléments très importants.

Jean-Jacques Bret, ancien élève de l'École polytechnique, puis professeur de mathématiques à la Faculté des sciences de Grenoble, observe[1] que les valeurs des trois racines carrées, qui figurent au second membre de (7), indiquées par Euler, ne sont pas toujours exactes, ce que Lagrange avoue, dans l'édition de ses leçons susdites.[2]

Wronski, ci-devant officier supérieur de l'artillerie russe, dans une brochure publiée à Paris en 1812, donne comme nouvelle la méthode d'Euler propre à résoudre les équations algébriques de tous les degrés, et il se vante beaucoup d'avoir donné ainsi une résolution nouvelle de l'équation du quatrième degré. Chose curieuse, Gergonne[3], l'éminent

[1] Correspondance sur l'École polytechnique II, 217—219; 1811.

[2] Journal de l'École polytechnique cah. VII—VIII, 173—278; juin 1812 (voyez p. 239).

[3] Annales de mathématiques pures et appliquées III, 51—59, 137—139; IX, 213—214.

rédacteur des Annales, remarquant avec raison que cette solution de l'équation du quatrième degré n'est point nouvelle, l'attribue à Bezout.

Euler[1] a donné encore une méthode générale »propre à résoudre les équations de tous les degrés«, méthode qui est analogue à celle indiquée presque en même temps par Bezout.[2]

En effet, le fondement de ces deux méthodes est l'élimination de x entre les deux équations

$$y^n + D = 0, \quad \alpha_0 y^{n-1} + \alpha_1 y^{n-2} + \cdots + \alpha_{n-2} y + x = 0,$$

puis la détermination des inconnues D et α_μ, de sorte que la réduite de ces deux équations soit identique à l'équation proposée (5), sans le second terme.

Euler suppose $\alpha_{n-2} = 1$, tandis que Bezout fait $D = -1$; néanmoins ces deux méthodes très analogues conduisent à des résultats assez différents, nous le verrons dans les recherches de Lagrange sur l'équation du quatrième degré. Quant à la méthode de Bezout, elle conduira à des résultats curieux, nous le verrons dans l'article suivant.

IV. Problèmes d'élimination de Bezout.

Nous possédons, de la main de Bezout, trois mémoires algébriques, dont le premier[3] donne une résolution de l'équation du troisième degré

$$x^3 + mx^2 + nx + p = 0, \tag{1}$$

en éliminant y entre les deux équations

$$y^3 + h = 0, \quad y = \frac{x+a}{x+b}, \tag{2}$$

[1] Novi Commentarii Academiæ Petropolitanæ IX, 70—98; 1764.

[2] Mémoire de l'Académie des sciences 1765, 533—552.

[3] Ibid. 1762, 17—52.

puis déterminant les constantes inconnues a, b, h, de sorte que la réduite et l'équation proposée (1) deviennent identiques.

Puis Bezout détermine les équations d'un degré quelconque, qui deviennent binomes en y, determinée par la dernière des équations (2), ce qui conduira à beaucoup d'équations des degrés supérieurs résolubles à l'aide des radicaux.

Dans le même mémoire, Bezout étudie les équations du degré n, qui admettent une racine de la forme

$$x = \sqrt[n]{a^{n-1} b} + \sqrt[n]{a b^{n-1}}, \tag{3}$$

et il détermine les équations de ce genre qui correspondent à $n = 3, 4, 5, 6, 7$, ce qui donnera aussi beaucoup d'équations des degrés supérieurs résolubles à l'aide des radicaux.

Chose curieuse, Bezout énonce le problème qui correspond à la première méthode d'Euler, mais il n'étudie que l'expression spéciale (3).

Dans le second mémoire[1], Bezout mentionne que l'on croyait que la réduite de trois équations, chacune du troisième degré, était du degré 81, puis il montre que ce degré ne peut pas être supérieur à 49. Il a évidemment coûté à Bezout un longue et pénible travail de pénétrer jusqu'à son théorème général sur le degré général de la réduite d'un nombre quelconque d'équations étant de degrés quelconques.[2]

Bezout étudie l'élimination de n inconnues entre n équations linéaires et homogènes; il introduit par conséquent les déterminants, assez imparfaitement définis, ce me semble, et il donne la réduite qui correspond à $n = 5$; c'est-à-dire qu'il a calculé le déterminant général du cinquième ordre.

[1] Mémoires de l'Académie des sciences 1764, 288—338.

[2] Théorie générale des équations algébriques. Paris 1779.

Quant au troisième mémoire[1] de Bezout, nous avons à mentionner son élimination de y entre les deux équations

$$(4) \qquad y^n = 1$$

$$(5) \qquad g(x, y) = \alpha_0 y^{n-1} + \alpha_1 y^{n-2} + \cdots + \alpha_{n-2} y + x = 0.$$

En effet, multiplions par y cette dernière équation, cette multiplication effectue une permutation cyclique des termes, savoir

$$\alpha \varphi(x, y) = \alpha_1 y^{n-1} + \alpha_2 y^{n-2} + \cdots + \alpha_{n-2} y^2 + xy + \alpha_0 .$$

Répétant la même opération, on aura finalement n équations de cette forme, et Bezout trouve la réduite

$$(6) \qquad F_n(x) = 0,$$

en déterminant les puissances de y à l'aide de $n-1$ de ces équations, puis introduisant les résultats, dans la n-ième des équations susdites; c'est-à-dire que nous aurons

$$(7) \qquad F_n(x) = \begin{vmatrix} \alpha_0 & \alpha_1 \ldots & \alpha_{n-2} & x \\ \alpha_1 & \alpha_2 \ldots & x & \alpha_0 \\ \ldots & \ldots & \ldots & \ldots \\ \alpha_{n-2} & x \ \ldots & \alpha_{n-4} & \alpha_{n-3} \\ x & \alpha_0 \ldots & \alpha_{n-3} & \alpha_{n-2} \end{vmatrix} .$$

Bezout détermine les expressions explicites de ces déterminants cycliques qui correspondent à $n = 3, 4, 5, 6$; c'est-à-dire que l'éminent algébriste a évidemment étudié, le premier, de tels déterminants.

En se rappelant la formation de la réduite (6), il est évident que $F_n(x)$ est divisible par $\varphi(x, y)$, y étant une racine quelconque de l'équation binome (4); soit donc $y_1 y_2 \ldots y_n$ toutes les racines de cette équation, on aura

[1] Mémoire de l'Académie des sciences 1765, 533—552.

$$(8) \qquad F_n(x) = (-1)^n \varphi(x, y_1)\, \varphi(x, y_2) \ldots \varphi(x, y_n),$$

ce qui est précisément le produit qui représente le déterminant cyclique[1], trouvé presque un siècle après la publication du mémoire de Bezout, qui, poursuivant son but beaucoup plus élévé, ne s'est pas arrêté à de telles choses.

Ce mémoire de Bezout semble très peu connu aujourd'hui; Baltzer, qui cite plusieurs fois les deux mémoires de 1762 et de 1764, ne mentionne point celui de 1765, et M. Kovalewski[2] ne cite qu'en passant les recherches de Bezout.

V. Recherches de Lagrange.

Cauchy écrit à la tête de son célèbre mémoire qui donne les fondements de la théorie des substitutions[3]:

»MM. Lagrange et Vandermonde sont, je crois, les premiers qui aient considéré les fonctions de plusieurs variables relativement au nombre de valeurs qu'elles peuvent obtenir, lorsqu'on substitue ces variables à la place les unes des autres. Ils ont donné plusieurs théorèmes intéressants relatifs à ce sujet, dans deux mémoires imprimés en 1771[4], l'un à Berlin l'autre à Paris.«

Remarquons que les variables susdites, permutées par Lagrange et Vandermonde, sont l'ensemble des racines d'une équation algébrique, il résulte de la remarque de Cauchy, qu'une nouvelle ère dans l'histoire de l'algèbre commence avec ces mémoires de Lagrange et de Vandermonde.

Or, on a de Lagrange un mémoire antérieur[5] à celui

[1] Voyez par exemple Baltzer: Theorie und Anwendung der Determinanten, 109; 5e éd. Leipsic 1881.

[2] Einführung in die Determinantentheorie. Leipsic 1909.

[3] Journal de l'École polytechnique, cah. XVII, 1—28; janvier 1815.

[4] C'est-à-dire publiés dans les Mémoires des deux Académies pour l'année 1771.

[5] Lu au courant de l'année 1771 et publié dans le volume des Nouveaux Mémoires de l'Académie de Berlin pour l'année 1770, paru en 1772.

cité par Cauchy et étudiant le même sujet, mémoire qui présente un intérêt spécial, à notre point de vue, parce qu'il donne une analyse profonde des résolutions connues des équations du troisième et du quatrième degré.

L'analyse de la méthode de Hudde conduira à déterminer une fonction linéaire des trois variables $x_1\, x_2\, x_3$ dont le cube n'a que deux valeurs. Soit α une quelconque des deux racines imaginaires de l'équation binome $x^3 = 1$, la fonction cherchée deviendra

$$y_1 = x_1 + \alpha x_2 + \alpha^2 x_3; \tag{1}$$

car les valeurs de cette fonction, obtenues en la soumettant à une permutation cyclique, sont

$$y_1,\ \alpha y_1,\ \alpha^2 y_1,$$

et c'est la même chose pour la fonction

$$y_2 = x_1 + \alpha^2 x_2 + \alpha x_3. \tag{2}$$

Appliquons ensuite l'identité

$$(y - y_1)(y - \alpha y_1)(y - \alpha^2 y_1) = y^3 - y_1^3$$

et l'identité analogue pour y_2, nous aurons immédiatement la réduite de Hudde.

Lagrange remarque en passant que la fonction rationnelle

$$\frac{x_1^r + \alpha x_2^r + \alpha^2 x_3^r}{x_1^s + \alpha x_2^s + \alpha^2 x_3^s}, \tag{3}$$

r et s étant des positifs entiers quelconques, n'a que deux valeurs inégales.

La méthode de Bezout de 1762 est identique à celle de Tschirnhausen.

Suit une étude intéressante de l'équation binome

$$x^{2n+1} - 1 = 0. \tag{4}$$

Supprimons le facteur $x-1$, puis posons

$$y = x + \frac{1}{x},$$

l'équation (4) se réduit à une équation du degré n par rapport à y; Lagrange étend cette méthode à une équation réciproque quelconque d'un degré pair, et il remarque que cette réduction est due à Moivre.[1]

Soit, dans (4), $n = 3$, l'équation en y devient

$$y^3 + y^2 - 2y - 1 = 0,$$

de sorte que l'équation binome $x^n - 1 = 0$ est résoluble à l'aide des radicaux, pourvu que $n = 2^\alpha \cdot 3^\beta \cdot 5^\gamma \cdot 7^\delta$, mais l'équation $x^{11} - 1$ conduira à l'équation du cinquième degré

(5) $$y^5 + y^4 - 4y^3 - 3y^2 + 3y + 1 = 0,$$ [2]

qui arrête ici Lagrange dans ses réflexions sur la résolution, à l'aide des radicaux, de l'équation binome.

Chose curieuse, Vandermonde avait déjà résolu, à l'aide des radicaux, l'équation (5), avant la publication du mémoire de Lagrange, nous le verrons dans ce qui suit.

La réduite de l'équation du quatrième degré étant du troisième degré, il s'agit de trouver une fonction rationnelle des quatre variables $x_1\, x_2\, x_3\, x_4$, qui n'ait que trois valeurs.

Lagrange étudie d'abord la fonction

(6) $$u = x_1 x_2 + x_3 x_3$$

et forme, par un calcul direct, l'équation du troisième degré qui a comme racines les trois valeurs possibles de u, ce qui conduira précisément à la réduite de Ferrari.

La fonction rationnelle

[1] Miscellanea analytica de seriebus et quadraturis, Londres 1730.

[2] L'équation indiquée par Lagrange est défigurée par une faute d'impression, car l'avant-dernier terme est indiqué comme $-3y$.

$$(7) \qquad t = (x_1 + x_2 - x_3 - x_4)^2$$

a aussi trois valeurs, et, soit l'équation proposée

$$(8) \qquad x^4 + ax^3 + bx^2 + cx + d = 0,$$

un calcul direct donnera

$$t = a^2 - 4b + 4u.$$

Posons donc, dans la réduite de Ferrari,

$$u = \frac{t - a^2}{4} + b,$$

nous aurons la réduite cherchée qui, pour $a = 0$, est précisément celle de Descartes.

Soit ensuite $t_1\ t_2\ t_3$ les racines de la réduite ainsi obtenue, Lagrange remarque qu'il faut choisir convenablement les racines carrées de ces trois quantités, restriction nécessaire qui est bien curieuse, parce que Lagrange admit plus tard, dans ses leçons aux Écoles normales, les expressions d'Euler, qui ne sont pas toujours exactes, nous l'avons déjà remarqué dans l'article III.

Suit une étude des méthodes d'Euler et de Bezout, qui sont en réalité les mêmes que celle de Tschirnhausen, parce que l'élimination de y entre les deux équations

$$x = \alpha + \beta y + \gamma y^2 + \delta y^3, \quad y^4 + D = 0$$

donnera l'équation proposée en x.

Euler suppose $\gamma = 1$ et trouve une réduite du troisième degré par rapport à D, réduite qui est celle de Ferrari.

Bezout, au contraire, suppose $D = -1$ et trouve une équation du troisième degré par rapport à γ^2, ce qui est la réduite de Descartes, tandis que la détermination de α ou de β conduira à une réduite du sixième degré par rapport à α^4 ou à β^4.

Ce mémoire contient aussi un théorème essentiel sur la méthode de Tschirnhausen.

En effet, soit

$$(9) \qquad y^n + A_1 y^{n-1} + A_2 y^{n-2} + \cdots + A_{n-1} y + A_n = 0$$

la réduite de Tchirnhausen, obtenue par l'élimination de y entre les deux équations

$$(10) \qquad x^n + a_1 x^{n-1} + a_2 x^{n-2} + \cdots + a_{n-1} x + a_n = 0$$

$$(11) \qquad y = \alpha_0 + \alpha_1 y + \alpha_2 y^2 + \cdots + \alpha_{n-1} y^{n-1},$$

Lagrange a démontré le premier[1], je crois, que la réduite des équations

$$A_1 = 0, \quad A_2 = 0, \ldots, \quad A_{n-1} = 0$$

est du degré $(n-1)!$.

Du reste, Lagrange remarque que, dans le cas spécial $n = 4$, on aura une réduite du troisième degré, en supposant $A_1 = 0$ et $A_3 = 0$.

Le mémoire de Lagrange, cité par Cauchy et publié en 1773[2], donne une véritable théorie de la méthode de Tschirnhausen.

Lagrange remarque tout d'abord que les inventeurs des méthodes »propres à résoudre algébriquement toutes les équations«, se sont généralement bornés à l'étude des équations du troisième et du quatrième degré.

Cependant il remarque que Bezout a aussi regardé l'équation du cinquième degré et trouvé une réduite du degré 120, dans laquelle tous les exposants de l'inconnue sont multiples de 5, de sorte que cette équation est en vérité

[1] Il faut se rappeler que le théorème général de Bezout n'était pas connu en 1771.

[2] Nouveaux Mémoires de l'Académie de Berlin pour l'année 1771, 138—253.

du degré 24 par rapport à la cinquième puissance de l'inconnue susdite.

Nous pouvons ajouter que l'existence de cette cinquième racine a évidemment affirmé Bezout dans son opinion que cette réduite du degré 24 soit algébriquement résoluble, comme la réduite du sixième degré obtenue pour l'équation du quatrième degré.

Quant au mémoire de Lagrange, nous savons que la méthode de Tschirnhausen conduira à une réduite du degré $(n-1)!$.

Mais soit n un nombre composé, savoir $n = p \cdot q$, nous pouvons faire disparaître, dans (9), toutes les puissances de y, dont les exposants ne sont pas divisibles par q, ce qui conduira à une réduite du degré

$$\frac{(n-1)!}{(p-1)!\, q^{p-1}}.$$

Soit par exemple $n = 6$, $p = 3$, $q = 2$, la réduite deviendra du quinzième degré.

Partant des équations (9), (10), (11), Lagrange démontre l'existence d'une inversion de la formule (11), savoir

$$x = \beta_0 + \beta_1 y + \beta_2 y^2 + \cdots + \beta_{n-1} y^{n-1}.$$

On trouve aussi une étude de l'équation du degré $\lambda = \varphi(n)$, qui détermine les racines primitives de l'équation binome $x^n = 1$, et il est démontré que, α étant une racine primitive de cette équation binome, la fonction rationnelle des n variables $x_1\, x_2 \ldots x_n$

$$(x_1 + \alpha x_2 + \alpha^2 x_3 + \cdots + \alpha^{n-1} x_n)^n$$

n'a que $(n-1)!$ valeurs différentes.

La seconde partie de ce mémoire de Lagrange est consacrée à la recherche des fonctions rationnelles des n variables $x_1 \, x_2 \cdots x_n$, étant déterminables à l'aide d'une équation d'un degré inférieur à $n!$.

VI. Recherches de Vandermonde.

Le mémoire de Vandermonde, cité par Cauchy, a été présenté à l'Académie des sciences au mois de novembre 1770 et paraphé par le secrétaire perpétuel le 28 du même mois, mais l'auteur n'étant pas alors membre de l'Académie, l'impression de son ouvrage fut retardée jusqu'à 1774.[1]

Vandermonde écrit que son but principal est l'étude des trois problèmes suivants:

1° Trouver une fonction des racines [de l'équation proposée], de laquelle on puisse dire, dans un certain sens, qu'elle égale telle de ces racines que l'on voudra;

2° Mettre cette fonction sous une forme telle qu'il soit de plus indifférent d'y changer les variables entre elles;

3° Y substituer les valeurs en somme de ces racines, somme de leurs produits deux à deux, etc.

Quant au premier de ces problèmes, soient $x_1 \, x_2 \ldots x_m$ les variables en question, et soient $\varepsilon_1 \, \varepsilon_2 \ldots \varepsilon_m$ les racines de l'équation binome $x^m = 1$, Vandermonde remarque que l'expression des racines de l'équation du troisième degré l'a conduit à la fonction

$$(1) \qquad \frac{1}{m} \sum_{s=0}^{s=m-1} \sqrt[m]{(\varepsilon_1^s x_1 + \varepsilon_2^s x_2 + \cdots + \varepsilon_m^s x_m)^m}.$$

En effet, si l'on attribue au terme sommatoire la valeur

[1] Mémoires de l'Académie des sciences pour l'année 1771, 365—416. Dans le temps intermédiaire, les deux mémoires de Lagrange ont parus.

$$\varepsilon_{\mu}^{-s}(\varepsilon_1^s x_1 + \varepsilon_2^s x_2 + \cdots + \varepsilon_m^s x_m),$$

l'expression (1) aura la valeur x_{μ}, où $\mu = 1, 2, \ldots, m$.

Soit m un nombre composé, l'expression (1) est susceptible de simplifications différentes, assez profondément étudiées par Vandermonde, notamment en ce qui concerne de petites valeurs de m.

On voit que les racines des équations binomes jouent un rôle fondamental aussi dans les recherches de Vandermonde; soit m un nombre impair, savoir $m = 2n + 1$, il suppose

$$r^m - 1 = (r-1)(r^2 + x_1 r + 1)(r^2 + x_2 r + 1) \cdots (r^2 + x_n r + 1),$$

et il indique l'équation du degré n, dont les racines sont les x_{μ}, équation que Lagrange n'avait pas donnée explicitement, savoir

$$(2) \quad \left\{ \begin{aligned} & x^n - x^{n-1} + (n-2)x^{n-3} - \binom{n-3}{2} x^{n-5} + \cdots \\ & -(n-1)x^{n-2} + \binom{n-2}{2} x^{n-4} - \binom{n-3}{3} x^{n-6} + \cdots = 0, \end{aligned} \right.$$

ce qui donnera, pour $m = 11$

$$(3) \qquad x^5 - x^4 - 4x^3 + 3x^2 + 3x - 1 = 0,$$ [1]

savoir la fameuse équation qui a arrêté Lagrange dans sa résolution, par des radicaux, des équations binomes.

Les développements de Vandermonde lui permettent de résoudre les équations du troisième et du quatrième degré contenant tous les termes.

Vandermonde mentionne la réduite du vingt-quatrième degré de l'équation du cinquième degré; il dit qu'il n'avait

[1] Posons, dans cette équation, $x = -y$, nous aurons précisément celle de Lagrange.

pu trouver aucune fonction linéaire de cinq variables, qui se détermine à l'aide d'une équation du troisième ou du quatrième degré, et qu'il ne croit pas à l'existence d'une telle fonction.

Comme application de ses formules, Vandermonde résout, à l'aide des radicaux, l'équation (3), savoir une équation irréductible du cinquième degré. A cet effet, Vandermonde applique, outre son expression générale (1), des relations entre les racines de (3)

$$x_r = -2\cos\frac{2r\pi}{11}, \quad r = 1, 2, 3, 4, 5,$$

par exemple

$$x_1^2 = -x_2 + 2, \quad x_1 x_2 = -x_1 - x_3.$$

Vandermonde indique la valeur

$$x = \frac{1}{5}(1 + A_1 + A_2 + A_3 + A_4),$$

où est posé pour abréger

$$A_1 = \sqrt[5]{\frac{11}{4}\left(89 + 25\sqrt{5} - 5\sqrt{-5+2\sqrt{5}} + 45\sqrt{-5-2\sqrt{5}}\right)},$$

$$A_2 = \sqrt[5]{\frac{11}{4}\left(89 + 25\sqrt{5} + 5\sqrt{-5+2\sqrt{5}} - 45\sqrt{-5-2\sqrt{5}}\right)},$$

$$A_3 = \sqrt[5]{\frac{11}{4}\left(89 - 25\sqrt{5} - 5\sqrt{-5+2\sqrt{5}} - 45\sqrt{-5-2\sqrt{5}}\right)},$$

$$A_4 = \sqrt[5]{\frac{11}{4}\left(89 - 25\sqrt{5} + 5\sqrt{-5+2\sqrt{5}} + 45\sqrt{-5-2\sqrt{5}}\right)};$$

mais il ne dit pas comment on doit déterminer ces quatre radicaux pour obtenir chacune des cinq racines.

On voit que les racines, qui sont toutes réelles, se présentent sous forme apparemment imaginaire, comme dans le cas irréductible du troisième degré, ce qui est une conséquence du théorème de M. Hölder cité dans l'article I.

De plus, Vandermonde écrit que la résolution, à l'aide des radicaux, de l'équation générale (2) est très facile.

Kronecker apprécia à sa juste valeur ce remarquable mémoire; il disait dans une de ses leçons: »L'essor moderne de l'algèbre commence avec le mémoire présenté par Vandermonde à l'Académie de Paris dans l'année 1770[1] et intitulé: Sur la résolution des équations; la profondeur des conceptions, si clairement exprimées dans cet ouvrage, nous semble vraiment surprenante.«

M. Carl Itzigsohn a donné (Berlin 1888) une édition allemande des trois mémoires de Vandermonde, en ce qui concerne les mathématiques pures[2], et la célébrité de cet éminent algébriste est maintenant reconnue en Allemagne, mais il faut que son nom figure, à côté de celui de Lagrange, comme des dignes prédécesseurs d'Abel.

VII. Nombre des racines réelles.

L'abbé de Gua[3] a démontré le premier, je crois, la célèbre règle de Descartes, vivement contestée par Fermat qui était, d'après Condorcet[4], adversaire acharné de Descartes:

L'équation algébrique aux coefficients réels

$$f(x) \equiv x^n + a_1 x^{n-1} + a_2 x^{n-2} + \cdots + a_{n-1} x + a_n = 0$$

n'ayant pas de racines imaginaires, le nombre de ses racines positives est égal au nombre des variations des signes de ses coefficients a_μ, et le nom-

[1] Il faut se rappeler que le mémoire de Vandermonde est, en vérité, plus ancien que celui de Lagrange.

[2] Chose curieuse, la publication mathématique de Vandermonde ne contient, outre ces trois mémoires, qu'une petite note intitulée Remarques sur des problèmes de situation, publiée dans les Mémoires 1771, 556—574.

[3] Mémoires de l'Académie des sciences 1741, 72—96.

[4] Éloge de l'abbé de Gua, Histoire de l'Académie 1786, 25.

bre de ses racines négatives est égal au nombre des permanences de ces mêmes signes.

De Gua donne deux démonstrations de cette règle, dont la première est une conclusion de n à $n+1$.

En effet, soit p un nombre positif, de Gua montre que les signes des coefficients de l'équation $(x-p)f(x) = 0$ forment précisément une variation de plus que ceux de $f(x) = 0$.

On trouve, dans ce mémoire, des extensions de la règle de Descartes; nous nous bornerons à indiquer une seule de ces extensions.

Soit α une racine réelle de $f^{(r)}(x) = 0$, r étant un quelconque des indices $1, 2, 3, \ldots, n-1$, l'équation (1) a toutes ses racines réelles, pourvu que $f^{(r-1)}(\alpha)\, f^{(r+1)}(\alpha) < 0$, quelle que soit la racine réelle α de l'équation susdite.

Soit, au contraire, $f^{(r-1)}(\alpha)\, f^{(r+1)}(\alpha) > 0$, l'équation $f(x) = 0$ a deux racines imaginaires, ce qui a lieu pour toutes les valeurs de r et α satisfaisant à cette dernière condition.

Les théorèmes indiqués par Fontaine[1], en ce qui concerne le nombre des racines positives ou négatives d'une équation algébrique, nous semblent si insignifiants que nous les passerons sous silence.

Fourier connaissait déjà en 1787 son célèbre théorème:

Étant donnée une équation algébrique $f(x) = 0$ du degré n et aux coefficients réels, si dans les $n+1$ fonctions

$$f(x), \quad f'(x), \quad f''(x), \ldots, f^{(n)}(x),$$

on substitue deux quantités réelles a et b, $b > a$, et si, après chaque substitution, on compte les

[1] Mémoires de l'Académie des sciences 1747, 665—677.

variations des signes que présente la suite des résultats, le nombre des racines de $f(x) = 0$ comprises entre a et b ne peut jamais surpasser celui des variations perdues de $x = a$ à $x = b$, et, quand il est moindre, la différence est toujours un nombre pair.[1]

L'illustre géomètre a communiqué son théorème dans ses cours à l'École polytechnique en 1796, en 1797 et en 1803. Néanmoins on[2] l'attribue souvent à Budan de Boislaurent, qui l'a communiqué dans un mémoire présenté à l'Institut en 1811, ce qui est aussi absurde qu'attribuer à Poisson ou à Chasles la célèbre sphère de Monge.

Darboux[3] croit que la raison principale de cette confusion, en ce qui concerne la priorité du théorème de Fourier, est à chercher dans la biographie de Fourier par Arago, qui décide, aussi facilement que superficiellement, la question en la faveur de Budan.

Quant à la détermination du nombre des racines réelles d'une equation algébrique aux coefficients réels, nous avons encore à mentionner des règles curieuses dues à Joseph-Balthazar Bérard, principal du collège de Briançon, qui élimine x entre les deux équations

$$f(x) = y, \quad f'(x) = 0.$$

Soit $\varphi(y) = 0$ la réduite de ces équations, étant du degré $n-1$, et soit n un nombre pair, savoir $n = 2m$, une des règles de Bérard est exprimée par le théorème:

Si $f(x) = 0$ n'a pas des racines multiples, elle

[1] On sait que Sturm, dans son beau théorème, a donné un moyen pour déterminer précisément le nombre des racines d'une équation proposée, comprises dans un intervalle quelconque.

[2] Voyez par exemple Serret Cours d'algèbre supérieure I, 267; Paris 1885 (5e éd.). Weber Lehrbuch der Algebra I, 201; Brunswick 1885.

[3] Œuvres de Fourier II, 310—313.

aura 0, 2, 4, ..., $2m$ racines imaginaires, suivant que $\varphi(y) = 0$ aura m, $m \pm 1$, $m \pm 2, \ldots$, $m \pm m$ permanences de signes.[1]

L'idée d'attacher le nombre des racines réelles de $f(x) = 0$ à celui des permanences des signes de $\varphi(y) = 0$ est hardie, trop hardie.

En effet, partons de l'équation trinome

$$x^n + ax - b = 0,$$

a et b étant positifs, la réduite $\varphi(y) = 0$ deviendra

$$\left(\frac{y+b}{n-1}\right)^{n-1} + \left(\frac{a}{n}\right)^n = 0,$$

équation qui n'a que des permanences de signes.

Soit ensuite n un nombre pair, l'équation trinome aurait, d'après la règle de Bérard, toutes ses racines imaginaires; mais cette équation a, dans ce cas, précisément deux racines réelles.

Du reste, des géomètres contemporains, entre autres F.-J. Servois, ancien officier d'artillerie, puis professeur de mathématiques à l'ecole régimentaire de l'artillerie à Lafère, ont démontré la fausseté du théorème susdit, immédiatement après sa publication.

Ces géomètres ont appliqué des équations spéciales aux coefficients numériques.[2]

VIII. Le théorème de d'Alembert.

L'illustre d'Alembert a énoncé le premier, je crois, le théorème fondamental[3]:

Si l'équation algébrique aux coefficients réels

[1] Voyez Annales de mathématiques pures et appliquées IX, 36; 1818.
[2] Ibid. IX, 223—227; 1819.
[3] Mémoires de l'Académie de Berlin 1746, 182—224.

$f(x) = 0$ n'a aucune racine réelle, il existe un nombre imaginaire $\alpha + i\beta$, de sorte que $f(\alpha + i\beta) = 0$.

On voit que ce théorème donnera immédiatement cet autre:

L'équation algébrique du degré n et aux coefficients réels a précisément n racines.

Le fondement de la démonstration de d'Alembert est l'étude de la courbe $y = f(x)$, démonstration qui est reproduite dans l'introduction au premier volume du Calcul intégral de Bougainville, Paris 1754.

Euler[1] a essayé de démontrer ce théorème fondamental, en démontrant que le polynome $f(x)$, du degré $n > 2$ et aux coefficients réels, peut toujours être décomposé en deux facteurs ayant aussi des coefficients réels:

$$f(x) = \varphi(x)\,\psi(x),$$

où

$$\varphi(x) = x^p + b_1 x^{p-1} + b_2 x^{p-2} + \cdots + b_{p-1} x + b_p$$
$$\psi(x) = x^{n-p} + c_1 x^{n-p-1} + c_2 x^{n-p-2} + \cdots + c_{n-p-1} x + c_{n-p}.$$

Il s'agit donc de démontrer que les équations

$$a_r = c_r + c_{r-1} b_1 + c_{r-2} b_2 + \cdots + c_1 b_{r-1} + b_r$$

admettent toujours des valeurs réelles des b_μ et des c_ν, à l'aide des deux propriétés connues d'une équation algébrique aux coefficients réels, savoir que cette équation a toujours une racine réelle, pourvu que son degré soit impair, et qu'elle en a toujours deux, pourvu que son degré soit pair et que son dernier terme soit négatif.

Lagrange[2] remarque avec raison que la démonstration d'Euler est incomplète, et il donne sous cette forme le théorème dont il s'agit:

[1] Mémoires de l'Académie de Berlin 1749.
[2] Nouveaux Mémoires de l'Académie de Berlin 1772, 222—258.

Un polynome aux coefficients réels et d'un degré plus élevé que 2 peut toujours être décomposé en facteurs du premier ou du second degré, ayant aussi leurs coefficients réels.

On sait que Gauss a donné plusieurs démonstrations de ce dernier théorème, la première dans sa thèse de doctorat.[1]

Argand[2] a essayé de démontrer le théorème de d'Alembert, comme il suit:

Introduisons, dans le polynome $f(x)$, un nombre quelconque x_0 à la place de x; si $f(x_0) = 0$, le théorème est démontré; si, au contraire, $f(x_0) \neq 0$, il est possible de déterminer h, étant généralement imaginaire, de sorte que

$$|f(x_0 + h)| < |f(x_0)|;$$

et, partant de cette inégalité, Argand conclut à l'existence d'un nombre α, tel que $f(\alpha) = 0$.

Chose curieuse, Servois[3] fait immédiatement contre cette conclusion une objection très grave et étonnante pour son temps, savoir:

»Ce n'est pas assez, ce me semble, de trouver des valeurs de x qui donnent au polynome des valeurs sans cesse décroissantes; il faut de plus que la loi des décroissements amène nécessairement le polynome à zéro, ou quelle soit telle que zéro ne soit pas, si l'on peut s'exprimer ainsi, l'asymptote du polynome«.

Ni Argand ni les géomètres contemporains n'ont compris la profondeur de cette objection qui, traduite en notre

[1] Helmstedt 1799.

[2] Annales de mathématiques pures et appliquées IV, 133—147; 1813

[3] Ibid. 231.

terminologie, dit qu'il n'est pas suffisant de démontrer que la limite inférieure de $|f(x)|$ est égale à zéro.

On sait que cette difficulté peut être immédiatement détournée par le théorème de Weierstrass sur la limite supérieure ou inférieure d'une fonction réelle et continue des variables réelles.

Ce supplément ajouté, on voit que la démonstration d'Argand est applicable quels que soient les coefficients de l'équation proposée, réels ou imaginaires.

On désigne parfois comme théorème d'Argand le théorème de d'Alembert, ce qui me semble injuste bien que la démonstration d'Argand soit infiniment supérieure à des soi-disant démonstrations plus récentes ou contemporaines.

Parmi ces essais nous nous bornerons à mentionner celui de Daniel Encontre, alors professeur de mathématiques à la Faculté des sciences de Montpellier, essai amélioré par Gergonne, l'éminent rédacteur des Annales.[1]

Soit, dans l'équation algébrique du degré pair

$$(1)\qquad x^{2n} + a_1 x^{2n-1} + a_2 x^{2n-2} + \cdots + a_{2n-1} x - \alpha^2 = 0,$$

le nombre α et les coefficients a_μ tous des nombres réels, cette équation a toujours deux racines réelles; c'est-à-dire qu'il existe une fonction réelle de ces quantités

$$(2)\qquad \varphi(a_1, a_2, \ldots, a_{2n-1}, \alpha)$$

qui, introduite dans (1) à la place de x, satisfait à cette équation.

Posons maintenant, dans (1), $\alpha = i\beta$, β étant un nouveau nombre réel, cette équation deviendra

[1] Voyez ce recueil t. IV, 201—222; 1814.

$$x^{2n} + a_1 x^{2n-1} + a_2 x^{2n-2} + \cdots + a_{2n-1} x + \beta^2 = 0,$$

équation qui est satisfaite par le nombre

$$\varphi(a_1, a_2, \ldots, a_{2n-1}, i\beta),$$

peut-être imaginaire d'une forme très compliquée.

On voit que ces opinions, en ce qui concerne les fondements de l'algèbre, diffèrent essentiellement des nôtres.

TABLE DES MATIÈRES

Forelagt paa Mødet den 30. Marts 1928.
Færdig fra Trykkeriet den 30. Juni 1928.

MATHEMATISK-FYSISKE MEDDELELSER

UDGIVNE AF

DET KGL. DANSKE VIDENSKABERNES SELSKAB

6. BIND (Kr. 17,00):

Kr. Ø.

1. Nielsen, Niels: Sur l'opération itérative des Équations de Lagrange. 1924 .... 3.10
2. Urey, H. C.: On the Effect of perturbing Electric Fields on the Zeeman Effect of the Hydrogen Spectrum. 1924 .... 0.65
3. Bøggild, O. B.: On the Labradorization of the Feldspars. With one plate. 1924 .... 3.00
4. Pedersen, P. O.: Om elektriske Gnister. II. Eksperimentelle Undersøgelser over Gnistforsinkelse og Gnistdannelse. Med 7 Tavler. 1924 .... 4.30
5. Juel, C.: Über Flächen von Maximalindex. 1924 .... 1.25
6. Nielsen, Niels: Sur une Équation de Lagrange. 1924 .... 1.25
7. Hevesy, G. de: Recherches sur les propriétés du Hafnium. Avec 2 planches. 1925 .... 6.25
8. Bohr, Harald: Neuer Beweis eines allgemeinen Kronecker'schen Approximationssatzes. 1924 .... 0.50
9. Bjerrum Niels and Ebert, Ludwig: On some recent Investigations concerning Mixtures of Strong Electrolytes (Transference Numbers and Amalgam Equilibria). 1925 .... 0.75
10. Landau, Edm.: Die Ungleichungen für zweimal differentiierbare Funktionen. 1925 .... 1.60

7. BIND (Kr. 17,70):

1. Bohr, Harald: Unendlich viele lineare Kongruenzen mit unendlich vielen Unbekannten. 1925 .... 1.40
2. Hartmann, Jul., and Trolle, Birgit: On Beat-phenomena in Cylindrical Tubes exposed to Sound-waves. With three plates. 1925 .... 2.85
3. Pauli, W. jr.: Ueber die Intensitäten der im elektrischen Feld erscheinenden Kombinationslinien. 1925 .... 0.65
4. Hardy, G. H. and Littlewood, J. E.: A theorem concerning series of positive terms, with applications to the theory of functions. 1925 .... 0.90
5. Steffensen, J. F.: On a Generalization of Nörlund's Polynomials. 1926 .... 1.00
6. Hartmann, Jul., and Trolle, Birgit: New investigation on the air jet generator for acoustic wawes. 1926 .... 2.40
7. Mollerup, Johs.: Sur l'approximation d'un nombre irrationnel par des carrés rationnels. 1926 .... 0.80
8. Nielsen, Niels: Sur certains développements d'une fonction holomorphe. 1926 .... 0.75

		Kr. Ø.
9.	BJERRUM, NIELS: Untersuchungen über Ionenassoziation. I. Der Einfluss der Ionenassoziation auf die Aktivität der Ionen bei mittleren Assoziationsgraden. 1926	2.00
10.	NIELSEN, NIELS: Recherches sur les fonctions cylindriques et sur certaines fonctions analogues. 1926	1.35
11.	HEVESY, G.: On the Missing Element 87. 1926	0.60
12.	KUHN, W.: Die Stärke der anomalen Dispersion in nicht leuchtendem Dampfe von Thallium und Cadmium. 1926	3.75
13.	SUGIURA, Y. and UREY, H. C.: On the Quantum Theory explanation of the Anomalies in the 6th and 7th Periods of the Periodic Table. 1926	1.15
14.	BURRAU, ØYVIND: Berechnung des Energiewertes des Wasserstoffmolekel-Ions (H_2^+) im Normalzustand. 1927	0.90
15.	KNUDSEN, MARTIN: The Hot-wire Manometer. 1927	0.80
16.	JUEL, C.: Om v. Staudt's Definitioner. 1927	0.85
17.	PÓLYA, G.: Über die algebraisch-funktionentheoretischen Untersuchungen von J. L. W. V. Jensen. 1927	1.40

8. BIND (under Pressen):

1.	HILLE, EINAR: On the Logarithmic Derivatives of the Gamma Function. 1927	2.40
2.	RASCH, G.: Beitrag zur Theorie der unvollständigen Gammafunktionen. Nach hinterlassenen Papieren von J. L. W. V. Jensen. 1927	0.85
3.	KNUDSEN, MARTIN: Thermal Molecular Pressure in Tubes. 1927	1.50
4.	PEDERSEN, P. O.: Den højere Atmosfæres Sammensætning, Tryk, Temperatur og elektriske Ledningsevne i Belysning af Radiobølgernes Udbredelsesforhold. 1927	0.60
5.	BESICOVITCH, A. and BOHR, H.: Some remarks on Generalisations of almost Periodic Functions. 1927	1.50
6.	PEDERSEN, ELLEN: Über einige besondere Klassen von fastperiodischen Funktionen. 1928	1.00
7.	STEFFENSEN, J. F.: A general Summation Formula. 1928	1.25
8.	NIELSEN, NIELS: Observations sur des recherches algébriques plus anciennes que le théorème d'Abel. 1928	1.40

Det Kgl. Danske Videnskabernes Selskab.
Mathematisk-fysiske Meddelelser. **VIII,** 9.

UNDERSØGELSER OVER FARADAYEFFEKTEN HOS VANDIGE OPLØSNINGER AF NOGLE UNI-UNIVALENTE ELEKTROLYTER

AF

E. BUCH ANDERSEN OG R. W. ASMUSSEN

KØBENHAVN
HOVEDKOMMISSIONÆR: ANDR. FRED. HØST & SØN, KGL. HOF-BOGHANDEL
BIANCO LUNOS BOGTRYKKERI
1928

Polarisationsplanets magnetiske Drejning hos sammensatte Stoffer er i Analogi med Brydningsforhold, Varmefylde o. a. til en vis Grad en additiv Egenskab, saaledes at et Atom, der indgaar som Bestanddel i et Molekyle, med Tilnærmelse altid giver samme Bidrag til Molekylets magnetiske Drejning, uafhængig af Molekylets øvrige Bestanddele, men ofte afhængig af den Maade, paa hvilken Atomet er bundet. Drejningen kan derfor i nogen Grad være konstitutivt bestemt og, som det vil kendes fra Perkins Undersøgelser over organiske Stoffer, undertiden være egnet til Afgørelse af konstitutive Spørgsmaal.

Medens Iagttagelsesmaterialet for organiske Stoffers Vedkommende maa siges at foreligge i baade fyldig og alsidig Form, gælder dette paa ingen Maade de uorganiske Forbindelser. De uorganiske Stoffer, hvis Drejning er maalt, bærer Præg af at være tilfældig valgt eller undersøgt med specielle fysiske Formaal for Øje. Et systematisk Iagttagelsesmateriale over uorganiske Forbindelser eksisterer ikke.

Af de foreliggende Maalinger kan det skønnes, at Drejningen af opløste Salte med Tilnærmelse kan beregnes additivt som Summen af Ionernes Drejning. En Blanding af to Salte i Opløsning kan give en Drejning, der er Summen af Bestanddelenes. Men ogsaa her har Konstitutionen Betydning, thi hvis der ved Blandingen finder en Kompleksdannelse Sted, bliver Resultatet ganske anderledes. Eksempelvis kan nævnes, at Drejningen for 2KJ, HgJ_2 beregnes additivt til 84, medens K_2HgJ_4 ved Maaling giver Værdien 136.

Det skulde derfor synes muligt ogsaa indenfor den uorganiske Kemis Omraade at benytte denne Egenskab til Belysning af forskellige Konstitutionsproblemer i Kompleksforbindelsernes Kemi.

Men ved Siden heraf giver Maalingen af Polarisationsplanets magnetiske Drejning visse Oplysninger (ganske vist foreløbig i meget implicit Form), der kan ventes at faa Betydning for den atomteoretiske Forstaaelse af Molekylernes Dannelse og Bygning. Polarisationsplanets magnetiske Drejning er et Dispersionsfænomen, og det, som egentlig maales ved Bestemmelsen af Drejningen, er Zeemanneffekten paa de for Dispersionen virksomme Overgange hos Molekylet. Ligesom Undersøgelsen af Zeemanneffekten paa en Spektrallinie fra et Atom i Dampform kan give Oplysning om Arten af de stationære Tilstande hos Atomet, mellem hvilke Overgangen foregaar, vil Faradayeffekten i Forbindelse med Kendskab til Stoffets Absorptionsspektrum og Dispersion kunne blive et vigtigt Hjælpemiddel ved Studiet af Molekylernes stationære Tilstande.

Drejningsvinklen α, som maales, naar en monokromatisk, planpolariseret Straale passerer gennem et Stof i Lagtykkelsen l, er givet ved:

$$\alpha = V \cdot \int_0^l H\,dl,$$

hvor H er den magnetiske Feltstyrke i Lysstraalens Retning. V er en for Stoffet karakteristisk Størrelse (Verdets Konstant) og er altsaa Stoffets Drejning i 1 cm's Lagtykkelse og i Feltet 1 Gauss i Straaleretningen. Den samlede Drejning α er proportional med Forskellen i magnetisk Potential mellem Stedet, hvor Lyset træder ind i Stoffet, og Stedet, hvor det træder ud.

BECQUEREL[1] har paa klassisk Grundlag vist, at Verdets Konstant kan udtrykkes ved:

$$V = \frac{e}{2mc^2} \cdot \lambda \cdot \frac{dn}{d\lambda},$$

hvor $\frac{e}{m}$ er Forholdet mellem Elektronens Ladning og Masse, c Lyshastigheden, λ Bølgebredden og n Brydningsforholdet. Ogsaa paa mere moderne Grundlag kommer DARWIN og WATSON[2] under visse Forudsætninger til ganske den samme Formel. Naar Faradayeffekten kan beskrives klassisk, betyder det, at man har at gøre med en normal Zeemanneffekt, og under disse Omstændigheder kan man altsaa ved Maaling af $\frac{dn}{d\lambda}$ beregne V. DARWIN og WATSON har udført saadanne Beregninger for en Del Stoffer, hvor de fornødne Data var til Raadighed, og derved fundet, at de beregnede Værdier for Verdets Konstant i Almindelighed kun hvad Størrelsesordenen angaar stemmer med de fundne. Forholdet $\frac{V_{\text{obs.}}}{V_{\text{ber.}}}$ kalder de den magnetiske Anomali, og dennes Afvigelse fra 1 kan betragtes som et Maal for, i hvilken Udstrækning de virksomme Overgange hos de paagældende Stoffer viser anomal Zeemanneffekt.

En dyberegaaende Udredelse af Forholdene kræver imidlertid Analyse af Stoffernes Dispersion i Relation til deres Absorptionslinier. Dette foreligger kun i et ringe Antal Tilfælde. Men ogsaa Maalinger af den magnetiske Drejning foreligger, som nævnt, for uorganiske Stoffer kun spredt. Vi har derfor i dette Arbejde søgt at tilvejebringe Begyndelsen til et mere systematisk Materiale for en Række simple Stoffers Vedkommende. Undersøgelsen omfatter forskellige Forbindelser af Kationerne H, Li, Na, K, Rb, Cs og

[1] Compt. rend. 125. p. 679. 1897.

[2] Proc. Roy. Soc. A. 114. p. 474. 1927.

NH_4 med Anionerne F, Cl, Br, J, OH, ClO_3, BrO_3 og JO_3. Disse Ioner er i deres Normaltilstand alle diamagnetiske (ved paramagnetiske Stoffer intræder mere komplicerede Forhold, Ladenburg[1]), og da Maalingerne kun omfatter Farver i det synlige Spektrum, kan disse alle regnes at være udført i Spektralomraader, som ligger langt fra Stoffernes Absorptionsbaand. Vi har ikke haft noget Præcisionsapparat til Raadighed. Vore Tal er derfor at betragte som et Oversigtsmateriale og ifølge Sagens hele Stilling mere beregnet paa at drage Sammenligninger mellem disse Ioner indbyrdes end en dybtgaaende Undersøgelse af den enkelte Ions Forhold. Usikkerheden paa Resultaterne vil senere blive diskuteret.

De undersøgte Opløsningers Drejning er, saavidt Forholdene tillod det, maalt for tre forskellige Farver ($578\mu\mu$, $546\mu\mu$ og $436\mu\mu$) og alle med Vand som Sammenligningssubstans. Vi har fulgt den sædvanlige Praksis og som Resultat opført Stoffernes »molekylære Drejning« (M), hvis Sammenhæng med Verdets Konstant er:

$$M = \frac{V \cdot m}{d},$$

hvor m og d er det undersøgte Stofs Molekylvægt og Vægtfylde. Da M er angivet med Vandets molekylære Drejning under samme Forhold som Enhed, kan Drejningernes Absolutværdier altsaa findes ved Hjælp af Verdets Konstant for Vand, som ifølge Roger og Watson[2] for Na-Lys er:

$$V_{\text{Vand}} = 0{,}01311 - 0{,}0_6 4 \cdot t - 0{,}0_7 4 \cdot t^2.$$

V er her udtrykt i Bueminutter, og t er en Temperatur mellem 4° og 98°. (Som det er at vente, er V hos diamag-

[1] Zs. f. Phys. 46. p. 168. 1927.

[2] Zs. f. Phys. Ch. 19. p. 357. 1896.

netiske Stoffer kun i ringe Grad afhængig af Temperaturen, og hos Vand gælder dette i endnu højere Grad for *M*, da Temperaturens Indflydelse paa *V* og *d* delvis kompenserer hinanden).

Nu har vi ikke anstillet Maalinger med Na-Lys. Men ved Hjælp af tidligere Bestemmelser af Vandets Drejning ved forskellige Bølgebredder (med hvilke vore Maalinger er i god Overensstemmelse) kan vi af vore Tal beregne nedenstaaende Værdier for Vandets Drejning, naar dets Drejning for Na-Lys under ellers samme Forhold sættes lig 1,000:

λ	V
589 $\mu\mu$	1,000
578 —	1,043
546 —	1,183
436 —	1,960

For at kunne beregne et opløst Stofs molekylære Drejning af Opløsningens maalte Drejning antages det, at opløst Stof og Opløsningsmiddel drejer uafhængigt af hinanden, og at den maalte Drejning er sammensat additivt af disse to. Herved ledes man til følgende Udtryk for det opløste Stofs molekylære Drejning (med Vandets molekylære Drejning lig 1), naar Opløsningsmidlet er Vand:

$$M = \frac{D_2}{D_1} \cdot \frac{(\mu m_1 + m_2)}{d m_1} - \mu, \tag{1}$$

idet

M er det opløste Stofs mol. Drejning i Forhold til Vand,
D_1 er maalt Drejningsvinkel for Vand,
D_2 er maalt Drejningsvinkel for Opløsning,
} ved samme Lagtykkelse, Temp. og Magnetfelt.
m_1 er Molekylvægten af Vand,
m_2 er Molekylvægten af det opløste Stof,

μ er Antallet af g-Mol. Vand pr. g-Mol. opløst Stof,
d er Vægtfylden af Opløsningen i Forhold til Vand ved den Temp., ved hvilken Drejningsvinklen maales.

At det opløste Stof og Opløsningsmidlet altid skulde dreje helt uafhængigt af hinanden, er jo ikke at vente, og det viser sig da ogsaa, at de fundne mol. Drejninger for de opløste Stoffer varierer med Opløsningernes Koncentration, undertiden endda ret stærkt. Der maa her være en Vej til Studiet af det opløste Stofs Tilstand i Opløsningen.

Naar Opløselighedsforholdene tillod det, har vi i Almindelighed maalt de paagældende Opløsningers Drejning i 3 Koncentrationer, nemlig omtrent mættet Opløsning, $\mu = 15-20$ og $\mu = 30-35$. Ved den sidste Koncentration var dog Usikkerheden paa Resultatet saa stor, at vi ikke i vor Tabel har medtaget disse Bestemmelser. Alle de maalte Opløsninger er temmelig koncentrerede.

Eksperimentalteknik.

Det anvendte Apparat blev bygget sammen af Dele, som fandtes i Laboratoriet. Selve Polarisationsapparatet var af Laurent-Typen (Synsfeltet delt i to Halvdele langs en lodret Diameter) indrettet til Na-Lys. Da vi ved Maalingerne benyttede Lys af andre Bølgebredder, kunde den ene Felthalvdel ikke give fuldstændigt Mørke; men Indstilling paa lige stor Klarhed i de to Halvdele af Synsfeltet lod sig dog uden Vanskelighed gennemføre.

Polarisationsrøret til Opløsningerne var 40 cm langt (Rmf. ca. 35 cm^3). Det var omtrent i hele sin Længde (ca. 38 cm) omgivet af en Solenoide, viklet af Kobbertraad med Diameter 1,5 mm. Den samlede Kobbervægt androg ca. 20 kg, og Solenoidens Modstand ved Stuetemperatur ca. 7 Ohm. For at faa nogenlunde store Drejningsvinkler at maale, anvendtes Strømstyrker paa 20—23 Amp., hvilket svarer til, at der i Solenoiden afsattes en Effekt paa 3000—3700 Watt. Dette medførte naturligvis en hurtig Opvarmning af Spolen.

Af Hensyn til Drejningens Temperaturafhængighed maatte der træffes særlige Foranstaltninger for at holde Opløsningerne paa en konstant og veldefineret Temperatur under Maalingerne. Til dette Formaal var Polarisationsrøret omgivet af en Kappe med Vand og revet Is, idet vi fandt, at dette under de givne Forhold var den simpleste og mest virkningsfulde Metode til at forhindre en Temperaturstigning i den undersøgte Opløsning. Alle vore Maalinger er altsaa foretaget ved 0° C. I den varmere Aarstid, hvor Luftens Fugtighedsindhold er større, kunde det være nødvendigt at fugte de udvendige Sider af Dækglassene i Polarisationsrøret med et ganske tyndt Lag Glycerin for at forhindre, at de blev matte af Dug. I den kolde Aarstid var dette næsten aldrig nødvendigt. Yderligere blæste vi jævnlig under Maalingerne en tør Luftstrøm hen over den Ende af Polarisationsrøret, som vendte mod Iagttageren. Der var naturligvis truffet Foranstaltninger til at sikre, at Røret altid blev anbragt i samme Stilling i Forhold til Apparatets øvrige Dele.

Som Lyskilde anvendtes en Kvarts-Kviksølvlampe (135 Watt), og af Kviksølvspektret udblendedes de ønskede Straaler ved Hjælp af Vædskefiltre. Filtrene (delvis efter O. Warburg, Stählers Haandbog, II. 2, p. 1530) var følgende (for Vædskernes Vedkommende i 1 cm's Lagtykkelse):

578 $\mu\mu$ { a) 1/4 molær $CuSO_4$.
b) 0,02 g Tartrazin + 0,02 g Erythrosin i 100 cm^3 Vand.

546 $\mu\mu$ { a) Wratten Filter Nr. 77 a. (Special green Hg-line).
b) 18 g Didymklorid i 50 cm^3 Vand.

436 $\mu\mu$ { a) 1/8 molær Ni $(NO_3)_2$.
b) 50 cm^3 1/2 molær $CuSO_4$ + 10 cm^3 konc. NH_4OH.

Ved Hjælp af disse Filterkombinationer opnaaedes en til vort Formaal tilstrækkelig monokromatisk Straaling.

De undersøgte Opløsninger blev fremstillet af de renest mulige Udgangsmaterialer (eventuelt efter yderligere Rensning af disse), og deres Koncentration bestemt ved Analyse (mindst to Bestemmelser paa hver Opløsning). Vægtfylden ved 0° blev maalt ved Udvejning i et kalibreret Pyknometer (ca. 15 cm^3).

De optiske Maalinger blev udført paa følgende Maade. Polarisationsrøret med Opløsningen blev henstillet til Afkøling med revet Is i Kappen i ca. 1 Time, idet der stadig blev sørget for Tilstedeværelsen af rigelige Mængder Is. Derefter blev det med rigelig Isforsyning anbragt i Polarisationsapparatet. Paa Grund af Spolens Opvarmning under Maalingerne ændrede Strømstyrken (og dermed Magnetfeltstyrken) ret hurtigt sin Værdi, og det viste sig at være for besværligt at holde Strømstyrken konstant ved Hjælp af Regulermodstande. Vi valgte derfor at lade den ene Iagttager foretage Indstillingen paa det optiske Apparat, medens den anden paa givet Signal aflæste Strømstyrken i Indstillingsøjeblikket paa et indskudt Præcisionsampèremeter. Hver Opløsning blev ved hver Bølgebredde maalt ialt 4 Gange (2 Gange af hver Iagttager), ligeledes blev Apparatets Nulstilling før og efter hvert saadant Sæt Maalinger bestemt 2 Gange af hver Iagttager. De fundne Drejningsvinkler blev først korrigeret for det tomme Rørs Drejning i Magnetfeltet (5—10 Bueminutter), derpaa reduceret til samme Strømstyrke (idet Drejningen er ligefrem proportional med Strømstyrken); endelig blev de fire Enkeltbestemmelser forenet til et Middeltal, og det er dette, der fremtræder som Resultat i Tabellerne.

De maalte Drejningsvinkler laa mellem Grænserne 18° og 140°. Polarisationsapparatet var forsynet med Nonius, som tillod Vinkelaflæsning med 1 Bueminuts Nøjagtighed. Indstillingsusikkerheden var dog i Almindelighed paa Grund af Ændringen i Feltstyrken en Del større. Tabellen Side 13 med Maalingerne af Vandets Drejning giver et Indtryk af Tallenes Paalidelighed. Ogsaa med en Del Opløsninger har vi for Kontrollens Skyld gentaget Bestemmelserne og har hver Gang kunnet reproducere Resultaterne med nogle faa Promilles Afvigelse. Vi mener at turde sige, at Usikkerheden paa Vinkelmaalingerne i Almindelighed ikke overstiger $^1/_2\,^0/_0$ af Værdien.

For at undersøge, hvilken Virkning de forskellige Fejlkilder kan have paa det endelige Resultat, dannes af Ligning (1) Udtrykket:

$$\varDelta M = \frac{\partial M}{\partial D_1} \cdot \varDelta D_1 + \frac{\partial M}{\partial D_2} \cdot \varDelta D_2 + \frac{\partial M}{\partial d} \cdot \varDelta d + \frac{\partial M}{\partial \mu} \cdot \varDelta \mu,$$

der giver:

$$\varDelta M = \frac{D_2}{D_1} \cdot A \cdot \frac{\varDelta D_1}{D_1} + A \cdot \frac{\varDelta D_2}{D_1} + \frac{D_2}{D_1} \cdot A \cdot \frac{\varDelta d}{d} + \left(\frac{D_2}{D_1} \cdot \frac{1}{d} - 1\right) \cdot \varDelta \mu, \quad (2)$$

hvor

$$A = \frac{\mu m_1 + m_2}{d m_1} \text{ eller i Reglen ca. } 1{,}15 \cdot \mu.$$

Det første Led i (2) repræsenterer Usikkerheden i M hidrørende fra Usikkerheden paa Bestemmelsen af Sammenligningsstoffets, Vandets, Drejning. Dette Led har kun Interesse, naar det gælder om at bestemme Absolutværdien for M saa nøjagtig som muligt. For stærke Opløsninger, hvor A er mindre end 10 og $\frac{D_2}{D_1}$ lig 2 — 3, medens $\frac{\varDelta D_1}{D_1}$ eksperimentelt er bestemt til 0,001, vil Usikkerheden paa D_1 kun kunne give et Par Enheders Fejl i 2. Decimal for M. Ved de tyndere Opløsninger bliver Produktet $A \cdot \frac{D_2}{D_1}$ noget større, men Fejlen hidrørende fra dette Led maa alligevel kaldes uvæsentlig i Sammenligning med andre. Vore Resultater er alle beregnede under Benyttelse af samme Værdi for D_1 og kan derfor uden videre sammenlignes.

Det næste Led i Formel (2) giver Usikkerheden paa M hidrørende fra Bestemmelsen af D_2. $\frac{\Delta D_2}{D_1}$ andrager højst $3{,}66 \cdot 0{,}005$ (i en meget koncentreret Opløsning), ellers er det mindre, for de tyndeste Opløsninger under 0,01; men da det multipliceres med A, ses det at kunne give Anledning til en Fejl paa et Par Enheder i første Decimal paa M. Dette Led, som inkluderer Fejlen paa Strømmaalingen, repræsenterer langt den alvorligste Kilde til Usikkerhed paa den endelige M-Værdi.

For Ligning (2)'s tredie Led gør lignende Betragtninger sig gældende som for det første Led; men da Vægtfyldebestemmelsen er den mest nøjagtige af alle Maalingerne, ses det, at Usikkerheden i denne ingen praktisk Betydning faar ved Siden af Usikkerheden paa Maalingen af Drejningsvinklen.

Hvis x er en Opløsnings procentiske Indhold af opløst Stof, bestemmes μ ved:

$$\mu = \frac{(100 - x) \cdot m_2}{x \cdot m_1},$$

hvoraf

$$\frac{\Delta \mu}{\mu} = \frac{100}{100 - x} \cdot \frac{\Delta x}{x}.$$

$\frac{\Delta x}{x}$ kan regnes at være 0,005; $\frac{\Delta \mu}{\mu}$ vil altsaa i de fleste Tilfælde ikke overstige 0,01. Da Udtrykket i Parentesen i Ligningens sidste Led ikke overstiger Værdien 1 og i Reglen er en lille ægte Brøk, kan Usikkerheden paa M hidrørende fra μ-Bestemmelsen ogsaa kun beløbe sig til et Par Enheder i 2. Decimal.

Alt i alt andrager Usikkerheden paa vore endelige Resultater altsaa et Par Procent af deres Værdi. Maalingerne i grønt anser vi for de bedste. Vi har anført disse Betragtninger over Fejlkilderne saa udførligt, fordi man i tidligere Arbejder over Polarisationsplanets magnetiske Drejning kan finde Resultaterne angivet med et Antal Cifre, der — ligesom de deraf dragne Konklusioner — tyder paa en ikke ringe Overvurdering af Resultaternes Paalidelighed.

Maaleresultater.

Alle vore Maalinger er, som nævnt, udført ved 0°C. Da Vandets Drejning er benyttet som Enhed, har vi maalt Effekten for dette Stof adskillige Gange, jævnt fordelt over det Tidsrum, Arbejdet strakte sig over. Vi har dermed tilstræbt dels at faa en god Bestemmelse af denne Størrelse og dels at kontrollere Forsøgsbetingelsernes Konstans.

Vandets Drejning ved 0°C.

	$\lambda = 578\,\mu\mu$	$\lambda = 546\,\mu\mu$	$\lambda = 436\,\mu\mu$
	18°,26	20°,76	34°,28
	18°,37	20°,68	34°,26
	18°,30	20°,69	34°,31
	18°,27	20°,63	34°,32
	18°,21	20°,76	34°,29
	18°,24	20°,72	34°,51
	18°,31	20°,71	34°,45
	18°,26	20°,76	34°,41
	18°,26	20°,74	34°,23
	18°,21	20°,57	34°,36
Middel:	18°,27 ± 0°,015	20°,70 ± 0°,020	34°,34 ± 0°,028

Alle disse Drejningsvinkler er reduceret til en vilkaarlig valgt, men bestemt Strømstyrke (i = 100 Skaladele paa Ampèremeteret, svarende til ca. 19 Amp.). Under Maalingerne var Strømstyrken i Reglen større. Af Vandets Drejning kan det beregnes, at de reducerede Drejningsvinkler svarer til en magnetisk Potentialforskel paa ca. 80000 Gauss. cm mellem Polarisationsrørets Endeflader eller en gennemsnitlig Feltstyrke paa ca. 2000 Gauss langs Røret.

I den følgende Tabel er endelig vore Maaleresultater for Opløsningerne anført. De benyttede Bogstaver har samme Betydning som angivet Side 7. Kolonnen »$^0/_0$« giver Opløsningernes procentiske Indhold af det anførte Stof. Drejningsvinklerne er reduceret til samme Strømstyrke som ved Vandet.

Som det fremgaar af Tabellen, kunde vi ikke altid maale Drejningen for alle tre Farver. Selv en meget svag Gulfarvning gør sig i 40 cm's Lagtykkelse saa stærkt gældende, at Synsfeltet i blaat, for hvilken Farve Øjet i Forvejen ikke er særlig følsomt, bliver for mørkt til Maaling. Ved LiJ gjaldt dette ogsaa grønt. Af samme Grund maatte vi trods gentagne Forsøg helt opgive at maale HJ.

Det vides ikke, hvorfra Gulfarvningen i CsCl-Opløsningerne stammer. Det anvendte Præparat var venligst stillet til Raadighed fra Læreanstaltens fysisk-kemiske Laboratorium og havde ved Prøve (Bestemmelse af Forholdet: $\frac{Cl}{CsCl}$) vist sig analytisk rent. Foruden dette Stof benyttede vi et andet CsCl-Præparat, som gav helt farveløs Opløsning; men af dette fandtes kun saa lidt, at Opløsningen blev temmelig fortyndet og Maaleresultatet i tilsvarende Grad usikkert. For dog at kunne angive en Bestemmelse af Cæsiumionens Drejning i blaat, har vi i Tabellen medtaget Maalingerne af denne Opløsning.

Stof	m_2	%	μ	d	$\lambda = 578\,\mu\mu$		$\lambda = 546\,\mu\mu$		$\lambda = 436\,\mu\mu$	
					D_2	M	D_2	M	D_2	M
KF	58,10	45,96	3,791	1,3815	18°,93	1,47	21°,33	1,44	35°,21	1,42
KF	—	21,46	11,80	1,2034	18°,85	1,08	21°,20	0,99	35°,24	1,02
HCl . . .	36,47	16,22	10,46	1,0826	23°,67	4,48	26°,86	4,51	45°,05	4,67
HCl . . .	—	7,612	24,57	1,0402	20°,92	4,71	23°,57	4,55	39°,41	4,78
LiCl . . .	42,40	29,31	5,675	1,2403	28°,22	4,32	31°,93	4,31	53°,65	4,44
LiCl . . .	—	13,05	15,69	1,0785	22°,23	4,67	25°,25	4,72	42°,13	4,84
NaCl . .	58,46	25,41	9,526	1,2040	25°,36	5,20	28°,66	5,16	48°,14	5,35
NaCl . .	—	12,80	22,10	1,0992	22°,11	5,81	24°,52	5,22	41°,03	5,45
KCl . . .	74,56	21,24	15,35	1,1480	22°,61	5,66	25°,31	5,41	42°,75	5,79
KCl . . .	—	15,34	22,84	1,1059	21°,31	5,62	24°,10	5,57	40°,24	5,75
RbCl . .	120,96	30,17	15,54	1,2833	22°,73	6,04	25°,81	6,09	43°,23	6,29
RbCl . .	—	21,77	24,13	1,1846	21°,13	5,99	23°,86	5,89	39°,73	6,00
CsCl . . .	168,26	59,62	6,325	1,7425	27°,83	7,37	31°,72	7,45	—	—
CsCl . . .	—	36,27	16,41	1,3779	22°,74	6,85	25°,96	7,03	—	—
CsCl . . .	—	14,06	57,01	1,1206	19°,79	6,60	22°,39	6,51	37°,52	7,16
NH_4Cl .	53,50	19,54	12,25	1,0596	23°,36	6,12	26°,74	6,31	44°,51	6,37
NH_4Cl .	—	12,54	20,71	1,0407	21°,44	5,63	24°,44	5,79	40°,62	5,84
HBr . . .	80,93	43,41	5,855	1,4327	36°,52	8,58	41°,51	8,63	—	—
HBr . . .	—	20,60	17,31	1,1629	25°,26	8,62	28°,88	8,85	—	—
LiBr . .	86,86	45,27	5,828	1,4663	36°,52	8,69	41°,49	8,73	71°,78	9,36
LiBr . .	—	23,70	15,52	1,2029	26°,33	8,86	29°,88	8,89	50°,86	9,53
NaBr . .	102,92	40,06	8,543	1,4278	32°,15	9,03	36°,79	9,21	62°,90	9,75
NaBr . .	—	25,58	16,62	1,2407	25°,85	8,85	29°,62	9,14	50°,11	9,65
KBr . . .	119,02	33,32	13,23	1,3033	27°,21	9,44	30°,97	9,54	52°,77	10,16
KBr . . .	—	24,30	20,58	1,2058	24°,30	9,41	27°,54	9,42	46°,62	10,03
RbBr . .	165,42	45,92	10,81	1,4907	28°,51	10,12	32°,29	10,11	55°,21	10,75
RbBr . .	—	30,94	20,48	1,2882	24°,42	10,30	27°,77	10,41	46°,69	10,83
NH_4Br	97,96	26,79	14,86	1,1749	26°,51	10,21	30°,11	10,27	51°,40	11,00
NH_4Br	—	20,05	21,68	1,1244	24°,25	10,34	27°,50	10,37	46°,75	11,16
LiJ	133,86	23,04	24,82	1,2009	29°,40	18,40	—	—	—	—
NaJ . . .	149,92	60,98	5,325	1,8532	60°,52	19,07	69°,52	19,41	125°,73	21,64
NaJ . . .	—	34,88	15,53	1,3624	35°,81	18,79	40°,76	18,95	72°,06	21,21
KJ	166,02	51,51	8,674	1,5799	44°,79	19,09	51°,45	19,48	91°,60	21,54
KJ	—	39,75	13,96	1,3987	36°,12	19,80	40°,98	19,85	72°,29	20,92
NH_4J . .	144,96	55,10	6,555	1,5166	50°,87	20,25	58°,26	20,55	103°,70	22,52
NH_4J . .	—	35,13	14,86	1,2793	35°,49	19,93	40°,34	20,04	—	—
NaOH .	40,01	12,98	14,88	1,1613	21°,42	2,39	24°,33	2,43	—	—
KOH . .	56,11	14,78	17,96	1,1474	20°,71	2,87	23°,34	2,75	38°,96	2,88
NH_4OH	35,05	10,25	17,03	0,9790	18°,83	2,95	21°,39	3,00	34°,79	2,61
$NaClO_3$	106,46	43,26	7,757	1,3771	20°,51	3,39	23°,16	3,35	38°,13	3,26
$NaClO_3$	—	23,91	18,80	1,1941	19°,34	3,11	21°,90	3,10	36°,33	3,10
$NaBrO_3$	150,92	19,86	33,78	1,1918	20°,86	6,61	23°,40	6,21	38°,58	5,97
$LiJO_3$ . .	181,86	28,03	25,90	1,3034	23°,66	9,87	27°,07	10,22	45°,42	10,63

Opløsningerne af Natrium- og Kaliumhydroxyd blev fremstillet af Mercks Præparater i Tabletter, som skulde have et særlig ringe Indhold af Karbonat. Opløsningerne blev analyseret umiddelbart efter de optiske Maalinger, og de anførte Analyseresultater giver hele Alkalimængden beregnet som Hydroxyd. Særlige Undersøgelser viste, at Karbonatindholdet i de to Opløsninger var under 2% af Hydroxydmængden. Dette Forhold har vi ikke korrigeret for ved Beregningen af *M*, da Karbonationen og Hydroxylionen omtrent har samme Drejning. Aarsagen til Natriumhydroxydopløsningens Uigennemsigtighed i blaat var en svag Opalescens.

$LiJO_3$-Opløsningen gav efter Fremstillingen en ringe Udskillelse (formentlig af et basisk Salt). Af denne Grund bestemte vi Indholdet af baade Li og JO_3 i den ved Maalingerne benyttede Opløsning. Li-Indholdet viste sig at være lidt for lavt i Forhold til Jodatmængden (2% af Værdien). Den i Tabellen angivne Koncentration refererer sig til Jodatbestemmelsen, og ved Beregningen af *M* har vi ligeledes lagt denne til Grund, idet Lithiumionens Drejning i Sammenligning med Jodationens er meget ringe.

Alle Stoffer, som er undersøgt i Opløsninger af forskellig Koncentration, viser, at *M* varierer med Koncentrationen (i en Del Tilfælde dog indenfor Maaleusikkerheden). Særlig stor er Variationen hos KF. Endvidere bemærkes det, at Variationen ikke altid er lige stor (ja end ikke altid har samme Fortegn) for de tre Farver. Hos Saltsyre er Ændringen i *M* ret stor i gult, hvilket stemmer med tidligere Maalinger, i grønt og blaat er den ubetydelig. Lignende Forhold genfindes hos NaCl. Et Fænomen, der synes at være reelt, er, at *M* for gult Lys i Rækken af Klorider vokser med voksende Fortynding, naar Kationen har lavt

Atomnummer, men aftager med voksende Fortynding, naar Kationens Atomnummer er højt; midt i Rækken er Variationen omtrent 0.

I Fig. 1 er Maaleresultaterne afbildet grafisk. Op ad Ordinaten er afsat Forbindelsernes molekylære magnetiske Drejning for grønt Lys i Opløsninger med μ lig 15—25

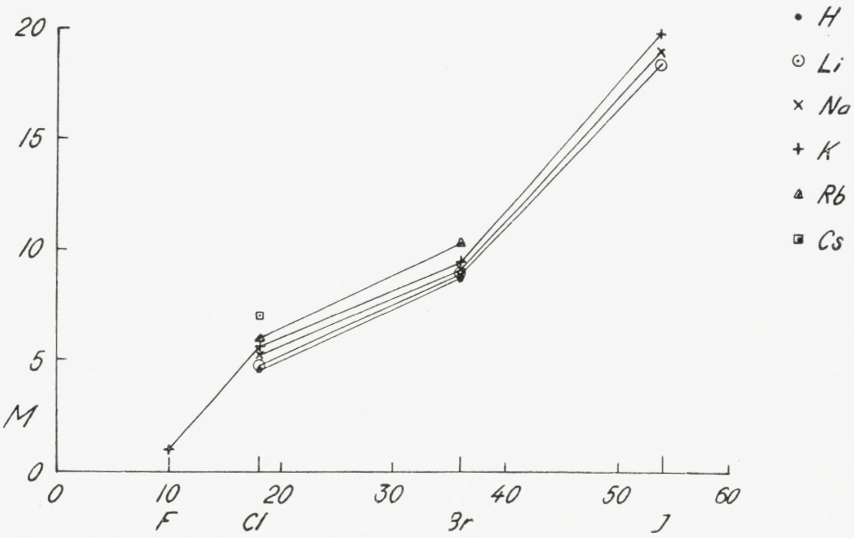


Fig. 1.

(Ved LiJ er Værdien for gult benyttet). Ud ad Abscissen er Anionens Atomnummer afsat. Det fremgaar tydelig af Figuren, at Drejningen hos disse Forbindelser med Tilnærmelse kan sammensættes additivt af de to Ioners Drejning. Endvidere ses det, at Ændring af Anionen med samme Kation giver en stor Forandring i Drejningen, medens Ændring af Kationen med samme Anion kun giver en ringe Variation i M.

De enkelte Ioners molekylære magnetiske Drejning kan kun beregnes, hvis man vilkaarlig fastsætter Værdien for en af dem. Vi kan vælge at sætte Brintionens Drejning lig O. (Dette vilde være strengt rigtigt, hvis Brintionen i

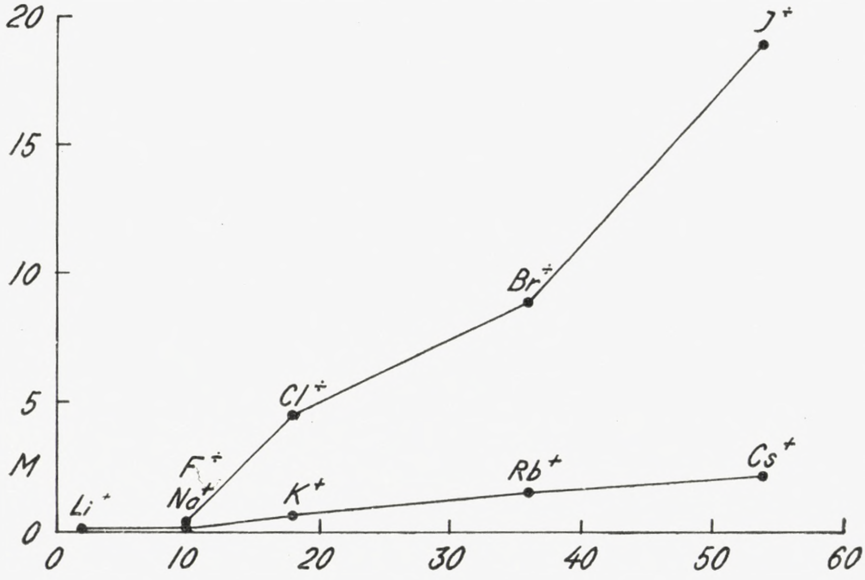


Fig. 2.

en Opløsning var identisk med en Brintkerne). Af Maalingerne findes da følgende Værdier for Ionernes Drejning i grønt.

Kation	*M*	Anion	*M*
H	0,00	F	0,39
Li	0,11	Cl	4,43
Na	0,13	Br	8,97
K	0,60	J	19,04
Rb	1,45	OH	2,15
Cs	2,08	ClO_3	1,76
NH_4	1,00	BrO_3	4,77
		JO_3	10,12

Disse Tal er afbildet grafisk i Fig. 2. Kurverne svarer i deres Form og Beliggenhed ganske til de analoge Kurver over Ionernes Volumina og deres elektriske Polariserbarhed (Molekylarrefraktion). De additive Relationer er dog, som nævnt, kun gyldige med Tilnærmelse. En Prøve viser, at der optræder Afvigelser af systematisk Karakter mellem de observerede og de af den sidst anførte Tabels Tal beregnede Værdier for *M*.

Vort Talmateriale giver et Par gode Eksempler paa, hvorledes et enkelt Atoms Drejning kan være afhængig af den Maade, paa hvilken Atomet er bundet i den paagældende Forbindelse. I nedenstaaende Tabel er angivet, hvor meget en Halogenforbindelses molekylære Drejning forøges, naar man i Molekylet erstatter henholdsvis et Kloratom med Brom eller et Bromatom med Jod. Tallene for Halogeniderne og Iltsyrerne er taget fra vore egne Maalinger, Tallene for organiske Forbindelser fra tidligere Bestemmelser af Perkin og andre.

	MX	MXO_3	Org. Forb.
Br — Cl	4,53	3,01	1,83
J — Br	10,07	5,35	4,20

Vi skal ikke paa dette Sted forsøge paa at diskutere disse Forhold nærmere, men vil blot til Slut gøre opmærksom paa et Par bemærkelsesværdige Kendsgerninger, som Undersøgelsen har bragt for Dagen, nemlig at Vand kun drejer omtrent halvt saa meget som Summen af Brint- og Hydroxylion, medens Ammoniumhydroxyd drejer som Summen af Ammonium- og Hydroxylion.

Laboratoriets Chef, Hr. Prof., Dr. phil. J. Petersen, takker vi for venlig Interesse for vort Arbejde.

Den polytekniske Læreanstalts kemiske Laboratorium A,
August 1928.

Færdig fra Trykkeriet den 31. December 1928.

Det Kgl. Danske Videnskabernes Selskab.
Mathematisk-fysiske Meddelelser. **VIII,** 10.

ON THE LICHTENBERG FIGURES

PART III. THE POSITIVE FIGURES

BY

P. O. PEDERSEN

WITH 28 PLATES

KØBENHAVN
HOVEDKOMMISSIONÆR: ANDR. FRED. HØST & SØN, KGL. HOF-BOGHANDEL
BIANCO LUNOS BOGTRYKKERI
1929

Pris: Kr. 10,00.

Det Kgl. Danske Videnskabernes Selskabs videnskabelige Meddelelser udkommer fra 1917 indtil videre i følgende Rækker:

Historisk-filologiske Meddelelser,
Filosofiske Meddelelser,
Mathematisk-fysiske Meddelelser,
Biologiske Meddelelser.

Hele Bind af disse Rækker sælges 25 pCt. billigere end Summen af Bogladepriserne for de enkelte Hefter.

Selskabets Hovedkommissionær er *Andr. Fred. Høst & Søn*, Kgl. Hof-Boghandel, København.

Det Kgl. Danske Videnskabernes Selskab.
Mathematisk-fysiske Meddelelser. **VIII**, 10.

ON THE LICHTENBERG FIGURES

PART III. THE POSITIVE FIGURES

BY

P. O. PEDERSEN

WITH 28 PLATES

KØBENHAVN
HOVEDKOMMISSIONÆR: ANDR. FRED. HØST & SØN, KGL. HOF-BOGHANDEL
BIANCO LUNOS BOGTRYKKERI
1929

CHAPTER I

Introduction.

The determination of the spreading-out-velocity of the LICHTENBERG Figures has been discussed in some previous publications which have also touched upon the problem of the formation of these figures[1].

In the meantime the photographic LICHTENBERG Figures have been successfully applied to the study of surges on high tension lines, especially the kind of surges due to lightning[2]. They have also been used for the measurement

[1] P. O. PEDERSEN: "On the Lichtenberg Figures": Part I. Vidensk. Selsk. Math.-fys. Medd. Vol. I, No. 11. Copenhagen (February 1919); Part II. Vol. IV, No. 7, Copenhagen 1922; referred to as L. F. I and L. F. II respectively. — "Die Ausbreitungsgeschwindigkeit der Lichtenbergschen Figuren und ihre Verwendung zur Messung sehr kurzer Zeiten". Ann. d. Physik (IV) Bd. 69, p. 205—230, 1922.

[2] J. F. PETERS: "The Klydonograph" "El. World" Vol. 83, p. 769—773, 1924. — J. H. COX and J. W. LEGG: "Trans. A. I. E. E." p. 857—870, 1925. — K. B. MCEACHRON: "Trans. A. I. E. E.", p. 712—717, 1926. — J. H. COX, P. H. MCAULEY and L. GALE HUGGINS: l. c. p. 315—329, 1927. — J. H. COX: l. c. p. 330—338, 1927. — R. J. C. WOOD: "Trans. A. I. E. E." p. 961—968, 1925. — EVERETT S. LEE and C. M. FOUST: "Trans. A. I. E. E." p. 339—348, 1927 and "Gen. Elec. Review" Vol. 30, p. 135—145, 1927. — W. W. LEWIS: "Trans. A. I. E. E." p. 1111—1121, 1928. — E. W. DILLARD: l. c., p. 1122—1124, 1928. — J. G. HEMSTREET and J. R. EATON: l. c., p. 1125—1131, 1928. — PHILIP SPORN: l. c., p. 1132—1139, 1928. — N. N. SMELOFF: l. c., p. 1140—1147, 1928. — H. MÜLLER: Mitteil. d. Hermsdorf Schomburg Isolatoren G. m. b. H., Heft 27, p. 813—829, 1926. — P. O. PEDERSEN: "Ingeniøren", p. 201—209, 1928. "Danmarks Naturvidenskabelige Samfunds Skrifter", A. No. 18, Copenhagen 1928. — MÜLLER-HILLEBRAND: "Siemens Zeitschr." 7, p. 547—551, 605—612, 1927. — E. BECK: "The Electric Journal", p. 591—595, 1928; p. 50—53, 1929.

of very short intervals of time, down to 10^{-10} sec. and even less[1]. Such measurements have been used extensively by the writer[2], by M. Iwatake[3] and others for the determination of time lag in electric sparks. The problem of spark lag and spark formation will, however, be treated of elsewhere.

Fig. 1 shows a sketch of the diagram of connections used for obtaining photographic *L. F.*, compare L. F. I, Fig. 10.

In L. F. I the previous theories of the formation of the L. F. have been mentioned and it is hardly necessary to

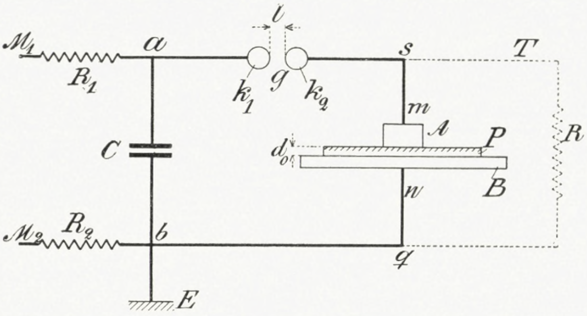


Fig. 1. Diagram of connections for obtaining photographic L. F. M_1 and M_2 are leads from the high tension source. R_1, R_2 and R_3 high resistances (slate pencils or the like). *P* a photographic plate, *B* a metal plate connected to earth *E*.

renew this discussion, especially because no sound theory could be worked out before the velocity of the figures was known.

[1] P. Heymans and N. H. Franck: "Phys. Review" (II). Vol. 25, p. 865—869, 1925.

[2] P. O. Pedersen: (a): Vidensk. Selsk. Math.-fys. Medd. Vol. IV, No. 10, Copenhagen 1922. — (b): l. c. Vol. VI, No. 4, Copenhagen 1924. — (c): "Teknisk Tidskrift (Elektroteknik)", p. 174—184, Stockholm 1923. — (d): Ann. d. Physik (IV). Bd. 71, p. 317—376, 1923. — (e): l. c., Bd. 75, p. 827—847, 1924.

[3] M. Iwataka: "Technology Reports Tôhoku Imp. University" Vol. 7, Nr. 1, p. 57—86, 1927. This paper contains an extensive bibliography on the time lag of electric sparks.

There are, however, one or two exceptions to be mentioned below.

U. Yoshida[1] has given a theory of the formation of the negative figures and of the characteristic dark radii in these figures which is in the main satisfactory. His theory is, with one important exception to be mentioned later, identical with that given by the writer in L. F. I[2]. According to Yoshida the negative figures are due to negative ions which the electric field drives away from the electrode and which cause ionization by collision along their paths. Yoshida does not state the nature of these ions and he does not point out that they must necessarily be electrons, as shown in L. F. I. But this necessity did not exist at the time when Yoshida worked out his theory because the great velocity of the spreading out of the L. F. was not known then.

K. Przibram, who has contributed a long series of important papers[3] on the L. F. has accepted the same view of the formation of the negative figures[4], while M. Toepler[5, 6] has been led to a somewhat different interpretation of the negative figures in his important and long continued investigations of gliding discharges.

All circumstances considered, the main points of the

[1] U. Yoshida: (a): Mem. Kyoto Imp. University, Vol. II, p. 105—116. 1917. — (b): l. c., p. 315—319. 1917.

[2] The writer did not know of the two mentioned papers of Yoshida at the time he wrote L. F. I, which paper was presented to the Royal Danish Soc. of Science on March 8, 1917.

[3] See bibliography in "L. F. II", p. 35.

[4] K. Przibram: (a): Phys. Zeitschr. Bd. 20, p. 299—303. July 1919. — (b): Die elektrischen Figuren in Handb. d. Physik Bd. XIV, p. 391—404, 1927.

[5] M. Toepler: (a): Phys. Zeitschr. Bd. 21, p. 706—711, 1920. — (b): Arch. f. Elektrotechnik Bd. 10, p. 157—185, 1921.

[6] K. Przibram: l. c., (b): p. 403—404.

above mentioned theory of the formation of the negative figures are so well founded that it is hardly necessary to discuss this problem further. In what follows we therefore only treat of the negative figures in so far as it is necessary in order to throw light on the formation of the positive figures.

The theory of the secondary and tertiary figures given in L. F. I has lately been fully corroborated in some interesting experiments of U. Yoshida[1] who has also given experimental proof of some further consequences of that theory. M. Toepler[2] has also investigated these secondary and tertiary figures and his results agree fairly well with those obtained in L. F. I and by Yoshida. We therefore need not further discuss the theory of these figures.

The only outstanding problem is therefore the formation of the positive figures, but this question is, no doubt, the most important and also the most difficult of all the problems connected with the theory of the L. F. In the course of the last 10 years we have made many experiments aiming at the elucidation of the nature of the positive figures. Most of the experimental material — containing among other things some 2500 photographic L. F. — was collected in the years 1918—20. But it was only about a year ago that the writer succeeded in putting forth a hypothesis which made it possible to establish a coherent theory of the formation of the positive figures explaining all their peculiarities in a satisfactory manner.

Before entering upon the detailed discussion of the experimental results and their theoretical explanation, it

[1] U. Yoshida and G. Tanaka: Mem. Kyoto Imp. University. Vol. V, No. 2, 145—152, 1921.

[2] M. Toepler: Phys. Zeitschr. Bd. 22, p. 78—80, 1921.

will be convenient to dwell a little on some preliminary questions, namely concerning the nature of the photographic impressions of L. F. in various gases.

With regard to the terminology adopted in the following we have to lay stress on the fact, that LICHTENBERG figures and discharges only refer to the well known, regular figures

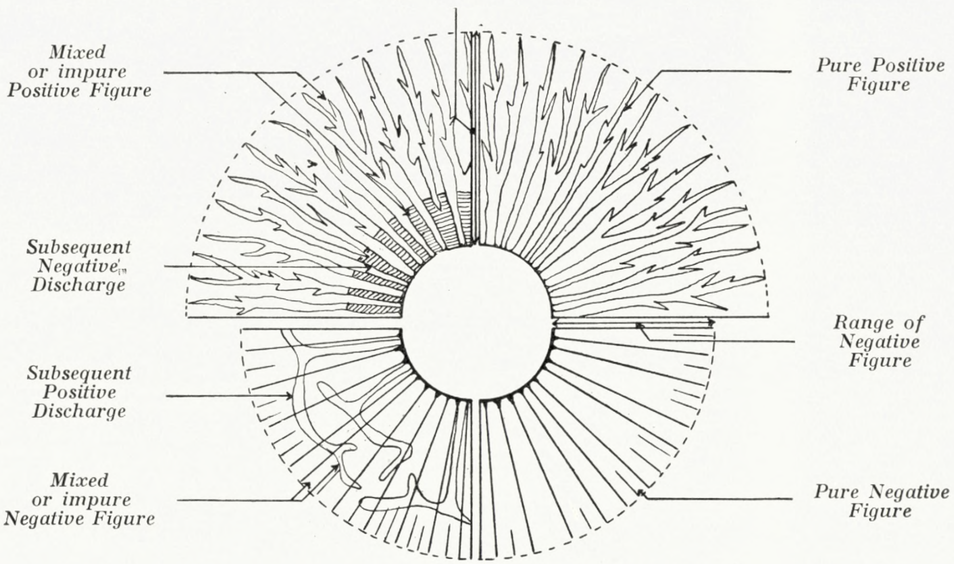


Fig. 2. Regular or simple Lichtenberg Figures. Upper part Positive and lower part Negative Figures. Right hand part pure, left hand part impure or mixed Figures.

showing the characteristic differences between positive and negative discharges and of relatively feeble luminosity but not to the bright sparks or spark tracks which occur if the *p. d.* is sufficiently high and of sufficient duration. Photographs of such spark tracks are to be seen on plate 1, parts VI—VII, and one single track on part III, plates 3 I, 7 II, and 11 I and IV.

The Lichtenberg discharges and figures may start either directly from the electrodes or from the above named bright spark tracks. In the following we will call the first kind regular the latter kind irregular figures and discharges. Samples of the first kind are shown in L. F. I figs. 2, 3, 6, 7, 8, 9, 12, 13, 19, 26, 27, 28 and in this paper on plate 1, parts I and V, plate 2, parts I—III and

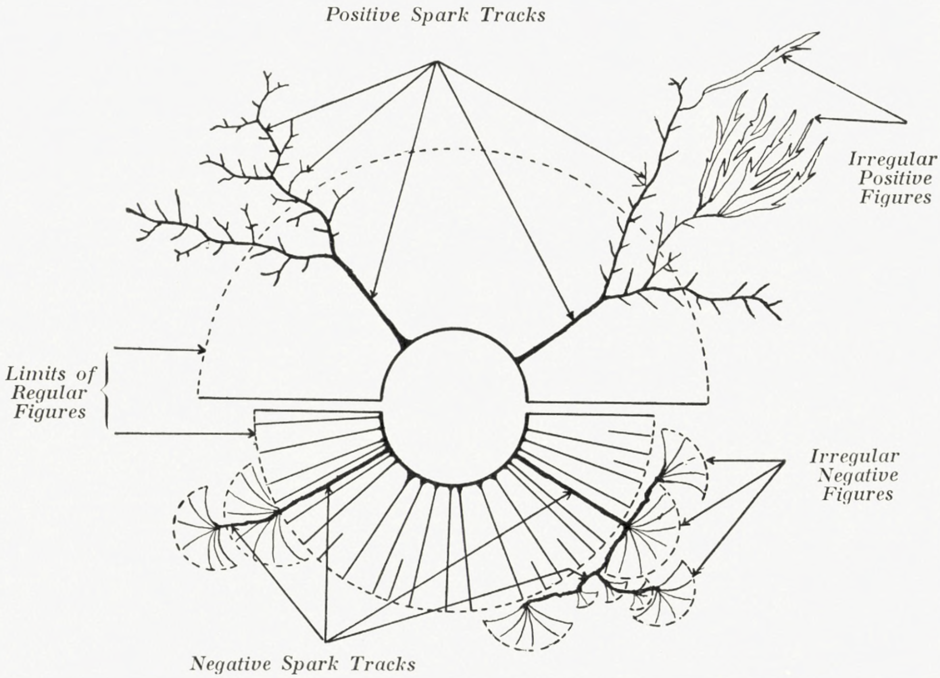


Fig. 3. Composite Lichtenberg Figures.

plate 3, parts I—III. Samples of the latter kind are seen in L. F. I figs. 4 and 5 and in this paper plate 1, parts III, VI and VII and plate 3 part I.

The regular figures will normally always be formed by a potential flash of very short duration. If the *p. d.* is not of very short duration, bright spark tracks will be formed, and from these will again start irregular figures which — especially in case of positive ones — will cover and obscure the regular figure.

A figure consisting of only a regular Lichtenberg figure is often called a simple figure. Such a simple figure is said to be pure if only one single discharge — positive or negative — has taken place; see f. inst. plate 1, parts IV and V, plate 2, part III, plate 3, parts II and III, plate 14, part II and plate 25, part III. If the first discharge has been followed by a second one — generally somewhat weaker — of opposite sign, the figure is said to be a mixed or impure one; see f. inst. plate 1, part II, plate 2, parts I and II, plate 14, part I, plate 15, parts I and II, plate 18, part II and plate 25, part I.

The schematical figures 2 and 3 illustrate this terminology.

CHAPTER II

1. Photographic or "Electrographic" Action?

Are the L. F. due to the photographic effect of the luminosity of the discharge, or to some more direct action of the discharge on the sensitive film — to what may be called an "electrographic" action?

J. Brown[1] considered the luminosity of the discharge to be too feeble to cause the photographic images in the ordinary way and he quoted some experiments in support of this view. It has been proved, however, by U. Yoshida[2] that Brown's experiments are not conclusive.

A. A. Campbell Swinton[3] tried to settle the question by placing discs of non-actinic ruby glass and clear glass on the sensitive film, the small electrode resting on these discs. The ruby glass stopped all action, the clear glass allowed the action to take place all over the plate, though on account of the thickness of the glass, and the consequent intervening distance between the discharge and the film, the details of the figure produced were somewhat blurred and indistinct. Subsequently, using very thin glass, this indistinctness was almost entirely eliminated.

From these results Campbell Swinton draws the con-

[1] J. Brown: Phil. Mag. (5) Vol. 26, p. 503—505, 1888.

[2] U. Yoshida: Mem. Coll. Sc. Kyoto Imp. University Vol. II, No. 2, p. 105—116, 1917.

[3] A. A. Campbell Swinton: "The Electrical Review" Vol. 31, p. 273—275, 1892.

clusion that "the action is due to the ordinary photochemical effect of the light produced by the spark, which though feeble in intensity to the eye, is blue, and must be remarkably actinic".

Even if it is ultimately proved that CAMPBELL SWINTON's opinion is right we cannot consider his experiments conclusive. There is thus no doubt whatever that the light emitted from the spark tracks is strong enough to give the ordinary photographic image, and it is also evident that this light will be cut off by the inactinic ruby glass disc. And with the high potential differences used by BROWN and CAMPBELL SWINTON such spark tracks will occur in every case. The question here discussed only concerns the LICHTENBERG figures proper, and in this case it is much more difficult to attain a decision. With this aim in view we have made a number of experiments of which we shall quote some in the following.

Ebonite discs 0.3 mm thick cut off completely. This is in agreement with CAMPBELL SWINTON's Experiments.

The following experiment will show, however, that the conditions are completely altered with very thin plates. In the figure plate 7, part I the electrode was placed on a mica plate 0.05 mm thick, and the mica covered the photographic film below the line marked *mn*. The mica plate was covered by a dry layer of inactinic red ink. The photographic image shows clearly the Lichtenberg discharges, while the strong light from the spark tracks — as f. inst. *bc, de—f—g—h and hij* — is completely cut off. These spark tracks were certainly on the upper side of the mica plate and the Lichtenberg discharges are seen to start from these tracks. (At a few points, f. inst. that marked *g*, the red ink coating has been defective and an image of a short portion of the spark track is to be seen).

As the strong light from the spark tracks in this case has been unable to penetrate the inactinic coating of the mica plate, it is altogether impossible that the feeble luminosity of the Lichtenberg discharges on the upper side of the mica plate can be the cause of the photographic L. F. below this plate.

The photographic L. F. may in this case be due to: (a) the influence of the strong electric field on the photographic film; or (b) the ordinary photographic effect of the light emitted from points where there is a strong ionization by collision, such ionization taking place on the lower side of the mica plate directly below the Lichtenberg Figures which are formed on the upper side of the mica plate; or (c) possibly a combination of (a) and (b), the sensibility of the photographic film being increased by the strong electric field.

The proposition (a) cannot be true because an electric field does not in itself give any photographic image, but without further evidence it is not possible to choose between (b) and (c). This point is illustrated by plate 7, part I. From an inspection of this figure it appears that the Lichtenberg figures cross the boundary *mn* without any discontinuity[1]. It is also evident that the Lichtenberg discharges in the mica-covered part of the figure have started from certain spark tracks along the upper surface of the mica plate. The photographic L. F. cannot be due to an ordinary discharge between the mica plate and the photographic film, since if this were the case, the spark

[1] This fact is quite evident in the original photographs, somewhat less so in the reproduction.

tracks *abc*, *de—f—g—h* and *hij* would have been very bright in the image, and actually they do not show at all. The photographic L. F. below the mica plate must therefore be due to one of the effects (b) or (c).

Plate 7, part II shows the result of another experiment. The parts *VP 1* and *VP 2* of the photographic film were covered respectively with one and two layers of thin (0.02 mm) violet transparent paper, and other parts, *BP 1* and *BP 2*, respectively with one and two layers of 0.04 mm thick opaque, black paper.

The paper strips were soaked in clear vaseline and pressed against the photographic film, no air being left between the strips and the film. All superfluous vaseline was removed before exposing the film to the discharge, and all strips and vaseline removed before developing the photographic plate. The resulting figure is seen in plate 7, part II.

Below the violet paper both the L. F. and the spark tracks are to be seen, clearest of course with only one layer of the paper. Below the black paper strips, on the contrary, there is no image of either L. F. or spark tracks. (In the case of one layer there are some faint spots of light beneath the spark which has passed over the upper surface of the strip, these spots being no doubt due to small holes in the papers).

This experiment proves that a strong electric field does not give any image in the case where there is no air, and therefore no ionization by collisions at the surface of the photographic film. But the experiment does not absolutely prove that the image of the L. F. is due solely to the light emitted by the Lichtenberg discharge, because the photographic film is subjected to a strong electric field simultaneously with the exposure to the light from the discharge. This question can be settled, however, by means of the experiments illustrated in Fig. 4. Upon the film of the ordinary photographic plate *P* were placed some small pieces of photographic plates with the film downward, either as at P', with a small distance δ between the films, or as at *a*, *b*, *c*, with the two films in direct contact. Even in this case the films only touch each other in a number of points since the films of small broken pieces *a*, *b*, *c* have somewhat projecting edges. In all these cases the Lichtenberg discharges have taken place in the space between the two films. Plate 8, part I shows the result of such an experiment with three pieces: *a*, *b*, *c*, the main plate *P* and the three pieces of plate being developed in exactly the same manner. It appears that there is very little difference between the photographic intensity of the image on the main plate and on the small plates. Even in cases such as P' in Fig. 4, where there is a considerable distance between the two films, the intensity of the image on the upper film may be almost as great as on the main film, see plate 7, part III. Up to $\delta = 1$ to 1.5 mm the image on P' is quite distinct; for greater distances it becomes blurred.

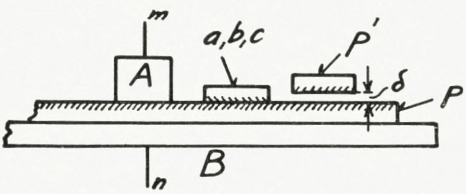


Fig. 4. Ordinary arrangement for obtaining photographic L. F., compare Fig. 1. P' and *a*, *b*, *c* are small pieces of photographic plates with the film downwards.

Since the intensity of the electric field at the film of the pieces P', *a*, *b*, *c*, is very small in comparison with the field at the film of the plate *P*, and since the photographic intensity is almost the same in the two cases, it is proved that the photographic L. F. are due to the ordinary photographic effect of the light emitted by the discharge. But the light may come from discharges in a very thin layer of air between the photographic film and the covering plate, these discharges being either ordinary Lichtenberg discharges —, as in plate 7, part III and plate 8, parts I—II, — or in cases where the covering plates are very thin, very intense ionizations due to very strong fields at right angles to the film, as in plate 7, part I.

This point of view is in accordance with all the previously known facts and with a number of further experiments and observations of which only a few will be mentioned in the following.

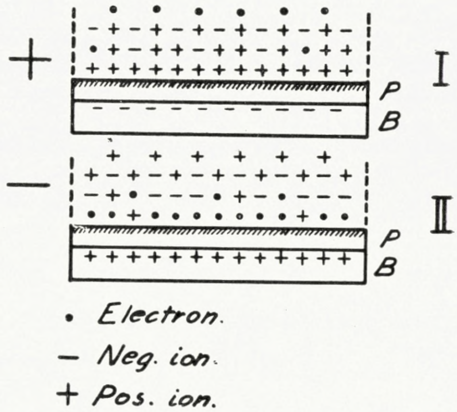


Fig. 5. Schematic representation of the distribution of electrons and positive and negative ions over the cross-section of positive and negative streamers.

With regard to the distribution of the photographic intensity it is to be remembered that the emission of light is mostly caused by the recombination of positive ions with either electrons or negative ions. Fig. 5 gives a schematical sketch of the distribution of electrons and positive and negative ions over the cross-section of positive and negative streamers. But the question of intensity-distribution will be taken up later on in Chap. IV 1 (d).

2. Lichtenberg Figures in Various Gases.

Air, nitrogen and argon give strong photographic L. F. both positive and negative, the last being by far the strongest. In oxygen the luminosity of the Lichtenberg discharges — both the visual and the photographic — is very small. It has not been possible to get positive photographic figures in almost pure oxygen — containing about one per cent of hydrogen —. In oxygen containing small amounts of atmospheric air or nitrogen very feeble positive photographic figures have been obtained, as for inst. plate 1, part II[1]. Part III is the corresponding negative figure.

It may be supposed that the positive discharge in oxygen is very feeble or even that there is no discharge at all in this gas, thus explaining the failure to obtain positive figures. But this supposition would be wrong, for there are strong positive discharges even in the purest oxygen, as may be proved by means of the dust method, using an electrically clean plate of ebonite instead of the photographic plate *P*. A photograph of such a dust figure is shown in plate 1, part I.

The photographic negative figure in oxygen is similar to the negative figure in air, but fainter, see plate 1, part IV. Owing to the faintness there is some difficulty in determining the range of the figures in oxygen. Table 1 contains some approximate values of such ranges in air and in oxygen[2].

[1] This positive figure is impure. The strong light in the neighbourhood of the electrode is due to a negative discharge taking place after the formation of the positive figure, see Chap. III 1 (d) and Chap. IV 1 (d).

[2] K. Przibram. (Wien. Ber. (II a) (a) Bd. 127, p. 395—404, 1918 and (b) Bd. 129, p. 151—160. 1920; (c) Phys. Zeitsch. Bd. 20, p. 299—303. 1919) found that the ratio $\frac{R+}{R-}$ was smaller in oxygen than in air (l. c.) (a) p. 402, (b) p. 151—2) while the writer previously came to the opposite result (L. F. I, p. 35 and 42). The new investigations have fully con-

Table 1. $p = 350$ mm Hg.; $l = 3$ mm; $d_0 = 1.5$ mm.

Air	$R+ = 34$ mm[1]	$R- = 17$ mm	$\frac{R+}{R-} = 2.0$
Oxygen ($98\,O_2 + 2\,H_2$)	$R+ = 35$ mm[1]	$R- = 13$ mm	$\frac{R+}{R-} = 2.7$

The photographic positive figures are also very faint in hydrogen, see f. inst. plate 2, parts I and II. But large and finely branched dust figures may be obtained in this gas. The photographic negative figures are also fainter in hydrogen than in air or nitrogen, see plate 2, part III.

All these circumstances agree well with our previous result, namely that the photographic L. F. are caused by the light emitted by the discharge, the visual luminosity in hydrogen being considerably less than in air but greater than in oxygen.

For general information, some figures in other gases have been included. Thus plate 3, parts II and III show regular positive and negative figures in CO_2[2] while part I shows some positive spark tracks in the same gas[3].

firmed our previous results. The difference between PRZIBRAM's and our results is perhaps due to the circumstance that our measurements refer to simple, regular Lichtenberg figures both in air and in oxygen — compare plate 1, part IV — while the oxygen figures in PRZIBRAM's papers (l. c. (a) Figs. 4 and 5; (c) Figs. 5 and 6) are complicated figures with strong spark tracks similar to the figures in plate 1, parts VI and VII. On the other hand PRZIBRAM's positive nitrogen figure (l. c. (a) Fig. 3) is a regular simple L. F.

[1] Other experiments also indicate that the range $R+$ is almost the same in air and oxygen.

[2] S. MIKOLA (Phys. Zeitsch. Bd. 18, p. 161, 1917) says: "In den anderen untersuchten Gasen (O, H, CO_2 und Leuchtgas) entwickeln sich die Strahlungsfiguren kaum sichtbar". This is strictly speaking only true of the positive figures in oxygen.

[3] The positive figure by PRZIBRAM (Wien. Ber. (II. a). Bd. 108, 1161—1171, 1899) indicates the presence of some air or nitrogen besides the gas CO_2.

We have not succeeded in obtaining photographic L. F. in pure helium. Plate I, part V shows a figure taken in a mixture of air and helium and plate 4, part II a figure from a mixture of air and argon.

The above remarks concerning the intensity of the luminosity refer only to the regular, pure L. F. The light from the spark tracks is very strong in all gases[1].

Plates 4—6 show parts of a number of positive figures in mixtures of N_2, O_2 and H_2 in various ratios. These figures will, however, be discussed later (Chap. IV. 3 (b)).

[1] Compare also Przibram l. c. (a) p. 399.

CHAPTER III

The Properties of the Positive Figures.

1. Differences of Form and other Features between Positive and Negative Figures.

With regard to the main points of these differences we may refer to the L. F. I and for ease of reference we repeat here figs. 6 and 7, which show respectively a positive and a negative photographic L. F. obtained by means of the experimental arrangement shown in Fig. 1.

Besides the very obvious differences between the two kinds of figures, which need no further comment, we shall in the following give some particulars about some specific points.

(a) Width of the Positive and Negative Spreaders.

The width of the negative spreaders varies greatly, the broadest ones having often ten times the width of the narrowest ones, compare f. inst. plate 25, part I and L. F. I Table 3, p. 28.

The positive spreaders on the contrary have in all cases almost the same width at the same distance from the tip, and the width is very nearly inversely proportional to the pressure of the gas if measured at distances from the tip which are also inversely proportional to the pressure. This relation is illustrated by plate 6, part I and by the figures in the following table.

The width of the positive spreaders measured at a certain distance from the tip depends very little upon the voltage or upon the thickness of the insulating plate.

Table 2. Effect of Gas Pressure on the Width of the Positive Spreaders.

Pressure p in mm Hg	Width of Spreaders t in mm	Distance from Tip to Test Point in mm	$p \cdot t$
5×760	0.027	0.10	103
3×760	0.043	0.17	99
2×760	0.07	0.25	108
760[1]	0.12	0.5	91
300[1]	0.29	1.3	87
150[1]	0.59	2.6	89
75[1]	1.42	5.2	106
34[1]	3.60	11.0	119
30	3.60	13.0	108
17	7.0	22	119
		Mean value	103

[1] From Table 7 in L. F. I p. 36.

(b) The Ramification of Positive and Negative Figures.

The negative spreaders show no ramification at all[1]. The positive spreaders ramify extensively and the number of branches per unit length of spreader is very nearly proportional to gas pressure, see Table 6, L. F. I p. 35.

This number of branches is greatest in H_2 and decreases in the following order: A, N_2, Air, CO_2 and O_2,

[1] In very broad negative spreaders the end is often divided by dark radii, see f. inst. plates 21 II, 25 II—III and 27 VI. These dark lines are quite similar to those issuing from the electrode, and the explanation of them offers no further difficulty. U. Yoshida (Mem. Coll. Sc. Kyoto Imp. Univ. Vol. II, No. 2, p. 114—15, 1917) has treated this question in a very thorough and convincing manner.

compare plates 1—6. The number of small branches about the middle of the spreaders is comparatively great in mixtures of H_2 and N_2, see plate 5, XI—XVI. An addition of some O_2 to such mixtures reduces the number of branches very considerably, see f. inst. plate 4 I and 5 I—V and XVII—XVIII.

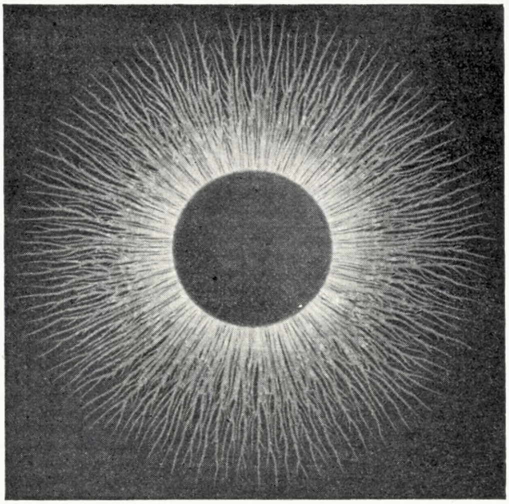


Fig. 6. Impure Positive Figure in Air.

(c) Boundary of the Positive and Negative Spreaders.

The photographic intensity of the negative figures falls off gradually at all points of the boundary but especially so at the outer edge. For the positive figures, on the contrary, the intensity drops down very abruptly to zero all along the boundary line. These properties of the positive and negative figures are illustrated by part II on plate 25.

(d) Distribution of the Photographic Intensity over the Area of the Positive and Negative Spreaders.

The photographic intensity of the negative spreaders has its greatest value at the electrode and on passing outwards it decreases gradually, becoming almost zero at the outer edge, see f. inst. plates I, parts III—IV, 2 III, 3 III, 12 I—IV, 25 I—IV, 26 I—IV, and 27 I—VI.

For pure positive spreaders, however, the photographic intensity has very nearly the same value over the whole area of the spreaders, see f. inst. plates 1 V, 4 I—II, 5 I—XVIII, 6 I, III, 10 I—IV, 12 V—VII, 13 II, 14 I—II, 21 I, 22, 23 and 25 II.

This intensity of the positive spreaders is much less than the maximum intensity of the corresponding negative spreaders.

The innermost parts of the positive spreaders very often, however, show a comparatively very strong intensity which at some distance from the electrode drops down rather abruptly to the normal positive intensity, see plates 2 I—II, 6 II, 12 VIII, 14 I, 15 I—II, 18 I (lowest part), 20 I, III, and 21 I and III. This high intensity is due to a subsequent negative discharge caused by electrical oscillations in the discharge circuit. Such oscillations may also cause a subsequent positive discharge in a previously formed negative figure, see f. inst. plate 18 II (the right hand part) and plate 25 I.

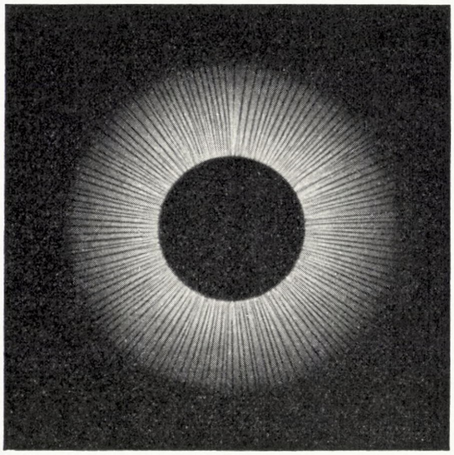


Fig. 7. Pure Negative Figure in Air.

In a positive figure, a subsequent negative discharge will evidently take place mainly out along the positive spreaders, and the resulting high intensity will therefore, be confined mainly to the area of these spreaders. By increasing the damping of the discharge circuit and by making the conditions unfavourable for the formation of negative figures, these subsequent negative discharges may be partly or completely eliminated. In the last case pure positive spreaders and figures are obtained. The damping is increased by reducing the leak resistance *R*, see Fig. 1, or by increasing the area of the positive figure, f. inst. by reducing the air pressure or by increasing — up to a certain point — the thickness of the insulating plate, compare L. F. I Fig. 34, p. 30. The range of the negative figure is, on the contrary reduced by increasing this thickness, as also appears from Fig. 34 in L. F. I.

It thus appears that pure positive figures are most easily obtained at reduced pressures and with great thickness of the insulating plate. These conclusions are in complete agreement with the experimental results.

With regard to the formation of subsequent positive discharges in previously formed negative figures the conditions are quite otherwise. We shall see later that negative discharges can start even with very low voltages — and from sharp points or edges probably down to almost zero voltage — while positive discharges do not start before the voltage has reached a certain minimum value, which decreases with decreasing pressure. The fact that impure negative figures generally only appear at low pressures is in complete agreement with this idea, for only at low pressures will the succeeding positive voltage be high enough to start a discharge.

U. Yoshida[1] considers the innermost bright part of the positive spreaders as due to an ionization caused by positive ions, while the outermost faint part is due to an ionization caused by negative ions, both these ionizations being essential to the formation of the positive discharge. We have seen, however, that the bright part is not a ne-

[1] U. Yoshida: Mem. Coll. Sc. Kyoto Imp. Univ. Vol. II, No. 2, p. 113, 1917.

cessary feature of the positive spreaders but may be eliminated altogether, and we think there can be no doubt that the bright part of the positive spreader here considered is due to a subsequent negative discharge.

On the other hand, if the voltage across the Lichtenberg gap is kept on long enough, both the positive and the negative ions will cause ionization, but the corresponding discharges and their images differ considerably from the Lichtenberg discharges and figures. We shall return to this question later on.

We thus come to the conclusion that the normal simple positive figure is pure and has almost the same photographic intensity over the whole area of the spreaders.

2. Range and Spreading-out-Velocity of the Positive and Negative Figures.

(a) Velocity of Positive and Negative Discharges.

The spreading-out-velocity is considerably greater for the positive than for the negative figures, f. inst. 2 to 4 times as great, see L. F. I p. 60.

(b) Relation between Air Pressure and Velocity.

The velocity, U, increases with decreasing pressure, p, for both positive and negative figures, but the manner in which U depends upon p is otherwise very different for the two kinds of figures, see fig. 8. At low pressures the velocity of the negative figures increases rapidly with decreasing pressure and this rapid increase continues down to a pressure of about 20 mm, which is about the lowest possible pressure at which the velocity can be measured in this manner[1]. For high pressures the velocity of the negative figures decreases slowly with increasing pressure.

The velocity of the positive figures, on the other hand, approaches a certain maximum value U_m with decreasing

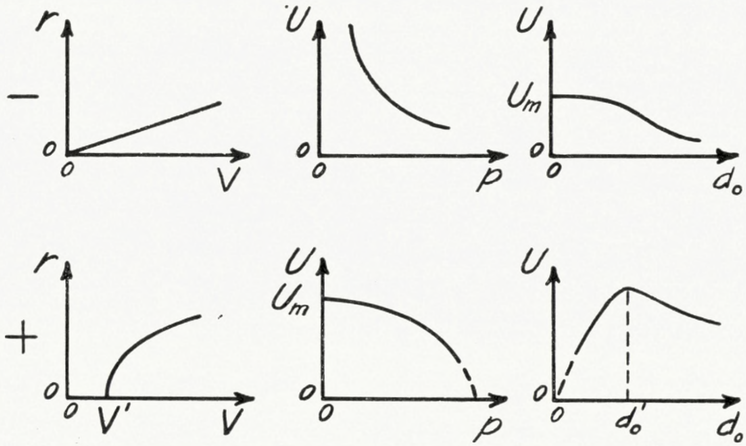


Fig. 8. Schematic representation of (1) the Range r against the Voltage V; (2) the Velocity U against the Pressure p; (3) the Velocity U against the total Thickness d_0 of the insulating Plate.

pressure and drops down to zero if the pressure is increased above a certain critical value, depending mainly on the voltage.

(c) Relation between Thickness of the Insulating Plate and Velocity.

For negative figures the velocity decreases with increasing thickness of the insulating plate; having its greatest value U_m for $d_0 = 0$. For positive figures the velocity seems to be zero for $d_0 = 0$, and for small values of d_0 the velocity increases rapidly with increasing values of d_0. At a certain thickness the velocity attains its highest value and decreases with further increase of d_0.

[1] See figure 24, p. 53.

(d) Relation between Voltage and Range of the Figures.

For negative figures the range, *r*, seems to go down to zero together with the voltage. In the case of positive figures there is a certain minimum voltage below which there is no discharge, and below which the range is accordingly equal to zero.

3. The Starting of the Positive and of the Negative Figures.

(a) Starting of the Negative Discharges.

From a finely pointed electrode placed directly on the photographic plate, even the smallest negative potential seems to start a figure at least if the potential is above some two hundred volts but the figures are very small at low potentials where the radius of the figure is proportional to the applied potential. This is shown in plate 21, part II to which again corresponds the straight line marked (—) in fig. 9.

(b) Starting of the Positive Discharges.

Under similar conditions, positive figures are only started if the potential is above a certain limit, the value of which increases with increasing air pressure. On the other hand, if a positive figure starts at all, it has always a finite and not inconsiderable range. For small potentials there is thus no proportionality between size of figure and potential. This is shown in plate 21, part I which again corresponds to curve (+) fig. 9.

This question is of such importance for the understanding of the formation of the positive figures that a closer consideration was necessary.

The production of a positive discharge — a positive figure — through the influence of a transitory potential depends not only upon potential and the air pressure, but also upon a series of other conditions, although the two first named factors are the most influential. We shall therefore first consider the influence of these two.

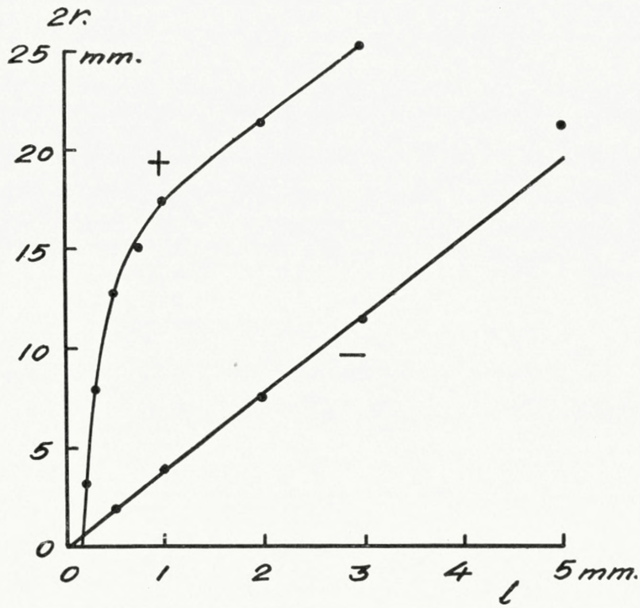


Fig. 9. Range r of positive (+) and negative (—) Figures against the spark length l.

Plate 21, part I shows how the size of the positive figures depends upon the potential at atmospheric pressure, but the conditions near the limit where figures may or may not appear are more favourably elucidated at lower pressures.

For general information plate 22 shows a number of figures produced at constant pressure ($p = 100$ mm Hg) but applying different potentials.

Where potentials here and in what follows are stated in volts, they are generated by a high-voltage D. C. dynamo, and the spark gap g in fig. 1 is replaced by a discharge key, the design of which we shall come back to later. The potential stated is the voltage across the discharge key before this is closed. Different forms of electrodes have been used; the figures reproduced on plate 22 and 23 were taken with a sharp-edged 3 mm cylindrical brass rod resting directly on the photographic plate.

At 100 mm pressure figures only appeared at potentials above 1160 volts. $V = 1366$ produces a pronounced figure having a radius of about 15 mm, as is also the case with the voltages 1382, 1400, 1410, 1430, 1446, 1478, 1536, 1541, 1581 and 1593 volts. At 1556 volts no discharge appeared at all, while all potentials above 1600 volts produced a figure. An inspection of the pictures on plate 22 shows that all the branches are of nearly equal length at potentials from 1366 to 1593 volts, and that the number of branches increases, although somewhat irregularly, with increasing potential. At 1556 volts the number of branches dropped to zero, no discharge appearing. This indicates a certain irregularity in regard to number of branches, whereas their range is nearly independent of the potential value within the critical interval.

The figures on plate 23 correspond to the constant potential 1190 volts, while the air pressure is varied from 40 to 101 mm Hg. At $p < 80$ mm the discharge appears as a uniform disc near the electrode with a number of teeth or branches stretching outward from the disc edge. At $p > 80$ the branches start directly from the electrode and they are of practically equal length at pressures from 80 to 99 mm, while their number decreases, though somewhat irregularly, with increasing pressure. At $p > 100$ the number of branches is zero, i. e. no discharge occurs at all.

Those series of figures presented on plates 22 and 23 are not especially selected or arranged but include all records taken in these particular series, which series again have been selected arbitrarily from a greater number.

To elucidate these conditions further we have in figs. 10—12 presented graphically the range *r* and the number of branches *n* as functions of either potential or air pressure. These curves also confirm the fact that the discharge — i. e. formation of the image — does not fail to appear because the length of the branches decreases toward zero but, on the contrary, because their number becomes zero. They also confirm the above mentioned irregularity in the dependency of *n* on pressure or potential.

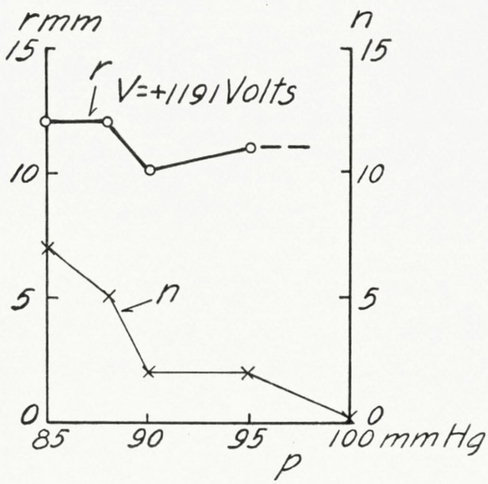


Fig. 10. Range *r* and number of spreaders *n* against the pressure *p*. ($V = +1191$ Volts).

The electrode used in the tests recorded in figs. 10—12 was a 3 mm round brass-rod the end of which was a plane surface, but nothing was especially done in order to keep the electrode sharp-edged. On the other hand a series of investigations have been carried out in order to ascertain how the shape and the state of the electrode may influence the formation of figures within the critical interval. The results of some of these investigations are shown in figs. 13—16. Four different forms of electrodes have mainly been tried: a 6 mm brass ball has been used for some of the tests in all of the figures; a rounded 3 mm brass rod is used in some of the tests in fig. 13; a sharp-edged electrode was used in some of the tests in figs. 13—15[1]; an electrode consisting of a

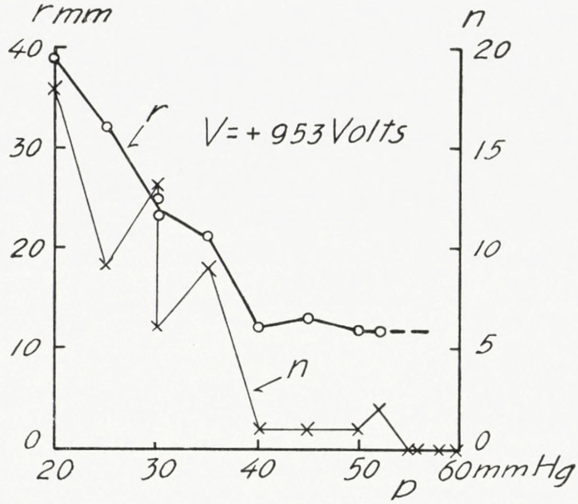


Fig. 11. Range r and number of spreaders n against the pressure p. ($V = +933$ Volts).

brass tube with very thin walls — 0.15 mm thick — which was kept very sharp edged by grinding, was used in some of the tests in fig. 16; finally we have for some tests in

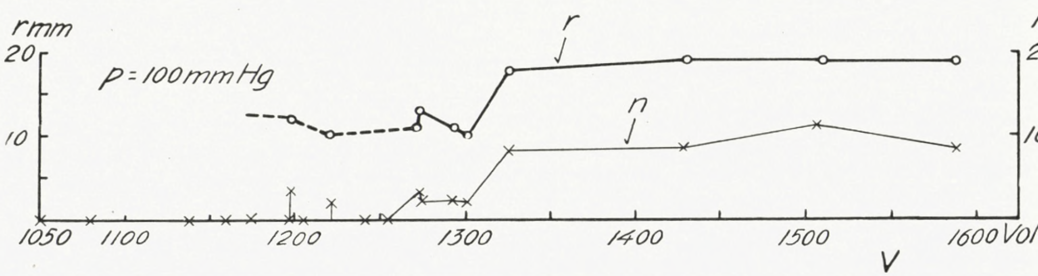


Fig. 12. Range r and number of spreaders n as functions of the p. d. V. ($p = 100$ mm Hg).

fig. 16 used a pointed electrode kept carefully pointed by repeated grinding. If nothing is stated to the contrary all

[1] This shape of electrode will gradually lose its sharp edge by the repeated cleaning process. For this reason we adopted the next electrode shape, a thin-walled tube.

the electrodes were kept "clean" by rubbing with carborundum[1]. The shape of the electrodes is stated in each figure. In some of them is indicated, in the same manner as in fig. 13, whether the figures consist of a uniform disc near the electrode. Finally in fig. 15, for some cases, the number of spreaders is indicated by figures. In fig. 16 we have only stated the number of spreaders n but not their length r, the values of which are given in figs. 13—15.

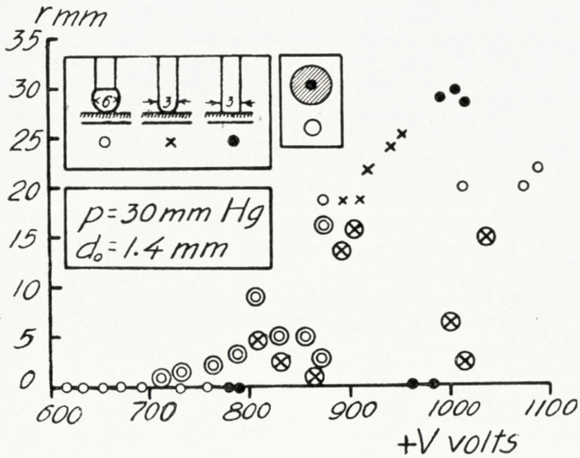


Fig. 13. Range r as function of the $p.d.$ V for three different shapes of the electrode. The large circles indicate that the figure has a continuous disc surrounding the electrode. (p = 30 mm Hg; d_0 = 1.4 mm).

Fig. 13, which refers to the conditions at 30 mm pressure, indicates that a figure is formed somewhat more easily from a spherical electrode than from a rounded or a sharp-edged rod, although the difference between the two first mentioned is not very pronounced. These tests also prove that with sharp-edged electrodes there are formed either spreaders of considerable length or no spreaders at all. With the spherical electrode and with the rounded rod the conditions change more gradually, since with these elec-

[1] See P. O. Pedersen: l. c., (a) p. 25, (d) p. 336.

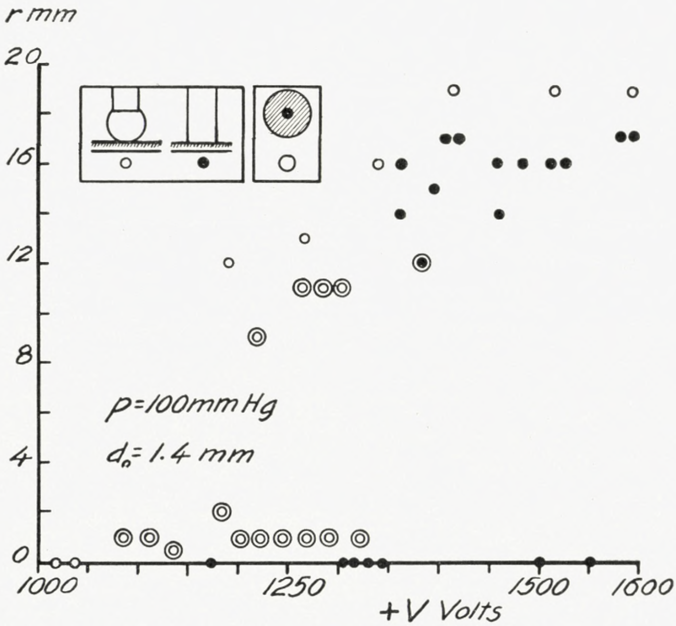


Fig. 14. Relation between Range r and Voltage V.

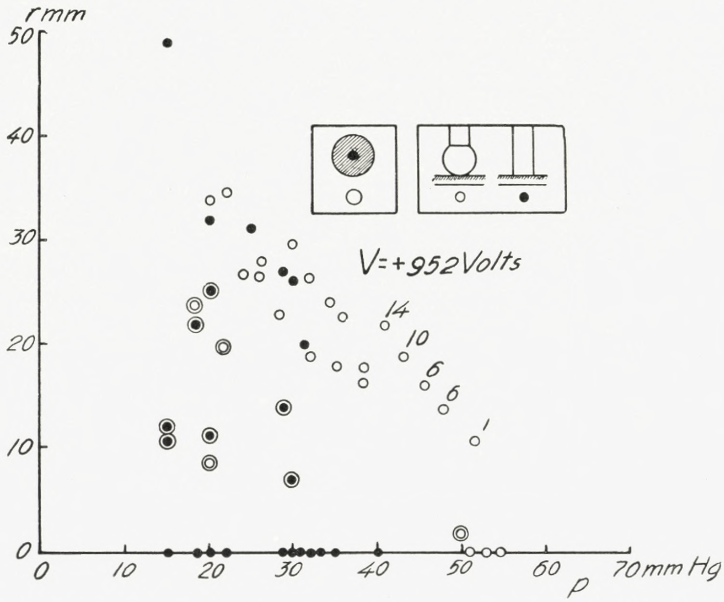


Fig. 15. Range r as function of the pressure p. ($V = + 952$ volts). The figures at some of the representative points indicate the number of spreaders.

trodes figures of small range may be obtained within the critical interval. We will come back to this question later on.

Fig. 14, referring to tests at 100 mm pressure applying different potentials, and fig. 15, where the potential is kept constant at 952 volts but the air pressure is varied, show exactly similar relations with regard to the spherical elec-

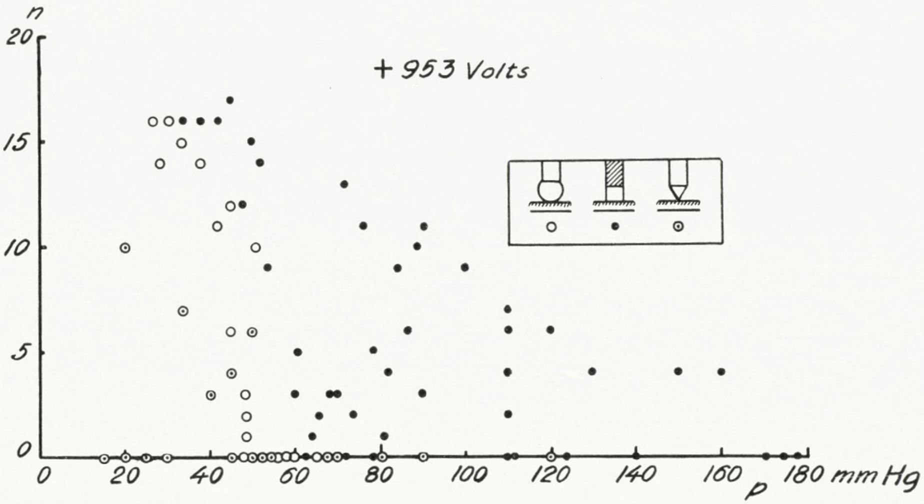


Fig. 16. Number of spreaders n as function of the pressure p. ($V = + 953$ volts).

trode and the sharp-edged rod. Finally, in fig. 16 is shown the number of spreaders (n) formed with spherical, tubular and pointed electrodes, for $V = 953$ volts and varying air pressures. It appears from this figure that figures are most easily formed from the tube electrode, less easily from the spherical one and least easily from the pointed electrode. For negative figures the reverse is the case; they are formed by far more easily from a pointed than from a spherical electrode.

The small range sometimes attained by the figures within the critical interval with the spherical electrode and with the rounded rod may possibly be due to the vertical electric field which forces the discharges, started beneath these electrodes, down towards the film of the photographic plate, thus preventing them from spreading out in the normal way. The considerable decrease in range of the positive figures when the plate thickness d_0 tends to zero

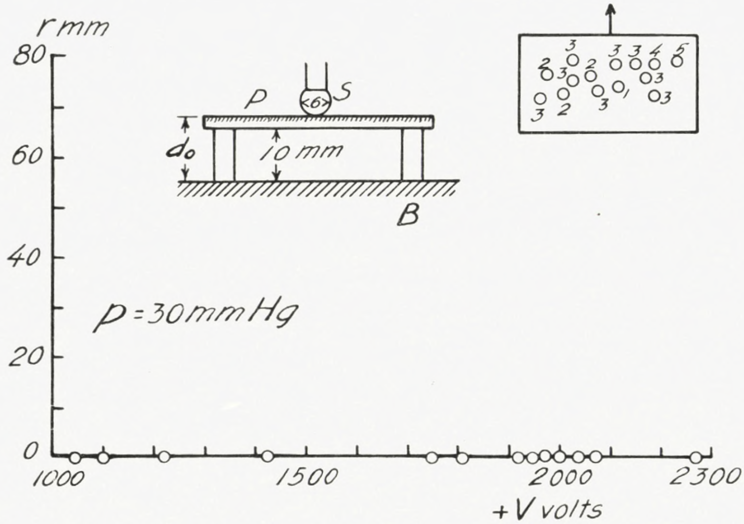


Fig. 17. Range r of positive figures from a spherical electrode. Thickness d_0 of insulating medium 11.4 mm. Experimental points marked by small circles. The figures at the experimental points denote the number of spreaders.

(see L. F. I fig. 34) suggests such an explanation. To elucidate this behaviour we have among others made the test shown in fig. 17.

The spherical electrode rests as usual directly on the film of the photographic plate P which, however, does not rest directly on the earthed plate B, but is raised 10 mm by means of the ebonite blocks shown. In this case the vertical field beneath the electrode will not be very strong, and, as expected, we find that under these conditions the spherical electrode gives either no discharge at all or forms comparatively very long spreaders as with the sharp-edged electrodes[1].

We have also investigated whether the design of the discharge key may influence the starting of the figures. Among others we have tried ordinary discharge keys and also mercury vacuum keys which in other respects show very special behaviour[2]. We have, however, not been able to ascertain any difference in the effect of the various keys.

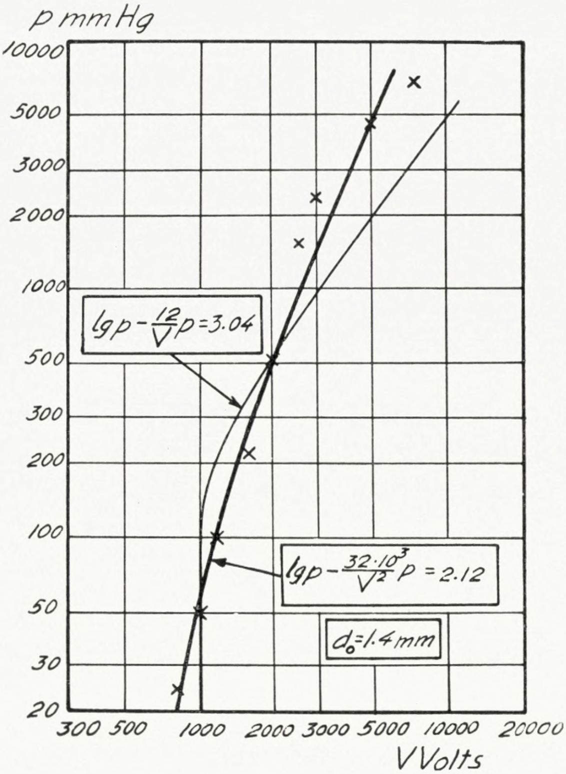


Fig. 18. Corresponding critical values of pressures p and $p. d. V$.

From the foregoing it appears that it is not possible to state with any great certainty the maximum air pressure at which a given potential may start a discharge, or conversely, at what minimum potential a discharge may be started at a given air pressure. The uncertain or fortuitous nature of the formation of figures

[1] The ranges represented by the points in the square frame are really greater than stated but the full range could not be determined as the figures — when discharges were started at all — passed beyond the edges of the photographic plates.

[2] P. O. Pedersen: Videnskabernes Selskabs Math.-fys. Medd. IV, Nr. 5, 1922. — Proc. Inst. Radio Eng. Vol. 13, p. 215—243, 1925.

within the critical interval prevents a true determination of a (p, V)-curve for this critical transitory interval. It is possible, however, on account of the large amount of experimental material available, to fix fairly correct corresponding values of p and V within the critical interval. Such sets of values are marked with crosses in fig. 18 for values of p from 22 to 7000 mm and of V from 700 to 7000 volts. We shall subsequently take up a discussion of this figure.

So far we have only considered "clean" electrodes but we have found that within the critical interval the formation of positive figures is independent of whether the electrodes are clean or slightly greased (unclean). Corresponding tests with negative figures show that these are formed much more easily from "clean" than from unclean (slightly greased) electrodes.

4. The Conductivity and the Spreading-out-Conditions of the Positive and Negative Figures.

There is a very pronounced difference in the manner in which the positive and the negative spreaders conduct electricity and in their ability to promote and initiate spark formation.

To elucidate this behaviour we have made the experiments described in the following.

(a) Conductivity of Positive and Negative Spreaders.

As soon as the edge of a negative figure reaches an outer electrode, a spark will pass between the two electrodes even where the outer electrode comes only very slightly inside the final range of the figure; see for example plate 26, parts I—IV. The spark discharge brings the two electrodes to a nearly equal potential.

In the case of positive figures, spark formation occurs only when the outer electrode comes far inside the range of the figure; see for instance plates 14 II and 15 I—II and also "L. F. I" figs. 45—50, L. F. II fig. 18 and "A. d. Ph." fig. 18. These relations were previously observed by U. Yoshida[1] and are also mentioned in "L. F. II" p. 22 and "A. d. Ph." p. 220.

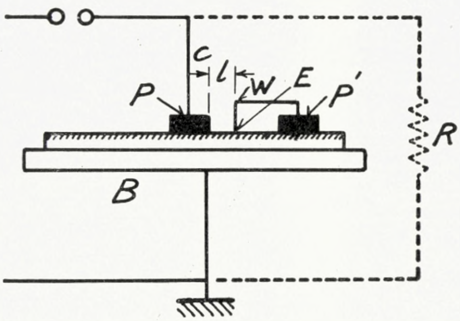


Fig. 19. Diagramatical representation of the electrical connections used in the experiments referred to in fig. 20 and in plate 12, parts I—VII.

For the further investigation of these relations, we have made among others the following experiments: Besides the usual electrode P, see fig. 19, a free electrode P' was placed on the film of the photographic plate. From P' a metal wire W is run to a point E of the film distant l from P. An account of the experimental results is given in fig. 20, parts III—V, and here are also shown the shapes of the electrodes employed, parts I—II. In part III the abscissa shows the ratio $\frac{l}{R}$, where l is the distance from the end E of the wire W to the electrode P, and R is the length of the spreaders from this electrode. The ordinate shows the ratio $\frac{R'}{R}$ of the spreaders from P and P' and for the positive figures also the ratio $\frac{n'}{n}$ of the number of spreaders from P' and P respectively.

[1] U. Yoshida: Mem. Coll. Sc. Kyoto Imp. Univ. Vol. II, No. 2, p. 115—116. 1917.

From parts III—V it appears that the conditions are essentially different for positive and negative figures with

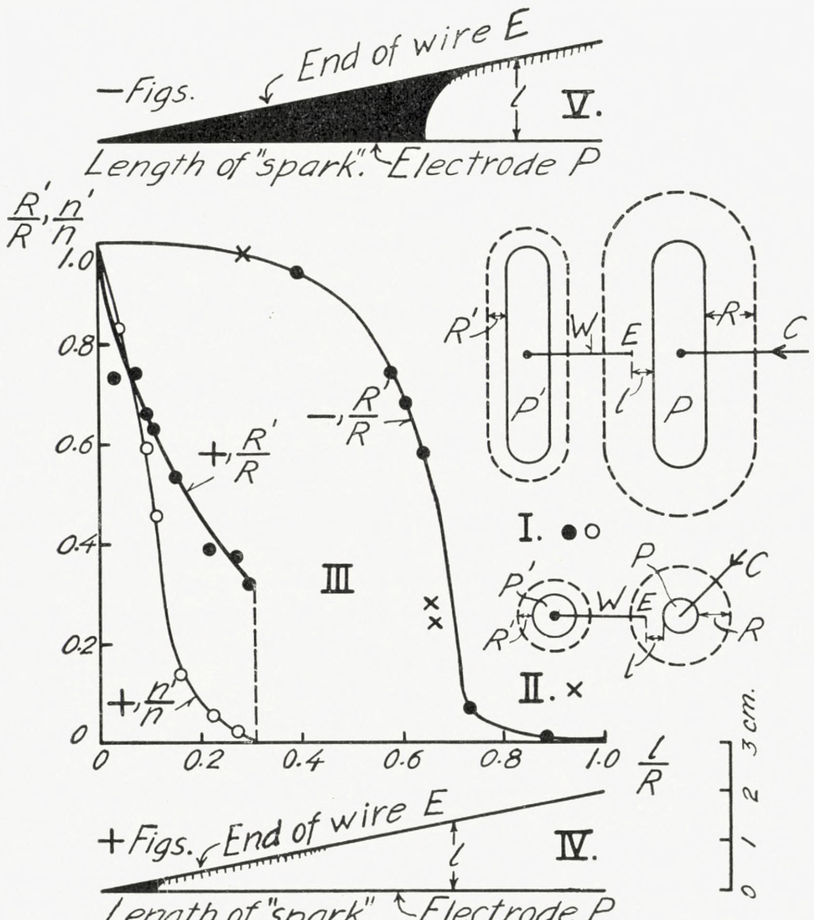


Fig. 20. Conductivity and sparking within positive and negative figures. In parts IV and V the black portions indicate that sparking has taken place. Air; $p = 760$ mm. Hg.

regard to conductivity and spark formation. With negative figures, pronounced spark formation occurs between the end E of the wire W and the electrode P when $\frac{l}{R} < 0.65$, and the ratio $\frac{R'}{R} > 0.9$ as long as $\frac{l}{R} < 0.5$. Not until

$\frac{l}{R} > 0.7$ does R' attain very small values and still at $\frac{l}{R} = 0.8$ to 0.9, R' has not decreased quite to zero. For positive figures conditions are entirely different. Here the formation of figures from P' stops completely when $\frac{l}{R} \cong 0.3$, and sparks are only formed when $\frac{l}{R} < 0.12$. Further the value of $\frac{R'}{R}$ decreases quite rapidly with increasing values of $\frac{l}{R}$ so that $\frac{R'}{R}$ has gone down to about 0.3 at $\frac{l}{R} = 0.3$ while for negative figures $\frac{R'}{R}$ remains nearly constant ($\cong 1$) up to $\frac{l}{R} = 0.4$.

In plate 12, parts I—VII, are reproduced some figures taken with the electrode arrangement shown in fig. 20, part I, though they only show the conditions existing near the end E of the wire W. (In these pictures the electrode P is placed to the left of E). From these pictures it appears that the spark formation is started at E, see for example plate 12, parts I and V, and 13 II. This relation is indicated schematically in fig. 20, parts IV and V.

From fig. 20, part III, it thus appears that "negative" sparks formed in this manner have a comparatively high conductivity, since the subsidiary figure starting from P' has nearly the same range as the one from P, as long as $\frac{l}{R} < 0.5$. Contrary to this the conductivity of the "positive" sparks is comparatively small, since both the length and the number of spreaders decreases very rapidly with increasing values of $\frac{l}{R}$ even at very small values of this quantity. A comparison of parts III and IV, fig. 20, on the other hand, shows that positive spreaders may allow a certain passage of electricity without the formation of a spark, since under these particular circumstances positive sparks are only formed when $\frac{l}{R} < 0.12$ while the subsidiary figure does not fail to appear until $\frac{l}{R} > 0.3$.

Fig. 20 at all events shows that both conductivity as well as tendency to spark formation is much smaller for positive than for negative figures.

(b) Influence of the Duration of the Pulse upon the Positive and the Negative Spreaders.

A corresponding difference of very pronounced character is found in another connection, namely the manner in

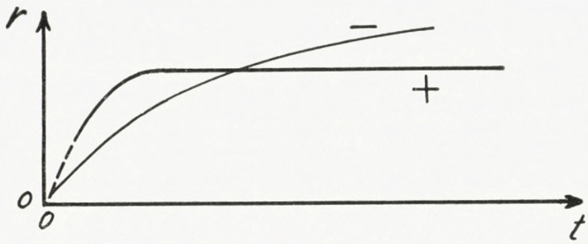


Fig. 21. Schematical representation of the range r of positive and negative figures as functions of the duration t of the $p.$ $d.$

which the positive and the negative figures behave when exposed to potentials of comparatively long duration. This may be varied f. inst. by varying the length L_0 of the lead to the electrode from which the figure starts, L_0 being reckoned from the condenser C fig. 1. The longer L_0 is, the longer will the electrode be subjected to the potential. This point has already been investigated in "L. F. I" p. 32 (figs. 36 and 37). It is found that the radius of the negative figures increases with increasing length of the wire L_0 — at all events for lengths of wire up to about 25 m — as shown in fig. 21. This ability of the negative figures to grow larger is also illustrated in plate 28, part II. The needle-shaped Lichtenberg electrode in the diagram (part III) denoted by K, is first exposed to the potential corresponding to the spark length L of the spark gap shown. A surge travels out along the open line 8 m long and the $p.\ d.$ is nearly doubled by reflection at the end. The Lichtenberg gap K is consequently exposed to a $p.\ d.$ corresponding to the spark length L during $\frac{16}{3 \cdot 10^8} = 5.3 \cdot 10^{-8}$ sec. after which the $p.\ d.$ momentarily increases to about the double value. In the negative figure, the growth of which had

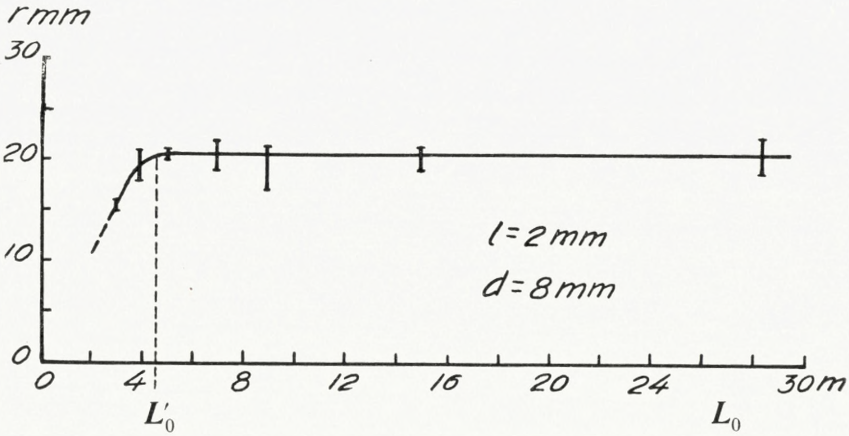


Fig. 22. Range r of positive figures as a function of the length L_0 of the connecting wire. (All experimental determinations of r fall within the heavy vertical lines shown).

practically ceased after the lapse of $5.3 \cdot 10^{-8}$ seconds at the lower $p.\ d.$, this is evinced by a continued growth under the influence of the higher $p.\ d.$ A darker ring in the figure marks clearly the boundary between the original and the subsequently formed part of the figure. It is clearly seen that the increased growth occurs exclusively as a continuation of the original spreaders.

We shall see later that positive figures formed under similar conditions behave entirely differently; they have already attained their full range at $L_0' = 4.5$ m, as appears from fig. 22, which shows the results of experiments carried out recently.

If we assume the *p. d.* in the wave front to increase gradually from zero to its maximum value V_0 over the length L' of the wave front, then the time τ during which the Lichtenberg gap is exposed to the maximum *p. d.* $2\,V_0$ will be determined by

$$\tau = \frac{2\,L'_0 - L'}{3 \cdot 10^{10}} \text{ sec.,} \qquad (1)$$

L'_0 and L' being measured in cm and assuming the capacity of the Lichtenberg gap to be negligible.

If we assume the range of the positive figures to depend only on the value of the maximum *p. d.* and to be entirely independent of its duration then, according to (1), the length of the wave front must be determined by $L' = 2\,L'_0 = 9$ m.

If, on the contrary, the wave front were perfectly steep i. e. $L' = 0$, then a value of $\tau = 3 \cdot 10^{-8}$ sec. would correspond to $L'_0 = 4.5$ m.

Actually L' must have a finite length and the duration τ_0 of the maximum *p. d.* necessary for the positive figures to attain their full range must thus have its value between

$$0 < \tau_0 < 3 \cdot 10^{-8} \text{ sec.} \qquad (2)$$

If the value of L' is known, then τ_0 may be determined by means of (1), but this question we will return to elsewhere.

(c) Irregular Figures caused by Pulses of long Duration.

Beside the regular, simple figures mentioned, there will, however, be formed other peculiar figures if the *p. d.* is maintained for a sufficiently long time. In this respect too positive and negative figures show great differences.

In the negative figures the discharge tracks formed when the *p. d.* is maintained long enough will follow the already formed negative spreaders. In no case have we observed such discharge tracks in the spaces between the original spreaders. Plate 27 shows some enlarged reproductions of negative discharge tracks; they mainly follow the centre line of the spreader concerned, but many show some smaller irregular bends. In some cases they may jump from one spreader to another. Such a case is seen in part VI. Here the spreader in which the discharge track starts is very short and separated from its neighbours by narrow and little ionized spaces. If the discharge track followed this spreader, it would find areas having little pre-ionization, and it is therefore found easier to jump to one of the neighbouring spreaders and then to continue along this. It is such jumps which form the sharp bends in the negative spark tracks, see f. inst. plate 26, part I and L. F. I fig. 5.

These discharge tracks will develop into sparks if the *p. d.* is maintained somewhat longer, but if a spark has been formed, the light from it will generally blur the figure and make it impossible to discern the details of the L. F. which existed before the formation of the spark.

From plate 27 it appears that the negative discharge tracks are very narrow.

With positive figures the conditions are entirely different. If the positive *p. d.* is maintained for some time, a new figure is often formed with its spreaders — trunks as well as branches — fitting themselves in proper order into the spaces between the trunks and branches of the first figure.

The new figure, figure no. 2, may, according to circumstances, pass beyond the boundary of figure 1, but it may also stop earlier.

An example is shown in plate 28, part I. This figure is taken under conditions quite analogous to those obtaining for the formerly mentioned negative figure shown in part II. Under the influence of the *d. p.* corresponding to the spark length L the figure has reached its full range (corresponding to this *p. d.*), within the $5.3 \cdot 10^{-8}$ sec. during which this potential was maintained. The outer boundary of the corresponding regular figure is marked by the drawn circle. The maintaining of the *p. d.* during the time mentioned, and its further increase for a short space of time has caused both some special discharges around and near the electrode, and also the starting of a number of new spreaders, which have had to "squeeze" themselves out between the original spreaders as is clearly recognised in the figure. Many of these subsequently formed spreaders have a greater range than those first formed.

The particular discharge phenomena which take place directly at the electrode will be treated of later.

At low pressures figure number 1 will often cover practically the whole surface, particularly after it has passed a little away from the electrode. In that case figure number 2 is not easily formed. Under especially favourable conditions, figure number 2 may be formed even down to a pressure as low as 150 mm. An example is shown in plate 6, part III. This figure shows clearly 2 sets of spreaders: a set of shorter ones which was formed first and a set of longer ones formed afterwards. That the order of formation is as stated may be inferred with certainty from the form of the spreaders, since the trunks and the branches of figure no. 2 "squeeze" themselves into the spaces between the trunks and branches of figure no. 1.

From the tracks followed by the trunks and branches of figure no. 2, it appears clearly that the trunks and branches of figure no. 1 were present with a sharply localized positive charge at the moment when figure no. 2 was formed. On the other hand, the spreaders of figure no. 1 cannot have possessed ability to conduct to any greater extent at the moment in question, since, if they had, they would simply have continued to grow in length because the field would then be strongest at their tips. The new spreaders start directly from the electrode in the spaces between the former ones, although the electrical conditions here are very unfavourable for a start because the existing charge on the first spreaders reduces the field at the starting points.

At atmospheric pressure this phenomenon is produced very easily and examples of such are given in plate 10, part III, plates 16 and 17. In these figures $a, b, c \ldots$ mark spreaders of figure no. 1 while $a', b', c' \ldots$ refer to figure no. 2, a'', b'', c'' to figure no. 3 and so on. On plate 17 4 successive discharges are clearly seen.

Beside this number of successive Lichtenberg discharges, another kind of positive discharge, essentially different from the Lichtenberg ones, will occur if the positive potential is maintained for a sufficient length of time. For instance, it has not their regular and sharply defined forms but is of a blurred character, while it has a higher brilliancy and has no well defined range, but spreads out further the longer the potential is maintained. Such discharges are seen near the electrode on plate 9, part I (especially at the electrode A_2), plate 16, parts II and IV, plate 17, plate 18, part II, and plate 20 II.

This positive discharge, (which is of a kind quite different to the Lichtenberg discharges, and will be treated of elsewhere in connexion with the question of spark formation) develops into highly luminous positive spark tracks if the *p. d.* is maintained long enough, and such sparks are shown schematically fig. 3 and examples of actual figures are shown plate 3, part I, plate 7, plate 11 I and IV and L. F. I fig. 4.

This discharge form must not be confounded with the previously mentioned flow of negative electricity from the electrode out along the positive spreaders, which often occurs when the discharge is so slightly damped that oscillations take place in the discharge circuit; see above under 1 (d).

Comparison shows that the character of this subsequent negative discharge is entirely different from the irregular positive discharge just mentioned: The negative flow occurs only out along the already positively charged spreaders, while the spaces between them remain untouched. This is not so in the case of the positive discharge. The boundary of the subsequent negative discharge is fairly sharp; this, also, is not the case with the positive one. The irregular positive discharge is often the starting base for the Lichtenberg discharges no. 2, 3 ....

Very often subsequent negative discharges and irregular positive discharges appear in the same figure, giving the innermost part of the figure a highly luminous and irregular appearance; see plate 9, part I (especially electrode A_2), plate 16 IV and 20 II.

It is of importance in this connexion again to emphasize that neither the subsequent negative nor the irregular positive discharges are necessary parts of the positive Lichtenberg figures; they are only complications which have nothing to do with the regular L. F., and which may easily be avoided by suitable arrangements, see thus plate 1, part V, plates 4, 5, 14 II, 22 and 23.

U. YOSHIDA[1] seems to consider both the subsequent negative discharge and the irregular and comparatively slow positive discharge as a normal and essential part of the positive Lichtenberg figures, in that he assumes the manner of formation to be somewhat different for the inner and for the outer part of the figure. Thus he says in his paper (b) p. 315: "In the anode figures (Fig. 3 and 13 of the former paper) we notice that every branch consists of two parts; namely a more intense portion near the electrode, and a weaker portion more removed. When a celluloid film is used instead of a common photographic plate, these two portions are more clearly distinguishable as seen in fig. 18 of the former paper, and fig. 1 of this paper, the ends of the intense portions of the branches terminating with a continuous outline".

That the subsequent negative discharge should in the latter case be predominant over the positive one is fully in agreement with the explanation we have given of the phenomenon in section 1 (d) above.

It is emphasized above that we cannot agree with U. YOSHIDA on this point.

(d) Resumé.

The results set forth above may be briefly summarized as follows:

[1] U. YOSHIDA: (a) Mem. Coll. Sc. Kyoto Imp. Univ. Vol. II, No. 2, p. 115—116, 1917. — (b) l. c. No. 6, p. 315—319, 1917.

For negative figures the range increases with increasing duration of the *p. d.* and, when the *p. d.* is maintained for a sufficient length of time, spark tracks will be formed along and inside the original spreaders — but never in the spaces between them. These spark tracks begin as fine threads with a somewhat irregular course.

For the positive figures the range of the first formed figure is independent of the duration of the *p. d.* — at all events when its value exceeds $3 \cdot 10^{-8}$ sec. If the *p. d.* is maintained long enough, there may be formed a new positive figure having a range greater or smaller than that of the first one, and in which the spreaders of the new figure fit themselve into the spaces between the spreaders of the first figure.

If the *p. d.* is maintained still longer, irregular positive spreaders, of a kind entirely different to Lichtenberg ones, may be formed. From these irregular positive spreaders there may be started new Lichtenberg spreaders which then fit themselves into the spaces between those already present.

5. Various Questions relating to the Formation of the Lichtenberg Figures, especially the Positive ones.

(a) Influence of Initial Ionization.

In order to solve the question of how positive figures are formed, it is most essential to know what influence an ionization within the area to be covered by the figure may have upon the formation and appearance of the figure. To show the importance of this question we may cite the following extract from U. Yoshida[1]: "If the potential of the anode is increased sufficiently, the negative ions, which were present just before the formation of the photographic impression by the ionisation commences, will be pushed toward the portions of the electrode where the electric force is strong, and will ionise the gas molecules with which they collide. The negative ions thus produced will also do the same thing; and many positive and negative ions will be produced. The group of positive ions, which will be left behind as negative ions are pushed toward the anode, will now act as a portion of the anode; and, repeating the same process as before, further branching and elongation of the anode branch will take place. With this explanation, the properties of the anode branches (that they are irregular in their branchings and elongations, and that their branches end in sharp points) will be immediately understood; because the formation of these branches is due to the presence of negative ions which would be distributed irregularly".

We have therefore made this question the subject of a fairly thorough investigation, the result of which was that in no case were we able to ascertain any influence of an existing ionization on the formation of the regular positive (or negative) figures — setting aside such unimportant cases in which the conductivity became so large that the electric fields, determining the course of formation, were liable to considerable distortion.

In the following we shall describe some of these investigations. Plate 16, I shows the result of such a test. Immediately prior to (10^{-6} sec. before) the formation of

[1] U. Yoshida: l. c. (a), p. 113.

the positive figure shown, the spot S was illuminated by a strong spark. By this means a strong ionization is caused photo-electrically within the area of the spot, which ionization cannot have vanished entirely before the figure was formed, and consequently the ionization within the spot must have been very much stronger than outside. In spite of this no influence of the ionization on the course or the appearance of the spreaders is to be observed.

Against the validity of this test it may possibly be said — although in our opinion unjustly — that the ionization is located at a place in the middle of the positive spreaders where the spreading out of these is going on with great strength and speed while the influence of such an ionization may be expected to be particularly obvious at the ends of the positive spreaders. Plate 10 III therefore shows the result of another test where the ionized spot A is located at the tips of the spreaders e' and f', but the presence of the ionized spot had also in this case no influence upon the form and course of the spreaders.

We have further investigated the formation of figures over an area where a strong ionization takes place simultaneously with formation of the figure. The results of a couple of these tests are shown in plate 10 I and II. Here the ionization was effected by means of a powerful Radium preparation resting a few millimeters above the photographic plate during the formation of the figure. The action of the radio-active rays was in the main confined to a limited area by means of shielding. The pictures show the ionized portions to be rather intensely luminous but it is also clearly seen that there is no difference in the formation of the positive figure inside and outside the ionized part of the plate.

An ionization by collision of the character mentioned by U. Yoshida, but in which the active negative ions are no doubt mainly electrons, may be brought about in various ways of which a few will be mentioned in the following.

With the arrangement shown in fig. 23 a series of pictures has been taken the general character of which appears from plate 6 II. The distance between the electrodes A_1 and A_2 is chosen so that the positive and the negative spreaders only just meet one another. Plate 13 I shows an example of this. As the positive and the negative discharge here take place exactly simultaneously plenty of free electrons will be present at the outer ends of those positive spreaders which reach over — or nearly over — to the outer ends of the negative spreaders. An ionization by collision will then occur at this place, as the positively charged tips of the positive spreaders will attract the free electrons with great strength. But the pictures show clearly that the discharge tracks thereby formed have a character altogether different to that of the positive spreaders[1].

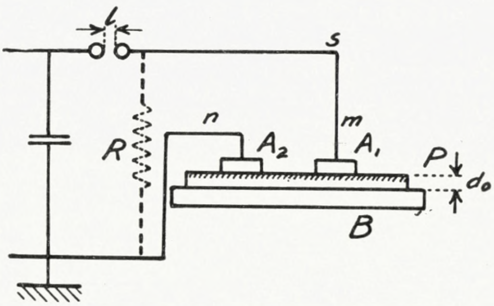


Fig. 23. Arrangement for simultaneous registration of positive and negative figures on one photographic plate. The metal plate *B* is insulated.

These discharge tracks go partly directly from the ends of the positive to the ends of the negative spreaders, and partly they connect the ends of neighbouring positive spreaders. The outer ends of a few of the positive spreaders have

[1] This feature is already mentioned in L. F. I, figs. 62 and 63.

apparently been bent by the electric field towards the nearest part of the negative figure.

Plate 24 I—IV shows enlarged portions of similar figures taken at atmospheric pressure. We see also here that the presence of a strong ionization, (in which, no doubt, there are a great number of free electrons), may be the cause of discharge tracks starting from the ends of the positive spreaders, but also that these discharges are not similar to regular positive spreaders. We will come back later to the formation and appearance of these discharge tracks; see appendix 3.

Similar conditions are found in the pictures shown in plate 18 I—II. In part I the positive spreaders go far into the area of the negative figure but in spite of this they fully preserve their character. In part II a positive discharge is subsequently started from the negative electrode and the course of this discharge is determined mainly by the electric force from the electric charge in the negative figure. (The same applies to the subsequent positive discharge in the negative figure shown plate 25 I. In these cases also the positive spreaders preserve their typical character.

In very many cases ionization by collision occurs by mutual action between the positive spreaders. This is thus the case in plate 18 I at the points marked 1—9. See also plate 14 I and II which are taken with the apparatus disposed as in fig. 24 — the same as fig. 43 in L. F. I.

In plate 14 I this ionization by collision occurs especially in such places where, after the formation of the positive spreaders, a strong electric field arises at right angles to their direction of propagation. In part II these discharges occur especially at the meeting line of the two figures.

The same applies to the ionization by collision in the neighbourhood of the meeting line of the positive figures on plate 15 I and II.

In all cases where such ionization occurs, the discharge has a character essentially different

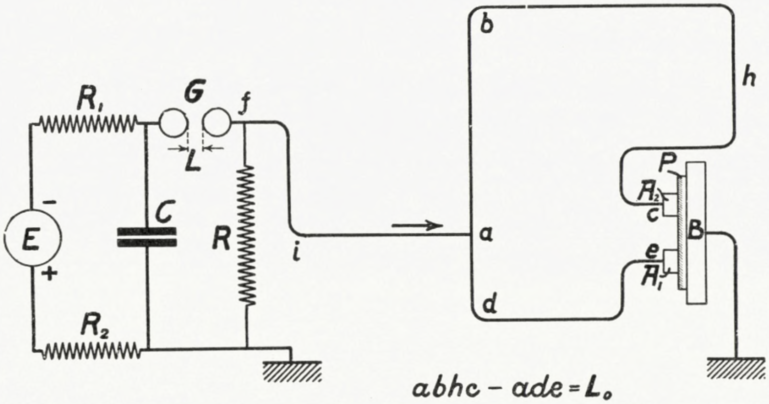


Fig. 24. Arrangement for determining the spreading-out-velocity of the Positive Figures.

to that typical of the positive Lichtenberg spreaders — to mistake the one for the other is impossible.

(b) Influence of Strong Simultaneous Ionization.

If a great number of electrons are set free at the same time and place as that at which spreaders are formed, then a very peculiar phenomenon occurs which we will examine a little more closely. Such a release of electrons may be effected by illumination of the photographic plate by an intense electric spark, preferably of short duration, since the plate will be blackened too much if the spark lasts too long. The number of electrons released will also only increase with the duration of illumination during a very short time interval, as the released electrons are for the greater part captured within 10^{-6} sec. by oxygen or water molecules, or the like, thus being[1] transformed into heavier, less active, ions, whose number will further decrease very rapidly owing to recombination between positive and negative ions.

The most simple experimental method is to apply a spark potential so high that a spark will pass from electrode *A* to the metal plate *B* below, see fig. 1. Another equally simple method is to let a spark pass between the electrodes A_1 and A_2 in fig. 23. In plate 11, I and IV show a few examples produced by the first method, II and III by the second. We will first examine the second. In that part of the photographic plate which is exposed to light from the spark, the positive spreaders are surrounded by a "soft", comparatively intensely luminous "veil" or rim, while those parts which are in the shadow have the normal appearance. The results are exactly similar in the first named case where the spark passed between the electrode *A* and the metal plate below. Part I shows such a picture taken in atmospheric air; part of the positive spreaders have here become strongly distinguished by the very strong rim-formation. Part IV shows a similar discharge in oxygen and is even more significant, since in this case the positive Lichtenberg discharges themselves are so faintly luminous that they cannot be photographed, see p. 16 above. But here they appear very distinctly owing to the strong rim-formation although the latter is weaker than in atmospheric air. This smaller intensity is caused by the fact that negative figures which are formed by ionization by collision are also less luminous in oxygen

[1] P. O. Pedersen: "Propagation of Radio Waves". Copenhagen 1927 (Fig. IV 7, p. 46).

than in atmospheric air. With negative figures — see parts II and III — a luminous ring appears a little outside the photographic boundary of the figure.

The luminous edges around the positive spreaders are caused by ionization by collision, which takes place in the strong fields immediately outside their boundary owing to the presence of numerous free electrons which may instantaneously start such ionization. We shall see later that there is reason to assume that the positive charge is of about equal intensity over the entire cross-section of positive spreaders. The electric field is therefore strongest at the very edge. The fact that the light intensity is greatest here is in good agreement with this.

In case of negative figures the outer boundary is not quite sharp, for, as stated before, the photographic intensity decreases gradually towards zero. We shall show later that there is reason to assume that the negative "electric" figure, a very short time after its formation, is slightly larger than the photographic figure, and that the charge-gradient is steepest at the edge of the "electric" figure. Consequently the strongest field and the most luminous rim is also found here.

It is thus possible fully to explain the source of these luminous rims. The foregoing discussion further shows that the figure formation — positive as well as negative — proceeds in the normal manner in spite of the presence of a great number of electrons and ions, while the luminous rims are caused by ionization by collision analogous to that which U. Yoshida and others assume to be the cause of the formation of the positive spreaders. But the discharge caused by this ionization by collision has, as appears from the positive pictures, a character entirely different to that of the positive spreaders.

U. Yoshida[1] has also treated this rim effect but has only tried to find an explanation in accordance with the one given for "dark sparks" ("dunkle Funken") i. e. founded on the Clayden effect[2], but he comes to the conclusion: "that the phenomenae observed are not photographic reversals of any kind ever known" (p. 319), which result is at all events not in contradiction to the one arrived at by us.

Some of the features touched upon in the preceding part are also treated rather more thoroughly in Appendix 3.

(c) Influence of Various Irregularities in the Photographic Film.

We have also investigated the effect of various irregularities prepared in or on the surface of the photographic film over which the positive figure spreads out. For example, plate 10 IV shows a case in which thick, electrically non-conducting ink lines were drawn over the film. They do not appear to affect the spreading-out of the figure. Plate 9 II shows the effect of some pencil lines over the film. It is clearly seen from the appearance of the figure at line 2 that this line is a fairly good con-

[1] U. Yoshida: l. c. (b).

[2] With reference to the formation of these "dark sparks" see P. Metzner (Verh. d. D. phys. Ges. Bd. 13, p. 612—616, 1911). Extraordinary beautiful samples of such dark sparks are found in Lord Armstrong: "Electric Movement in Air and Water", plates nos. 18 and 34 (London 1899). With regard to an explanation of these dark sparks on the basis of the Clayden effect we may refer to R. W. Wood (Phil. Mag. vol. 6, p. 577—590, 1903), K. Schaum (Verh. d. D. phys. Ges. Bd. 13, p. 676—679, 1911) and to M. Volmer und K. Schaum (Zeitschr. f. wiss. Phot. Bd. 14, p. 1—14, 1914).

ductor. On the other hand the positive points ending near line 1 are perfectly normal, proving that a considerable conductivity produced in this way has no effect on the formation of the positive spreaders, as long as these do not come into contact with the conducting lines.

(d) Ionization and Negative Discharges.

In the preceding paragraphs we have mainly treated only the spreading-out-conditions of the positive figures. The formation of negative figures is also, however, within very wide limits, independent of an existing ionization. Our reason for not entering much into this question is that we consider the formation of the negative figures as already fully explained in L. F. I.

As the main result of the investigations described and referred to in this section, we may set forth the conclusion that the formation of regular Lichtenberg figures is independent within very wide limits of the intensity of an existing ionization.

CHAPTER IV

The Formation of the Positive Figures.

A. Preliminary Discussion of various Hypotheses.

At the outset various possibilities can be assumed for the process of formation of positive figures, and several hypotheses to this effect have already been set forth[1]. However, in our opinion at least, none of the hypotheses already advanced can be brought into agreement with the experimental results. In this section we will give a brief account of the reasons which have led us to this view; in section *B* we will then give an account of the hypotheses we have arrived at through our investigations, and we shall here have occasion to give further reasons why we cannot accept the older hypotheses.

1. Formation of Positive Figures as due to Positive Ions moving away from the Electrode.

The process of formation of positive figures may be assumed to proceed analogously to the formation of the

[1] With regard to the various hypotheses for the formation of Lichtenberg figures see L. F. I chapt. I—II and K. Przibram: "Die ionentheoretische Deutung der elektrischen Figuren", Handb. d. Physik. Bd. 14, p. 402 ff. 1927.

negative ones, except that here positive ions, and not electrons, are moving outwards from the electrode, and by ionization by collision produce the necessary number of new positive and negative ions and electrons, the positive charge of the figure being mainly due to the movement of the latter towards the electrode.

The active positive ions may be either ordinary atomic or molecular ions, corresponding to the particular gas in which the discharge takes place[1], or they may be H^+-particles (Protons).

If positive figures were formed in this manner, then (α) their appearance should in the main features agree with that of the negative ones[2], (β) the relation between the spreading-out-velocity and the thickness d_0 of the plate should in the main be the same for positive and for negative figures, and finally (γ) the spreading-out-velocity should be considerably smaller for positive than for negative figures.

In chap. III it is, however, shown that none of these consequences are in agreement with the facts, in fact the spreading-out-velocity is even considerably greater for positive than for negative figures. In section *B* of the following it will further be shown that the spreading-out-velocity of the positive discharges is so great that the figure formation cannot take place in the manner here considered.

[1] Also S. Mikola (Phys. Zeitschr. Bd. 18, p. 161. a. f. 1917) seems mainly to have this view, though he considers both "Die korpuskulare Strahlung des Kondensators" and further "Die impulsive Strahlung des Kondensators" as active in the formation of figures and the second of these rays he considers to have electro-magnetic character.

[2] We have been unable to feel convinced by the reasons given by K. Przibram (l. c., p. 403) for the great difference in appearance between the positive and the negative figures on the basis of this hypothesis.

2. Formation of Positive Figures due to Negative Ions (Electrons) moving Inwards to the Electrode.

The formation of positive figures could also be assumed to proceed in the following manner: Negative ions from the outer edge of the figure are drawn inward towards the electrode, the necessary number of negative ions being produced by ionization by collision at the outermost ends of the spreaders. The positive charge on the figure is also in this case in the main due to the negative ions moving toward the electrode.

Here again there are several possibilities, in that the ionization by collision may be initiated either (α) by means of the (natural) ionization present in the air directly in front of the outer edge of the figure while this is being formed, or (β) by means of ionization by collision initiated by some positive particles driven out of the tips of the positive spreaders by the electric field.

(α). U. Yoshida is — as mentioned in chap. III 5 (a) — a follower of the first one of the here named theories. But since the experiments discussed in chap. III 5 (a) and (b) show that the circumstances of formation and the appearance of positive figures are independent within very wide limits of an initial ionization either previous to or simultaneous with, the formation of the figure, we cannot suppose the figures to be formed in the manner set forth under (α).

But even setting aside the said experiments — which we indeed consider as conclusive — the hypothesis set forth under (α) would meet with a number of difficulties. It would thus be difficult to give a satisfactory explanation of the very marked difference in appearance of the positive and the negative figures, of the difference in their spreading-out-velocity and initial conditions and of the dependency of the range on the plate thickness. Further, it would be difficult to explain the sharp boundary lines found, and the fact that the width of the positive spreaders is inversely proportional, and their number directly proportional, to the applied air pressure.

Completely decisive for the dismissal of this hypothesis are also the examples given in chap. III 5 (a) and (b), which show that even though such suction or drawing in of negative ions (or electrons) may occur under particular conditions, the discharge-tracks thereby formed have a character entirely different from that of the positive spreaders.

By this means we have, in our opinion, proved the untenability of the hypothesis in question.

(β). There remains now only the hypothesis mentioned under (β), which we shall subject to a more thorough discussion.

B. Theory of the Positive Figure according to the Hypothesis 2 (β).

0. Further Specification of this Hypothesis.

The foremost positive particle p^+, see fig. 25, is assumed to travel with the velocity U in the front of the positive spreader Q_+. The foremost particle has a sufficiently high velocity to act in a strongly ionizing manner, and releases a considerable number of electrons, indicated by dots in fig. 25. These electrons are pulled toward the tip of the spreader by the very strong field existing here, and on their way they release by collision a further large number of electrons, which are then pulled towards the electrode by the field along the spreader.

In this manner the strong electric field will automatically follow the foremost particle. The field strength at the

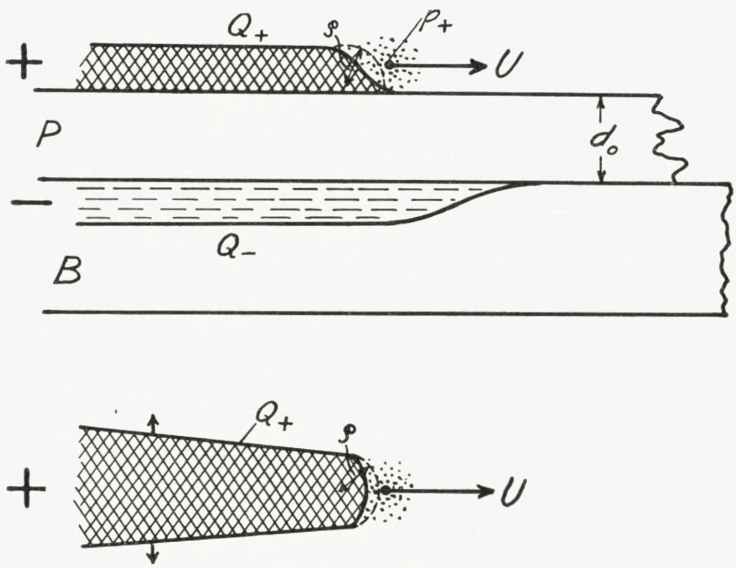


Fig. 25. Schematical Representation of the conditions at the Tip of a Positive Spreader. *P* is the photographic Plate, *B* the earthed metal Plate, see fig. 1.

tip of the spreader cannot be directly measured, but its approximate value may be estimated by the following considerations.

Assuming the *p. d.* between the positive electrode and the grounded plate *B* to be 15 000 volts, this will correspond to a spark length of about 2 mm. A drop in potential will, however, occur out along the positive spreader, and the potential at the tip of the spreader at the moment in question is assumed to be 9 000 volts. It is further assumed that the outermost tip has a spherical shape with a radius ϱ cm. The foremost particle p^+ at a distance $(\varrho + \Delta\varrho)$

from the centre of the spherical spreader tip will then be exposed to an electric field $E_{\varrho+\Delta\varrho}$ which may with approximation be put equal to

$$E_{\varrho+\Delta\varrho} = \frac{\varrho}{(\varrho+\Delta\varrho)^2} \cdot V_0\,, \tag{1}$$

which for $\Delta\varrho = 0$ reduces to

$$E_\varrho = \frac{1}{\varrho} \cdot V_0. \tag{2}$$

If the front particle is inside the spherical part and the charge density is assumed to be constant, then the field at a distance $(\varrho - \Delta\varrho)$ from the centre will be determined by

$$E_{\varrho-\Delta\varrho} = \frac{\varrho - \Delta\varrho}{\varrho^2} \cdot V_0. \tag{3}$$

If, on the other hand, the total charge is located on the surface of the sphere, we have

$$E_{\varrho-\Delta\varrho} = 0. \tag{4}$$

$E_{\varrho-\Delta\varrho}$ will in general have a value somewhere between those determined by (3) and (4).

$E_\varrho = \frac{V_0}{\varrho}$ is thus the highest value attainable for the field intensity.

If $E_{\varrho+\Delta\varrho}$ at a given moment is greater than the field value E_U, necessary to maintain the velocity U of the front particle, then this velocity will increase. The distance $(\varrho+\Delta\varrho)$ will thereby become greater, but this is again — according to (1) — followed by a decrease in the field intensity.

If, on the contrary, $E_{\varrho+\Delta\varrho} < E_U$ the velocity of the particle will decrease but the distance $\varrho+\Delta\varrho$ will thereby also be decreased and $E_{\varrho+\Delta\varrho}$ will consequently increase. From this it appears that the particle will adjust itself to a distance where the field intensity and the velocity have corresponding values.

This, however, assumes that

$$E_{\varrho} = \frac{V_0}{\varrho} > E_{U_0}, \tag{5}$$

where U_0 is the smallest velocity at which the front particle can be controlled in the manner referred to.

The positive front particle will in fact almost, or possibly completely, lose its ionizing ability if its velocity decreases below a certain value, see App. 1, sections 6 and 7. At this limit the strong ionization at the tip will cease and the intensity of the electric field in which the front particle is located will become much weaker. The particle will then lose its velocity within a very short distance. This process will actually already have set in at a velocity at which the particle still possesses some ionizing ability, and the critical value of U_0 will therefore be so high that considerable ionization still takes place. The actual value of U_0 cannot, however, be calculated beforehand, though such a limiting value no doubt exists.

The front particle may, however, be brought to a standstill by another cause. It will not always contain a positive charge but will sometimes be positive and sometimes neutral, as is well known from investigations of positive rays and other high-speed positive particles, see App. 1. If we call the distance travelled by the particles in a charged state l_1, and the distance travelled in a neutral state l_2, then we know from experience that the ratio $\frac{l_1}{l_2}$ decreases with decreasing velocity. This ratio seems, however, especially in cases where the positive particle is a proton, to approach a limiting value which does not further decrease with decreasing velocity, see for example fig. 2 in App. 1. These relations are, however, not quite certain, compare App. 1, section 6.

The deduced values for the ratio $\frac{l_1}{l_2}$ are at all events only valid for the average values of these distances, and the actual distances may show very considerable differences. If the particle travels 10^{-4} cm without charge, then it will decrease considerably in velocity. If it has been retarded so much that it has come inside the strongly charged spherical front, then the field intensity drops so considerably that the particle, even if it becomes positively charged again, has only very little chance of regaining the necessary velocity; if it does not, it will be drawn back and more into the charged sphere, and completely lose its velocity.

In both of the above described cases the front-ionization ceases and the growth of the positive spreader stops. The strong electric field at the tip decreases simultaneously because the positive charge is spreading out, although comparatively slowly, by which means V_0 decreases and the radius ϱ increases. The drop in V_0 is due to the spreading out of the positive electricity available over a larger area, and as no perceptible ionization now occurs at the tip, no negative electricity is carried away toward the electrode. V_0 must therefore decrease.

We shall now proceed to give a more detailed explanation of the characteristics of the positive figures mentioned in chap. III, but we may remark here that in part 2 of the following we are led to the assumption that the positive front-particles are protons.

1. Differences in Shape of Positive and Negative Figures.

(a) Width of Positive Spreaders.

The width of the positive spreaders is, as was shown in chapter III 1 (a), inversely proportional to the air pressure p[1].

Immediately after the passage of the front-particle the spreader will be very thin and this the more so the smaller is the velocity of the particle. This primary ionization channel is strongly positively charged, because part of the electrons set free through the ionization by collision in the tip of the channel are drawn toward the electrode. The value of the positive charge is mainly determined by the capacity of the channel and by the fact that the potential in the channel is smaller than the potential of the electrode by an amount corresponding to the ohmic drop of potential along the spreader. We will return to this point later. The ionization along the spreader is so intense that the electrons pulled toward the electrode only represent a fraction of the total number set free by the ionization by collision.

Repulsion between the resulting positive charges in the spreader will cause an increase in thickness of the spreader, and this increase will occur with a rapidity inversely proportional to the air pressure, as is more thoroughly shown in App. 2. The thickness — or width — of the spreader a given time after the passage of the front-particle will therefore — ceteris paribus — also be inversely proportional to the air pressure.

As we shall see later the luminous effect must be

[1] Parts of the following accounts apply also to discharges in the free atmosphere — as is easily seen.

assumed to arise a certain time after the original ionization, and the intensity of the photographic picture of the spreader must therefore also as a whole be inversely proportional to the air pressure.

The rate of increase of the thickness of the spreader must also — ceteris paribus — be proportional to the resulting charge per unit length of the spreader, and this charge must have a value increasing with increasing potential across the Lichtenberg spark gap[1]. The *p. d.* between the spreaders and the earthed plate *B* decreases, as said before, from the electrode towards the edge of the figure, and the drop in potential is the greater the higher the air pressure. But as the width t of the spreaders is measured at a distance from the tip inversely proportional to the air pressure, it may be assumed that the spark potential will have no appreciable influence on the value of t thus measured.

The preceding further shows that the width must increase fairly regularly from the tip towards the electrode. This regular increase in width is only found, however, at such distances from the electrode where the spreaders are so wide apart that they do not interfere with each other's lateral spreading. For Lichtenberg figures, the discharge takes place over the surface of an insulating plate. Here the plate surface will form a geometric boundary to the figure and further the induced charge Q_- on the metal plate *B*, see fig. 25, will, in proportion to the thickness of the insulating plate, exert a greater or smaller influence upon the electric field immediately in the neighbourhood of the spreader Q_+. In the following we shall consider the

[1] In the neighbourhood of the electrode the field from neighbouring spreaders will hinder a free lateral development of the spreader.

influence which the air pressure and other factors exert upon the formation of the spreader.

We will first, on the basis of the schematic representation in fig. 26, look at the conditions at very low pressures.

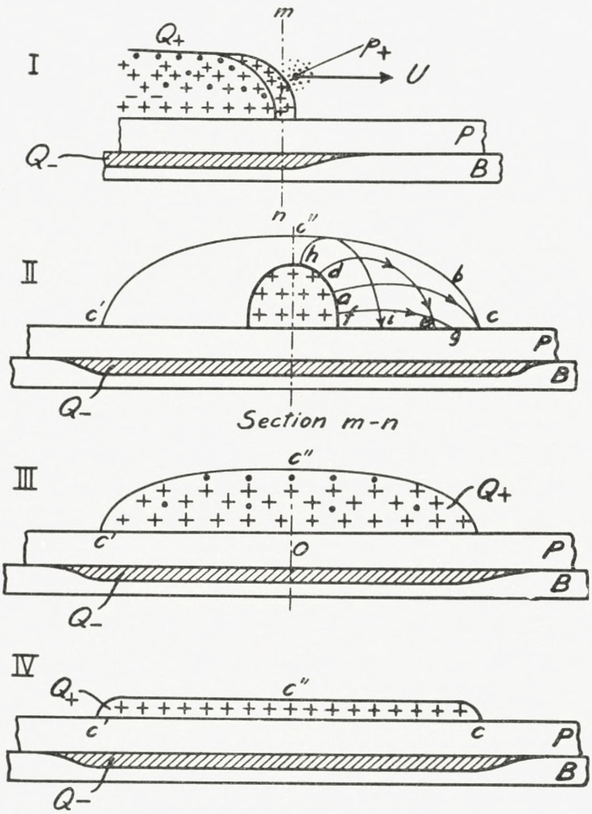


Fig. 26. Schematical Representation of the manner of formation of Positive Spreaders at low Pressures.

Part I represents the tip of a positive spreader. Immediately around the front-particle p_+ will be found some electrons and a corresponding number of positive ions. The number, however, is comparatively small because the air pressure is low. Both electrons and positive ions are indicated by dots in the figure. Immediately behind the front-particle will exist, within a limited layer, a very dense positive space charge, due to the very strong ionization by collision occurring in the very strong field at the tip of the spreader, combined with the great mobility of the electrons which allows their quick removal by the field toward the left — in the direction of the electrode — while the comparatively heavy positive ions may be considered as nearly stationary within the very short intervals of time here in question.

Due to mutual repulsion the positive ions will during the time following move outward — f. inst. *a* in part II along the path *a-c* — and the attraction from the induced charge Q_- will finally force the ion down against the surface of the plate *P*. The smaller the air pressure, and the greater the velocity of the ion, the fewer collisions will it undergo on its way.

Even if the front travels so rapidly that no considerable number of free electrons get time to recombine with positive ions to form neutral molecules, or with neutral molecules to form heavy negative ions, nevertheless, at such low pressures a sufficient number of electrons may easily be carried away toward the electrode, thus enabling the capacitive positive charge on the spreader — and the corresponding induced negative charge Q_- — to increase in accordance with the width of the spreader. The charge density per unit length of the spreader will therefore increase very rapidly. By this means also the forces tending to drive the positive ions outward will become great. Since the ions have at the same time a high mobility, owing to the low pressure, the width of the spreaders must also become great.

In part II it is assumed that the ion *a* travels the farthest before it impinges upon the plate *P* at *c*. Ions which start from higher points such as d and h or from lower points such as f will then impinge upon P at shorter distances from the centre of the spreader.

The positive ions which hit the surface of the plate will be pulled down against it by the electric field and will then be retained by the electric forces between the ion and the molecules in the surface of the solid plate. These positive ions will thus almost completely lose their mobility.

After a short time the state of charge will pass through the state III and approach the state IV, where the density per unit area σ of Q_+ and Q_- (which have the same numerical value) is very nearly constant over the entire width of the spreader in that

$$\sigma = \frac{\varepsilon}{4\pi d_0} \cdot V,$$

where ε is the dielectric constant of the plate and V the *p. d.* between the point of the spreader considered and the plate B.

Those positive ions which are not retained in this way will recombine with a corresponding number of negative ions or electrons and this recombination will in the main occur evenly distributed over the cross section $cc''c'$ — see part II — since the above mentioned scattering of the positive ions occurs with such great velocity[1], and the density of these ions is so great, that the air particles within this cross section are set in such violent motion that effective mixing takes place.

The spreading out occurs in an analogous manner at higher air pressures, but the scattering velocity of the

[1] See also later.

positive ions is smaller and the dispersal of the electrons is slower so that the width of the spreader becomes smaller. Another contributory cause is that the original ionization channel, formed by the front-particle, has a smaller cross-section and accordingly also a smaller charge density. The width of the spreader must therefore decrease with increasing air pressure.

Fig. 27 shows schematically various states of the formation of a positive spreader.

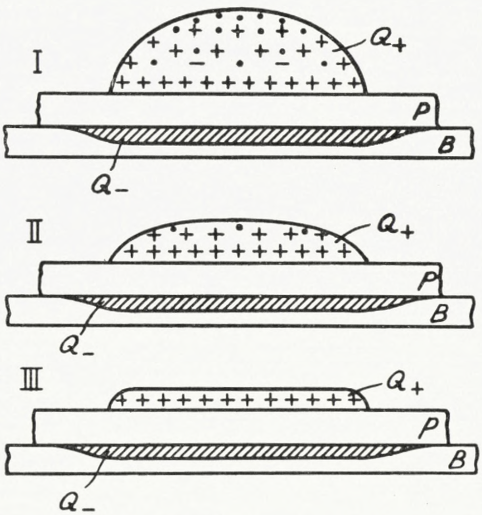


Fig. 27. Schematical Representation of various States of a Cross-Section of a Positive Spreader at Atmospheric Pressure.

With regard to the influence of the spark length on the width of the spreaders, we have already mentioned that the width at a distance from the tip inversely proportional to the air pressure is practically independent of the potential value on the electrode.

Generally speaking, the width will increase with increasing spark length, but to a smaller degree, because the induced charge Q_- increases at the same rate as Q_+, so that the direction of the field will be independent of the value of the two charges. Added to this, we here have near the electrode the above mentioned additional factor that neighbouring spreaders prevent the free spreading out of the single spreader.

The dependency of the width on the thickness d_0 of the insulating plate is evidently partly due to the fact that the induced charge Q_- at thin plates comes up near Q_+, so that the electric field at the edge of Q_+ has a powerful component perpendicular to the surface of the plate, and partly to the fact that Q_+ and Q_- at a given voltage will increase rapidly with decreasing values of d_0. The positive ions will therefore move with great velocity, but owing to the geometrical form of the field they will nevertheless impinge on the plate at a comparatively short distance from the centre of the spreader. It is not easy to say off-hand which of these two causes has the greatest influence. The value of d_0 has actually very little influence upon the width as long as d_0 is not too small. For very small values of d_0, however, the width of the spreaders will decrease with decreasing d_0 but as the original ionization channel has a finite thickness the width will not become quite zero as $d_0 \rightarrow 0$.

In the foregoing we have first discussed the ionization by collision at the tip and thereupon the increase in width of the spreaders. These processes actually occur simultaneously and the total ionization will be almost proportional to the final width of the spreaders.

We shall see later — under section (d) — that the photographic width must be very nearly the same as the width cc' of the charge; see fig. 26 and 27.

The hypothesis set forth thus fully explains the fact that the width of the positive spreaders increases with decreasing pressure, with increasing potential and with increasing plate thickness. An exact quantitative calculation cannot be carried out, but the simple considerations set forth indicate that the width must be inversely proportional to the air pressure. All the results mentioned in chap. III 1 (a) with regard to the dependency of the width on the various parameters are thus explained in a satisfactory manner by the hypothesis set forth.

(b) Number of Branches.

The number of branches per unit of length of a spreader must, ceteris paribus, be proportional to the air pressure. The number of suitable particles at those points where the field is strongest — i. e. at the edges of the spreaders — must in fact be proportional to the number of such particles per unit volume, i. e. to the air pressure.

The number of branches will also depend on the nature of the gas, but this we will return to later.

(c) Boundary-line of the Positive Spreaders.

The limits of the photographic image of the positive spreaders are determined by the limits to which the positive ions have spread at the moment when emission of light occurs. According to the foregoing, the spreading out of the positive ions is coterminous with the resulting positive charge; compare sect. (a) and figs. 26 and 27.

It appears from Appendix 2 that the outer boundary of a charge which spreads out owing to mutual repulsion will have a decided inclination to become sharp. The various conditions mentioned above under sect. (a) regarding the sidewards spreading out of the spreaders also contribute to the same effect.

These circumstances, combined with what will be set forth under the following sect. (d), offers a satisfactory explanation of the comparatively sharp boundary lines of the positive spreaders.

(d) Photographic Intensity.

The difference in luminosity of positive and negative figures is due to the fact that in the case of negative discharges, some of the positive ions set free by ionization by collision travel towards the electrode, while the main part of the negative charge is formed by electrons moving outward from the electrode or its immediate vicinity. This — as we shall see later — results in an exceedingly strong ionization at and near the electrode.

The light is produced in the main by recombination between the positive ions and the electrons and to a much smaller degree by recombination between positive ions and negative ions produced by neutral molecules combining with electrons. Where no positive ions are present no light appears. To this circumstance may be ascribed the fact that the luminosity at the outer boundary of the negative figures decreases gradually towards zero, since here the ionizing intensity, and consequently also the density of positive ions, decreases gradually.

Even when the field at the outer boundary has become so weak that ionization by collision has ceased, the electrons will be driven further out by the field but will in the mean time have combined with neutral molecules to form negative ions. But since no ionization occurs, no positive ions are produced, and therefore no light appears. We shall show later, under sect. 5 (a), that we are able to confirm by other means the fact that the charge has spread beyond the luminous part of the figure.

Corresponding conditions exist with regard to the dark lines of the negative figures. In this case also electrons and negative ions will pass in from the neighbouring parts of the spreader after the ionization by collision has ceased, but no, or at least extremely few, positive ions. The photographic luminosity will therefore also be extremely small.

For positive figures the conditions are entirely different. Here the highly mobile electrons set free by ionization by collision will be driven up from the photographic plate to the surface of the spreader by the field — compare fig. 27 I—II — and the field will then set them moving toward the electrode. The number of electrons absorbed by the electrode will be just sufficient to give the spreader a charge corresponding to the *p. d.* between the spreader and the plate *B*. Part of the rest of the electrons combine with positive ions, while others combine with neutral molecules to give negative ions, which again later combine with positive ones. The light is mainly produced by combination of positive ions with electrons or with negative ions.

The ionizing intensity is much smaller near the electrode than in the case of negative figures. The photographic intensity is therefore also comparatively small but will, on the other hand, be nearly uniform over the entire length of the spreader. The ionization will in fact, as said before, be very nearly proportional to the width of the spreader and will not be influenced by the electrons carrying negative charge from the spreader tip toward the electrode. The voltage drop along the positive spreaders is in fact so small that these electrons will cause no appreciable ionization by collision.

In single cases, the outermost tips or branches may be less luminous than the rest of the spreaders, see plates 4 I, 5 and 6 I which are taken at low pressures. At atmospheric pressure the corresponding branches are comparatively shorter. Possibly here the velocity of the front particles has become so small that the initial ionization which they cause is insignificant.

In this way we have accounted for the very peculiar difference between positive and negative figures with regard to light distribution. But it may still be asked: Can the spreaders really increase noticeably in thickness in the very short time elapsing before the positive and negative ions have recombined and before the light resulting therefrom is emitted?

The greater part of the recombination no doubt occurs within 1 to $2 \cdot 10^{-7}$ seconds. The emission of light probably occurs within a similar time interval after the recombination[1].

If, for example, the *p. d.* between the point in question of the spreader and plate *B* is 9 000 volts, and the thickness of the photographic plate about 1 mm, then the intensity of the electric field at the edge of the spreader will be of the order of 100 000 volt cm^{-1}. If we put the mobility of the ions at atmospheric pressure equal to 1.8 cm sec^{-1} per volt cm^{-1}, then the velocity of the edge will be $1.8 \cdot 10^{-5}$ cm sec^{-1}. The edge of the spreaders will then in the course of $1 \cdot 10^{-7}$ seconds move a distance of 0.18 mm. The increase in width of the spreader may thus attain the necessary velocity, and this velocity, moreover, is so high that the ionization at the tip may be assumed to be proportional to the width *t*.

This great velocity, combined with the great density of ions will, as said above, cause very vigorous movements inside the air volume occupied by the spreaders. By this means a very effective mixing occurs through the

[1] J. Franch u. P. Jordan: Anregung von Quantensprünge durch Stösse, p. 201, 1926.

entire cross section of the spreader whereby the density of the positive ions as such and negative ions as such will be practically the same over a cross section. The luminosity will therefore also very nearly be the same over the whole width of the spreader.

The occurrence of this vigorous motion further offers an explanation of various phenomena of a mechanical nature in connection with the formation of the figures. We may just mention the formation of dust figures or the depressions produced in the surface by the spreaders when the figure is formed over water or other liquids. Of a different nature are the permanent indentations made by Lichtenberg figures in the surface of melting pitch, which were noticed by E. W. Blake[1], and which are no doubt due to the electric forces between the charges of the spreaders and of the plate B.

2. Range and Velocity of the Positive Figures.

(a) Relation between Range and Velocity and the *p. d.*

According to L. F. II formulae (I)—(II) we have that:

$$x = R(1 - e^{-\alpha t}), \qquad \text{(I)}$$

and

$$U = \alpha R e^{-\alpha t} = U_0 \cdot e^{-\alpha t}, \quad U_0 = \alpha R \qquad \text{(II)}$$

where R is the range of the figures and x the length of the spreaders at time t.

From (I) and (II) we have that

$$U = U_0 \cdot \frac{R - x}{R}. \qquad \text{(III)}$$

[1] E. W. Blake: Am. Journ. Sc. and Arts (2) Vol. 49, p. 289—94, 1870.

According to App. 1, formula (7) the maintenance of a velocity U requires an electric force X determined by:

$$X = k\sqrt{U} \tag{IV}$$

where k is a constant dependent upon the nature and the pressure of the gas and also on the nature of the particle.

From (III) and (IV) we have that

$$X = k\sqrt{U_0} \cdot \sqrt{\frac{R-x}{R}}. \tag{1}$$

From this we find the corresponding *p. d.* distribution by putting, according to formula (2) p. 63.

$$X = \frac{1}{\varrho} V, \tag{2}$$

where ϱ is the radius of the strongly charged tip of the spreader. The length of this radius cannot be determined beforehand but it is reasonable to assume ϱ to be proportional to the range of the front particle at the velocity in question. This range, according to App. 1 formula (2), is proportional to $U^{\frac{3}{2}}$ so that the above formula may be written

$$X = \frac{aV}{U^{\frac{3}{2}}}, \tag{2'}$$

where a is the constant from formula (2) in App. 1.

From the above formulae we have that

$$V = \frac{k}{a} U_0^{\frac{1}{2}} U^{\frac{3}{2}} \sqrt{\frac{R-x}{R}} = \frac{k}{a} \cdot U_0^2 \left(\frac{R-x}{R}\right)^{\frac{3}{2}}. \tag{3}$$

When $x \to 0$ then $V \to V_0$ and $U \to U_0$ and we therefore get

$$U_0 = \sqrt{\frac{a}{k} \cdot V_0}. \tag{4}$$

This expression agrees in the main with the experimental results; compare L. F. I p. 50, fig. 52.

According to (II) we have

$$R = \frac{U_0}{\alpha},$$

and since according to Ann. d. Ph. Bd. 69, p. 214 α is very nearly independent of the *p. d.* V_0, it then follows from the above formulae that

$$R \cong c\sqrt{V_0}, \tag{5}$$

which also on the whole agrees with experiment; compare L. F. I p. 39.

Formula (3) determines the dependency of the *p. d.* on the distance from the electrode at the instant when the front particle is passing the point in question. We will then proceed to investigate how the resistance of the spreaders must depend upon the time elapsed since the start of the ionization, in order that by application of Ohm's law we may come to the said expression for the *p. d.* We may here emphasize, however, that these calculations are only rough estimates. The conditions in the discharge tracks are so complicated and so little known, that a theory satisfactory in all details cannot be developed at present.

For the sake of simplicity we will therefore assume that the initial ionization following the tip of the spreader has the same value all along the discharge track. For a length element dx of the spreader the resistance may be $r \cdot dx$. We further assume that after a lapse of time t' reckoned from the commencement of the ionization the resistance has increased to $e^{\beta t'} \cdot r \cdot dx$.

The resistance M_y of the length y of the spreader at the moment when this has the range y is consequently determined by

$$M_y = r\int_0^y e^{\beta t'} \cdot dx = r\int_0^y e^{\beta(t_1-t_2)} \cdot dx\,,$$

where t_1 is the time from the start of the discharge until when it has reached the distance y, while t_2 is the time at which the spreader has attained the length x.

From (I) we have that

$$e^t = \left(\frac{R}{R-x}\right)^{\frac{1}{\alpha}}, \tag{6}$$

where t and x are corresponding values of time and spreader length.

Consequently we have

$$M_y = r\int_0^y \left(\frac{R-x}{R-y}\right)^{\frac{\beta}{\alpha}} \cdot dx = \frac{r}{\frac{\beta}{\alpha}+1}\left(\frac{R^{\frac{\beta+1}{\alpha}+1}}{(R-y)^{\frac{\beta}{\alpha}}} - (R-y)\right). \tag{7}$$

For the current i we have the following expressions

$$i = \frac{V_0 - V_y}{M_y} \quad \text{and} \quad i = h\,V_y\,U\,, \tag{8}$$

where h is proportional to the capacity of the spreader per unit of length and therefore also approximately proportional to the width t of the spreader.

If we put the two quantities in (8) equal we get

$$V_y = V_0 \cdot \frac{\left(\frac{R-y}{R}\right)^n}{\left(\frac{R-y}{R}\right)^n + K_0 R\left(1 - \left(\frac{R-y}{R}\right)^{n+2}\right)}, \tag{9}$$

where

$$K_0 = \frac{hr}{\frac{\beta}{\alpha}+1} \quad \text{and} \quad n = \frac{\beta}{\alpha} - 1\,. \tag{10}$$

Consequently, for $y = 0$, $V_y = V_0$, and for $y = R$, $V_y = 0$.

As h is very nearly proportional, and r inversely proportional, to the width of the spreader, then K_0 is very nearly independent of this width.

From (9) we have

$$\frac{dV_y}{dy} = -V_0 \cdot \left(\frac{R-y}{R}\right)^{n-1} \cdot \frac{K_0\left(n+2\left(\frac{R-y}{R}\right)^{n+2}\right)}{\left[\left(\frac{R-y}{R}\right)^n + K_0 R\left(1-\left(\frac{R-y}{R}\right)^{n+2}\right)\right]^2}. \quad (11)$$

For

$$K_0 R(n+2) = n \quad (12)$$

we have for $y \to 0$ and $y \to R$

$$\left(\frac{dV_y}{dy}\right)_{y \to 0} = \left(\frac{dV_y}{dy}\right)_{y \to R} = -V_0 \cdot \frac{n}{R} \cdot \left(\frac{R-y}{R}\right)^{n-1}. \quad (13)$$

If we further put $n = \frac{3}{2}$ then we see not only that V_y at $y = 0$ and $y = R$ has values equal to those determined for V by (3) but also that $\left(\frac{dV}{dy}\right)$ at the points considered has the same value in the two cases since, according to (4), we have

$$V_0 \cdot \frac{n}{R} = \frac{3}{2} \cdot \frac{R}{a} \cdot U_0^2.$$

With regard to the various assumptions on which the above results are deduced, we must still add a few remarks.

The assumption that the original resistance per unit of length is constant along the entire spreader is not correct. There is reason to assume the resistance to be inversely proportional to the width, but the quantity of electricity consumed by charging of the spreader is at the same time nearly directly proportional to the width. It will, however, in the main be the product of resistance and capacity of the spreader which is the deciding factor, and this product will, as mentioned, be very little dependent on the width. It is, however, to be assumed that the capacity decreases somewhat more slowly than the conductivity with decreasing width, and the product $(r \cdot h)$ will therefore increase a little under these conditions. The error caused by the slightly tapered form of the spreaders will, all things considered, hardly have any considerable value, but for the thinner spreaders there will be a tendency to a smaller velocity and a smaller range than for the wide ones, a result which is confirmed by plate 4 I.

We will now discuss the assumption that the conductivity decreases at the rate $e^{-\beta t}$. The conductivity is mainly due to the free electrons the number of which per cm^3 we will call N. We therefore simply write the conductivity of the spreader as

$$\frac{1}{r} = \lambda = \frac{c}{A} \cdot N, \tag{14}$$

where A is the cross-sectional area of the spreader and c is a constant.

The free electrons may be lost by recombination with positive ions or by forming negative ions with neutral molecules.

In the cases where the ionization is not extremely strong and where no strong outer field is found, we shall have in the first case

$$\frac{dN}{dt} = -\alpha N^2 \quad \text{or} \quad N = \frac{N_0}{1 + \alpha t N_0}, \tag{15}$$

where, at atmospheric pressure, $\alpha \cong 1 \cdot 6 \cdot 10^{-6}$.

In the second case we have

$$\frac{dN}{dt} = -bN \quad \text{or} \quad N = N_0 \cdot e^{-bt}, \tag{16}$$

where b is a constant which is proportional to the air pressure but also to a large extent depends upon the nature of the gas. For hydrogen and nitrogen b will thus be very nearly equal to zero, while for oxygen b will be of the the order of 10^6 at a pressure of 760 mm Hg.

A rough calculation of the maximum resistance shown by the spreaders and the number of ions per cm^2 of a cross section of the spreader — in both cases on the assumption that the thickness of the spreaders (perpendicular to the plate) is about $^1/_4$ mm — leads, however, to the result that the density must be 10^{12} to 10^{13} electrons per cm^3. We have, however, further assumed that we can apply a value of the conductivity calculated on the basis of the kinetic gas-theory. But at such intense ionization and such strong fields as are here considered, neither formula (15) nor (16) can be applied; nor will the general kinetic theory lead to results of even approximately the right order. The conditions in highly conductive air spaces are as a whole very little known; compare f. inst. the efforts to establish a plausible theory for the conductivity of the electric arc[1]. We shall therefore not enter further into this question here but shall merely remark that, considering the agreement found with experiment, when assuming the conductivity to decrease at the rate $e^{-\beta t}$, there is reason to believe that the conductivity of strongly ionized spaces in the main decreases in the said manner, the value of β being about $5 \cdot 10^{-7}$. Presumably this result cannot be dismissed at the outset as unreasonable.

A few more remarks may be made with regard to the decrease of the conductivity as a function of time. First, the conductivity will maintain a considerable value as long as

[1] H. Hagenbach: Handb. d. Phys. Bd. 14, p. 346 1927.

a great number of electrons are set free at the tip of the spreader. We will come back later to this question under 4 (a). Secondly, even when all of the free electrons have disappeared, the discharge track will retain some conductivity owing to the heavy ions still present. But this conductivity is so inconsiderable, at least at higher pressures, that it is of no significance for the passage of the comparatively strong currents existing in the spreaders during their formation.

At comparatively low pressures the conductivity due to the ions may be of such value that it may have some influence upon the formation of the spreaders. The conductivity due to the ions will decrease at a slower rate than the conductivity due to the electrons, and the following conditions may then occur: the growth of the spreader stops completely, or very nearly completely, because the ohmic voltage-drop becomes too high, after which the part of the spreader already formed will be charged comparatively slowly to a higher potential owing to conductivity due to the ions. Especially at the tip the potential may increase considerably. The result of this is that a new spreader starts from the tip of the old one, or that the tip of the latter gains renewed speed. Examples of such cases are quite frequent; see thus plate 21 IV and 23 (p = 50 mm Hg.). In some few cases this process may be repeated several times; see plate 12 IX.

The conductivity due to ions at all events plays an important rôle for the continued growth of the spreaders if the potential on the electrode is maintained for a comparatively long time, but this question will be treated later. The leading away of the charge from the figure through the shunt R — see fig. 1 — also mainly depends on the conductivity due to ions.

Although it has not been possible to work out an explicit theory for the conditions in the strongly conductive spreaders, the considerations set forth above show that the formation of the spreaders, as known from experience, may be explained in a reasonable manner by means of the hypothesis set forth about the manner of formation of the positive spreaders.

As for the negative figures there exists in the main no doubt with regard to their mode of formation, and a detailed analysis of the above treated questions is therefore of less interest for these figures. We shall therefore make only a few remarks concerning their mode of formation.

Since their conductivity is mainly due to free electrons, the amount of electricity neccessary to charge the negative figure must for the most part originate in electrons travelling outward from the electrode. A strong ionization must therefore occur at, and in the neighbourhood of, the electrode, with intensity decreasing outwards. A minor part of the charge is due to positive ions moving toward the electrode. On their way outward the electrons will cause ionization and this will be the stronger the stronger is the field. Since the latter increases towards the electrode, this will also be the case for the ionization by collision. The outer edge of the figure is determined by the condition that the field becomes too weak here to maintain ionization by collision.

(b) Relation between Range and Velocity and the Thickness d_0 of the Insulating Plate.

With decreasing thickness d_0 of the insulating plate, the electric field at the tip of the positive spreaders will increase, but its direction will at the same time be bent down towards the plate; the track of the front particle will therefore also have a downward direction as indicated by the arrow U in fig. 28.

When the front particle hits the plate it will be stopped. The tendency to impinge on the plate will increase with decreasing thickness of the plate and consequently the range of the figure will go toward zero simultaneously

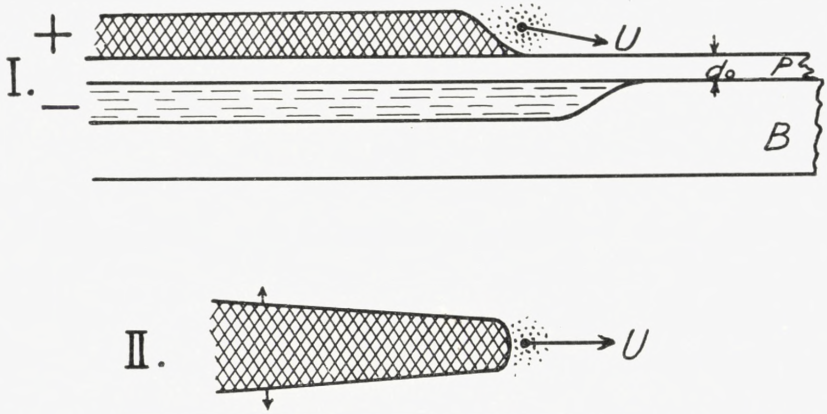


Fig. 28. Schematical Representation of the Conditions at the Tip of a Positive Spreader for small Thicknesses of the Insulating Plate.

with the plate-thickness. For very great values of d_0 the electric field, and therefore also the velocity, will become comparatively small. The velocity consequently attains a maximum value at an intermediate value of d_0. This is in complete agreement with the experimentally determined relation between velocity and plate-thickness; see fig. 8 in this paper and L. F. I, p. 56, fig. 59.

The range of the positive figures as a function of d_0 — if, as is to be expected, α remains constant — must be governed by a similar relation, and this is also in agreement with the experimental results set forth in L. F. I, p. 30, fig. 34.

In the case of negative figures the electrons at the end of the spreader will also be drawn down towards the plate, and the thinner the plate, the stronger will be the field. The front-ionization will therefore also be the stronger, the thinner the plate. This strong ionization will very rapidly set free electrons in sufficient number to supply the proper negative charge to the negative spreader. When this is effected, the surplus of electrons coming from the electrode and its vicinity will not be pulled down toward the plate, but can move freely in the air under the influence of the field. Since the field at the edge increases with decreasing thickness of the plate, the velocity, and simultaneously the range, will approach a maximum when the plate thickness decreases toward zero. This result, also, is in agreement with experience; compare L. F. I, p. 30, fig. 34 and fig. 8 in this paper.

(c) Relation between Range and Velocity and the Pressure. Nature of the Positive Front Particle.

It appears from equation (IV) in section (a) above and equations (2), (6), and (7) in App. 1 that

$$k = k_0 \cdot p, \tag{1}$$

where p is the air pressure and k_0 a constant.

From equation (2) in App. 1 we find that

$$a = \frac{a_0}{p}. \tag{2}$$

Equation (4) in section (a) will consequently give

$$U_0 = \sqrt{\frac{a}{k} \cdot V_0} = \frac{1}{p}\sqrt{\frac{a_0}{k_0} \cdot V_0}, \tag{3}$$

so that according to the theory the velocity of the positive spreaders should be inversely proportional to the air pressure.

Their range is accordingly determined by

$$R = \frac{U_0}{\alpha} = \frac{1}{\alpha p} \cdot \sqrt{\frac{a_0}{k_0} \cdot V_0}. \tag{4}$$

Since α decreases somewhat with decreasing pressure, the range will, according to equation (4), increase somewhat more rapidly with decreasing pressure than it would if R were proportional to $\frac{1}{p}$. This is in good agreement with the formulae given by v. Bezold and S. Mikola[1] and our experimental determination of the dependency of the range on the air pressure, at not too low pressures is also fully in agreement with this.[2]

But from fig. 42 L. F. I it further appears that at very low pressures R increases more slowly with decreasing pressure than it should according to equation (4). Fig. 58 in L. F. I further shows that the velocity approaches a finite limiting value — U_m in fig. 8 above — when the pressure approaches zero.

It is also easily seen that at extremely low pressures the manner of formation must be different from that assumed here. If we f. inst. assume that in the limit the pressure becomes zero, then the front particle will nevertheless by unable to attain a velocity greater than that corresponding to a free fall through the entire $p.d.$ V_0. We consequently have

$$U_{p \to 0} < 1.38 \cdot 10^6 \sqrt{\frac{z}{M} \cdot V_0}, \tag{5}$$

where M is the mass of the particle with the hydrogen atom as unity, and $(z \cdot e)$ its charge.

The value given by the right hand term in (5) is not attained before the particle has travelled through the total

[1] L. F. I, p. 41. [2] l. c., Fig. 42.

p. d. V_0. Right at the electrode the velocity will be zero. On the other hand the method used in these investigations for measuring the velocity, (see L. F. I), determines particularly the velocity near the electrode. Applying this method to the case of freely moving particles will consequently lead to values of U considerably lower than those given by the right hand term of equation (5). Finally, the spreading-out-velocity of the figure at small but finite pressures will certainly be smaller than the velocities found for freely moving particles.

The experimentally obtained values of the velocity must therefore, at low pressures, be supposed to have a limiting value considerably lower than the one determined by equation (5); in fact, hardly more than half that value. We therefore put this limiting value to be

$$U_{m,\, p \to 0} = 7 \cdot 10^5 \sqrt{\frac{z}{M} \cdot V_0}. \tag{6}$$

Fig. 42 in L. F. I shows the results of a series of such measurements. The limiting value found is about $5.2 \cdot 10^7$ cm sec.$^{-1}$ and the *p. d.* used was $V \cong 7000$ volts. By inserting these values in (6) we get

$$\frac{z}{M} = 0.79.$$

For H^+-particles or protons $\frac{z}{M} = 1$, for α-particles this ratio is $\frac{1}{2}$, and for all other known positive particles considerably smaller.

We are therefore led to the supposition that the front-particles are protons.

In the negative figures, where the active outward-travelling particles are electrons, there exists no such limiting value for the velocity at decreasing pressures, or more correctly, the limiting value is so high that it falls outside the range here considered. This is in complete agreement with experiment, which shows that at decreasing pressures the spreading out velocity of the negative figures may attain very high values, see f. inst. fig. 8 above and L. F. I, fig. 38.

3. The Start of Positive and of Negative Discharges.

(a) The Start of Negative Discharges.

The start of a negative figure presents no remarkable features. With a needle-point as electrode placed directly on the photographic plate the range of the figure will, at low potentials, be proportional to the *p. d.*, and this holds right down to a few hundred volts, when the figure becomes so small that it cannot be seen with the naked eye. The explanation is obvious. In the immediate vicinity of the electrode-point the field has, even at small potentials, such a value that the ionization by collision sets in, i. e. figure formation starts. But at these low potentials the electrode must neccessarily be point-shaped or have sharp edges, and must further be clean in the sense defined elsewhere by the author[1]. If the electrode is unclean, greasy, for example, the air does not come in direct contact with the sharp edges or points, and ionization by collision will thus not occur at low potentials.

There is thus no difficulty in explaining the experimental results, that negative figures start most easily from a clean point, that at low potentials their range is proportional to the *p. d.*, and that no lower limit exists for the effective potential.

[1] P. O. Pedersen: Vidensk. Selsk. Math.-fys. Medd. Vol. IV, No. 10. Copenhagen 1922; Ann. d. Physik (VI) Bd. 71, p. 317—376, 1923.

(b) The Start of the Positive Discharges.

The start of a positive figure, however, presents very peculiar features, as appears from chap. III 3 (b). We shall now proceed to discuss these features in the light of the hypothesis already set forth, and further defined in the preceding section, where we stated that the front particles must be protons.

We will for the present assume that these protons originate from hydrogen molecules, and we will put their number N per cm^3 equal to

$$N = N_0 \cdot p, \tag{1}$$

where p is as usual the air pressure.

The volume of that space near the electrode within which the electric field has sufficient strength to split off protons from hydrogen molecules we will call A. The total number of hydrogen molecules available for the possible giving up of protons is consequently

$$n = AN_0 p. \tag{2}$$

The probability s that a hydrogen molecule will give up a proton, depends upon the field, and therefore also upon the potential V. We put

$$s = s(V), \tag{3}$$

and the number n_p of protons available is consequently

$$n_p = sAN_0 p. \tag{4}$$

We shall next discuss under what conditions an available proton becomes active, that is, causes the start of a positive spreader. These conditions are undoubtedly rather complicated. However, it is excluded from the chance of becoming active if it is first rendered “hors de combat” by any means: for example it may adhere to the first neutral molecule it strikes, thus giving rise to a positive ion, or it may be neutralized by an electron from the molecule with which it collides. The conditions that are neccessary to prevent such occurrences cannot be stated with any certainty, but the probability that the proton remains free after a collision is the greater the greater its velocity. We assume for simplicity that the proton remains free if — and only if — its velocity at the collision is greater than that corresponding to V' volts. Denoting the distance travelled by the proton in the direction of the field since its liberation by x, and the electric field strength by E, the proton will remain free after the first collision if

$$Ex > V'. \tag{5}$$

In order to simplify the calculations we take a mean value of E in the active space A by putting

$$E = \frac{V}{L}, \tag{6}$$

where L is proportional to the thickness of the active space.

Then equation (5) will read

$$x > \frac{L}{V} \cdot V'. \tag{5'}$$

The mean value of the free path l of the proton is inversely proportional to the pressure, that is

$$l = \frac{l_0}{p} \tag{7}$$

and the probability S that the first free path of one of the available protons is greater than x is

$$S = e^{-\frac{x}{l}} = e^{-\frac{LV'}{l_0 V} \cdot p} = e^{-c\frac{p}{V}}, \quad c = \frac{L}{l_0} V_0. \tag{8}$$

The total number n^0 of active protons is consequently

$$n^0 = np \cdot S = sAN_0 pe^{-c\frac{p}{V}}. \tag{9}$$

Until anything is stated to the contrary, we shall assume that we have the same electrode and the same kind of gas; in that case A and N_0 are constants and (9) may be written

$$n^0 = sKpe^{-c\frac{p}{V}}. \tag{9'}$$

We will further assume quite provisionally that $s = s(V)$ is a constant. This assumption is, no doubt, not justified, but it will simplify the following considerations, and we shall investigate later the effect of the dependency of s on V.

We therefore write provisionally

$$n^0 = K_0 \cdot p \cdot e^{-c\frac{p}{V}}, \tag{9''}$$

where K_0 is considered a constant.

A positive discharge starts only if $n^0 \geqq 1$. This is the case when

$$\frac{1}{p} \cdot e^{c\frac{p}{V}} \geqq K_0. \tag{10}$$

Here we do not know — and we are unable to calculate beforehand with sufficient accuracy — the value of K_0, but we may try to find out under what conditions the left hand term of (10) is independent of the air pressure. We will write this condition in the form

$$lg\, p - c\frac{p}{V} = \text{constant}. \tag{11}$$

Now the question is: are we able to select such a value of $c = \frac{L}{l_0} V'$ that a (p, V)-curve determined by means of

(11) shows with sufficient accuracy a relation between p and V corresponding to the relation found by experiment? Further, will the value of c applied in such a case be a reasonable one?

In fig. 18 are shown by crosses the corresponding values of potential and air pressure obtained in atmospheric air and with a plate thickness of about 1.4 mm.

The thin curve in fig. 18 represents the equation

$$lg\, p - \frac{12}{V} \cdot p = 3.04 \,. \tag{12}$$

We see that equation (12) does not give any very good representation of the experimental results since — especially at higher pressures — the experimentally determined potentials have somewhat lower values than the calculated potentials. Equation (12), on the other hand, represents values of the right order of magnitude for the relation between p and V over a wide range, and it may, therefore, be of some interest to see whether the value given for c is also of the right order of magnitude. If in (7) the free path is given in cm and the pressure in mm, then $l_0 \cong 2.6 \cdot 10^{-2}$ cm. For electrodes resting directly on the photographic plate the value of L in (6) will presumably be about 0.025 à 0.05 cm. V' must presumably have a value of about 10 volts. The values of c corresponding to this are between 9.6 to 19.2, and these limiting values agree very well with the value 12 in equation (12).

It is obvious why equation (12) can give only a comparatively rough approximation. The probability $s(V)$ that a hydrogen molecule gives up a proton is not constant, as we have assumed in the equations (9″) to (12), but must, on the contrary, increase quite rapidly with increasing values of the potential. However, we do not know this relation in any detail, and for the sake of simplicity we therefore replace (9′) by the following equation:

$$n_0 = sKpe^{-c\frac{p}{V}} = K^0 \cdot p \cdot e^{-c^0\frac{p}{V^2}}. \tag{13}$$

This expression is, no doubt, not theoretically correct but — as compared with (9″) — has the advantage that the value of the probability s increases with increasing value of the potential.

In fig. (18) the heavy curve corresponds to the equation

$$lg\, p - 32 \cdot 10^3 \cdot \frac{p}{V^2} = 2.12\,, \tag{14}$$

and we see that this curve furnishes a representation of the experimental results as satisfactory as may be expected when we take into account the considerable uncertainty in the experimental determination.

Fully in agreement with the considerations here set forth is the fact that the positive figure, at potentials near the critical minimum value, often consists of only one single or a few spreaders. But from this it again follows that it is a matter of chance whether or not a figure is formed with potentials near the critical value.

According to (9) the number of active protons is proportional to the volume of the active space A. It is difficult to state a definite value for this, but it is obvious that of the four different forms of electrodes shown in figs. 13 and 16 the spherical one will have the greatest value of A, the pointed one the smallest value and the rounded rod a value between the two. This explains the fact that within the critical interval the spherical electrode shows a greater tendency to produce figures than the point-electrode. The tube-formed sharp-edged electrode shown in fig. 16, however, shows an ability to produce figures, at least equal to the spherical one. For this electrode, not only is the field very strong at the sharp edge, and consequently the probability *s* great, but also the active space has a considerable volume owing to the considerable length of the edge.

Uncleanness (grease or the like), will cause a slight diminution in the active space immediately around edges and points, but if the uncleanness is only small, its influence in this respect will be insignificant. For positive discharges the electrode will therefore be very little sensitive to contaminations even within the critical interval.

The difference between positive and negative electrodes in sensitiveness to contaminations is quite analogous to the peculiar relations which have been stated with regard to retardation of spark-formation as dependent upon the state of the electrodes[1]. This question will, however, be treated in more detail elsewhere.

In chapt. III 3 (b) we have mentioned that the spherical electrode, within the critical interval, if placed on a not too thick plate, shows a tendency to produce abnormally short spreaders, while sharp-edged electrodes do not show this tendency. This is easily understood from the considerations set forth above, section 2 (c), according to which for spherical electrodes the electric field will have a powerful vertical component at the starting points of the spreaders, which drives the protons down into the plate.

A final consequence of the hypothesis here set forth is that the front-particle must have a certain, rather high velocity if a discharge of the kind here considered shall

[1] P. O. Pedersen: "Vid. Selsk. Math.-fys. Medd." IV, 10. Copenhagen, 1922; "Ann. d. Phys." (IV). Bd. 71, p. 317—376, 1923.

be started. The positive spreaders must consequently attain at all events a certain, finite length, if they start at all.

There then remains the question of the origin of the protons. It is well known that protons are set free in discharge tubes containing even the smallest trace of hydrogen[1], and in many cases also *H*-atoms in considerable numbers[2]. In an apparatus such as that by means of which Lichtenberg figures are produced there will be found — even in cases where the use of pure gases as f. inst. N_2 or O_2 is aimed at — a not insignificant number of hydrogen molecules, and these are also present in considerable numbers in the ordinary air of the laboratory. Even if we reckon with an admixture of only 0.0001 pCt. hydrogen there will be present — at atmospheric pressure — $3 \cdot 10^{13}$ hydrogen molecules per cm^3. Hydrogen molecules in sufficient numbers will consequently always be at hand.

Another feature may possibly seem peculiar at first sight. The formation of positive figures ceases — as we have seen — because the number of spreaders approaches zero. It is comparatively seldom, however, that figures are obtained having only 1 or 2 spreaders, generally there will be f. inst. 5—10. On the other hand, the number will not exceed f. example 30—40, according to the existing conditions, even if we raise the potential or diminish the pressure very greatly. The above calculated probability increases, however, to a very great extent, and accordingly the number of spreaders should increase highly.

There are, however, other causes which automatically set a limit to the number n of the spreaders. Out of a probable number of spreaders, the individual members will

[1] J. J. Thomson: "Rays of Positive Electricity", see f. inst. p. 15—20.
[2] K. F. Bonhoeffer: Erg. d. ex. Naturwiss. Bd. 6, p. 204, 207—8, 1927.

not start quite simultaneously and those first started will consequently come ahead of the others. When so many have started that at some distance from the electrode they cover practically the entire circumference, then — on account of their charge — they will counteract the formation of further new spreaders, or at least stop any such in their growth.

If we are far outside the critical interval, the number of spreaders will therefore not be determined by the available number of protons, but simply by the available space.

At not too high pressures we therefore often observe also a number of small spreaders which are stopped in their growth by the electrical charge of those first formed, see for example plates 12 V—VII and IX, 13 I—II and 14 I—II.

According to the hypothesis set forth, it may be expected that the number of available protons will be the greatest in hydrogen or in compounds of hydrogen and nitrogen, smaller in nitrogen, still smaller in compounds of nitrogen and oxygen and the smallest in gases having great electron-affinity as for example oxygen.

From the foregoing it will be understood that this difference will not make itself apparent by the number of spreaders of the normal figures, but it manifests itself quite clearly by the number of side-branches and smaller side-spreaders. The greater the number of available protons the greater number of these will be observed. A look over the figures shown in plates 4, 5 and 6 I, which are taken in various gaseous compounds will verify this fact.

The small side-spreaders and branches are presumably formed in this manner: The original spreaders, which are in the process of formation, act as electrodes with a potential somewhat lower than that of the main electrode, but still of a value sufficient to start available protons with the necessary velocity.

Inside the spreaders protons will generally be set free to a great extent and these will be driven by the electric field to the surface of the spreader, so that here there will always be available protons. Spreaders will, however, be formed only if the electric field in the surrounding air is of sufficient value to give the protons the necessary velocity. The field outside in the air will be strongest at the edge of the spreader immediately on the surface of the photographic film. The protons driven out here will therefore also have the greatest probability of becoming active. If the spreader is formed in a narrow space between two planes — for example if there is a second photographic plate as shown in plate 8 I—II — then the lateral spreading out will be small but both the field as well as the proton-density will be great at the edges and therefore the tendency to form side-spreaders will be great.

If the side-spreaders start in directions which are nearly the same as that of the original spreaders, they will be squeezed in between these, and if they start perpendicular to the original ones, they will become very short as they are stopped by the neighbouring spreaders. Such short side-spreaders are observed in great numbers in hydrogen-nitrogen compounds. (In pure hydrogen this phenomenon is very difficult to study because the luminosity is so small). In some cases these side-spreaders may point inward — in the direction of the electrode — if there is more free space in that direction, see for example plate 5 XVI.

In cases where the spreaders are stopped abruptly by an opposing field as f. inst. at the meeting line in velocity-measurements, the potential at the tip of the spreader will become comparatively high and the tendency to form side-spreaders will therefore also be great. This conclusion is also verified by experiments, see for example plate 14 I, especially just above the dividing line some little distance outside both ends of the lower electrode.

The hypothesis set forth thus explains in a simple manner all the peculiar features of the start of the positive figures.

4. Conductivity of Positive and of Negative Spreaders.

(a) Conductivity of Negative Spreaders.

In fig. 29, *A* represents a negative electrode resting on a photographic plate *P*, while *a b* represents, schematically, a negative spreader which at the point *b* runs into a free — or earthed — electrode *B*, the potential of which is zero to begin with.

There will be present in the spreader positive ions as well as negative ions and electrons. The sum of the two last named will be greater than the number of positive ions because the spreader has a negative charge. Since the electrons have a much greater mobility than the ions only electrons are shown in I and II of fig. 29.

We assume the conductivity of the spreaders to be so high that at a point just a little to the left of *b* there is a potential which is a considerable fraction of the value of the potential of *A*. There will consequently be a strong field immediately near *b* and this field will pull the electrons toward and into *B*. This results in a decrease in the negative charge near *b* as indicated at *c* in II. A strong field will then be developed near *c*, which will pull a further number of electrons in the direction of *B*. The location of the strong field will therefore — owing to the great mobility of the electrons — move to the left with great velocity. This is indicated by the double arrow at *c*.

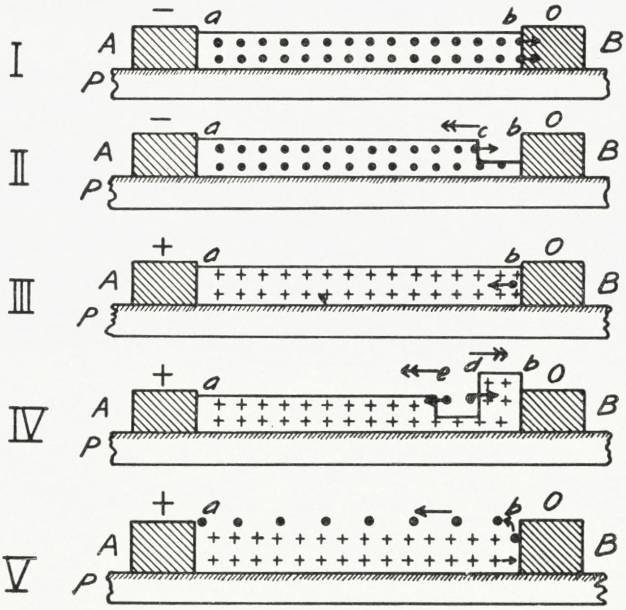


Fig. 29. Schematical Representation of the Conductivity of Negative Figures (I—II) and of Positive Figures (III—IV).

The strong field at *b* and later at *c* will cause ionization by collision, by which means the conductivity between *c* and the electrode *B* will be comparatively high, even though the negative charge between these two points will be comparatively small. From what is stated here, it appears that there will be a strong tendency to spark-formation between *A* and *B*, and the spark will begin at *B*. This is verified by our experiments, see chapt. III 4 (a).

(b) Conductivity of Positive Spreaders.

Here we find entirely different conditions. Tbe spreader *a b* in III fig. 29 has a positive charge, the number of positive ions being greater than the sum of electrons and negative ions. The electrons are mainly present in the upper part of the spreader.

When the spreader *a b* gets near to *B* there will arise a strong field at *b*, which will drive the positive ions into the electrode *B*. Owing, however, to the low mobility of the ions they will move comparatively very slowly, and we will therefore consider them as insignificant at present. We might perhaps assume that the strong field at *b* would drive the electrons with great velocity towards *A*. Such a movement would, however, very soon be stopped. If f. inst. most of the electrons in the space *b d* in IV are carried over to the space *d e* then this would give a charge distribution as indicated in the figure. But in such a case those of the electrons within the space *d e* which were near *d* would be carried back again with great force toward *B*, while those near *e* would be carried along in the direction of *A*, but with a smaller force. An "electron-wave" such as *d e* will therefore be dispersed and cannot travel from *B* to *A*. Consequently no spark can be formed.

A flow of electrons may, however, very well occur from *B* to *A* i. e. an electric current passing from *A* to *B*, provided only that a suitable number of electrons are set free at *B*, and such a current may be quite strong without actually forming a spark. The strong field at *b* will cause the positive ions to impinge on *B* with great velocity and thus set free a sufficient supply of electrons. These will then, as shown in V of fig. 29, drift to the surface of the spreader and here be carried toward the electrode *A*.

If the *p. d.* between *A* and *B* is not of sufficient value to cause this current to ionize by collision, no light is emitted and no spark is initiated.

If, finally, we maintain the *p. d.* between *A* and *B* so long that the positive ions get time to move sufficient distances, then we get a discharge analogous to the one shown in fig. 28 II, except that here the moving particles are positive ions instead of electrons and in consequence the discharge moves relatively slowly.

The features explained here are in full agreement with the experimental results set forth in chapt. III 4.

A theoretical treatment of the discharge conditions in the case when the potential is maintained long enough to allow the heavier ions to play a deciding rôle in the formation of the discharge, f. inst. the question of spark formation, will be taken up elsewhere. We have mentioned these conditions in chap. III 4 (b) and (c) only to emphasize that they have nothing to do with the formation of the regular Lichtenberg figures.

5. Various Circumstances in Connection with the Formation of the Lichtenberg Figures.

(a) No Influence of an Initial Ionization.

It will be understood without further explanation that an initial ionization has no appreciable influence on the spreading-out of the positive figures as long as this ionization is not strong enough to give rise to a conductivity which will materially alter the electric field active in the formation of the figure. The hypothesis thus gives a perfectly satisfactory explanation of the features mentioned in chap. III 5 (a).

(b) The Bright Border-Line.

The bright rim-formations mentioned in chap. III 5 (b) are easily explained by means of what is set forth in 1 (d) above and in Appendix 2.

In the case of negative figures the electrons will, as is mentioned above, continue their outward movements even after the field at the edge has decreased so much that no ionization by collision takes place. The said electrons will then, however, generally combine with neutral molecules into heavier negative ions, and the outermost edge of this ion-stream will, as pointed out in Appendix 2, become comparatively sharp.

If now, simultaneously with this outward movement, a great number of electrons are set free over the area in question, then ionization by collision will occur at the sharp edge of the figure, and this ionization will cause the emission of light a little outside the edge which, owing to the former mentioned reasons, is situated a little outside the boundary of the photographic image.

Somewhat similar conditions occur in the case of positive figures, except that here the positive ions are so firmly placed that no appreciable spreading out of the figure will occur after the formation of the photographic image has stopped. A rim-formation will in this case occur close to the edge of the photographic image.

Concluding Remarks.

A theory of the formation of negative figures, which is in the main satisfactory has — as stated in the introduction — already been given earlier. When we have also treated of negative figures in the investigations described here our object has been to throw further light upon certain features of the theory already set forth.

The theory developed above of the formation of positive Lichtenberg figures, namely, that formation of positive spreaders is due to protons which the strong field at the tip of the spreaders drive outwards with great velocity, by which means electrons are set free in sufficient number to initiate a sudden and strong ionization by collision which again sets free electrons in sufficient numbers necessary to carry away the charge towards the electrode is found to explain throughout in a satisfactory manner the many peculiar features presented by the positive spreaders.

The investigations here described further indicate that protons play an important rôle not only in the formation of positive Lichtenberg figures, but that their importance in connection with spark formation is much greater than hitherto assumed. None of the theories so far proposed for spark formation offer a satisfactory explanation of a number of peculiar spark phenomena which have been observed and pointed out during recent years[1]. The theory of the formation of the positive figures given here will presumably also be useful for the solution of those problems, but this question will be treated at a later occasion.

Presumably, also, protons play a more important rôle than has hitherto been assumed in a number of other discharge phenomena.

Finally I wish to express my cordial thanks to Carlsberg Fondet for its valuable support of these investig-

[1] P. O. PEDERSEN: The papers (a)—(e) mentioned p. 4.

ations. To the Rask-Ørsted Fond for contributing to the publication. And also to the various cooperators who have kindly assisted me in carrying out the experimental investigations: namely, Mr. J. P. Christensen, Mr. Chr. Nyholm and Mr. B. B. Rud who have taken most of the photographic Lichtenberg figures here used. A few of the pictures were taken by Mr. C. Schou and Mr. N. E. Holmblad while Mr. F. Nielsen has carried out the photographic enlargements. Mr. J. P. Christensen has further assisted me in preparing this publication for the press.

APPENDIX I

On Positive Particles.

Within the velocity range — $1 \cdot 10^7$ to $10 \cdot 10^7$ cm sec^{-1} — which is of particular interest for positive Lichtenberg figures only very little has been published about investigations of how the range, the ionization, and the charge etc. etc. of the positive particles depend upon the velocity.

For particles at somewhat higher velocities a number of investigations have been carried out and these we have mentioned below under sections 1—5 in so far as they are judged to be of interest for the understanding of the positive Lichtenberg figures. Finally, A. J. DEMPSTER has recently carried out investigations with H^+ particles in hydrogen and helium at velocities within the said range, and these and a few other investigations we will discuss under section 6.

1. Range, Ionization and Velocity of Positive Particles.

The range R of high-speed α-particles is known[1] to be proportional to the third power of the velocity U that is

$$U^3 = K \cdot R = 1.076 \cdot 10^{27} R \quad (1)$$

This relation holds good as long as the range is more than 2 cm *i. e.* at velocities above $1.3 \cdot 10^9$ cm sec.$^{-1}$.

At small ranges and velocities the range is, on the con-

[1] H. GEIGER: Proc. Roy. Soc. (A) Vol. 83, p. 505—515, 1910. E. MARSDEN and T. S. TAYLOR: Proc. Roy. Soc. (A) Vol. 88, p. 443—454. 1913. GEIGER und SCHEEL: Handbuch d. Phys. Bd. 24, p. 152, 1927.

trary, proportional to the velocity in the power $\frac{3}{2}$ as shown f. inst. by P. M. S. Blackett[1]. That is

$$U^{\frac{3}{2}} = aR, \tag{2}$$

which holds good with sufficient accuracy for ranges from 1 cm down to somewhat less than 1 mm.

The equation (2) may also be written as

$$R = b \cdot V_0^{\frac{3}{4}}, \tag{2'}$$

where V_0 is the *p. d.* through which the particle must fall to obtain the velocity U. The constant b has the value

$$b = \frac{1}{a}\left(2\frac{e'}{m}\right)^{\frac{3}{4}}, \tag{2''}$$

where m is the mass of the particle and e' its charge.

Equation (2) holds good — as shown by Blackett — not only for α-particles but also for positive atomic ions of hydrogen, argon and "atmospheric air". The corresponding values of a and some other factors of importance for the following — all based on the Blacketts measurements — are set up in the following table A.

Table A. Various properties of positive particles.

Particle	a	$\frac{a}{a_H}$	Atomic weight m	$\sqrt{m}$	Average loss of energy per unit length of track $\frac{\frac{m}{R}}{\frac{m_H}{R_H}}$	Relative charges $\eta = \frac{\sqrt{\frac{m}{R}}}{\sqrt{\frac{m_H}{R_H}}}$
H	$6.2 \cdot 10^{13}$	1.0	1.0	1.0	1.0	1.0
He (α-Part)	$3.3 \cdot 10^{13}$	1.9	4.0	2.8	2.1	1.5
"Air"	$1.94 \cdot 10^{13}$	3.2	14.4	3.8	4.5	2.1
A	$1.21 \cdot 10^{13}$	5.1	40	6.3	7.8	2.8

[1] P. M. S. Blackett: Proc. Roy. Soc. (A) Vol. 103, p. 62—78, 1923.

The table contains also the average value of the relative average loss of energy suffered by the various particles per cm of the track, which loss BLACKETT takes to be proportional to $\frac{m}{R}$. Further BLACKETT assumes this loss to be proportional to the number of pairs of ions set free per cm, and this number again with approximation proportional to the square of the effective charge of the particle. The latter he therefore takes to be proportionat to $\sqrt{\frac{m}{R}}$, and the value of this quantity is given in the last column of the table. As the charge of the particles is by no means constant over the entire track this assumption may not be quite justified but will nevertheless presumably lead to an approximately correct result.

If the average energy neccessary to set free one pair of ions is w ergs, and if the entire kinetic energy of the particle is spent in the formation of ions, then — denoting the number of pairs of ions set free per cm by I — we have the following relations:

$$wIdR = \frac{1}{2} m \left[U^2 - (U - dU)^2\right] = mUdU.$$

or

$$wI = mU \frac{dU}{dR}, \tag{3}$$

but according to (2) we have

$$\frac{dU}{dR} = \frac{2a}{3U^{\frac{1}{2}}}. \tag{4}$$

From the equations (2)—(4) we get

$$I = \frac{2ma}{3w} \cdot U^{\frac{1}{2}} = \frac{2mU^2}{3wR} = \frac{4}{3} \cdot \frac{V_0}{RV'} = \frac{4}{3} \cdot \frac{V_0^{\frac{1}{4}}}{b \cdot V'}, \tag{5}$$

where V' is the ionization-voltage of the gas in which the particle moves.

The measurements carried out by Blackett unfortunately do not go down to such small velocities and ranges as those in which we are directly interested, but until further we will assume that the law expressed by equation (2) is also applicable to velocities from $2 \cdot 10^7$ to $10 \cdot 10^7$ cm sec.$^{-1}$. In order to give an idea of the magnitude of

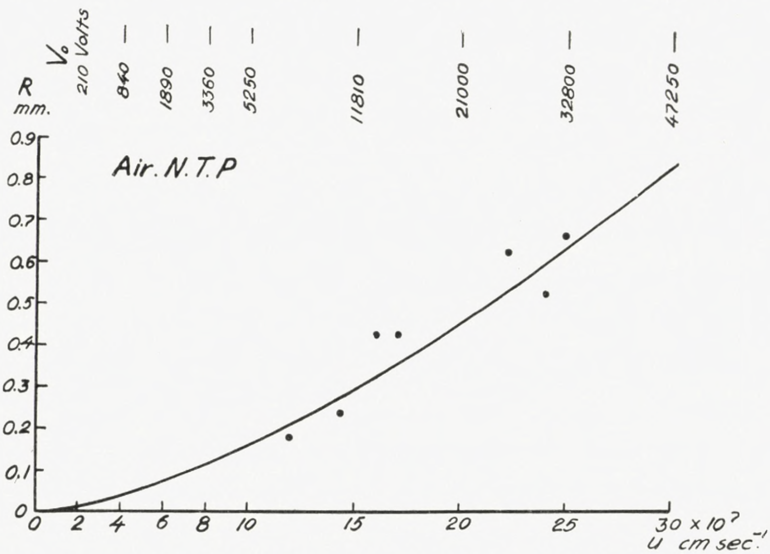


Fig. 1. Range of H^+ particles in Air N. T. P. according to Blackett.

the extrapolation thereby performed we have in fig. 1 marked by dots the corresponding values of velocity and range found by Blackett for H^+ particles in atmospheric air, while the curve shown has the equation

$$U^{\frac{3}{2}} = 6.2 \cdot 10^{13} R.$$

2. Velocity in Strong Fields.

Further we have

$$U = \sqrt{2 \frac{e'}{m} V_0} = c\sqrt{V_0}. \qquad (6)$$

From this expression and from equation (2) follows

$$\frac{dV_0}{dR} = X = \frac{2\,V_0^{\frac{1}{2}}}{c} \cdot \frac{dU}{dR} = \frac{4\,a\,V_0^{\frac{1}{2}}}{3\,c\,U^{\frac{1}{2}}} = \frac{4\,a}{3\,c^2} \cdot U^{\frac{1}{2}}. \qquad (7)$$

Here X is the intensity of the electric field neccessary and sufficient to maintain the velocity U of the particle, assuming that the particle all the time has the charge e'.

In the following table B is shown among other things the value of c in equation (6) when V_0 is measured in volts for the particles investigated by BLACKETT. The figures given for α-particles refer actually to a helium atom with one positive charge.

Table B. Various properties of positive particles in air.

Particle	c	c^2	a	$\frac{a}{c^2}$	Relative Charge η	$\xi = \frac{4}{3\eta} \cdot \frac{a}{c^2}$	X for $U = 2 \cdot 10^7$ cm sec^{-1}
$H+$	$13.8 \cdot 10^5$	$190 \cdot 10^{10}$	$6.2 \cdot 10^{13}$	32.6	1.0	43.5	$195 \cdot 10^3$ volt cm^{-1}
α-Part (H_e^+)	$6.9 \cdot 10^5$	$47.5 \cdot 10^{10}$	$3.3 \cdot 10^{13}$	69.6	1.5	61.8	$259 \cdot 10^3$ volt cm^{-1}
(Air)+	$3.6 \cdot 10^5$	$12.9 \cdot 10^{10}$	$1.9 \cdot 10^{13}$	148	2.1	94.2	$442 \cdot 10^3$ volt cm^{-1}
$A+$	$2.2 \cdot 10^5$	$4.6 \cdot 10^{10}$	$1.2 \cdot 10^{13}$	263	2.8	125	$559 \cdot 10^3$ volt cm^{-1}

ξ indicates the factor by which $U^{\frac{1}{2}}$ is to be multiplied in order to obtain the intensity X of the electric field necessary to maintain the velocity of the particle in question in the gas in question, which latter in the above table is air at N. T. P. It must, however, be remembered that for this calculation of X it is assumed that the charge of the H-particle is equal to $+e$ over the entire track.

3. The Mean Value of the Charge of the Hydrogen Atom.

If l_1 is the average length of those parts of the track (in the direction of the force) along whick the H-particle has the charge $+e$, i. e. is a proton, while l_2 is the corresponding average length of track along which the H-particle is a neutral hydrogen atom — and its charge consequently is zero — then the values found for X must evidently be multiplied by the factor

$$\mu = \frac{l_1 + l_2}{l_1} = 1 + \frac{l_2}{l_1}. \tag{8}$$

For the value of $\frac{l_1}{l_2}$ there exist a number of rather contradictory determinations. In the following we will apply only those given by E. Rüchardt[1]. Fig. 2 shows Rüchardt's

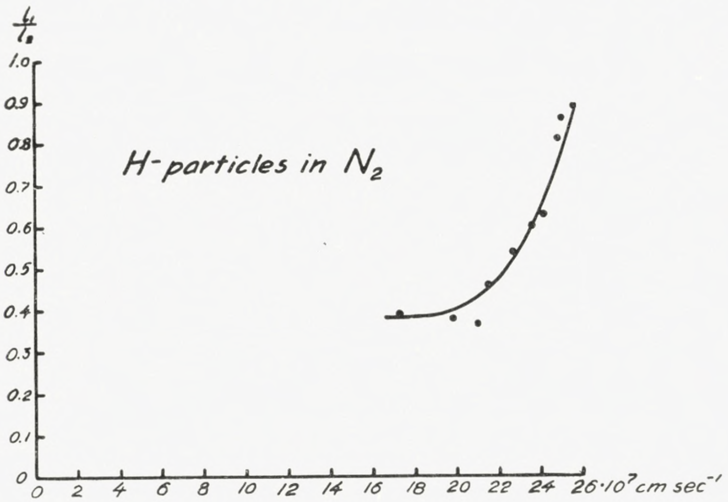


Fig. 2. Value of $\frac{l_1}{l_2}$ for H-particles in N_2 according to Rüchardt.

results with H-rays in N_2. For H-particles in 0_2 it seems that $\frac{l_1}{l_2}$ is somewhat smaller. Unfortunately none of the measurements deal with such relatively small velocities as those with which we have to do in the case of the Lichtenberg figures. An extrapolation must therefore be very uncertain. But Rüchardt's results seem to indicate

[1] E. Rüchardt: Ann. d. Phys. (IV). Bd. 71, p. 377—423, 1923.

that for H-particles in both H_2 and N_2 $\frac{l_1}{l_2}$ has decreased to its minimum value at velocities about $1.7 \cdot 10^8$ cm sec.$^{-1}$. Since no other information on this point is available, we will for the present assume that $\frac{l_1}{l_2}$ is constant at velocities from 10^7 to 10^8 cm sec.$^{-1}$ in which we are here interested, and we will put, for H-particles in atmospheric air

$$\frac{l_1}{l_2} = 0.4. \tag{9}$$

Although this value for $\frac{l_1}{l_2}$ is a little higher than the one found for N_2 and still more so in comparison to the one determined by Rüchardt for O_2, we have selected it because in all the cases in which we are particularly interested the particle will move in an exceedingly strong electric field which will tend to increase the relative velocity between the H-particle and the electrons set free by its collisions. This increase in relative velocity will decrease the probability that an electron is again captured by an H^+-particle, and consequently l_1 will be greater than it is without such a strong field, compare Rüchardt's work dealing with this question.[1]

4. The Value of the Ratio $\frac{l_1}{l_2}$ depends upon the Intensity of the Electric Field.

An exact treatment of this question is not possible at present owing to insufficient knowledge of the ionization — and recombination — processes. But in order to form an idea of the influence of a strong field on these pro-

[1] E. Rüchardt: Zeitschr. f. Phys. Bd. 15, p. 164—171, 1923. E. Rüchardt: Ann. d. Phys. (IV). Bd. 73, p. 228—236, 1924.

cesses we will subject the matter to some simple considerations based upon the ideas outlined by RÜCHARDT.

At first we assume the electric field to be zero. The particle having the charge $+e'$ moves with the velocity U_0 while the electrons are assumed to be at rest. With regard to the question of recombination we may as well consider the particle to be at rest and the electrons to move with the velocity U_0 relatively to the particle. The electrons that in a given moment are set free with the velocity zero — relative to the particle consequently with the velocity U_0 — and that are located around the particle within a sphere having the radius ϱ_0, will all move in eliptical orbits around the particle provided that ϱ_0 satisfies the following conditions.

$$\frac{1}{2} m U_0^2 = \frac{e \cdot e'}{\varrho_0}, \tag{10}$$

m denoting the mass of the electrons and $-e$ their charge.

Electrons set free outside this sphere will move in hyperbolic orbits and will consequently not be captured by the particle, while all those set free inside the sphere will be captured.

If we put $e = e' = 4.774 \cdot 10^{-10}$ E. S. E. and $m = 9 \cdot 10^{-28}\, g$ then

$$\varrho_0 = \frac{5.05 \cdot 10^8}{U_0^2}. \tag{10'}$$

For

$U_0 = 1 \cdot 10^7 \quad 2 \cdot 10^7 \quad 3 \cdot 10^7 \quad 4 \cdot 10^7 \quad 5 \cdot 10^7$ cm sec^{-1}

we get

$\varrho_0 = 5.05 \cdot 10^{-6} \quad 1.26 \cdot 10^{-6} \quad 5.6 \cdot 10^{-7} \quad 3.16 \cdot 10^{-7} \quad 2.02 \cdot 10^{-7}$ cm.

We will next consider the case where a constant field X (E. S. E.) acts parallel to the velocity U_0. Here an exact treatment is very difficult, we will therefore take only a simple approximation.

Referring to figure 3, the particle is at a given moment located at P and an electron is set free at A. To begin with the electron has — in relation to the particle — the velocity U_0 in the positive direction of the X-axis. It is further acted upon by the electric field X which tends to

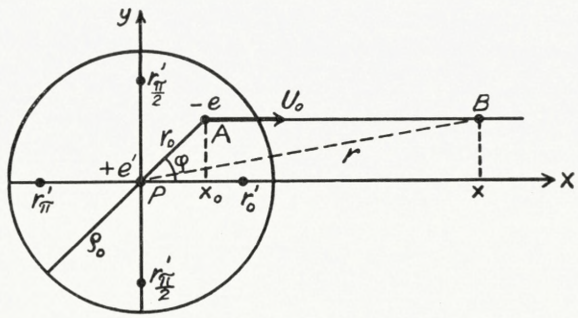


Fig. 3. Recapture of Electrons in Strong Electric Field.

increase the relative velocity. We now assume that the electron is continually moving out along the line AB which is parallel to the X-axis. This is of course not correct, since the track will be curved. But we are here not interested in the shape of the track itself but only whether the electron will infinitely continue its movement away from the particle, or return to its vicinity. In the latter case we consider it to be captured, in the first case not.

If the starting point A were located on the x-axis then the electron would remain there and would only be captured if at some distance or other its velocity decreases to zero. If this occurs then the electron will return to the particle, in the reverse case it would travel away. The conditions are quite analogous if the electron is bound to move along the line AB.

At the point B the velocity U_x of the electron is determined by

$$\frac{1}{2} m U_x^2 = \frac{1}{2} m U_0^2 - \frac{ee'}{r_0} + \frac{ee'}{r} + eX(x - x_0)$$

$$\cong \frac{1}{2} m U_0^2 - \frac{ee'}{r_0} + \frac{ee'}{x} + eX(x - x_0)$$

$$= ee'\left(\frac{1}{\varrho_0} - \frac{1}{r_0} + \frac{1}{x}\right) + eX(x - x_0). \tag{11}$$

The velocity U_x will be zero for $x = x'$ if

$$e'\left(\frac{1}{\varrho_0} - \frac{1}{r_0} + \frac{1}{x'}\right) + X(x' - x_0) = 0 \tag{12}$$

or when

$$x'^2 - x'\left[x_0 + \frac{e'}{X}\left(\frac{1}{r_0} - \frac{1}{\varrho_0}\right)\right] + \frac{e'}{X} = 0 \tag{13}$$

and consequently for

$$x' = \frac{1}{2}\left[x_0 + \frac{e'}{X}\left(\frac{1}{r_0} - \frac{1}{\varrho_0}\right)\right] \pm \sqrt{\frac{1}{4}\left[x_0 + \frac{e'}{X}\left(\frac{1}{r_0} - \frac{1}{\varrho_0}\right)\right]^2 - \frac{e'}{X}}. \tag{14}$$

A necessary condition for the velocity to become zero at all is that the value of x' as determined by (14) is real, that is, we must have

$$x_0 + \frac{e'}{X}\left(\frac{1}{r_0} - \frac{1}{\varrho_0}\right) \geqq 2\sqrt{\frac{e'}{X}}. \tag{15}$$

In the limiting case, to which the lowest sign of equation (15) applies, we have

$$x' = \sqrt{\frac{e'}{X}} \quad \text{or} \quad X = \frac{e'}{X'^2}, \tag{16}$$

so that the electric force in the point x' is zero. If the velocity has not gone quite to zero when the electron has attained this point then the velocity can never be zero.

Since $x_0 = r_0 \cos\varphi$ then the equality in (15) gives the following equation for the determination of the maximum value r' of r_0, for which the electron will return to the vicinity of the particle:

$$r'^2 - r'\frac{2}{\cos\varphi}\left(\sqrt{\frac{e'}{X}} + \frac{e'}{2\varrho_0 X}\right) + \frac{1}{\cos\varphi}\cdot\frac{e'}{X} = 0 \qquad (17)$$

and therefore

$$r' = \left[\frac{1}{\cos\varphi}\cdot\left(1 + \frac{1}{2\varrho_0}\sqrt{\frac{e'}{X}}\right) \pm\sqrt{\frac{1}{\cos^2\varphi}\left(1+\frac{1}{2\varrho_0}\sqrt{\frac{e'}{X}}\right)^2 - \frac{1}{\cos\varphi}}\,\right]\cdot\sqrt{\frac{e'}{X}}. \qquad (18)$$

For $\varphi = 0$ we have

$$r' = \left[1 + \frac{1}{2\varrho_0}\sqrt{\frac{e'}{X}} - \sqrt{\frac{1}{\varrho_0}\sqrt{\frac{e'}{X}} + \frac{1}{4\varrho_0^2}\cdot\frac{e'}{X}}\,\right]\cdot\sqrt{\frac{e'}{X}} = r'_0,$$

and for $\varphi = \frac{\pi}{2}$ we have

$$r' = \frac{\frac{1}{2}\sqrt{\frac{e'}{X}}}{1 + \frac{1}{2\varrho_0}\sqrt{\frac{e'}{X}}} = r'_{\frac{\pi}{2}}, \qquad (19)$$

and for $\varphi = \pi$ we have

$$r' = \left[-\left(1 + \frac{1}{2\varrho_0}\sqrt{\frac{e'}{X}}\right) + \sqrt{2 + \frac{1}{\varrho_0}\sqrt{\frac{e'}{X}} + \frac{1}{4\varrho_0^2}\cdot\frac{e'}{X}}\,\right]\cdot\sqrt{\frac{e'}{X}} = r'_\pi.$$

The space within which the electrons return and are captured may then with approximation be reckoned to be a sphere having the radius ϱ' determined by

$$\varrho' = \frac{1}{4}\left(r'_0 + 2\,r'_{\frac{\pi}{2}} + r'_\pi\right). \qquad (20)$$

The electric field has thus decreased the space within which capturing occurs in the ratio

$$\nu = \left(\frac{\varrho'}{\varrho_0}\right)^3. \tag{21}$$

In table C below are given values of ν calculated for a series of velocities from $1 \cdot 10^7$ to $1 \cdot 10^8$ cm sec.$^{-1}$ assuming there is a field intensity $X = 10^3$ E. S. U. $= 3 \cdot 10^5$ Volt cm^{-1}.

Table C. Values of ν for various velocities and for $X = 3 \cdot 10^5$ volts cm^{-1}; $\sqrt{\frac{e'}{X}} = 6.9 \cdot 10^{-7}$ cm.

U_0	ϱ_0	r'_0	$r'_{\frac{\pi}{2}}$	r'_{π}	ϱ'	$\nu = \left(\frac{\varrho'}{\varrho_0}\right)^3$
	cm	cm	cm	cm	cm	
$1 \cdot 10^7$	$5.05 \cdot 10^{-6}$	$4.82 \cdot 10^{-7}$	$3.23 \cdot 10^{-7}$	$2.70 \cdot 10^{-7}$	$3.50 \cdot 10^{-7}$	0.00033
$2 \cdot 10^7$	$1.26 \cdot 10^{-6}$	$3.34 \cdot 10^{-7}$	$2.71 \cdot 10^{-7}$	$2.39 \cdot 10^{-7}$	$2.79 \cdot 10^{-7}$	0.011
$3 \cdot 10^7$	$5.6 \cdot 10^{-7}$	$2.38 \cdot 10^{-7}$	$2.14 \cdot 10^{-7}$	$1.96 \cdot 10^{-7}$	$2.15 \cdot 10^{-7}$	0.057
$4 \cdot 10^7$	$3.13 \cdot 10^{-7}$	$1.72 \cdot 10^{-7}$	$1.63 \cdot 10^{-7}$	$1.52 \cdot 10^{-7}$	$1.62 \cdot 10^{-7}$	0.137
$5 \cdot 10^7$	$2.02 \cdot 10^{-7}$	$1.31 \cdot 10^{-7}$	$1.27 \cdot 10^{-7}$	$1.24 \cdot 10^{-7}$	$1.27 \cdot 10^{-7}$	0.262
$1 \cdot 10^8$	$5.05 \cdot 10^{-8}$	$4.83 \cdot 10^{-8}$	$4.43 \cdot 10^{-8}$	$4.14 \cdot 10^{-8}$	$4.14 \cdot 10^{-8}$	0.69

For small velocities, l_1 must accordingly be many times smaller with the strong field than without such a field, while l_2 will presumably be nearly independent of the field. This relation will, however, hardly be so pronounced as appears from the above table, because the density of the electrons will presumably be the greatest near the particle. In the table we have reckoned with a uniform density of released electrons.

A H^+-particle will presumably need less energy in order to penetrate the air than a H-atom. Since the strong field increases the value of $\frac{l_1}{l_2}$ it tends to reduce the energy necessary to maintain a certain velocity for a H-particle as compared to the value arrived at by calculations based upon BLACKETT's investigations where no field was employed.

5. Relation between the Intensity of the Electric Field and the Velocity of Positive Particles.

According to table B and the formulae (8) and (9) the field intensity necessary to maintain the velocity $U_0 = 2 \cdot 10^7$ cm sec.$^{-1}$ should be

$$X = 195\,000 \left(1 + \frac{1}{0.4}\right) = 683\,000 \text{ volts cm}^{-1}. \qquad (22)$$

Considering the influence of the strong field on the ability of the H^+-particle to capture electrons it will presumably be reasonable to reduce the value given for the necessary field intensity, and as a rough estimate we put it at

$$X = 300\,000 \text{ volts cm}^{-1} \text{ for } U_0 = 2 \cdot 10^7 \text{ cm sec.}^{-1}. \quad (23)$$

Although this value is only based upon a rough estimate it may presumably be of the right order of magnitude.

6. Collisions between Slow Positive Particles and neutral Molecules.

It may certainly be considered doubtful whether positive ions having a velocity less than that corresponding to 20—30 volts are able to ionize common gases at all. The ionizing ability of these slow positive ions is at all events extremely small[1]. W. J. HOOPER (l. c.) is furthermore of the opinion that even at velocities corresponding

[1] J. FRANCK und P. JORDAN: Handb. d. Phys. Bd. 23, p. 730—733, 1926. W. J. HOOPER: Phys. Rev. (II) vol. 27, p. 109, 1926.

to 925 volts, positive ions in hydrogen at low pressures (0.012 mm) produce very little or no ionization at all, while at this velocity the ionization seems to be quite considerable at higher pressures.

From experiments carried out by W. Aich[1] it appears that the cross-sectional area of hydrogen molecules, determined for movements of protons in hydrogen, is very nearly the same as the gas-kinetic cross-sectional area if the proton is considered to have infinitesimal dimensions.

Similar relations presumably exist in connection with slow protons in other gases.

From investigations with protons having a velocity corresponding to about 900 volts A. J. Dempster[2] draws the conclusion that the effective cross-sectional area of hydrogen- and helium-molecules in case of collisions with protons of this velocity is very nearly equal to zero, and the protons therefore — just as electrons — should show the Ramsauer-effect if their velocity is about $4 \cdot 10^7$ cm sec.$^{-1}$. In that case also, no alternations should occur in the charge of the proton.

It will hardly be possible to take a definite stand-point with regard to these questions at the present time.[3] There is at all events hardly any reason to assume that nitrogen, oxygen, carbonic acid, and other gases that do not show the Ramsauer-effect in connection with electrons, should do so in connection with protons, and what we have set forth above in Chapt. IV indicates, in our opinion definitely, that this will not be the case.

[1] W. Aich: Zeitschr. f. Phys. Bd. 9, p. 372—378, 1922.

[2] A. J. Dempster: Proc. Nat. Acad. Sc. Amer. vol. 11, p. 552—554, 1925; vol. 12, p. 96—98, 1926.

[3] Compare for example: E. Rüchardt in Handb. d. Phys. Bd. 24, p. 99—101, 1927 and W. Wien: Handb. d. Experimentalphysik Bd. 14, p. 527, 1927.

The conditions are different with the noble gases which we have had no opportunity to investigate in a perfectly pure state. Here our investigations are not decisive, but some observations in connection with helium indicate that the results found for this gas by Dempster are correct.

7. Resume.

Even though the material of experimental results at hand is very incomplete we are, no doubt, on the safe side when we assume that:

At velocities which are smaller than that corresponding to 20—30 volts, the positive ions — including the protons — will only in an extremely small degree have an ionizing effect on those gases referred to in the investigations here discussed, while the molecules of those gases, for collisions with positive ions, show nearly the same cross-sectional area as they should have according to the kinetic gas-theory.

At velocities corresponding to more than 10 000 volts, any kind of positive ions have a strongly ionizing effect in any kind of gas.

At velocities corresponding to more than a few hundred volts all positive ions — except the protons — will have a strongly ionizing effect in any kind of gas.

At velocities corresponding to more than a few hundred volts the protons will also have a strongly ionizing effect in all gases except the noble ones and possibly hydrogen. In the noble gases — and possibly in hydrogen — the protons do not cause ionization up to velocities corresponding to about 900 volts.

APPENDIX II

1. Dispersion of Positive or Negative Charges.

We put the velocity v of the ions equal to

$$v = kE, \tag{1}$$

where E is the electric field intensity and where k may be considered to be constant within very wide limits; for example for air at N. T. P. from $E = 0.1$ to $E = 20\,000$ volts cm.^{-1}.[1] Within very wide limits, k is inversely proportional to the air pressure p, so that we may write

$$k = \frac{k_0}{p}, \tag{1'}$$

where k_0 is a constant independent of the air pressure.

If we are not taking into account the mutual repulsion between the single particles, then the density of these will not be varied by their movement in an outer electric field. To understand this, we may just consider a space-element, the outer boundaries of which are a thin tube of lines of force representing the outer field, and two equipotential surfaces (1) and (2) having the areas f_1 and f_2 (see fig. 1). The electric field at these surfaces are respectively E_1 and E_2. We then have:

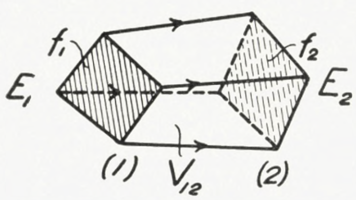


Fig. 1. Element of space having for boundaries a tube of lines of force as shown, and the two equipotential surfaces having the areas f_1 and f_2.

[1] See L. B. Loeb: "Kinetic Theory of Gases", p. 440, 1927.

$$f_1 E_1 = f_2 E_2. \tag{2}$$

During the time dt the volume of the element of charge at the area (1) will, owing to displacement of the charge be reduced by $f_1 v_1 dt = f_1 k E_1 dt$, and at the area (2) be increased by $f_2 v_2 dt = f_2 k E_2 dt$, but since according to equation (2) $f_1 v_1 = f_2 v_2$ the volume of the element of charge will remain unaltered. Variations in the volume — and consequently also in the density — can therefore be due only to the repulsive forces within the element of charge in question, and we shall therefore proceed to discuss the effect of this repulsion.

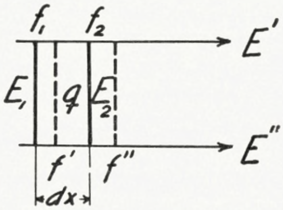


Fig. 2. E' and E'' denotes lines of force; f_1 and f_2 equipotential surfaces.

The two equipotential surfaces of which f_1 and f_2 are sections, are separated by the infinitesimal distance dx, and for the areas f_1 and f_2, circumscribed by the tube of lines of force, we therefore have that $f_1 = f_2$. During the time dt, f_1 covers the volume $f_1 v_1 dt$ while f_2 covers the volume $f_2 v_2 dt = f_1 v_2 dt$.

The element of space is consequently increased from $f_1 \cdot dx$ to $\left(f_1 \ (dx + \frac{v_2 - v_1}{dx} \cdot dx\, dt\right)$.

The relative increase in volume is consequently

$$\frac{dV}{V} = \frac{v_2 - v_1}{dx} \cdot dt, \tag{3}$$

where V is the volume of the space-element.

If the electric density at the point in question is q, then we have

$$\frac{v_2 - v_1}{dx} = k \frac{E_2 - E_1}{dx} = 4 \pi q k\,. \tag{4}$$

We consequently get

$$\frac{1}{V} \cdot \frac{dV}{dt} = 4\pi q k, \tag{I}$$

or

$$\frac{dV}{dt} = 4\pi k Q, \tag{II}$$

where $Q = q \cdot V$ is the total charge of the space-element V.

Since $Vdq + qdV = 0$ then (I) may be written

$$-\frac{dq}{dt} = 4\pi k q^2. \tag{III}$$

From this equation we see that if the density was originally constant, $q = q_0$, then the density will always have the same value all over but this value itself will decrease at the rate

$$\frac{1}{q} - \frac{1}{q_0} = 4\pi k t. \tag{IV}$$

For an infinitely long cylinder, no movement occurs in the direction of the axis of the cylinder and consequently in (II) we can replace V by the cross-sectional area A of the cylinder, if at the same time instead of Q we insert the quantity of charge per cm length of the cylinder, Q_1.

$$\frac{dA}{dt} = 4\pi k Q_1. \tag{II'}$$

If we put $A = \pi R^2$ then equation (II') will be

$$R\frac{dR}{dt} = 2kQ_1 \tag{II''}$$

or $R^2 - R_0^2 = 4kQ_1 t$, where R_0 is the radius for $t = 0$.

If mn (fig. 3) is a plane surface on which the charge q is distributed, and if the ordinates to the curve $abcd$ are equal to the corresponding values of q, then the electric strength acting out along the surface will be smaller at the point d than at the point c' at which latter point the charge-curve is assumed to have a point of inflection. If we draw a curve $d'c$ symmetrical to cd with regard to $c'c$ then the outwardly directed field in c' will originate

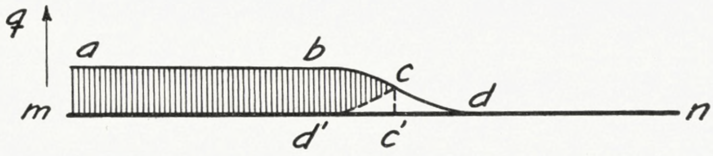


Fig. 3. The spreading out of a charge q along a plane surface mn.

from the shaded part of the front of the charge. The electric strength will therefore have very nearly its highest value at the point of inflection c'. The charge at c' will consequently move outward with greater velocity than will the charge outside this point. The steepness of the outermost front will consequently gradually increase as the charge is spreading out.

Near the centre of a large, plane charge — see fig. 4 — we shall have

$$D - D_0 = 4\pi k q_1 t. \qquad \text{(V)}$$

At the edge, the charge will disperse with a somewhat smaller velocity, so that after a while its outer boundary will have the form shown by the dotted line, and then it will gradually approach a spherical form.

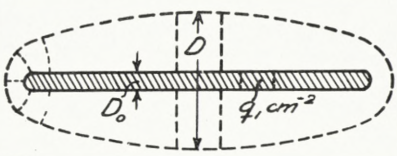


Fig 4. The spreading out of a plane electric charge.

In order to form an idea of the tendency of an originally flat front of an electric charge to assume a steeper form we will treat a couple of geometrically simple cases where the calculations offer no difficulty. The first one is a spherically distributed charge in which the density q

depends only on the time and on the radius r from a fixed centre. In the second case — a charge having cylindrical form — q depends only on the distance r from the straight line representing the axis of the cylinder.

According to (III) the charge-density q at a distance r decreases during the time dt from q to $q - 4\pi k q^2 dt$ while the charge-density $\left(q + \frac{dq}{dr} \cdot dr\right)$ at a distance $r + dr$ during the same time interval decreases to $(q + dq) - 4\pi k (q + dq)^2 \cdot dt$.

The difference $\varDelta q$ between the charge-densities, (which originally was $\varDelta q_{(t=0)} = dq$), is after the time dt.

$$\varDelta q_{(t=dt)} = dq\,(1 - 8\pi k q\, dt). \tag{5}$$

In case of the spherical charge we have according to (1)

$$v = k \cdot \frac{Q_r}{r^2}, \tag{6}$$

where Q_r is the total charge within the sphere of radius r.

During the time dt the distance $\varDelta r_{(t=0)} = dr$ between the two charge-particles will change to

$$\varDelta r_{(t=dt)} = dr + \frac{dv}{dr} \cdot dt = dr\left(1 + 2k\frac{2\pi q r^3 - Q_r}{r^3} \cdot dt\right). \tag{7}$$

From (5) and (7) we get

$$\left(\frac{\varDelta q}{\varDelta r}\right)_{(t=dt)} = \frac{dq}{dr} \cdot \left[1 - 2k\left(6\pi q - \frac{Q_r}{r^3}\right) dt\right]. \tag{8}$$

In the preceding the products $(k \cdot Q_r)$ and $(k \cdot q)$ are always positive (compare equation (1)), and from equation (8) it therefore appears that the steepness of the front will remain unaltered for

$$6\pi q = \frac{Q_r}{r^3} = \frac{4}{3}\pi \cdot q_0 \quad \text{or for} \quad q = \frac{2}{9} \cdot q_0, \tag{9}$$

where q_0 is the mean density of the charge within the sphere of radius r.

From (8) and (9) appears that the steepness of the charge-curve increases for $q < \frac{2}{9} \cdot q_0$ and decreases for $q > \frac{2}{9} \cdot q_0$.

For the cylindrical charge distribution we get, in a corresponding manner,

$$\left(\frac{\Delta q}{\Delta r}\right)_{(t = dt)} = \frac{dq}{dr} \cdot \left[1 - 2k\left(6\pi q - \frac{Q_r}{r^3}\right) \cdot dt\right], \qquad (10)$$

so that the steepness of the charge-curve increases when

$$6\pi q < \frac{Q_r}{r^2} = \pi q_0 \quad \text{or} \quad q < \frac{1}{6} q_0, \qquad (11)$$

while in the opposite case it decreases.

These calculations confirm the fact that the moving front of an electric charge will have a tendency to increase in steepness.

We shall further give some examples where the results from the foregoing may be applied.

Examples:

(1) A homogeneous sphere has a radius R and the total charge Q. We have

$$\frac{dR}{dt} = k \cdot E = k \frac{Q}{R^2} \qquad (12)$$

and

$$\frac{dV}{dt} = \frac{d\left(\frac{4\pi}{3} R^3\right)}{dt} = 4\pi k Q,$$

which agrees with equation (II).

(2) For a homogeneous circular infinitely long cylinder having the radius R and the charge Q_1 per cm of length we have

$$\frac{dR}{dt} = k \cdot \frac{2\,Q_1}{R}$$

and

$$\frac{dV}{dt} = \frac{d(\pi R^2)}{dt} = 4\,\pi\,k\,Q_1. \tag{13}$$

2. Dispersion of Positive and Negative Charges.

If there are ions of both signs present, and if the charge-densities are respectively q_+ and q_- then the resultant density will be $q = q_+ - q_-$.

In the preceding, we have assumed that only ions of the one sign were present, but the results found are of course also valid with approximation in the case when ions of both signs are present, provided that those of the one sign are extremely few compared to those of the op posite sign.

Finally, if q_+ and q_- have such great values that the ionized area may be considered to have infinitely good conductivity, then the surplus of charge Q will collect on the surface of the area, and a further extension of the area can then be determined by means of simple electrostatic considerations.

A sphere or a cylinder of such a conductivity will thus expand in accordance with the formulae (5) and (6) above.

Also in this case, the outer boundary of the charge will have a tendency to become sharp, since stray ions found outside the charge, but being of the same sign as this, will move more slowly than the surface of the charge.

APPENDIX III

1. The Shape of the Positive and Negative Spreaders at the Meeting Point.

There is still a single phenomenon — appearing quite peculiar at first sight — which we will just touch upon, although it does not directly belong to the regular Lichtenberg figures the theory of which has been our sole object in the preceding. Our purpose is to prevent that some readers should find the phenomenon in question to be contradictory to the above evolved theory of the formation of the positive Lichtenberg figures.

If, in the arrangement outlined in fig. 23, the conditions are so selected that the positive and the negative spreaders only just reach each other, very peculiar features are often observed at the place where the two figures meet. An example is shown in L. F. I, fig. 63 and enlarged reproductions of others are shown on plate 24. Of these latter, we will first discuss the upper meeting point in part II.

From the end of the positive spreader, a strongly luminous thread projects over to the edge of the negative figure, where it ends in a strongly luminous spot at the outer end of a negative spreader. This luminous thread has the same characteristics as the subsequent negative discharges in — or along — positive spreaders. (See for example plates 12 VIII, 18 I and 20 I). A closer inspection of the examples reproduced in plate 24 shows that those negative spreaders which have come into contact with the positive ones are more luminous than those that have no connection. There can be no doubt that in the cases considered negative electricity flows over to the end of the positive spreaders. So far the phenomenon is clear enough.

A number of peculiar features may, however, be observed on this link between the positive and the negative

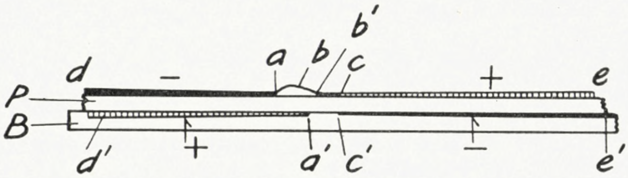


Fig. 1.

spreaders. All of the four junctions on plate 24 show such a strongly luminous spot at the end of the respective negative spreaders, but in all of the four cases the luminosity is comparatively faint in the immediate vicinity of the spot and this is so on both the positive and the negative side. Something similar is apparent on the other figures shown in plate 24.

In order to explain the faintness in luminosity on the positive side of the spot we may, by referring to fig. 1, propose the following. The regular negative figure *da* has just reached the point *a* at the same moment when the regular positive figure has reached the point *c*, and would normally have stopped here but for the negative figure which exerts an attraction on it. Under this influence, the positive spreader proceeds to *b'*. The electric field extending from the negative charge is here still stronger than at *c* and the positive spreader will proceed further, but the field has here a slightly different direction because the induced positive charge $d'a'$ on the metal plate B at a' projects a little beyond the edge a' of the negative discharge. This positive charge at a' changes the direction of the electric field at the tip of the positive spreader from the usual downward to a somewhat upward direction at the point b. A little further forward near the point a the electric field from the negative spreaders will predominate and the positive spreader will "strike down" in the end-point a of the negative spreader.

In the case where the positive and negative figures are placed nearer together, so that near the meeting line both of the figures are still moving on at a considerable velocity, the positive spreaders will be held down against the photographic plate all the way until where they join with the negative ones. Examples of this are shown plate 24 I. The positive spreaders are in this case of no greater luminosity than usual although the corresponding negative spreaders show that a flow of negative electricity has taken place in the direction of the positive electrode, but this circumstance is quite analogous to what is mentioned in chap. III 4 (a) (see also fig. 20).

This terminates our discussion of the relations shown by positive figures where they join with negative ones.

2. Distribution of the Photographic Intensity at the Meeting Point.

In this connection also we observe certain peculiarities, as for instance the formation of luminous wings, stretching out from the point of junction. A satisfactory and thorough explanation of this light distribution can hardly be given at present, although the main features may be explained without difficulty.

The luminous effect is in the main due to three different causes. It is partly due to the comparatively strong light emitted from the luminous thread $abb'c$ (fig. 1) and especially from the part abb' which is raised a little above the photographic film; and partly to the light emitted by the recombination of the positive and negative ions which are present immediately at the surface of the film. The most active factor here is presumably the recombination of electrons with positive ions, while the recombination between positive and negative atomic or molecular ions seems to occur without the emission of any great amount of actinic light.

We shall next discuss some particulars concerning the distribution of the light. In those cases where the positive and negative spreaders just get into contact with each other, a strongly luminous junction-point is visible in the outer faint part of the negative spreader; see for example the upper point of junction in part II and the two middle ones in part III plate 24. This feature is less pronounced in the cases where the positive and negative eleetrodes are placed nearer to each other; see for example the lower points of junction in II and III and all of the points of junction in IV plate 24.

In the first mentioned cases the formation of the negative figure is completed before the positive figure reaches over to it. The junction will then occur at the edge of that area outside the strongly luminous negative figure to which the charge has further moved after the photographic figure was in reality completed, and at a moment when the electrons have already combined with oxygen-, water- or other molecules to form molecular ions. In this case the point of junction will be located outside the strongly luminous part of the negative spreaders.

If, on the contrary, the electrodes are placed so near each other that the positive spreaders reach the negative ones before the formation of the latter is completed, then the point of junction will be located in that part of the negative figure where ionization by collision occurs and where light is consequently produced.

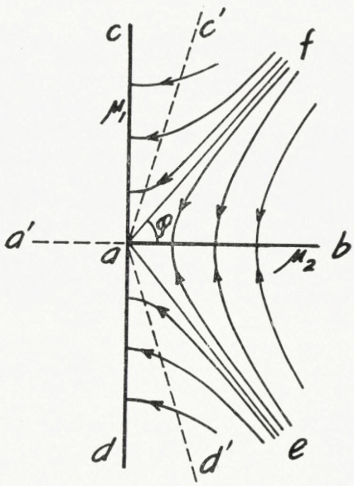


Fig. 2. Charge- and field-distribution after the joining of a positive spreader *ba* with a negative spreader *a'a*.

At the moment the positive spreader has joined with the negative one at *a*, electrons will start out along the positively charged spreader where they will cause ionization by collision, which is followed by recombinations which will produce the strongly luminous tracks visible in the figures, while the charge of these tracks will simultaneously change from positive to negative.

The charge- and the field-distribution thus obtained are outlined in fig. 2. The line *cad* marks the boundary of the negative charge at the moment when the positive spreader reaches the point *a*. The track *ab* will then, as mentioned, rapidly become negatively charged, whereupon we shall have the field-distribution shown.

We may with approximation assume the field from the negative figure to be produced by a charged line *cad* having the charge density $-\mu_1$, since the influence of that part of the charge which is located to the left of *cad* is mainly compensated by the influence of the induced positive charge on the plate B below (see fig. 1). We will further assume the charge density along *ab* to be equal to μ_2. For the sake of simplicity we will further assume that the charge along *ba* projects to the left from *a* as indicated by the dotted line *aa'*.

Under these assumptions the lines of force will take the form of hyperbolae with the axis *ab*, *ac* and *ad* while the asymptotes *af* and *ae* with *ab* form the angle φ determined by

$$tg\ \varphi = \sqrt{\frac{\mu_2}{\mu_1}}.$$

The negative charge at great distances from *a* will move on with practically unaltered velocity, while its spreading out will be completely stopped at *a*. The boundary line of the negative charge will consequently acquire a bend at *a* and after a while it will follow the dotted line *c'ad'*. If this line is reached by the negative charge at the moment when the light from the luminous track *abb'* (fig. 1) is releasing a considerable number of electrons by photoelectric effect, then a strong ionization by collision will occur along the line *c'ad'* since here, according to appendix 2, the charge-density changes quite suddenly and therefore a strong field will exist. The subsequent recombination will then produce the luminous "wings".

CONTENTS

CHAPTER I

CHAPTER II

CHAPTER III
The Properties of the Positive Figures.

CHAPTER IV
The Formation of the Positive Figures.

APPENDIX I
On Positive Particles.

APPENDIX II

APPENDIX III

Forelagt paa Mødet den 20. Januar 1928.
Færdig fra Trykkeriet den 28. Maj 1929.

MATHEMATISK-FYSISKE MEDDELELSER

UDGIVNE AF

DET KGL. DANSKE VIDENSKABERNES SELSKAB

7. BIND (Kr. 17,70):

1. Bohr, Harald: Unendlich viele lineare Kongruenzen mit unendlich vielen Unbekannten. 1925 .... 1.40
2. Hartmann, Jul., and Trolle, Birgit: On Beat-phenomena in Cylindrical Tubes exposed to Sound-waves. With three plates. 1925 .... 2.85
3. Pauli, W. jr.: Ueber die Intensitäten der im elektrischen Feld erscheinenden Kombinationslinien. 1925 .... 0.65
4. Hardy, G. H. and Littlewood, J. E.: A theorem concerning series of positive terms, with applications to the theory of functions. 1925 .... 0.90
5. Steffensen, J. F.: On a Generalization of Nörlund's Polynomials. 1926 .... 1.00
6. Hartmann, Jul., and Trolle, Birgit: New investigation on the air jet generator for acoustic wawes. 1926 .... 2.40
7. Mollerup, Johs.: Sur l'approximation d'un nombre irrationnel par des carrés rationnels. 1926 .... 0.80
8. Nielsen, Niels: Sur certains développements d'une fonction holomorphe. 1926 .... 0.75
9. Bjerrum, Niels: Untersuchungen über Ionenassoziation. I. Der Einfluss der Ionenassoziation auf die Aktivität der Ionen bei mittleren Assoziationsgraden. 1926 .... 2.00
10. Nielsen, Niels: Recherches sur les fonctions cylindriques et sur certaines fonctions analogues. 1926 .... 1.35
11. Hevesy, G.: On the Missing Element 87. 1926 .... 0.60
12. Kuhn, W.: Die Stärke der anomalen Dispersion in nicht leuchtendem Dampfe von Thallium und Cadmium. 1926 .... 3.75
13. Sugiura, Y. and Urey, H. C.: On the Quantum Theory explanation of the Anomalies in the 6th and 7th Periods of the Periodic Table. 1926 .... 1.15
14. Burrau, Øyvind: Berechnung des Energiewertes des Wasserstoffmolekel-Ions (H_2^+) im Normalzustand. 1927 .... 0.90
15. Knudsen, Martin: The Hot-wire Manometer. 1927 .... 0.80
16. Juel, C.: Om v. Staudt's Definitioner. 1927 .... 0.85
17. Pólya, G.: Über die algebraisch-funktionentheoretischen Untersuchungen von J. L. W. V. Jensen. 1927 .... 1.40

8. BIND (Kr. 17,20):

1. Hille, Einar: On the Logarithmic Derivatives of the Gamma Function. 1927 .... 2.40

Kr. Ø.

2. RASCH, G.: Beitrag zur Theorie der unvollständigen Gammafunktionen. Nach hinterlassenen Papieren von J. L. W. V. Jensen. 1927 .... 0.85
3. KNUDSEN, MARTIN: Thermal Molecular Pressure in Tubes. 1927 1.50
4. PEDERSEN, P. O.: Den højere Atmosfæres Sammensætning, Tryk, Temperatur og elektriske Ledningsevne i Belysning af Radiobølgernes Udbredelsesforhold. 1927 .... 0.60
5. BESICOVITCH, A. and BOHR, H.: Some remarks on Generalisations of almost Periodic Functions. 1927 .... 1.50
6. PEDERSEN, ELLEN: Über einige besondere Klassen von fastperiodischen Funktionen. 1928 .... 1.00
7. STEFFENSEN, J. F.: A general Summation Formula. 1928 .... 1.25
8. NIELSEN, NIELS: Observations sur des recherches algébriques plus anciennes que le théorème d'Abel. 1928 .... 1.40
9. BUCH ANDERSEN, E. og ASMUSSEN, R. W.: Undersøgelser over Faradayeffekten hos vandige Opløsninger af nogle uni-univalente Elektrolyter. 1928 .... 0.80
10. PEDERSEN, P. O.: On the Lichtenberg Figures. Part III. The positive Figures. With 28 Plates. 1929 .... 10.00
11. HJELMSLEV, JOHANNES: Einleitung in die allgemeine Kongruenzlehre. Erste Mitteilung. 1929 .... 1.60

9. BIND (under Pressen):

1. BJERRUM, NIELS, und UNMACK, AUGUSTA: Elektrometrische Messungen von Wasserstoffelektroden in Mischungen von Säuren und Basen mit Salzen. (Under Pressen).
2. HARTMANN, JUL.: The Jet-Wave. Theory of the Periodic Jet-Wave. (Under Pressen).
3. STRÖMGREN, ELIS: Asymptotische Lösungen im restringierten Dreikörperproblem (Problème restreint). Mit 1 Tafel. 1929 .... 2.00
4. HARTMANN, JUL.: The Jet-Chain- and the Jet-Wave-Vibrator. (Under Pressen).
5. PEDERSEN, P. O.: Wireles Echoes of Long Delay. 1929 .... 2.40
6. STRÖMGREN, ELIS: Fortsetzung und Abschluss der Librationen um L_2 und L_3 im restringierten Dreikörperproblem (Problème restreint). Mit 1 Tafel. 1929 .... 2.00

Det Kgl. Danske Videnskabernes Selskab.
Mathematisk-fysiske Meddelelser. **VIII,** 11.

EINLEITUNG IN DIE ALLGEMEINE KONGRUENZLEHRE

VON

JOHANNES HJELMSLEV

ERSTE MITTEILUNG

KØBENHAVN
HOVEDKOMMISSIONÆR: ANDR. FRED. HØST & SØN, KGL. HOF-BOGHANDEL
BIANCO LUNOS BOGTRYKKERI
1929

Pris: Kr. 1,60.

Det Kgl. Danske Videnskabernes Selskabs videnskabelige Meddelelser udkommer fra 1917 indtil videre i følgende Rækker:

Historisk-filologiske Meddelelser,
Filosofiske Meddelelser,
Mathematisk-fysiske Meddelelser,
Biologiske Meddelelser.

Hele Bind af disse Rækker sælges 25 pCt. billigere end Summen af Bogladepriserne for de enkelte Hefter.

Selskabets Hovedkommissionær er *Andr. Fred. Høst & Søn*, Kgl. Hof-Boghandel, København.

Det Kgl. Danske Videnskabernes Selskab.
Mathematisk-fysiske Meddelelser. **VIII,** 11.

EINLEITUNG IN DIE ALLGEMEINE KONGRUENZLEHRE

VON

JOHANNES HJELMSLEV

ERSTE MITTEILUNG

KØBENHAVN
HOVEDKOMMISSIONÆR: ANDR. FRED. HØST & SØN, KGL. HOF-BOGHANDEL
BIANCO LUNOS BOGTRYKKERI
1929

Einleitung.

In meiner Arbeit »Neue Begründung der ebenen Geometrie« (Math. Ann. 64. Bd. 1907) wurde gezeigt, dass die ebene Geometrie unter ausschliesslicher Benutzung ebener Axiome ohne Stetigkeitsbetrachtungen, ganz unabhängig von der Parallelenfrage aufgebaut werden kann. Und tatsächlich bestehen meines Wissens bis jetzt keine andere Grundlagen, um die Sätze der ebenen Geometrie (z. B. den Höhenschnittpunktsatz, den Medianenschnittpunktsatz, u. dgl.), ohne räumliche Betrachtungen und ohne Stetigkeitsaxiome, unabhängig von der Parallelenfrage zu beweisen.

Schon in dieser Arbeit war sehr auffallend, wie geringe Rolle die Beziehungen der Anordnung spielten, und es hat sich schliesslich gezeigt, dass hierher gehörige Axiome ganz ausser Spiel gesetzt werden können[1]. In einer Reihe von Arbeiten soll nun gezeigt werden, wie das allgemeine Kongruenzproblem, das darin besteht, alle Geometrien zu erforschen, die eine Kongruenzlehre gestatten, wo aber sowohl die Anordnungsaxiome als auch das Axiom betreffs der eindeutigen Bestimmung der geraden Linie durch zwei Punkte (das Eindeutigkeitsaxiom) fortgelassen werden, gelöst werden kann.

Der Weg wurde angebahnt durch die Studien der

[1] Beretning om d. 2. skand. Matematikerkongres i Kjøbenhavn 1911 (cf. Jahrb. ü. d. Fortschr. d. Math. 43, S. 560).

1*

Geometrie der Wirklichkeit, wo man das nächstliegende Beispiel einer hierher gehörigen Geometrie vor Augen hat, indem eben hier die Eindeutigkeit der Bestimmung der geraden Linie durch zwei Punkte keine unbedingte Gültigkeit hat. Diejenige Form der Kongruenzaxiome, welche sich als die fruchtbarste für die Behandlung des allgemeinen Kongruenzproblems herausgestellt hat, ist eben aus den einfachsten Tatsachen der Wirklichkeit entsprungen. Umgekehrt werden die hier dargestellten allgemeinen Untersuchungen geeignet sein, über die Probleme der Geometrie der Wirklichkeit neues Licht zu verbreiten.

Die in Rede stehenden sehr allgemeinen Geometrien sind als Nicht-Eudoxische Geometrien (sogenannte Nicht-Archimedische Geometrien) zu bezeichnen. Sie sind allerdings viel allgemeiner als die Nicht-Eudoxischen Geometrien im gewöhnlichen Sinne, insofern die Axiome der Anordnung ausser Betracht gestellt worden sind. Lässt man die Axiome der Anordnung zu, und nimmt man überdies auch das Eudoxische Axiom (das sogenannte Archimedische Axiom) an, so wird man im Stande, das Eindeutigkeitsaxiom zu beweisen.

Sehr überraschend wirkt in unserer allgemeinen Kongruenzlehre das Resultat betreffs der Existenz des Rechtecks: Auf Grund der Existenz zweier Geraden mit mehr als einem gemeinsamen Punkt lässt sich beweisen, dass Rechtecke existieren. Derjenige Beweis, der mit dem Eindeutigkeitsaxiom an der Spitze der Geometrie, nicht gelingen wollte, und nicht gelingen konnte, gelingt so in einfachster Weise, wenn man die gerade Linie ihre natürlichen Eigenschaften behalten lässt!

Es wird sozusagen hierdurch die Gültigkeit der Euklidischen Geometrie »im Kleinen« festgestellt.

1. Das Axiomsystem.

1. Die allgemeine Kongruenzlehre wird auf folgendes Axiomsystem gegründet.

I. Es gibt *Punkte.* Es gibt Punktmengen, die *gerade Linien* (*Geraden*) heissen. Es gibt Transformationen welche *Bewegungen* heissen. Jede Bewegung ist eine Zuordnung, durch welche jeder Geraden und jedem in ihr gelegenen Punkte eine Gerade und ein in ihr gelegener Punkt umkehrbar eindeutig entspricht. Die Umkehrung einer Bewegung ist auch eine Bewegung. Die Bewegungen bilden eine Gruppe. Zwei Figuren, die durch eine Bewegung auseinander abgeleitet werden können, sollen *kongruent* heissen.

II. Ausser der Identität gibt es eine und nur eine Bewegung, welche alle Punkte einer geraden Linie stehen lässt. Diese Bewegung heisst eine *Spiegelung* an der geraden Linie. Die Linie wird als Achse der Spiegelung bezeichnet. Jede Gerade ist die Achse einer Spiegelung. Jeder Punkt ausserhalb der Achse geht bei der Spiegelung in einen von ihm verschiedenen Punkt über.

III. Eine Gerade b heisst *senkrecht* zu einer von ihr verschiedenen Geraden a (in Zeichen: $b \perp a$), wenn b bei der Spiegelung an a in sich selbst übergeht. Durch jeden Punkt geht eine und nur eine Gerade b senkrecht zu einer gegebenen Geraden a; die beiden Geraden haben stets einen und nur einen Punkt gemein.

IV. Wenn zwei Punkte A und B einer geraden Linie l angehören, haben sie immer eine *Spiegelungsachse* m, derart dass A und B bei der Spiegelung an m ineinander übergehen, während l in sich selbst übergeführt wird ($l \perp m$). Diese Spiegelungsachse schneidet l in einem Punkt M, welcher als *Mittelpunkt* von A und B (oder von AB) auf l bezeichnet wird.

V. *Zwei kongruente Punktreihen* $ABC\ldots$ und $AB'C'\ldots$ auf einer oder auf zwei Geraden, mit dem gemeinsamen Punkt A, können immer durch eine Spiegelung ineinander übergeführt werden.

Es werden keine Voraussetzungen über eindeutige Bestimmung einer Geraden durch zwei Punkte aufgestellt. Es wird nicht einmal die Existenz einer Geraden durch zwei beliebig gewählte Punkte gefordert. Die Axiome der Anordnung sind ganz ausgeschaltet worden.

2. Beispiele. 1°. In der Cartesischen Ebene wählen wir die rationalen Punkte und die rationalen Geraden aus, und die Bewegungen lassen wir durch rationale orthogonale Substitutionen definiert sein. Alle unsere Axiome sind dann befriedigt. Es sei aber hervorgehoben, dass nicht jede Gerade in jede andere Gerade durch Bewegung übergehen kann. Man betrachte z. B. die beiden Geraden $y = 0$, $y = x$.

2°. In der Cartesischen Ebene nehme man die Punkte, deren Koordinaten geschlossene Dualbrüche sind, und die Geraden, deren Gleichungen auf eine der folgenden vier Formen geschrieben werden können: $x = a$, $y = b$, $y = \pm x + c$, wobei a, b, c geschlossene Dualbrüche bezeichnen. Die Bewegungen seien aus Spiegelungen an diesen Geraden zusammengesetzt. Für die so definierte Geometrie wird auch unser ganzes Axiomsystem erfüllt sein.

3°. In der gewöhnlichen komplexen Ebene nehmen wir alle Punkte, und alle Geraden mit Ausnahme der isotropen Linien $y = \pm ix + q$. Die Bewegungen seien wie gewöhnlich durch orthogonale Substitutionen definiert. Es gilt dann unser ganzes Axiomsystem. Es gibt aber Punkte ohne Verbindungsgerade. Die Axiome der Anordnung sind ungültig.

4°. In einer Koordinatengeometrie, wo die Koordinaten (x, y) Zahlen der Form $a + \varepsilon b$ darstellen, mit a reell, b reell, $\varepsilon^2 = 0$, wird auch unser Axiomsystem erfüllt sein, wenn die Geraden und die Bewegungen wie üblich definiert werden. Es gibt hier Punktepaare mit unendlich vielen Verbindungsgeraden (Beispiel: $(0, 0)$, $(\varepsilon, 0)$).

Erweitert man das System dahin, dass auch komplexe Werte für a und b in Betracht gezogen werden, können auch Punktepaare ohne Verbindungsgerade vorkommen.

5°. Dem vorhergehenden Beispiel kann man eine kinematische Deutung geben. In der Euklidischen Ebene denkt man sich »Punkt mit Geschwindigkeit« als Punkt einer Geometrie. Der Abstand zweier solchen Punkte wird dann »Strecke mit Streckengeschwindigkeit«. Mit anderen Worten ausgedrückt: Jeder Punkt der Grundebene wird mit einem Geschwindigkeitsvektor versehen, jede Grösse (Abstand, Winkel, ...) der Grundebene wird so mit einem Differential versehen. Die hierdurch hergestellte Geometrie, d. h. die ebene Kinematik, wird so in der allgemeinen Kongruenzlehre einbegriffen. Anwendungen auf höhere Differentiale erfolgen in ähnlicher Weise[1].

3. Schliesslich sei nur noch darauf hingewiesen, dass die Tragweite unseres Systems noch erheblich vergrössert werden kann, wenn die Forderung betreffs der Eindeutigkeit des Senkrechtfällens (auf eine Gerade von einem beliebigen Punkte aus) fortgelassen wird. Es werden dann auch Anwendungen auf die Liniengeometrie in derjenigen Form, wie sie von E. Study in seiner Geometrie der Dynamen behandelt worden ist, und auch Anwendungen auf ähnliche höhere Zweige der Geometrie in Betracht kommen.

[1] Vgl. hierzu meine Arbeit: Die Nicht-Eudoxische Mathematik (Den sjette Skandinaviske Matematikerkongres i København 1925).

Wir müssen aber zunächst das einfachere Problem, wie es hier aufgestellt wurde, eingehend behandeln.

4. Es sei endlich noch hervorgehoben, dass die besonderen Geometrieformen, wo die Axiome der Anordnung oder das Eindeutigkeitsaxiom oder andere spezielle Axiome, welche in unser System nicht aufgenommen sind, Gültigkeit haben, auch mit Vorteil mit den hier im folgenden dargestellten allgemeinen Hilfsmitteln behandelt werden können.

2. Zeichensprache.

5. Bei Rechnungen mit Transformationen wollen wir die Aufeinanderfolge der auszuführenden Transformationen von links nach rechts schreiben. Statt $U^{-1}\,T\,U$ schreiben wir oft T^U (die Transformierte von T durch U). Es wird dann

$$(T^U)^V = T^{UV}, \quad T^U \cdot V^U = (T\,V)^U.$$

Die in unserer Arbeit vorkommenden Transformationen werden gewöhnlich aus Spiegelungen zusammengesetzt. Gerade Linien bezeichnen wir mit den Buchstaben $a, b, c, d, \ldots$ Die zugehörenden [Spiegelungen bezeichnen wir mit denselben Buchstaben. Z. B. soll abc eine Bewegung bezeichnen, welche dadurch entsteht, dass die Spiegelungen a, b, c nacheinander in der angegebenen Reihenfolge ausgeführt werden.

Jede Spiegelung ist involutorisch. Es wird somit

$$a^{-1} = a, \quad a^2 = 1, \quad b^a = aba,$$

b^a bezeichnet eine Spiegelung, deren Achse bei der Spiege lung an a der Geraden b entspricht.

6. Ist $b \perp a$, wird

$$b^a = b, \quad ab = ba,$$

und umgekehrt:

ist $$ab = ba, \quad \text{also} \quad b^a = b,$$

und b von a verschieden, so folgt $b \perp a$. Ist $b \perp a$, ist somit auch $a \perp b$.

Ist $ab = ba$, hat man entweder $a \perp b$ (und $b \perp a$), oder die beiden Geraden sind identisch.

7. Es ist
$$(abc)^a = a(abc)a = bca,$$
$$(abcd)^a = bcda, \text{ u. s. w.}$$

8. Die Aufeinanderfolge von zwei Spiegelungen kann nicht durch eine einzige Spiegelung ersetzt werden.

Aus $ab = c$, würde folgen $ab = ba$, $bc = cb$, $ca = ac$, d. h. (indem die drei Möglichkeiten $a = b$, $b = c$, $c = a$, ausgeschlossen sind)

$$a \perp b, \quad b \perp c, \quad c \perp a,$$

was unmöglich ist, weil zwei zu einander senkrechte Geraden immer einen Schnittpunkt aufweisen, und von einem Punkt nur eine Senkrechte auf eine Gerade gefällt werden kann.

Es folgt sofort, dass 3 Spiegelungen einander nicht aufheben können.

9. Zwei zueinander senkrechte Geraden a, b werden durch eine Bewegung U in zwei zueinander senkrechte Geraden transformiert.

Es ist nämlich

$$ab = ba,$$
$$(ab)^U = (ba)^U,$$
$$a^U b^U = b^U a^U.$$

3. Bewegungen mit einem festen Punkt.

10. Wenn bei einer Bewegung der Punkt A fest ist, so wird eine gerade Punktreihe $ABC\ldots$ auf einer Geraden g durch A in eine gerade Punktreihe $AB'C'\ldots$ auf einer Geraden g' durch A hinübergeführt. Die beiden Reihen haben eine Spiegelungsachse a. Die vorgelegte Bewegung muss dann entweder mit der Spiegelung a oder mit der Bewegung ga gleichwertig sein.

Also:

Wenn eine Bewegung einen Punkt fest lässt, muss sie entweder einer einzigen Spiegelung a durch diesen Punkt oder der Aufeinanderfolge zweier Spiegelungen g, a durch den Punkt gleichwertig sein. Im letzteren Falle kann die erste Gerade g ganz beliebig durch den festen Punkt gewählt werden.

Nach 8 schliessen die beiden Fälle einander aus.

11. Es folgt nun unmittelbar: Wenn 3 Geraden a, b, c durch denselben Punkt O gehen, lässt sich immer eine vierte Gerade x durch O finden, derart dass $ab = cx$, oder $x = cab$. Es gilt also auch der Satz:

Die Aufeinanderfolge von drei Spiegelungen a, b, c, deren Achsen durch denselben Punkt O hindurch gehen, kann durch eine einzige Spiegelung ersetzt werden. Die Achse der Spiegelung geht durch O.

Es folgt nun auch, dass die Gleichung $ab = xc$ lösbar ist; es wird nämlich $x = abc$.

12. Ist $a \perp b$, so wird auch $x \perp c$; weil $ab = ba$, ist nämlich auch $xc = cx$.

Hieraus schliesst man:

Zwei beliebige zueinander senkrechte Geraden

a, b durch den Punkt O, bestimmen eine Bewegung ab, welche jede Gerade c durch O stehen lässt. Die Bewegung ist involutorisch, weil $ab = ba$.

Diese Bewegung soll als Umwendung um den Punkt O bezeichnet werden. Die Umwendung selbst soll mit O bezeichnet werden.

13. Die Umwendung O lässt keinen von O verschiedenen Punkt stehen. Wäre P ein fester Punkt, könnte die Bewegung durch die Aufeinanderfolge von zwei Spiegelungen r und s durch P ersetzt werden, und die Gerade l von O senkrecht auf r müsste bei der Bewegung rs fest bleiben, d. h. sie müsste auch senkrecht auf s stehen, was unmöglig ist, weil r und s nicht zusammenfallen. Der Fall $l = s$ ist ausgeschlossen, weil zwei zueinander senkrechte Geraden nur einen Punkt gemein haben.

14. Folgerung: Ist $P^Q = P$, oder $PQ = QP$, hat man notwendig $Q = P$.

15. Bei Rechnungen mit Spiegelungen und Umwendungen ist es nützlich, die folgenden einfachen Sätze zu kennen:

Ist $P^a = P$, muss P auf a liegen.

Der Satz besagt nur, dass die Spiegelung a keine anderen Punkte stehen lässt, als die Punkte von a.

16. Ist $a^P = a$, liegt P auf a.

Es ist nämlich

$$PaP = a, \quad aPa = P, \quad P^a = P,$$

d. h. P liegt auf a (15).

17. Ist Pa involutorisch, muss P auf a liegen.

Es ist nämlich

$$P^{aP} = P^{Pa},$$
$$(P^a)^P = (P^P)^a = P^a$$

also $$P^a = P,$$

d. h. P liegt auf a.

4. Bewegungen, die eine Gerade fest lassen.

18. Wenn die Gerade l fest bleibt, muss eine Punktreihe $ABC\ldots$ auf l in eine entsprechende Punktreihe $A'B'C'\ldots$ auf l übergehen. Die Senkrechte auf l in A und die Mittelsenkrechte der Punkte AA' auf l seien mit a bezw. m bezeichnet. Die Punktreihe $ABC\ldots$ muss dann in $A'B'C'\ldots$ hinübergehen entweder durch die Spiegelung m allein, oder durch die Aufeinanderfolge von a und m. Die vorgelegte Bewegung lässt sich dann sicher auf eine der folgenden Formen darstellen:

$$m, \quad ml, \quad am, \quad aml.$$

Es gibt also nur 4 Möglichkeiten:

Eine Bewegung, die eine Gerade l stehen lässt, ist entweder eine Spiegelung an einer Achse senkrecht zu l, oder eine Umwendung um einen Punkt von l, oder sie kann durch zwei Spiegelungen, deren Achsen senkrecht zu l sind, oder durch zwei solche Spiegelungen und eine nachfolgende Spiegelung an l selbst ersetzt werden.

Fallen a und m zusammen, gehen die letzten zwei Formen in die Identität bezw. die Spiegelung l hinüber.

19. Die Aufeinanderfolge von 3 Spiegelungen a, b, c, deren Achsen senkrecht zu einer Geraden l stehen, ist eine involutorische Bewegung.

Nach dem vorhergehenden Satz bestehen nämlich nur folgende Möglichkeiten:

$$abc = \begin{cases} m \\ ml \end{cases} \qquad abc = \begin{cases} am \\ aml \end{cases} \qquad m \perp l.$$

Die beiden letzten Möglichkeiten können durch die folgenden ersetzt werden

$$bc = \begin{cases} m \\ ml \end{cases}$$

also bc involutorisch, $b \perp c$, was unmöglich ist, oder $b = c$, was kein Interesse hat.

Es bleiben also nur die folgenden Möglichkeiten übrig:

$$abc = \begin{cases} m \\ ml \end{cases}$$

d. h. eine Spiegelung, oder eine Umwendung, womit der Satz bewiesen ist.

20. Nach diesem vorbereitenden Satz können wir nun den folgenden Hauptsatz beweisen:

Die Aufeinanderfolge von drei Spiegelungen a, b, c, deren Achsen senkrecht zu einer Geraden l sind, lässt sich durch eine einzige Spiegelung ersetzen, deren Achse senkrecht auf l steht.

Zunächst bestimmen wir eine Gerade x dergestalt, dass

$$ax = xc;$$

x ist senkrecht zu l und Spiegelungsachse für a und c.

Wir bestimmen ferner eine Gerade $d \perp l$ derart, dass die Gleichung

$$bx = xd$$

erfüllt wird; d wird durch Spiegelung von b an x erzeugt

Dann folgt

$$(ax)(xb) = (xc)(dx),$$

oder

$$ab = (xcd)x;$$

da ferner xcd involutorisch ist (19), also

$$xcd = dcx,$$

so ergibt sich

$$ab = (dcx)\,x = dc\,,$$

d. h.

$$abc = d\,,$$

w. z. b. w.

21. Zwei Spiegelungen a und b, deren Achsen senkrecht auf einer Geraden l stehen, können durch zwei andere Spiegelungen dergestalt ersetzt werden, dass die Achse der ersten oder der zweiten dieser Spiegelungen in irgendeine gegebene Gerade $c \perp l$ fällt.

Die Gleichungen $ab = cx$, und $ab = xc$ sind in der Tat gleichbedeutend mit

$$cab = x, \text{ bzw. } abc = x\,.$$

22. Zwei Umwendungen A und B um zwei Punkte einer Geraden l können durch zwei Spiegelungen, deren Achsen in die Lote von l in A und B fallen, ersetzt werden.

Die beiden Lote a und b erfüllen nämlich die folgenden Gleichungen:

$$al = A\,, \quad lb = B,$$

woraus folgt

$$allb = AB\,,$$

oder

$$ab = AB.$$

23. Die Aufeinanderfolge von drei Umwendungen A, B, C, um Punkte einer Geraden l, kann durch eine einzige Umwendung um einen Punkt derselben Geraden ersetzt werden,

Die Lote von l und A, B, C seien mit a, b, c bezeichnet. Es wird dann

$$ABC = allbcl = (abc)\,l,$$

wo

$$abc = n \perp l,$$

also

$$ABC = nl = D,$$

indem D den Schnittpunkt von n und l bezeichnet. Die Gleichungen $AB = CX$, $AB = DX$, lassen sich hiernach lösen.

24. Wenn A und B auf l liegen, und $r \perp l$, so ist ABr eine Spiegelung, deren Achse senkrecht auf l steht.

Die Lote von l in A und B seien mit a bezw. b bezeichnet. Es ist dann

$$A = al, \quad B = lb, \quad ABr = abr.$$

25. Die Bewegung Op lässt sich in eine ihr gleichwertige Bewegung $O'p'$ (oder $p'O'$) umändern dergestalt, dass p' durch einen beliebig vorgeschriebenen Punkt P geht.

Das Lot Op (d. h. das Lot von O auf p) sei q, und das Lot Pq sei p'. Es ist dann möglich, einen Punkt O' auf q zu finden, dergestalt, dass $Op = O'p'$ oder $p'O'$.

26. Wenn ABr eine involutorische Bewegung darstellt, so muss das Lot von A auf r durch den Punkt B laufen.

Beweis. Statt Br kann man $r'B'$ schreiben, wo r' durch A geht. Es wird dann

$$ABr = Ar'B' = sB' \quad (s \perp r' \text{ in } A).$$

Soll diese Bewegung involutorisch sein, muss B' auf s liegen (17), d. h. das Lot Br (oder $r'B'$) fällt mit s zusammen, und enthält somit den Punkt A.

27. Wenn $a \perp n$ $(an = A)$, $b \perp n$ $(bn = B)$, und abc involutorisch ist, so gibt es eine Gerade durch A und B senkrecht auf c.

Es ist nämlich

$$a = An, \quad b = Bn, \quad abc = AnnBc = ABc,$$

wodurch der Satz auf den vorigen zurückgeführt wird.

5. Der Mittelpunkt.

28. Zwei Punkte A, B einer Geraden l haben nach Voraussetzung immer einen Mittelpunkt auf dieser Geraden. Wenn mehrere Verbindungsgeraden der beiden Punkte A, B vorhanden sind, könnte die Möglichkeit bestehen, dass auch mehrere Mittelpunkte den verschiedenen Verbindungsgeraden entsprechend existieren. Es ist aber dem nicht so. In der Tat, haben die beiden Punkte A, B die Mittelpunkte M, M_1, den beiden Verbindungsgeraden l, l_1 entsprechend, müssen die Punkte A und B bei der Bewegung MM_1 fest bleiben. Die Bewegung MM_1 ist nun der Aufeinanderfolge von zwei Spiegelungen l, r (r durch A) gleichwertig (10, 14). Es folgt dann

$$lr = MM_1, \quad M = ls \ (s \perp l \text{ in } M),$$
$$lr = lsM_1, \ r = sM_1,$$
$$M_1 \text{ auf } s, \quad r \perp s, \quad r = l, \quad M = M_1.$$

29. Haben die beiden Punkte A, B keine Verbindungsgerade, gibt es doch einen Mittelpunkt, d. h. es gibt eine Umwendung, durch welche A und B in einander übergehen. Um das zu beweisen, ziehen wir durch A und B zwei zueinander senkrechte Geraden a, b, die einander in C kreuzen. n ist die Mittelsenkrechte von A, C auf der Geraden a, P der Mittelpunkt von B, C auf der Geraden b. s ist das Lot von P auf n. Die Bewegung nP, welche die Gerade s stehen lässt, führt nun A in B über; sie führt deshalb auch die senkrechte AA_1 auf s in die senkrechte BB_1 auf s hinüber. Die Punkte A_1B_1 haben auf s den Mittelpunkt M. Die Bewegung nP lässt sich nun in eine Bewegung xM umändern, wo x senkrecht auf s steht, und da A_1 bei dieser Bewegung in B_1 übergehen muss, so fällt x mit der senkrechten AA_1 auf s zusammen. A geht nun bei der Bewegung xM in B über; die Spiegelung x lässt aber

A ungeändert, und die Umwendung M allein muss also A nach B führen.

Wir haben also den Satz:

Zwei Punkte haben immer einen eindeutig bestimmten Mittelpunkt, unabhängig davon, ob die Punkte eine, mehrere oder gar keine Verbindungsgerade haben.

30. Haben zwei gerade Linien zwei Punkte A und B gemein, haben sie auch den Punkt A^B (d. h. den Punkt, nach welchem A durch die Umwendung B geführt wird) gemein. Durch fortgesetzte Umwendung wird dann eine ganze Reihe von gemeinsamen Punkten erzeugt. Ebenso müssen die beiden Geraden auch den Mittelpunkt von A, B enthalten, und durch fortgesetzte Mittelpunktkonstruktion entsteht ein ganze Menge von Punkten, welche alle in den beiden Geraden enthalten sind. Wir kommen später auf diese Frage zurück.

31. Wenn man die Axiome der Anordnung in geeigneter Fassung zu unserem Axiomsystem hinzufügt, wird sich zeigen, dass eine Strecke durch ihre Endpunkte eindeutig bestimmt wird. Der Beweis ist sehr einfach: Wenn die Endpunkte A, B einer Strecke gemeinsame Projektion auf eine gerade Linie haben, so müssen alle Punkte der Strecke dieselbe Projektion haben; es würde sonst durch die Projektion die Ordnung der Punkte gestört. Würde man ausserdem das Eudoxische Axiom (sogenannte Archimedische Axiom) heranziehen, würden auch die Verlängerungen der Strecke eindeutig bestimmt. Also:

Bei Annahme der Anordnungsaxiome und des Eudoxischen Axioms ist die Eindeutigkeit der Bestimmung der geraden Linie durch zwei Punkte beweisbar.

6. Das Rechteck.

32. Zwei Geraden l, m haben die beiden Punkte A, B gemein. Auf l wählen wir einen Punkt C, welcher nicht auf m liegt. Wir fällen das Lot p von C auf m; der Schnittpunkt mit m werde mit D bezeichnet.

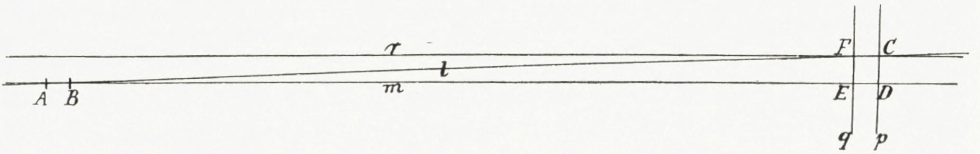


Fig. 1.

Es wird nun

$$ABD = E,$$

wo E ein Punkt von m ist. Hieraus folgt

$$ABC = EDC,$$

und da A, B, C einer Geraden angehören, ist ABC, also EDC eine involutorische Bewegung (23).

Errichtet man nun das Lot q auf m in E, und das Lot r auf p in C, so wird

$$E = qm, \quad D = mp, \quad C = pr,$$

also

$$EDC = qmmppr = qr.$$

qr ist also eine involutorische Bewegung, d. h. $q \perp r$. Setzt man $qr = F$, haben wir ein Rechteck $CDEF$ konstruiert.

Die Existenz des Rechtecks ist hiermit gesichert.

33. Durch Spiegelung des gefundenen Rechtecks an einer der Seiten m, p, q, r wird ein neues Rechteck gebildet, und durch fortgesetzte Spiegelungen dieser Art bildet man zwei Reihen äquidistanter Linien, die einander rechtwinklig durchkreuzen. Das System lässt sich so erweitern, dass

man die Spiegelungsachsen je zweier Geraden jeder Reihe hinzufügt. Es gilt nämlich der folgende Satz:

34. Jedes Rechteck hat zwei Spiegelungsachsen.

Beweis: Es sei $ABCD$ ein Rechteck (AB auf a, BC auf p, CD auf b, DA auf q; $p \perp a$, $p \perp b$, $q \perp a$, $q \perp b$). r und s seien die Mittelsenkrechten von AB (auf a) bezw. CD (auf b). Es soll dann gezeigt werden, dass r und s zusammenfallen müssen. Die Bewegung rs lässt sowohl p wie q ungeändert. Sie ist keine Spiegelung, weil zwei Spiegelungen sich überhaupt nicht durch eine Spiegelung ersetzen lassen. Sie ist auch keine Umwendung, weil sie zwei einander nicht schneidende Geraden stehen lässt. Es bleiben dann nur noch folgende Möglichkeiten übrig (18):

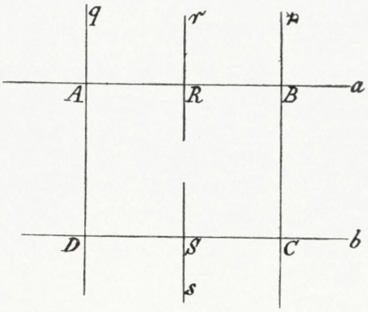


Fig. 2.

$$rs = ak, \quad \text{oder} \quad rs = akq,$$

wo k ein Lot von q bezeichnet. Aus diesen Gleichungen folgt aber:

$$ars = k, \quad \text{bezw.} \quad ars = kq,$$

d. h.

$$Rs = \begin{cases} k, \\ K, \end{cases}$$

indem $ar = R$, $kq = K$ gesetzt wird.

Nur der erste Fall ist möglich, und es wird

$$Rs = k, \quad R = ks,$$

$$R \text{ auf } s, \quad R \text{ auf } k, \quad k = a,$$

$$ars = a, \quad r = s.$$

35. Die beiden Spiegelungsachsen des Rechtecks sind senkrecht zueinander. Ihr Schnittpunkt ist Mittelpunkt

einer Umwendung, welche die Gegenecken in einander überführt.

36. Haben die beiden Ecken A und B nur eine Verbindungsgerade, dann gilt dasselbe von den Ecken C und D. Jedes Lot x von a ist dann auch Lot von b. pqx ist nämlich eine involutorische Bewegung. Es folgt nun der Satz:

Gibt es ein Rechteck $ABCD$, wo die geraden Linien AB und AD eindeutig bestimmt sind, so gibt es ein Rechteck mit der Ecke A und mit zwei anderen Ecken X und Y in beliebig gewählten Punkten von diesen beiden Geraden.

7. Die allgemeine Bewegung.

37. Eine Bewegung, durch welche der Punkt A in den Punkt A_1 übergeht, kann durch eine Umwendung um den Mittelpunkt M von A und A_1 und eine nachfolgende Bewegung, die den Punkt A_1 stehen lässt, ersetzt werden. Die letztere Bewegung ist entweder durch eine oder zwei Spiegelungen darstellbar. Hieraus folgt:

Jede Bewegung lässt sich ersetzen durch eine Umwendung M und eine einzige Spiegelung a oder durch eine Umwendung M und zwei Spiegelungen a, b, deren Achsen einen Punkt gemein haben.

Im ersteren Falle kann man auch die Ordnung der Umwendung und der Spiegelung umkehren, indem

$$Ma = aaMa = a(aMa) = aM',$$

wo $M' = M^a$ (Spiegelbild von M durch die Achse a).

Es folgt hieraus, dass man auch im letzteren Falle die Ordnung der Umwendung und der beiden Spiegelungen ändern kann.

Der Typus Ma soll als Um-Bewegung bezeichnet werden, der Typus Mab als In-Bewegung. Die Spiegelung ist eine Um-Bewegung, die Identität eine In-Bewegung.

38. Um zu zeigen, dass die beiden Typen wirklich verschieden sind, wird es nötig zu beweisen, dass eine In-Bewegung und eine Um-Bewegung einander nicht gleichwertig sein können. Wir müssen also zeigen, dass die Gleichung

$$Mab = On$$

unmöglich ist, indem die Geraden a und b einen gemeinsamen Punkt P haben. Ohne Beschränkung der Allgemeinheit können wir voraussetzen, dass n durch P geht. Wäre dies nämlich nicht der Fall, könnten wir On durch $O'n'$ ersetzen, wo n' das Lot von P auf das Lot On bezeichnet. Hiernach wird die obige Gleichung

$$M(abn) = O,$$

$$Mc = O,$$

was unmöglich ist.

39. Die Aufeinanderfolge von zwei Spiegelungen a, b ist eine In-Bewegung.

Beweis. Vom Punkte B auf b fällen wir die senkrechte c auf a. Es ist dann

$$ab = (ac)cb = Ocb.$$

40. Die Aufeinanderfolge von drei Spiegelungen a, b, c ist eine Um-Bewegung.

Wir schreiben

$$bc = bmC,$$

wo b und m einen gemeinsamen Punkt P haben, $m \perp c$. Statt bm schreiben wir xy, wo x (und y) durch P geht und $x \perp a$. Est ist nun

$$abc = axyC = AyC,$$

wo A der Schnittpunkt (ax) bezeichnet. Ay lässt sich nun in $A'y'$, wo y' durch C geht, umändern, und es wird dann

$$abc = A'y'C = A'z.$$

41. Die Aufeinanderfolge von vier Spiegelungen ist eine In-Bewegung.

Folgt sofort aus dem vorigen Satz, indem $A'zd$ immer so umgeändert werden kann, dass $A'z = A''z'$, wo z' durch einen Punkt von d gelegt wird.

42. Wenn man zu einer Um-Bewegung eine Spiegelung hinzufügt, erhält man eine In-Bewegung. Und wenn man umgekehrt zu einer In-Bewegung eine Spiegelung hinzufügt, kommt eine Um-Bewegung heraus. In der Tat ist

$$(Mab)c = M(abc) = MA'z = Mz'A'',$$

wo z' durch M gelegt ist, und hieraus folgt:

$$Mz' = s,$$
$$(Mab)c = sA''.$$

Ferner:

Die Aufeinanderfolge einer beliebigen Anzahl von Spiegelungen ist eine In-Bewegung oder Um-Bewegung, je nachdem die Anzahl der Spiegelungen gerade oder ungerade ist.

43. Jede Um-Bewegung, welche keine Spiegelung ist, lässt eine und nur eine gerade Linie fest bleiben. Diese Gerade enthält den Mittelpunkt eines jeden Paares bei der Bewegung einander entsprechender Punkte.

Die Bewegung sei Op. Die senkrechte l von O auf p ist eine feste Gerade. Dass keine andere Gerade fest bleiben kann, folgt aus folgender Betrachtung. Soll die Gerade x bei einer Um-Bewegung fest bleiben, und einfache Spiegelungen nicht in Betracht kommen, so muss die Bewegung aus einer Umwendung um ein Punkt O_1 auf l und einer Spiegelung an eine Achse p_1 senkrecht zu l hervorgehen.

Da nun $$Op = O_1p_1, \quad O_1Op = p_1,$$

so folgt, dass die senkrechte von O auf p auch durch O_1 geht und senkrecht auf p_1 steht, d. h. $x = l$. Es gibt also nur eine feste Gerade.

Der zweite Teil des Satzes wird folgendermassen bewiesen. Der Punkt A gehe bei unserer Bewegung in A_1 über. Durch A_1 ziehen wir das Lot n auf l. Es lässt sich dann ein Punkt P auf l finden, derart dass

$$Op = Pn.$$

Da nun A bei diesen Bewegung in A_1 übergeht, muss schon die Umwendung P den Punkt A nach A_1 führen, d. h. P ist der Mittelpunkt von A und A_1. Der Mittelpunkt von A und A_1 liegt also auf l.

44. Die Mittelpunkte entsprechender Punkte in zwei kongruenten geradlinigen Punktreihen liegen immer auf einer Geraden; speziell können sie in einen einzigen Punkt zusammenfallen.

Die beiden geradlinigen Punktreihen können zur Deckung gebracht werden, durch eine Umwendung, welche einen Punkt der einen Reihe in den entsprechenden Punkt der anderen Reihe überführt, und eine Spiegelung, wo letzterer Punkt fest bleibt. Der Satz wird hiermit auf den vorhergehenden zurückgeführt.

45. Wenn eine Um-Bewegung einen Punkt fest stehen lässt, ist sie einer einfachen Spiegelung gleichwertig.

Die bei der Bewegung fest stehende Gerade g muss nach dem vorigen Satz den Punkt O enthalten, und die Bewegung ist somit eine Spiegelung, deren Achse senkrecht auf g in O steht.

8. Involution von drei Spiegelungen.

46. Wir haben schon gesehen, dass die Aufeinanderfolge von drei Spiegelungen, deren Achsen durch denselben Punkt hindurchgehen, oder senkrecht auf derselben Gerade stehen, einer einzigen Spiegelung gleichwertig ist. Wir wollen jetzt den allgemeinen Fall untersuchen, wo die Aufeinanderfolge von 3 Spiegelungen a, b, c eine involutorische Bewegung darstellt. Diese Bewegung muss dann eine Spiegelung sein, weil es keine andere involutorische Um-Bewegung gibt.

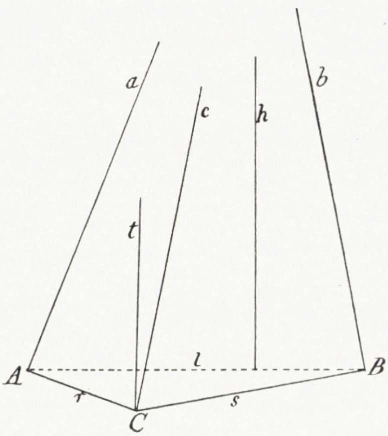


Fig. 3.

Durch den Punkt C auf c fällen wir die Lote r, s auf a bezw. b. Sie schneiden a und b in A bezw. B. Es gilt dann

$$acb = arrcssb = (ar)(rcs)(sb) = A(rcs)B.$$

rcs ist einer Spiegelung t gleichwertig ($rcs = t$, $rt = cs$), und es folgt

$$acb = AtB.$$

Diese Transformation soll nun involutorisch sein, und das wird nach 26 bedeuten, dass das Lot durch A auf t durch B hindurchgehen muss.

Die Achse h der Spiegelung acb ist senkrecht auf l, und es gilt $At = hB$. Die beiden entsprechenden »Abstände« sind einander gleich.

47. Es seien zwei Geraden a, b und ein Punkt C beliebig vorgelegt. Durch C wollen wir eine Gerade c legen, welche mit a und b in Involution ist (d. h. die Bewegung abc — und somit abc, bca, u. s. w. — soll involutorisch sein). Durch C fällen wir die Lote r, s auf a, b. Die Schnittpunkte mit a, b sind A, B. Ist nun l eine Verbindungsgerade von A und B, und fällen wir durch C das Lot t zu l, und konstruieren wir ferner die Gerade c so, dass

$$c = rts,$$

wird diese Gerade der vorgeschriebenen Bedingung genügen. Es wird nämlich

$$acb = AtB,$$

was eben eine involutorische Bewegung vorstellt. Umgekehrt weiss man, dass es keine anderen Linien gibt als diejenigen, die auf diese Weise erhalten werden.

48. Haben die beiden Punkte A, B keine Verbindungsgerade, hat die Aufgabe keine Lösung. Haben sie mehrere Verbindungsgeraden, entspricht jeder von diesen eine Gerade c, welche der Aufgabe genügt.

In dem Falle, wo a und b einen Punkt O und nur diesen Punkt, gemein haben, werden die Verbindungsgeraden $c = OC$ und $l = AB$ einander so entsprechen, dass die Geraden rcs und l paarweise aufeinander senkrecht stehen.

49. Schliesslich wollen wir noch eine andere Konstruktion von der gesuchten Gerade c ableiten: Aus

$$abc = cba,$$

folgt

$$C^{abc} = C^{cba}$$

d. h.

$$(C^{ab})^c = C^{ba};$$

c wird also Spiegelungsachse der beiden Punkte C^{ab} und C^{ba}.[1]

[1] Vgl. die grundlegende Konstruktion in Math. Ann. 64, S. 450.

9. Das Dreieck.

50. Es sei vorgelegt ein Dreieck *ABC*, d. h. ein System von drei Punkten *A*, *B*, *C* mit bestimmten festgelegten Verbindungsgeraden *BC*, *CA*, *AB*. Diese Verbindungsgeraden (die Seiten des Dreiecks) sollen mit *a*, *b*, *c* bezeichnet werden. Es gibt bestimmte Mittelpunkte *R*, *S*, *T* für *BC*, *CA*, *AB*. Diese Mittelpunkte sind immer eindeutig durch die Punkte *A*, *B*, *C* bestimmt (unabhängig von den Verbindungsgeraden, die mehrdeutig sein können). Die Mittelsenkrechten durch *R*, *S*, *T*, den Seiten *a*, *b*, *c* entsprechend, bezeichnen wir mit *r*, *s*, *t*. Die Bewegung *rst* ist eine Um-Bewegung, und da der Punkt *B* bei dieser Bewegung fest steht, kann die Bewegung durch eine Spiegelung ersetzt werden (45). Die Bewegung *rst* ist also involutorisch. Haben zwei von den drei Geraden *r*, *s*, *t* einen eindeutig bestimmten Schnittpunkt, muss die dritte durch denselben Punkt hindurch gehen.

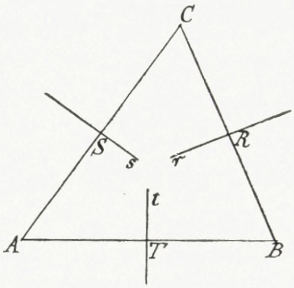


Fig. 4.

Die Bewegung *RtS* ist eine Um-Bewegung, und zwar, da der Punkt *C* fest bleibt, eine Spiegelung. Es gibt sonach eine Gerade durch *R* und *S*, welche senkrecht auf *t* steht. Dies ist die allgemeine Form des Mittelpunktstransversalensatzes.

51. *ABC* ist ein Dreieck mit festgelegten Seiten *a*, *b*, *c*. Ob zwei beliebige von diesen Seiten, z. B. *b* und *c* überhaupt eine Spiegelungsachse haben, können wir nach unseren Voraussetzungen nicht wissen. Haben sie aber eine, haben sie auch eine andere senkrecht zu der ersteren. Wir nehmen nun an, dass jedes Paar *bc*, *ca*, *ab* zwei Spiegelungsachsen hat: Durch *A* gehen zwei Spiegelungsachsen x, x_1 für *b* und *c*; durch *B* zwei Achsen y, y_1 für *c* und *a*,

und durch C zwei Achsen z, z_1 für a und b. Es seien nun die Bezeichnungen so gewählt, dass x, y, z nicht involutorisch sind. Es lässt sich dann zeigen, dass x, y, z_1 notwendig involutorisch sein müssen. Die Aufeinanderfolge der Spiegelungen xyz führt b in sich selbst über. Und die Aufeinanderfolge xyz_1 ebenso. Da nun ferner

$$xyz_1 = xyzC,$$

muss jedenfalls eine der Bewegungen xyz und xyz_1 einen festen Punkt auf b haben; das heisst, die Bewegung muss eine Spiegelung darstellen.

Ist also xyz keine Spiegelung, muss dies mit xyz_1 der Fall sein, und umgekehrt.

Ist also xyz nicht involutorisch, so wird xyz_1 involutorisch, xy_1z_1 hingegen nicht, während wiederum $x_1y_1z_1$ involutorisch wird. Hiermit haben wir dann die 4 Gruppen von involutorischen Linien, welche bei den 6 Spiegelungsachsen des Dreiecks auftreten:

$$xyz_1, \quad xy_1z, \quad x_1yz, \quad x_1y_1z_1.$$

52. Um die Höhen des Dreiecks, d. h. die Geraden, welche von den Ecken senkrecht zu den festgelegten gegenüberliegenden Seiten gefällt werden, zu untersuchen, betrachten wir die Aufeinanderfolge der Spiegelungen a, b, c. Die Bewegungen abc, bca, cab werden auseinander durch Spiegelungen gebildet, indem

$$(abc)^a = bca, \quad (bca)^b = cab.$$

Die erste Bewegung abc hat eine feste Gerade y, die zweite eine feste Gerade z, und die dritte cab eine feste Gerade x, und den obigen Gleichungen zufolge ergibt sich

$$y^a = z, \quad z^b = x, \quad x^c = y,$$

d. h. y und z haben die Spiegelungsachse a, z und x die Spiegelungsachse b, x und y die Spiegelungsachse c.

Die Gerade x muss durch die beiden Höhenfusspunkte N und P hindurchgehen. In der Tat, wenn C durch die Umwendung P in C' übergeht, so wird C' durch die Bewegung cab nach C geführt. Die feste Gerade der Bewegung cab ist x. Sie muss also den Mittelpunkt P von CC' enthalten. Ebenso zeigt sich, dass sie auch den Punkt N enthalten muss. Ähnlich zeigt sich, dass y durch M und P, z durch M und N hindurchlaufen. Die drei Geraden xyz bilden demnach die Seiten eines Dreiecks MNP, wo a, b, c Spiegelungsachsen der Seitenpaare sind. Die anderen Spiegelungsachsen sind die Höhen des gegebenen Dreiecks; und müssen nach dem vorhergehenden Satz involutorisch sein.

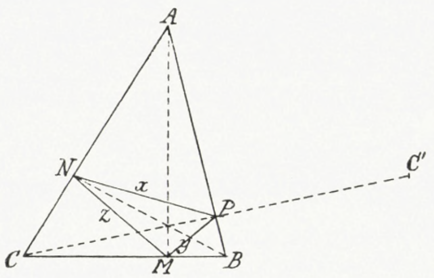


Fig. 5.

In der allgemeinen Kongruenzlehre gilt also folgender Satz:

Die Höhen eines in beliebiger Weise fixierten Dreiecks sind immer involutorisch.

Haben zwei von den Höhen einen eindeutig bestimmten Schnittpunkt, geht die dritte Höhe durch diesen Punkt hindurch.

53. Der Satz kann auch folgendermassen formuliert werden: Es werden 3 Punkte A, B, C vorgelegt (sie können untereinander verschieden sein oder nicht). Eine Gerade a gehe durch B und C, eine Gerade b durch C und A, eine Gerade c durch A und B (diese Geraden können vielleicht in mannigfacher Weise gewählt werden; wir setzen voraus, dass wir für jedes Paar der Punkte eine Verbindungs-

gerade herausgewählt haben). Die drei senkrechten von A, B, C auf bezw. a, b, c, werden dann immer involutorisch sein. Der Satz gilt für jede Lage von A, B, C; auch bei zusammenfallenden Punkten hat der Satz seine Gültigkeit. Überhaupt ist es bemerkenswert, dass die Sätze der Geometrie eine grössere Allgemeinheit erhalten, wenn das Eindeutigkeitsaxiom wegfällt, in dem Sinne, dass der Ausdruck »zwei Punkte« nunmehr nicht notwendig »zwei verschiedene Punkte« bedeutet.

10. Halbdrehungen.

54. Eine Halbdrehung um den Punkt O ist eine Transformation, welche folgendermassen durch zwei gerade Linien a, a_1, welche beide durch O gehen und nicht zueinander senkrecht sind, bestimmt wird. Jeder geraden Linie l soll eine neue gerade Linie l_1 entsprechen derart, dass das Lot m von O auf l, l in einem Punkt trifft, durch welchen l_1 gehen soll. Ferner soll l_1 zu derjenigen Geraden m_1 durch O senkrecht stehen, welche durch die Gleichung $mm_1 = aa_1$ bestimmt wird. Auf Grund dieser Festsetzung wird jeder geraden Linie l eine eindeutig entsprechende Linie l_1 zugewiesen. Speziell wird sich hieraus ergeben, dass jeder geraden Linie p durch O eine gerade Linie p_1 durch O entspricht, derart dass $pp_1 = aa_1$.

Ferner: Jedem Punkt P wird ein entsprechender Punkt P_1 zugewiesen, derart dass P_1 als Mittelpunkt von P und P', wo P' der zu P bei der Bewegung aa_1 entsprechende Punkt ist, bestimmt wird. Bei der so erklärten Transformation, wo jeder Geraden eine Gerade und jedem Punkt ein Punkt entsprechen wird, zeigt sich nun, das eine Gerade l und ein Punkt P in ihr, in eine Gerade l_1 und einen in dieser gelegenen Punkt P_1 übergehen werden. l_1 ist nämlich die feste Gerade bei der Bewegung $lmm_1 = laa_1$, und es folgt dann, dass P_1 in l_1 enthalten ist (43).

Der Punkt P_1 lässt sich aus P in folgender Weise ableiten: Man zieht eine Gerade $OP = p$, bestimmt die Gerade p_1 mittels der Gleichung $pp_1 = aa_1$; P_1 wird nun als Fusspunkt des Lotes von P auf p_1 bestimmt. Die Konstruktion setzt voraus, dass die Punkte O und P wenigstens eine Verbindungsgerade haben. Haben die Punkte O und P mehrere Verbindungsgeraden $p, q, r, \ldots.$, und die entsprechenden Linien bei der Halbdrehung mit $p_1, q_1, r_1, \ldots.$ bezeichnet werden, so müssen diese Linien alle den Punkt P_1 enthalten.

55. Zwei Halbdrehungen um denselben Punkt O sind miteinander vertauschbar. Dies folgt unmittelbar aus der Untersuchung in 46—48. Wir zeichnen hier die Figur auf (Fig. 6); aus

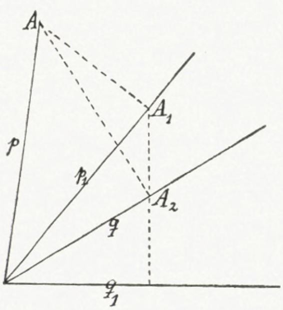


Fig. 6.

$$pp_1 = qq_1,$$

$$AA_1 = a_1 \perp p_1, \qquad AA_2 = a_2 \perp q,$$

folgt

$$A_1 A_2 q_1 = p_1 a_1 a_2 q q_1 = p_1 (a_1 a_2 p) p_1;$$

diese Bewegung ist eine involutorische Bewegung (weil $a_1 a_2 p$ eine Spiegelung ist). Nach 26 müssen dann die von A_1 und A_2 auf q_1 gefällten Lote zusammenfallen.

56. Die inverse Transformation einer Halbdrehung ist nicht immer eindeutig, und es wird nicht immer möglich sein, die Transformation auf beliebige Punkte oder gerade Linien anzuwenden. Wenn die in Rede stehenden Transformationen aber eindeutig und möglich sind, wird die Reihenfolge von beliebigen direkten oder inversen Halbdrehungen um denselben Punkt willkürlich sein, natürlich

unter der Voraussetzung, dass die hierdurch entstehenden Transformationen der in Betracht kommenden Punkten oder geraden Linien immer möglich und eindeutig sind.

11. Die Fixpunkte einer Bewegung.

56. Eine Um-Bewegung ist entweder eine Spiegelung oder nicht. Im ersten Falle gibt es unendlich viele Fixpunkte, nämlich die Punkte der Spiegelungsachse. Im zweiten Falle gibt es keinen Fixpunkt; es gibt wohl eine feste Gerade, aber keinen festen Punkt.

57. Hat eine In-Bewegung zwei Fixpunkte A und B, so lässt sich folgendes aussagen. Zunächst muss der Mittelpunkt von A und B auch fest bleiben, und der Punkt A^B (und B^A) ebenfalls; auf dieser Weise findet man schon eine ganze Reihe von Fixpunkten. Wir legen nun durch A und B zwei Geraden, bezw. a und b, welche einen eindeutig bestimmten Schnittpunkt C haben (a und b können z. B. senkrecht zueinander gewählt werden, es gibt aber auch andere Möglichkeiten). Der Punkt C wird bei unserer Bewegung notwendig fest bleiben. In der Tat weil A fest ist, lässt sich die vorgelegte Bewegung durch zwei Spiegelungen a und a_1 (durch A) ersetzen, ebenso durch zwei Spiegelungen b und b_1 (durch B). Aus $aa_1 = bb_1$, folgt aber

$$baa_1 = b_1,$$

und da b und a der Voraussetzung zufolge einen eindeutig bestimmten Schnittpunkt C haben, muss also a_1 durch diesen Punkt gehen (ebenso natürlich auch b_1). Der Punkt wird somit fest. Gleichzeitig hat sich herausgestellt, dass unsere Bewegung durch zwei Spiegelungen a, a_1 dargestellt werden kann, deren Achsen die Punkte A und C, also mehrere Punkte gemein haben.

Wir können deshalb folgenden Satz aussprechen: Hat eine In-Bewegung zwei Fixpunkte (*A*, *B*), ist sie entweder die Identität, oder sie lässt sich durch zwei verschiedene Spiegelungen a, a_1 darstellen, deren Achsen unendlich viele Punkte gemein haben. Die gemeinsamen Punkte von a und a_1 sind alle Fixpunkte. Jede Gerade, welche einen Fixpunkt enthält, enthält unendlich viele Fixpunkte. Zwei Geraden, welche Fixpunkte enthalten, schneiden sich, wenn der Schnittpunkt eindeutig ist, in einem Fixpunkt.

58. Unter die Fixpunktmenge (*A*, *B*) verstehen wir nun alle Punkte, die notwendig fest bleiben bei jeder Bewegung, wo *A* und *B* fest bleiben.

Nach den eben erwähnten Eigenschaften leuchtet unmittelbar ein, dass die Fixpunktmenge eine Geometrie ausmachen, für welche unser ursprünglich aufgestelltes Axiomsystem gültig wird, wenn wir unter gerade Linie die Punkte der Menge verstehen, welche einer ursprünglichen geraden Linie angehören, und unter Bewegung eine beliebige Transformation der Menge (*A*, *B*) in sich, die durch eine ursprüngliche Bewegung erzeugt wird.

59. Es folgt nun auch der folgende Satz:

Wenn zwei Punkte *A* und *B* mehr als eine Verbindungsgerade aufweisen, so gibt es eine ganze Reihe von Punkten, welche allen Geraden durch *A* und *B* angehören. Diese Reihe von Punkten wird durch den Schnitt der Fixpunktmenge (*A*, *B*) mit einer beliebigen Geraden durch *A* und *B* erzeugt.

Ist g eine beliebige Gerade durch *A* und *B*, und *C* ein

Punkt der Menge (A, B) ausserhalb g, so wird jede Gerade h durch C, welche g eindeutig schneidet, im Schnittpunkt mit g einen Punkt der genannten Reihe erzeugen.

60. Von anderen aus unseren Untersuchungen sofort fliessenden Tatsachen sollen nur noch hervorgehoben werden:

Wenn die Punkte A und B mehr als eine Verbindungsgerade haben, und zwei Geraden a und b, welche durch A bezw. B hindurchgehen, einander eindeutig in C schneiden, so haben A und C (bezw. B und C) mehrere Verbindungsgeraden.

61. Wenn 3 Geraden a, b, c in Involution sind, und a und b einander eindeutig in C schneiden, so wird c notwendig durch C gehen. Wenn aber a und b zwei Punkte C und D gemein haben (also eine ganze Reihe von gemeinsamen Punkten haben) so lässt sich von c nur behaupten, dass sie Fixpunkte der Bewegung ab enthalten muss. Fällt man von C das Lot a_1 auf c (Schnittpunkt E), so lässt sich schreiben

$$ab = a_1 b_1,$$

wo b_1 eine Gerade durch C bedeutet.

Es wird nun

$$abc = a_1 b_1 c,$$

und weil $a_1 \perp c$, haben die Geraden a_1 und c einen eindeutigen Schnittpunkt E; b_1 muss dann durch E gehen, und E wird somit ein Fixpunkt der Bewegung $a_1 b_1$, d. h. der Bewegung ab.

62. Haben zwei Geraden p, q mehrere Punkte $(A, B, \ldots)$ gemein, so werden die Lote x, y, welche von einem beliebigen Punkte P aus auf p und q

gefällt werden, auch mehrere Punkte gemein haben.

Es ist nämlich

$$ABx = yyABx = y(yAB)x = yzx,$$

wo $z \perp q$. Da nun die Bewegung ABx involutorisch ist (eine Spiegelung), wird auch yzx involutorisch. Hätten also

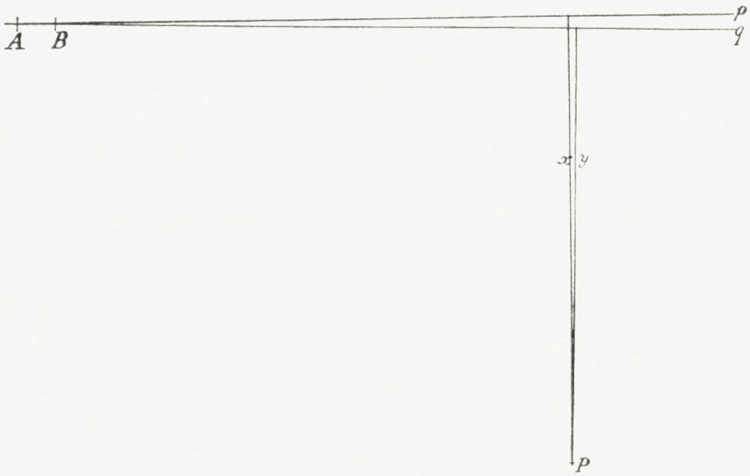


Fig. 7.

x und y nur den einzigen Punkt P gemein, müsste z durch P gehen, d. h. es gäbe zwei Lote y, z von P auf q, was unmöglich ist. Die beiden Geraden x, y müssen deshalb mehrere Punkte gemein haben.

In ähnlicher Weise zeigt man den folgenden Satz:

Haben zwei Geraden p, q zwei gemeinsame Lote r, s (also unendlich viele), so müssen die Lote x, y, welche von einem beliebigen Punkt P auf p und q gefällt werden, (entweder ganz zusammenfallen oder) unendlich viele Punkte gemein haben.

Es ist nämlich

$$rsx = yyrsx = y(yrs)x = yzx.$$

Schlusswort der ersten Mitteilung.

Die im vorhergehenden entwickelten allgemeinen Hilfsmittel werden nun zunächst für die Begründung der Geometrie in dem Falle, wo das Eindeutigkeitsaxiom erfüllt ist, d. h. wo zwei Punkte eine und nur eine Verbindungsgerade haben (oder etwa: höchstens eine Verbindungsgerade haben), anzuwenden sein. Es handelt sich hier von einer leichten Revision meiner Arbeit aus 1907 (Math. Ann. 64). Diese Revision ist aber sehr wichtig, und soll deshalb in ihren wesentlichen Einzelheiten in einer folgenden Mitteilung gegeben werden. Das wesentlich Neue wird das vollständige Unterdrücken der Axiome der Anordnung; aber auch andere Fragen kommen in Betracht.

Der nächste Schritt soll darauf hinausgehen, die Geometrie in dem Falle zu entwickeln, wo es sowohl Punktepaare AB mit einer eindeutig bestimmten Verbindungsgerade als auch Punktepaare AB mit mehreren Verbindungsgeraden vorkommen. Im ersten Falle wollen wir sagen, dass der Abstand AB gross ist, im zweiten Falle, dass der Abstand klein ist (oder, dass A und B »Nachbarpunkte« sind). Wir haben durch diese Namen schon angedeutet, in welche Richtung hin die Lösung sich gestalten wird. Jedem Punkt A wird ein Nachbargebiet Ⓐ zugewiesen, welches aus allen Nachbarpunkten des Punktes A besteht, und jeder Geraden g wird ein Nachbargebiet Ⓖ zugewiesen, welches aus allen Nachbarpunkten der Punkte von g bestehen. Bezeichnet man nun Ⓐ als »Grosspunkt«, Ⓖ als »Grossgerade«, wird man eine »Grossgeometrie« mit diesen Elementen entwickeln können, wo genau dasselbe Axiomsystem, wie das für unsere ursprüngliche Geometrie aufgestellte, erhalten werden kann. Zudem ergibt sich, dass für diese »Grossgeometrie« das Eindeutigkeitsaxiom gültig wird. Die Grossgeometrie wird also unmittelbar durch die früheren Untersuchungen zugänglich. Aus jedem Satz der gewöhnlichen projektiven Geometrie (und projektiven Trigonometrie) lässt sich sofort einen Satz unserer allgemeinen Geometrie ableiten, nämlich einen Satz von »Grosspunkten« und »Grosslinien«. Der letzte Teil unserer Untersuchungen wird nun darin bestehen, die Geometrie innerhalb des »Grosspunktes« und innerhalb der »Grosslinie« zu entwickeln. Hiervon soll nur an dieser Stelle gesagt werden, dass die Untersuchungen in mancher Hinsicht an infinitesimalgeometrische Untersuchungen erinnern werden, obgleich sie weit allgemeiner sind als die gewöhnlich bekannte Infinitesimalgeometrie, die, wie in der Einleitung angedeutet, als Anwendungen in unsere allgemeine Geometrie eingehen können.

Der Fall, wo alle Punktepaare unserer Geometrie mehrdeutige Verbindungsgeraden haben, wird durch die letztgenannten Untersuchungen auch erledigt werden.

Forelagt paa Mødet den 19. Oktober 1928.
Færdig fra Trykkeriet den 8. Maj 1929.

MATHEMATISK-FYSISKE MEDDELELSER

UDGIVNE AF

DET KGL. DANSKE VIDENSKABERNES SELSKAB

7. BIND (Kr. 17,70):

8. BIND (Kr. 17,20):

Kr. Ø.

2. Rasch, G.: Beitrag zur Theorie der unvollständigen Gammafunktionen. Nach hinterlassenen Papieren von J. L. W. V. Jensen. 1927 .... 0.85
3. Knudsen, Martin: Thermal Molecular Pressure in Tubes. 1927 1.50
4. Pedersen, P. O.: Den højere Atmosfæres Sammensætning, Tryk, Temperatur og elektriske Ledningsevne i Belysning af Radiobølgernes Udbredelsesforhold. 1927 .... 0.60
5. Besicovitch, A. and Bohr, H.: Some remarks on Generalisations of almost Periodic Functions. 1927 .... 1.50
6. Pedersen, Ellen: Über einige besondere Klassen von fastperiodischen Funktionen. 1928 .... 1.00
7. Steffensen, J. F.: A general Summation Formula. 1928 ...... 1.25
8. Nielsen, Niels: Observations sur des recherches algébriques plus anciennes que le théorème d'Abel. 1928 .... 1.40
9. Buch Andersen, E. og Asmussen, R. W.: Undersøgelser over Faradayeffekten hos vandige Opløsninger af nogle uni-univalente Elektrolyter. 1928 .... 0.80
10. Pedersen, P. O.: On the Lichtenberg Figures. Part III. The positive Figures. With 28 Plates. 1929 .... 10.00
11. Hjelmslev, Johannes: Einleitung in die allgemeine Kongruenzlehre. Erste Mitteilung. 1929 .... 1.60

9. BIND (under Pressen):

1. Bjerrum, Niels, und Unmack, Augusta: Elektrometrische Messungen von Wasserstoffelektroden in Mischungen von Säuren und Basen mit Salzen. (Under Pressen).
2. Hartmann, Jul.: The Jet-Wave. Theory of the Periodic Jet-Wave. (Under Pressen).
3. Strömgren, Elis: Asymptotische Lösungen im restringierten Dreikörperproblem (Problème restreint). Mit 1 Tafel. 1929 .. 2.00
4. Hartmann, Jul.: The Jet-Chain- and the Jet-Wave-Vibrator. (Under Pressen).
5. Pedersen, P. O.: Wireles Echoes of Long Delay. (Under Pressen).
6. Strömgren, Elis: Fortsetzung und Abschluss der Librationen um L_2 und L_3 im restringierten Dreikörperproblem (Problème restreint). Mit 1 Tafel. (Under Pressen).