

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

---

DIE STRUKTUR PERIODISCHER  
TRANSFORMATIONEN  
VON FLÄCHEN

VON

JAKOB NIELSEN



KØBENHAVN  
LEVIN & MUNKSGAARD  
EJNAR MUNKSGAARD

1937

Printed in Denmark.  
Bianco Lunos Bogtrykkeri A/S.

Wenn eine umkehrbar eindeutige und stetige Transformation  $\tau$  einer Mannigfaltigkeit  $\varphi$  in sich die Eigenschaft hat, dass die  $n$  Male iterierte Transformation  $\tau^n$  jeden Punkt von  $\varphi$  in sich transformiert, während es andererseits Punkte  $P$  von  $\varphi$  gibt, für welche das Punktsystem  $P, \tau P, \dots, \tau^{n-1}P$  aus  $n$  verschiedenen Punkten besteht, nennt man  $\tau$  eine *periodische Transformation  $n$ -ter Ordnung*. Ist  $\mathcal{G}$  eine topologische Abbildung von  $\varphi$  auf eine mit  $\varphi$  homöomorphe Mannigfaltigkeit  $\varphi'$ , so ist  $\tau' = \mathcal{G}\tau\mathcal{G}^{-1}$  eine periodische Transformation  $n$ -ter Ordnung von  $\varphi'$  in sich, die als mit der Transformation  $\tau$  von  $\varphi$  *topologisch äquivalent* bezeichnet wird. Eine Gesamtheit topologisch äquivalenter fassen wir zu einem *topologischen Typus periodischer Transformationen* zusammen.

Periodische Transformationen treten bei mannigfachen Fragen der Geometrie und der Funktionentheorie auf. Beispielsweise stellt die birationale Transformation einer algebraischen Kurve von einem Geschlecht grösser als 1 in sich eine periodische Transformation der zugehörigen Riemannschen Fläche dar. BROUWER griff die Untersuchung periodischer Transformationen (und endlicher Gruppen von Transformationen) auf allgemeiner topologischer Grundlage an und untersuchte insbesondere Kugelfläche und Torus [1, 2, 3, 4]<sup>1</sup>. B. v. KERÉKJÁRTÓ wies die einfache Struktur der periodischen Transformationen der Kreisscheibe und

<sup>1</sup> Zahlen in eckigen Klammern weisen auf das am Schluss der Arbeit angeführte Literaturverzeichnis hin.

Kugelfläche nach, wobei sich als mögliche topologische Typen nur die der Drehungen, Spiegelungen und Drehspiegelungen ergaben [5]; siehe auch [6], Abschnitt 6. In [7] behandelte er die endlichen Gruppen topologischer Transformationen der Kugelfläche mit Erhaltung der Orientierung. W. SCHERRER [8, 9, 10] gab eine ausführliche Entwicklung der Hilfsmittel der Untersuchung periodischer Transformationen von Flächen (auch nicht-orientierbaren) und der endlichen Gruppen von Transformationen, teils auf Grund einer ihm eigenen Methode, teils im Anschluss an den von BROUWER geschaffenen Begriff der *Modulfläche* und unter systematischer Verwendung der *Monodromiegruppe*. Von seinen Resultaten seien hier hervorgehoben einerseits die Aufstellung aller endlichen Gruppen für Kugelfläche und projektive Ebene, andererseits die erschöpfende Behandlung des Spezialfalles der Ordnung 2. In dem letzteren Fall verschafft sein Begriff der »minimalen halbierenden Schnittsysteme« einen Einblick in die Struktur dieser involutorischen Transformationen und gestattet, die topologische Äquivalenz für diese zu entscheiden. Im Anschluss an seine Arbeiten konnte F. STEIGER [11] als topologische Verallgemeinerung eines von WIMAN für algebraische Kurven gefundenen Satzes den maximalen Wert der Ordnung  $n$  periodischer Transformationen geschlossener Flächen in Abhängigkeit von dem topologischen Charakter der Fläche herleiten. Dass dieser Satz über die maximale Ordnung sich für Transformationen mit Erhaltung der Orientierung als ein Satz über *Abbildungsklassen* erweist, habe ich in [12] dargetan.

Die vorliegende Abhandlung setzt sich das Ziel, für orientierbare, geschlossene oder berandete Flächen die Struktur periodischer Transformationen mit Erhaltung der Orientierung zu untersuchen. Mit dem Ausgangspunkt in den Be-

griffen der Modulfläche und der Monodromiegruppe wird für den in der Fundamentalgruppe durch die periodische Flächenabbildung induzierten Automorphismus eine *Normaldarstellung* gegeben, die alle Fälle umfasst, und aus der in jedem Einzelfall die Struktur der Abbildung durch Vertauschung von Bereichen auf der universellen Überlagerungsfläche, und damit auf der Fläche selbst, erkennbar ist. Hieraus ergibt sich weiter eine notwendige und hinreichende Bedingung für die topologische Äquivalenz zweier periodischer Transformationen und damit die Möglichkeit einer vollständigen *Aufzählung aller topologischen Typen* periodischer Transformationen. Einen direkten Weg zur Entscheidung dieses Äquivalenzproblems habe ich in der dänisch geschriebenen Arbeit [13] dargestellt. — Aus der Normaldarstellung des Automorphismus der Fundamentalgruppe liest man weiter den zugehörigen Automorphismus endlicher Ordnung der *Homologiegruppe* ab, der sich in einer gegenüber Auswahl einer Homologiebasis invarianten Form durch Angabe des Systems der Eigenwerte, also des zugehörigen *charakteristischen Polynoms* vollständig beschreiben lässt. Dieses Polynom hängt nur ab von der Ordnung der Transformation, der Anzahl der Fixpunkte ihrer Potenzen und dem Geschlecht und der Ränderzahl der Modulfläche; und umgekehrt gestattet es, für geschlossene Flächen diese Zahlen zu bestimmen. Insbesondere setzt dieses Polynom die Aussage der bekannten *Spurformel* von ALEXANDER [16], LEFSCHETZ [17] und HOPF [18] für den Fall periodischer Transformationen von Flächen in Evidenz. Man gewinnt dadurch die Erkenntnis, dass der Automorphismus der Homologiegruppe, der wegen seines algebraischen Charakters besonders zugänglich ist, die periodische Flächentransformation sehr weitgehend, wenn auch nicht vollständig, festlegt.

Die Arbeit gliedert sich wie folgt:

1. Die Modulfläche.
2. Die Valenz von Kurven und Verzweigungspunkten.
3. Zurückführung auf den Fall der Unverzweigkeit.
4. Homomorphismus der Fundamentalgruppe von  $M$ .
5. Spezielle Wahl der Erzeugenden von  $T$  für  $q > 0$ .
6. Automorphismus der Fundamentalgruppe von  $\varphi$ .
7. Normaldarstellung im Falle  $q > 0$ .
8. Normaldarstellung im Falle  $q = 0$ .
9. Zusätzliche Relationen bei Verzweigkeit.
10. Strukturbeschreibung auf Grund der Normaldarstellung.
11. Der Äquivalenzsatz.
12. Automorphismus der Homologiegruppe für  $q > 0$ .
13. Automorphismus der Homologiegruppe für  $q = 0$ .
14. Charakteristisches Polynom und Spurformel.
15. Kennzeichnung der Abbildung durch das charakteristische Polynom.

Literaturverzeichnis.

**1. Die Modulfläche.** Eine geschlossene oder von endlich vielen Randkurven berandete, orientierbare Fläche  $\varphi$  sei einer periodischen, topologischen und die Orientierung erhaltenden Abbildung  $\tau$  auf sich unterworfen.  $n > 1$  sei die *Ordnung* von  $\tau$ , d. h. die kleinste positive ganze Zahl, für welche  $\tau^n$  die identische Abbildung von  $\varphi$  auf sich ist. Ist  $P$  ein Punkt von  $\varphi$ , so besteht das Punktsystem

$$(1.1) \quad P, \tau P, \tau^2 P, \dots, \tau^{n-1} P$$

aus einer gewissen Anzahl  $m$  verschiedener Punkte von  $\varphi$ . Dabei ist  $1 \leq m \leq n$  und  $m$  ein Teiler von  $n$ . Es gibt Punkte von  $\varphi$ , für welche  $m = n$  ist; diese mögen *einfache*

Punkte von  $\varphi$  heissen. Durchläuft  $P$  ganz  $\varphi$ , so lässt sich die Menge der Punktsysteme (1.1) nach BROUWER [1] umkehrbar eindeutig und stetig auf alle Punkte  $\{P\}$  einer ebenfalls orientierbaren Fläche  $M$  abbilden, die als *Modulfläche* der periodischen Abbildung  $\tau\varphi$  bezeichnet wird.  $M$  ist geschlossen oder berandet, je nachdem  $\varphi$  geschlossen oder berandet ist. Durch die angegebene Beziehung ist  $\varphi$  als eine  $n$ -blättrige Überlagerungsfläche von  $M$ , eine Riemannsche Fläche über  $M$ , erklärt.

Die über dem Spurpunkt  $\{P\}$  von  $M$  liegenden insgesamt  $m$  verschiedenen Punkte des Systems (1.1) lassen sich durch

$$(1.2) \quad P, \tau P, \tau^2 P, \dots, \tau^{m-1} P$$

repräsentieren. Diese  $m$  Punkte von  $\varphi$  sind Fixpunkte bei  $\tau^m$  und den Potenzen dieser Abbildung und werden, falls  $m > 1$  ist, durch alle anderen Potenzen von  $\tau$  zyklisch vertauscht. Wir setzen

$$(1.3) \quad n = m\lambda$$

und sprechen den Punkten des Systems (1.2) die *Multiplizität*  $\lambda$  zu. Die gleiche Bezeichnung werde auch ihrem Spurpunkt  $\{P\}$  auf  $M$  beigelegt, wie wir überhaupt im Folgenden Benennungen, die sich auf die Überlagerung von  $M$  durch  $\varphi$  beziehen, in gleicher Weise den Punkten oder Kurven auf  $\varphi$  und ihren Spurpunkten oder Spurkurven auf  $M$  beilegen.

Einfache Punkte haben die Multiplizität 1. *Multiple Punkte*, d. h. Punkte mit einer Multiplizität  $\lambda > 1$ , liegen, wie BROUWER gezeigt hat, isoliert und sind *Verzweigungspunkte* von  $\varphi$  über  $M$ . Solche können nicht auf Randkurven liegen. Rückt ein variabler Punkt  $\{P\}$  auf  $M$  innerhalb

einer ganz aus einfachen Punkten bestehenden Umgebung eines Verzweigungspunktes in diesen hinein, so rücken auf  $\varphi$  je  $\lambda$  von den darüberliegenden  $n$  Punkten zusammen und vereinigen sich zu einem Punkt der Multiplizität  $\lambda$ . Man pflegt die Zahl  $\lambda - 1$  als die zugehörige »Verzweigungsordnung« zu bezeichnen. — *Fixpunkte* bei  $\tau$  sind Punkte mit  $m = 1$ ,  $\lambda = n$ , und nur solche. — Falls es nur einfache Punkte ( $m = n$ ,  $\lambda = 1$ ) gibt, heisst  $\varphi$  *unverzweigt* über  $M$ ; in diesem Falle sind die Abbildungen  $\tau$ ,  $\tau^2$ , ...,  $\tau^{n-1}$  sämtlich fixpunktfrei.

Zur näheren Begründung der Eigenschaften der Modulfläche, an die in diesem Paragraphen erinnert wurde, vergleiche man ausser [1] auch den Abschnitt 6, § 6 des Lehrbuches [6] von B. v. KERÉKJÁRTÓ.

## 2. Die Valenz von Kurven und Verzweigungspunkten.

Es sei  $\alpha$  eine einfache, geschlossene, gerichtete Kurve auf  $M$ , die eventuell vorhandene Verzweigungspunkte vermeidet, und  $P$  ein Punkt von  $\varphi$ , dessen Spurpunkt  $\{P\}$  auf  $\alpha$  liegt. Man durchlaufe von  $P$  aus einen Weg auf  $\varphi$  über der Spur  $\alpha$ . Dieser möge sich erstmalig nach  $\lambda$ -fachem Umlauf um  $\alpha$  schliessen; es entsteht dabei eine einfache, geschlossene, gerichtete Kurve  $k$  auf  $\varphi$ , die eine  $\lambda$ -fache Überdeckung von  $\alpha$  darstellt. Die Zahl  $\lambda$  heisse die *Multiplizität* der Kurven  $\alpha$  oder  $k$ , und im Falle  $\lambda > 1$  werden diese als *multiple Kurven* bezeichnet. Das Bild  $\tau k$  von  $k$  bei  $\tau$  ist ebenfalls eine einfache, geschlossene, gerichtete Kurve auf  $\varphi$ , die eine  $\lambda$ -fache Überdeckung von  $\alpha$  darstellt. Da auf  $k$  genau  $\lambda$  Punkte über  $\{P\}$  liegen, ist  $\tau k$  dann und nur dann mit  $k$  identisch, wenn  $\lambda = n$ . Für  $\lambda < n$  ist  $\tau k$  zu  $k$  fremd. Ist auch  $2\lambda < n$ , so ist  $\tau^2 k$  zu  $k$  und  $\tau k$  fremd. Durch Fortsetzung dieses Schlusses sieht man, dass  $\lambda$  ein Teiler



von  $n$  ist; setzt man  $n = m\lambda$  wie in (1.3), so liegen über  $z$  die  $m$  getrennten Kurven

$$(2.1) \quad k, \tau k, \tau^2 k, \dots, \tau^{m-1} k,$$

und es ist  $\tau^m k = k$ . Die Deckkurven von  $z$  haben also alle die gleiche Multiplizität  $\lambda$ , und die Multiplizität ist daher eine  $z$  zukommende Eigenschaft. Man drückt bekanntlich diese Eigenschaft der Überlagerung dadurch aus, dass man  $\varphi$  als eine *reguläre* Riemannsche Fläche über  $M$  bezeichnet.

Auf der Kurve  $k$  liegen die  $\lambda$  Punkte

$$(2.2) \quad P, \tau^m P, \tau^{2m} P, \dots, \tau^{(\lambda-1)m} P$$

über dem Spurpunkt  $\{P\}$ . Es sei  $\tau^{\sigma m} P$  derjenige unter den Punkten (2.2), der bei positivem Umlauf um  $k$  zunächst auf  $P$  folgt. Bei  $\tau^{\sigma m}$  wird  $k$  so auf sich abgebildet, dass das Punktsystem (2.2) in seiner Anordnung auf  $k$  eine zyklische Vertauschung um einen Schritt erleidet. Diese Anordnung auf  $k$  wird daher durch die Reihenfolge

$$(2.3) \quad P, \tau^{\sigma m} P, \tau^{2\sigma m} P, \dots, \tau^{(\lambda-1)\sigma m} P$$

angegeben, wobei Exponenten von  $\tau$  nur mod  $n$  in Betracht kommen. Da  $\tau^m P$  unter den Punkten (2.3) vorkommt, muss

$$(2.4) \quad (\sigma, \lambda) = 1$$

sein, indem dies Klammersymbol den grössten gemeinsamen Teiler bezeichnet. Man bestimme  $\delta$  so, dass

$$(2.5) \quad \delta\sigma \equiv 1 \pmod{\lambda}$$

ist; dadurch ergibt sich die Stelle, an welcher der Punkt  $\tau^m P = \tau^{\delta\sigma m} P$  in der Reihe (2.3) auftritt. Die Operation  $\tau^m$ ,

bei welcher die Kurve  $k$  aus dem Kurvensystem (2.1) erstmalig auf sich abgebildet wird, ist für diese Kurve eine periodische Transformation der Ordnung  $\lambda$ , und zwar ist  $\tau^m k$  offenbar topologisch äquivalent mit der Drehung einer Kreislinie in sich um  $\frac{\delta}{\lambda}$  einer Volldrehung, wobei  $\delta$  wegen (2.5) zu  $\lambda$  teilerfremd ist. Da nun  $\tau^m \cdot \tau P = \tau \cdot \tau^m P$  ist, folgt, dass zu der periodischen Transformation  $\tau^m$  von  $\tau k$  und weiter von allen Kurven aus (2.1) die gleiche Drehzahl  $\frac{\delta}{\lambda}$  gehört. Daher ist auch  $\sigma$  eine unabhängig von der Auswahl von  $P$  zu  $z$  gehörige charakteristische Zahl. Die Zahlen  $\sigma$  und  $\delta$  kommen dabei nur als Restklassen mod  $\lambda$  in Betracht. Wir nennen  $\sigma m$  den zu der gerichteten Kurve  $z$  gehörigen *Monodromieexponenten* und schreiben

$$(2.6) \quad \mu(z) \equiv \sigma m \pmod{n}$$

oder auch nur

$$(2.7) \quad \mu(z) = \sigma m.$$

Kennt man den Monodromieexponenten  $\mu(z)$  einer Kurve  $z$  von  $M$ , so berechnet man

$$(2.8) \quad m = (\mu(z), n), \quad \lambda = \frac{n}{m}, \quad \sigma = \frac{\mu(z)}{m}.$$

Dann gilt (2.4), und mittels (2.5) kann man die Drehzahl berechnen. Denjenigen Charakter der Kurve  $z$  oder der Kurven des Systems (2.1), der sich in dem Monodromieexponenten, also ausführlicher in dem Zahlensystem (2.8) zu erkennen gibt, nennen wir die *Valenz* dieser Kurven.

Für nicht-multiple Kurven hat man  $\lambda = 1$ ,  $m = n$ , und man kann  $\sigma = 0$  setzen. Für  $\lambda = n$ ,  $m = 1$  hat man eine *Fixkurve* bei  $\tau\varphi$ , und der Monodromieexponent ist eine zu  $n$  teilerfremde Restklasse  $\sigma \pmod{n}$ .

Kehrt man die Durchlaufungsrichtung von  $z$  um, so

ändern sich  $m$  und  $\lambda$  nicht, während die Restklasse  $\sigma \bmod \lambda$  das Vorzeichen wechselt: Die neue Valenz ist der früheren *entgegengesetzt*.

Ersetzt man  $\tau$  durch  $\tau^\varrho$ ,  $(\varrho, n) = 1$ , so ändern sich  $m$  und  $\lambda$  nicht, und  $\sigma$  ist durch die Restklasse  $\sigma\varrho^{-1} \bmod \lambda$  zu ersetzen; denn die Drehzahl ist mit  $\varrho$  zu multiplizieren.

Deformiert man die Kurve  $\varkappa$  auf  $M$ , ohne sie über Verzweigungspunkte hinwegzuschieben, so bleibt die Valenz aus Gründen der Stetigkeit ungeändert. Ist  $\varkappa$  auf diese Weise auf einen Punkt von  $M$  zusammenziehbar, der nicht Verzweigungspunkt ist, so ist  $\varkappa$  nicht multipel.

Für den Fall berandeter orientierter Flächen treffen wir die Verabredung, die Randkurven immer so zu orientieren, dass sie die Fläche positiv beranden. Dann kann man in eindeutiger Weise von der Valenz der Ränder sprechen. Kehrt man die Flächenorientierung um, so geht die Valenz einer Randkurve in die entgegengesetzte über. — Ist  $M$  homöomorph einem Kreisring und  $\varphi$  unverzweigt über  $M$ , so haben die Randkurven entgegengesetzte Valenz.

Die letzte Bemerkung gibt uns die Möglichkeit, von der *Valenz eines Verzweigungspunktes* auf einer orientierten Fläche zu sprechen: Es ist die Valenz einer einfachen, geschlossenen Kurve, die den Verzweigungspunkt beliebig nahe auf  $M$  positiv umkreist. Bei Wechsel der Orientierung auf  $\varphi$  geht die Valenz eines Verzweigungspunktes in die entgegengesetzte über. Die Zahlen  $m$  und  $\lambda$  haben für den Verzweigungspunkt dieselbe Bedeutung wie in 1, und auch die Bedeutung der Drehzahl ist unmittelbar anschaulich zu erfassen. Vgl. den Rotationssatz von B. v. KERÉKJÁRTÓ [5, 6].

Geht man mittels einer topologischen Abbildung  $\mathcal{P}$  von  $\varphi$  auf eine Fläche  $\varphi'$  zu einer topologisch äquivalenten periodischen Transformation

$$(2.9) \quad \tau' = \mathcal{D}\tau\mathcal{D}^{-1}$$

von  $\varphi'$  über, so erhält die gerichtete Kurve  $k' = \mathcal{D}k$  von  $\varphi'$  offenbar dieselbe Valenz, wie die Kurve  $k$  von  $\varphi$ . Überträgt man die Orientierung von  $\varphi$  durch  $\mathcal{D}$  nach  $\varphi'$ , so wird eine Randkurve bzw. ein Verzweigungspunkt von  $\varphi$  durch  $\mathcal{D}$  in eine Randkurve bzw. einen Verzweigungspunkt von  $\varphi'$  mit gleicher Valenz abgebildet.

Zu den in diesem Paragraphen behandelten Begriffen der regulären Überlagerungsfläche und des Monodromie-exponenten vergleiche man [9] und [14].

### 3. Zurückführung auf den Fall der Unverzweigtheit.

Es sei  $V$  ein multipler Punkt von  $\varphi$  mit der Valenz  $[m, \lambda, \sigma]$ . Um den Spurpunkt  $\{V\}$  auf  $M$  grenzen wir durch eine Kurve  $\alpha$  ein Elementarflächenstück  $E$  von  $M$  ab, das abgesehen von  $\{V\}$  nur einfache Punkte enthält und nicht an den eventuell vorhandenen Rand von  $M$  reicht. Dabei sei die Kurve  $\alpha$  so orientiert, dass sie  $M - E$  positiv berandet, also  $\{V\}$  negativ umkreist. Die Valenz von  $\alpha$  ist also durch  $[m, \lambda, -\sigma]$  gegeben. Entfernt man die inneren Punkte von  $E$  aus  $M$  und die entsprechenden  $m$  Elementarflächenstücke um  $V, \tau V, \dots, \tau^{m-1}V$  aus  $\varphi$ , so entsteht eine neue Fläche  $\bar{\varphi}$  mit einer neuen Modulfläche  $\bar{M} = M - E$ . Dem Übergang von  $\varphi$  zu  $\bar{\varphi}$  entspricht die Ersetzung von  $m$  multiplen Punkten von  $\varphi$  durch  $m$  multiple Ränder von  $\bar{\varphi}$ . Nun genügt es offenbar, die Struktur der  $n$ -periodischen Transformation der Teilfläche  $\bar{\varphi}$  von  $\varphi$  zu untersuchen. Denn da die Struktur der  $n$ -periodischen Transformation des Überlagerungsgebildes von  $E$  durch das Valenzsymbol  $[m, \lambda, \sigma]$  vollständig beschrieben wird, kann man nach-

träglich wieder »die Löcher schliessen«, also die Transformation von  $\bar{\varphi}$  auf  $\varphi$  erweitern.

Indem wir auf diese Weise alle multiplen Punkte durch multiple Ränder ersetzen, reduzieren wir die Strukturuntersuchung auf den Fall der Unverzweigtheit, bei dem es nur einfache Punkte gibt. Die Abbildungen  $\tau$ ,  $\tau^2$ ,  $\dots$ ,  $\tau^{n-1}$  sind also sämtlich fixpunktfrei. Dabei legen wir nun folgende Bezeichnungen für die weitere Untersuchung fest:

$n$  die Ordnung der periodischen Transformation  $\tau$ ,

$p$  das Geschlecht der orientierbaren Fläche  $\varphi$ ,

$r$  die Ränderzahl von  $\varphi$ ,

$q$  das Geschlecht der Modulfläche  $M$ ,

$s$  die Ränderzahl von  $M$ ,

$[m_\nu, \lambda_\nu, \sigma_\nu]$  die Valenz der  $\nu$ -ten Randkurve von  $M$ .  
 $\nu = 1, 2, \dots, s$ .

Dabei sind diese Symbole nicht willkürlich wählbar: Da die  $\nu$ -te Randkurve von  $M$  durch  $m_\nu$  Ränder von  $\varphi$  überlagert wird, hat man

$$(3.1) \quad r = m_1 + m_2 + \dots + m_s.$$

Ferner hat man wegen der unverzweigten Überlagerung die folgende Beziehung zwischen den Charakteristiken:

$$(3.2) \quad 2p + r - 2 = n(2q + s - 2).$$

(Denkt man sich  $M$  irgendwie trianguliert, so ist die Anzahl der Seiten vermindert um die Anzahl der Ecken und Dreiecke gleich der Charakteristik  $2q + s - 2$ . Bei Durchdrückung dieser Triangulierung auf eine unverzweigte  $n$ -blättrige Überlagerungsfläche sind diese Anzahlen einfach mit  $n$  zu multiplizieren.)

Später ergeben sich noch weitere Relationen, insbesondere (4.6).

Falls alle Ränder in der oben beschriebenen Weise durch Ausschneidung von Verzweigungspunkten entstanden sind, erhält man durch Eintragung von (3.1) in (3.2) die Formel von RIEMANN-HURWITZ für geschlossene Riemannsche Flächen.

Kommen keine multiplen Ränder vor, so ist  $r = ns$  wegen (3.1) (speziell auch  $r = s = 0$ ), und aus (3.2) wird

$$(3.3) \quad p - 1 = n(q - 1) \quad (\text{für } r = ns).$$

#### 4. Homomorphismus der Fundamentalgruppe von $M$ .

Nach Wahl eines Punktes  $O$  auf  $M$  zeichne man auf  $M$  mit  $O$  als Anfangs- und Endpunkt ein kanonisches System von  $2q + s$  Kurven wie in Fig. 1. Dann stellen die Wege

$$(4.1) \quad \alpha_1, \beta_1, \alpha_2, \beta_2, \dots, \alpha_q, \beta_q, \gamma_1, \gamma_2, \dots, \gamma_s$$

ein Erzeugendensystem der Fundamentalgruppe  $T$  von  $M$  dar. Dabei hat man als einzige definierende Relation von  $T$  die Relation

$$(4.2) \quad k_1 k_2 \dots k_q \gamma_1 \gamma_2 \dots \gamma_s = 1,$$

wobei zur Abkürzung

$$(4.3) \quad k_i = \alpha_i \beta_i \alpha_i^{-1} \beta_i^{-1} \quad (i = 1, 2, \dots, q)$$

gesetzt ist.

Ist  $o$  ein Punkt von  $\varphi$  über der Spur  $O$ , so gelangt man bei einem Umlauf z. B. über der Spur  $\alpha_1$  von  $o$  nach  $\tau^{\mu(\alpha_1)} o$ ; vgl. die Definition der Operation  $\mu$  in (2.6). Einem geschlossenen Weg wie z. B.  $\alpha_1 \alpha_2$  entspricht dabei der Monodromieexponent  $\mu(\alpha_1) + \mu(\alpha_2)$ . Die Operation  $\mu$  stellt

also eine homomorphe Beziehung der Fundamentalgruppe  $T$  von  $M$  auf die additive Gruppe der Restklassen mod  $n$  dar. Die zu den Erzeugenden (4.1) gehörigen Monodromie-substitutionen müssen die ganze Monodromiegruppe er-

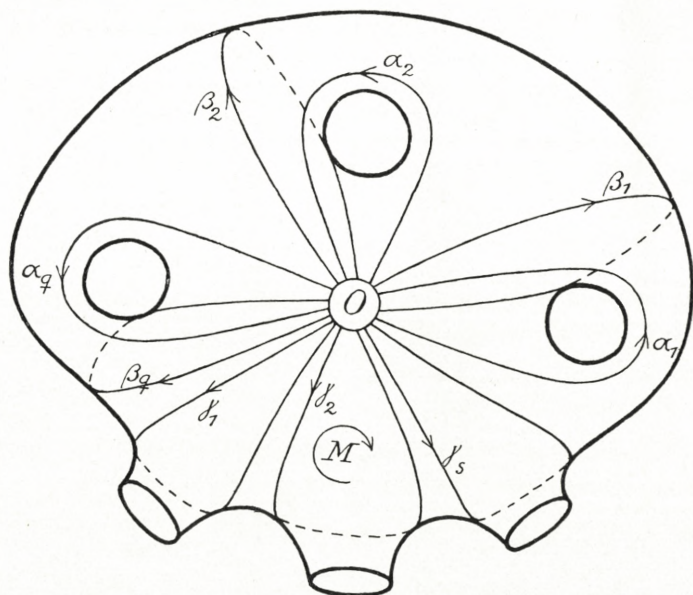


Fig. 1.

zeugen, da man ja  $o$  mit  $\tau o$  auf  $\varphi$  verbinden und den zugehörigen Spurweg auf  $M$  in der Fundamentalgruppe darstellen kann. Also muss

$$(4.4) \quad (\mu(\alpha_1), \dots, \mu(\beta_q), \mu(\gamma_1), \dots, \mu(\gamma_s), n) = 1$$

sein. Nach den Bezeichnungen von 3 ist dabei

$$(4.5) \quad \mu(\gamma_\nu) \equiv \sigma_\nu m_\nu \pmod{n} \quad (\nu = 1, 2, \dots, s)$$

gesetzt.

Die Operation  $\mu$  ist kommutativ. Bei ihrer Anwendung auf ein Element von  $T$  kommt es also nur darauf an, in

welche Restklasse nach der Kommutatorgruppe von  $T$  dies Element gehört. Insbesondere ist  $\mu(k_i) \equiv 0 \pmod n$ . Aus (4.2) und (4.5) hat man daher

$$(4.6) \quad \sigma_1 m_1 + \sigma_2 m_2 + \cdots + \sigma_s m_s \equiv 0 \pmod n$$

als weitere Relation zwischen den in 3 definierten Symbolen. Hat die Modulfläche das Geschlecht Null, so folgt aus (4.4) überdies

$$(4.7) \quad (\sigma_1 m_1, \sigma_2 m_2, \dots, \sigma_s m_s, n) = 1 \quad (\text{für } q = 0).$$

### 5. Spezielle Wahl der Erzeugenden von $T$ für $q > 0$ .

Die Anbringung des kanonischen Kurvensystems der Fig. 1 auf der Modulfläche  $M$  ist mit weitgehender Willkür verbunden. Wir werden im Folgenden topologische Abbildungen von  $M$  auf sich betrachten, die die Orientierung von  $M$  erhalten und  $O$  in sich überführen, und dabei untersuchen, wie sich die  $\mu$ -Werte des Erzeugendensystems (4.1) ändern. Diese Abbildungen von  $M$  auf sich setzen wir wiederum aus 8 erzeugenden Abbildungen  $S_1, \dots, S_8$  zusammen. Wir werden diese Abbildungen durch isomorphe Transformationen der Fundamentalgruppe  $T$  von  $M$  in sich beschreiben, indem wir angeben, welche Elemente von  $T$  jeweils für die ursprünglichen Erzeugenden zu substituieren sind; Erzeugende, die ungeändert bleiben, werden dabei nicht mit aufgeführt.

$$S_1: \quad \alpha_1 \rightarrow \alpha_1 \beta_1$$

$$S_2: \quad \beta_1 \rightarrow \beta_1 \alpha_1$$

$$S_3: \quad \alpha_1 \rightarrow \alpha_2$$

$$\beta_1 \rightarrow \beta_2$$

$$\alpha_2 \rightarrow k_2^{-1} \alpha_1 k_2$$

$$\beta_2 \rightarrow k_2^{-1} \beta_1 k_2$$



$$\begin{aligned}
 S_4: \quad & \alpha_1 \rightarrow \alpha_2 \\
 & \beta_1 \rightarrow \beta_2 \\
 & \alpha_2 \rightarrow \alpha_3 \\
 & \text{---} \\
 & \beta_{q-1} \rightarrow \beta_q \\
 & \alpha_q \rightarrow k_q^{-1} \dots k_2^{-1} \alpha_1 k_2 \dots k_q \\
 & \beta_q \rightarrow k_q^{-1} \dots k_2^{-1} \beta_1 k_2 \dots k_q \\
 \\
 S_5: \quad & \alpha_1 \rightarrow \alpha_1 \alpha_2 \beta_2^{-1} \\
 & \beta_1 \rightarrow \beta_2 \alpha_2^{-1} \beta_1 \alpha_2 \beta_2^{-1} \\
 & \alpha_2 \rightarrow \beta_2 \alpha_2^{-1} \beta_1 \alpha_2 \beta_2^{-1} \beta_1^{-1} \alpha_2 \beta_2 \alpha_2^{-1} \beta_1^{-1} \alpha_2 \beta_2^{-1} \\
 & \beta_2 \rightarrow \beta_2 \beta_2 \alpha_2^{-1} \beta_1^{-1} \alpha_2 \beta_2^{-1} \\
 \\
 S_6: \quad & \gamma_1 \rightarrow \gamma_1 \gamma_2 \gamma_1^{-1} \\
 & \gamma_2 \rightarrow \gamma_1 \\
 \\
 S_7: \quad & \gamma_1 \rightarrow \gamma_2 \\
 & \gamma_2 \rightarrow \gamma_3 \\
 & \text{---} \\
 & \gamma_{s-1} \rightarrow \gamma_s \\
 & \gamma_s \rightarrow \gamma_s^{-1} \dots \gamma_2^{-1} \gamma_1 \gamma_2 \dots \gamma_s \\
 \\
 S_8: \quad & \alpha_q \rightarrow \gamma_1^{-1} \alpha_q \gamma_1 \\
 & \beta_q \rightarrow \gamma_1^{-1} \alpha_q^{-1} \gamma_1 \alpha_q \beta_q \alpha_q \alpha_q \gamma_1 \\
 & \gamma_1 \rightarrow \gamma_1^{-1} \alpha_q^{-1} \gamma_1 \alpha_q \gamma_1.
 \end{aligned}$$

Man stellt zunächst mit geringer Mühe fest, dass  $S_1, \dots, S_8$  Automorphismen der durch die Elemente (4.1) erzeugten *freien* Gruppe sind, indem die ursprünglichen Erzeugenden sich durch die ihnen bei den Substitutionen  $S$  entsprechenden ausdrücken lassen, ohne dass man von der Relation (4.2) Gebrauch macht. Ferner geht das *Relationswort* auf der linken Seite von (4.2) bei allen  $S$  in sich über, wie man durch unmittelbare Einsetzung feststellt.

Also sind die  $S$  auch Automorphismen von  $T$ . Dass ihnen dann auch topologische Abbildungen von  $M$  auf sich mit Erhaltung des Punktes  $O$  und der Orientierung entsprechen,

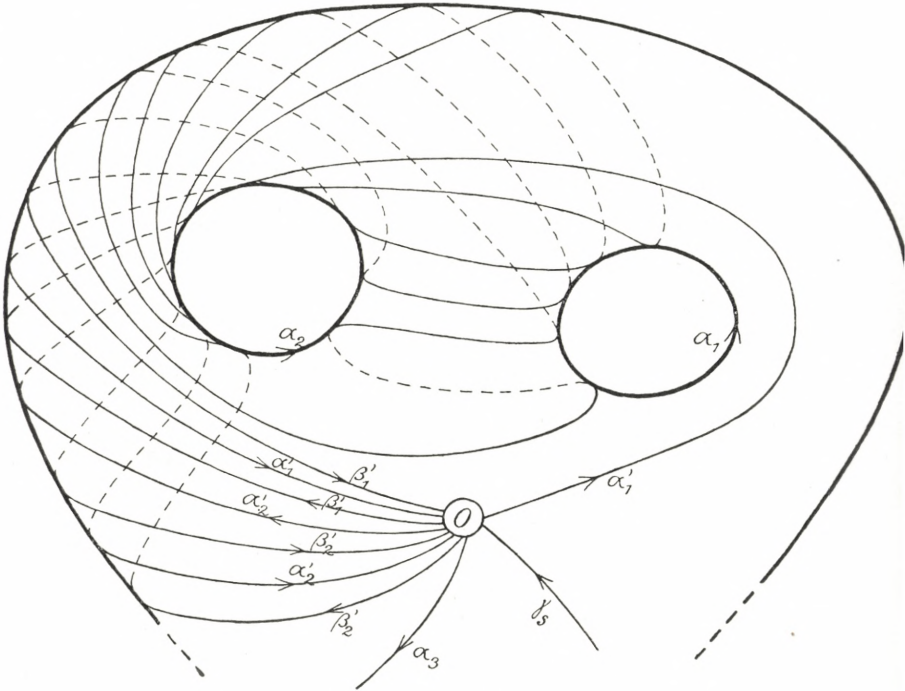


Fig. 2.

bestätigt man leicht durch Zeichnung des entsprechenden Systems von Bildkurven. Es wird genügen, dies in den zwei verwickeltesten Fällen, nämlich für  $S_5$  bzw.  $S_8$ , durch die Figuren 2 bzw. 3 auszuführen. Dabei ist in Fig. 2 nur derjenige Teil der Fläche angedeutet, der zu den ersten beiden Rückkehrschnittpaaren gehört, und in Fig. 3 derjenige Teil, der zu dem  $q$ -ten Rückkehrschnittpaar und der ersten Randkurve gehört. Die neuen Fundamentalkurven sind dabei durch einen Akzent bezeichnet.

Für  $q = 2$  ist  $S_3 = S_4$ . Für  $s = 2$  ist  $S_6 = S_7$ . Für  $q = 1$  fallen  $S_3, S_4$  und  $S_5$  fort. Für  $s = 1$  fallen  $S_6$  und  $S_7$ , und für  $s = 0$  überdies  $S_8$  fort. Für  $q = 0$  fallen alle  $S$  fort, die Symbole  $\alpha$  und  $\beta$  enthalten.

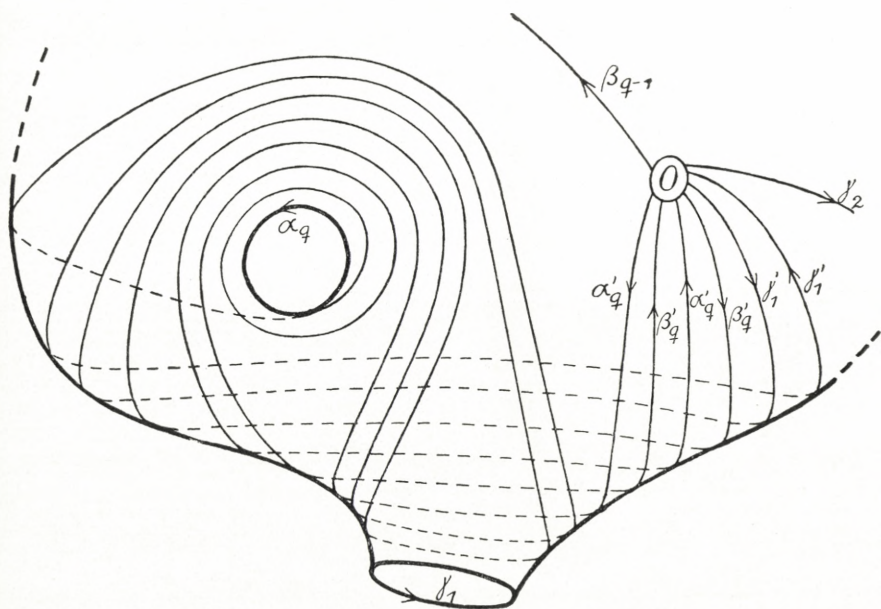


Fig. 3.

Die  $\mu$ -Werte der Erzeugenden  $\gamma_1, \dots, \gamma_s$ , sind durch die Valenzen der Ränder gegeben; siehe (4.5). Für diese besteht die Wirkung der  $S$  nur darin, dass  $S_6$  und  $S_7$  eine beliebige Änderung derjenigen Reihenfolge herzustellen gestatten, in der die Ränder für die Darstellung von  $T$  berücksichtigt werden. Für den speziellen Fall  $q = 0$  ist dies somit überhaupt die einzige Wirkung der  $S$ .

In dem Rest dieses Paragraphen handelt es sich somit um den Fall  $q > 0$ . Wir stellen die  $\mu$ -Werte der  $\alpha_i$  und  $\beta_i$  in einer Matrix  $A$  mit  $q$  Zeilen und zwei Spalten zusammen:

$$(5.1) \quad A = \begin{pmatrix} \mu(\alpha_1) & \mu(\beta_1) \\ \mu(\alpha_2) & \mu(\beta_2) \\ \dots & \dots \\ \mu(\alpha_q) & \mu(\beta_q) \end{pmatrix}$$

Es soll nun gezeigt werden, dass man durch Ausübung der  $S$  erreichen kann, dass an einer beliebig gewählten Stelle von  $A$  die Restklasse 1 mod  $n$  und an den übrigen  $2q-1$  Stellen die Restklasse 0 mod  $n$  zu stehen kommt.

Durch  $S_1$  wird  $\mu(\alpha_1)$  durch  $\mu(\alpha_1) + \mu(\beta_1)$ , und durch  $S_2$  wird  $\mu(\beta_1)$  durch  $\mu(\beta_1) + \mu(\alpha_1)$  ersetzt, während sich im übrigen in  $A$  nichts ändert. Durch fortgesetzte Anwendung von  $S_1$  und  $S_2$  kann man daher erreichen, dass  $\mu(\beta_1)$  durch die Restklasse 0 und  $\mu(\alpha_1)$  durch die Restklasse der Zahl  $(\mu(\alpha_1), \mu(\beta_1), n)$  ersetzt wird. Ist  $q > 1$ , so kann man durch  $S_3$  und  $S_4$  eine beliebige Permutation der  $q$  Zeilen von  $A$  herstellen, ohne im übrigen an den Elementen der Matrix etwas zu ändern. Also kann man durch  $S_1, \dots, S_4$  erzielen, dass die zweite Spalte von  $A$  aus Restklassen 0 besteht. Nachdem dies eingetreten ist, zieht man  $S_5$  heran: Die Wirkung von  $S_5$  besteht dann nur darin,  $\mu(\alpha_1)$  durch  $\mu(\alpha_1) + \mu(\alpha_2)$  zu ersetzen, ohne  $\mu(\alpha_2), \dots, \mu(\alpha_q)$  zu ändern und ohne etwas daran zu ändern, dass  $\mu(\beta_1), \dots, \mu(\beta_q)$  die Restklasse 0 sind. Durch fortgesetzte Anwendung von  $S_5$  zusammen mit den durch  $S_3$  und  $S_4$  bewirkten Zeilenpermutationen kann man daher noch erreichen, dass  $q-1$  der  $\mu(\alpha_i)$ , also  $2q-1$  der Elemente von  $A$  die Restklasse 0 sind. Als das letzte Element kann man den grössten gemeinsamen Teiler aller ursprünglichen Elemente von  $A$ , d. h. die durch die Zahl

$$(5.2) \quad (\mu(\alpha_1), \dots, \mu(\alpha_q), \mu(\beta_1), \dots, \mu(\beta_q), n) = z$$

bestimmte Restklasse erzielen.

Ist  $z = 1$ , so ist unser Ziel erreicht. Ist  $z = n$ , so ist  $\mathcal{A}$  die Nullmatrix und ist daher von vorne herein die Nullmatrix gewesen, sodass dar kein  $S$  angewendet ist. Ist  $z > 1$  und ein echter Teiler von  $n$ , so kann man noch die Restklasse  $z$  durch  $yz$  ersetzen, wobei  $y$  eine beliebig vorgeschriebene, zu  $\frac{n}{z}$  teilerfremde Zahl ist. Denn ist  $\{z, 0\}$  die erste Zeile von  $\mathcal{A}$ , so wird sie durch  $S_2^y$  in  $\{z, yz\}$  und weiter durch  $S_1^x$  in  $\{z(1+xy), yz\}$  verwandelt, und das ist gleich  $\{0, yz\}$ , wenn man  $x$  so wählt, dass  $xy \equiv -1 \pmod{\frac{n}{z}}$  ist. Diese Bemerkung zeigt zugleich (für  $y=1$ ), dass man die von 0 verschiedene Restklasse nach Belieben in der ersten oder zweiten Spalte von  $\mathcal{A}$ , also mittels  $S_1, \dots, S_4$  an beliebiger Stelle von  $\mathcal{A}$  anbringen kann.

Ist nun  $z > 1$ , so muss  $s > 0$  sein wegen (4.4). Dann sei

$$u = (\mu(\gamma_1), \dots, \mu(\gamma_s), n).$$

Wegen (4.4) gibt es Zahlen  $x$  und  $y$  so, dass

$$xu + yz = 1$$

ist. Für  $z = n$  muss  $u = 1$  sein, und man kann  $y = 0, x = 1$  wählen. Für  $1 < z < n$  ist  $y$  nur als Restklasse mod  $u$  bestimmt, und wegen  $(y, u) = 1$  kann man  $y$  so wählen, dass  $\left(y, \frac{n}{z}\right) = 1$  ist. Man kann also in allen Fällen eine Zahl  $y$ , die für  $z = n$  gleich Null und für  $1 < z < n$  zu  $\frac{n}{z}$  teilerfremd ist, und eine Linearkombination  $L$  von  $\mu(\gamma_1), \dots, \mu(\gamma_s)$  so wählen, dass

$$(5.3) \quad yz + L \equiv 1 \pmod{n}$$

ist.

Nun richte man für  $z > 1$  die Matrix  $\mathcal{A}$  so ein, dass  $\mu(\beta_q) = yz$  und alle übrigen Elemente 0 sind, wobei  $y$

gemäss (5.3) gewählt ist. Dann hat  $S_8$  die Wirkung,  $\mu(\beta_q)$  durch  $\mu(\beta_q) + \mu(\gamma_1)$  zu ersetzen und im übrigen nichts in  $\mathcal{A}$  zu ändern. Durch  $S_6$  und  $S_7$  kann man ein beliebiges  $\gamma_\nu$  an die Stelle von  $\gamma_1$  bringen und daher durch fortgesetzte Anwendung von  $S_6$ ,  $S_7$  und  $S_8$  eine beliebige Linearkombination von  $\mu(\gamma_1), \dots, \mu(\gamma_s)$  zu  $\mu(\beta_q)$  hinzufügen. Nach (5.3) kann man daher an der Stelle  $\mu(\beta_q)$  die Restklasse  $1 \bmod n$  erzielen, und diese lässt sich dann, wie schon genannt, an jede Stelle von  $\mathcal{A}$  verlegen. — Damit ist der Beweis für den obigen Satz erbracht.

*Im Falle  $q > 0$  lassen sich also die kanonischen Kurven  $\alpha_1, \dots, \beta_q$  der Fig. 1 so wählen, dass eine (beliebig gewählte) unter ihnen die Multiplizität  $n$  hat und die übrigen  $2q - 1$  nicht multipl sind, und dass überdies einem Umlauf um die multiple die Monodromiesubstitution  $\tau^1$  entspricht.*

## 6. Automorphismus der Fundamentalgruppe von $\varphi$ .

Die universelle Überlagerungsfläche von  $M$  lässt sich bekanntlich topologisch auf einen Bereich  $\Phi$  der euklidischen oder hyperbolischen Ebene so abbilden, dass die Fundamentalgruppe  $T$  von  $M$  dabei als diskontinuierliche Bewegungsgruppe von  $\Phi$  in sich dargestellt wird.

Der euklidische Fall liegt vor, wenn  $T$  abelsch ist, und das tritt nur in zwei Fällen ein:

- a) Wenn  $q = 0$  und  $s = 2$ , wobei beide Ränder multipl sind,
- b) Wenn  $q = 1$  und  $s = 0$ .

Denn wenn kein Rand multipl ist, so ist  $q > 0$  wegen  $n > 1$  und (3.3), und  $T$  ist dann nur für  $q = 1$ ,  $s = 0$  abelsch; und wenn multiple Ränder vorkommen, so ist ihre Anzahl mindestens 2 wegen (4.6), und andererseits ist  $T$  im Falle  $s \geq 2$  nur für  $s = 2$ ,  $q = 0$  abelsch. — Im Falle

a) ist  $M$  und damit auch  $\varphi$  einem Kreisring homöomorph, wobei  $n$  beliebige Werte haben kann. Je nachdem von den beiden Randkurven keine, eine oder beide durch Ausschneiden eines Verzweigungspunktes entstanden sind, ist die betrachtete Transformation der ursprünglichen Fläche  $\varphi$  topologisch äquivalent mit der  $n$ -periodischen Drehung  $a_1$ ) eines Kreisringes um seinen Mittelpunkt,  $a_2$ ) einer Kreisscheibe um ihren Mittelpunkt,  $a_3$ ) einer Kugel um einen Durchmesser; dabei hat man in jedem der drei Fälle die Drehzahl als einzige kennzeichnende Invariante. Das sind die durch BROUWER und v. KERÉKJÁRTÓ zuerst erledigten Fälle [2, 5]. — Im Falle b) kann man nach 5 ein kanonisches Schnittpaar  $\alpha_1, \beta_1$  so wählen, dass zu ihnen die Substitutionen  $\tau^1$  und  $\tau^0$  gehören. Wegen (3.3) ist  $p = 1$  und  $n$  beliebig. Die betrachtete Transformation von  $\varphi$  ist also topologisch äquivalent mit der Drehung eines metrisch regulären Torus in sich  $\frac{1}{n}$  einer vollen Drehung. Schon in einem so einfachen Fall ergibt also die Betrachtung von 5 das anschaulich zuerst wohl überraschende Resultat, dass z. B. eine  $\frac{2}{5}$  Drehung eines metrisch regulären Torus in sich mit der  $\frac{1}{5}$  Drehung topologisch äquivalent ist. Dadurch vereinfacht sich die von BROUWER in [3, § 1] gegebene Aufzählung, indem die dort mit  $a$  bezeichnete, zu  $n$  teilerfremde Zahl immer durch 1 ersetzt werden kann.

Da man somit im euklidischen Fall die Struktur der periodischen Transformation und das zugehörige Äquivalenzproblem völlig übersieht und eine Aufzählung aller möglichen topologischen Typen angeben kann, können wir uns im Folgenden auf den hyperbolischen Fall beschränken, in welchem  $T$  nicht kommutativ ist.

Die hyperbolische Ebene sei in bekannter Weise durch

das Innere des Einheitskreises  $E$  der Ebene einer komplexen Variablen  $x$  so dargestellt, dass ihre Bewegungen in sich durch linear-gebrochene Substitutionen in  $x$  beschrieben werden, die  $E$  in sich überführen. Der Bereich  $\Phi$  ist mit dem Inneren von  $E$  oder mit einem Teilbereich davon identisch, je nachdem  $M$  geschlossen oder berandet ist. In letzterem Fall kann man annehmen, dass die Ränder von  $M$  geschlossene geodätische Linien im Sinne der von der  $x$ -Ebene auf  $M$  übertragenen hyperbolischen Metrik sind. Dann wird der Bereich  $\Phi$  von »Geraden«, d. h. zu  $E$  orthogonalen Kreisbögen berandet, die das Innere von  $E$  durchziehen. Der kürzeren Ausdrucksweise zuliebe sprechen wir von  $\Phi$  direkt als der universellen Überlagerungsfläche von  $M$ , anstatt  $\Phi$  als topologisches Abbild der abstrakt definierten universellen Überlagerungsfläche zu bezeichnen.  $T$  ist die Gruppe der Decktransformationen von  $\Phi$  über  $M$ , was wir kurz durch die Schreibweise

$$(6.1) \quad M = \Phi \bmod T$$

zum Ausdruck bringen.

Nun lässt sich  $\Phi$  auch als universelle Überlagerungsfläche von  $\varphi$  auffassen und zwar auf genau  $n$  verschiedene Weisen, von denen wir eine folgendermassen festlegen: Sei  $x = \omega$  ein Punkt von  $\Phi$  und  $o$  ein Punkt von  $\varphi$ , beide mit dem Punkt  $O$  der Fig. 1 als Spurpunkt auf  $M$ . Einer geschlossenen Kurve von  $M$  mit dem Anfangs- und Endpunkt  $O$ , die einem Element  $z$  der Fundamentalgruppe  $T$  von  $M$  entspricht, entspricht in  $\Phi$  ein Weg von  $\omega$  nach dem Punkte  $z\omega$  und auf  $\varphi$  ein Weg von  $o$  nach dem Punkte  $\tau^{\mu(z)}o$ . Der letztere ist dann und nur dann geschlossen, wenn  $\mu(z) \equiv 0 \pmod{n}$  ist, und ihm entspricht in diesem Fall ein Element der Fundamentalgruppe  $F$  von  $\varphi$  bei



Zugrundelegung von  $o$  als Anfangspunkt. Dadurch ergibt sich  $F$  als diejenige Untergruppe von  $T$ , für deren Elemente  $\mu \equiv 0 \pmod{n}$  ist. Wegen des kommutativen Charakters der Operation  $\mu$  ist  $F$  Normalteiler von  $T$ . Die Faktorgruppe  $\frac{T}{F}$  ist die zyklische Gruppe der Ordnung  $n$ . Die Festlegung von  $\Phi$  als universelle Überlagerungsfläche von  $\varphi$  geschieht nun eindeutig durch die Vorschrift, dass zwei Punkte von  $\Phi$  dann und nur dann denselben Punkt von  $\varphi$  überdecken, wenn sie sich bei einem Element der Untergruppe  $F$  von  $T$  entsprechen, — sie überdecken dann auch denselben Punkt von  $M$ , — und dass überdies  $\omega$  den Punkt  $o$  überdecken soll. Nach dieser Festsetzung schreiben wir entsprechend (6.1)

$$(6.2) \quad \varphi = \Phi \text{ mod } F.$$

Der durch ein beliebiges Element  $z$  von  $T$  bewirkten Bewegung von  $\Phi$  in sich entspricht nun die Abbildung  $\tau^{\mu(z)}$  von  $\varphi$  auf sich. Durch Transformation mit  $z$  wird der Normalteiler  $F$  von  $T$  isomorph auf sich abgebildet. Ist  $f$  ein Element von  $F$ , also eine Decktransformation von  $\Phi$  über  $\varphi$ , so entspricht auch  $fz$  der Abbildung  $\tau^{\mu(z)}$  von  $\varphi$ , und der zugehörige Automorphismus von  $F$  geht aus dem vorigen durch Hinzufügung eines inneren Automorphismus von  $F$ , nämlich der Transformation aller Elemente von  $F$  mit  $f$ , hervor. Zwei solche Automorphismen nennen wir *verwandt* und eine Gesamtheit verwandter eine *Automorphismenfamilie* von  $F$ . Den  $n$  Potenzen von  $\tau$  entsprechen also  $n$  Automorphismenfamilien von  $F$ .

Nun sei  $t$  ein fest gewähltes Element von  $T$ , für welches  $\mu(t) \equiv 1 \pmod{n}$  ist. Da man für  $T$  die Restklassenzerlegung

$$(6.3) \quad F, Ft, Ft^2, \dots, Ft^{n-1}$$

nach  $F$  hat, lässt sich jedes Element von  $T$  in der Form

$$(6.4) \quad ft^x, \quad (f \text{ in } F, 0 \leq x \leq n-1)$$

schreiben. Das Element  $t^n$  liegt in  $F$  und sei mit

$$(6.5) \quad t^n = a$$

bezeichnet. Durch Transformation von  $F$  mit  $t$  erhält man einen zu der Abbildung  $\tau$  gehörigen Automorphismus von  $F$ . Iteriert man diesen  $n$  Male, so erhält man denjenigen inneren Automorphismus von  $F$ , bei dem jedes Element von  $F$  mit dem ebenfalls zu  $F$  gehörigen Element  $a$  transformiert wird, und dieser gehört zu der identischen Abbildung  $\tau^n$  von  $\varphi$ .

Nun ist  $F$  als Fundamentalgruppe einer Fläche vom Geschlecht  $p$  mit  $r \geq 0$  Randkurven durch endlich viele Elemente erzeugbar, etwa durch die Elemente

$$(6.6) \quad f_1, f_2, \dots, f_h$$

mit gewissen definierenden Relationen. Man erhält dann eine neue Darstellung von  $T$ , indem man zu den Erzeugenden (6.6) mit den für sie geltenden Relationen noch die Erzeugende  $t$  mit der Relation (6.5) hinzufügt und als weitere Relationen hinschreibt, in welche Elemente die Erzeugenden (6.6) durch Transformation mit  $t$  übergehen.

Für diese Beschreibung der Abbildung  $\tau$  von  $\varphi$  durch den in  $F$  induzierten Automorphismus soll nun in den folgenden Paragraphen eine *Normaldarstellung* gesucht werden. Dabei ist zwischen den Fällen  $q > 0$  und  $q = 0$  zu unterscheiden. Wir erhalten diese Normaldarstellung zunächst für unverzweigte Überlagerung von  $M$  durch  $\varphi$  und können daraus nachträglich ohne Mühe auch für den allgemeinen Fall der verzweigten Überlagerung eine Normal-

darstellung gewinnen. Es wird sich dabei zeigen, dass es zweckmässig ist, die Erzeugendenzahl  $h$  von  $F$  in (6.6) nicht immer so klein wie möglich zu halten.

**7. Normaldarstellung im Falle  $q > 0$ .** Wenn das Geschlecht  $q$  der Modulfläche  $M$  positiv ist, so sei das Erzeugendensystem (4.1) der Fundamentalgruppe  $T$  von  $M$  nach 5 so gewählt, dass  $\mu(\alpha_1) \equiv 1 \pmod n$  ist und  $\beta_1, \alpha_2, \dots, \beta_q$  nicht multipl sind. Dann kann  $\alpha_1$  die Rolle des in 6 mit  $t$  bezeichneten Elements von  $T$  übernehmen und soll dementsprechend mit  $t$  bezeichnet werden. Dabei ist nach der Bezeichnung (6.5)

$$(7.1) \quad t^n = a.$$

Die Elemente  $\beta_1, \alpha_2, \dots, \beta_q$  gehören zu  $F$  und sollen hier zunächst mit  $b_1, \alpha_2, \dots, b_q$  bezeichnet werden. Die  $\mu$ -Werte der  $s \geq 0$  Randkurven sind durch (4.5) gegeben. Dabei ist  $\sigma_\nu$  nur als Restklasse nach dem Modul  $\lambda_\nu = \frac{n}{m_\nu}$  bestimmt. Wir wollen hier für diese Restklassen ein System von  $\nu$  repräsentierenden Zahlen, die wieder mit  $\sigma_\nu$  bezeichnet werden, so auswählen, dass man an Stelle von (4.6)

$$(7.2) \quad \sigma_1 m_1 + \sigma_2 m_2 + \dots + \sigma_s m_s = 0$$

hat. Statt der Erzeugenden  $\gamma_\nu$  führen wir neue Erzeugende

$$(7.3) \quad c_\nu = \gamma_\nu t^{-\sigma_\nu m_\nu} \quad \nu = 1, 2, \dots, s$$

ein, die zu  $F$  gehören, da  $\mu(c_\nu) \equiv 0 \pmod n$  wird. Dann ist

$$(7.4) \quad \gamma_\nu = c_\nu t^{\sigma_\nu m_\nu} \quad \nu = 1, 2, \dots, s.$$

Somit hat man für  $T$  das Erzeugendensystem

$$(7.5) \quad t, b_1, \alpha_2, b_2, \dots, \alpha_q, b_q, c_1, c_2, \dots, c_s$$

und als einzige definierende Relation entsprechend (4.2)

$$(7.6) \quad \begin{aligned} &tb_1 t^{-1} b_1^{-1} a_2 b_2 a_2^{-1} b_2^{-1} \dots \\ &\dots a_q b_q a_q^{-1} b_q^{-1} c_1 t^{\sigma_1 m_1} c_2 t^{\sigma_2 m_2} \dots c_s t^{\sigma_s m_s} = 1. \end{aligned}$$

Im Falle  $s > 0$  könnte man mittels (7.6) eine Erzeugende, etwa  $c_s$ , eliminieren und dadurch  $T$  durch ein System freier Erzeugender darstellen; aber teils würde dadurch die Gleichberechtigung der Randkurven nicht mehr in der formalen Darstellung zum Ausdruck kommen, teils würde der Fall  $s = 0$  eine gesonderte Behandlung erfordern. Wir bleiben daher bei der Darstellung von  $T$  durch (7.5) und (7.6).

Nun suchen wir eine Darstellung des Normalteilers  $F$  von  $T$ . Unter den Erzeugenden (7.5) gehört nur  $t$  nicht zu  $F$ . Jedes Element von  $T$  lässt sich schreiben als ein Produkt von Transformaten der Erzeugenden  $b_1, \dots, c_s$  von (7.5) mit Potenzen von  $t$ , multipliziert mit einer Potenz  $t^x$ , und es gehört dann und nur dann zu  $F$ , wenn der Exponent  $x$  durch  $n$  teilbar ist.

Für diese Transformaten führen wir folgende Bezeichnungen ein:

$$(7.7) \quad \begin{aligned} a_{i,j} &= t^j a_i t^{-j} & i &= 2, 3, \dots, q \\ b_{i,j} &= t^j b_i t^{-j} & i &= 1, 2, 3, \dots, q \\ c_{\nu,j} &= t^j c_\nu t^{-j} & \nu &= 1, 2, 3, \dots, s. \end{aligned}$$

Dabei nimmt  $j$  alle Werte an, und  $a_{i,0}, b_{i,0}, c_{\nu,0}$  bedeuten dasselbe wie  $a_i, b_i, c_\nu$ . Ein in den Erzeugenden (7.5) geschriebener Ausdruck  $e$  ist dann und nur dann gleich 1 in  $T$ , wenn er sich identisch in diesen Erzeugenden in der Form eines Produktes

$$(7.8) \quad e \equiv \Pi(\psi R^{\pm 1} \psi^{-1})$$

schreiben lässt, wobei  $R$  das Relationswort in (7.6) und die Transformatoren  $\psi$  Ausdrücke in den Erzeugenden (7.5) sind.  $e$  hat also wegen (7.6) und (7.2) die Exponentensumme 0 in  $t$ . Da sich die  $\psi$  in der Form

$$\psi = \Phi(a_{i,j}, b_{i,j}, c_{\nu,j}) \cdot t^x$$

schreiben lassen, lässt sich (7.8) in die Form

$$(7.9) \quad e \equiv \Pi(\Phi R_x^{\pm 1} \Phi^{-1})$$

bringen, wobei

$$(7.10) \quad R_x = t^x R t^{-x}$$

gesetzt ist. Dabei sind die  $R_x$  ebenfalls Ausdrücke in den Elementen (7.7), weil  $t$  in  $R$  wegen (7.2) die Exponentensumme 0 hat.

Man hat also für  $F$  die Erzeugenden

$$(7.11) \quad t^n, a_{i,j}, b_{i,j}, c_{\nu,j}.$$

Ein Ausdruck in diesen, der gleich 1 sein soll, muss in  $t^n$  die Exponentensumme 0 haben, da er als Element von  $T$  in  $t$  die Exponentensumme 0 haben muss. Er lässt sich daher zunächst mittels der Relationen

$$(7.12) \quad \begin{aligned} t^n a_{i,j} t^{-n} &= a_{i,j+n} \\ t^n b_{i,j} t^{-n} &= b_{i,j+n} \\ t^n c_{\nu,j} t^{-n} &= c_{\nu,j+n} \end{aligned}$$

durch die  $a_{i,j}$ ,  $b_{i,j}$ ,  $c_{\nu,j}$  allein darstellen und dann iden-

tisch in diesen in die Form (7.9) bringen. Also sind die Relationen (7.12) zusammen mit den Relationen

$$(7.13) \quad R_x(a_{i,j}, b_{i,j}, c_{\nu,j}) = 1$$

vollständig für  $F$ . Benutzt man die Abkürzung

$$(7.14) \quad k_{i,j} = a_{i,j} b_{i,j} a_{i,j}^{-1} b_{i,j}^{-1} \quad i = 2, 3, \dots, q,$$

so erhält man explizit die Relationen (7.13) für  $0 \leq x \leq n-1$  aus (7.6) und unter Benutzung von (7.2)

$$(7.15) \quad \begin{aligned} b_{1,1}^{-1} &= b_{1,0}^{-1} k_{2,0} \dots k_{q,0} c_{1,0} c_{2,\sigma_1 m_1} c_{3,\sigma_1 m_1 + \sigma_2 m_2} \dots c_{s,-\sigma_s m_s} \\ b_{1,2}^{-1} &= b_{1,1}^{-1} k_{2,1} \dots k_{q,1} c_{1,1} c_{2,\sigma_1 m_1 + 1} \dots c_{s,-\sigma_s m_s + 1} \\ &--- \\ b_{1,n-1}^{-1} &= b_{1,n-2}^{-1} k_{2,n-2} \dots k_{q,n-2} c_{1,n-2} \dots c_{s,-\sigma_s m_s + n-2} \\ b_{1,n}^{-1} &= b_{1,n-1}^{-1} k_{2,n-1} \dots k_{q,n-1} c_{1,n-1} \dots c_{s,-\sigma_s m_s + n-1} \end{aligned}$$

Diese Darstellung von  $F$  lässt sich nun durch Elimination überflüssiger Erzeugender folgendermassen vereinfachen: Mittels (7.12) kann man alle diejenigen Erzeugenden (7.7) eliminieren, für welche  $j$  ausserhalb des Wertbereichs  $0 \leq j \leq n-1$  liegt, und damit fallen die Relationen (7.12) fort. Weiter kann man sich dann auf die Auswahl (7.15) der Relationen (7.13) beschränken, da alle übrigen aus den  $n$  Relationen (7.15) durch Transformation mit Potenzen des in  $F$  gelegenen Elementes  $t^n$  hervorgehen. Ferner kann man mittels der ersten  $n-1$  Relationen (7.15) successive  $b_{1,1}, b_{1,2}, \dots, b_{1,n-1}$  eliminieren und in die letzte Relation (7.15) einsetzen. Dabei ist in dieser  $b_{1,n}$  durch  $t^n b_{1,0} t^{-n}$

zu ersetzen. Wir schreiben nun  $a$  statt  $l^n$  nach (7.1) und  $b$  statt  $b_{1,0}$  und haben dadurch folgende Darstellung von  $F$  gewonnen:

*Erzeugende von F:*

$$\begin{array}{l}
 a \\
 b \\
 (7.16) \quad \left. \begin{array}{l} a_{i,j} \\ b_{i,j} \\ c_{\nu,j} \end{array} \right\} \begin{array}{l} i = 2, 3, \dots, q \\ \\ \nu = 1, 2, \dots, s \end{array} \left\} \begin{array}{l} j = 0, 1, 2, \dots, n-1. \\ \\ \end{array}
 \end{array}$$

*Definierende Relation von F:*

$$\begin{aligned}
 (7.17) \quad &aba^{-1}b^{-1}k_{2,0} \dots k_{q,0}c_{1,0}c_{2,\sigma_1 m_1}c_{3,\sigma_1 m_1 + \sigma_2 m_2} \dots c_{s,-\sigma_s m_s}k_{2,1} \dots \\
 &\dots k_{q,1}c_{1,1} \dots c_{s,-\sigma_s m_s + 1}k_{2,2} \dots \dots \dots \\
 &\dots k_{2,n-1} \dots k_{q,n-1}c_{1,n-1} \dots c_{s,-\sigma_s m_s + n-1} = 1.
 \end{aligned}$$

In dieser Schreibweise der Relation (7.17) wird zwar der zweite Index der Symbole  $c$  nicht immer dem Wertebereich von  $0$  bis  $n-1$  angehören, aber ein solches  $c$  ist dann nach der Formel

$$(7.18) \quad c_{\nu,j+xn} = a^x c_{\nu,j} a^{-x} \quad 0 \leqq j \leqq n-1$$

zu lesen. Diese Bemerkung ist auch im Folgenden für die Symbole  $c$  zu beachten.

$F$  ist Fundamentalgruppe der Fläche  $\varphi$  vom Geschlecht  $p$  mit  $r$  Randkurven. Die Anzahl der Erzeugenden (7.16) von  $F$  ist  $n[2(q-1) + s] + 2$ , und das ist wegen (3.2) gleich  $2p + r$ . Die  $ns$  Symbole  $c_{\nu,j}$  kommen in (7.17) genau je einmal vor. Ist  $s > 0$  (also auch  $r > 0$  und  $\varphi$  berandet), so kann man mittels (7.17) genau eines der  $c_{\nu,j}$  eliminieren und dadurch  $F$  als freie Gruppe mit  $2p + r - 1$  freien Er-

zeugenden darstellen. Ist  $s = 0$  (also auch  $r = 0$  und  $q$  geschlossen), so fallen die Symbole  $c$  überall fort; dann hat man  $2p$  Erzeugende in (7.16) und als Relation (7.17) das Produkt der  $p$  Kommutatoren gleich 1, also die bekannte kanonische Form für  $F$ . — Aus diesen Bemerkungen ergibt sich die Vollständigkeit der Relation (7.17) für die Erzeugenden (7.16) auch unmittelbar aus der Bedeutung von  $F$ , ohne Bezugnahme auf die vorhergehende formale Rechnung.

Für den in dem Normalteiler  $F$  von  $T$  durch die Transformation mit  $t$  induzierten Automorphismus  $I$  erhält man die folgende Darstellung:

$$\begin{aligned}
 (7.19) \quad & ta\bar{t}^{-1} = a \\
 & tbt^{-1} = [b^{-1}k_{2,0} \dots k_{q,0}c_{1,0}c_{2,\sigma_1 m_1} \dots c_{s,-\sigma_s m_s}]^{-1} \\
 & \left. \begin{aligned} ta_{i,j}\bar{t}^{-1} &= a_{i,j+1} \\ tb_{i,j}\bar{t}^{-1} &= b_{i,j+1} \end{aligned} \right\} \begin{aligned} & i = 2, 3, \dots, q \\ & j = 0, 1, 2, \dots, n-2 \end{aligned} \\
 & tc_{\nu,j}\bar{t}^{-1} = c_{\nu,j+1} \quad \nu = 1, 2, \dots, s \\
 & \left. \begin{aligned} ta_{i,n-1}\bar{t}^{-1} &= aa_{i,0}a^{-1} \\ tb_{i,n-1}\bar{t}^{-1} &= ab_{i,0}a^{-1} \end{aligned} \right\} i = 2, 3, \dots, q \\
 & tc_{\nu,n-1}\bar{t}^{-1} = ac_{\nu,0}a^{-1} \quad \nu = 1, 2, \dots, s.
 \end{aligned}$$

Iteriert man diesen Automorphismus  $n$  Male, so sieht man, dass  $I^n$  derjenige *innere* Automorphismus von  $F$  ist, bei welchem jedes Element  $f$  von  $F$  in  $afa^{-1}$  übergeht. Für das Element  $b$  ist dabei die Relation (7.17) zu beachten.

Durch Transformation mit  $t$  geht das Relationswort (7.17) in eine transformierte seiner selbst, nämlich in den mit

$$[ak_{2,0} \dots k_{q,0}c_{1,0} \dots c_{s,-\sigma_s m_s} a^{-1}]^{-1}$$

transformierten Ausdruck über.



Damit ist die gewünschte Normaldarstellung des zu der Abbildung  $\tau$  von  $\varphi$  gehörigen Automorphismus von  $F$  für  $q > 0$  gefunden. Ist  $\varphi$  geschlossen, also  $s = 0$ , so sind die Symbole  $c_{\nu,j}$  einfach überall fortzulassen.

Stellt man die Gesamtgruppe  $T$  dadurch dar, dass man als Erzeugende die Elemente (7.16) und dazu noch  $t$  nimmt, so bilden die Relationen (7.1), (7.17) und (7.19) ein vollständiges System definierender Relationen für diese Erzeugendenwahl. Denn mittels (7.19) und (7.1) kann man jedes Potenzprodukt dieser Erzeugenden in die Form (6.4) bringen, und für  $F$  allein ist (7.17) vollständig.

Im Falle  $s > 0$  wollen wir noch unser Augenmerk auf die Elemente  $\gamma_\nu$ ,  $\nu = 1, 2, \dots, s$ , von  $T$  richten, von denen wir am Anfang dieses Paragraphen zu den neuen Erzeugenden  $c_\nu$  übergangen. Wegen (7.4) ist  $\gamma_\nu^{\lambda_\nu}$  die niedrigste Potenz von  $\gamma_\nu$ , die zu  $F$  gehört, und es ist

$$(7.20) \quad \gamma_\nu^{\lambda_\nu} = \left( c_\nu t^{\sigma_\nu m_\nu} \right)^{\lambda_\nu} = c_{\nu,0} c_{\nu,\sigma_\nu m_\nu} c_{\nu,2\sigma_\nu m_\nu} \dots c_{\nu,(\lambda_\nu-1)\sigma_\nu m_\nu} t^{\sigma_\nu m_\nu \lambda_\nu}.$$

Der letzte Faktor ist  $a^{\sigma_\nu}$  wegen (1.3) und (7.1). Das Element (7.20) werde mit  $g_{\nu,0}$  und seine Transformaten durch  $t^l$  für  $l = 1, 2, \dots, m_\nu - 1$  mit  $g_{\nu,1}, \dots, g_{\nu,m_\nu-1}$  bezeichnet:

$$(7.21) \quad g_{\nu,l} = t^l g_{\nu,0} t^{-l} = t^l \gamma_\nu^{\lambda_\nu} t^{-l} \quad \begin{array}{l} \nu = 1, 2, \dots, s \\ l = 0, 1, 2, \dots, m_\nu - 1. \end{array}$$

Zufolge (7.20) findet man für diese Elemente von  $F$  die Darstellungen:

$$(7.22) \quad \begin{array}{l} g_{\nu,0} = c_{\nu,0} c_{\nu,\sigma_\nu m_\nu} c_{\nu,2\sigma_\nu m_\nu} \dots c_{\nu,(\lambda_\nu-1)\sigma_\nu m_\nu} \cdot a^{\sigma_\nu} \\ g_{\nu,1} = c_{\nu,1} c_{\nu,\sigma_\nu m_\nu+1} c_{\nu,2\sigma_\nu m_\nu+1} \dots c_{\nu,(\lambda_\nu-1)\sigma_\nu m_\nu+1} \cdot a^{\sigma_\nu} \\ \text{---} \\ g_{\nu,m_\nu-1} = c_{\nu,m_\nu-1} c_{\nu,\sigma_\nu m_\nu+m_\nu-1} \dots c_{\nu,(\lambda_\nu-1)\sigma_\nu m_\nu+m_\nu-1} \cdot a^{\sigma_\nu}. \end{array}$$

Transformiert man dies letzte Element mit  $t$ , so erhält man

$$(7.23) \quad tg_{\nu, m_\nu - 1} t^{-1} = c_{\nu, m_\nu} c_{\nu, m_\nu + \sigma_\nu, m_\nu} c_{\nu, m_\nu + 2\sigma_\nu, m_\nu} \cdots \\ \cdots c_{\nu, m_\nu + (\lambda_\nu - 1)\sigma_\nu, m_\nu} \cdot a^{\sigma_\nu}.$$

Ist, entsprechend (2.5),

$$\delta_\nu \sigma_\nu = 1 + x_\nu \lambda_\nu, \quad 0 < \delta_\nu < \lambda_\nu,$$

so kommt in  $g_{\nu, 0}$  der Faktor

$$c_{\nu, \delta_\nu \sigma_\nu, m_\nu} = c_{\nu, m_\nu + x_\nu n} = a^{x_\nu} c_{\nu, m_\nu} a^{-x_\nu}$$

vor, und man erkennt, dass

$$a^{x_\nu} tg_{\nu, m_\nu - 1} t^{-1} a^{-x_\nu} \\ = [c_{\nu, 0} \cdots c_{\nu, (\delta_\nu - 1)\sigma_\nu, m_\nu}]^{-1} g_{\nu, 0} [c_{\nu, 0} \cdots c_{\nu, (\delta_\nu - 1)\sigma_\nu, m_\nu}]$$

ist. Also ist (7.23) die Transformierte von  $g_{\nu, 0}$  mit dem zu  $F$  gehörigen Element

$$(7.24) \quad \left[ c_{\nu, 0} c_{\nu, \sigma_\nu, m_\nu} \cdots c_{\nu, (\delta_\nu - 1)\sigma_\nu, m_\nu} a^{\frac{\delta_\nu \sigma_\nu - 1}{\lambda_\nu}} \right]^{-1} = d_\nu.$$

Man hat also:

$$(7.25) \quad \left. \begin{aligned} tg_{\nu, l} t^{-1} &= g_{\nu, l+1} & 0 \leq l \leq m_\nu - 2 \\ tg_{\nu, m_\nu - 1} t^{-1} &= d_\nu g_{\nu, 0} d_\nu^{-1} \end{aligned} \right\} \nu = 1, 2, \dots, s.$$

Diese Formeln drücken den folgenden Sachverhalt aus: Über der  $\nu$ -ten Randkurve von  $M$ , die in Fig. 1 durch die Kurve  $\gamma_\nu$  umlaufen wurde, liegen  $m_\nu$  Kurven von  $\varphi$ , die sich je erst nach  $\lambda_\nu$ -fachem Umlauf um die Spurkurve schliessen: Die diesen  $m_\nu$  Kurven von  $\varphi$  entsprechenden  $m_\nu$

Elementklassen von  $F$  werden durch die Elemente (7.22) repräsentiert. Bei der Abbildung  $\tau$  von  $\mathcal{g}$  vertauschen sich die  $m_\nu$  Kurven zyklisch: Bei Transformation mit  $t$  geht  $g_{\nu,l}$  nach (7.25) in  $g_{\nu,l+1}$  über für  $l = 0, \dots, m_\nu - 2$  und  $g_{\nu,m_\nu-1}$  in ein Element derselben Elementklasse von  $F$  wie  $g_{\nu,0}$ . Bei der Abbildung  $\tau^m$  werden die  $m_\nu$  Kurven Fixkurven mit der Drehzahl  $\frac{\delta_\nu}{\lambda_\nu}$ : Bei Transformation mit  $t^m$  bleiben die Elemente  $g_{\nu,l}$  innerhalb ihrer Elementklasse in  $F$ , und die Drehzahl geht aus der Gestalt des Transformators (siehe (7.24)) hervor.

Die beiden extremen Fälle  $m_\nu = 1$ ,  $\lambda_\nu = n$  und  $m_\nu = n$ ,  $\lambda_\nu = 1$  werden hierbei sinngemäss mit umfasst.

Die  $m_\nu$  Elemente (7.22) enthalten zusammen mit Rücksicht auf (7.18) jede der Erzeugenden  $c_{\nu,j}$ ,  $0 \leq j \leq n-1$ , genau einmal.

**8. Normaldarstellung im Falle  $q = 0$ .** Dass der Fall, in welchem die Modulfläche  $M$  das Geschlecht 0 hat, eine gesonderte Behandlung erfordert, hat folgenden Grund: Im Falle  $q > 0$  konnte man ein für die Zerlegung (6.3) brauchbares Element  $t$  von  $T$  ausserhalb der durch das Teilsystem  $\gamma_1, \dots, \gamma_s$  von (4.1) erzeugten Untergruppe von  $T$  wählen; im Falle  $q = 0$  macht diese Untergruppe ganz  $T$  aus.

Es sei  $q = 0$ . Dann gibt es, wie schon in 6 benutzt, multiple Randkurven, und zwar wegen (4.6) mindestens zwei. Nach 6 können wir dabei sogleich  $s \geq 3$  annehmen. Nach (4.1) und (4.2) hat man für  $T$  das Erzeugendensystem

$$(8.1) \quad \gamma_1, \gamma_2, \dots, \gamma_s$$

mit der definierenden Relation

$$(8.2) \quad \gamma_1 \gamma_2 \dots \gamma_s = 1.$$

$T$  ist somit eine freie Gruppe, und wir wählen die Elemente

$$(8.3) \quad \gamma_1, \gamma_2, \dots, \gamma_{s-1}$$

als ein System freier Erzeugender für  $T$ . Wegen (4.5), (4.6) und (4.7) ist dabei

$$(8.4) \quad (\sigma_1 m_1, \sigma_2 m_2, \dots, \sigma_{s-1} m_{s-1}, n) = 1.$$

Da es nach 5 auf die Reihenfolge der  $\gamma_\nu$  nicht ankommt und  $s \geq 3$  angenommen wurde, kann man die Bezeichnung so wählen, dass auch unter den (8.3) entsprechenden Kurven mindestens zwei multiple vorkommen, also solche, deren  $m$ -Wert eine in  $n$  aufgehende Zahl kleiner als  $n$  ist. Sei etwa  $m_1$  die kleinste der Zahlen  $m_1, m_2, \dots, m_{s-1}$ , und es sei etwa auch  $m_2 < n$ . Ersetzt man dann  $\gamma_2$  durch  $\gamma_2 \gamma_1^x$  und lässt die übrigen Erzeugenden (8.3) ungeändert, wodurch man zu einem neuen System freier Erzeugender für  $T$  übergeht, so kann man durch passende Wahl von  $x$  erreichen, dass  $\mu(\gamma_2 \gamma_1^x) \equiv 0 \pmod{n}$  ist, falls  $m_1$  in  $m_2$  aufgeht, und dass  $(\mu(\gamma_2 \gamma_1^x), n) < m_1$  ist, falls  $m_1$  nicht in  $m_2$  aufgeht. Durch fortgesetzte Anwendung solcher Erzeugendenwechsel kann man daher wegen (8.4) offenbar zu einem neuen System freier Erzeugender für  $T$  gelangen, bei dem der  $\mu$ -Wert einer Erzeugenden  $\equiv 1$  und derjenige der übrigen  $\equiv 0 \pmod{n}$  ist. Die erstere wählen wir als das  $t$  von (6.3) und bezeichnen das Erzeugendensystem mit

$$(8.5) \quad t, f_{1,0}, f_{2,0}, \dots, f_{s-2,0}.$$

Die Elemente  $f_{i,0}$  gehören dabei zu  $F$ . Um alle Elemente von  $T$  auf die Form (6.4) zu bringen, hat man noch die Transformierten

$$(8.6) \quad f_{i,j} = t^j f_{i,0} t^{-j} \quad i = 1, 2, \dots, s-2$$

für alle Werte von  $j$  einzuführen. Dann ist das  $f$  der Formel

(6.4) ein Produkt der  $f_{i,j}$ , und das Element (6.4) gehört dann und nur dann zu  $F$ , wenn  $n$  in  $x$  aufgeht. Setzen wir noch wie in (6.5)

$$(8.7) \quad t^n = a,$$

so hat man ebenso wie in (7.18)

$$(8.8) \quad f_{i,j+xn} = a^x f_{i,j} a^{-x} \quad 0 \leq j \leq n-1,$$

und kann daher nach Einführung des Elements  $a$  von  $F$  alle diejenigen Elemente (8.6), deren zweiter Index ausserhalb des Wertbereiches von 0 bis  $n-1$  liegt, mittels (8.8) eliminieren. Man gewinnt dadurch das folgende

*System freier Erzeugender von  $F$ :*

$$(8.9) \quad \begin{matrix} a \\ f_{i,j} \end{matrix} \quad i = 1, 2, \dots, s-2; \quad j = 0, 1, \dots, n-1.$$

Dass in der Tat keine Relationen zwischen den Erzeugenden (8.9) von  $F$  bestehen, kann man formal aus dem Verfahren bei ihrer Herstellung ähnlich wie in 7 nachweisen. Einfacher erschliesst man es aus der Bedeutung von  $F$ , wie dies auch in 7 angedeutet wurde:  $F$  ist als Fundamentalgruppe einer Fläche  $\varphi$  vom Geschlecht  $p$  mit  $r > 0$  Randkurven eine freie Gruppe mit  $2p + r - 1$  freien Erzeugenden. Die Anzahl der in (8.9) verwendeten Erzeugenden von  $F$  ist  $n(s-2) + 1$ , und wegen  $q = 0$  und (3.2) ist dies gleich der Minimalzahl  $2p + r - 1$ , woraus nach einer bekannten Eigenschaft freier Gruppen die Freiheit des Systems (8.9) folgt. Vgl. [15, § 5].

Für den in dem Normalteiler  $F$  von  $T$  durch die Transformation mit  $t$  induzierten Automorphismus  $I$  erhält man die Darstellung:

$$(8.10) \quad \left. \begin{aligned} taf^{-1} &= a \\ tf_{i,j}t^{-1} &= f_{i,j+1} \quad 0 \leq j \leq n-2 \\ tf_{i,n-1}t^{-1} &= af_{i,0}a^{-1} \end{aligned} \right\} i = 1, 2, \dots, s-2.$$

$I^n$  ist wieder der innere Automorphismus  $f \rightarrow af a^{-1}$  von  $F$ .

Damit ist die gewünschte Normaldarstellung des zu der Abbildung  $\tau$  von  $\varphi$  gehörigen Automorphismus von  $F$  für  $q = 0$  gefunden.

Stellt man  $T$  dadurch dar, dass man zu den Erzeugenden (8.9) noch  $t$  hinzunimmt, so bilden (8.7) und (8.10) ein vollständiges System definierender Relationen.

Wie in (7.21) sollen nun noch die Elemente

$$(8.11) \quad g_{\nu,l} = t^l g_{\nu,0} t^{-l} = t^l \gamma_{\nu}^l t^{-l} \quad \begin{array}{l} l = 0, 1, 2, \dots, m_{\nu}-1 \\ \nu = 1, 2, \dots, s \end{array}$$

aufgestellt werden. Die Elemente  $\gamma_{\nu}$  sind eindeutig bestimmte Ausdrücke in den Erzeugenden (8.5) von  $T$ :

$$(8.12) \quad \gamma_{\nu} = II_{\nu}(t, f_{i,0}).$$

Durch Einführung der Elemente (8.6) ergeben sich aus (8.12) eindeutige Darstellungen

$$(8.13) \quad \gamma_{\nu} = c_{\nu}(f_{i,j}) t^{\sigma_{\nu} m_{\nu}}.$$

Diese Gleichung entspricht (7.4). Hier bilden jedoch die Elemente  $c_{\nu}$  von  $F$  nicht einen Teil der Erzeugenden von  $F$  wie in 7, sondern werden nur als Abkürzung für gewisse Potenzprodukte der Erzeugenden (8.9) von  $F$  benutzt; man bedenke dabei, dass der Index  $j$  in (8.13) nicht beschränkt ist, dass man aber jederzeit mittels (8.8) auf das Erzeugendensystem (8.9) von  $F$  zurückgehen kann. Ferner beachte man, dass der Exponent  $\sigma_{\nu} m_{\nu}$  von  $t$  in (8.13) hier

eine bestimmte Zahl ist und nicht wie in 7 in einer bestimmten Restklasse mod  $n$  beliebig wählbar ist.

Nun ist nach (8.13) für beliebiges  $\varrho$

$$(8.14) \quad t^\varrho \gamma_\nu t^{-\varrho} = c_\nu (f_{i,j+\varrho}) t^{\sigma_\nu m_\nu}.$$

Um volle Übereinstimmung mit der Schreibweise von 7 zu erzielen, setzen wir

$$(8.15) \quad c_{\nu,\varrho}(f_{i,j}) = c_\nu(f_{i,j+\varrho}) \quad (c_{\nu,0} = c_\nu).$$

Dann genügen die  $c_{\nu,\varrho}$  der Gleichung (7.18). Nun werden die durch (8.11) definierten Elemente  $g_{\nu,i}$  von  $F$  durch (7.22) angegeben, indem sich an den Formeln (7.20–25) nichts ändert und auch die am Schluss von 7 beschriebene Beziehung zur Flächenabbildung  $\tau$  in Kraft bleibt.

Die Relation (8.2), die ja die einzige Relation zwischen den  $\gamma_\nu$  darstellt, hat Beziehungen einerseits zwischen den Zahlen  $\sigma_\nu m_\nu$ , andererseits zwischen den Elementen  $c_{\nu,\varrho}$  zur Folge. Nach (8.2) und (8.13) ist

$$\begin{aligned} 1 &= \gamma_1 \gamma_2 \cdots \gamma_s = c_{1,0} t^{\sigma_1 m_1} c_{2,0} t^{\sigma_2 m_2} \cdots c_{s,0} t^{\sigma_s m_s} \\ &= c_{1,0} c_{2,\sigma_1 m_1} c_{3,\sigma_1 m_1 + \sigma_2 m_2} \cdots c_{s,\sigma_1 m_1 + \cdots + \sigma_{s-1} m_{s-1}} t^{\sigma_1 m_1 + \cdots + \sigma_s m_s}. \end{aligned}$$

Diese Relation wird eine Identität, wenn man mittels (8.15) und (8.8) auf die freien Erzeugenden (8.9) von  $F$  zurückgeht, und da  $a$  dabei mit der Exponentensumme 0 auftreten muss, sieht man erstens, dass die Zahlen  $\sigma_\nu m_\nu$ , die sich, wie oben bemerkt, zwangsläufig ergeben haben, auch hier der Relation (7.2) genügen, und zweitens, dass sich zwischen den  $c_{\nu,\varrho}$  Relationen ergeben, die mit (7.15) übereinstimmen, wenn man dort alle  $b_{1,j}$  und alle  $k_{i,j}$  durch 1 ersetzt:

$$\begin{aligned}
 (8.16) \quad & c_{1,0} c_{2,\sigma_1 m_1} \cdots c_{s,-\sigma_s m_s} = 1 \\
 & c_{1,1} c_{2,\sigma_1 m_1 + 1} \cdots c_{s,-\sigma_s m_s + 1} = 1 \\
 & \text{---} \text{---} \text{---} \text{---} \text{---} \\
 & c_{1,n-1} c_{2,\sigma_1 m_1 + n-1} \cdots c_{s,-\sigma_s m_s + n-1} = 1.
 \end{aligned}$$

Eine weitere Beziehung zwischen den  $c_{\nu,q}$  erhält man daraus, dass das Element  $t$  von  $T$  sich durch die Elemente (8.3) ausdrücken lässt, so wie es am Anfang dieses Paragraphen durch fortgesetzten Erzeugendenwechsel geschah. Es sei

$$(8.17) \quad t = \Pi(\gamma_1, \gamma_2, \dots, \gamma_{s-1}),$$

also

$$(8.18) \quad \bar{t}^{-1} \Pi(c_1 t^{\sigma_1 m_1}, c_2 t^{\sigma_2 m_2}, \dots, c_{s-1} t^{\sigma_{s-1} m_{s-1}}) = 1.$$

Bei Übergang zu den Erzeugenden (8.5) sieht man, dass  $t$  in (8.18) die Exponentensumme 0 haben muss. Also kann (8.18) in der Form eines Produktes  $\Phi$  der  $c_{\nu,q}$

$$(8.19) \quad \Phi(c_{\nu,q}) = 1$$

geschrieben werden. Man erhält die  $n$  Relationen

$$(8.20) \quad \Phi(c_{\nu,q+z}) = 1 \quad 0 \leq z \leq n-1.$$

Hierin kommt der Wert  $s$  für  $q$  nicht vor. Weitere Werte von  $z$  ergeben nichts neues, wenn man den zweiten Index der  $c$  durch (7.18) reduziert.

Die  $2n$  Relationen (8.16) und (8.20) zwischen den  $c_{\nu,q}$ , oder zwischen  $a$  und den  $c_{\nu,q}$ , wenn man deren zweiten Index durch (7.18) reduziert, werden natürlich zu Identitäten, wenn man auf die Erzeugenden (8.9) von  $F$  zurückgeht. Sie stellen ein vollständiges System von Relationen zwischen den durch die Gleichungen (8.13) definierten Ele-



menten  $c_\nu$  von  $F$  und ihren Transformierten durch die Potenzen von  $t$  dar.

Daraus lässt sich eine neue Normalform des Automorphismus von  $F$  für  $q = 0$  gewinnen. Denn wegen (8.13) und (8.1) kann man  $T$  durch die Elemente

$$(8.21) \quad t, c_1, c_2, \dots, c_s,$$

also  $F$  durch die Elemente

$$(8.22) \quad a, c_{\nu, q} \quad \nu = 1, 2, \dots, s \quad q = 0, 1, \dots, n-1$$

darstellen. Dann bilden (8.16) und (8.20) ein vollständiges Relationssystem für die Erzeugenden (8.22), wobei man den zweiten Index der  $c$  zu reduzieren hat, und der Automorphismus von  $F$  ist durch

$$(8.23) \quad \left. \begin{aligned} taf^{-1} &= a \\ tc_{\nu, q}f^{-1} &= c_{\nu, q+1} \quad 0 \leq q \leq n-2 \\ tc_{\nu, n-1}f^{-1} &= ac_{\nu, 0}a^{-1} \end{aligned} \right\} \nu = 1, 2, \dots, s$$

gegeben. Dabei kann man, wenn man will, durch (8.16) z. B. die  $n$  Erzeugenden  $c_{s, q}$ ,  $0 \leq q \leq n-1$ , eliminieren.

Die Normalform (8.9–10) ist insofern einfacher, als sie auf einer relationsfreien Darstellung von  $F$  fusst. Bei (8.22–23) benutzt man als Erzeugende diejenigen Elemente  $c_{\nu, q}$ , die man in (7.22) zur Darstellung der  $g_{\nu, l}$  benötigt. Dabei tritt in den Relationen (8.20) ein Potenzprodukt  $\Phi$  auf, dessen Form nicht allgemein angebar ist, da sie von den Valenzen der  $s$  Randkurven abhängt. Sie kann natürlich in besonderen Fällen sehr einfach werden. Ist z. B.  $\mu(\gamma_1) \equiv 1 \pmod{n}$ , so kann man  $t = \gamma_1$  setzen, und die Relationen (8.20) lauten dann einfach  $c_{1, q} = 1$ ,  $0 \leq q \leq n-1$ . Man kommt dann sofort auf eine Darstellung von  $F$  mittels der freien Erzeugenden  $a$  und  $c_{\nu, q}$  für  $\nu = 2, \dots, s-1$ .

Zum näheren Vergleich der in 7 bzw. 8 behandelten Fälle  $q > 0$  bzw.  $q = 0$  sei noch folgende Bemerkung hinzugefügt: Wenn man für  $q > 0$  nicht von dem Ergebnis von 5 Gebrauch macht, sondern bei einem beliebigen Erzeugendensystem (4.1) von  $T$  mit der Relation (4.2) stehen bleibt, hat man eine (8.17) entsprechende Gleichung

$$t = II(\alpha_1, \dots, \gamma_s)$$

anzusetzen, und bekommt entsprechend (8.20) ein System von  $n$  Relationen

$$\Phi(a_{i,j+z}, b_{i,j+z}, c_{\nu,j+z}) = 1 \quad 0 \leq z \leq n-1,$$

und entsprechend (8.16) ein System von  $n$  Relationen, das eine Verallgemeinerung von (7.15) darstellt. Dann liesse sich eine für  $q > 0$  und  $q = 0$  gemeinsam gültige Normalform des Automorphismus von  $F$  angeben, sodass man nicht zwischen diesen beiden Fällen zu unterscheiden brauchte. Die Unterscheidung wird also dadurch nötig, dass wir die Normalformen so weit vereinfachen, dass sie keine nicht allgemein näher angebbaren Potenzprodukte  $\Phi$  enthalten.

**9. Zusätzliche Relationen bei Verzweigkeit.** In 3 wurde der Fall, dass Fixpunkte bei der Flächenabbildung  $\tau$  oder solchen Potenzen von  $\tau$ , deren Exponenten echte Teiler von  $n$  sind, auftreten, also dass  $\varphi$  über  $M$  verzweigt ist, dadurch auf den Fall der Unverzweigkeit zurückgeführt, dass Elementarflächenstücke um die Verzweigungspunkte aus  $M$  ausgeschnitten wurden, wodurch multiple Ränder statt der Verzweigungspunkte entstanden. Wir haben zu untersuchen, was es für die Fundamentalgruppe  $F$  und ihren in 7 und 8 dargestellten Automorphismus  $I$  bedeutet, wenn wir diesen Prozess wieder rückgängig machen.

Es sei etwa die  $\nu$ -te Randkurve von  $M$  durch Ausschneidung eines ursprünglichen Verzweigungspunktes von  $M$  entstanden. Sie ist daher multipel, also  $n \geq \lambda_\nu > 1$ ,  $1 \leq m_\nu < n$ . Die Schliessung dieser Randkurve von  $M$  durch ein Elementarflächenstück erfordert die Schliessung der sie überlagernden  $m_\nu$  Randkurven von  $\varphi$  durch  $m_\nu$  Elementarflächenstücke. Das bedeutet die Ersetzung der diesen  $m_\nu$  Randkurven von  $\varphi$  entsprechenden Elementklassen von  $F$  durch die Identität. Diese Elementklassen werden nun, wie am Schluss von 7 hervorgehoben, gerade durch die  $m_\nu$  Elemente (7.22) für den betreffenden Wert  $\nu$  repräsentiert. Die Wiedereinführung dieser Verzweigungsstelle von  $\varphi$  über  $M$  hat also für  $F$  nur die Wirkung, dass man unter Beibehaltung des benutzten Erzeugendensystems (7.16) bzw. (8.9) (für  $q > 0$  bzw.  $q = 0$ ) die Relationen

$$\begin{aligned}
 & c_{\nu,0} c_{\nu,\sigma_\nu m_\nu} \cdot \dots \cdot c_{\nu,(\lambda_\nu-1)\sigma_\nu m_\nu} \cdot a^{\sigma_\nu} = 1 \\
 (9.1) \quad & \text{--- --- ---} \\
 & c_{\nu,m_\nu-1} c_{\nu,\sigma_\nu m_\nu + m_\nu - 1} \cdot \dots \cdot c_{\nu,(\lambda_\nu-1)\sigma_\nu m_\nu + m_\nu - 1} \cdot a^{\sigma_\nu} = 1
 \end{aligned}$$

hinzuzufügen hat. Man hat dabei zu bedenken, dass die  $c_{\nu,q}$  im Falle  $q = 0$  zur Abkürzung für gewisse Ausdrücke in den Erzeugenden (8.9) stehen, während sie im Falle  $q > 0$  selbst zu den benutzten Erzeugenden gehören (oder Erzeugende transformiert mit Potenzen von  $a$  sind).

Stellt man in dieser Weise mehrere Verzweigungsstellen wieder her, so hat man Relationssysteme (9.1) für mehrere Werte von  $\nu$  hinzuzufügen. Ist  $\varphi$  geschlossen, so bestehen die Relationen (9.1) für alle Werte von  $\nu$ .

An der Darstellung des Automorphismus  $I$  von  $F$  durch (7.19) bzw. (8.10) ist dabei nichts zu ändern. *Daher sind (7.19) bzw. (8.10) auch Normaldarstellungen des Automor-*

phismus von  $F$  für den Fall, dass bei der periodischen Abbildung  $\tau$  der geschlossenen oder berandeten Fläche  $\varphi$ , oder bei den Potenzen von  $\tau$ , Fixpunkte auftreten.

Man sieht, dass die benutzte Darstellung von  $F$  durch Erzeugende und definierende Relationen der betreffenden periodischen Abbildung  $\tau$  so angepasst ist, dass sie die Fixpunkte in Evidenz setzt: Ein Fixpunkt bei  $\tau$  selbst hat eine Einzelrelation (9.1) zur Folge (also  $m_\nu = 1$ ,  $\lambda_\nu = n$ ); deren linke Seite geht bei dem Automorphismus  $I$  von  $F$  in eine Transformierte ihrer selbst über und lässt dabei die zu dem Fixpunkt bei  $\tau$  gehörige Drehzahl  $\frac{\delta}{n}$  erkennen; siehe (7.24) für  $m_\nu = 1$ ,  $\lambda_\nu = n$ . Ein Fixpunkt, der erstmalig bei  $\tau^m$ , wo  $m$  ein Teiler von  $n$  und  $1 < m < n$  ist, auftritt, gehört einem System von  $m$  Punkten von  $\varphi$  an, die sich bei  $\tau$ ,  $\tau^2$ ,  $\dots$ ,  $\tau^{m-1}$  vertauschen und alle gleichzeitig bei  $\tau^m$  Fixpunkte werden; sie haben  $m$  Relationen von  $F$  zur Folge, die ein System (9.1) bilden ( $m_\nu = m$ ,  $\lambda_\nu = \frac{n}{m}$ ). Deren linke Seiten vertauschen sich bei  $I$  so, wie durch (7.25) angegeben. Bei  $I^m$  geht jede in eine Transformierte von sich über und lässt dabei die zu den  $m$  Fixpunkten gehörige gemeinsame Drehzahl  $\frac{\delta}{\lambda}$  erkennen. Insbesondere sieht man:

*In der Normaldarstellung entspricht jedem multiplen Punkt von  $\varphi$  genau eine Relation von  $F$ . Für  $q=0$  gibt es keine weiteren Relationen als diese, für  $q > 0$  noch genau eine, nämlich (7.17).*

Der durch (7.19) bzw. (8.10) gegebene Automorphismus  $I$  von  $F$  hat in allen Fällen zugleich die Form eines Automorphismus der durch die betreffenden Erzeugenden hervorgebrachten freien Gruppe. Sein Charakter als Automorphismus der mit Relationen behafteten Gruppe  $F$  tritt

unmittelbar dadurch zu Tage, dass er offensichtlich das Relationssystem in ein äquivalentes überführt.

Wie in 7 und 8 hervorgehoben, ergibt sich aus der Darstellung von  $F$  und dem Automorphismus  $I$  auch eine Darstellung von  $T$ , indem man einerseits eine weitere Erzeugende, nämlich  $t$ , zu den Erzeugenden von  $F$  hinzunimmt, andererseits die Relation  $t^n = a$  und die Relationen (7.19) bzw. (8.10) zu den für  $F$  bestehenden Relationen hinzufügt. Deshalb bedeutet die Hinzunahme eines Relationssystems (9.1) nicht nur für  $F$  sondern auch für  $T$  den Übergang zu einer Faktorgruppe. Bezeichnen wir diese Faktorgruppen für den Augenblick mit  $F'$  und  $T'$ , so ist  $\frac{T'}{F'}$  die zyklische Gruppe  $n$ -ter Ordnung, ebenso wie  $\frac{T}{F}$ .

Nun zeigen (7.21) und (7.22), dass die Hinzufügung der  $m_\nu$  Relationen (9.1) für  $T$  dasselbe bedeutet, wie die Hinzufügung der einen Relation

$$(9.2) \quad \gamma_\nu^{\lambda_\nu} = 1,$$

also mit Benutzung der in der Normaldarstellung verwendeten Erzeugenden (siehe (7.4) und (8.13))

$$(9.3) \quad (c_\nu t^{\sigma_\nu m_\nu})^{\lambda_\nu} = 1.$$

Dies bedeutet das Auftreten von *Elementen endlicher Ordnung* in  $T'$ , während es in  $T$  keine gab.

$T'$  ist nicht etwa die Fundamentalgruppe der Modulfläche nach Schliessung der  $\nu$ -ten Randkurve; denn diese würde durch Hinzufügung der Relation  $\gamma_\nu = 1$  zu  $T$  entstehen. Statt dessen wird ja nur (9.2) gefordert, und es ist  $\lambda_\nu > 1$ . Man kann  $T'$  so erklären: Wie  $T$  die Decktransformationsgruppe der universellen Überlagerungsfläche  $\Phi$  von  $M$  über  $M$  ist, so ist  $T'$  die Decktransformations-

gruppe einer schwächeren Überlagerungsfläche  $\Phi'$  von  $M$ , bei der jede Deckkurve der  $\nu$ -ten Randkurve von  $M$  sich nach  $\lambda_\nu$  Umläufen schliesst. M. a. W.  $\Phi'$  entsteht aus  $\Phi$  durch Reduktion modulo derjenigen Untergruppe von  $T$ , die von allen Transformaten des Elements  $\gamma_\nu^{\lambda_\nu}$  erzeugt wird. Dann gilt analog mit (6.1) und (6.2)

$$(9.4) \quad M = \Phi' \text{ mod } T'$$

$$(9.5) \quad \varphi = \Phi' \text{ mod } F',$$

und  $\frac{T'}{F'}$  entspricht wieder der Decktransformationsgruppe von  $\varphi$  über  $M$ , d. h. der periodischen Abbildung  $\tau$ . Das bleibt richtig, wenn man nun das  $\nu$ -te Loch schliesst, und  $\Phi'$  kann dann wieder durch einen Bereich der hyperbolischen Ebene dargestellt werden, wie in 6. Dabei wird dann  $\gamma_\nu = c_\nu t^{\sigma_\nu m_\nu}$  in  $T'$  eine *elliptische Substitution* der Ordnung  $\lambda_\nu$  und mit der Drehzahl  $\frac{1}{\lambda_\nu}$ , während es in  $T$  nur hyperbolische Substitutionen gab.

Diese Betrachtungen wiederholen sich bei Schliessung weiterer Randkurven, indem man dadurch wieder zu Faktorgruppen  $F''$  und  $T''$  von  $F'$  und  $T'$  übergeht. Um der einfachen Ausdrucksweise willen behalten wir die Bezeichnung  $F$  und  $T$  für diese Gruppen ebenso wie  $\Phi$ ,  $\varphi$  und  $M$  für die Flächen auch nach Schliessung von Rändern bei. Dann ist immer  $F$  die Fundamentalgruppe von  $\varphi$ , und das ist die Decktransformationsgruppe von  $\Phi$  über  $\varphi$ , und  $\frac{T}{F}$  stellt die Decktransformationsgruppe von  $\varphi$  über  $M$ , also die periodische Abbildung  $\tau$ , dar. Aber  $T$  ist nur dann die Fundamentalgruppe von  $M$  und  $\Phi$  nur dann die universelle Überlagerungsfläche von  $M$ , wenn keine Schliessung multipler Randkurven stattgefunden hat, m. a. W., wenn es keine multiplen Punkte, also keine Relationen (9.1) gibt.

Eine besondere Erwähnung verdient der Fall, dass es Punkte der Multiplizität  $n$  gibt. Es entspreche z. B. dem Werte  $\nu = 1$  ein solcher, also  $m_1 = 1$ ,  $\lambda_1 = n$ . Die Relation (9.1) lautet dann

$$c_{1,0} c_{1,\sigma_1} \cdots c_{1,(n-1)\sigma_1} \cdot a^{\sigma_1} = 1$$

oder in der Form (9.3)

$$(c_1 l^{\sigma_1})^n = 1.$$

Es gibt also Elemente  $n$ -ter Ordnung in  $T$  und daher auch in der Restklasse  $Ft$  der Zerlegung (6.3). Es gibt demnach ein  $f$  in  $F$  so, dass

$$(9.6) \quad (ft)^n = 1$$

ist. Erstattet man  $t$  durch  $ft$ , so bedeutet das den Übergang von  $I$  zu einem damit verwandten  $I_1$ , nämlich die Hinzufügung des zu  $f$  gehörigen inneren Automorphismus von  $F$ . Und nun ist  $I_1^n$  der identische Automorphismus, da (9.6) an die Stelle von (6.5) tritt. Man sieht also:

*Dann und nur dann, wenn  $\tau\varphi$  Fixpunkte hat, gibt es in der zu  $\tau$  gehörigen Automorphismenfamilie Automorphismen endlicher Ordnung.*

Notwendige und hinreichende Bedingung dafür ist m. a. W., dass  $T$  Elemente  $n$ -ter Ordnung enthält. Da aus unserer Normaldarstellung zugleich eine Darstellung der ganzen Familie von  $I$  folgt, wird daher speziell eine Normaldarstellung aller Automorphismen endlicher Ordnung der Fundamentalgruppen geschlossener oder berandeter Flächen, soweit sie durch periodische Abbildungen induziert werden können, umfasst. Ob zu allen Automorphismen endlicher Ordnung von  $F$  periodische Abbildungen  $\tau$  von  $\varphi$  gehören, wird dadurch nicht entschieden.

**10. Strukturbeschreibung auf Grund der Normaldarstellung.** Man übersieht die Struktur einer  $n$ -periodischen Abbildung  $\tau$  von  $\varphi$ , wenn man  $\varphi$  in  $n$  Bereiche einteilen und die zyklische Vertauschung dieser Bereiche bei  $\tau$  in einfacher Weise beschreiben kann. Jeder der  $n$  Bereiche stellt dabei eine genau einmalige Überdeckung der Modulfläche  $M$  dar. Dabei genügt es, wie schon in 3 erwähnt, diese Strukturbeschreibung für den Fall der Unverzweichtigkeit, also bei Abwesenheit von multiplen Punkten, durchzuführen, da die Erweiterung der Abbildung auf die Elementarflächenstücke, durch die man eventuell multiple Randkurven zu schliessen hat, eine wohlbekannt einfache Struktur hat (Rotationssatz von B. v. KERÉKJÁRTÓ [5, 6]). Man benutzt nun am einfachsten die Überlagerungsfläche  $\Phi$  von  $\varphi$  und  $M$ , in der diese Bereiche kongruente Fundamentalbereiche für die Gruppe  $T$  von hyperbolischen Bewegungen sind. Diese Bereiche sind in  $\Phi$  in unendlicher Zahl vorhanden, aber wenn man unter ihnen  $n$  auswählt, die mod  $F$  verschieden sind und einen (nicht notwendig zusammenhängenden) Teilbereich  $\bar{\varphi}$  von  $\Phi$  ausmachen, so stellen sie  $\varphi$  genau einmal dar, indem diejenigen Seiten dieses Teilbereichs  $\bar{\varphi}$ , die dem Inneren von  $\Phi$  angehören, paarweise bei  $F$  äquivalent sind.

Den zu benutzenden Fundamentalbereich von  $T$  stellt man dadurch her, dass man die Modulfläche  $M$  der Fig. 1 teils längs der  $2q$  kanonischen Rückkehrschnitte  $\alpha_1, \beta_1, \dots, \alpha_q, \beta_q$ , teils längs  $s$  Querschnitten, die von  $O$  aus nach den  $s$  Randkurven von  $M$  geführt werden, aufschneidet. Die Rückkehrschnitte gibt es nur für  $q > 0$ , die Querschnitte nur für  $s > 0$ . Die aufgeschnittene Fläche entspricht in  $\Phi$  einem  $(4q + 3s)$ -Eck  $\bar{M}$ , von dessen Seiten  $s$  auf Randseiten von  $\Phi$  liegen, während die übrigen  $4q + 2s$  Seiten von  $\bar{M}$  dem Inneren von  $\Phi$  angehören und einander paarweise



bei den  $2q + s$  erzeugenden Operationen (4.1) von  $T$  entsprechen.

Nun sei zunächst  $q > 0$ . Dann sei das Rückkehrschnittsystem nach 5 und die Bezeichnung wie in 7 gewählt, also  $t$  statt  $\alpha_1$  und  $b_1, a_2, \dots, b_q$  statt  $\beta_1, \alpha_2, \dots, \beta_q$ . Die  $n$  Bereiche

$$(10.1) \quad \overline{M}, t\overline{M}, t^2\overline{M}, \dots, t^{n-1}\overline{M}$$

bilden einen zusammenhängenden Teilbereich  $\overline{\varphi}$  von  $\mathcal{O}$ , indem jeder der ersten  $n-1$  Bereiche der Reihe (10.1) längs einer dem Schnitt  $b_1$  entsprechenden Seite mit dem nächstfolgenden zusammenhängt.  $\overline{\varphi}$  hat also noch  $n(4q + 2s) - 2(n-1) = 2[n(2q + s - 1) + 1]$  freie Seiten im Inneren von  $\mathcal{O}$ , und diese müssen einander paarweise bei  $n(2q + s - 1) + 1$  Elementen von  $F$  entsprechen. Diese Zahl ist  $2p + r + n - 1$ , während wir in (7.16) nur  $2p + r$  Erzeugende von  $F$  benutzen. Wir betrachten die Herstellung von  $\varphi$  aus  $\overline{\varphi}$  durch paarweise Korrespondenz der inneren Seiten in folgender Reihenfolge, wobei die Normalform (7.19) des Automorphismus  $I$  als leitendes Prinzip dient:

Die beiden dem Schnitt  $b_2$  entsprechenden Seiten von  $\overline{M}$  korrespondieren bei  $a_2 = a_{2,0}$ , welches ja ein Element von  $F$  ist. Dann korrespondieren die beiden aus ihnen durch  $t$  hervorgehenden Seiten von  $t\overline{M}$  bei  $ta_2t^{-1}$ , und das ist nach (7.19) das Element  $a_{2,1}$  von  $F$ . So setzt man durch die ganze Reihe (10.1) fort und wiederholt das Verfahren mit den Elementen  $b_{2,j}, a_{3,j}, \dots, b_{q,j}, 0 \leq j \leq n-1$ .

Bei der Operation  $\gamma_\nu$  von  $T$  korrespondieren die beiden dem  $\nu$ -ten Querschnitt entsprechenden Seiten  $S_\nu$  und  $\gamma_\nu S_\nu$  von  $\overline{M}$ . Aber  $\gamma_\nu$  gehört im Allgemeinen nicht zu  $F$ , und dann sind diese beiden Seiten nicht bei der Herstellung von

$\varphi$  zu vereinigen. In jedem Fall kann man die Gleichung (7.3) in der Form

$$(10.2) \quad c_\nu a^x = \gamma_\nu t^{nx - \sigma_\nu m_\nu}$$

schreiben und dabei  $x$  so bestimmen, dass

$$(10.3) \quad 0 \leq \sigma_\nu m_\nu - xn \leq n-1$$

ist. Dann ist  $t^{\sigma_\nu m_\nu - xn} S_\nu$  eine Seite von  $\overline{\varphi}$ , die dem Bereich  $t^{\sigma_\nu m_\nu - xn} \overline{M}$  angehört, und diese Seite wird durch das zu  $F$  gehörige Element (10.2) auf die Seite  $\gamma_\nu S_\nu$  von  $\overline{M}$  abgebildet. Durch Transformation mit den ersten  $n-1$  Potenzen von  $t$  unter Beachtung von (7.18) und (7.19) erhält man die Korrespondenz der damit nach  $T$  äquivalenten Seiten in den übrigen Bereichen (10.1). Dies ist für alle Werte von  $\nu$  durchzuführen.

Nun ist die Korrespondenz nach  $F$  für  $n[2(q-1)+s]$  Seitenpaare von  $\overline{\varphi}$  hergestellt, und dabei sind die  $n[2(q-1)+s]$  Erzeugenden  $a_{i,j}$ ,  $b_{i,j}$ ,  $c_{\nu,j}$  von  $F$  verwendet worden. Es bleibt noch die Korrespondenz für die restlichen  $n+1$  Seitenpaare von  $\overline{\varphi}$  herzustellen.  $\overline{\varphi}$  besitzt noch genau 2 dem Schnitt  $b_1$  entsprechende Seiten, die eine in  $\overline{M}$ , die zweite in  $t^{n-1}\overline{M}$ ; die erste wird in die zweite durch  $t^n = a$  übergeführt. Es bleiben noch die  $n$  Seitenpaare, die dem Schnitt  $t (= \alpha_1)$  entsprechen. Die beiden zu  $\overline{M}$  gehörigen korrespondieren bei  $b = b_{1,0}$ , das ja zu  $F$  gehört. Die beiden zu  $t\overline{M}$  gehörigen korrespondieren daher bei  $tb t^{-1}$ , und die zweite Gleichung von (7.19) zeigt, dass dies Element von  $F$  durch die Erzeugenden (7.16) ausgedrückt werden kann. Das gleiche gilt daher auch für die Elemente  $t^2 b t^{-2}$ ,  $\dots$ ,  $t^{n-1} b t^{-(n-1)}$ , die die noch fehlenden Korrespondenzen in den übrigen Bereichen der Reihe

(10.1) herstellen. Man verwendet also hier die  $(n-1)$  ersten der Reihe der Elemente (7.15), die wir ursprünglich bei der Erzeugung von  $F$  mit verwendeten, aber nachträglich wieder eliminieren konnten.

Damit ist im Falle  $q > 0$  die Einteilung der Fläche  $\varphi$  in  $n$  Bereiche, ihr gegenseitiges Aneinandergrenzen auf  $\varphi$  und ihre zyklische Vertauschung bei  $\tau$  aufgezeigt. Ist dabei  $s > 0$ , so gehören  $ns$  Seiten von  $\overline{\varphi}$  zu Randseiten von  $\Phi$ . Die  $n$  Seiten von  $\overline{\varphi}$ , die aus dem  $\nu$ -ten Rand von  $\overline{M}$  entstanden sind, bilden bei dem Zusammenschluss von  $\overline{\varphi}$  zu  $\varphi$  zu je  $\lambda_\nu$  einen Rand von  $\varphi$ , und von solchen entstehen dabei  $m_\nu$ . Im ganzen entstehen dabei also  $r$  Ränder von  $\varphi$ , wobei  $r$  durch (3.1) ausgedrückt ist. Wie man einen Teil von diesen, oder alle, durch Elementarflächenstücke mit einem multiplen Punkt schliessen kann, ist schon erwähnt worden.

Kommen keine multiplen Ränder vor, so sind alle  $\gamma_\nu$  — falls überhaupt  $s > 0$ , also Ränder vorkommen — Elemente von  $F$ , und die Seiten  $S_\nu$  und  $\gamma_\nu S_\nu$  von  $\overline{M}$  korrespondieren daher bei einem Element von  $F$ . Durch die Vereinigung bei  $F$  korrespondierender Seiten von  $\overline{M}$  entsteht daher eine Fläche vom Geschlecht  $q-1$  mit  $s+2$  Randkurven, indem nämlich 2 Randkurven von den beiden dem Schnitt  $b_1$  entsprechenden Seiten stammen. Dasselbe wiederholt sich für die übrigen Bereiche der Reihe (10.1). Diese  $n$  Flächen ordnen sich durch zyklische Vereinigung der dem Schnitt  $b_1$  entsprechenden Randkurven zu einer ringartigen Fläche  $\varphi$  zusammen. Man kann die Fläche  $\varphi$  metrisch regulär so einrichten, dass sie eine Rotationsgruppe der Ordnung  $n$  gestattet. Dann ist  $\tau$  die  $\frac{1}{n}$  Drehung von  $\varphi$  in sich. Dabei ist  $p = n(q-1) + 1$ , wie in (3.3) ausgedrückt ist. — Dieser

Fall liegt insbesondere stets für  $s = 0$ , also für periodische Transformationen geschlossener Flächen ohne multiple Punkte, vor.

Für  $q = 0$  ist das Verfahren analog wie für  $q > 0$ . Für  $q = 0$  ist  $\bar{M}$  ein  $3s$ -Eck, von dessen Seiten  $s$  auf Randseiten von  $\mathcal{D}$  liegen, während die übrigen  $2s$  dem Inneren von  $\mathcal{D}$  angehören. Man bildet wieder mittels des in 8 hergestellten Elements  $t$  von  $T$  die Reihe der Bereiche (10.1), die zusammen  $\bar{\varphi}$  ausmachen. Hierbei ist  $\bar{\varphi}$  nicht notwendig zusammenhängend, aber die  $2ns$  Seiten von  $\bar{\varphi}$ , die nicht auf Randseiten von  $\mathcal{D}$  liegen, korrespondieren wieder paarweise bei Elementen aus  $F$ . Sind wieder  $S_\nu$  und  $\gamma_\nu S_\nu$  ein Seitenpaar von  $\bar{M}$ , so verschafft man sich durch (10.2) und (10.3) diejenige (eindeutig bestimmte) Seite  $t^{\sigma_\nu m_\nu - \alpha n} S_\nu$  von  $\bar{\varphi}$ , die durch das zu  $F$  gehörige Element (10.2) auf  $\gamma_\nu S_\nu$  abgebildet wird. Dann hat man wieder mit den Potenzen von  $t$  zu transformieren und dabei auf (7.18) und (8.23) zu achten. Führt man dies für alle Werte von  $\nu$  durch, so erhält man vollen Aufschluss über die Lagerung der  $n$  Bereiche auf  $\varphi$  und ihre zyklische Vertauschung bei  $\tau$ .

Der Aufbau von  $\varphi$  ist im Falle  $q = 0$  besonders anschaulich, wenn Randkurven von der Multiplizität  $n$  vorkommen. Nehmen wir speziell an, es sei  $\mu(\gamma_1) \equiv 1 \pmod{n}$ . Dann kann man  $\gamma_1$  als das  $t$  von 8 und (10.1) nehmen. In diesem Fall wird  $\bar{\varphi}$  zusammenhängend, indem die  $n$  Bereiche (10.1) längs der  $\gamma_1$  entsprechenden Randkurve von  $\varphi$  (wegen  $m_1 = 1$  gibt es nur die eine  $\gamma_1$  entsprechende) zyklisch angeordnet sind. Man stellt dann leicht die Korrespondenz der übrigen nicht auf Randseiten von  $\mathcal{D}$  liegenden Seiten von  $\bar{\varphi}$  her. — Ist der erste Rand aus einem multiplen Punkt entstanden, so hat man nach Schliessung das bekannte Bild eines Sternes von  $n$  Bereichen

um den Drehpunkt einer elliptischen Substitution  $n$ -ter Ordnung.

**11. Der Äquivalenzsatz:** *Wenn zwei homöomorphe, orientierbare Flächen  $\varphi$  und  $\varphi'$  periodischen, die Orientierung erhaltenden Transformationen  $\tau$  bzw.  $\tau'$  in sich von gleicher Ordnung unterworfen sind, so sind  $\tau$  und  $\tau'$  dann und nur dann topologisch äquivalent, wenn man — nötigenfalls nach Wechsel der Orientierung auf der einen Fläche — die Randkurven und die bei  $\tau$  multiplen Punkte von  $\varphi$  den Randkurven bzw. den bei  $\tau'$  multiplen Punkten von  $\varphi'$  so zuordnen kann, dass zugeordnete Randkurven und zugeordnete multiple Punkte die gleiche Valenz haben.*

Dass die Bedingung notwendig ist, geht aus der am Schluss von 2 hervorgehobenen Invarianz der Valenz bei topologischer Abbildung hervor.

Den Beweis dafür, dass die Bedingung hinreichend ist, kann man zunächst dadurch erleichtern, dass man eventuell vorhandene multiple Punkte auf  $\varphi$  und  $\varphi'$  nach dem Verfahren von 3 austrennt. Zwei einander zugeordnete multiple Punkte von  $\varphi$  und  $\varphi'$  ergeben dabei zwei multiple Ränder von  $\varphi$  und  $\varphi'$ , die wiederum gleiche Valenz haben, also einander zugeordnet werden können. Es ist wohl klar, wie man aus einer gegebenen Äquivalenz von  $\tau$  und  $\tau'$  vor der Austrennung eine ebensolche nach der Austrennung, und umgekehrt, herleiten kann. Wir können also annehmen, dass bei  $\tau$  und  $\tau'$  keine multiplen Punkte vorkommen.

Nun seien  $p$  und  $r$  Geschlecht und Ränderzahl von  $\varphi$  und damit auch von der mit  $\varphi$  homöomorphen Fläche  $\varphi'$ , und  $n$  die gemeinsame Ordnung von  $\tau$  und  $\tau'$ . Sei  $R$  eine Randkurve von  $\varphi$  und  $R'$  die ihr zugeordnete von  $\varphi'$ , beide

mit der Valenz  $[m, \lambda, \sigma]$ . Dann gehört  $R$  zu einem System von  $m$  Rändern  $R, \tau R, \dots, \tau^{m-1}R$  von  $\varphi$ , die die gleiche Randkurve auf der Modulfläche  $M$   $\lambda$ -fach überdecken, ebenso  $R'$  zu einem System von  $m$  Rändern  $R', \tau' R', \dots, \tau'^{m-1}R'$  von  $\varphi'$ , die die gleiche Randkurve auf der Modulfläche  $M'$   $\lambda$ -fach überdecken. Man kann die gegebene Zuordnung von Rändern gleicher Valenz nötigenfalls so abändern, dass sie jene  $m$  Kurven von  $\varphi$  diesen  $m$  Kurven von  $\varphi'$  zuordnet, und in der gleichen Weise fortfahren. Man hat daher auf beiden Flächen die gleiche Zusammensetzung der Zahl  $r$  nach (3.1), also die gleiche Ränderzahl  $s$  der Modulflächen  $M$  und  $M'$ . Nach (3.2) haben  $M$  und  $M'$  dann auch das gleiche Geschlecht  $g$ .

Nun bringe man, falls  $g > 0$  ist, auf  $M$  und  $M'$  ein Rückkehrschnittsystem  $\alpha_1, \dots, \beta_q$  bzw.  $\alpha'_1, \dots, \beta'_q$  wie in Fig. 1 so an, dass  $\alpha_1$  und  $\alpha'_1$  den  $\mu$ -Wert 1 und alle übrigen Rückkehrschnitte den  $\mu$ -Wert 0 haben, ordne, falls  $s > 0$  ist, die  $s$  Randkurven von  $M$  und  $M'$  so, dass die Valenzen von Rändern mit der gleichen Nummer  $\nu$  übereinstimmen, und ziehe wie in 10 die Querschnitte von  $O$  bzw.  $O'$  nach den Rändern von  $M$  bzw.  $M'$ . Dann sei  $\mathcal{G}$  eine topologische Abbildung von  $M$  auf  $M'$ , die das Schnittsystem von  $M$  auf das gleichbezeichnete von  $M'$  abbildet.  $\mathcal{G}$  ist zugleich eine topologische Abbildung des Fundamentalbereichs  $\bar{M}$  von  $T$  in  $\mathcal{O}$  (siehe 10) auf einen Fundamentalbereich  $\bar{M}'$  von  $T'$  in  $\mathcal{O}'$ . Zwei Randpunkte von  $\bar{M}$ , die sich bei einem Element  $z$  aus der Erzeugendenreihe (4.1) von  $T$  entsprechen, werden durch  $\mathcal{G}$  in zwei Randpunkte von  $\bar{M}'$  abgebildet, die sich bei dem entsprechenden Element  $z'$  aus der Erzeugendenreihe

$$(11.1) \quad \alpha'_1, \beta'_1, \dots, \beta'_q, \gamma'_1, \dots, \gamma'_s$$

von  $T'$  entsprechen, d. h. für einen solchen Randpunkt  $x$  von  $\overline{M}$  gilt die Beziehung

$$(11.2) \quad \mathcal{J}zx = x'\mathcal{J}x.$$

Dann wird aber  $\mathcal{J}$  durch die Forderung der unbeschränkten Gültigkeit der Funktionalgleichung (11.2) für alle Erzeugenden  $z$ , und damit für alle Elemente von  $T$  überhaupt, zu einer topologischen Abbildung von ganz  $\mathcal{O}$  auf ganz  $\mathcal{O}'$ .

Nun haben wir die Fundamentalgruppen  $F$  und  $F'$  von  $\mathcal{O}$  bzw.  $\mathcal{O}'$  herzustellen. Dabei war im Falle  $q > 0$  in 7 ein System von  $s$  Zahlen  $\sigma_\nu$  so auszuwählen, dass (7.2) galt. Da wir bei  $T'$  wegen der Gleichheit der Rändervalenz die Zahl  $\sigma_\nu$  in derselben Restklasse mod  $\lambda_\nu$  zu wählen haben, wie bei  $T$ , können wir bei  $T'$  dieselben Zahlen  $\sigma_\nu$  wählen, wie bei  $T$ . Im Falle  $q = 0$  hatten wir in 8 in anderer Weise einen nicht ganz eindeutig festgelegten Prozess auszuführen, indem wir von dem System (8.3) von freien Erzeugenden von  $T$  durch eine Reihe von Erzeugendenwechseln zu dem System (8.5) übergingen. Da nun zu den Erzeugenden  $\gamma'_1, \gamma'_2, \dots, \gamma'_{s-1}$  von  $T'$  der Reihe nach dieselben Valenzen gehören, wie zu (8.3), können wir die Reihe der Erzeugendenwechsel für  $T'$  genau nach dem Schema derjenigen für  $T$  verlaufen lassen, und gelangen dadurch zu einem neuen Erzeugendensystem

$$(11.3) \quad t', f'_{1,0}, f'_{2,0}, \dots, f'_{s-2,0}$$

für  $T'$ , dessen Elemente dieselben Ausdrücke in  $\gamma'_1, \dots, \gamma'_{s-1}$  sind, wie (8.5) in (8.3). Bei Innehaltung dieser beiden Vorschriften geht dann alles zwangsläufig: Zwei gleichbezeichnete Elemente von  $F$  und  $F'$  entsprechen sich bei dem Isomorphismus zwischen  $T$  und  $T'$ , der durch die Zuordnung der Erzeugendenreihe (11.1) von  $T'$  zu der Erzeu-

gendenreihe (4.1) von  $T$  hergestellt wurde. Insbesondere gilt dann die Funktionalgleichung

$$\mathcal{J}tx = t'\mathcal{J}x,$$

indem man in (11.2)  $t$  für  $z$  setzt, und wenn man diese auf  $\varphi = \mathcal{O} \bmod F$  und  $\varphi' = \mathcal{O}' \bmod F'$  deutet,

$$\mathcal{J}tP = t'\mathcal{J}P$$

für jeden Punkt  $P$  von  $\varphi$ . Also gilt (2.9), wobei  $\mathcal{J}$  nun diejenige topologische Abbildung von  $\varphi$  auf  $\varphi'$  bedeutet, die sich aus  $\mathcal{J}\mathcal{O} = \mathcal{O}'$  und der Gültigkeit von (11.2) für alle Elemente  $z$  von  $F$  ergibt.

Damit ist der Beweis für den Äquivalenzsatz erbracht. In [13] habe ich diesen Satz ohne Benutzung der Normaldarstellung durch direktes Operieren mit  $\varphi$  als regulärer Überlagerungsfläche von  $M$  hergeleitet. Man erkennt, dass man durch Variieren des Zahlensystems

$$(11.4) \quad n, q, s, m_\nu, \lambda_\nu, \sigma_\nu \quad \nu = 1, 2, \dots, s,$$

das den topologischen Typus der Abbildung festlegt, unter Innehaltung der für diese Zahlen geltenden Relationen (1.3), (2.4), (4.6) und für  $q = 0$  (4.7) eine vollständige Aufzählung aller topologischen Typen periodischer Transformationen (mit Erhaltung der Orientierung) angeben kann. Vgl. hierzu die Ausführungen von W. SCHERRER [9]. Die Zahlen  $r$  und  $p$  ergeben sich für jeden Typus (11.4) aus (3.1) und (3.2). Bei der Typenaufzählung (11.4) ist nicht mit multiplen Punkten, sondern nur mit Randkurven gerechnet. Je nachdem, wie man unter den  $s$  Randkurven der Modulfläche eine ausgewählte Teilmenge von multiplen durch Elementarflächenstücke mit je einem multiplen Punkt schliesst, bekommt man für jeden Typus (11.4) eine Unterteilung in endlich viele topologisch verschiedene Typen.



Der zum Werte  $n = 2$  gehörige Spezialfall des Äquivalenzsatzes ist von W. SCHERRER in [8] und [10] aufgestellt worden. Da man für multiple Ränder und Punkte im Falle  $n = 2$  nur die eine Möglichkeit  $m = 1$ ,  $\lambda = 2$ ,  $\sigma \equiv 1 \pmod{2}$  hat, kann man dem Äquivalenzsatz für  $n = 2$  folgende Form geben: Indikatrizerhaltende Transformationen zweiter Ordnung homöomorpher Flächen sind dann und nur dann topologisch äquivalent, wenn sie in der Anzahl der Fixpunkte und in der Anzahl der invarianten Ränder übereinstimmen.

Hervorgehoben sei auch der folgende besonders einfache Spezialfall des Äquivalenzsatzes: Indikatrizerhaltende Transformationen  $n$ -ter Ordnung homöomorpher Flächen, bei denen vor der  $n$ -ten Potenz kein Fixpunkt und kein invarianter Rand auftritt, sind immer äquivalent.

## 12. Automorphismus der Homologiegruppe für $q > 0$ .

Die Homologiegruppe  $H$  von  $\varphi$  ist die Faktorgruppe der Kommutatorgruppe von  $F$ . Der durch die Transformation mit  $t$  in  $F$  induzierte Automorphismus  $I$  hat dabei einen Automorphismus  $J$  von  $H$  zur Folge. Da  $I^n$  ein innerer Automorphismus von  $F$  ist, also jedes Element von  $F$  in seiner Elementklasse lässt und daher jedes Element von  $H$  in sich überführt, ist  $J^n$  der identische Automorphismus von  $H$ . Jedes Element von  $F$  bestimmt eindeutig ein Element von  $H$ , und dieses geben wir durch dasselbe Zeichen wieder. Der Übergang von  $F$  zu  $H$  kommt dann einfach dadurch zum Ausdruck, dass man die Erzeugenden von  $F$  »vertauschbar macht«. Insbesondere wird ein Element von  $F$  in  $H$  identisch mit den Elementen, die aus ihm durch Transformation mit Elementen von  $F$  hervorgehen. Wegen des abelschen Charakters von  $H$  ist es bequem,  $H$  additiv

zu schreiben, sodass jedes Element von  $H$  durch eine lineare Verbindung der benutzten Erzeugenden dargestellt wird. Dann wird der Automorphismus  $J$  durch eine lineare Transformation des Erzeugendensystems, also durch eine quadratische Matrix, ausgedrückt, und weil  $J^n$  die Identität ist, kann man diese zu einer Diagonalmatrix umgestalten, indem man für die Linearverbindungen Koeffizienten aus dem mit der Zahl

$$(12.1) \quad \varepsilon = e^{\frac{2\pi i}{n}}$$

gebildeten Zahlkörper nimmt. Dabei ist  $\varepsilon$  Wurzel der Gleichung

$$(12.2) \quad x^n - 1 = 0$$

und für jede Faktorzerlegung  $n = m\lambda$  mit  $m > 1$  und jedes  $l$ ,  $0 < l < m$ , gilt

$$(12.3) \quad 1 + \varepsilon^{l\lambda} + \varepsilon^{2l\lambda} + \dots + \varepsilon^{(m-1)l\lambda} = 0.$$

Der Automorphismus  $J$  von  $H$  wird dann vollständig bestimmt durch das System der *Eigenwerte*, die in der Hauptdiagonale der Diagonalmatrix auftreten und mit denen sich die Elemente der speziell ausgewählten Homologiebasis der Reihe nach multiplizieren. Das System dieser Eigenwerte geben wir am einfachsten an durch das *charakteristische Polynom*, das genau diese Eigenwerte zu Wurzeln hat; dieses Polynom ist ja die einzige Invariante von  $J$  gegenüber beliebigem Wechsel der Homologiebasis, die man zur Erzeugung von  $H$  verwendet.

Den hiermit vorgezeichneten Weg zur Aufstellung von  $J$  führen wir nun zunächst für den Fall der Unverzweigkeit durch, und da wir dabei die in 7 und 8 gegebene

Normaldarstellung von  $F$  und  $I$  zu benutzen haben, behandeln wir die Fälle  $q > 0$  und  $q = 0$  getrennt. Für beide Fälle kann man die allgemeine Regel aufstellen, dass bei Verwendung von Symbolen  $a, b, c$  oder  $f$  mit doppelten Indizes der zweite Index nur nach seinem Restcharakter mod  $n$  in Betracht kommt, wenn das betreffende Symbol ein Element von  $H$  bezeichnet; siehe z. B. (7.18).

Im Falle  $q > 0$  verwenden wir (7.16) als Erzeugendensystem für  $F$  und haben dabei die definierende Relation (7.17). Wir können daher zunächst  $H$  durch dieselben Symbole (7.16) erzeugen, d. h. die Elemente von  $H$  als Linearverbindungen in diesen Symbolen schreiben, und bekommen dabei aus (7.17) die einzige Relation

$$(12.4) \quad \sum_{\nu=1}^s \sum_{j=0}^{n-1} c_{\nu,j} = 0.$$

Eliminiert man im Falle  $s > 0$  mittels (12.4) eine der Erzeugenden  $c_{\nu,j}$ , so bekommt man eine unabhängige Homologiebasis.

Nun gehen wir zu neuen Erzeugenden

$$(12.5) \quad a, B, A_{i,j}, B_{i,j}, C_{\nu,j}$$

mit gleichen Wertbereichen der Indizes wie in (7.16) über, indem wir von (7.16) nur  $a$  behalten und im übrigen folgendermassen substituieren:

$$(12.6) \quad \begin{aligned} A_{i,0} &= a_{i,0} + a_{i,1} + a_{i,2} + \dots + a_{i,n-1} \\ A_{i,1} &= a_{i,0} + \varepsilon^{-1} a_{i,1} + \varepsilon^{-2} a_{i,2} + \dots + \varepsilon^{-(n-1)} a_{i,n-1} \\ A_{i,2} &= a_{i,0} + \varepsilon^{-2} a_{i,1} + \varepsilon^{-4} a_{i,2} + \dots + \varepsilon^{-2(n-1)} a_{i,n-1} \\ &\text{---} \\ A_{i,n-1} &= a_{i,0} + \varepsilon^{-(n-1)} a_{i,1} + \varepsilon^{-2(n-1)} a_{i,2} + \dots + \varepsilon^{-(n-1)(n-1)} a_{i,n-1} \end{aligned}$$

und entsprechend

$$(12.7) \quad B_{i,j} = b_{i,0} + \varepsilon^{-j} b_{i,1} + \varepsilon^{-2j} b_{i,2} + \cdots + \varepsilon^{-(n-1)j} b_{i,n-1}$$

$$(12.8) \quad C_{\nu,j} = c_{\nu,0} + \varepsilon^{-j} c_{\nu,1} + \varepsilon^{-2j} c_{\nu,2} + \cdots + \varepsilon^{-(n-1)j} c_{\nu,n-1}$$

und endlich

$$(12.9) \quad B = nb - (n-1) [c_{1,0} + c_{2,\sigma_1 m_1} + c_{3,\sigma_1 m_1 + \sigma_2 m_2} + \cdots + c_{s,-\sigma_s m_s}] \\ - (n-2) [c_{1,1} + c_{2,\sigma_1 m_1 + 1} + \cdots + c_{s,-\sigma_s m_s + 1}] \\ \text{---} \\ - 1 \cdot [c_{1,n-2} + c_{2,\sigma_1 m_1 + n-2} + \cdots + c_{s,-\sigma_s m_s + n-2}].$$

In der Tat ist (12.5) ein neues Erzeugendensystem für  $H$ . Denn die Determinante von (12.6) ist die Differente der Gleichung (12.2), also von Null verschieden. Man kann somit die  $a_{i,j}$  aus den  $A_{i,j}$ , ebenso die  $b_{i,j}$  aus den  $B_{i,j}$  und die  $c_{\nu,j}$  aus den  $C_{\nu,j}$  ausdrücken und schliesslich  $b$  aus (12.9) herstellen. Dabei gilt für die neuen Erzeugenden statt (12.4) die einfachere Relation

$$(12.10) \quad \sum_{\nu=1}^s C_{\nu,0} = C_{1,0} + C_{2,0} + \cdots + C_{s,0} = 0.$$

Ist  $s > 0$ , so kann man mittels (12.10) z. B.  $C_{s,0}$  eliminieren und die übrigen  $n [2(q-1) + s] + 1 = 2p + r - 1$  Elemente (12.5) bilden eine unabhängige Homologiebasis. Ist  $s = 0$ , so fallen die  $C_{\nu,j}$  fort, da die  $c_{\nu,j}$  fortfallen, und damit fällt auch die Relation (12.10) fort. In diesem Fall enthält (12.5)  $2p$  Elemente, die eine unabhängige Homologiebasis bilden.

Bei Erzeugung von  $H$  durch (12.5) wird  $J$  in Diagonalform dargestellt. Denn auf Grund von (7.19) ergeben sich für die Elemente (12.5) folgende Ersetzungen bei  $J$ :

$$(12.11) \left. \begin{aligned} A_{i,j} &\rightarrow \varepsilon^j A_{i,j} \\ B_{i,j} &\rightarrow \varepsilon^j B_{i,j} \\ C_{\nu,j} &\rightarrow \varepsilon^j C_{\nu,j} \quad \nu = 1, 2, \dots, s \\ a &\rightarrow 1 \cdot a \\ B &\rightarrow 1 \cdot B. \end{aligned} \right\} \begin{array}{l} i = 2, 3, \dots, q \\ j = 0, 1, \dots, n-1. \end{array}$$

Denn  $B$  geht zunächst nach (7.19) über in

$$\begin{aligned} nb - n [c_{1,0} + c_{2,\sigma_1 m_1} + \dots + c_{s,-\sigma_s m_s}] \\ - (n-1) [c_{1,1} + \dots] \\ - (n-2) [c_{1,2} + \dots] \\ \dots \dots \dots \\ - 1 \cdot [c_{1,n-1} + \dots], \end{aligned}$$

und addiert man hierzu alle  $c_{\nu,j}$  einmal, was nach (12.4) statthaft ist, so bekommt man wieder  $B$ .

Da die  $\varepsilon^j$ ,  $0 \leq j \leq n-1$ , sämtliche Wurzeln von (12.2) sind, sind die in (12.11) auftretenden Eigenwerte sämtliche Wurzeln des Polynoms  $(x^n - 1)^{2q+s-2}(x-1)^2$ . Im Falle  $s > 0$  hat man aber, um auf eine unabhängige Homologiebasis zu kommen, etwa  $C_{s,0}$  fortzulassen, wodurch man einen Eigenwert 1 verliert. Also bekommt man für das zu  $J$  gehörige charakteristische Polynom  $P_H(x)$  im Falle  $q > 0$ :

$$(12.12) \quad P_H(x) = (x^n - 1)^{2q+s-2}(x-1) \quad \text{für } s > 0$$

$$(12.13) \quad P_H(x) = (x^n - 1)^{2q-2}(x-1)^2 \quad \text{für } s = 0.$$

Nun soll noch diejenige Untergruppe  $R$  von  $H$  näher bestimmt werden, deren Elemente als Linearkombination der den Randkurven von  $\varphi$  entsprechenden Elemente von  $H$  ausgedrückt werden können.  $R$  verschwindet nur im Falle  $s = 0$ . Denn ist  $s > 0$ , so ist  $r > 1$ ; sonst müsste wegen

(3.1)  $s = 1$ ,  $m_1 = 1$  sein, multiple Ränder können aber wegen (4.6) nicht in der Einzahl auftreten. Nun werden die den Rändern von  $\varphi$  entsprechenden Elementklassen von  $F$  nach 7 durch die Elemente (7.21) repräsentiert. Die  $g_{\nu,l}$  von (7.22) erzeugen also  $R$ . Statt ihrer führen wir das folgende linear transformierte System für  $\nu = 1, 2, \dots, s$  ein:

$$\begin{aligned}
 G_{\nu,0} &= g_{\nu,0} + g_{\nu,1} + g_{\nu,2} + \dots + g_{\nu,m_\nu-1} \\
 (12.14) \quad G_{\nu,1} &= g_{\nu,0} + \varepsilon^{-\lambda_\nu} g_{\nu,1} + \varepsilon^{-2\lambda_\nu} g_{\nu,2} + \dots + \varepsilon^{-(m_\nu-1)\lambda_\nu} g_{\nu,m_\nu-1} \\
 &\quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \\
 G_{\nu,m_\nu-1} &= g_{\nu,0} + \varepsilon^{-(m_\nu-1)\lambda_\nu} g_{\nu,1} + \varepsilon^{-2(m_\nu-1)\lambda_\nu} g_{\nu,2} + \dots \\
 &\quad \dots + \varepsilon^{-(m_\nu-1)(m_\nu-1)\lambda_\nu} g_{\nu,m_\nu-1}.
 \end{aligned}$$

Nun zeigen (12.14), (7.22) und (12.8), dass dieses Erzeugendensystem für die Untergruppe  $R$  von  $H$  sich in folgender Weise durch das Erzeugendensystem (12.5) für  $H$  ausdrückt:

$$(12.15) \quad G_{\nu,0} = C_{\nu,0} + \sigma_\nu m_\nu a$$

und, falls  $m_\nu > 1$  ist,

$$(12.16) \quad G_{\nu,l} = C_{\nu,l} \varepsilon^{\lambda_\nu} \quad l = 1, 2, \dots, m_\nu - 1,$$

da  $a$  in (12.16) wegen der durch (12.3) ausgedrückten Eigenschaft von  $\varepsilon$  nicht auftritt. Dabei ist wegen (7.2), (12.15) und (12.10)

$$(12.17) \quad \sum_{\nu=1}^s G_{\nu,0} = \sum_{\nu=1}^s \sum_{l=0}^{m_\nu-1} g_{\nu,l} = 0.$$

Diese Gleichung drückt einfach die Tatsache aus, dass die Summe aller Randkurven von  $\varphi$  homolog Null ist.

Ferner zeigen (7.25) und (12.14), dass man für die  $G_{\nu,l}$  folgende Ersetzungen bei  $J$  erhält:

$$(12.18) \quad G_{\nu,l} \rightarrow \varepsilon^{l\lambda_\nu} G_{\nu,l} \quad l = 0, 1, 2, \dots, m_\nu - 1.$$

Dies ergibt sich auch aus (12.15), (12.16) und (12.11). Berücksichtigt man noch, dass man, um eine unabhängige Basis für  $R$  zu erhalten, nach (12.17) etwa  $G_{s,0}$  fortzulassen hat, und dass  $\varepsilon^{\lambda_\nu}$  der Gleichung  $x^{m_\nu} = 1$  genügt, so findet man für  $s > 0$  für  $R$  das charakteristische Polynom

$$(12.19) \quad P_R(x) = \frac{(x^{m_1} - 1)(x^{m_2} - 1) \cdots (x^{m_s} - 1)}{x - 1} \quad (s > 0).$$

### 13. Automorphismus der Homologiegruppe für $q = 0$ .

Im Falle  $q = 0$  hat man die  $n(s-2) + 1 = 2p + r - 1$  Symbole (8.9) als Erzeugende der Homologiegruppe  $H$ , und diese bilden hier eine unabhängige Basis der abelschen Gruppe  $H$ . Ebenso wie in 12 gehen wir mit Benutzung von (12.1) zu einem neuen Erzeugendensystem

$$(13.1) \quad a, F_{i,j}$$

mit gleichen Wertbereichen der Indizes wie in (8.9) über, indem wir substituieren

$$(13.2) \quad F_{i,j} = f_{i,0} + \varepsilon^{-j} f_{i,1} + \varepsilon^{-2j} f_{i,2} + \cdots + \varepsilon^{-(n-1)j} f_{i,n-1},$$

sodass (13.1) wieder eine unabhängige Basis wird. Nach (8.10) hat man dann für die Basis (13.1) bei dem Automorphismus  $J$  von  $H$  die Ersetzung

$$(13.3) \quad \begin{aligned} & a \rightarrow 1 \cdot a \\ & F_{i,j} \rightarrow \varepsilon^j F_{i,j} \quad i = 1, 2, \dots, s-2; \quad j = 0, 1, \dots, n-1 \end{aligned}$$

Bei Benutzung der Basis (13.1) hat man also wieder die Transformationsmatrix in Diagonalf orm, und man erhält als charakteristisches Polynom

$$(13.4) \quad P_H(x) = (x^n - 1)^{s-2} (x - 1).$$

Das ist der  $q = 0$  entsprechende Spezialfall von (12.12).

Nun soll auch hier diejenige Untergruppe  $R$  von  $H$  aufgestellt werden, deren Elemente mit einer Linearkombination von Rändern homolog sind. Diese Ränder werden durch die Elementklassen der  $g_{\nu,l}$  (siehe (8.11)) in  $F$  dargestellt, und für die  $g_{\nu,l}$  gelten die Gleichungen (7.22), wobei die  $c_{\nu} = c_{\nu,0}$  durch (8.13) als gewisse Potenzprodukte in den  $f_{i,j}$  eingeführt sind.

Setzt man beim Übergang zu  $H$

$$(13.5) \quad c_{\nu,0} = \sum_{i=1}^{s-2} \sum_{j=0}^{n-1} \alpha_{\nu}^{i,j} f_{i,j},$$

so hat man nach (8.15)

$$(13.6) \quad c_{\nu,q} = \sum_{i=1}^{s-2} \sum_{j=0}^{n-1} \alpha_{\nu}^{i,j} f_{i,j+q},$$

wobei Indizes der  $c$  und  $f$ , die an zweiter Stelle stehen, in  $H$  nur nach ihrem Restcharakter mod  $n$  in Betracht kommen.

Nun ist nach (7.22) und (13.6) für  $k = 0, 1, 2, \dots, m_{\nu} - 1$ :

$$(13.7) \quad \begin{aligned} g_{\nu,k} &= c_{\nu,k} + c_{\nu,k+\sigma_{\nu}m_{\nu}} + c_{\nu,k+2\sigma_{\nu}m_{\nu}} + \dots + c_{\nu,k+(\lambda_{\nu}-1)\sigma_{\nu}m_{\nu}} + \sigma_{\nu}\alpha \\ &= \sum_{i,j} \alpha_{\nu}^{i,j} [f_{i,j+k} + f_{i,j+k+\sigma_{\nu}m_{\nu}} + \dots + f_{i,j+k+(\lambda_{\nu}-1)\sigma_{\nu}m_{\nu}}] + \sigma_{\nu}\alpha \end{aligned}$$

Indem man wieder durch die Transformation (12.14),



deren Determinante nicht verschwindet, von den Elementen  $g_{\nu,l}$  zu neuen Elementen  $G_{\nu,l}$  übergeht, findet man für  $l = 0, 1, 2, \dots, m_\nu - 1$ :

$$\begin{aligned}
 (13.8) \quad G_{\nu,l} &= \sum_{k=0}^{m_\nu-1} \varepsilon^{-lk\lambda_\nu} g_{\nu,k} \\
 &= \sum_{i,j} \alpha_\nu^{i,j} \left\{ \sum_k \varepsilon^{-lk\lambda_\nu} [f_{i,j+k} + f_{i,j+k+\sigma_\nu m_\nu} + \dots] \right\} \\
 &\quad + \sigma_\nu a \sum_k \varepsilon^{-lk\lambda_\nu}.
 \end{aligned}$$

Hierin hat das letzte Glied den Wert  $\sigma_\nu m_\nu a$  für  $l = 0$  und nach (12.3) den Wert 0 für  $l = 1, 2, \dots, m_\nu - 1$ . Die geschweifte Klammer im ersten Gliede wird, ausführlich geschrieben, indem  $k$  von 0 bis  $m_\nu - 1$  läuft,

$$\begin{aligned}
 (13.9) \quad \{ \} &= f_{i,j} + \varepsilon^{-l\lambda_\nu} f_{i,j+1} + \varepsilon^{-2l\lambda_\nu} f_{i,j+2} + \dots + \varepsilon^{-(m_\nu-1)l\lambda_\nu} f_{i,j+m_\nu-1} \\
 &\quad + f_{i,j+m_\nu} + \varepsilon^{-l\lambda_\nu} f_{i,j+m_\nu+1} + \dots \\
 &\quad + f_{i,j+2m_\nu} + \dots \\
 &\quad + \text{---} \\
 &= \varepsilon^{jl\lambda_\nu} F_{i,l\lambda_\nu}.
 \end{aligned}$$

Dabei ist benutzt, dass man die Summanden umordnen und im zweiten Index  $\delta_\nu \sigma_\nu m_\nu$  durch  $m_\nu$  ersetzen kann, wenn  $\delta_\nu \sigma_\nu \equiv 1 \pmod{\lambda_\nu}$  ist.

Trägt man (13.9) in (13.8) ein und bezeichnet

$$(13.10) \quad \sum_j \alpha_\nu^{i,j} \varepsilon^{jl\lambda_\nu} = \beta_\nu^{i,l},$$

so kommt

$$(13.11) \quad G_{\nu,0} = \sum_{i=1}^{s-2} \beta_{\nu}^{i,0} F_{i,0} + \sigma_{\nu} m_{\nu} a$$

$$(13.12) \quad G_{\nu,l} = \sum_{i=1}^{s-2} \beta_{\nu}^{i,l} F_{i,l} \quad l = 1, 2, \dots, m_{\nu} - 1.$$

Indem man die Relationen (8.16) in  $H$  additiv schreibt und addiert, erhält man aus (13.6), (13.2) und (13.10)

$$(13.13) \quad \begin{aligned} 0 &= \sum_{\nu=1}^s \sum_{\varrho=0}^{n-1} c_{\nu,\varrho} = \sum_{i,j,\nu,\varrho} \alpha_{\nu}^{i,j} f_{i,j+\varrho} = \\ &= \sum_{i,\varrho} \left( \sum_{\nu,j} \alpha_{\nu}^{i,j} \right) f_{i,\varrho} = \sum_i \left( \sum_{\nu,j} \alpha_{\nu}^{i,j} \right) F_{i,0} = \\ &= \sum_i \left( \sum_{\nu} \beta_{\nu}^{i,0} \right) F_{i,0}. \end{aligned}$$

Denn da  $j$  und  $\varrho$  unabhängig von einander von 0 bis  $n-1$  laufen, kann man im zweiten Index der  $f$  das  $j$  fortlassen, da es nur auf den Restcharakter mod  $n$  ankommt. Nun sind die  $F_{i,0}$  unabhängig, also erhält man

$$(13.14) \quad \sum_{\nu=1}^s \sum_{j=0}^{n-1} \alpha_{\nu}^{i,j} = \sum_{\nu=1}^s \beta_{\nu}^{i,0} = 0 \quad i = 1, 2, \dots, s-2,$$

als Relationen zwischen den Koeffizienten in (13.5) und (13.10).

Aus (13.11), (7.2) und (13.13) folgt nun, dass die Gleichung (12.17) auch im Falle  $q = 0$  ihre Gültigkeit behält.

Das Ergebnis (13.11), (13.12) und (12.17) hätte man direkt erschliessen können, ohne die formale Rechnung durchzuführen. Denn wie in (12.18) nimmt  $G_{\nu,l}$  bei dem

Automorphismus  $J$  den Faktor  $\varepsilon^{l\lambda_\nu}$  an. Also muss sich  $G_{\nu,l}$  linear durch diejenigen Basiselemente (13.3) ausdrücken lassen, die zu dem gleichen Eigenwert gehören, also  $G_{\nu,0}$  durch die  $F_{i,0}$  und  $a$ , und  $G_{\nu,l}$  für  $l > 0$  durch die  $F_{i,l\lambda_\nu}$ . In (13.11) und (13.12) werden überdies nur die Koeffizienten dieser Linearverbindungen auf die in (13.5) eingeführten Koeffizienten  $\alpha_\nu^{i,j}$  und die Zahlen  $\sigma_\nu m_\nu$  zurückgeführt. — Die Gleichung (12.17) bringt zum Ausdruck, dass die  $r$  Randkurven von  $\varphi$  zusammen homolog Null sind. Da  $r-1$  von diesen Randkurven in der Homologiegruppe unabhängig sind, besteht ausser (12.17) keine weitere Relation zwischen den  $g_{\nu,l}$  oder den  $G_{\nu,l}$ .

Man findet daher auch für  $q = 0$  das charakteristische Polynom (12.19) für die Untergruppe  $R$  von  $H$ .

**14. Charakteristisches Polynom und Spurformel.** Man kann die für  $q > 0$  bzw.  $q = 0$  benutzte Homologiebasis (12.5) bzw. (13.1), die zu einer Darstellung von  $J$  in Diagonalform führte, durch Benutzung der  $G_{\nu,l}$  als einen Teil der Basiselemente abändern, ohne dabei die Diagonalform zu verlassen wegen (12.18).

Für  $q > 0$ ,  $s = 0$  ist nichts zu ändern. Für  $q > 0$ ,  $s > 0$  ersetzt man  $C_{\nu,l\lambda_\nu}$  durch  $G_{\nu,l}$  für  $\nu = 1, 2, \dots, s$  und  $l = 0, 1, 2, \dots, m_\nu - 1$ . Für  $l > 0$  ist das wegen (12.16) einfach eine Umbenennung. Für  $l = 0$  ist es wegen (12.15) für diejenigen Werte  $\nu$ , die zu multiplen Randkurven gehören, eine geringfügige Abänderung der  $C_{\nu,0}$ . Entsprechend dem Umstande, dass man wegen (12.10) etwa  $C_{s,0}$  aus (12.5) fortlässt, um eine unabhängige Basis zu haben, lässt man wegen (12.17) auch  $G_{s,0}$  fort.

Für  $q = 0$  hat man  $s > 0$  und folgenden Sachverhalt:

Von den unabhängigen Basiselementen (13.1) erzeugen die  $s-1$  Elemente

$$(14.1) \quad a, F_{1,0}, F_{2,0}, \dots, F_{s-2,0}$$

die zum Eigenwert 1 gehörige Untergruppe  $U_1$  von  $H$ . Dieselbe Gruppe wird nach (13.11) von den  $s-1$  Elementen

$$(14.2) \quad G_{1,0}, G_{2,0}, \dots, G_{s-1,0}$$

erzeugt, indem man wieder wegen (12.17)  $G_{s,0}$  fortlässt und benutzt, dass weitere Relationen zwischen den  $G_{\nu,l}$  nicht bestehen. Die zu einem Eigenwert  $\varepsilon^k$ ,  $0 < k < n$ , gehörige Untergruppe  $U_k$  von  $H$  wird durch die  $s-2$  Basiselemente

$$(14.3) \quad F_{1,k}, F_{2,k}, \dots, F_{s-2,k}$$

erzeugt. Unter den  $G_{\nu,l}$  gehören höchstens  $s-2$  in diese Untergruppe  $U_k$ , da die  $G_{\nu,l}$  unabhängig sind. Man bestätigt das auch so: Für ein solches  $G_{\nu,l}$  muss nach (12.18)  $l\lambda_\nu = k$  sein. Wenn nun mindestens  $s-1$  der  $G_{\nu,l}$  zu  $U_k$  gehörten, so würden mindestens  $s-1$  der  $\lambda_\nu$  in  $n$  und  $k$ , also in dem Teiler  $(n, k) < n$  von  $n$  aufgehen, also mindestens  $s-1$  der  $m_\nu$  würden den Faktor  $\frac{n}{(n, k)} > 1$  enthalten, und das ergibt wegen (4.6) und (4.7) einen Widerspruch. — Man kann also solche Elemente  $G_{\nu,l}$ , für welche  $l\lambda_\nu = k$  ist, mittels (13.12) als Basiselemente für die Untergruppe  $U_k$  von  $H$  einführen und dafür ebenso viele der Basiselemente (14.3) (im Höchsthfall alle) eliminieren.

Diese Einführung der  $G_{\nu,l}$  mit alleiniger Ausnahme von  $G_{s,0}$  in die benutzte Homologiebasis sei nun im Falle  $s > 0$  vorgenommen. Damit ist alles vorbereitet, um den Automorphismus der Homologiegruppe auch im Fall der Ver-

zweigtheit sofort angeben zu können. Ist nämlich die  $\nu$ -te Randkurve von  $M$  durch Ausschneiden eines multiplen Punktes entstanden, so bedeutet ihre Schliessung nach 9 die Hinzufügung der Relationen (9.1), also nach (7.22) der Relationen  $g_{\nu,l} = 1$ ,  $l = 0, 1, 2, \dots, m_\nu - 1$ , zu  $F$ . Dem entspricht in  $H$  die Nullsetzung der  $g_{\nu,l}$  und damit nach (12.14) auch der  $G_{\nu,l}$ . Man hat also in der jetzt benutzten Basis für  $H$  einfach die  $G_{\nu,l}$  für den betreffenden Wert von  $\nu$  fortzulassen und im übrigen nichts zu ändern. Dadurch verliert man sowohl in dem charakteristischen Polynom  $P_H$  wie in  $P_R$  den Faktor  $(x^{m_\nu} - 1)$ . Die Darstellung von  $J$  verbleibt in Diagonalform.

Sind mehrere Ränder von  $M$  zu schliessen, so setze man dies entsprechend fort. Nur für den Fall, dass *alle*  $s$  Ränder von  $M$  geschlossen werden, ist noch folgende Bemerkung hinzuzufügen: Da  $G_{s,0}$  nicht in der Basis vorkommt, hat man beim Schliessen des letzten,  $s$ -ten, Randes nur  $m_s - 1$  Basiselemente zu filgen, und man verliert in den charakteristischen Polynomen nur den Faktor  $\frac{x^{m_s} - 1}{x - 1}$ .

Das Ergebnis ist also folgendes: Es sei  $s = u + v$ , indem die zu  $\nu = 1, 2, \dots, u$  gehörigen Ränder von  $M$  geschlossen werden und die zu  $\nu = u + 1, u + 2, \dots, u + v$  gehörigen Ränder von  $M$  offen bleiben, also von Randkurven von  $\varphi$  überdeckt werden.  $\varphi$  ist also berandet für  $v > 0$  und geschlossen für  $v = 0$ . Dann bekommt man wegen (12.12), (12.13), (12.19) und (13.4) die folgenden charakteristischen Polynome:

$$\left. \begin{aligned} (14.4) \quad P_H &= \frac{(x^n - 1)^{2q + s - 2} (x - 1)}{(x^{m_1} - 1)(x^{m_2} - 1) \dots (x^{m_u} - 1)} \\ (14.5) \quad P_R &= \frac{(x^{m_{u+1}} - 1)(x^{m_{u+2}} - 1) \dots (x^{m_{u+v}} - 1)}{x - 1} \end{aligned} \right\} \text{für berandetes } \varphi$$

$$(14.6) \quad P_H = \frac{(x^n - 1)^{2q+s-2} (x-1)^2}{(x^{m_1} - 1)(x^{m_2} - 1) \dots (x^{m_s} - 1)} \quad \text{für geschlossenes } \varphi.$$

Für geschlossenes  $\varphi$  fällt  $P_R$  fort, da  $R$  verschwindet. Diese Formeln gelten sowohl für  $q > 0$ ,  $s \geq 0$ , wie für  $q = 0$ , in welchem Fall ja  $s \geq 2$  ist. Die am Anfang von 6 vorweg behandelten Fälle a) und b) eines abelschen  $T$  werden mit umfasst: Für a) ist  $q = 0$ ,  $s = 2$  und dabei  $m_1 = m_2 = 1$ . Im Fall  $u = 0$ ,  $v = 2$  des Kreisringes hat  $H$  ein Basiselement, und (14.4) ergibt  $P_H = x - 1$ . Im Fall  $u = 1$ ,  $v = 1$  der Kreisscheibe und im Fall  $u = 2$ ,  $v = 0$  der Kugel verschwindet  $H$  und (14.4) bzw. (14.6) fallen fort. Für b) ist  $q = 1$ ,  $s = 0$ , und  $H$  wird die freie abelsche Gruppe mit 2 Erzeugenden und mit  $F$  identisch.  $J$  ist dabei, ebenso wie  $I$ , der identische Automorphismus, und (14.6) gibt  $P_H = (x - 1)^2$ .

Für die Faktorgruppe  $\frac{H}{R}$  bekommt man in allen Fällen (14.6) als charakteristisches Polynom.

Einem Fixpunkt von  $\tau\varphi$  entspricht der  $m$ -Wert 1. Die Zahl  $\Xi$ , die angibt, wie oft die Zahl 1 unter den Zahlen

$$(14.7) \quad m_1, m_2, \dots, m_u$$

vorkommt, ist also gleich der Anzahl von Fixpunkten von  $\tau\varphi$ . Die Ausdrücke (14.4—6) sind ganze rationale Funktionen. Die Spur  $S_H$  der zu  $J$  gehörigen Transformationsmatrix von  $H$  ist die Summe der Wurzeln von  $P_H$ , also  $1 - \Xi$  in (14.4) und  $2 - \Xi$  in (14.6). Also erhält man für die Anzahl  $\Xi$  von Fixpunkten bei  $\tau\varphi$

$$(14.8) \quad \Xi = 1 - S_H \quad \text{für berandetes } \varphi$$

$$(14.9) \quad \Xi = 2 - S_H \quad \text{für geschlossenes } \varphi.$$

In so einfacher Weise ergibt sich aus (14.4) und (14.6) für den von uns betrachteten Fall der periodischen Transformationen von Flächen die bekannte, sehr allgemeine *Spurformel von ALEXANDER* [16], *LEFSCHETZ* [17] und *HOPF* [18]. Zum Vergleich mit der von *HOPF* benutzten Form, siehe [18] Seite 498 am Schluss von § 2, sei bemerkt, dass die dort auftretenden Symbole in unserem Fall folgende Bedeutung haben: Der Index  $j_q$  jedes einzelnen Fixpunktes von  $\tau\varphi$  ist  $+1$ , die linke Seite der Formel ist also  $\Xi$ . Die Zahl  $n$  gibt die Dimension der Mannigfaltigkeit an und ist hier also 2. Die Spur  $S^0f$  der Homologiegruppe nullter Dimension bei der Abbildung  $f$  ist hier 1, da  $\varphi$  zusammenhängt.  $S^1f$  ist unser  $S_H$ .  $S^2f$  ist 0, falls  $\varphi$  berandet ist, und gleich dem Abbildungsgrad, hier also  $+1$ , wenn  $\varphi$  geschlossen ist. Die Formel ergibt somit (14.8) bzw. (14.9) für berandetes bzw. geschlossenes  $\varphi$ .

Die Anzahl  $\Psi$  der bei  $\tau$  invarianten Randkurven von  $\varphi$  ist gleich der Zahl, die angibt, wie oft die Zahl 1 unter den Zahlen

$$(14.10) \quad m_{u+1}, m_{u+2}, \dots, m_{u+v}$$

vorkommt, also für *berandetes*  $\varphi$  nach (14.5) durch

$$(14.11) \quad \Psi = 1 + S_R$$

gegeben, wenn  $S_R$  die Spur der zu  $J$  gehörigen Transformationsmatrix von  $R$ , also die Summe der Wurzeln von  $P_R$  bedeutet. Aus (14.8) und (14.11) erhält man

$$(14.12) \quad \Xi + \Psi = 2 - S_H + S_R = 2 - \frac{S_H}{R}.$$

Diese Formel zur Bestimmung der Gesamtanzahl von invarianten Punkten und invarianten Rändern gilt sowohl

für berandete wie für geschlossene Flächen. Denn für geschlossene Flächen verschwindet  $R$  und damit  $S_R$ , so dass man auf (14.9) kommt.

**15. Kennzeichnung der Abbildung durch das charakteristische Polynom.** Es soll untersucht werden, wie weitgehende Aussagen man über den topologischen Typus der periodischen Transformation allein auf Grund der Kenntnis des charakteristischen Polynoms  $P_H(x)$  machen kann.

Die im Nenner von (14.4) bzw. (14.6) auftretenden Faktoren  $x^{m_\nu} - 1$  entsprechen multiplen Punkten von  $\tau\varphi$ ; daher ist  $m_\nu < n$ . Diejenigen irreduziblen Faktoren des Zählers, die (im Falle  $2q + s > 2$ ) die primitiven  $n$ -ten Einheitswurzeln zu Wurzeln haben, können daher nicht bei der Division herausfallen.

Ist nun eine ganze rationale Funktion  $P_H(x)$  als charakteristisches Polynom einer periodischen Transformation vorgelegt, so spalte man sie in Faktoren, die im rationalen Zahlkörper irreduzibel sind. Dann haben die einzelnen irreduziblen Faktoren das vollständige System von primitiven Einheitswurzeln eines gewissen Grades zu Wurzeln. Kommt dabei kein Grad  $> 1$  vor, so muss man  $2q + s = 2$ , also einen der vorweg behandelten Fälle a) und b) von 6 haben, für die in 14 das  $P_H$  bestimmt wurde. Man erhält daher folgenden Aufschluss über die Transformation:

- 1) Enthält  $P_H$  überhaupt keinen Faktor, so liegt der Fall der Kreisscheibe oder der Kugel vor.
- 2) Für  $P_H = x - 1$  hat man den Fall des Kreisringes.
- 3) Für  $P_H = (x - 1)^2$  hat man den Fall des Torus ohne multiple Punkte.



In diesen Fällen lässt sich nichts über die Ordnung  $n$  der Transformation aus dem Polynom  $P_H$  erschliessen.

In allen anderen Fällen hat man  $2q + s > 2$ , und es kommen Grade  $> 1$  von Einheitswurzeln vor. Dann ergibt sich zunächst die Ordnung  $n$  der Transformation als der höchste vorkommende Grad von Einheitswurzeln. Die Multiplizität, mit der das irreduzible Polynom, welches die primitiven  $n$ -ten Einheitswurzeln zu Wurzeln hat, in  $P_H$  vorkommt, ist  $2q + s - 2$ , woraus sich die Anzahl  $2q + s$  bestimmt.

Nun bilde man das Polynom

$$(15.1) \quad H(x) = \frac{(x^n - 1)^{2q+s-2} (x-1)^2}{P_H(x)}.$$

Wegen (14.4) und (14.6) geht der Nenner von (15.1) jedenfalls im Zähler auf, und (15.1) lässt sich auf eine und nur eine Weise in die Form

$$(15.2) \quad H(x) = (x^{m_1} - 1)(x^{m_2} - 1) \dots (x^{m_w} - 1)$$

bringen. Dabei ist  $w \leq s \leq 2q + s$ . Wir unterscheiden nun zwischen zwei Fällen:

I. Ist keine der Zahlen  $m_1, m_2, \dots, m_w$  gleich 1, so lässt sich  $P_H$  nicht auf die Form (14.4) bringen, also muss  $\varphi$  notwendig geschlossen sein. Dann ist  $w = s$ , und da man oben  $2q + s$  fand, kennt man auch  $q$ . Aus dem nach (15.2) bekannten Zahlensystem  $m_1, m_2, \dots, m_s$  findet man  $r$  nach (3.1), und danach ergibt sich  $p$  aus (3.2). Im Falle I kann man also die Ordnung der Transformation, das Geschlecht der geschlossenen Fläche  $\varphi$  und ihrer Modulfläche  $M$  und die Verzweigung von  $\varphi$  über  $M$  vollständig bestimmen. Zur Charakterisierung des topologischen Typus der Transformation fehlen nur (in dem Unterfalle  $s > 0$ ) die

zu den einzelnen Verzweigungsstellen gehörigen Drehzahlen. Es ist von vorne herein klar, dass diese sich nicht aus  $P_H$  ergeben können. Denn ersetzt man  $\tau$  durch  $\tau^\rho$ , wo  $(\rho, n) = 1$  ist, so ändern sich die Drehzahlen, während sich die Wurzeln der einzelnen irreduziblen Faktoren von  $P_H$  nur unter einander vertauschen, so dass  $P_H$  ungeändert bleibt. — Hervorgehoben sei der Spezialfall: Der topologische Typus einer periodischen Transformation einer geschlossenen Fläche ohne multiple Punkte ergibt sich eindeutig aus dem zugehörigen charakteristischen Polynom (abgesehen davon, dass in dem obigen Spezialfall 3) des Torus die Ordnung unbestimmt bleibt).

II. Ist mindestens eine der Zahlen  $m$  die 1, etwa  $m_w = 1$ , so besteht dann und nur dann die Möglichkeit eines geschlossenen  $\varphi$ , wenn  $w$  dieselbe Parität hat, wie die oben bestimmte Zahl  $2q + s$ . Ist das der Fall, so setzt man wieder  $w = s$  und verfährt wie unter I. Andererseits kann man im Fall II immer  $P_H$  auf die Form (14.4) mit  $w = u + 1$  bringen (und man muss das tun, wenn  $w$  und  $2q + s$  ungleiche Parität haben) und kommt damit auf den Fall einer berandeten Fläche. Dann kann man, mit den Bezeichnungen von 14,  $(2q + s) - u = 2q + v$  berechnen. Nun bietet aber die Kenntnis von  $P_H$  keine Handhabe mehr, um die Verteilung der Einheiten dieser Zahl auf die Anzahlen  $q$  und  $v$  zu bestimmen, sodass in diesem Fall eine grössere Anzahl von topologischen Realisierungen zu dem gegebenen  $P_H$  gehören.

Wenn man indessen sowohl  $P_H$  wie  $P_R$  kennt, kann man  $v$  bestimmen und damit die Bestimmung der topologischen Realisierung ebensoweit treiben wie im Fall I, in dem  $P_R$  fortfiel.

Für geschlossene Flächen enthält der Nenner von  $P_H$  in (14.6) den Faktor  $(x-1)^s$ . Die Multiplizität der Wurzel 1 in  $P_H$  ist also eine gerade Zahl, und zwar  $2q$ , sodass sich das Geschlecht der Modulfläche unmittelbar aus  $P_H$  ergibt.

Man bemerkt, dass es, abgesehen von den Sonderfällen 1), 2) und 3), die  $n$  unbestimmt lassen, zu einem gegebenen charakteristischen Polynom  $P_H$  nur endlich viele topologisch nicht äquivalente Realisierungen durch periodische Transformationen gibt. Das verhältnismässig grobe Hilfsmittel der Homologiegruppe gestattet also eine so weitgehende Typeneinteilung der periodischen Transformationen von Flächen, dass das feinere Werkzeug der Fundamentalgruppe darüber hinaus nur die Unterteilung jedes Typus in endlich viele topologisch nicht äquivalente Untertypen liefert.

---

## LITERATURVERZEICHNIS

- [1] L. E. J. BROUWER: Über topologische Involutionen. Amsterdam Proceedings, Vol. 21, 1919.
- [2] — Über die periodischen Transformationen der Kugel. Math. Annalen, Bd. 80, 1919.
- [3] — Aufzählung der periodischen Transformationen des Torus. Amsterdam Proceedings, Vol. 21, 1919.
- [4] — Énumération des groupes finis de transformations topologiques du tore. Comptes rendus (Paris), t. 168 pg. 845, 1919.
- [5] B. v. KERÉKJÁRTÓ: Über die periodischen Transformationen der Kreisscheibe und der Kugelfläche. Math. Annalen, Bd. 80, 1919.
- [6] — Vorlesungen über Topologie. Berlin, Springer, 1923.
- [7] — Über die endlichen topologischen Gruppen der Kugelfläche. Amsterdam Proceedings, Vol. 22, 1919.
- [8] W. SCHERRER: Über topologische Involutionen. Vierteljahrsschrift der Naturforschenden Gesellschaft in Zürich, Bd. 70, 1925.
- [9] — Über periodische Transformationen von Flächen. Ibidem.
- [10] — Zur Theorie der endlichen Gruppen topologischer Abbildungen von geschlossenen Flächen in sich. Commentarii Math. Helv., 1929.
- [11] F. STEIGER: Die maximalen Ordnungen periodischer topologischer Abbildungen geschlossener Flächen in sich. Commentarii Math. Helv., Vol. 8, 1935.
- [12] J. NIELSEN: Topologischer Beweis eines Satzes von Wiman. Matematisk Tidsskrift B, 1936.
- [13] — Äkvivalensproblemet for periodiske Transformationer. Matematisk Tidsskrift B, 1937.
- [14] — Om regulære Riemannske Flader. Matematisk Tidsskrift B, 1932.

- [15] J. NIELSEN: Om Regning med ikke-kommutative Faktorer og dens Anvendelse i Gruppeteorien. Matematisk Tidsskrift B, 1921.
  - [16] J. W. ALEXANDER: Invariant points of a surface transformation of given class. Transactions of the American Math. Soc., Vol. 25, 1923.
  - [17] S. LEFSCHETZ: Intersections and transformations of complexes and manifolds. Transact. Amer. Math. Soc., Vol. 28, 1926.
  - [18] H. HOPF: Über die algebraische Anzahl von Fixpunkten. Math. Zeitschrift, Bd. 29, 1929.
-



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

---

FLÄCHEN ZWEITER ORDNUNG UND  
GESCHARTE KOLLINEATIONEN

VON

DAVID FOG



KØBENHAVN  
LEVIN & MUNKSGAARD  
EJNAR MUNKSGAARD

1937

Printed in Denmark.  
Bianco Lunos Bogtrykkeri A/S.



Es handelt sich in dieser Abhandlung um die Bestimmung der gescharten Kollineationen des Raumes, durch welche eine gegebene, nicht ausgeartete Fläche zweiter Ordnung in eine gegebene andere übergeführt wird.

Diese Aufgabe vergleichen wir zunächst mit der ebenen Aufgabe, diejenigen Homologien zu finden, bei welchen zwei Kegelschnitte einer und derselben Ebene einander entsprechen. Die letztgenannte Aufgabe hat bekanntlich eine endliche Anzahl von Lösungen (bei allgemeiner Lage der beiden Kurven genau 12), in Übereinstimmung damit, dass sowohl  $\infty^5$  Kegelschnitte als auch  $\infty^5$  Homologien in der Ebene vorhanden sind. Nun gibt es auch im Raume »gleich viele« Flächen zweiter Ordnung und gescharte Kollineationen, nämlich  $\infty^9$ ; man könnte deshalb erwarten, dass die oben angeführte räumliche Aufgabe ebenfalls im allgemeinen eine endliche Anzahl von Lösungen besitzt. Dies ist jedoch, wie wir sehen werden, nicht der Fall. Zwei Flächen zweiter Ordnung allgemeiner Lage entsprechen sich bei keiner gescharten Kollineation. Der scheinbare Widerspruch löst sich, wie aus dem folgenden hervorgehen wird, dadurch, dass unendlich viele gescharte Kollineationen existieren, die eine Fläche zweiter Ordnung in eine zweite überführen, sobald es eine solche Kollineation gibt.

Über den betrachteten Gegenstand ist dem Verfasser nur eine Arbeit von K. ROHN<sup>1</sup> bekannt. Es werden dort die har-

<sup>1</sup> K. ROHN: Geschart-involutorische Lage zweier Flächen 2. Grades. Leipz. Ber. 66 (1914), S. 60.

monischen gescharten Kollineationen bestimmt, welche zwei Flächen zweiter Ordnung vertauschen.

### § 1. Die drei Typen von Achsen.

Eine gescharte Kollineation ist eine Kollineation, bei der zwei windschiefe Geraden, die Achsen der Transformation, punktweise und ebenenweise festbleiben. Diese Achsen sind im allgemeinen getrennt, können aber auch »konsekutiv« sein und einen Streifen einer Regelfläche bestimmen. Man könnte die Achsen auch als zwei sich schneidende Geraden annehmen; dann würde aber die Transformation eine parabolische Homologie sein (d. h. eine Homologie, deren Zentrum in der Homologieebene enthalten ist). Solche Homologien wollen wir hier nicht aufsuchen, sondern nur insofern mitnehmen, als sie in natürlicher Weise mit den eigentlichen gescharten Kollineationen zusammen auftreten.

Sind bei einer gescharten Kollineation  $(M, M_1)$  ein Paar von entsprechenden Punkten,  $E$  und  $F$  die Schnittpunkte der Geraden  $MM_1$  mit den beiden Achsen, dann ist das Doppelverhältnis  $(EFMM_1)$  nur von der Transformation, nicht von dem Punktepaare  $(M, M_1)$  abhängig.

Die zwei Flächen zweiter Ordnung seien  $\varphi$  und  $\varphi_1$ ; wir suchen die gescharten Kollineationen, welche  $\varphi$  in  $\varphi_1$  überführen. Für eine Achse einer solchen Transformation muss gelten, dass jeder Punkt, welchen sie mit der einen Fläche gemein hat, auch der anderen angehört und dass jede durch sie gehende Tangentialebene an die eine Fläche auch die andere berührt. Es erweisen sich daher die folgenden drei Typen von Achsen als möglich:

I. Die Achse schneidet die Flächen in denselben zwei Punkten und liegt in zwei gemeinsamen Tangentialebenen.

II. Die Achse ist Tangente in einem Berührungspunkt der beiden Flächen.

III. Die Achse ist eine gemeinsame Erzeugende der beiden Flächen; diese müssen sich dann in allen Punkten dieser Erzeugenden berühren.

Wir zeigen nun:

Jede Gerade, welche die in I genannten Bedingungen erfüllt, ist Achse zweier gescharten Kollineationen, welche  $\varphi$  in  $\varphi_1$  überführen.

Die Achse sei  $a$ , ihre Schnittpunkte mit den Flächen  $A$  und  $B$ ; ferner möge die eine durch  $a$  gehende Tangentialebene  $\varphi$  und  $\varphi_1$  in  $P$  bzw.  $P_1$ , die andere in  $Q$  bzw.  $Q_1$  berühren. Die

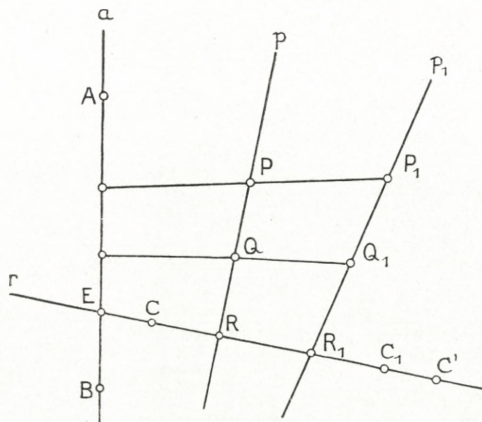


Fig. 1.

Gerade  $p = PQ$  bzw.  $p_1 = P_1Q_1$  ist dann die Polare von  $a$  in bezug auf  $\varphi$  bzw.  $\varphi_1$ ; sowohl  $p$  als auch  $p_1$  sind zu  $a$  windschief. Die folgenden zwei Fälle sind nun zu betrachten:

1)  $p$  und  $p_1$  sind zu einander windschief (Fig. 1). Die drei Geraden  $a, p, p_1$  bestimmen dann eine Fläche zweiter Ordnung, auf welcher die andere Achse zu suchen ist. Diese Fläche geht bei der Transformation in der Weise in sich selbst über, dass jede Erzeugende des von  $a, p, p_1$  verschiedenen Systems in sich selbst transformiert wird. Es sei  $r$  eine solche von  $PP_1$  und  $QQ_1$  verschiedene Gerade.

Ferner sei angenommen, sie gehe nicht durch  $A$  oder  $B$ ; sie ist dann weder Tangente an  $\varphi$  noch an  $\varphi_1$ . Diese Gerade  $r$  schneide die Fläche  $\varphi$  in  $C$  und  $C'$ , die Fläche  $\varphi_1$  in  $C_1$  und einem weiteren Punkt und die Geraden  $a, p, p_1$  in  $E, R$  bzw.  $R_1$ .

Die durch die gesuchte Transformation auf  $r$  erzeugte Projektivität ist nun dadurch festgelegt, dass sie den Punkt  $E$

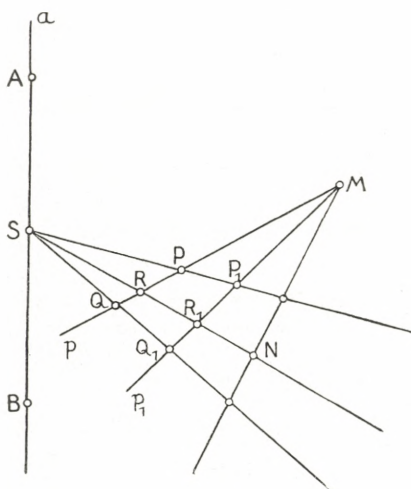


Fig. 2.

als Fixpunkt hat,  $R$  in  $R_1$  und  $C$  (oder  $C'$ ) in  $C_1$  überführt. Dadurch ist auch die andere Achse bestimmt; sie geht durch den zweiten Fixpunkt der Projektivität und ist ausserdem eine Erzeugende der genannten Fläche zweiter Ordnung von demselben System wie  $a, p$  und  $p_1$ . Wir erhalten auf diese Weise zwei gescharte Kollineationen; diese führen  $A, B, P$

und  $Q$  in bzw.  $A, B, P_1$  und  $Q_1$  sowie  $C$  (oder  $C'$ ) in  $C_1$  über, daher auch  $\varphi$  in  $\varphi_1$ . Die eine dieser Kollineationen kann konsekutive Achsen haben.

2)  $p$  und  $p_1$  schneiden einander in  $M$  (Fig. 2).  $PP_1$  und  $QQ_1$  schneiden sich dann in einem Punkt  $S$  von  $a$ . Eine beliebige durch  $a$  gehende Ebene, welche nicht Tangentialebene an  $\varphi$  und  $\varphi_1$  ist, schneidet diese Flächen in zwei Kegelschnitten,  $\varkappa$  und  $\varkappa_1$ . Die Pole  $R$  und  $R_1$  von  $a$  in bezug auf diese Kegelschnitte liegen auf  $p$  bzw.  $p_1$ . Die Kurven  $\varkappa$  und  $\varkappa_1$  sind auf zwei Weisen homolog mit  $a$  als Achse; die

entsprechenden Zentren seien  $N$  und  $N'$ ; sie liegen auf der Geraden  $SRR_1$ . Die gescharte Kollineation mit den Achsen  $a$  und  $MN$  (oder  $MN'$ ), welche  $p$  in  $p_1$  überführt, führt dann  $z$  in  $z_1$  und daher auch  $\varphi$  in  $\varphi_1$  über. Es gibt demnach auch in diesem Fall zwei gescharte Kollineationen der gesuchten Art. Doch kann insbesondere  $N$  (oder  $N'$ ) in  $S$  fallen; die eine der Kollineationen wird dann eine parabolische Homologie.

Der Beweis gilt bis auf unwesentliche Änderungen, wenn  $p$  mit  $p_1$  zusammenfällt; die beiden gescharten Kollineationen haben dann dieselben Achsen und nur verschiedene charakteristische Doppelverhältnisse.

Dem Achsentypus III entsprechend erhält man:

Wenn sich  $\varphi$  und  $\varphi_1$  längs einer Erzeugenden berühren, gibt es  $\infty^2$  gescharte Kollineationen mit dieser Erzeugenden als Achse, welche  $\varphi$  in  $\varphi_1$  überführen.

$\varphi$  und  $\varphi_1$  haben nämlich ausser der Berührungserzeugenden  $a$  noch zwei andere (evtl. konsekutive) gemeinsame Erzeugende,  $b$  und  $c$ . Benutzt man, dass  $b$  und  $c$  in Verbindung mit den beiden konsekutiven Erzeugenden  $a$  und noch einem Punkt von  $\varphi_1$  diese Fläche eindeutig festlegen, so sieht man, dass die andere Achse unter den Transversalen von  $b$  und  $c$  beliebig gewählt werden kann.

Im Gegensatz hierzu ist eine Gerade, welche den Bedingungen II genügt, im allgemeinen keine Achse einer gescharten Kollineation, die  $\varphi$  in  $\varphi_1$  überführt. Denn es gibt  $\infty^7$  Flächen zweiter Ordnung, welche eine gegebene Gerade in einem gegebenen ihrer Punkte berühren, aber nur  $\infty^5$  gescharte Kollineationen, welche die genannte Gerade als eine Achse haben.

Um die Achsen der eventuellen gescharten Kollineationen, die  $\varphi$  in  $\varphi_1$  überführen, näher zu bestimmen, stellen wir die folgende Überlegung an:

Es seien  $a$  und  $b$  ein Paar zusammengehöriger Achsen. Sie werden von (mindestens) einem Paar von Geraden,  $\alpha$  und  $\beta$ , geschnitten, die polar in bezug auf  $\varphi$  sind. Da  $a$  und  $\beta$  bei der gescharten Kollineation festbleiben, sind sie auch polar in bezug auf  $\varphi_1$ . Die Achsen  $a$  und  $b$  werden also von einem Paar gemeinsamer Polargeraden  $\alpha$  und  $\beta$  der beiden Flächen geschnitten. (Dies gilt auch, wenn  $a$  und  $b$  konsekutive Geraden sind.) Es erweisen sich drei verschiedene Lagen der Geraden  $\alpha$  und  $\beta$  als möglich:

1)  $\alpha$  und  $\beta$  sind zu einander windschief.

2)  $\alpha$  und  $\beta$  gehören derselben Ebene an; diese ist dann Tangentialebene in einem Berührungspunkt der Flächen. Eine der Achsen geht durch den Berührungspunkt, die andere liegt in der Tangentialebene. Diese Ebene geht in sich selbst über, und die Transformation in ihr ist eine Homologie mit dem genannten Berührungspunkt als Zentrum; die Ebene schneidet demnach  $\varphi$  und  $\varphi_1$  in denselben zwei Erzeugenden.  $\varphi$  und  $\varphi_1$  müssen also, wenn dieser Fall vorkommen soll, zwei einander schneidende Erzeugende gemein haben.

3)  $\alpha$  und  $\beta$  fallen zusammen;  $\varphi$  und  $\varphi_1$  haben dann eine Erzeugende gemein.

## § 2. Die gegenseitige Lage der Flächen.

Wir wollen im folgenden — von einem letzten § 6 abgesehen — die Untersuchung auf diejenigen Fälle beschränken, wo  $\varphi$  und  $\varphi_1$  (mindestens) ein gemeinsames Polar-

tetraeder haben. Es liegen dann 4 Möglichkeiten vor, die dem folgenden Aussehen der Schnittkurve entsprechen:

A. Eine Raumkurve 4. Ordnung ohne Doppelpunkte oder Spitze.

B. Zwei Kegelschnitte (in verschiedenen Ebenen) mit zwei gemeinsamen Punkten.

C. Vier Geraden, die ein windschiefes Viereck bilden.

D. Ein doppeltzählender Kegelschnitt.

Ehe wir die eigentliche Untersuchung weiterführen, wollen wir einige Einzelheiten über die genannten Lagen, insbesondere über die gemeinsamen Polartetraeder der beiden Flächen angeben.

Lage A. Die Flächen haben keinen Berührungspunkt; sie haben 1 gemeinsames Polartetraeder.

Lage B. Die Flächen berühren einander in 2 Punkten, etwa  $P$  und  $Q$  (die Schnittpunkte der Kegelschnitte); ihre Verbindungsgerade  $PQ = l$  hat dieselbe Polare  $l'$  in bezug auf beide Flächen. Es gibt  $\infty^1$  gemeinsame Polartetraeder; diese haben alle ein Paar gegenüberliegender Kanten auf  $l$  und  $l'$ . Die auf  $l'$  liegende Eckpunkte sind allen Tetraedern gemein; auf  $l$  können als Eckpunkte die Punkte jedes Paares auftreten, das von  $(P, Q)$  harmonisch getrennt wird.

Lage C. Die Flächen berühren einander in 4 Punkten (die Eckpunkte des Vierecks). Sie haben  $\infty^2$  gemeinsame Polartetraeder, die die Diagonalen des Vierecks als gemeinsame gegenüberliegende Kanten haben. Als Eckpunkte können auf jeder Diagonale die Punkte eines beliebigen Paares auftreten, das von dem entsprechenden Paar von Berührungspunkten harmonisch getrennt wird.

Lage D. Die Flächen berühren einander in allen Punkten des gemeinsamen Kegelschnittes; die Ebene des Kegelschnittes hat denselben Pol,  $O$ , in bezug auf beide Flächen.

Es gibt hier  $\infty^3$  gemeinsame Polartetraeder; diese haben alle einen Eckpunkt in  $O$ , und die drei anderen liegen in der Ebene des Kegelschnittes und können als Eckpunkte eines beliebigen Polardreieckes dieser Kurve gewählt werden.

Gegenüberliegende Kanten eines gemeinsamen Polartetraeders sind immer Polargeraden in bezug auf beide Flächen. Umgekehrt hat man bei den hier betrachteten 4 Lagen, dass jedes Paar windschiefer gemeinsamer Polaren der beiden Flächen Träger eines Paares von gegenüberliegenden Kanten eines gemeinsamen Polartetraeders sind.<sup>1</sup>

Wählt man ein gemeinsames Polartetraeder als Koordinatentetraeder, so können  $\varphi$  und  $\varphi_1$  durch die Gleichungen

$$(1) \quad x_1^2 + x_2^2 + x_3^2 + x_4^2 = 0 \quad \text{und} \quad \lambda_1 x_1^2 + \lambda_2 x_2^2 + \lambda_3 x_3^2 + \lambda_4 x_4^2 = 0$$

dargestellt werden. In den Fällen, wo mehrere Polartetraeder vorhanden sind, kann man für alle solchen Polartetraeder die Gleichungen (1) mit denselben  $\lambda$  benutzen. Der Lage A entspricht, dass alle vier  $\lambda$  verschieden sind; B erhält man, wenn 2 (und nur 2) einander gleich sind. C entspricht, dass sie paarweise einander gleich sind, z. B.  $\lambda_1 = \lambda_2$ ,  $\lambda_3 = \lambda_4$ , und D ergibt sich, wenn 3 (und nur 3)  $\lambda$  einander gleich sind.

### § 3. Die Bestimmung der Achsen.

Wir kehren nun zur Untersuchung der gescharten Kollineationen, die  $\varphi$  in  $\varphi_1$  überführen, zurück, und beginnen mit den Lagen A, B und D, da bei C besondere Verhält-

<sup>1</sup> Dagegen kann es bei anderen Lagen, wo kein gemeinsames Polartetraeder vorhanden ist, dennoch sehr wohl ein Paar windschiefer gemeinsamer Polaren geben.



nisse vorliegen; dieser Fall wird daher am besten für sich allein behandelt (§ 5).

Bei den drei angeführten Lagen von  $\varphi$  und  $\varphi_1$  sieht man, dass die Möglichkeiten 2) und 3) auf Seite 8 sich nicht realisieren lassen, so dass ein Paar zusammengehöriger Achsen notwendig ein Paar von windschiefen gemeinsamen Polaren der Flächen, also ein Paar gegenüberliegender Kanten eines gemeinsamen Polartetraeders schneiden müssen.

Wir suchen in diesem Paragraphen die Achsen des Typus I (siehe Seite 4). Die Gleichungen der Flächen in Punktkoordinaten seien die oben angeführten (1); in Ebenenkoordinaten lauten sie

$$(2) \quad u_1^2 + u_2^2 + u_3^2 + u_4^2 = 0 \quad \text{und} \quad \frac{u_1^2}{\lambda_1} + \frac{u_2^2}{\lambda_2} + \frac{u_3^2}{\lambda_3} + \frac{u_4^2}{\lambda_4} = 0.$$

Als eventuelle Achse betrachten wir eine Gerade, die in Punktkoordinaten durch die Gleichungen

$$(3) \quad x_1 = \alpha x_2, \quad x_3 = \beta x_4,$$

in Ebenenkoordinaten durch

$$(4) \quad \alpha u_1 + u_2 = 0, \quad \beta u_3 + u_4 = 0$$

dargestellt ist. Ihre Schnittpunkte mit den Flächen werden durch

$$\begin{aligned} (\alpha^2 + 1)x_2^2 + (\beta^2 + 1)x_4^2 &= 0, \\ (\alpha^2\lambda_1 + \lambda_2)x_2^2 + (\beta^2\lambda_3 + \lambda_4)x_4^2 &= 0 \end{aligned}$$

bestimmt. Die durch die Gerade gehenden Tangentialebenen ergeben sich auf duale Weise aus

$$\begin{aligned} (\alpha^2 + 1)u_1^2 + (\beta^2 + 1)u_3^2 &= 0, \\ \left(\frac{\alpha^2}{\lambda_2} + \frac{1}{\lambda_1}\right)u_1^2 + \left(\frac{\beta^2}{\lambda_4} + \frac{1}{\lambda_3}\right)u_3^2 &= 0. \end{aligned}$$

Soll nun die Gerade sowohl durch zwei (verschiedene) gemeinsame Punkte der Flächen gehen als auch in zwei (verschiedenen) gemeinsamen Tangentialebenen liegen, so erfordert dies das Bestehen der Gleichungen

$$(5) \quad \begin{cases} (\alpha^2 \lambda_1 + \lambda_2) (\beta^2 + 1) = (\beta^2 \lambda_3 + \lambda_4) (\alpha^2 + 1), \\ \lambda_3 \lambda_4 (\alpha^2 \lambda_1 + \lambda_2) (\beta^2 + 1) = \lambda_1 \lambda_2 (\beta^2 \lambda_3 + \lambda_4) (\alpha^2 + 1) \end{cases}$$

in der Weise, dass nicht beide Seiten null werden. Also muss

$$(6) \quad \lambda_1 \lambda_2 = \lambda_3 \lambda_4$$

sein.<sup>1</sup> Ist umgekehrt diese Gleichung erfüllt, so reduzieren sich die Gleichungen (5) auf eine,

$$(7) \quad (\alpha^2 \lambda_1 + \lambda_2) (\beta^2 + 1) = (\beta^2 \lambda_3 + \lambda_4) (\alpha^2 + 1),$$

und jedes Wertepaar  $(\alpha, \beta)$ , das (7) befriedigt, ohne beide Seiten zum Verschwinden zu bringen, erweist sich leicht als eine Gerade vom Typus I, also als eine Achse. Es gibt also unter der Bedingung (6)  $\infty^1$  gescharte Kollineationen, die  $\varphi$  in  $\varphi_1$  überführen. Die zugehörigen Achsen sind Erzeugende einer Regelfläche mit der Gleichung

$$(8) \quad (\lambda_1 x_1^2 + \lambda_2 x_2^2) (x_3^2 + x_4^2) = (\lambda_3 x_3^2 + \lambda_4 x_4^2) (x_1^2 + x_2^2)$$

oder in Ebenenkoordinaten

$$(9) \quad \left( \frac{u_1^2}{\lambda_1} + \frac{u_2^2}{\lambda_2} \right) (u_3^2 + u_4^2) = \left( \frac{u_3^2}{\lambda_3} + \frac{u_4^2}{\lambda_4} \right) (u_1^2 + u_2^2);$$

sie ist von 4. Ordnung und 4. Klasse und besitzt zwei windschiefe Doppelgeraden.<sup>2</sup>

<sup>1</sup> Die Relation (6) als Ausdruck der Bedingung I (Seite 4) findet sich bei ROHN, S. 73, ohne Verbindung mit dem hier betrachteten Problem der allgemeinen gescharten Kollineationen.

<sup>2</sup> Vgl. Encykl. d. math. Wiss. III, 2, 2 b, S. 1753.

Wir können der Gleichung (6) eine einfache geometrische Deutung geben. Wir bemerken zunächst, dass es 8 Kollineationen gibt, die die Eckpunkte des Koordinatentetraeders zu Fixpunkten haben, und die ausserdem  $\varphi$  in  $\varphi_1$  überführen. Sie werden durch

$$(10) \quad \begin{aligned} x_1 &= \pm \sqrt{\lambda_1} x'_1, & x_2 &= \pm \sqrt{\lambda_2} x'_2, & x_3 &= \pm \sqrt{\lambda_3} x'_3, \\ x_4 &= \pm \sqrt{\lambda_4} x'_4 \end{aligned}$$

dargestellt.

Betrachten wir nun die Kollineation

$$(11) \quad x_1 = x'_1, \quad x_2 = x'_2, \quad x_3 = m x'_3, \quad x_4 = \frac{1}{m} x'_4.$$

Sie lässt die Gerade  $x_3 = 0$ ,  $x_4 = 0$  punktweise und die Gerade  $x_1 = 0$ ,  $x_2 = 0$  ebenenweise fest. Sie kann als Produkt zweier harmonischen Homologien aufgefasst werden, deren Homologieebenen durch die erste der genannten Geraden gehen und deren Zentren auf der zweiten liegen, und mag deshalb als eine »projektive Drehung« bezeichnet werden. Umgekehrt hat jede projektive Drehung der betrachteten Art eine Darstellung der Form (11). Eine andere projektive Drehung, wo die Achsen ihre Rolle vertauscht haben, sei

$$x_1 = n x'_1, \quad x_2 = \frac{1}{n} x'_2, \quad x_3 = x'_3, \quad x_4 = x'_4.$$

Das Produkt dieser beiden Drehungen (in beliebiger Reihenfolge) ist die Kollineation

$$x_1 = n x'_1, \quad x_2 = \frac{1}{n} x'_2, \quad x_3 = m x'_3, \quad x_4 = \frac{1}{m} x'_4;$$

sie möge eine »projektive Schraubung« um die genannten

Achsen heissen. Wir fragen nun, wann die allgemeine Kollineation

$$x_1 = \mu_1 x'_1, \quad x_2 = \mu_2 x'_2, \quad x_3 = \mu_3 x'_3, \quad x_4 = \mu_4 x'_4$$

eine projektive Schraubung um die genannten Achsen ist. Als notwendig und hinreichend erweist sich

$$\mu_1 \mu_2 = \mu_3 \mu_4.$$

Wir sehen daher, dass das Erfülltsein von (6) bedeutet, dass 4 von den oben genannten 8 Kollineationen, die  $\varphi$  in  $\varphi_1$  überführen, projektive Schraubungen um das oben erwähnte Paar von gegenüberliegenden Kanten des Koordinatentetraeders sind.

Wir wollen nun die einzelnen Lagen A, B und D etwas näher betrachten.

Lage A. Im allgemeinen gibt es keine Lösungen; nur wenn (6) oder eine damit analoge Gleichung erfüllt ist, gibt es eine Schar von  $\infty^1$  Achsen. Denkt man sich die Flächen als konzentrische Ellipsoide mit gemeinsamen Symmetrieebenen, so dass die Gleichungen in rechtwinkligen Koordinaten

$$(12) \quad \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1 \quad \text{und} \quad \frac{x^2}{a_1^2} + \frac{y^2}{b_1^2} + \frac{z^2}{c_1^2} = 1$$

geschrieben werden können, so ist (6) mit

$$\frac{ab}{c} = \frac{a_1 b_1}{c_1}$$

gleichbedeutend.

Lage B. Es sei z. B.  $\lambda_3 = \lambda_4$ . Die Schnittkurve der Flächen zerfällt in zwei Kegelschnitte, die in den Ebenen

$$(13) \quad \sqrt{\lambda_1 - \lambda_3} x_1 \pm \sqrt{\lambda_3 - \lambda_2} x_2 = 0$$

liegen und sich in den Punkten  $(0, 0, 1, \pm i)$  schneiden. Wie oben gibt es im allgemeinen keine Lösungen; die Gleichung (6) erhält hier die Form

$$(14) \quad \lambda_1 \lambda_2 = \lambda_3^2,$$

und die Bedingung (7) für  $a$  und  $\beta$  wird nach Division durch  $\beta^2 + 1$ :

$$a^2(\lambda_1 - \lambda_3) + \lambda_2 - \lambda_3 = 0.$$

Hieraus erhält man

$$(15) \quad \alpha = \pm \frac{\sqrt{\lambda_3 - \lambda_2}}{\sqrt{\lambda_1 - \lambda_3}}, \quad \beta \text{ beliebig,}$$

also die Achsen

$$(16) \quad \sqrt{\lambda_1 - \lambda_3} x_1 \pm \sqrt{\lambda_3 - \lambda_2} x_2 = 0, \quad x_3 = kx_4.$$

Sie bilden zwei Geradenbüschel, die in den Ebenen (13) der Kegelschnitte liegen und ihre Zentren in den Punkten haben, wo diese Ebenen die Polare der gemeinsamen Sehne der Kegelschnitte schneiden.

Sind  $\varphi$  und  $\varphi_1$  durch die Gleichungen (12) dargestellt, so wird

$$c = c_1, \quad ab = a_1 b_1.$$

Hieraus folgt, dass die Flächen von jeder zur  $xy$ -Ebene parallelen Ebene in zwei inhaltsgleichen Ellipsen geschnitten werden, und sie entsprechen sich demnach in zwei involutorischen Affinitäten, deren Homologieebenen die Kegelschnittsebenen sind. Man entnimmt hieraus, dass die Bedingungen

$$(17) \quad \lambda_1 \lambda_2 = \lambda_3 \lambda_4 \text{ und } \lambda_3 = \lambda_4$$

damit gleichbedeutend sind, dass die Flächen  $\varphi$  und  $\varphi_1$  mit zwei Kugeln von demselben Radius (und mit nicht-isotroper Verbindungsgerade der beiden Zentren) projektiv äquivalent sind.

Die Bedingungen (17) können auch so gedeutet werden, dass 4 von den 8 Kollineationen, die die Eckpunkte des Koordinatentetraeders zu Fixpunkten haben und  $\varphi$  in  $\varphi_1$  überführen, projektive Drehungen sind, bei denen die Schnittlinie der Kegelschnittsebenen punktweise und ihre Polare ebenenweise festbleiben.

Es soll kurz erwähnt werden, dass man, statt die Bedingung  $\lambda_3 = \lambda_4$  mit (6) zu kombinieren, diese Bedingung auch mit einer zu (6) analogen Gleichung zusammen nehmen könnte, z. B. mit

$$\lambda_1 \lambda_3 = \lambda_2 \lambda_4.$$

Hieraus würde indessen  $\lambda_1 = \lambda_2$  folgen, und die Flächen hätten dann die Lage C. Dies werden wir später, in § 5, näher erörtern.

Wir bemerken noch, dass wir durch die obige Methode alle Achsen vom Typus I gefunden haben, obwohl die Flächen  $\infty^1$  gemeinsame Polartetraeder haben; denn die gefundenen Achsen sind in Beziehung zu den Flächen selbst, nicht nur zu dem gewählten Koordinatentetraeder, charakterisiert worden.

Lage D. Es gibt hier gar keine Achsen, weil (6) in Verbindung mit z. B.  $\lambda_1 = \lambda_2 = \lambda_3$  das vollständige Zusammenfallen der Flächen zur Folge haben würde.

## § 4. Zusammengehörige Achsen.

Wir haben in § 3 die brauchbaren Achsen bestimmt ohne Rücksicht darauf, wie sie in den gescharten Kollineationen zusammengehören. Um auch diese Frage zu beantworten, greifen wir gewissermassen das ganze Problem nochmals von vorn an, indem wir uns nur auf die §§ 1—2 stützen, und benutzen eine rein analytische Methode, weil diese — obgleich durch nicht ganz kurze Rechnungen — die Verhältnisse am besten beleuchtet und das ganze Formelsystem mit einem Schlag liefert. Wir wollen hier direkt nur gescharte Kollineationen mit getrennten Achsen aufsuchen, aber doch auch solche mit konsekutiven Achsen mitnehmen, wenn sie von selber in den Rechnungen auftreten. Dafür gibt die Methode auf einmal die Achsen vom Typus I und Typus II. Wie vorher nehmen wir an, dass die Flächen eine der Lagen A, B oder D haben; auf die Lage C gehen wir in § 5 ein.

Die Flächen  $\varphi$  und  $\varphi_1$  seien wie früher durch

$$(18) \quad x_1^2 + x_2^2 + x_3^2 + x_4^2 = 0$$

und

$$(19) \quad \lambda_1 x_1^2 + \lambda_2 x_2^2 + \lambda_3 x_3^2 + \lambda_4 x_4^2 = 0$$

gegeben. Zusammengehörige Achsen müssen, wie wir wissen, zwei gegenüberliegende Kanten eines gemeinsamen Polartetraeders schneiden; wir wollen nun die gescharten Kollineationen aufsuchen, deren Achsen die Geraden  $x_1 = 0$ ,  $x_2 = 0$  und  $x_3 = 0$ ,  $x_4 = 0$  schneiden. Die Achsen können dann durch

$$(20) \quad x_1 = \alpha x_2, \quad x_3 = \beta x_4 \quad \text{und} \quad x_1 = \alpha_1 x_2, \quad x_3 = \beta_1 x_4$$

dargestellt werden. Eine gescharte Kollineation mit diesen Achsen ist durch

$$\begin{aligned}
 x_1 - ax_2 &= p(x'_1 - ax'_2), \\
 x_3 - \beta x_4 &= p(x'_3 - \beta x'_4), \\
 x_1 - a_1 x_2 &= q(x'_1 - a_1 x'_2), \\
 x_3 - \beta_1 x_4 &= q(x'_3 - \beta_1 x'_4)
 \end{aligned}
 \quad (a \mp a_1, \beta \mp \beta_1, p \mp q)$$

oder, wenn die Gleichungen nach  $x_1, x_2, x_3$  und  $x_4$  gelöst werden, durch

$$(21) \quad \begin{cases} (a_1 - a)x_1 = (a_1 p - a q)x'_1 - (p - q)aa_1 x'_2, \\ (a_1 - a)x_2 = (p - q)x'_1 - (ap - a_1 q)x'_2, \\ (\beta_1 - \beta)x_3 = (\beta_1 p - \beta q)x'_3 - (p - q)\beta\beta_1 x'_4, \\ (\beta_1 - \beta)x_4 = (p - q)x'_3 - (\beta p - \beta_1 q)x'_4 \end{cases}$$

gegeben. Wir verlangen nun, dass die Transformation (21)  $\varphi$  in  $\varphi_1$  überführen soll. Bei der Einführung von (21) in (18) müssen dann zunächst die Glieder mit  $x'_1 x'_2$  und  $x'_3 x'_4$  verschwinden; dies erfordert

$$(22) \quad pa(1 + a_1^2) = qa_1(1 + a^2)$$

und

$$(23) \quad p\beta(1 + \beta_1^2) = q\beta_1(1 + \beta^2).$$

Das Einsetzen ergibt nun im übrigen, wenn wir die Striche weglassen und nur die erste Hälfte, also die Glieder mit  $x_1^2$  und  $x_2^2$ , aufschreiben:

$$(24) \quad \frac{1}{(a_1 - a)^2} \left\{ x_1^2 [p^2(1 + a_1^2) - 2pq(1 + aa_1) + q^2(1 + a^2)] + x_2^2 [p^2 a^2(1 + a_1^2) - 2pqaa_1(1 + aa_1) + q^2 a_1^2(1 + a^2)] \right\} + \dots = 0.$$

Führt man mittels der Relationen (22) und (23)  $pq$  in diejenigen Glieder von (24) als Faktor ein, wo  $p^2$  oder  $q^2$  steht, so findet man, dass sich die Gleichung (24) nach Kürzen von  $pq$  auf



$$(25) \quad \frac{x_1^2}{aa_1} + aa_1x_2^2 + \frac{x_3^2}{\beta\beta_1} + \beta\beta_1x_4^2 = 0$$

reduziert. Notwendig und hinreichend für die Übereinstimmung mit (19) ist:

$$(26) \quad \lambda_1 aa_1 = \frac{\lambda_2}{aa_1} = \lambda_3 \beta\beta_1 = \frac{\lambda_4}{\beta\beta_1}.$$

Die Gleichungen (22) und (23) schreiben wir so:

$$(27) \quad \frac{a_1(1+a^2)}{a(1+a_1^2)} = \frac{\beta_1(1+\beta^2)}{\beta(1+\beta_1^2)} = \frac{p}{q}.$$

Die Gleichungen (26) in Verbindung mit der ersten (27) bestimmen dann die brauchbaren Achsen und ihren Zusammenhang, während der gleichzeitig durch (27) festgelegte Wert von  $\frac{p}{q}$  das charakteristische Doppelverhältnis der Kollineation liefert.

Aus (26) erhält man zunächst die früher gefundene Bedingung

$$(28) \quad \lambda_1 \lambda_2 = \lambda_3 \lambda_4;$$

diese erweist sich also als notwendig nicht nur für die Existenz von Achsen vom Typus I, sondern auch vom Typus II. Ist die Bedingung erfüllt, so erhält man aus (26), wenn  $\sigma$  einen beliebigen, aber festen Wert von  $\sqrt{\lambda_1 \lambda_2} = \sqrt{\lambda_3 \lambda_4}$  bedeutet:

$$(29) \quad aa_1 = \frac{\sigma}{\lambda_1}, \quad \beta\beta_1 = \frac{\sigma}{\lambda_3}.$$

Führt man  $a_1$  und  $\beta_1$  mittels (29) in (27) ein, so ergibt sich

$$(30) \quad \frac{\sigma(1 + \alpha^2)}{\lambda_1 \alpha^2 + \lambda_2} = \frac{\sigma(1 + \beta^2)}{\lambda_3 \beta^2 + \lambda_4} = \frac{p}{q}.$$

Hierdurch ist das gesuchte Formelsystem gefunden.

Die erste Gleichung (30) liefert die schon bekannte Bedingung (7) dafür, dass eine Gerade  $(\alpha, \beta)$  als Achse brauchbar ist; auch diese gilt also sowohl für Achsen vom Typus I als auch vom Typus II. Gleichzeitig erhält man den Wert des Doppelverhältnisses ausgedrückt durch die  $\lambda$  und die Parameter der einen Achse.

Aus (29) ergibt sich, in Übereinstimmung mit Seite 5—7, dass eine gewählte Achse  $(\alpha, \beta)$  mit zwei anderen zusammengehört, da  $\sigma$  zwei Werte annehmen kann (gleich gross mit entgegengesetztem Vorzeichen). Aus (30) ersieht man dann, dass auch die zwei Doppelverhältnisse gleich gross mit entgegengesetztem Vorzeichen sind.

Weiter sieht man: Wenn  $(\alpha, \beta)$  und  $(\alpha_1, \beta_1)$  ein Paar zusammengehöriger Achsen ist, so gehören je zwei Achsen zusammen, die unter  $(\pm \alpha, \pm \beta)$  bzw.  $(\pm \alpha_1, \pm \beta_1)$  so gewählt sind, dass eine gerade Anzahl von Minuszeichen vorkommt. Den hierdurch gefundenen 8 Transformationen entsprechen insgesamt 2 Werte des charakteristischen Doppelverhältnisses, die gleich gross mit entgegengesetztem Vorzeichen sind; jeder Wert kommt viermal vor.

Übrigens hängt das charakteristische Doppelverhältnis, wenn  $\lambda_1 \mp \lambda_2$  und  $\lambda_3 \mp \lambda_4$  ist, von der Wahl der Achsen ab. Wir können fragen, wann es gleich  $-1$  wird; dies tritt nach (30) ein, wenn

$$\sigma(1 + \alpha^2) + \lambda_1 \alpha^2 + \lambda_2 = 0, \quad \sigma(1 + \beta^2) + \lambda_3 \beta^2 + \lambda_4 = 0.$$

Dies lässt sich, immer unter der Voraussetzung  $\lambda_1 \mp \lambda_2$ ,  $\lambda_3 \mp \lambda_4$ , auf

$$(31) \quad \alpha^2 = -\frac{\sigma}{\lambda_1}, \quad \beta^2 = -\frac{\sigma}{\lambda_3}$$

reduzieren. Wegen der zwei Werte von  $\sigma$  ergeben sich hieraus 8 Achsen ( $\alpha, \beta$ ). Durch Vergleich von (31) mit (29) sieht man, dass

$$\alpha_1 = -\alpha, \quad \beta_1 = -\beta$$

wird. Die 8 Achsen (31) gehören also paarweise zusammen, und wir erhalten 4 involutorische Transformationen.<sup>1</sup>

Setzt man  $\frac{p}{q} = 1$ , so erhält man wieder die 8 Achsen (31), aber in der Form

$$(32) \quad \alpha^2 = \frac{\sigma}{\lambda_1}, \quad \beta^2 = \frac{\sigma}{\lambda_3},$$

sodass

$$\alpha_1 = \alpha, \quad \beta_1 = \beta.$$

Hieraus kann man zunächst nichts schliessen; denn es war ja bei allen Rechnungen dieses Paragraphen vorausgesetzt, dass die Achsen verschieden waren. Doch wird man auf die Vermutung geführt, dass eine Gerade ( $\alpha, \beta$ ), welche (32) befriedigt, Achse einer gescharten Kollineation mit konsekutiven Achsen ist, die auf der Fläche (8)—(9) liegen. Dass dies wirklich der Fall ist, kann man z. B. durch einen Grenzübergang einsehen. Es gibt somit 8 solche Transformationen.

Wir betrachten nun besonders die Lage B. Es sei

$$(33) \quad \lambda_3 = \lambda_4, \quad \lambda_1 \lambda_2 = \lambda_3^2.$$

<sup>1</sup> Vgl. ROHN, S. 63.

Hier ist  $\sigma = \pm \lambda_3$ . Die brauchbaren Achsen werden aus der ersten Gleichung (30) erhalten und zerfallen in

$$(34) \quad \beta = \pm i, \quad \alpha \text{ beliebig}$$

und die früher in (15) angegebenen

$$(35) \quad \alpha = \pm \frac{\sqrt{\lambda_3 - \lambda_2}}{\sqrt{\lambda_1 - \lambda_3}}, \quad \beta \text{ beliebig,}$$

Den Achsen (34) entsprechen für  $\sigma = \lambda_3$  nach (29) die Achsen

$$\beta_1 = \mp i, \quad \alpha_1 = \frac{\lambda_3}{\lambda_1 \alpha}.$$

Hierdurch ist dargetan, dass die Tangenten an die Flächen  $\varphi$  und  $\varphi_1$  in ihren beiden Berührungspunkten paarweise als Achsen solcher gescharten Kollineationen zusammengehören, die  $\varphi$  in  $\varphi_1$  überführen.

Für  $\sigma = -\lambda_3$  erhält man dagegen

$$(36) \quad \beta_1 = \pm i, \quad \alpha_1 = -\frac{\lambda_3}{\lambda_1 \alpha}.$$

Nach (34) und (36) sollten nun entsprechende Achsen Tangenten in demselben Berührungspunkt der Flächen sein, also einander schneiden. Dem entsprechen keine Transformationen der betrachteten Art.

Die Achsen (35) können auch

$$(35 \text{ a}) \quad \alpha = \pm \frac{\sqrt{\lambda_3}}{\sqrt{\lambda_1}}, \quad \beta \text{ beliebig,}$$

geschrieben werden. Ihnen entsprechen für  $\sigma = \lambda_3$  nach (29)

$$(37) \quad \alpha_1 = \pm \frac{\sqrt{\lambda_3}}{\sqrt{\lambda_1}}, \quad \beta_1 = \frac{1}{\beta},$$

und ausserdem ergibt (30), dass  $\frac{p}{q} = 1$  ist. Dem entsprechen keine eigentlichen gescharten Kollineationen. Dagegen stehen die Formeln (35 a) und (37) in gewisser Verbindung damit, dass es 2 parabolische Homologien gibt, deren Homologieebenen mit den Kegelschnittsebenen zusammenfallen und die  $\varphi$  in  $\varphi_1$  überführen (man beachte nur, dass  $\varphi$  und  $\varphi_1$  mit zwei gleich grossen Kugeln projektiv äquivalent sind).

Für  $\sigma = -\lambda_3$  erhält man dagegen

$$\alpha_1 = \mp \frac{\sqrt{\lambda_3}}{\sqrt{\lambda_1}}, \quad \beta_1 = -\frac{1}{\beta},$$

und (30) ergibt nun  $\frac{p}{q} = -1$ . Hierdurch sind  $\infty^1$  harmonische gescharte Kollineationen bestimmt, deren Achsen in je einer der Kegelschnittsebenen liegen.

Bei der Lage D gibt es niemals Lösungen, auch nicht solche, deren Achsen vom Typus II sind; denn die Bedingung (28) zieht, wie schon früher erwähnt, das vollständige Zusammenfallen der Flächen nach sich.

### § 5. Die Schnittfigur der Flächen ist ein windschiefes Viereck.

Wir gehen nun dazu über, die bisher ausgeschlossene Lage C zu betrachten. Zunächst bemerken wir, dass die in §§ 3—4 gegebenen Entwicklungen auch in diesem Fall anwendbar sind; nur geben sie nicht — jedenfalls nicht ohne weiteres — sämtliche Transformationen der betrachteten Art, teils weil ein Übergang zu einem anderen gemein-

samen Polartetraeder als Koordinatentetraeder neue Transformationen liefert (im Gegensatz zum Verhalten bei der Lage B), teils weil es hier Achsenpaare gibt, die kein Paar gegenüberliegender Kanten eines solchen Tetraeders schneiden.

Wir beginnen nichtsdestoweniger damit, die früheren Resultate auf die Lage C anzuwenden. Es sei  $\lambda_1 = \lambda_3$ ,  $\lambda_2 = \lambda_4$ . Dann ist (6) erfüllt; (7) reduziert sich auf

$$(38) \quad a = \pm \beta,$$

und die Fläche (8) zerfällt in die beiden Flächen zweiter Ordnung

$$(39) \quad x_1 x_4 = \pm x_2 x_3.$$

Die Gleichungen (29) und (30) werden hier

$$(40) \quad a a_1 = \beta \beta_1 = \frac{\sigma}{\lambda_1}$$

und

$$(41) \quad \frac{\sigma(1 + \alpha^2)}{\lambda_1 \alpha^2 + \lambda_2} = \frac{p}{q}.$$

Die harmonischen Transformationen sind durch

$$\alpha^2 = \beta^2 = -\frac{\sigma}{\lambda_1}$$

in Verbindung mit  $a_1 = -a$ ,  $\beta_1 = -\beta$  bestimmt. Ferner findet man für  $\frac{p}{q} = 1$

$$\alpha^2 = \beta^2 = \frac{\sigma}{\lambda_1}$$

sowie  $a_1 = a$ ,  $\beta_1 = \beta$ . Dies ergibt Transformationen, deren

Achsen zwei konsekutive Erzeugende einer der Flächen (39) sind.

Um Achsen zu finden, die ein anderes Paar gegenüberliegender Kanten des Koordinatentetraeders schneiden, vertauscht man am einfachsten die Indizes der  $\lambda$ . Die Annahme  $\lambda_1 = \lambda_4, \lambda_2 = \lambda_3$  ergibt etwas mit dem früheren ganz Analoges; es sei nun aber  $\lambda_1 = \lambda_2, \lambda_3 = \lambda_4$ . Hierdurch werden diejenigen Achsen bestimmt, die die beiden Polaren schneiden, welche gegenüberliegende Kanten in allen gemeinsamen Polartetraedern sind. Soll diese letzte Annahme mit (6) verträglich sein, so muss

$$(42) \quad \lambda_1 = \lambda_2 = -\lambda_3 = -\lambda_4$$

sein. Ist dies erfüllt, so zerfällt (7) in

$$(43) \quad a = \pm i, \beta \text{ beliebig}$$

und

$$(44) \quad a \text{ beliebig, } \beta = \pm i.$$

Die entsprechenden Achsen sind die Tangentenbüschel in den 4 Berührungspunkten der Flächen. Ihre Zusammengehörigkeit erhält man aus (29) durch Einsetzen von  $\sigma = \pm \lambda_1$ . Indem wir auf die Ausführungen für die Lage B auf Seite 22 bezugnehmen, sehen wir, dass entsprechende Achsen Tangenten in gegenüberliegenden Berührungspunkten der Flächen sind. Die entsprechenden gescharten Kollineationen sind harmonisch.

Ein elementares Beispiel, wo die mittels (43) und (44) bestimmten Kollineationen existieren, hat man in den beiden Hyperboloiden

$$x^2 - y^2 + z^2 = 1, \quad -x^2 + y^2 + z^2 = 1$$

( $x, y, z$ , rechtwinklige Koordinaten).

Um nun einen vollständigen Überblick zu gewinnen und insbesondere alle Transformationen der gesuchten Art zu finden, wollen wir nach dieser Einleitung die Frage aufs neue angreifen, indem wir als Grundlage der Untersuchung nicht ein zufälliges der  $\infty^2$  gemeinsamen Polartetraeder benutzen, sondern die Schnittkurve der Flächen, die ja in ein windschiefes Viereck ausgeartet ist. Die gemeinsamen Tangentialebenen der beiden Flächen werden von 4 Ebenenbüscheln gebildet, deren Achsen die Seiten des Vierecks sind. Notwendig und hinreichend dafür, dass eine Gerade die Flächen in denselben zwei Punkten schneidet und in zwei gemeinsamen Tangentialebenen liegt, ist, wie man leicht einsieht, dass die Gerade ein Paar gegenüberliegender Seiten des windschiefen Vierecks schneidet (ohne ganz auf einer der Flächen zu liegen). Also:

Bei der Lage C gibt es  $\infty^2$  Achsen vom Typus I, nämlich alle Geraden, die ein Paar gegenüberliegender Seiten im Schnittviereck der Flächen schneiden, ohne ganz auf einer der Flächen zu liegen. Sie bilden, wenn man von den genannten Ausnahmen absieht, zwei lineare Kongruenzen.

Hierzu kommen in gewissen Fällen, wie wir gesehen haben, 4 Geradenbüschel von Achsen vom Typus II.

Wir gehen nun zur Frage der Paarung der Achsen über. Das windschiefe Schnittviereck sei  $ABCD$  (Fig. 3); ferner sei  $a$  eine Achse vom Typus I, die  $AB$  und  $CD$  schneidet. Die Gerade  $a$  ist weder Erzeugende von  $\varphi$  noch von  $\varphi_1$ . Die eine durch  $a$  gehende gemeinsame Tangentialebene ist



durch  $a$  und  $AB$  festgelegt; ihre Berührungspunkte mit den Flächen liegen beide auf  $AB$ . Diese Gerade muss dann bei der Transformation invariant sein und demnach auch die andere Achse,  $b$ , schneiden. Also:

Eine Achse vom Typus I schneidet dieselben Seiten des windschiefen Schnittvierecks wie die zugehörige Achse; diese ist demnach ebenfalls vom Typus I.

Die normale Lage zweier zusammengehörigen Achsen vom Typus I ist die in Fig. 3 angegebene. Speziell kann die Achse  $a$  durch  $C$  gehen, d. h. sie schneidet ausserdem  $BC$ ;

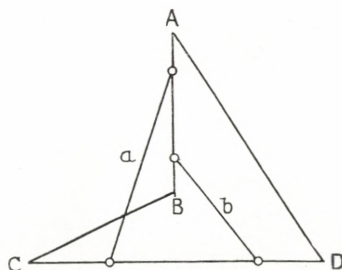


Fig. 3.

dies muss dann auch bei  $b$  der Fall sein, d. h.  $b$  muss durch  $B$  gehen. Noch spezieller kann  $a$  in die Diagonale  $AC$  fallen;  $b$  fällt dann in  $BD$ .

Um den Zusammenhang der Achsen vom Typus I durch eine analytische Behandlung näher festzustellen, wählen wir die Eckpunkte des windschiefen Vierecks als Grundpunkte des Koordinatensystems. Die Flächen  $\varphi$  und  $\varphi_1$  können dann

$$(45) \quad x_1 x_3 + x_2 x_4 = 0$$

bzw.

$$(46) \quad \lambda_1 x_1 x_3 + \lambda_2 x_2 x_4 = 0$$

geschrieben werden. Wir suchen diejenigen Achsen, die  $x_1 = 0$ ,  $x_2 = 0$  und  $x_3 = 0$ ,  $x_4 = 0$  schneiden. Ein Paar zusammengehöriger Achsen lässt sich dann wie früher

durch die Gleichungen (20) darstellen, und eine gescharte Kollineation mit diesen Achsen hat wieder das Aussehen (21). Führt man dies in (45) ein, so müssen zunächst die Glieder mit  $x'_1 x'_4$  und  $x'_2 x'_3$  verschwinden, also:

$$(47) \quad \begin{cases} p\beta(1 + \alpha_1\beta_1) = q\beta_1(1 + \alpha\beta), \\ p\alpha(1 + \alpha_1\beta_1) = q\alpha_1(1 + \alpha\beta). \end{cases}$$

Wir können hier voraussetzen, dass  $1 + \alpha\beta$  (und  $1 + \alpha_1\beta_1$ ) von 0 verschieden ist; denn sonst würde die entsprechende Gerade eine Erzeugende von  $\varphi$  sein. Demnach können die Gleichungen (47) auch in der Form

$$(48) \quad \frac{\alpha_1}{\alpha} = \frac{\beta_1}{\beta} = \frac{p(1 + \alpha_1\beta_1)}{q(1 + \alpha\beta)}$$

geschrieben werden.

Einsetzen von (20) in die linke Seite von (45) ergibt nun, wenn die Striche weggelassen werden:

$$\frac{1}{(\alpha_1 - \alpha)(\beta_1 - \beta)} \left\{ x_1 x_3 [p^2(1 + \alpha_1\beta_1) - pq(\alpha_1\beta + \alpha\beta_1 + 2) + q^2(1 + \alpha\beta)] + x_2 x_4 [p^2\alpha\beta(1 + \alpha_1\beta_1) - pq(\alpha\beta_1 + \alpha_1\beta + 2\alpha\alpha_1\beta\beta_1) + q^2\alpha_1\beta_1(1 + \alpha\beta)] \right\} = 0.$$

Benutzen wir wie in § 4 die Gleichungen (48) auf passende Weise dazu, überall den Faktor  $pq$  statt  $p^2$  oder  $q^2$  einzuführen, so finden wir, dass (45) durch die Transformation in

$$\frac{x_1 x_3}{\alpha\beta_1} + \alpha\beta_1 x_2 x_4 = 0$$

übergeht. Soll dies mit (46) übereinstimmen, so muss

$$(49) \quad \lambda_1^* \alpha\beta_1 = \frac{\lambda_2}{\alpha\beta_1}$$

sein.

Aus (48) und (49) lassen sich, wenn die Achse  $(\alpha, \beta)$

gegeben ist,  $\alpha_1$  und  $\beta_1$  sowie das charakteristische Doppelverhältnis  $\frac{p}{q}$  der Transformation bestimmen. Man erhält, wenn  $\sqrt{\lambda_1 \lambda_2} = \sigma$  gesetzt wird:

$$(50) \quad \alpha_1 \beta = a \beta_1 = \frac{\sigma}{\lambda_1}$$

und

$$(51) \quad \frac{p}{q} = \frac{\sigma(a\beta + 1)}{\lambda_1 \alpha \beta + \lambda_2},$$

wo  $\sigma$  zwei verschiedene Werte haben kann, die gleich gross mit entgegengesetztem Vorzeichen sind. Aus diesen Formeln kann man durch Grenzübergang auch die Fälle erhalten, wo die Achsen durch einige der Ecken des windschiefen Vierecks gehen.

Wir bemerken insbesondere, dass die erste Gleichung (48) bedeutet, dass zwei zusammengehörige Achsen auf einer und derselben Fläche des Büschels

$$(52) \quad x_1 x_4 = k x_2 x_3$$

liegen.

Wollen wir wie früher insbesondere die harmonischen gescharten Kollineationen finden, die  $\varphi$  in  $\varphi_1$  überführen, so setzen wir in (51)  $\frac{p}{q} = -1$ . Nach Reduktion erhält man

$$\alpha \beta = -\frac{\sigma}{\lambda_1}.$$

Für zusammengehörige Achsen gilt dann  $\alpha_1 = -\alpha$ ,  $\beta_1 = -\beta$ .

Setzt man in (51)  $\frac{p}{q} = 1$ , so findet man auf dieselbe Weise

$$\alpha \beta = \frac{\sigma}{\lambda_1},$$

woraus  $\alpha_1 = \alpha$ ,  $\beta_1 = \beta$  folgt. Dem entsprechen gescharte Kollineationen mit konsekutiven Achsen, die auf einer der Flächen (52) liegen. Einige von diesen sind schon auf Seite 24—25 gefunden worden.

Wir haben oben diejenigen Achsen betrachtet, welche das eine Paar gegenüberliegender Seiten des Schnittvierecks schneiden; die Achsen, die das andere Paar schneiden, verhalten sich in derselben Weise. Über eventuelle Achsen vom Typus II ist nur zu bemerken, dass man leicht sieht, dass sie die Flächen in einem Paar von gegenüberliegenden Eckpunkten des Schnittvierecks berühren müssen. Es kann dann keine anderen als die auf Seite 25 erwähnten geben, und diese existieren, wie wir sahen, nur in speziellen Fällen.

### § 6. Flächen mit einer gemeinsamen Erzeugenden.

Als Beispiel einer Lage, wo die Flächen  $\varphi$  und  $\varphi_1$  kein gemeinsames Polartetraeder haben, wollen wir zuletzt kurz den Fall betrachten, wo die Schnittkurve in eine Gerade

und eine Raumkurve dritter Ordnung ausartet, die einander in zwei getrennten Punkten schneiden.

Will man hier die Geradenpaare bestimmen, die gemeinsame Polaren der beiden Flächen sind, so zeigt es sich, dass es keine anderen gibt als die gemeinsame Erzeugende, die ihre eigene Polare in bezug auf beide Flächen ist. Nach § 1 müssen dann alle eventuellen Achsen diese Erzeugende schneiden.

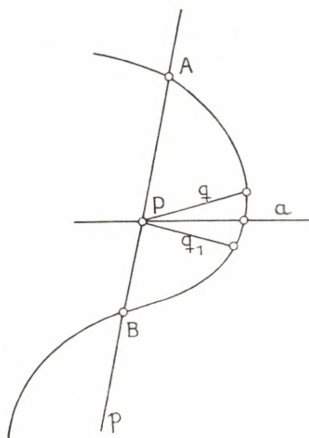


Fig. 4.

Die gemeinsame Erzeugende sei  $p$ , ihre Schnittpunkte mit der Kurve dritter Ordnung  $A$  und  $B$  (Fig. 4). Wir wollen die Achsen vom Typus I bestimmen. Da eine solche Achse in zwei gemeinsamen Tangentialebenen liegen soll, kann sie  $p$  weder in  $A$  noch in  $B$  schneiden. Wir suchen diejenigen Achsen, die  $p$  in einem beliebigen anderen Punkt  $P$  schneiden. Die durch  $P$  gehenden, von  $p$  verschiedenen Erzeugenden von  $\varphi$  und  $\varphi_1$  mögen  $q$  bzw.  $q_1$  heißen; die Ebene  $qq_1$  enthält  $p$  nicht. Eine durch  $P$  gehende, von  $p$  verschiedene Gerade, die in zwei gemeinsamen Tangentialebenen liegen soll, muss der Ebene  $qq_1$  angehören. Diese Ebene schneidet die Kurve dritter Ordnung in 3 Punkten, von denen der eine auf  $q$ , der zweite auf  $q_1$  liegt; der dritte ergibt, mit  $P$  verbunden, eine brauchbare Achse,  $a$ . Also:

Es gibt in dem hier betrachteten Fall  $\infty^1$  Achsen vom Typus I; durch jeden von  $A$  und  $B$  verschiedenen Punkt der gemeinsamen Erzeugenden geht eine.



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

---

THE DISSOCIATION CONSTANT OF THE  
ORTHO-CHLORO-ANILINIUM ION

BY

KAI JULIUS PEDERSEN



KØBENHAVN  
LEVIN & MUNKSGAARD  
EJNAR MUNKSGAARD

1937

Printed in Denmark.  
Bianco Lunos Bogtrykkeri A/S.



In this paper the dissociation constant of the *o*-chloro-anilinium ion

$$K = (\text{H}^+) (\text{B}) / (\text{BH}^+), \quad (1)$$

where  $(\text{H}^+)$ ,  $(\text{B})$ , and  $(\text{BH}^+)$  are the concentrations of hydrogen ions, *o*-chloro-aniline, and *o*-chloro-anilinium ions respectively, is measured by means of the glass electrode in the same way as the dissociation constant of the anilinium ion was determined in an earlier paper.<sup>1</sup>

Cells of the following composition were measured



The "solution" was either an *o*-chloro-aniline buffer (solution 1), or a hydrochloric acid solution (solution 2). The analytical composition of solution 1 was:  $a$  molar *o*-chloro-anilinium chloride,  $b$  molar *o*-chloro-aniline,  $(s-a)$  molar sodium chloride, where  $a \cong b$ . Throughout a series of measurements the total salt concentration  $s$  was constant while four different buffer concentrations were used:  $a \cong b \cong 0.025$ , and  $\frac{3}{4}$ ,  $\frac{1}{2}$ , and  $\frac{1}{4}$  of this concentration. Solution 2, which was measured every day, had the composition:  $0.01026 \text{ } m \text{ HCl}$ ,  $0.0900 \text{ } m \text{ NaCl}$ . The difference  $E$  volts between the e. m. f. measured for solutions 1 and 2 may be considered as the e. m. f. of the cell

<sup>1</sup> PEDERSEN, Kgl. Danske Vid. Selsk., Math.-fys. Medd. **14** (1937) no. 9.

H<sub>2</sub> | solution 1 | 3.5 m KCl | solution 2 | H<sub>2</sub>.

In the paper mentioned above it was shown that, while it is in principle impossible to find  $K$  for the actual buffer solutions without making an arbitrary assumption concerning the salt effect potential, we may by extrapolation find  $K$  for sodium chloride solutions containing the buffer constituents in infinite dilution. The values of  $K$  found in this way are independent of both the diffusion and the salt effect potentials.

We use the same symbols as in the earlier paper, and, in addition, the following abbreviations

$$\alpha \equiv \log \frac{f_1}{f_{(\text{NaCl})_1}} - \delta_1 + \delta_{(\text{NaCl})_1} \quad (2)$$

and

$$\beta \equiv \delta_2 - \delta_{(\text{NaCl})_1} - \log (\text{H}^+)_2 - \log \frac{f_2}{f_{(\text{NaCl})_1}}. \quad (3)$$

We then have

$$-\log K - \alpha = \frac{E}{q} + \beta - \log \frac{b + (\text{H}^+)_1}{a - (\text{H}^+)_1}. \quad (4)$$

Throughout a series of measurements at constant salt concentration  $\beta$  is the same. It can be calculated by means of equation 3 from the known hydrogen ion concentration  $(\text{H}^+)_2$  and the data given in the first paper. We are not able to calculate  $\alpha$ , but it disappears when solution 1 is pure sodium chloride solution.  $(\text{H}^+)_1$  and  $(-\log K - \alpha)$  are found from equations 1 and 4 by successive approximation, the final step being the calculation of  $(\text{H}^+)_1$  for each of the buffer solutions from  $a$ ,  $b$ , and the value of  $K$  for  $s$  molar sodium chloride solution given by the expressions 5.  $(\text{H}^+)_1$  is introduced into equation 4, and  $-\log K - \alpha$  is calculated. From a series of values of  $-\log K - \alpha$  corresponding to the same  $s$  but varying  $a$  we find by rectilinear extrapolation  $-\log K - \alpha$  when  $a = 0$ , that is,  $-\log K$  for

an *s* molar sodium chloride solution. The extrapolation is carried out by the method of least squares. Each determination is given a weight proportional to the concentration of the buffer constituents.

The buffer solutions were made from *o*-chloro-anilinium chloride and sodium hydroxide solution. The *o*-chloro-anilinium chloride was prepared from 50 g. *o*-chloro-aniline (The British Drug Houses, Ltd.), which was dissolved in 550 cc. boiling 10% hydrochloric acid. The salt obtained by cooling in ice was recrystallized from 10% hydrochloric acid, washed with water and alcohol, and dried over sulphuric acid (yield 47 g., preparation 1). According to KING and ORTON<sup>1</sup> the salt obtained in this way is free from the isomeric meta- and para-compounds, because these salts are much more soluble in water than the ortho-salt. 20 g. of the preparation was recrystallized from 65 cc. boiling water. The product (10.5 g.) is called preparation 2.

The two preparations were analyzed by titration with sodium hydroxide and by determination of the amount of bromine which they could take up. The latter analysis was carried out in the same way as was used for aniline in an earlier paper.<sup>2</sup> One mole of *o*-chloro-anilinium chloride needs four equivalents of bromine. From the titration with sodium hydroxide the molecular weight 164.4 was found for both preparations; from the bromine titration 164.7 was found for preparation 1 and 164.1 for preparation 2. From the formula 164.0 is calculated.

The results of the electrometric measurements, which were carried out at the three temperatures 14.8°, 25.0°, and 34.9°, are given in the table. Preparation 2 was used for

<sup>1</sup> KING and ORTON, J. Chem. Soc. **99** (1911) 1381.

<sup>2</sup> PEDERSEN, J. Amer. Chem. Soc. **56** (1934) 2615.

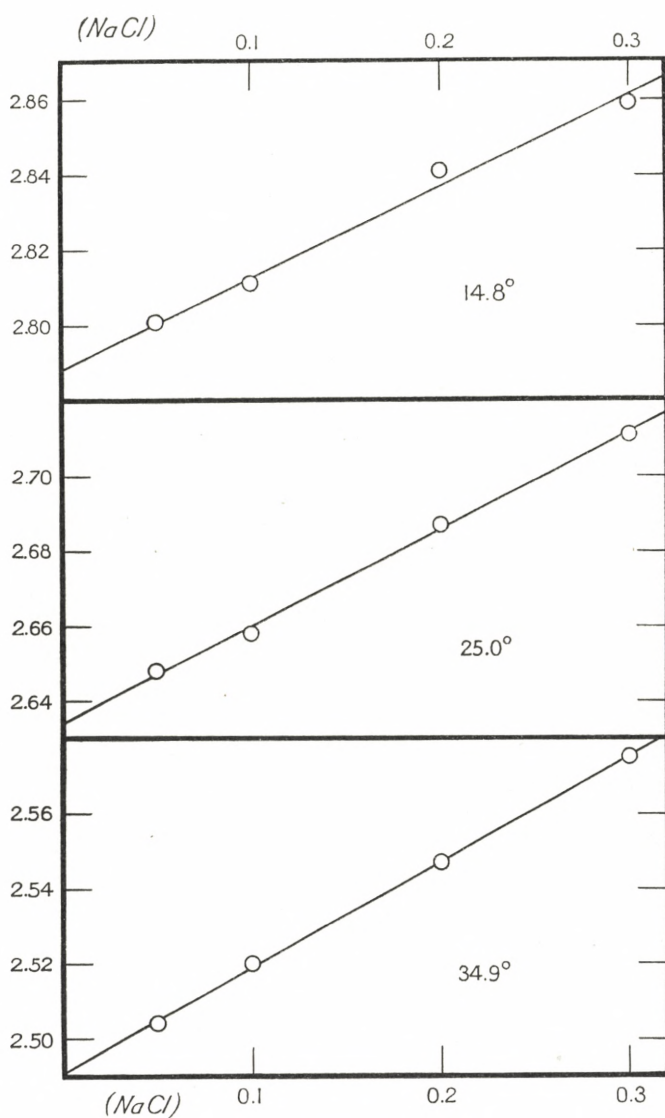


Fig. 1. The dissociation constant of the *o*-chloro-anilinium ion in sodium chloride solutions.

Table.

Electromotive force  $E$  volts of the cell

measured by means of the glass electrode.

	$10^3 a$	$10^3 b$	$10^3 E$	$10^3 (\text{H}^+)_1$	$-\log K - \alpha$
Temp. $14.8^\circ$ $s = 0.050$	* 25.02	24.93	47.54	1.416	2.788
	* 18.77	18.70	48.32	1.370	2.787
	* 12.51	12.46	50.35	1.290	2.796
	* 6.255	6.232	54.15	1.109	2.797
	0	0			(2.801)
Temp. $14.8^\circ$ $s = 0.100$	* 25.07	24.97	48.07	1.383	2.794
	* 18.80	18.73	49.03	1.339	2.797
	* 12.53	12.48	50.96	1.261	2.805
	* 6.267	6.242	54.58	1.087	2.804
	0	0			(2.810)
Temp. $14.8^\circ$ $s = 0.200$	25.00	25.00	49.49	1.310	2.824
	18.75	18.75	50.61	1.270	2.830
	12.50	12.50	52.13	1.200	2.832
	6.25	6.25	55.91	1.040	2.836
	0	0			(2.841)
Temp. $14.8^\circ$ $s = 0.300$	* 25.03	25.09	50.38	1.244	2.849
	* 18.77	18.82	51.28	1.208	2.852
	* 12.51	12.54	52.66	1.144	2.853
	* 6.258	6.273	56.40	0.997	2.858
	0	0			(2.859)

Table  
(continued).

	$10^3 a$	$10^3 b$	$10^3 E$	$10^3 (H^+)_1$	$-\log K - \alpha$
Temp. 25.0° $s = 0.050$	25.03	24.90	41.19	1.941	2.634
	24.98	25.01	41.37	1.931	2.635
	18.77	18.67	42.52	1.859	2.638
	18.73	18.75	43.0	1.850	2.644
	12.52	12.45	44.67	1.720	2.640
	12.49	12.50	44.65	1.711	2.638
	6.258	6.225	49.73	1.426	2.644
	6.245	6.252	49.96	1.420	2.646
	0	0			(2.648)
Temp. 25.0° $s = 0.100$	* 24.94	25.03	42.41	1.878	2.649
	24.95	25.09	42.59	1.874	2.652
	* 18.70	18.77	43.70	1.800	2.653
	18.71	18.82	43.74	1.797	2.653
	* 12.47	12.52	45.7	1.668	2.654
	12.47	12.54	45.71	1.666	2.653
	6.238	6.272	50.64	1.387	2.657
	0	0			(2.658)
	Temp. 25.0° $s = 0.200$	25.07	25.05	43.38	1.790
25.21		24.86	43.28	1.810	2.678
18.80		18.78	44.56	1.719	2.677
18.91		18.65	44.47	1.737	2.681
12.53		12.52	46.62	1.598	2.681
12.60		12.43	46.43	1.615	2.682
6.268		6.263	51.29	1.338	2.682
6.303		6.215	51.45	1.350	2.689
0		0			(2.687)
Temp. 25.0° $s = 0.300$	26.24	25.00	43.27	1.775	2.704
	19.68	18.75	44.24	1.707	2.703
	13.12	12.50	46.34	1.590	2.708
	6.56	6.25	50.89	1.337	2.710
	0	0			(2.711)

Table  
(continued).

	$10^3 a$	$10^3 b$	$10^3 E$	$10^3 (H^+)_1$	$-\log K - \alpha$
Temp. 34.9° $s = 0.050$	24.89	25.10	35.50	2.532	2.496
	18.67	18.83	37.08	2.398	2.498
	12.44	12.55	39.62	2.180	2.498
	6.222	6.275	45.8	1.746	2.503
	0	0			(2.504)
Temp. 34.9° $s = 0.100$	25.04	24.91	36.27	2.493	2.509
	18.78	18.68	37.76	2.364	2.510
	12.52	12.46	40.50	2.151	2.514
	* 6.242	6.258	46.7	1.718	2.518
	0	0			(2.520)
Temp. 34.9° $s = 0.200$	25.07	24.92	37.42	2.364	2.537
	18.80	18.69	38.82	2.246	2.538
	12.54	12.46	41.43	2.052	2.541
	6.268	6.230	47.40	1.659	2.547
	0	0			(2.547)
Temp. 34.9° $s = 0.300$	* 24.97	25.03	37.9	2.222	2.553
	24.92	24.92	38.39	2.226	2.562
	* 18.73	18.77	39.34	2.118	2.556
	18.69	18.69	39.23	2.120	2.554
	12.46	12.46	42.27	1.944	2.567
	* 6.242	6.258	48.15	1.583	2.573
	6.230	6.230	48.14	1.583	2.574
0	0			(2.575)	

most of the experiments. Only the measurements marked with an asterisk were carried out with preparation 1. The last two columns in the table contain  $(H^+)_{1}$  and  $-\log K - \alpha$ ,

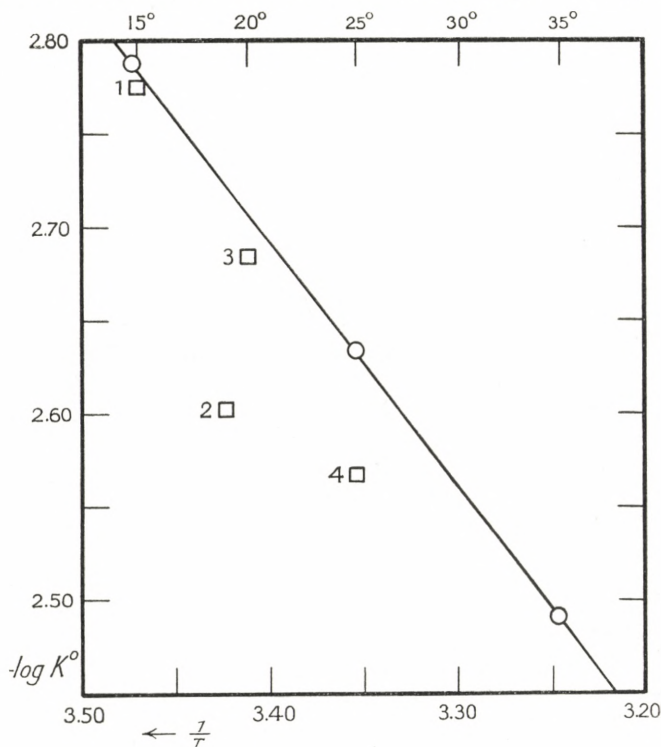


Fig. 2. The dissociation constant of the *o*-chloro-anilinium ion. Circles, this paper. Squares, other determinations.

<sup>1</sup> BRØNSTED and DUUS, catalysis. <sup>2</sup> BRØNSTED and DUUS, hydrogen electrode,

<sup>3</sup> BRØNSTED and BEHRNTS-JENSEN, hydrogen electrode. <sup>4</sup> WILLIAMS and SOPER, partition between water and benzene.

<sup>1,2</sup> Z. physik. Chem. **117** (1925) 299. <sup>3</sup> Private communication. <sup>4</sup> J. Chem. Soc. (1930) 2469.

computed as described above. For each series of experiments the value of  $-\log K$  for pure sodium chloride solution, found by extrapolation of  $-\log K - \alpha$  is given in parenthesis. In Fig. 1  $-\log K$  for the sodium chloride solutions is plotted against the concentration  $s$ .



The straight lines are those which are found by the method of least squares to fit closest to the points. They correspond to the following expressions for the dissociation constant  $K$  of the *o*-chloro-anilinium ion in an  $s$  molar sodium chloride solution ( $s < 0.3$ ):

$$\left. \begin{array}{l} \text{at } 14.8^\circ \quad -\log K = 2.788 + 0.243 s \\ \text{at } 25.0^\circ \quad -\log K = 2.634 + 0.260 s \\ \text{at } 34.9^\circ \quad -\log K = 2.491 + 0.281 s \end{array} \right\} \quad (5)$$

By setting  $s = 0$  we find  $K^\circ$ , the dissociation constant at infinite dilution. If we plot  $-\log K^\circ$  against  $\frac{1}{T}$ , the reciprocal of the absolute temperature, we get the points indicated by circles in Fig. 2. They fall on a straight line of equation

$$-\log K^\circ = 1.313 \frac{10^3}{T} - 1.772. \quad (6)$$

From the slope we find the heat of dissociation of the *o*-chloro-anilinium ion at infinite dilution,  $-6007$  cal./mole.

The dependence of  $K^\circ$  on the temperature may also be expressed by the following equation which is more convenient for interpolation than equation 6

$$-\log K^\circ = 2.634 - 0.0148(t - 25) + 0.00003(t - 25)^2.$$

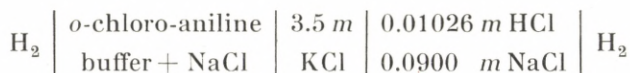
Here  $t$  is the temperature in degrees centigrade.

In Fig. 2 we have plotted both the determinations in this paper (circles) and those of other investigators (squares). The numbers refer to the list under the figure.

I wish to express my thanks to the head of the laboratory, Professor NIELS BJERRUM, for his kind interest in my work.

### Summary.

The e. m. f. of cells of the composition



were measured by means of the glass electrode at 14.8°, 25.0°, and 34.9°.

The dissociation constant of the *o*-chloro-anilinium ion at infinite dilution in 0—0.3 molar sodium chloride was calculated from the measurements (expressions 5).

From the variation with temperature the heat of dissociation of the *o*-chloro-anilinium ion at infinite dilution was calculated, and found to be —6007 cal./mole.

*Chemical Laboratory of the Royal Veterinary  
and Agricultural College, Copenhagen.*

April 1937.

Det Kgl. Danske Videnskabernes Selskab.

Mathematisk-fysiske Meddelelser **XV**, 4.

---

OM RELATIONEN MELLEM VARME  
OG ARBEJDE

AF

J. N. BRØNSTED



KØBENHAVN  
LEVIN & MUNKSGAARD

EJNAR MUNKSGAARD

1937

Printed in Denmark.  
Bianco Lunos Bogtrykkeri A/S.

## 1. Indledning.

De thermodynamiske Grundsætningers særlige Karakter beror som bekendt paa deres Uafhængighed af specielle Antagelser om Stoffets Natur og deres direkte Udledning udfra et makroskopisk Erfaringsmateriale. Indholdet af Thermodynamiken er dog ikke simpelthen en Ophobning af Erfaringskendsgerninger. Hvis vi undersøger den historiske Udvikling, vil vi hos Grundlæggerne af Thermodynamiken finde dybt rodfæstet visse konstruktive Ideer, hvis Oprindelse paa Grund af deres aksiomatiske Karakter vanskeligt kan paapeges, men som danner det egentlige Udgangspunkt for Opbygningen af Thermodynamikens System.

Dette er saaledes i høj Grad betegnende for den første betydningsfulde thermodynamiske Indsats, der knytter sig til Navnet CARNOT. I sit berømte Arbejde af 1824<sup>1</sup>, der indeholder de væsentlige Træk af det, der senere betegnedes som Varmetheoriens 2den Hovedsætning, har CARNOT udviklet Theorien for »Varmens bevægende Kraft« paa Grundlag af fysisk-tekniske Erfaringer uden nogen Støtte i de eksisterende Forestillinger om Stoffets molekulære Struktur. Men denne Tilknytning til Iagttagelse og Erfaring er forbundet med en ledende Grundtanke, der kunde udtrykkes som Ideen om Temperaturens Potentialkarakter. For Grundlæggerne af 1ste Hovedsætning R. J. MAYER og J. P. JOULE, hvis

<sup>1</sup> S. CARNOT: Reflexions sur la puissance motrice du feu, 1824.

Arbejdet indledes knap 20 Aar efter CARNOT, er Forholdet af lignende Art. Begge er besjælede af Forestillingen om »Kræfternes Enhed« og deres gensidige Omdannelsesmuligheder. For begges Vedkommende fremtræder disse Ideer som aprioriske Sandheder, tilkendegivende en vis metafysisk Indstilling overfor Problemerne. Dette kan med ikke mindre Ret siges at være betegnende for COLDING<sup>1</sup>, der har bidraget uafhængigt til Energiprincipets Udvikling. Om dem alle gælder dog uomtvisteligt, at de Forestillinger, der har dannet det ide-mæssige Udgangspunkt, ikke er hævdet som et fyldestgørende Bevismateriale, men har været forbundet med en klar Forstaaelse af Nødvendigheden af en induktiv erfaringsmæssig Bevisførelse for de opstillede Postulater.

Fremtræder Thermodynamiken saaledes saavel i sin Grundlæggelse som i sin videre Udvikling som en Erfaringsvidenskab, hvis Læresætninger deduceres udfra ganske faa empiriske Postulater eller Principer, saa synes det dog ved en nærmere Undersøgelse utvivlsomt, at Elementer af hypothetisk Natur i visse Henseender har præget Fortolkningen af det energetiske Erfaringsmateriale. Ved Opstillingen af den »mekaniske Varmetheoris« Varmebegreb og det deraf følgende Ækvivalensprincip er det i højere Grad den kvantitative Sammenhæng mellem Energiformerne, der har været Genstand for Problemopstillingen, end en gennemført kritisk Analyse af Energiomdannelsernes Gennemførlighed. Der har her, specielt hvad angaar Forholdet mellem Varme og mekanisk Arbejde, været Plads for hypothetisk Argumentation udfra molekular-theoretiske Forestillinger. HELMHOLTZ udtaler med Henblik paa dette Problem<sup>2</sup>: »Ob bei der Erzeugung mechanischer

<sup>1</sup> L. A. COLDING: Undersøgelse om de almindelige Naturkræfter, Kgl. Danske Vidensk. Selsk. Skr. 2, 121 (1851.)

<sup>2</sup> H. v. HELMHOLTZ: Über die Erhaltung der Kraft, 33 (1847).

Kraft Wärme verschwindet, was ein notwendiges Postulat der Erhaltung der Kraft sein würde, ist noch niemals gefragt worden«. Maaske vilde det være rigtigere at sige, at Problemet til det paagældende Tidspunkt endnu aldrig havde været underkastet en maalbevidst Prøvelse. Men han giver ved den videre Drøftelse heraf sin Tilslutning til den Opfattelse, at Sammenhængen mellem Varme og Arbejde kan udledes ved Betragtning af Processer, ved hvilke »die ganze lebendige Kraft der hingetretenen Wärme wirklich in Arbeitskraft abgegeben wird«.

Denne Udtalelse af en af Energetikens ledende Skikkelser er overordentlig karakteristisk for Stillingen til Spørgsmaalet om Relationen mellem Varme og Arbejde. Den deri udtrykte Mulighed for en fuldstændig Afgivelse af Varmens »levende Kraft« i Form af tilgængelig »Arbejdskraft« er ikke blot knyttet til den fremspirende Ide om Kræfternes Enhed, men er en fast Ingrediens i den efterfølgende Tids Formuleringer af et formentligt Erfaringsmateriale. Naar et System forandrer sig under Arbejdsydelse og Varmeoptagelse, betragtes Varmens Funktion som bestaaende dels i en Temperaturforøgelse, dels i Frembringelsen af et »indre Arbejde«, ogsaa betegnet som »latent Varme«, dels endelig i Frembringelsen af et ydre Arbejde, der i Kraft af den mekaniske Varmetheoris Naturopfattelse anses for at opstaa paa Bekostning af den kinetiske Molekularenergi, hvormed Varmen identificeres. Det er aabenbart, at det ved denne Fortolkning er den ledende Ide om Varmens Natur og dens mekaniske Ækvi-valens og ikke den nøgterne Henholdelse til et praktisk Iagttagelsesmateriale, der er den afgørende Faktor.

Paa ganske anden Maade forankret i det teknisk-erfaringsmæssige optræder den mere specialiserede CARNOT'ske Ide. Det er klart, at denne før eller senere maatte kollidere med

Ækvivalensprincippet i dets oprindelige Form, da hvert af disse Principer gav sin Løsning af Problemet om Forholdet mellem Varme og Arbejde. Dette er først klart opfattet og udtrykt af WILLIAM THOMSON. I den traditionelle Thermodynamik tilskriver man R. CLAUDIUS Forligelsen af de stridende Anskuelser paa dette Omraade.

Om der ved den CLAUDIUS'ske Løsning er opnaaet den endelige og modsigelsesfri Afgørelse af Varme-Arbejds-Problemet, er et andet Spørgsmaal. Den omhandler jo i ethvert Tilfælde kun et enkelt af de Fænomener, der er omfattet af de ovenfor nævnte Forestillinger om Relationen mellem Varme og Arbejde. Det vil være af betydelig Interesse at betragte Problemet i sin Almindelighed fra den genetiske Sammenhængs Synspunkt ved en konsekvent Gennemførelse og Generalisering af den CARNOT'ske Ide. Dertil maa knytte sig en nærmere Analyse af Varmebegrebet fænomenologisk betragtet, uafhængig af Molekylkinetiken.

Ved Behandlingen af denne Opgave, som er nærværende Afhandlings, vil det vise sig, at en skarpere Formulering af de energetiske Grundbegreber og Grundfænomener er i væsentlig Grad paakrævet. Dette gælder ikke blot de Begreber og Fænomener, i hvilke Varmen eller i det hele taget de thermiske Parametre indgaar. Vi vil derfor betegne vor Opgave som henhørende under Energetiken i RANKINE-OSTWALD'sk Betydning snarere end under Thermodynamikens mere begrænsede Omraade.

## 2. De energetiske Grundfænomener.

Det Begreb, der mest umiddelbart fremstiller sig som det grundlæggende i Energetiken, er Arbejdsbegrebet. Visse Fænomener som Løftningen af et Lod, Spændingen af en Fjeder etc. foregaar ikke spontant, men der maa,



for at de skal kunne fremkaldes, tilvejebringes visse fysiske Betingelser for deres Realisation. Man siger, at der maa tilføres »Arbejde«. Det vil paa dette Sted være overflødigt at søge en dybere erkendelsestheoretisk Analyse af Arbejdsbegrebet, idet det nødvendige for den foreliggende Opgave er dets praktisk akcepterede og utvetydige Betydning. Men der er Grund til at indlede vor Undersøgelse med Udpegelsen af de elementære Bestanddele, hvoraf Arbejdet og overhovedet det, vi betegner som »Energi«, kan sammensættes, og ved hvis Hjælp det alene er muligt at give en fyldestgørende Beskrivelse af de i Naturen foregaaende energetiske Forandringer.

Som allerede fremhævet af MAXWELL<sup>1</sup> for enkelte Energi-formers Vedkommende, lader disse energetiske Grundbegreber sig dele i to Hovedgrupper, der hver for sig omfatter Led med sideordnede Egenskaber. Vi betegner med G. HELM<sup>2</sup> de i disse Grupper indgaaende Grundbegreber som Kvantitets- og Intensitetsstørrelser. De førstnævnte, f. Ex. Masse, Rumfang, Elektricitetsmængde, Stoffmængde og Entropi, er, i Overensstemmelse med deres Betegnelse som Kvantiteter, udmærkede ved Delelighed, Summabilitet og Transportabilitet. De til den anden Gruppe hørende, f. Ex. Gravitationspotential, Tryk, elektrisk Potential, kemisk Potential og Temperatur, har et Potentials almindelige Egenskaber og er bestemmende for den mulige Indtræden af energetiske Forandringer. De parvis korresponderende Kvantitets- og Intensitetsstørrelser kan kombineres til Produkter, der har karakteristiske Fællesegen-

<sup>1</sup> C. MAXWELL: Theory of Heat, 5th Ed., 194 (1877).

<sup>2</sup> G. HELM: Die Lehre von der Energie, 56 (1887). Jfr. ogsaa: W. OSTWALD: Studien zur Energetik, Z. f. physik. Chem. 9, 563 (1892), 10, 363, (1892).

skaber og sædvanligvis siges at fremstille »Energi«. Disse Produkter er ligesom de energetiske Kvantiteter kendetegnet af Delelighed og Summabilitet, men ikke af Transportabilitet i den almindelige Forstand, hvori denne Egenskab er til Stede for Kvantiteter.

Det er af afgørende Betydning for de energetiske Problemers Behandling, at man ved Hjælp af de nævnte Energifaktorer kan fastslaa de Elementer, hvori de energetiske Processer kan opløses. Kvantitetsfaktorerne kan i al Almindelighed hverken opstaa eller tilintetgøres. I Overensstemmelse hermed kan de simpleste energetiske Forandringer beskrives som Flytning af en Kvantitet af ovennævnte Art fra et Potential til et andet. En saadan Transport, der kendetegner og er en Bestanddel af enhver Art af energetisk Forandring, vil vi betegne som den energetiske Grundproces. Den kan bestaa f. Ex. i Flytning af et Rumfang fra et Tryk til et andet, af en Stofmængde fra et kemisk Potential til et andet eller en Entropimængde fra en Temperatur til en anden o. s. v. De Potentialer, hvormellem Kvantiteten bevæger sig, kan eventuelt være lige store, i hvilket Tilfælde Processen kunde betegnes som en neutral Grundproces, medens Kvantitetsforskydningen mellem indbyrdes forskellige Potentialer, naar det er paa krævet for Tydelighedens Skyld, vil blive betegnet som en aktiv Grundproces. Den neutrale Grundproces fremstiller et Grænsetilfælde af den aktive for uendelig lille Potentialdifferens, men der er for den førstnævntes Forløb iøvrigt ingen særlige Betingelser til Stede. For den sidstnævntes Vedkommende gælder, at den ikke kan forløbe isoleret, d. v. s. der kendes ikke i Naturen Processer, der alene bestaar i Flytning af en enkelt Kvantitet mellem forskellige Potentialer. En saadan Flytning vil enten være

knyttet til en anden jævnsides dermed forløbende aktiv Grundproces, i hvilket Tilfælde der imellem dem er etableret en Kobling, der nødvendiggør Samtidighed i deres Forløb. Eller den vil, hvis en saadan Kobling mangler, være ledsaget af en Frembringelse af Entropi. Ved den sidstnævnte Proces har vi at gøre med en Kvantitetsdannelse, der aabenbart afviger fra den ellers almengyldige Sætning om Kvantitetens Konstans. Man kan saaledes opstille to Grupper af indbyrdes forskelligartede aktive Grundprocesser, de koblede og de ukoblede. De første tilhører Systemer, der kan betegnes som værende i Ligevægt, idet der i Systemet som Helhed ikke er nogen »Tendens« til Stede til en Forandring af Tilstanden. De Processer, der forløber under disse Betingelser, er altsaa ikke retningsbestemte, de forløber lige let i modsatte Retninger og betegnes almindeligt som reversible. Det er muligt at opfatte den paagældende Ligevægtstilstand som en afbalanceret Ligevægt, hvori de enkelte aktive Grundprocessers Tilbøjelighed til at forløbe ophæver hinanden. Herom giver dog en Betragtning af koblede Systemer ingensomhelst Oplysning. Spørgsmaalet om, hvorvidt en given Grundproces har Tendens til at forløbe, kan kun afgøres ved Betragtning af ikke koblede Systemer. Heri vil de enkelte Grundprocesser forløbe uafhængigt af hinanden, og der vil da, som nævnt, tillige foregaa en Entropidannelse. Saadanne Processer, karakteriserede af Uligevægt og Entropidannelse, er retningsbestemte og betegnes som irreversible eller frivillige Processer.

Det simpleste Tilfælde af en irreversible Grundproces fremstilles ved Flytning af en enkelt Kvantitetsart mellem to Potentialer. Denne Flytning kan siges at foregaa fra et højere til et lavere Potential, idet den relative Placering af to Potentialer, foretaget efter dette Princip, er entydig.

Er Retningen den nævnte, kaldes Processen positiv, uafhængig af den Maade, hvorpaa den iøvrigt forløber. Er Retningen modsat, kaldes Processen negativ. Man har saaledes i Betragtningen af den irreversible Proces et Middel til Ordning af Potentialer af samme Art efter Størrelse og Bestemmelse af en aktiv Grundproces's Fortegn. For den reversible Proces betyder dette, at visse Kvantitetsforskudninger maa foregaa fra et højere til et lavere Potential, medens andre forløber i modsat Retning. Med andre Ord: den reversible Proces er, i Overensstemmelse med den Ligevægtstilstand, der som nævnt hersker i koblede Systemer, sammensat af Grundprocesser af forskelligt Fortegn.

De Energifaktorer, som saaledes er anvendelige til Karakterisering af de energetiske Processer, kan, som allerede antydtes, tillige tjene til Formulering af Arbejdsbegrebet. Dette er paa det nøjeste knyttet til den energetiske Grundproces. Medens Kvantitetens Bevægelse fra et Potential til et andet beskriver Processen, er den paagældende Potentialdifferens, multipliceret med den transporterede Kvantitet, et Udtryk for Arbejdet. Ved den positive Proces, d. v. s. hvis Flytningen af Kvantiteten sker fra et højere til et lavere Potential, siges der i Systemet at foregaa et indre Arbejdstab, medens den negative Proces, d. v. s. Flytning i modsat Retning, er forbundet med en indre Arbejdsgevinst. Arbejdstabet har altsaa samme Fortegn som Grundprocessen. Er  $P_1$  det Potential, hvorfra Kvantiteten  $\delta K$  flyttes, og  $P_2$  det Potential, hvortil den føres, kan man for Arbejdstabet  $\delta A$  ved den paagældende Grundproces saaledes skrive:

$$\delta A = (P_1 - P_2) \delta K. \quad (1)$$

Man kan ogsaa sige, at Systemets »potentielle Energi« ved Grundprocessens Forløb er formindsket med denne Størrelse.

Til hvert Par af korresponderende Energifaktorer svarer en Arbejdsart, hvis Størrelse udtrykkes ved Ligninger af Form som (1). For de Arbejdsarter, der almindeligt optræder i Energetiken, er disse Ligninger opført nedenfor, hvorved Betydningen af de for de indgaaende Kvantiteter og Potentiale anvendte Symboler umiddelbart vil fremgaa.

$$\text{Gravitationsarbejde} \dots \delta A = (\Phi_1 - \Phi_2) \delta m \quad (2)$$

$$\text{Rumfangsarbejde} \dots \delta A = -(p_1 - p_2) \delta v \quad (3)$$

$$\text{Elektrisk Arbejde} \dots \delta A = (\psi_1 - \psi_2) \delta \epsilon \quad (4)$$

$$\text{Kemisk Arbejde} \dots \delta A = (\mu_1 - \mu_2) \delta n \quad (5)$$

$$\text{Thermisk Arbejde} \dots \delta A = (T_1 - T_2) \delta S \quad (6)$$

Det negative Fortegn til Udtrykket for Rumfangsarbejdet betegner Trykket som identisk med Rumfangspotentialet efter Fortegnsforandring. Den positive Grundproces bestaar altsaa for Rumfangsarbejdet i en Kvantitetsforskydning fra det mindre til det større Tryk.

Hvis man deler et foreliggende energetisk System i to Afsnit, der paa ovennævnte Maade er koblede til hinanden, fremkommer to Del-Systemer, som vi vil kalde arbejds-kommunicerende Systemer. En særlig Simpelhed, der er af Betydning for energetiske Problemers formelle Behandling, fremkommer, hvis et af disse arbejds-kommunicerende Systemer er saaledes afgrænset, at der deri kun kan forløbe en enkelt aktiv Grundproces, f. Ex. Forskydningen af et Lod, der har Adgang til at bevæge sig vinkelret paa Tyngdekraftens Retning. Et saadant Del-System vil vi kalde et Arbejdsreservoir. Den positive Grundproces foregaar heri frivilligt og irreversibelt ved Løsning af den bestaaende Kobling, idet f. Ex. ved Sænkning af Loddet Masse flyttes fra et højere til et lavere Gravitationspotential. Ved reversibel Kobling vil Reservoirret herved »afgive« og Systemet

»modtage« Arbejde, medens der til Løftning af Loddet kræves Arbejde tilført Reservoirret. En saadan Anordning, der er et vigtigt energetisk Hjælpemiddel, tjener til Opsamling og Registrering af det Arbejde, der ydes af et vilkaarligt dertil koblet System, d. v. s. af det Arbejdstab, som foregaar deri ved Forløbet af de energetiske Grundprocesser.

Paa lignende Maade vil det være muligt at opsamle den irreversibelt frembragte Entropi i Entropireservoirer, d. v. s. afsluttede Omraader, hvori der ikke foregaar aktive Grundprocesser eller yderligere Kvantitetstilførsel. Det er i denne Sammehæng væsentligt at bemærke, at den irreversible Proces ikke er fuldstændig karakteriseret ved Beskrivelse af Grundprocessen og Paapegelse af Irreversibiliteten, men til fuldstændig Karakterisation kræver en Lokalisering af den frembragte Entropi og Fastsættelse af den Temperatur, ved hvilken den bliver opsamlet. Afhængende af dette Valg vil der aabenbart til en given irreversibel Grundproces svare forskellige Totalprocesser, et Forhold, der frembyder en væsentlig Distinktion i Beskrivelsen af de irreversible Forandringer i Naturen.

Paa tilsvarende Maade vil man i Energetiken anvende Reservoirer for andre Kvantitetsstørrelser end Entropien. Et Systems rumlige Omgivelser maa saaledes opfattes som et Rumfangsreservoir, hvormed Systemet kommunikerer, idet dets Rumfang ændres.

I visse Tilfælde vil Forløbet af den energetiske Grundproces ikke medføre nogen Ændring i de indgaaende Potentialer. F. Ex. vil dette med stor Tilnærmelse være rigtigt ved Bevægelsen af et Lod i Forhold til Jordoverfladen. Almindeligvis vil Potentialdifferensen dog kun ved den differentiale Proces kunne sættes konstant. Ved integrale Processer, d. v. s. Processer af endelig Udstrækning,

vil Potentialerne som Regel forandre sig under Processens Forløb saavel for den Arbejdsarts Vedkommende, hvis Kvantitet transporteres, som ogsaa for andre i Systemet optrædende Potentialer. Specielt er dette af Vigtighed for Systemer, i hvilke kemisk, thermisk og Rumfangsarbejde optræder. Der opstaar dog ikke derved Forhold, der principielt influerer paa Anvendelsen af den ovenfor givne Skemativering af de energetiske Grundfænomener.

Anvendelsen af Begrebet: »den energetiske Grundproces«, hvormed vi har betegnet en Flytning af Kvantitet mellem forskellige Potentialer, vil frembyde væsentlige Fordele ved Studiet af de energetiske Fænomener. Betydningen af denne Kvantitetsflytning for Arbejdsbegrebet er tidligere fremhævet af MEYERHOFFER<sup>1</sup>. Hvor Synspunkter af lignende Art i øvrigt er fremkommet i Thermodynamiken, er det almindeligt en Energioverføring, ikke en Kvantitetsoverføring, der opstilles som den elementære Proces. Saaledes udtrykker HELM<sup>2</sup> Energetikens Fundamentallov paa følgende Maade: »Jede Energieform hat das Bestreben von Stellen, in welchen sie in höherer Intensität vorhanden ist, zu Stellen von niederer Intensität überzugehen«. Hertil maa dog bemærkes, at naar en saadan Proces gennemføres, vil der samtidig foregaa en Energiomdannelse, der hindrer Processen i at begrænses til en simpel Energi-transport. I Overensstemmelse hermed er Energi i Indledningen til dette Afsnit betegnet som utransportabel.

### 3. Kvantitetsvariable Systemer.

Den ovenstaaende Beskrivelse af den energetiske Proces er udtømmende i den Forstand, at det for Systemer i Hvil

<sup>1</sup> W. MEYERHOFFER, Z. f. physik. Chemie, 7, 544 (1891).

<sup>2</sup> G. HELM: Die Lehre von der Energie, 62 (1887).

ikke er muligt at paapege energetiske Fænomener, der ikke er dækket af den givne Fremstilling. Det er derimod muligt at begrænse Betragtningen af Grundprocessen til en Del af de faktisk indtrædende Forandringer og derved at opnaa en Behandling af energetiske Problemer, der overfor visse Tilfælde i formel Henseende kan frembyde Fordele. Dette kan ske derved, at man spalter Forsøgssystemet i to Dele, der simplest hver for sig tænkes at være i indre Ligevægt, men som kan have indbyrdes forskellige Værdier for et eller flere Potentialer, saaledes at det samlede System kan være i Uligevægt. Den energetiske Funktion af et saadant System som Helhed betragtet kan kun bestaa i en Vandring af Kvantitet fra det ene Delsystem til det andet. For hvert af disses Vedkommende vil det, der foregaar, derimod bestaa i en Tilførsel eller Bortførsel af Kvantitet. Et System, der har en saadan Adgang til vilkaarlig Ændring af sit Kvantitetsindhold, vil vi betegne som et kvantitetsvariabelt eller aabent System i Modsætning til Totalsystemet, der er kvantitetsafsluttet eller lukket. Det bemærkes udtrykkeligt, at et saadant lukket System meget vel kan være arbejds-kommunicerende.

Det er almindeligt i Thermodynamiken at betragte en Kvantitetsoptagelse som ensbetydende med Energiop-tagelse. Hvis Systemet optager Kvantiteten  $dK$ , og det tilsvarende Potential i Systemet er  $P$ , siges Energioptagelsen eller Forøgelsen i Energi i Systemet at være:

$$dE = PdK, \quad (7)$$

medens den totale Energiforøgelse, naar flere Kvantitetsformer optages, er:

$$dE = \sum P_1 dK_1. \quad (8)$$



Dette harmonerer f. Ex. med den GIBBS'ske Energirelation<sup>1</sup>, men de anførte Ligninger kan i Virkeligheden kun betragtes som Definitioner af Energitilvæksten og er ikke Udtryk for iagttagelig eller verificerbar Realitet. Men det er klart, at man i Kraft af disse Ligninger og den dertil knyttede Konvention maa tillægge ethvert System et Indhold af saavel Energi som Kvantitet. Kvantitetstilførslen kan ikke identificeres med Tilførsel af Arbejde eller potentiel Energi i nogen Form, idet der hertil kræves en Flytning af Kvantitet indenfor Systemets Grænser. Det vil derfor være naturligt at betegne den saaledes ved konstant Potential tilstedeværende eller ved Kvantitetsvariation tilførte Energi som ækvipotentiel.

Specielt maa det iagttages, at den fuldkomne Parallelisme, der hersker imellem Udtrykkene for de forskellige Arbejdsformer, genfindes ved Betragtning af de her nævnte energetiske Forandringer i aabne Systemer. Dette er en simpel Følge af den identiske Dimensionering, der karakteriserer de i Ligningerne (1) og (7) for henholdsvis  $\delta A$  og  $dE$  indgaaende Produkter. Naar det er almindeligt i Thermodynamiken at tillægge visse af disse Produkter en anden Karakter end andre, idet f. Ex.  $p dv$  identificeres med »Arbejde«, medens Produktet  $T dS$  identificeres med »Varme«, saa maa en saadan Terminologi derfor betegnes som værende uden nogen principiel Begrundelse. Da denne Uoverensstemmelse med den traditionelle thermodynamiske Opfattelse er af væsentlig Betydning for nærværende Fremstilling, vil det være formaalstjenligt, at den underkastes en nøjere Betragtning.

Vi vil i dette Øjemed først undersøge en Forandring, ved hvilken Rumfanget er den ændrede Kvantitet. Naar det betragtede System af Trykket  $p$  undergaar en Rumfangsforøgelse  $dv$ , siges det at udrette Arbejdet  $p dv$ , fordi der

<sup>1</sup> W. GIBBS, Collected Works I, 63 (1928).

ved den paagældende Rumfangsforøgelse kan opnaas, at der i et Arbejdsreservoir — f. Ex. derved, at et Lod løftes — opsamles et Arbejde af netop denne Størrelse. Denne Arbejdsydelse i Reservoiret opnaas imidlertid ikke forudsætningsløst. Der kræves dertil, at Rumfanget  $dv$ , der ved Processen opnaar Trykket  $p$ , tages fra et Vakuum. Men i saa Fald maa Systemet omfatte et saadant Vakuum. Processen antager imidlertid derved Karakteren af arbejdsyde Grundproces, idet den bestaar af en Forskydning af et Rumfang mellem to Tryk indenfor et afsluttet, arbejdskommunicerende System. Da Rumfangsflytningen foregaar fra det mindre til det større Tryk, er der til en saadan Proces knyttet et Arbejdstab. Men det er naturligvis yderligere en Forudsætning for Arbejdspræstationen i Reservoiret, at Grundprocessen forløber reversibelt.

Naar et Rumfangsarbejde  $pdv$  præsteres eller modtages af Systemet, sker det altsaa ved en heri forløbende veldefineret Grundproces. For Delsystemet alene er Arbejdsbegrebet uden Mening. Det samme gælder om Reversibilitetsbegrebet. Derimod vil Optagelsen af Rumfanget  $dv$  definitions-mæssigt medføre en Energi-optagelse, der er —  $pdv$ , og som ifølge Sagens Natur i Overensstemmelse med vor Terminologi maa betegnes som ækvipotentiel Energi.

Er ikke Rumfanget, men Entropien den ændrede Kvantitet, vil man ganske analogt kunne betegne Optagelse af Entropimængden  $dS$  ved Temperaturen  $T$  af Systemet som ensbetydende med Optagelse af Energimængden  $TdS$ . At opfatte denne Optagelse som Arbejde er, trods den fuldkomne Analogi mellem Produkterne  $pdv$  og  $TdS$ , en Thermodynamiken fjerntliggende Tanke, ja man har ved Indførelsen af Benævnelsen »Varme« for Produktet  $TdS$  endog indskærpet dets Modsætningsforhold til Arbejdet. Tillige har den særlige

Betegnelse det Formaal i al Almindelighed at pointere en Særstilling for denne Energiart. Det er imidlertid klart, at Indførelsen af Betegnelsen »Varme« for den her optrædende ækvipotentielle thermiske Energi i sig selv ikke kan medføre nogen som helst Særstilling for den mellem de ækvipotentielle Energiformer, lige saa lidt som det thermiske Arbejde, der er udtrykt ved (6), indtager nogen Særstilling mellem de forskellige Arbejdsformer. Der er i denne Henseende en fuldstændig Parallelisme mellem Rumfangs- og thermisk Energi. Men Analogien mellem disse Energier viser sig end yderligere derved, at Entropioptagelsen — om man vil være konsekvent — kan identificeres med Arbejdsydelse paa samme tvivlsomme Maade, som dette almindeligvis sker for Rumfangsoptagelse. Hvis Entropien nemlig hentes paa reversibel Maade fra et Omraade, hvor  $T = 0$ , og som indbefattes i Systemet, er efter Formel (6) Arbejdet  $= TdS$ .

Hvad der her er udledet med Hensyn til Rumfangs- og thermisk Energi, har ganske almen Gyldighed. Et System vil ved Kvantitetsvariation optage eller afgive Energi i ækvipotentiel Form, uden at denne Optagelse eller Afgivelse for nogen Energiforms Vedkommende kan vurderes som en Form for Arbejde. Men det er naturligvis muligt, at der jævnsides saadanne Forandringer i det aabne System, naar dette er eller bringes i en Uligevægtstilstand, foregaar koblede Grundprocesser, hvorved det i Kommunikation med et Arbejdsreservoir afgiver eller modtager Arbejde.

Da Arbejde, som paa denne Maade tilføres et System, ogsaa betegnes som Energioptagelse, ses dette Udtryk at dække over væsentlig forskelligartede Fænomener. Betegnelserne potentiel og ækvipotentiel Energi er indført i det foregaaende til Haandhævelse af den ganske nødvendige Distinktion mellem disse ulige Energibegreber.

#### 4. De energetiske Grundlove.

De i de foregaaende Afsnit omtalte energetiske Grundbegreber, Kvantitet og Potential, er i de fleste Tilfælde tilgængelige for direkte Maaling, og man kan, naar dette er muligt, umiddelbart bestemme det Arbejde, som er knyttet til den energetiske Grundproces, ved en Ligning som den ved (1) fremstillede. Dette Arbejdes Talværdi afhænger naturligvis af de Enheder, hvori de to Faktorer er udtrykt, og der er gennem Maalinger af disse kun Mulighed for en indbyrdes Sammenligning af Arbejder af samme Art. Men det har ved Undersøgelse af koblede Systemer vist sig, at visse positive Koefficienter, karakteristiske for de enkelte Grundprocesser og afhængige af de benyttede Enheder, kan bestemmes saaledes, at Relationen:

$$\sum I(P_1 - P_2) \delta K = 0, \quad (9)$$

hvor  $P_1$  og  $P_2$  er henholdsvis Begyndelses- og Slutningspotentialer for den transporterede Kvantitet  $\delta K$ , og  $I$  er den ovennævnte positive Koefficient, har almen Gyldighed, uafhængigt af Fortegnsfordelingen, naar Summationen foretages over alle de i Forandringen deltagende Grundprocesser.

De i (9) indgaaende Led udtrykker ifølge det foregaaende forskellige Former af Arbejdstab eller Formindskelse i potentiel Energi, for hvilke Fortegnet er det samme som Fortegnet for den tilhørende Grundproces. Ligning (9) kan derfor siges at udtrykke den Sætning, at Arbejdssummen eller Ændringen i potentiel Energi i en reversibel eller koblet Proces er Nul. Eller at der, naar der ved den reversible Proces foregaar et Arbejdstab ved positive Grundprocesser, i alle Tilfælde viser sig en lige saa stor Arbejdsgevinst ved

negative Grundprocesser. Denne Sætning, der er Energetikens vigtigste erfaringsmæssige Grundlov, vil vi betegne som Sætningen om Arbejdsformernes numeriske og genetiske Ækvivalens, eller kortere som Arbejdsprincippet.

Det vil utvivlsomt kunne hævdes, at dette Princip er en logisk Følge af de reversible Processers og Ligevægts-tilstandes Existens, saaledes at Principets Gyldighed kan betragtes som ensbetydende med Paavisningen af disse Processers og Tilstandes Forekomst i Virkeligheden. Man benytter derfor ogsaa Arbejdsprincippets Konsekvenser, uden at Principet specielt fremhæves som Grundlag, ligesom der overhovedet ofte kan konstateres en vis Vilkaarlighed i Behandlingen af de Grundsætninger, der er af empirisk-axiomatisk Karakter.

Hvis Koblingen omfatter et Arbejdsreservoir, vil det heri vundne Arbejde, der er let tilgængeligt for Maaling, være identisk med det i Resten af Systemet foregaaende Arbejdstab. Er Processen i Arbejdsreservoirret positiv, vil Systemet modtage, er den negativ, vil det afgive eller udrette Arbejde. Tilstedeværelsen af den numeriske Ækvivalens kræver naturligvis stadig, at der for Arbejdscoeffcienten I er indført den rigtige, med (9) overensstemmende Værdi, men der er naturligvis den specielle Mulighed til Stede, at man vælger Enhederne for P og K saaledes, at I i alle Tilfælde er 1. Man har da alment for Arbejdstabet:

$$\delta A = (P_1 - P_2) \delta K, \quad (10)$$

og kan i Stedet for Ligning (9) skrive:

$$\delta A = \sum (P_1 - P_2) \delta K = 0. \quad (11)$$

Hvis en Faktor i Arbejdsudtrykket (10) ikke kan maales direkte, vil Arbejdsprincippet, idet det postuleres gyldigt ogsaa

for dette Tilfælde, kunne anvendes til Bestemmelse af vedkommende Faktor. Dette er Tilfældet for det kemiske Potentials Vedkommende. Ved etableret Kobling f. Ex. mellem en elektrisk og en kemisk Grundproces vil man saaledes kunne sætte:

$$(\psi_1 - \psi_2) \delta \varepsilon + (\mu_1 - \mu_2) \delta n = 0 \quad (12)$$

i Overensstemmelse med (11) og derved bestemme den kemiske Potentialforskel  $\mu_1 - \mu_2$  paa Grundlag af Maalinger af de 3 andre i Ligningen indgaaende Faktorer.

Det maa ved Bestemmelser af denne Art bemærkes, at energetiske Kvantiteter kan være indbyrdes uløseligt forbundne, saaledes at Forløbet af een Grundproces automatisk vil medføre Forløbet af en anden. F. Ex. vil en Stofmængde altid indeholde saavel Masse som Entropi og en kemisk Grundproces fra dette Synspunkt derfor altid kunne siges at være ledsaget af en mekanisk og en thermisk. I hvilken Grad dette vil influere paa Grundprocessens Arbejde, vil fremgaa ved nærmere Analyse af den Definition, der er lagt til Grund for det paagældende Potential.

Det er i Virkeligheden kun for det kemiske Arbejde, at der er Tale om noget Definitionsvalg, da de almindeligt anvendte indgaaende Arbejdsarter i øvrigt er definerede i Tilknytning til en umiddelbar Maalingsanordning, muliggjort ved de tilsvarende energetiske Grundprocessers Uafhængighed af hinanden. Specielt maa det her nævnes m. H. t. Bestemmelsen af det thermiske Arbejde, at en Entropimængde  $\delta S$  er maalt simpelthen ved den Temperaturstigning  $dT$ , der viser sig i et givet Standardkalorimeter ved Tilføjelse hertil af  $\delta S$ . Naar en Stofmængde flyttes fra et kemisk Potential til et andet, maa Spørgsmaalet om Betydningen af den samtidig transporterede Entropi og det

samtidig transporterede Rumfang ved eksisterende Temperatur- og Trykdifferens derimod tages i Betragtning.

Vi vil til Undersøgelse af dette Forhold betragte det Arbejde, som kan vindes i et Arbejdsreservoir ved den reversible kemiske Grundproces, d. v. s. ved den reversible Overføring af en Stofmængde fra en Tilstand I til en anden Tilstand II, idet det samlede System I + II iøvrigt naturligvis er kvantitetsafsluttet. Vi kan for Simpelt Skyld antage, at de to Tilstande fremstilles af et rent Stof, og vil tillige til Systemet medregne et omgivende Vakuüm.

Lad Tilstanden I og II være givet henholdsvis ved:

$$p_1, T_1, \mu_1, e_1, s_1, v_1$$

$$p_2, T_2, \mu_2, e_2, s_2, v_2,$$

hvor Kvantitetsstørrelserne refererer sig til Stofenheden. Til den reversible Overføring vil vi tillige benytte en Mellemtilstand III givet ved:

$$p_3, T_2, \mu_3, e_3, s_1, v_3.$$

Overføringen kan da ske ved følgende Operationer:

1. Udtagelse af Stofenheden fra System I.
2. Adiabatisk Tilstandsændring I  $\rightarrow$  III.
3. Isotherm Tilstandsændring III  $\rightarrow$  II.
4. Indførelse af Stofenheden i System II,

idet den til Proces 3 nødvendige Entropi tages fra System II. De vundne Arbejder herved vil da være henholdsvis:

1.  $p_1 v_1$
2.  $e_1 - e_3$
3.  $e_3 - e_2 + T_2 (s_2 - s_1)$
4.  $-p_2 v_2,$

altsaa ialt:

$$A = e_1 - e_2 + p_1 v_1 - p_2 v_2 + T_2 (s_2 - s_1),$$

eller, idet  $\mu = e - Ts + pv$ :

$$A = \mu_1 - \mu_2 + (T_1 - T_2)s_1. \quad (13)$$

Det er let at se, at Arbejdet, saaledes bestemt, vil være uafhængigt af Vejen, ad hvilken Overføringen foregaar. F. Ex. kunde man som Mellemtilstand III anvende:

$$p_3, T_1, \mu_3, e_3, s_2, v_3,$$

hvorved Proces 2 blev isotherm og Proces 3 adiabatisk. I Stedet for Arbejdet 2 fik man nu:  $e_1 - e_3 + T_1(s_2 - s_1)$  og i Stedet for Arbejdet 3:  $e_3 - e_2$ . Ialt altsaa:

$$A' = \mu_1 - \mu_2 + (T_1 - T_2)s_2.$$

Da Entropien  $s_2 - s_1$ , som tilføres under den isotherme Proces, imidlertid maa tages reversibelt fra Tilstand II, hvis Totalprocessen skal være uforandret ved de to betragtede Overføringsmetoder, kommer hertil yderligere Overføringsarbejdet for denne Entropimængde fra  $T_2$  til  $T_1$ , nemlig:

$$A'' = (T_2 - T_1)(s_2 - s_1).$$

Heraf følger da umiddelbart Arbejdets Uafhængighed af Vejen, idet, som man let ser:

$$A = A' + A''.$$

Naar det erindres, at den Proces, ved hvilken et Arbejde, givet ved (13), vindes i et Arbejdsreservoir, kan fuldstændig beskrives som en Stoftransport fra I til II, og tillige at denne Stoftransport omfatter Transport af den i Formel (13) indgaaende Entropimængde  $s_1$ , og at alene denne Entropimængde er blevet transporteret, idet Entropimængden  $s_2 - s_1$ , som System II afgav under den isotherme Delproces, jo genindførtes i dette System sammen med Stoffets Indførelse, saa ses Formel (13) at give et instruktivt og an-



skueligt Udtryk for Arbejdsgevinsten ved automatisk forbundne Grundprocesser: I Overensstemmelse med at Stofoverføringen er uløseligt forbunden med Entropioverføring, vil Arbejdet ved Transporten være Summen af et kemisk og et thermisk Arbejde, idet begge disse Arbejder fremstilles ved de samme Grundligninger som de, der er opstillede for de simple Grundprocesser. Kompenseres Entropioverføringen  $s_1$  ved en Overføring af samme Mængde fri Entropi fra II til I, kan man sige, at ingen Entropitransport har fundet Sted, og Arbejdet er da det rent kemiske  $\mu_1 - \mu_2$ .

Det er let at se, at disse Resultater er almene, altsaa gyldige ogsaa for de Tilfælde, hvor de kemiske Systemer, der anvendes, er mere sammensatte end de ovenfor betragtede.

Ligning (13) giver tillige meget direkte en anden Oplysning af Interesse. Entropimængden i et Stof er i den sædvanlige Thermodynamik gjort afhængig af en konventional Standardtilstand, og  $s_1$  vil derfor variere med Valget af denne Tilstand. Det samme maa da ogsaa være Tilfældet med  $\mu_1 - \mu_2$ , hvis  $T_1 \neq T_2$ . Begrebet den kemiske Potentialforskel er altsaa, saaledes som det naturligvis ogsaa kan indses ved Betragtning af elementære thermodynamiske Formler, f. Ex.:  $\left(\frac{\partial \mu}{\partial T}\right)_p = -S$ , ubestemt for Systemer af forskellig Temperatur. Dette er dog uden Betydning for Anvendelsen af vore Grundligninger, da den kemiske Overføring under disse Forhold, d. v. s. naar Temperaturen er forskellig i de to Systemer, ikke kan gennemføres isoleret, men maa ledsages af den aktive thermiske Grundproces, hvis ligeledes ubestemte Arbejdsværdi ved Tilføjelse til det kemiske Arbejde giver den definerede Sum, der optræder i (13).

Saafrømt man paa Grundlag af Thermodynamikens 3die Hovedsætning fixerer Entropiindholdet i de transporterede Stoffer, vil naturligvis ogsaa den kemiske Potentialforskel  $\mu_1 - \mu_2$  være alment fixeret.

Beregningen af Overføringsarbejdet for et Stof, der transporteres reversibelt fra et System til et andet, er her foretaget efter den klassiske Methode, som man kunde betegne som den CARNOT-VAN'T HOFF'ske Methode. Ved denne overholdes og dokumenteres Processens Reversibilitet ved Fastlæggelse af den detaljerede Mekanisme. Det herved opnaaede Udtryk (13) er forsaavidt utilfredsstillende, som den fuldkomne Symmetri, der karakteriserer de forskellige Energi- og Arbejdsstørrelser, slet ikke fremtræder heri, idet Ligningen foruden et kemisk Led vel indeholder et thermisk, der hidrører fra Entropitransport, men intet mekanisk Led, betegnende for en Rumfangstransport. Grunden hertil er den simple, at der ingen Nettotransport af Rumfang er foregaaet i denne Overføringsproces, saaledes at selve Processen er karakteriseret af Usymmetri. Den Rumfangstransport, som følger af Forskellen i Rumfang for de to Tilstande af den transporterede Stofmængde, er nemlig kompenseret af en modsat rettet Transport af frit Rumfang. Muligheden herfor er naturligvis betinget af Tilgængeligheden for direkte Udmaaling af Rumfang, knyttet til Stof, i Modsætning til Entropi, knyttet til Stof. Det er da nærmest praktiske, ikke principielle Forskelle mellem Kvantitetsstørrelserne, der er Aarsagen til Usymmetrien i Formel (13).

Opstilling af en Arbejdsformel, der tilfredsstillende det nævnte Symmetrikrav, sker rationelt paa Grundlag af de i Afsnit 3 opstillede Synspunkter, hvorefter Kvantitets-Potential-Produkterne PdK uafhængigt af deres Art alment fremstiller Energimængder, ikke Arbejds-mængder. Det vil tillige

herved være muligt at gennemføre Overføringsprocessen uden Betragtning af den detaljerede Mekanisme, som man hidtil har benyttet sig af i den kemiske Thermodynamik. Som Grundlag for Arbejdsberegningen vil vi simpelthen betjene os af den Sætning, at en hvilkensomhelst Grundproces ogsaa i et kvantitetsafsluttet System ved Kobling til et Arbejdsreservoir kan gennemføres reversibelt.

Udfra dette Synspunkt kan da Beregning af Overføringsarbejdet fra System I til System II foregaa paa følgende Maade:

1. Afgrænsning af Stofenheden indenfor System I.
2. Reversibel Ændring af den afgrænsede Stofenhed fra Tilstand I til Tilstand II, idet den hertil udkrævede Entropi og det hertil udkrævede Rumfang tages fra System II.
3. Afgrænsning af System II, saaledes at det kommer til at omfatte den saaledes ændrede Stofenhed.

Energiændringen i Totalsystemet ved disse Operationer faas ved Summation af de enkelte, i Systemerne I og II optrædende Energiændringer. Disse er følgende:

1. System I afgiver Energien  $e_1$ .
2. System II afgiver Entropien  $s_2 - s_1$  og derfor Energien  $T_2(s_2 - s_1)$ . Endvidere afgiver System II Rumfanget  $v_2 - v_1$  og vinder derved Energien  $p_2(v_2 - v_1)$ .
3. System II vinder Energien  $e_2$ .

For det totale Energitab, som i det afsluttede System er identisk med Arbejdstabet i Systemet og det vundne ydre Arbejde, faas altsaa:

$$A = e_1 - e_2 + T_2(s_2 - s_1) - p_2(v_2 - v_1)$$

eller:

$$A = \mu_1 - \mu_2 + (T_1 - T_2)s_1 - (p_1 - p_2)v_1. \quad (14)$$

Formel (14), der saaledes er udledet uden Betragtning af den specielle Omdannelsesmekanisme, tilfredsstiller den nødvendige Symmetribetingelse, idet de tre Kvantiteter, som deltager i Processen, figurerer i Formlen. Som det fremgaar af Udledningen, fremstiller den Arbejdet ved Overføring af Stofenheden fra System I til System II, idet samtidig den deri indeholdte Entropi og det dertil knyttede Rumfang transporteres. Formlen, der let kan generaliseres ved Tilføjeelse af analoge Led, saafremt der til Stofvandringen er knyttet andre Kvantiteter, viser alment, at de for de frie Kvantiteter anvendte Arbejdsudtryk bibeholder deres Gyldighed, naar Kvantiteten er knyttet til Stof.

Unddrages en af de til Stoffet knyttede Kvantiteter fra Overføring ved kompenserende Transport af den frie Kvantitet i modsat Retning, forsvinder det paagældende Led af Formel (14). Dette er, som nævnt, en let gennemførlig Sag for Rumfangets Vedkommende, hvorved (14) reduceres til (13).

Har vi at gøre med ikke-koblede eller ufuldstændigt koblede Systemer, vil Arbejdsprincippet, udtrykt ved (11), ikke mere være gyldigt, idet de energetiske Grundprocesser forløber irreversibelt og er ledsaget af Entropidannelse. Hvis vi antager, at Entropien dannes i et iøvrigt afsluttet og i Ligevægt værende Omraade af ensartet Temperatur — hvad vi i det foregaaende har betegnet som Entropireservoir — vil Entropidannelsen  $dS$  være identisk med Frembringelsen af Energimængden  $dQ$ , givet ved Relationen:

$$dQ = TdS, \quad (15)$$

hvor  $T$  er Reservoirrets Temperatur. En Energiform af denne Art betegnes som nævnt som Varme. Ligning (15) vil iøvrigt have definitions-mæssig Gyldighed, uafhængigt af Arten af

det entropioptagende System og af Oprindelsen af den deri opstaaede Entropimængde. Den vil saaledes ogsaa være gyldig for Entropitilvæksten, der hidrører fra irreversible Processer i Systemet.

Medens det for de øvrige Kvantitetsstørrelser gælder alment, at de kun kan tilføres et System ved Flytning der-til fra et andet Omraade, hvori de i Forvejen er til Stede, er det altsaa ejendommeligt for Entropien at kunne frembringes i et System af deri foregaaende energetiske Grundprocesser ved disses ukoblede Forløb. Paa tilsvarende Maade adskiller Varmen sig fra andre ækvipotentielle Energiformer. Det maa nøje iagttages, at denne Forskel paa Varme og andre ækvipotentielle Energiformer alene fremtræder ved de ukoblede Processer og fuldstændig udslettes i hele det Afsnit af Energetiken, der omhandler de reversible Tilstandsforandringer.

Undersøgelsen af de kvantitative Relationer indenfor de irreversible Grundprocessers Omraade ved MAYER, JOULE og COLDING har ført til Loven om Varmens mekaniske Ækvivalens. Som en almindelig og exakt Formulering af denne Lov kan udtales, at den til et vist Arbejdstab svarende Varmemængde er uafhængig saavel af Arten af den paagældende Grundproces som af den Temperatur, ved hvilken den paagældende Entropi dannes. Forstaaelsen af Indholdet af denne Sætning, der danner en væsentlig Del af det, der i Thermodynamiken betegnes som »Energiprincippet« eller »1ste Hovedsætning«, kræver dog en nærmere Betragtning af de principielle Synspunkter, der ligger til Grund for Varmemængdens kvantitative Bestemmelse, og Behandlingen af dette Spørgsmaal skal derfor forbeholdes et senere Kapitel.

Med Hensyn til de energetiske Grundlove for aabne

Systemer er det indlysende, at almene Love af samme indholdsmæssige Kvalitet som f. Ex. Arbejdsprincippet paa Grund af disse Systemers uafsluttede Karakter ikke vil kunne formuleres. Der vil her kun være Mulighed for den mere formelle Behandling, saaledes som kendt fra Læren om de thermodynamiske Funktioner.

Er et lukket System opbygget af to aabne Delsystemer, mellem hvilke Kvantitetstransporten i en energetisk Grundproces foregaar, og er Begyndelses- og Slutningspotentialet for den positive Kvantitet  $\delta K$  henholdsvis  $P_1$  og  $P_2$ , vil det Delsystem, der afgiver Kvantiteten, efter den ovennævnte ved Ligning (7) udtrykte Opfattelse miste Energimængden:

$$-dE_1 = -P_1 dK_1 = P_1 \delta K,$$

medens det Delsystem, der modtager Kvantiteten, modtager Energimængden:

$$dE_2 = P_2 dK_2 = P_2 \delta K.$$

Ialt er den af Systemet som Helhed afgivne Energimængde, hidrørende fra Grundprocessens Forløb:

$$-dE = -dE_1 - dE_2 = (P_1 - P_2) \delta K, \quad (16)$$

hvilket ifølge (10) er identisk med det ved den betragtede Grundproces tabte Arbejde. Er Systemet koblet til et Arbejdsreservoir, vil det heri opsamlede Arbejde  $\delta A$  være lige saa stort. Vi kan altsaa skrive:

$$-dE = \delta A.$$

Den opstillede Formel for Energertilvæksten i et aabent System fører altsaa til kvantitativ Ækvivalens mellem Energital og Arbejdstab eller lidt anderledes udtrykt: mellem Ændringerne i potentiel og ækvipotentiel Energi i et lukket System. Definitionerne af Arbejde og ækvipotentiel Energi

er altsaa, hvor en saadan Sammenligning er mulig, i indbyrdes Overensstemmelse. Vi har i Virkeligheden benyttet os heraf ved Betragtningerne over den tvungne Kobling tidligere i dette Afsnit.

### 5. Varmebegrebet.

Begrebet »Varme« og »Varmemængde« optræder i den klassiske Thermodynamik uden nogen egentlig eller fyldestgørende Definition. Saalænge det kun drejer sig om Maaling af Varmemængde ved en given Temperatur, f. Ex. af den af et givet Standardkalorimeter optagne Varmemængde, vil Bestemmelsen ikke indeholde noget principielt Problem, idet Temperaturstigningen i Kalorimeteret, naar den gøres tilstrækkelig lille, vil kunne betragtes som et Maal for denne Varmemængde. De særlig simple Forhold, der er knyttet til Maalingen af Varmemængde ved konstant Temperatur, beror paa, at Varmeoverføringen mellem to Varmereservoirir af samme Temperatur er ensbetydende med en Entropi-overføring, altsaa med den Proces, der ovenfor er blevet betegnet som en neutral Grundproces og derfor har den elementære Karakter, som i al Almindelighed er et Kendemærke for de energetiske Grundprocesser.

Hvis derimod Varmeoverføring finder Sted ved »Ledning« mellem to Kalorimetre med forskellig Temperatur, kan det Spørgsmaal tænkes at opstaa, hvorvidt den af det ene Kalorimeter afgivne Varmemængde er lig den af det andet optagne. Dette Spørgsmaal har dog naturligvis kun en Mening og kan kun underkastes en experimentel Prøvelse, naar Varmemængden er defineret uafhængig af Temperaturen. Den klassiske Thermodynamiks Behandling af Problemet bestaar imidlertid deri, at de to Varmemængder, den afgivne og den optagne, simpelthen »sættes« eller »antages at være«

lige store. Saaledes hedder det hos MAXWELL<sup>1</sup> efter Beskrivelse af Varmeledningsprocessen: »Heat, then, is something, which may be transferred from one body to another, so as to diminish the quantity of heat in the first and increase that in the second by the same amount«. Og PLANCK<sup>2</sup> skriver: »Man ... nimmt als Masz der von einem Körper bei seiner Abkühlung abgegebenen Wärmemenge diejenige Temperaturerhöhung, welche ein mit dem sich abkühlenden Körper in Berührung gebrachter Normalkörper (Wasser) erfährt, ... Zugleich setzt man dabei die von dem Körper abgegebene Wärmemenge gleich der von dem Normalkörper aufgenommenen Wärmemenge.«

Denne Antagelse kunde synes at harmonere med Opfattelsen af Varmen som en transportabel Størrelse i Lighed med de energetiske Kvantiteter. Det er dog allerede i 1ste Afsnit udtalt, at en saadan Transportabilitet ikke foreligger for de med Energidimensioner udstyrede Produkter af Kvantitet og Potential. Specielt for det foreliggende Tilfælde kan det bemærkes, at Varmeledningsprocessen i alle Tilfælde er en irreversibel Proces og som saadan knyttet til en Entropi- og ifølge (15) derfor ogsaa til en Varmefrembringelse. Der optræder altsaa ved denne Proces andre Varmefænomener end Ledningsfænomenet, og den simple Opfattelse af Varmen som en transportabel Kvantitet kan derfor ikke opretholdes.

Det er utvivlsomt, at dette særlige Fænomen i den klassiske Thermodynamik ikke er blevet betragtet under den rette Synsvinkel, og at den derved skabte Opfattelse af Varmen har været en Hindring for klar Forstaaelse af væsentlige Fænomener i Thermodynamiken.

<sup>1</sup> C. MAXWELL: Theory of Heat, 5th Ed. 6 (1877).

<sup>2</sup> M. PLANCK: Thermodynamik, 8te Aufl. 31 (1927).



En tilfredsstillende Tydning af Begrebet Varmemængde saavel som af Varmeledningsfænomenet faas paa Grundlag af den i Afsnit 2 opstillede og ved Ligning (15) formulerede Definition af Varmen som en ækvipotentiel Energiform, udtrykt ved Produktet af Entropi og Temperatur. Herefter er Varmemængden paa entydig Maade defineret som en Funktion af Entropien, idet Optagelsen af Entropimængden  $dS$  i et Kalorimeter vil være ensbetydende med Optagelsen af Varmemængden  $TdS$ , naar  $T$  er Kalorimeterets Temperatur. Ved denne Formulering er Varmen, saaledes som det er blevet betonet, sideordnet de andre ækvipotentielle Energiformer.

Det er ligeledes blevet understreget, at den potentielle thermiske Energi eller det thermiske Arbejde er ganske sideordnet andre potentielle Energi- eller Arbejdsformer. Naar der overhovedet kan foretages en principiel Adskillelse mellem thermisk Energi og andre Energiformer, er dette alene muligt paa Grundlag af de ved den irreversible Ledningsproces og andre irreversible Processer iagttagne Fænomener.

Efter det ovenstaaende vil den Mængde Varme, der bortledes fra et Varmereservoir I af Temperaturen  $T_1$ , i Overensstemmelse med (15) være:

$$\delta Q_1 = -T_1 dS_1,$$

hvor  $-dS_1$  er den tabte Entropimængde, medens:

$$\delta Q_2 = T_2 dS_2$$

med analoge Betegnelser fremstiller den af et Varmereservoir II optagne Varmemængde. Sker den thermiske Proces ved Ledning fra I til II, viser Erfaringen, at:

$$-T_1 dS_1 = T_2 dS_2.$$

Forholdet mellem de numeriske Værdier af de to Entropi-

mængder  $dS_1$  og  $dS_2$ , hvis Bestemmelse er nødvendig til Beviset for denne Identitet, faas — i Overensstemmelse med det i Afsnit 4 m. H. t. saadanne Maalinger angivne — som Forholdet mellem de Temperaturændringer  $dT_1$  og  $dT_2$ , der viser sig i et Standardkalorimeter, naar dette ved reversibel thermisk Kommunikation med Reservoirerne I og II modtager eller afgiver Entropimængder, der frembringer Temperatureffekter i disse Reservoirer, som er identiske med de ved den irreversible Ledning frembragte. Ved Maalinger af denne Art er i Virkeligheden WILLIAM THOMSON'S absolute Temperaturskala (KELVIN-Skalaen), hvorefter T i ovenstaaende Ligninger og i Grundligningen (6) bestemmes, bleven fastlagt.

Definitionen af Varmen som Produktet af Temperatur og Entropi leverer altsaa umiddelbart en Basis for den experimentelle Paavisning af den irreversible Varmetransportabilitet. Til Forstaaelse af Varmeledningens egentlige Natur, betragtet fra et energetisk Synspunkt, maa man først gøre sig klart, at den energetiske Grundproces, der her som ved enhver anden energetisk Forandring maa være en Bestanddel af det faktisk foregaaende, er en positiv Entropivandring, d. v. s. en Vandring af Entropi fra en højere til en lavere Temperatur. Er den transporterede Entropimængde  $\delta S$ , vil Omraadet I af højere Temperatur  $T_1$  miste Varmemængden  $T_1\delta S$ , medens Omraadet II af lavere Temperatur  $T_2$  vinder Varmemængden  $T_2\delta S$ . Disse Varmemængder er aabenbart ulige store. Hertil kommer imidlertid ved Entropipassagen et Arbejdstab, som efter vor Grundanskuelse er  $(T_1 - T_2)\delta S$ . Da dette paa Grund af Grundprocessens Irreversibilitet, naar Ledningsprocessen er tilendebragt, fremtræder som Varme i Omraadet II, finder man altsaa her en Varmetilvækst, som er:

$$T_2\delta S + (T_1 - T_2)\delta S = T_1\delta S, \quad (17)$$

d. v. s. en Størrelse, som er identisk med den fra Omraadet I med Temperatur  $T_1$  bortledede Varme.

Varmeledningsprocessen er altsaa af mere sammensat Natur end Kvantitetstransporten, den energetiske Grundproces. Om sidstnævnte gælder det, at den kan gennemføres paa to Maader, enten reversibelt eller irreversibelt, idet den i sidste Tilfælde er ledsaget af en Varmedannelse. Om det Fænomen, der kaldes en Varmetransport, gælder det derimod, at det i sig omfatter en irreversibel Varmedannelse, og at det derfor aldrig vil kunne ledes reversibelt saaledes som Grundprocessen eller gøres til en Bestanddel af en reversibel Proces i et lukket System. Varmetransporten er et Exempel paa, hvad ovenfor er kaldt den irreversible Totalproces. Vi kan derfor heller ikke samstemme med HELM's Udtalelse<sup>1</sup>, at »nur die Wärme kann von höherer zu niederer Intensität übergehen ohne dasz . . . überhaupt Umformung eintritt«.

Denne Betragtning er af væsentlig Betydning for Varmetransportens Karakterisation, den bringer umiddelbart Løsningen af det WILLIAM THOMSON'ske Problem om Varmeledningsprocessens Arbejdsækvivalens, et Spørgsmaal, som er berørt allerede i Indledningen og er nærmere omtalt i Begyndelsen af Afsnit 6. Dette Spørgsmaal, hvorom E. MACH udtaler, at det »auch heute (d. v. s. i 1899) keine unberechtigte ist«<sup>2</sup>, vil nærmere blive behandlet i Afsnit 7, der omhandler Varmekraftmaskinen og CLAUSIUS's Princip.

At Varmen rent fænomenologisk kan betragtes som irreversibelt transportabel medfører efter den ovenstaaende Fortolkning nødvendigvis Rigtigheden af Relationen:

$$\delta Q = (T_1 - T_2) \delta S, \quad (18)$$

<sup>1</sup> G. HELM: Die Lehre von der Energie, 63 (1887).

<sup>2</sup> E. MACH: Die Principien der Wärmelehre, 4te Aufl. 271 (1923).

som benyttedes i denne Fortolkning. Da denne Ligning fremstiller et specielt Tilfælde af den almindelige for irreversible Processer antagne Betingelse:

$$\delta Q = (P_1 - P_2) \delta K, \quad (19)$$

som, da Produktet paa højre Side udtrykker et Arbejde, er den klassiske Thermodynamiks Energiprincip, anvendt paa irreversible Processer, er det aabenbart, at den klassiske »Antagelse« om Ligestorhed af de ved Ledning afgivne og modtagne Varmemængder i Virkeligheden implicite indeholder en Forudsætning om Energiprincipets Korrekthed. Dette fører os direkte ind i vort Hovedproblem om Relationen mellem Varme og Arbejde.

## 6. Relationen mellem Varme og Arbejde.

Studiet af den irreversible Varmefrembringelse ved mekaniske og elektriske Processer har ført til det MAYER-JOULE'ske Energiprincip. Det kan, som allerede nævnt, med den i nærværende Afhandling benyttede Terminologi udtrykkes ved den Sætning, at Mængden af Varme, dannet ved forskelligartede irreversibelt forløbende Grundprocesser, er proportional med Arbejdstabet og uafhængig af den omhandlede Grundproces's særlige Art. Det maa dog straks bemærkes, at Anvendelsesomraadet for dette Princip i den sædvanlige Thermodynamik i Overensstemmelse med dets historiske Grundlag er det mekaniske og elektriske, og at specielt de kemiske og thermiske Grundprocessers Behandling i Almindelighed sker uden en direkte Indførelse af Energi-principet i den her anvendte Form. Den klassiske Thermodynamik beskæftiger sig ikke med Varmedannelsen ved den thermiske Grundproces, og hvad angaar Thermokemiens »Varmetoning«, saa har denne, som det allerede fremgaar af dens

ubestemte Fortegn, en ganske anden Betydning end den her omtalte, irreversible MAYER-JOULE'ske Varmefrembringelse. Denne forskelligartede Behandlingsmaade, som har medført en vis uheldigt virkende Usystematik, har, som vi senere skal se, sin let paaviselige Aarsag.

Den ved den mekaniske og elektriske Grundproces frembragte Varme kan let vilkaarligt lokaliseres og opsamles i Reservoirer af forskellig Temperatur. Af Principets almindelige Gyldighed for Varmedannelse i et Standardkalorimeter ved konstant Temperatur følger, paa Grundlag af Betragtningerne i forrige Kapitel, at den dannede Varmemængde ogsaa ved varierende Temperatur vil være den samme. Principet er derfor ogsaa ensbetydende med, at den ved en given irreversibel Grundproces dannede Entropimængde er omvendt proportional med den Temperatur, hvorved den dannes.

Sammenhængen mellem Entropi og Varme viser, at Entropidannelsen ved den irreversible Proces altid er ensbetydende med en Varmedannelse og omvendt. Der kunde tænkes at opstaa det Spørgsmaal, hvilken af disse Størrelser der er den bedst egnede som Maal for Irreversibiliteten af en Grundproces. I Varmens Favør som Repræsentant for Irreversibiliteten kunde henvises til, at den dannede Varmemængde er lige saa stor som Arbejdstabet  $= (P_1 - P_2) dK$ . Herved maa dog følgende Forhold tages i Betragtning: Hvis en vis Arbejdsmængde omdannes til Varme, saaledes at ved en vis Forsøgsanordning Varmen dannes i et Reservoir I af den højere Temperatur  $T_1$ , medens den ved en anden Forsøgsanordning dannes i Reservoiret II ved Temperaturen  $T_2$ , hvorved:

$$\delta Q_1 = \delta Q_2 = T_1 dS_1 = T_2 dS_2,$$

saa vil det, skønt de to Processer begge er fuldstændig

irreversible, idet alt Arbejde tabes under Varmefrembringelsen, dog kunne synes rationelt at definere den 2den Proces som »mere irreversibel« end den 1ste, idet det nemlig er muligt, naar den 1ste Proces er forløbet, at lede den dannede Entropimængde  $\delta S = dS$ , reversibelt til Reservoir II under Dannelse af Arbejdet  $(T_1 - T_2)dS_1$ . Det totale Arbejdstab bliver i saa Fald:

$$T_1 dS_1 - (T_1 - T_2) dS_1 = T_2 dS_1,$$

altsaa i Forholdet  $T_2/T_1$  mindre end Arbejdstabet  $T_2 dS_2$  ved den direkte Varmedannelse i Reservoir II.

Udfra dette Synspunkt vilde det være naturligt at anse den dannede Entropimængde for et Maal for Irreversibiliteten, saaledes at en Varmedannelse ved meget høj Temperatur, svarende til meget ringe Entropidannelse, blev betragtet som kun i ringe Grad irreversibel. Hvis to faste Legemer gnider mod hinanden, saaledes at Temperaturen paa deres Berøringspunkter stiger meget stærkt, vilde man kunne sige, at Irreversibiliteten af Gnidningsprocessen i mindre Grad var bestemt af Tilintetgørelsen af det mekaniske Arbejde end af den irreversible Udbredelse af Entropien til de tilgrænsende Dele af de gnidende Legemer med lavere Temperatur. Ganske analogt gælder f. Ex. for Entropidannelse i en elektrisk Glødetraad eller paa Berøringsfladen ved heterogene kemiske Reaktionen. Saadanne Betragtninger savner dog tilstrækkelig Præcision og omhandler mere Spørgsmaal af praktisk og terminologisk Betydning. Til det egentlige Principspørgsmaal vil der senere blive vendt tilbage.

Angaaende det ovenfor berørte Spørgsmaal om Behandlingen af den kemiske Grundproces udfra Energiprincipets Synspunkt kan det udtales, at saafremt vi postulerer dette Princip Rigtighed ogsaa for kemiske Processer, vil Varme-

udviklingen ved den irreversible kemiske Proces være givet ved:

$$\delta Q = (\mu_1 - \mu_2) \delta n,$$

hvor  $\delta n$  er den Stofmængde, der overføres irreversibelt eller ukoblet fra Potentialet  $\mu_1$  til Potentialet  $\mu_2$ . I Almindelighed vil dog denne Varmedannelse ikke udtrykkes paa simpel Maade ved en i det kemisk reagerende System optrædende Temperaturstigning, idet enhver kemisk Forandring, hvad enten den bestaar i en aktiv eller en neutral Grundproces, vil være ledsaget af Temperaturfænomener, der er uafhængige af den irreversible Varmeproduktion. F. Ex. vil den neutrale Grundproces, der bestaar i en Stofvandring fra Vædske til Damp eller omvendt, naar disse Faser er i indbyrdes Ligevægt, fremkalde en Temperaturændring, der efter Processens Retning er henholdsvis negativ eller positiv, saaledes at man, hvis Temperaturen under Fordampning og Kondensation skal holdes konstant maa henholdsvis tilføre eller fjerne Entropi. Disse Faseforandringer siges at forløbe henholdsvis under »Varmeabsorption» og »Varmeutvikling«. Men allerede det negative Fortegn, der saaledes kan vise sig for Varmeproduktionen, godtgør, som nævnt, at disse thermiske Fænomener er af ganske anden Oprindelse end den Varme, der hidtil har figureret i vore Betragtninger over den irreversible Grundproces. Skal Størrelsen af denne Varmedannelse ved den kemiske Grundproces's irreversible Forløb konstateres, maa den paagældende Reaktion først bringes til reversibelt Forløb, idet det kemiske Arbejdstab herved bestemmes ved Kobling f. Ex. med et mekanisk eller elektrisk Arbejdsreservoir. Ved det ukoblede Forløb, hvor Arbejdsgevinst af anden Art ikke bliver opnaaet som Ækvivalent for det tabte kemiske Ar-

bejde, faas da en irreversibel Varmedannelse, hvis Størrelse er identisk med Arbejdstabets.

Medens man for den mekaniske Grundproces og ligeledes med stor Tilnærmelse for den elektriske kan maale den irreversible Varmedannelse som den Varme, der fra det reagerende System maa bortføres — og f. Ex. overføres til et Varmereservoir — for at disse Grundprocesser skal forløbe isothermt, saa finder man i Henhold til ovenstaaende for den kemiske Grundproces, at den paa tilsvarende Maade bestemte Varmemængde almindeligvis er ganske forskellig fra den kemisk dannede Varme. Aarsagen er den aabenbare, at der ved kemiske Grundprocesser, der forløber i selve det stofflige Systemindre, sker Forandringer i de energetiske Potentialer, specielt Temperaturen, saaledes at Temperaturændringen ikke længere kan afgive et Maal for den Varmefrembringelse, der er en Følge af Grundprocessens Irreversibilitet. Dette Forhold er af saa dominerende Indflydelse paa de kemiske Processers energetiske Behandling, at den irreversible Varmedannelse, der fremtræder som det elementære og anskuelige Vidnesbyrd om Arbejdstabet ved mekaniske og elektriske Processer, slet ikke optræder som et selvstændigt Fænomen indenfor den kemiske Thermodynamiks Omraade. En Proces, bestaaende f. Ex. i Opløsning af Kaliumnitrat i Vand, betragtes utvivlsomt som en entropidannende Proces, men den Varmedannelse, som dog Entropidannelsen er ensbetydende med, forsvinder i den traditionelle Behandling ganske under Skjul af et andet og forskelligartet thermisk Fænomen, der betegnes som en Opløsningsprocessen ledsagende »Varmeabsorption«.

Det er en utilstrækkelig Adskillelse imellem disse to forskelligartede Varmefænomer, der har ligget til Grund f. Ex. for den urigtige thermokemiske Beregningsmaade for



den elektromotoriske Kraft af galvaniske Elementer udfra den WILLIAM THOMSON'ske Regel eller for BERTHELOT's »principe du travail maximum«. Der vil derfor være god Grund til at skelne de to Varmefænomer fra hinanden ved særlige Betegnelser, f. Ex. den energetiske Varmeudvikling i Modsætning til den thermometriske Varmetoning.

Naar Fejltagelser som de nævnte har kunnet opstaa indenfor den kemiske Thermodynamik, er Aarsagen i sidste Instans det tidligere berørte Forhold, at det kemiske Potential ikke paa samme Maade som andre energetiske Potentiale, f. Ex. Temperatur og Tryk, kan gøres til Genstand for direkte Maaling. For den thermiske Energis Vedkommende har Vanskeligheden for en adækvat Behandling rent praktisk været bestemt af Entropiens Bevægelighed, d. v. s. Materiens diatermane Karakter, medens den fra et principielt Synspunkt har bestaaet i den Tilsøring af Varmetransportens egentlige Natur, som Identiteten af den transporterede og den irreversibelt frembragte Kvantitet har været Aarsagen til.

En Analyse af Energiprincipet omfatter imidlertid ikke alene Omdannelsen af Arbejde til Varme, det Fænomen, som vi hidtil har betragtet i dette Afsnit, men i ikke mindre Grad de eksisterende Synspunkter med Hensyn til den modsatte Proces, Varmens Omdannelse til Arbejde. Muligheden for en saadan Omdannelse betragtes i Almindelighed som henhørende under »Thermodynamikens 2den Hovedsætning«, medens Lovene for Arbejdets Omdannelse til Varme betragtes som henhørende under Energiprincipet eller »1ste Hovedsætning«. Vi vil senere undersøge Hensigtsmæssigheden af dette Inddelingsprincip og i det hele af de energetiske Fænomeners Skematisering og Rubricering. Hvad det

foreløbig her drejer sig om, er en Undersøgelse af de Anskuelser, der i Tidernes Løb har udviklet sig paa dette Omraade, og en kritisk Vurdering af den Opfattelse af »Varmens bevægende Kraft«, som nu — og i Virkeligheden forlængst — har rodfæstet sig som det endelige Udtryk for denne Udvikling.

At en Proces, der udelukkende bestaar i en Omdannelse af Varme til Arbejde, ikke er mulig, er en Erfaringssætning, der kan betegnes som alment akcepteret. PLANCK<sup>1</sup> stiller saaledes følgende Sætning i Spidsen for sit Bevis for 2den Hovedsætning: »Es ist unmöglich, eine periodisch funktionierende Maschine zu konstruieren, die weiter nichts bewirkt als Hebung einer Last und Abkühlung eines Wärmereservoirs«. Paa lignende Maade er denne Grundsætning formuleret af Grundlæggerne af Thermodynamiken, CLAUDIUS og WILLIAM THOMSON, for sidstnævntes Vedkommende ved Opstilling af Postulater om Umuligheden af, hvad der senere af OSTWALD betegnedes som et »Perpetuum mobile af 2den Art«, d. v. s. en Maskine, der præsterer Arbejde paa Bekostning af Varme, uden at denne Proces er ledsaget af andre i Naturen foregaaende Forandringer<sup>2</sup>. Rigtigheden heraf er almindelig anerkendt indenfor Thermodynamiken.

Til Karakterisering af dette Postulat er der Grund til med Eftertryk at pege paa den faktiske Dobbeltthed i Postulaterets Indhold, idet dette baade formelt og reelt gaar ud ikke alene paa Umuligheden af en vis Konstruktion, men ogsaa paa Gennemførligheden af en anden. I Realiteten er der Tale om to Postulater, et 1ste Varmepostulat, der udtaler, at Varme ikke ukompenseret kan gaa over til Ar-

<sup>1</sup> M. PLANCK: Thermodynamik, 8te Aufl. 87 (1927).

<sup>2</sup> WILLIAM THOMSON: On the Dynamic Theory of Heat, Edinb. Trans. 20, 261 (1851).

bejde, og et 2det Varmepostulat, der udtaler, at en saadan Omdannelse er mulig, saafremt den forbindes med visse »kompenserende« Forandringer. Det er i særlig Grad nødvendigt at understrege denne Dobbeltthed i Varmepostulatet, som i Almindelighed i den traditionelle Fremstilling af Thermodynamiken ikke fremtræder tilstrækkelig klart formuleret.

Skal der opnaas en Kontrol m. H. t. det Bevismateriale, der føres i Marken for det, vi ovenfor har betegnet som 2det Varmepostulat, er det til Fjernelse af enhver Usikkerhed nødvendigt klart at fixere de experimentelle Betingelser, paa Basis af hvilke der kan drages Konklusioner vedrørende det nævnte Postulat. Da det drejer sig om Tilintetgørelse af Varme, Frembringelse af Arbejde samt »Kompensationer«, hvis Art indtil videre er ubestemt, vil der til et Forsøg, der tilsigter den omtalte Varmeomdannelse, kræves dels et Varmereservoir, hvorfra Varmen hentes, og som alene er ansvarligt for den for Forsøget nødvendige Varmekommunikation, dels et Arbejdsreservoir, hvori Arbejdet opsamles, samt endelig et Reaktionssystem, hvori de til Omdannelsen knyttede kompenserende Forandringer finder Sted, og som er saaledes afgrænset, at alle Forandringer, der har Relation til Varmeomdannelsen, foregaar indenfor dets Rammer. Vi bemærker, at der ved denne Afgrænsning af Reaktionssystemet er foreskrevet dette Kommunikation med Varmereservoiret gennem saadanne Dele af Systemet, som har Varmereservoirets Temperatur, da i modsat Fald Overgang af Entropi fra Varmereservoiret til Systemet vilde medføre en udenfor Systemet foregaaende Forandring, hvilket vilde være i Modstrid med den for dette fastlagte Afgrænsning.

Undersøgelser over 2det Varmepostulat vil altid kunne

foretages og vil altid være henvist til at foretages i en tredledet Forsøgsanordning som den ovenfor beskrevne. De enkelte Tilfælde adskilles kun fra hinanden ved Arten af de »kompenserende« Ledsagefænomener, der foregaar indenfor Reaktionsystemet. Disse kan enten være af thermisk Art, i hvilket Tilfælde vi har at gøre med »Varmekraftmaskinen«s Problem, hvis Løsning ved CLAUSIUS har ført til Opstillingen af Thermodynamikens 2den Hovedsætning. Vi vil reservere Behandlingen af dette Punkt for et særligt Kapitel. Eller de kompenserende Ledsagefænomener kan være af ikke-thermisk Art.

I alle Tilfælde vil det være klart, at en Frembringelse af Arbejde paa Bekostning af Varme, altsaa af en Form for ækvipotentiel Energi, vil være i Strid med de grundlæggende Synspunkter for nærværende Afhandling, idet en saadan Omdannelse ikke vil kunne finde nogen Plads i det System af Processer, der er opstillet i 1ste Afsnit som energetiske Muligheder. Det vil derfor være Opgaven for den efterfølgende Fremstilling at paavise det Fejlsyn i den klassiske Behandling af disse Problemer, der har ført til Anerkendelse af Varmens Omdannelse til Arbejde som et tilladeligt Naturfænomen.

At man ved en arbejdsydende Foranstaltning, der foreløber under Varmeoptagelse, konstaterer, at Varme »forsvinder« og saaledes, efter Energi princippet, fremtræder i Form af en ækvivalent Arbejds mængde, er en saa alment udbredt og formuleret Opfattelse, at en detaljeret Behandling af Emnet vil føre for vidt. Men det vil være naturligt at gennemføre denne Paavisning ved Betragtning af PLANCK's Behandling af den isotherme Gasexpansion, idet vi her har at gøre med en særlig autoritativ Fremstilling af en Proces, der opfattes som et overbevisende Exempel paa Varmens

Omdannelse til Arbejde. Om denne Proces skriver PLANCK følgende<sup>1</sup>:

»So findet man auch heute noch manchmal den zweiten Hauptsatz dahin charakterisiert, dass die Verwandlung von Arbeit in Wärme vollständig, die von Wärme in Arbeit dagegen nur unvollständig stattfinden könne, in der Weise, dass jedesmal, wenn ein Quantum Wärme in Arbeit verwandelt wird, zugleich notwendigerweise ein anderes Quantum Wärme eine entsprechende, als Kompensation dienende Verwandlung, z. B. Übergang von höherer in tiefere Temperatur, durchmachen müsse. Dieser Ausspruch ist in gewissen ganz speziellen Fällen richtig; ganz allgemein genommen trifft er aber durchaus nicht das Wesen der Sache, wie der Deutlichkeit halber an einem einfachen Beispiel gezeigt werden soll. Eine der allerwichtigsten mit der Entdeckung des Energieprinzips verknüpften Errungenschaften für die Wärmetheorie ist der in der Gleichung (19) (§ 70) ausgesprochene Satz, dass die gesamte innere Energie eines idealen Gases lediglich von der Temperatur abhängt und nicht vom Volumen. Lässt man nun ein ideales Gas sich unter Arbeitsleistung ausdehnen, und verhindert man die Abkühlung des Gases durch gleichzeitige Zuleitung von Wärme aus einem Wärmebehälter von höherer Temperatur, so behält das Gas mit seiner Temperatur zugleich auch seine Energie unverändert bei, und man kann sagen, dass durch den Prozess die vom Reservoir abgegebene Wärme vollständig in Arbeit verwandelt worden ist, ohne dass sonst irgendwo ein Energieumsatz stattgefunden hat. Gegen diesen Ausspruch lässt sich nicht das mindeste Tatsächliche einwenden«.

Som det fremgaar af dette Citat, er det Formaalet at vise, at der til en Overgang af Varme til Arbejde ikke

M. PLANCK: Thermodynamik, 8te Aufl. 80 (1927).

nødvendigvis er knyttet en kompenserende Transport af en anden Varmemængde fra højere til lavere Temperatur, saaledes som Tilfældet er i Varmekraftmaskinen, men at den til Disposition værende Varme meget vel fuldstændig kan nyttiggøres som Arbejde. At i Stedet for en Varmetransport her Gasexpansionen optræder som »Kompensation« for den foregaaende Varmetransformation, omtales, som det ovenstaaende Citat viser, ikke, men man maa dog forudsætte, at dette er PLANCK's Opfattelse, da Arbejdet i modsat Fald vilde være opstaaet ukompenseret, hvilket er i Strid med det alment anerkendte 1ste Varmepostulat. I den ovenfor omtalte 3-leddede Transformationsanordning udgøres Reaktionssystemet altsaa i dette Tilfælde af et Rumfang, hvori en ideal Luftart ved Hjælp af et bevægeligt Stempel er adskilt fra et Vakuum. Ved dette Reaktionssystems Funktion vil Varmemængden  $Q$  forsvinde fra Varmereservoiret og Arbejds­mængden  $A$  dannes i Arbejdsreservoiret, og man vil ved Maaling af disse to Størrelser finde, at  $Q = A$ .

Af den numeriske Identitet af  $A$  og  $Q$ , der saaledes lader sig paavise ved denne Proces, er det imidlertid utilladeligt at slutte til den genetiske Sammenhæng mellem disse Størrelser, som Paastanden om Varmens Overgang til Arbejde indeholder. Saafremt Reaktionssystemet var forblevet uforandret under Processen, vilde en saadan Paastand naturligvis kunne opretholdes, idet der da ikke vilde være nogen anden Kilde til Arbejdsdannelsen end den forsvundne Varme. Men i saa Fald vilde hele Fænomenet være i Konflikt med det utvivlsomt anerkendte 1ste Varmepostulat. Naar der derimod, saaledes som Tilfældet er, foregaar en faktisk Forandring i Reaktionssystemet, bestaaende i den som Rumfangsflytning beskrevne energetiske Grundproces — en Forandring, som PLANCK mærkværdigvis lader ganske ude

af Betragtning i sin Fortolkning af Fænomenet — saa vil Arbejdsproduktionen rent logisk betragtet lige saa vel kunne henføres til denne Forandring. Man har med andre Ord, rent logisk betragtet, to Alternativer i Forsøget paa at redegøre for det opstaaede Arbejdes Oprindelse. Dels Varmen, den ækvipotentielle Energiform, der alene erkendes at være uden Arbejdseffekt. Dels den nævnte Forandring i Reaktions-systemet, bestaaende i en Grundproces, der er af aktiv Karakter, idet den bestaar i en Rumfangsflytning fra lavere til højere Tryk, d. v. s. fra højere til lavere Rumfangspotential. Da en saadan aktiv Grundproces ved reversibelt Forløb, hvorom Talen jo er paa dette Sted, ifølge Arbejdsprincippet er tvungen til Produktion af en ækvivalent Arbejds-mængde udenfor Systemet, altsaa netop hvad man iagttager Dannelsen af, kan der imidlertid ikke herske nogen som helst Tvivl om, hvilket af de to Alternativer, der maa anvendes i Fortolkningen. Fænomenet maa tydes som en koblet Proces, hvor Arbejde præsteres i et Arbejdsreservoir paa Bekostning af en Formindskelse i potentiel Energi i det aktive Reaktionssystem.

Paa meget slaaende Maade fremgaar Rigtigheden af denne Opfattelse af det let konstaterbare Faktum, at Arbejdet, hvis Størrelse jo kan udtrykkes ved:

$$\delta A = p dv,$$

forbliver ganske uforandret, ogsaa hvis Forbindelsen med Varmereservoiret under den differentiale Proces fuldstændig afbrydes.

Vi vil iøvrigt bemærke, at samme Effekt som den, der opnaas ved den ideale Gasexpansion, kan erholdes ved talrige andre Processer. Et idealt Koncentrationselement vil saaledes præstere et elektrisk Arbejde, som er numerisk iden-

tisk med den af Elementet ved isotherm Funktion optagne Varmemængde. Vil man ogsaa her sige, at det er Varmetilførslen, der producerer det elektriske Arbejde? Fremdeles er den Omstændighed, at Temperaturen af det paagældende Reaktionssystem, saaledes som i de to nævnte Specialtilfælde, under Omdannelsen holder sig konstant, uden konklusionsmæssig Betydning, idet denne Temperaturkonstans slet ikke er benyttet i Ræsonnementet. Man vilde derfor, hvis PLANCK'S Opfattelse af Varmetransformationen ved Gasexpansionen kunde opretholdes, være berettiget til at hævde, at den ved en hvilken som helst kemisk eller fysisk Reaktion præsterede Arbejds mængde var dannet af Varme, naar blot en ækvi-valent Varmemængde under Arbejdsproduktionen pumpedes ind i Systemet.

Paa Trods af den aabenbare Uigennemførlighed af en saadan Opfattelse er den almindeligt herskende selv i anerkendte, autoritative Fremstillinger af Thermodynamiken. Det er saaledes en hyppigt optrædende Formulering af Relationen mellem Varme og Arbejde, at der ikke kan konstrueres en »periodisk virkende« Maskine, der fremstiller Arbejde af Varme, eller en Anordning, der »vedblivende« har en saadan Funktion. Den PLANCK'ske Udtalelse: »Es ist unmöglich eine periodisch wirkende Maschine zu konstruieren, die weiter nichts bewirkt, als einen Wärmehälter abzukühlen und eine Arbeit zu leisten«, akcepteres som et grundlæggende Postulat<sup>1</sup>. Eller det hedder: »Wir gelangen so zu dem Satze, dass eine Vorrichtung, welche die Wärme fortwährend in äussere Arbeit umzusetzen vermöchte, eine Unmöglichkeit ist.«<sup>2</sup> Fremhævelsen af Anordningens »perio-

<sup>1</sup> Jfr. f. Ex.: B. WEINSTEIN: Thermodynamik und Kinetik der Körper, 20, (1901). A. WASSMUTH: Müller-Pouillet's Lehrbuch der Physik, 672 (1907).

<sup>2</sup> W. NERNST: Theoretische Chemie, 5te Aufl. 17 (1907).



diske« Funktion og Processens »vedvarende« Karakter som nødvendig for Konklusionen tilkendegiver paa aabenbar Maade, at en Varmeomdannelse til Arbejde, som er undtaget fra de her fremhævede Betingelser, opfattes som en realiserbar Proces.

Overfor saadanne Anskuelser vil vi opstille Arbejdsprincippet som det alment vejledende Erfaringspostulat ved alle Arter af Arbejdsfrembringelse. Allevegne, hvor der opstaar et Arbejde, der er ledsaget af et Varmeforbrug, vil der i Overensstemmelse med 1ste Varmepostulat samtidig forløbe en »kompenserende Proces«. Men det vil vise sig, at denne kompenserende Proces altid for sig alene er i Stand til at producere det paagældende Arbejde, og at Afspærring til Varmereservoiret er uden mærkbar Virkning.

## 7. Det Clausius'ske Princip.

Det Afsnit af Thermodynamiken, i hvilket Theorien om Varmens Omdannelse til Arbejde har spillet den mest fremtrædende Rolle, og hvorved denne formentlige Energitransformation er opfattet som en særlig haandgribelig Realitet, er det, der fremstiller den saakaldte »Varmekraftmaskine«. Det var paa dette Omraade, at CARNOT gjorde sin skarp-sindige Indsats, der i væsentlig Grad vil kunne berettige, at han betragtes som Grundlæggeren af det Afsnit i Thermodynamiken, der senere betegnedes som dens 2den Hovedsætning. Men den 1ste Hovedsætning, der almindeligt betragtes som simplere, var ham ikke bekendt. Sammenknytningen af denne Sætning med det CARNOT'ske Princip skyldes CLAUDIUS. Herved opnaaede Thermodynamiken en Omformning, som Eftertiden har betragtet som af afgørende og dybtgaaende Betydning, og hvis Ide og Grundlag er forblevet uantastet indtil Nutiden.

Som Indledning til en Betragtning af den CLAUDIUS'ske Udvikling vil det være naturligt at berøre de Vanskeligheder, som paa det Tidspunkt frembød sig for en modsigelsesfri Løsning af de thermodynamiske Grundproblemer. Opgaven bestod i at forlige CARNOT's Opfattelse af »Varmestoffet«s Bevægelse fra højere til lavere Temperatur som Aarsag til det producerede Arbejde med MAYER-JOULE's Paavisning af Varmens og Arbejdets Ækvivalens. MACH udtaler herom<sup>1</sup>, at man maaske kunde have ventet, at der ikke vilde have været særlige Hindringer for en Forening af de to Principer og Sammenfatningen af dem til en konsekvent Naturopfattelse. Men, fortsætter han: »Dieser Schritt erforderte noch eine bedeutende geistige Anstrengung. Vertieft man sich in die damalige intellektuelle Situation, so begreift man dies psychologisch ganz wohl. Verhielt sich die Wärme bei der Arbeitsleistung wie das Wasser auf einer Mühle, welches nach gethaner Arbeit noch vorhanden ist, nur auf einem tiefern Niveau? Oder verhielt sich die Wärme wie die Kohle, welche beim Heizen der arbeitenden Dampfmaschine verbraucht wird? Diese beiden Auffassungen schienen sich durchaus zu widersprechen, man hielt sie für unvereinbar«. Meget klart kommer Problemet frem hos WILLIAM THOMSON i et Arbejde<sup>2</sup>, hvortil er hentydet allerede i Indledningen til denne Afhandling. Hans Standpunkt udtrykkes bl. a. i følgende karakteristiske Udtalelse om Varmeledningsprocessen: »When 'thermal agency' is thus spent in conducting heat through a solid, what becomes of the mechanical effect which it might produce? Nothing can be lost in the operations of nature — no energy can be destroyed.

<sup>1</sup> E. MACH: Principien der Wärmelehre, 4te Aufl. 269 (1923)

<sup>2</sup> WILLIAM THOMSON: An Account of CARNOT's Theory of the Motive Power of Heat, Edinb. Trans. 16, 541 (1849).

What effect then is produced in place of the mechanical effect which is lost? A perfect theory of heat imperatively demands an answer to this question; yet no answer can be given in the present state of science«.

Den CLAUSIUS'ske Løsning af det saaledes fremtrædende Problem beror paa hans Spaltning af den ved Arbejdsydelsen disponible Varmeenergi i to Dele med forskellig Funktion. For det ved Varmekraftmaskinen leverede Arbejde søges et Ækivalent i samtidig forbrugt Varme, saaledes at 1ste Hovedsætning herigennem respekteres. Da Varmen ikke umiddelbart og »ukompenseret« kan omdannes til Arbejde, er den nødvendige Kompensation knyttet til Varmeomdannelsen gennem en Vandring af Varme fra højere til lavere Temperatur. Men denne Proces indtager ikke hos CLAUSIUS Stillingen som Arbejdsdannelsens egentlige Aarsag, saaledes som Tilfældet er ved den CARNOT'ske Betragtning. Ja, det kan snarere siges, at Indførelsen af 1ste Hovedsætning og Paapegelsen af den numeriske Ækivalens i væsentlig Grad har tilsløret det, der efter CARNOT er Grundlaget for Varmens Omdannelse til »bevægende Kraft«, ligesom det paa afgørende Maade er i Modsætning til det, der efter nærværende Fremstilling er det bestemmende for al Energitransformation, nemlig den i det fungerende System tilstedeværende Potentialforskel. Selvom der ved den CLAUSIUS'ske Behandling er skabt en uangribelig Basis for thermodynamisk Beregning og Korrelation af energetiske Fænomener, saa synes den derfor ikke at kunne opfattes, principielt set, som et Fremskridt i Opfattelsen og Forstaaelsen af de thermodynamiske Grundfænomener.

Til Belysning af dette Spørgsmaal vil vi først nærmere paapege, hvad der er Grundtanken i den CLAUSIUS'ske Opfattelse. Der kan ikke være Tvivl om, at det, som CLAUSIUS

indfører som Aarsagen til og Ækvivalensen for Arbejdspræstationen i Varmekraftmaskinen, er den Varmemængde, der tilføres denne fra et Varmereservoir af konstant Temperatur, og at den Overgang af Varme, som samtidig siges at foregaa fra højere til lavere Temperatur, er opstillet som et ledsagende Fænomen, der vel er en nødvendig »Kompensation«, men ikke har nogen genetisk Betydning for Transformationen. At dette er Tilfældet, fremgaar klart af hans hele Behandlingsmaade og vises paa særlig prægnant Maade af nedenstaaende Citater. Han udtaler saaledes i sin første Afhandling om Emnet<sup>1</sup>, idet han først vender sig imod CARNOT's Anskuelse om Varmemængdens Uforanderlighed: »Wenn man annimmt, die Wärme könne, ebenso wie ein Stoff, nicht an Quantität geringer werden, so musz man auch annehmen, dasz sie sich nicht vermehren könne. Es ist aber fast unmöglich z. B. die durch Reibung verursachte Erwärmung ohne eine Vermehrung der Wärmequantität zu erklären, und durch die sorgfältigen Versuche von JOULE, bei welchen auf sehr verschiedene Weisen unter Anwendung von mechanischer Arbeit Erwärmung hervorgerufen wurde, ist auszer der Möglichkeit, die Wärmequantität überhaupt zu vermehren, auch der Satz, dasz die Menge der neu erzeugten Wärme der dazu angewandten Arbeit proportional sei, fast zur Gewiszheit geworden. Dazu kommt noch, dasz in neuerer Zeit immer noch mehr Thatsachen bekannt werden, welche dafür sprechen, dasz die Wärme nicht ein Stoff sei, sondern in einer Bewegung der kleinsten Theile der Körper bestehe. Wenn dieses richtig ist, so musz sich auf die Wärme auch der allgemeine Satz der Mechanik anwenden lassen, dasz eine vorhandene Bewegung sich in Arbeit um-

<sup>1</sup> R. CLAUDIUS: Über die bewegende Kraft der Wärme, Pogg. Ann. 79, 369, (1850).

setzen kann, und zwar so, dasz der Verlust an lebendiger Kraft der geleisteten Arbeit proportional ist«. Og i Arbejdets anden Del<sup>1</sup>: »Denn wenn wir auch eines eigenthümliches Aequivalentes für die erzeugte Arbeit nicht mehr bedürfen, nachdem wir als solches einen wirklichen Verbrauch von Wärme angenommen haben, so bleibt es doch möglich. . . . I en senere Afhandling<sup>2</sup> hedder det angaaende Sætningen om Ækvivalensen af Varme og Arbejde: »Es lässt sich Arbeit in Wärme und umgekehrt Wärme in Arbeit verwandeln, wobei stets die Grösze der einen der der anderen proportional ist.« I »Die mechanische Wärmetheorie«<sup>3</sup>: »Demgemäss können wir setzen:

$$Q_1 = Q_2 + Q,$$

und können somit in der Wärmemenge  $Q_1$ , welche der veränderliche Körper von dem Körper  $K_1$  erhalten hat, zwei Teile unterscheiden. . . . . Die eine aus dem Körper  $K_1$  stammende Wärmemenge  $Q$  ist in Arbeit verwandelt und die andere Wärmemenge  $Q_2$  ist aus dem Körper  $K_1$  in den kälteren Körper  $K_2$  übergegangen«. Og ibidem p. 80: »Diese Betrachtungsweise stimmt aber mit unseren jetzigen Anschauungen nicht überein, indem wir vielmehr annehmen, dasz zur Hervorbringung von Arbeit eine entsprechende Menge Wärme verbraucht werde, und dasz demnach die während des Kreisproceszes nach Auszen abgegebene Wärmemenge geringer sei, als die von Auszen aufgenommene. Wenn nun aber zur Hervorbringung von Arbeit Wärme verbraucht wird, so kann natürlich, mag neben dem Verbräuche von Wärme noch gleichzeitig ein Übergang einer anderen Wärmemenge

<sup>1</sup> R. CLAUSIUS, Pogg. Ann. **79**, 500 (1850).

<sup>2</sup> R. CLAUSIUS, Pogg. Ann. **93**, 482 (1854).

<sup>3</sup> R. CLAUSIUS: Die mechanische Wärmetheorie I, 75 (1876).

von einem wärmeren zu einem kälteren Körper stattfinden, oder nicht, doch keinesfalls die Rede sein, dasz die Arbeit von Nichts entstanden sei. Demnach bedurfte nicht nur der Satz, welchen CARNOT aussprochen hatte, einer Änderung, sondern es muszte auch für den Beweis eine andere Basis gesucht werden, als diejenige, auf welche CARNOT den seinen gegründet hatte«.

Den CARNOT'ske Opfattelse af Temperaturdifferensens potentielle Betydning genfinder vi saaledes ikke hos CLAUDIUS, men ser det MAYER-JOULE'ske Ækvivalensprincip indtræde som den egentlige Begrundelse for Arbejdsgevinsten ved Varmekraftmaskinens Funktion. Den kinetiske Opfattelse af Varmen som Molekylernes Bevægelsesenergi er her en afgørende Faktor. I alle efterfølgende Fremstillinger af Thermodynamiken er denne Opfattelse bibeholdt i uforandret Skikkelse.

Ved en kritisk Betragtning af den CLAUDIUS'ske Fortolkning af den kompenserede Varmeomdannelse skal det først fremhæves, at det fra et alment Synspunkt maa erklæres for uforstaaeligt, at en Varmeomdannelse, der som isoleret Proces er anerkendt umulig, skulde blive gennemførlig, blot derved, at der dertil knytttes visse sideløbende Processer, der er uden genetisk Relation til det frembragte Arbejde. At disse Processer — iøvrigt med en Betegnelse, der synes at antyde deres irrationelle Karakter — benævnes »Kompensationer«, gør ikke Forstaaelsen simplere. Et Forsøg paa en Begrundelse heraf findes da heller ikke hos CLAUDIUS, som alene beskæftiger sig med den kvantitative Ækvivalens. Men det er ikke nok med, at Tingen fra et alment Synspunkt er uforstaaelig. Hvis man nærmere undersøger den egentlige Natur af det, der indføres som »Kompensationer«, nemlig Overgangen af en Varmemængde fra en Temperatur til en anden, saa maa man erkende, at en

saadan Kompensation ved en reversibel Proces, som Varmekraftmaskinens Funktion bestaar af, er uantagelig, da, som vist i Afsnit 5, Varmetransporten omfatter en irreversibel Varmedannelse og derfor aldrig vil kunne ledes reversibelt eller gøres til en Bestanddel af en reversibel Proces, uden at denne irreversible Varme erstattes af Varme, tilført udefra. Den hertil disponible Varmemængde  $Q_1 - Q_2$  er imidlertid allerede bleven beslaglagt som Ækvivalens for Arbejdsydelsen. Det utvivlsomme Fejlsyn, der ligger til Grund for den CLAUSIUS'ske Tydning, hænger saaledes paa det nøjeste sammen med den fejlagtige Opfattelse af Varmen som en alment transportabel Kvantitet eller af Varmetransporten som et Exempel paa, hvad der i nærværende Afhandling er betegnet som energetiske Grundprocesser. At der ved den CLAUSIUS'ske Betragtningssmaaede overhovedet fremkommer en Totalproces, der har Plads i Virkelighedens Verden, beror alene paa det Forhold, at det, hvormed den udførlige Kompensationsproces sammenknyttes, er en ligeledes uigennemførlig Hovedreaktion. Det er naturligvis altid muligt rent formelt at opbygge en virkelig Reaktion af to uigennemførlige Delreaktioner, ligesom man i al Almindelighed kan korrigere een Misopfattelse ved Indførelse af en anden. Men en saadan Operation kan ikke akcepteres som en Tydning eller Fortolkning, fører ikke til nogen Forstaaelse, indeholder intet Fremskridt.

Udfra de i nærværende Afhandling udviklede Synspunkter for Arbejdsproduktion, specielt Opfattelsen af de energetiske Grundreaktioner som alle ligestillede i denne Henseende, volder Tydningen af Varmekraftprocessens Mechanisme ingensomhelst Vanskeligheder. Det thermiske Arbejde, hvormom Talen er her, kræver til sin Udfoldelse to Omraader, som vi betegner som Varme- eller Entropi-

reservoirer, med forskellig Temperatur. Den positive thermiske Grundproces består i en Transport af Entropi fra det ene Reservoir I med den højere Temperatur  $T_1$  til det andet Reservoir II med den lavere Temperatur  $T_2$ . Det thermiske Arbejdstab er herved efter Formel (6), naar  $\delta S$  er den transporterede Entropimængde:

$$(T_1 - T_2) \delta S,$$

og ligesaa stort er ifølge Arbejdsprincippet den ved tilstedeværende Kobling i et Arbejdsreservoir opsamlede Arbejds- mængde. Altsaa faas:

$$\delta A = (T_1 - T_2) \delta S. \quad (20)$$

Denne Formel, der er i Overensstemmelse med Grund- ligningen (1) og saaledes hviler paa et alment, simpelt og modsigelsesfrit Grundlag, er, som man let ser, i fuldstændig Overensstemmelse med det CLAUDIUS'ske Udtryk, hvis kvan- titative Korrekthed er fastslaaet gennem et meget stort For- søgsmateriale, nemlig:

$$\delta A = \frac{T_1 - T_2}{T_1} \delta Q_1, \quad (21)$$

idet vi til Paavisning af disse Ligningers Identitet kun be- høver at bemærke, at ifølge (15) den fra Varmereservoir I bortledelede Varme er givet ved:

$$\delta Q_1 = T_1 \delta S.$$

Skulde man indpasse den her skildrede Arbejdsme- kanisme i den i foregaaende Kapitel beskrevne treleddede An- ordning, der i al Almindelighed maa anvendes til Under- søgelse over Muligheden for Varmens Omdannelse til Ar- bejde, saa vilde det af Varmereservoirerne I og II dannede System, indenfor hvilket Entropitransporten foregaar, na- turligvis være det der optrædende »Reaktionssystem«. Da



dette, som vi nu har set, alene yder det i Arbejdsreservoiret opsamlede Arbejde, vil det som Varmekilde i det treleddede System optrædende Varme- eller Entropireservoir, hvis Funktion alene bestaar i en Afgivelse af Varme eller Entropi til Dele af Reaktionssystemet, som har Reservoirets Temperatur, være uden Betydning for Systemets Arbejdspræstation. Det er da ogsaa af denne Grund udelukket ved Virksomheden i Varmekraftmaskinen at finde nogetsomhelst Holdepunkt for Muligheden af en Overgang af Varme til Arbejde.

Til Belysning af Uholdbarheden af den ved CLAUDIUS fastslaaede Opfattelse vil det være nyttigt at fremholde Konsekvensen for denne af den fuldkomne Analogi og Symmetri, som karakteriserer Udtrykkene for Arbejdsdannelsen ved alle aktive koblede Grundprocesser. Den med (20) analoge, almindelige Ligning:

$$\delta A = (P_1 - P_2) \delta K$$

kan jo nemlig for alle Arbejdsarter ved Indførelse af:

$$dE = PdK$$

omskrives til:

$$\delta A = \frac{P_1 - P_2}{P_1} \delta E_1, \quad (22)$$

hvor E er den til Potentialet P og Kvantiteten K svarende ækvipotientielle Energiart. Af denne Ligning, som er fuldstændig analog med den CLAUDIUS'ske (21), kunde man da med samme Ret som for dennes Vedkommende uddrage den Slutning, at af en foreliggende Mængde ækvipotentiel Energi af vilkaarlig Art  $dE_1 = P_1 dK$  en vis Brøkdel kan omdannes til Arbejde, naar en anden Brøkdel transporteres til et lavere Potential, efter Skemaet:

$$\delta E_1(P_1) \begin{cases} \nearrow (P_1 - P_2) \delta K = \frac{P_1 - P_2}{P_1} \delta E_1 \rightarrow \delta A \\ \searrow P_2 \delta K = \frac{P_2}{P_1} \delta E_1 \rightarrow \delta E_2(P_2). \end{cases}$$

Herefter vilde man f. Ex. ved Transport af elektrisk Energi fra et elektrisk Potential til et lavere kunne faa en anden Mængde elektrisk Energi omdannet til Arbejde.

Naturligvis er et saadant Regnestykke numerisk korrekt og i Overensstemmelse med den kvantitative Ækivalens mellem Energitab og Arbejdspræstation, som er paapeget i Slutningen af 4de Afsnit af nærværende Afhandling. Det skal ogsaa nævnes, at enkelte Forfattere har fremhævet den Analogi, der saaledes kan hævdes at foreligge mellem den CLAUSIUS'ske Behandling af det thermiske Arbejde og andre Arbejdsformer<sup>1</sup>. En Generalisation af Formel (22) til Omfatning af Arbejdsdannelse i al Almindelighed vil dog utvivlsomt bidrage til at klargøre det ubrugbare i en Opfattelse, der søger at sammensætte en realiserbar Proces af uigennemførlige Delprocesser.

Der skal endelig peges paa de Forhold — af rent praktisk Natur — der har gjort Opstillingen af den CLAUSIUS'ske Tydning af Varmekraftmaskinens Funktion og den ubetingede Tilslutning, den hidtil har nydt, psykologisk forstaaelig. Det tekniske Problem at frembringe Arbejde ad thermisk Vej bestaar jo ved en idealt fungerende Mekanisme ikke i Tilvejebringelsen en Gang for alle af Varmereservoirer med forskellig Temperatur, men i Opretholdelsen af den paa-gældende Temperaturtilstand i de indgaaende Reservoirer under Systemets vedvarende Funktion. Tilførsel af Entropi

<sup>1</sup> E. MACH: Die Geschichte und die Wurzel des Satzes von der Erhaltung der Arbeit, 54 (1909). Principien der Wärmelehre, 329 (1923). W. MEYERHOFFER, Z. f. physik. Chemie 7, 544 (1891).

eller Varme til Reservoiret af høj Temperatur kan betegnes som den første af de Operationer, der er uomgængelig til Opretholdelsen af Varmekraftmaskinens stationære Arbejdstilstand, og Betydningen heraf har givet sig Udtryk i Opstilling af Begrebet »Nyttevirkning« som Forholdet imellem det udrettede Arbejde og den Varmemængde, der til dette Formaal maa tilføres Reservoiret I:

$$\eta = \frac{A}{Q_1}. \quad (23)$$

Som allerede CARNOT med stor Klarhed udtalte, bestaar imidlertid den tekniske Opgave i ikke mindre Grad i Fjernelsen af den i Reservoiret II opsamlede Varme, saaledes at en »Nyttevirkning«, udtrykt ved:

$$\eta = \frac{A}{Q_2}, \quad (24)$$

principielt set med samme Ret vilde kunne indføres som et Maal for den tekniske Effektivitet. Naar alligevel alene den ved (23) udtrykte Koefficient og den Varmemængde, som skal tilføres Systemet, er blevet tillagt Betydning, beror det paa det rent praktiske Forhold, at Temperaturen  $T_1$  altid er højere end Jordoverfladens normale Temperatur, saaledes at der til Opretholdelse af  $T_1$  maa tilføres Varme, f. Ex. ved Hjælp af exotherme kemiske Processer, medens tilsvarende Foranstaltninger ikke udkræves af Temperaturen  $T_2$ , da denne er Jordoverfladens. Hvis man som højere Temperatur i Varmekraftmaskinen benyttede Jordoverfladens Temperatur, vilde det tekniske Problem ikke bestaa i Tilvejebringelse af  $Q_1$ , men i Fjernelse af  $Q_2$  ved Hjælp af endotherme Processer, i hvilket Tilfælde der næppe vilde være

den samme Betingelse til Stede for Opfattelsen af Varmen som en til andre Energiformer forvandlelig Energiart<sup>1</sup>.

Nærmest ved dette Synspunkt af alle tidligere Forskere er i Virkeligheden CARNOT. OSTWALD<sup>2</sup> har med Rette gjort opmærksom paa, at Betegnelsen »calorique«, der almindeligt identificeres med »Varmestof«, af CARNOT stedse anvendes, hvor Varmens bevægende Kraft sammenlignes med den bevægende Kraft af et Vandfald, medens han ved almindeligere Betragtninger anvender Ordet »chaleur«. Dette Forhold kunde tyde paa en rent intuitiv Opfattelse af en væsentlig Forskel i disse to thermiske Betegnelser. Der kræves aabenbart kun en Ombytning af Begrebet »calorique« med Begrebet Entropi, for at den CARNOT'ske Behandling af Varme-Arbejds-Relationen skal opnaa en fuldstændig principiel Overensstemmelse med nærværende Fremstillings.

Om der ved de i CARNOT's efterladte Dagbog<sup>3</sup> anførte Betragtninger over Varmens mekaniske Ækvivalent er medført en Ændring i den oprindelige Opfattelse af Varmekraft-maskinens Funktion, kan ikke her afgøres.

Læren om den kompenserede Varmetransformation, udviklet af CLAUSIUS og fremstillet i beslægtet Formulering af WILLIAM THOMSON, har som nævnt fuldkommen præget Eftertidens thermodynamiske Opfattelse. Som et enkelt Exempel herpaa skal citeres en Udtalelse i MAXWELL's bekendte Værk om Varmens Theori<sup>4</sup>: »It is impossible, by the unaided action of natural processes, to transform any part of the heat of a body into mechanical work, except by

<sup>1</sup> Jfr. J. N. BRØNSTED: Lærebog i fysisk Kemi, 29 (1936).

<sup>2</sup> W. OSTWALD: Ostwald's Klassiker der exakten Wissenschaften Nr. 37, 70 (1909).

<sup>3</sup> Jfr. E. MACH: Principien der Wärmelehre, 242 (1923), F. ROSENBERGER: Geschichte der Physik, 229 (1887—1890).

<sup>4</sup> C. MAXWELL: Theory of Heat, 5th Ed. 153 (1877).

allowing heat to pass from that body into another at a lower temperature«. Det vil dog paa Grundlag af det i dette og det foregaaende Afsnit fremførte formentlig være berettiget at hævde, at der i den traditionelle Thermodynamiks Behandling af Spørgsmaalet om Varmens betingede Omdannelse til Arbejde ikke kan paapeges noget overbevisende Argument for Muligheden af en saadan Proces, men at tværtimod Antagelsen af dens ubetingede Uigennemførlighed fremgaar som en tvangfri og harmonisk Fortolkning af Energetikens Erfaringsmateriale.

---

Det maa ventes, at det ovenfor meddelte ikke vil være uden Indflydelse paa Formuleringen af de almene Postulater, der er Indholdet af selve de thermodynamiske Hovedsætninger. Det drejer sig i Virkeligheden herved i Hovedsagen om en Skematisering af allerede vundne Resultater. Da der ved Fremstillingen heraf ogsaa maa inddrages under Betragtningen visse Synspunkter, der fra anden Side er opstillet for Hovedsætningernes Formulering paa et fra den klassiske Thermodynamiks væsensforskelligt Grundlag, vil det imidlertid være formaalstjenligt at give disse Problemer en samlet Behandling i et efterfølgende Arbejde.

---

## INDHOLD

	Side
1. Indledning.....	3
2. De energetiske Grundfænomener.....	6
3. Kvantitetsvariable Systemer.....	13
4. De energetiske Grundlove.....	18
5. Varmebegrebet.....	29
6. Relationen mellem Varme og Arbejde.....	34
7. Det CLAUDIUS'ske Princip.....	47

Det Kgl. Danske Videnskabernes Selskab.

Mathematisk-fysiske Meddelelser. **XV**, 5.

---

# ÜBER POTENTIALUNTERSCHIEDE IN MEMBRANHALTIGEN SYSTEMEN

VON

AUGUSTA UNMACK



KØBENHAVN

LEVIN & MUNKSGAARD

EJNAR MUNKSGAARD

1937

Printed in Denmark.  
Bianco Lunos Bogtrykkeri A/S.



## Übersicht.

Es ist schon seit langem bekannt, dass das Diffusionspotential zwischen zwei Lösungen durch die Zwischenschaltung einer Membran geändert werden kann. Das Ergebnis ist besonders von MICHAELIS untersucht worden, der gezeigt hat, dass die Wirkung der Membran dadurch erklärt werden kann, dass die Überführungszahlen der Ionen dadurch beeinflusst werden, dass die Ionen sich nicht in einem ausgedehnten Raum, sondern in den engen Poren der Membran bewegen. Durch den Nachweis, dass die hervorgerufene Änderung des Potentials von der Orientierung des Systems zu der Richtung der Schwerkraft abhängig ist, hat BRAUNER die Beobachtungen MICHAELIS's ergänzt. Eine zufriedenstellende Erklärung dieses sogenannten geoelektrischen Effektes hat BRAUNER nicht gegeben.

In der vorliegenden Arbeit ist gezeigt worden, dass der geoelektrische Effekt dadurch entsteht, dass die von MICHAELIS beobachtete Membranwirkung nach und nach deshalb abnimmt, weil eine immer stärkere Verlegung des Konzentrationsgefälles von der Membran nach der Lösung stattfindet, und die Geschwindigkeit dieser Verlegung von der Dialyse-richtung abhängt. Ist die Dialyse-richtung eine solche, dass der hindurchdialysierende Stoff nicht nur durch Diffusion, sondern auch durch Konvektionsströme entfernt wird, so wird die ursprüngliche Konzentrationsverteilung mit dem Konzentrationsgefälle ganz in der Membran besser aufrechterhalten werden, als wenn es keine Konvektionsströme

gibt. Diese Erklärung wird durch den Nachweis eines ähnlichen Schwerkrafteffektes auf die Dialysegeschwindigkeit durch die Membran, sowie durch den Nachweis der Abhängigkeit der Abnahme des Potentials von dem spezifischen Gewicht der Lösungen, gestützt.

Findet die Entfernung des Stoffes nur durch Diffusion statt, so können die gewöhnlichen Diffusionsgesetze angewandt werden. Eine Berechnung des Potentials für irgend eine durch Spülung hervorgerufene, stationäre Konzentrationsverteilung wird dadurch möglich. Für den Fall, dass die Konzentrationsverteilung nicht eine stationäre ist, bietet die exakte mathematische Behandlung leider so grosse Schwierigkeiten, dass auf die Durchführung einer solchen hier verzichtet worden ist. Für grosse Versuchszeiten ist eine annähernde Berechnung möglich. Die Übereinstimmung mit den Versuchsdaten ist befriedigend.

Die grösste Änderung des Potentials durch die Zwischenschaltung einer Membran findet man für undurchlässige Membranen. Für gut durchlässige Membranen ist die Membranwirkung immer klein. Für eine gegebene Membran ist die Wirkung für verdünnte Lösungen am grössten.

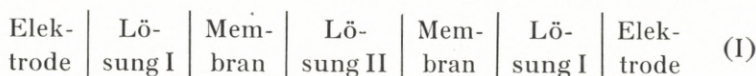
Der geoelektrische Effekt muss immer kleiner als die maximale Membranwirkung sein. Für gut durchlässige Membranen ist er deshalb klein. Für sehr undurchlässige Membranen muss dieser Effekt aber auch klein sein, und zwar deshalb, weil der Unterschied zwischen auf- und abwärtsgerichteter Dialyse belanglos wird, wenn die Menge des hindurchdialysierenden Stoffes minimal wird. Der geoelektrische Effekt erreicht deshalb seine grössten Werte für Membranen von einer gewissen mittelgrossen Durchlässigkeit.

Die Bedeutung der besprochenen Effekte für Potentialmessungen in membranhaltigen Systemen ist an drei Bei-

spielen diskutiert worden: 1. Messungen von DONNAN-Potentialen; 2. Potentialmessungen von galvanischen Ketten mit eingeschaltetem Agar-Heber; 3. Versuche zur experimentellen Realisation der Verhältnisse, für welche die PLANCK'sche Formel für das Diffusionspotential zwischen zwei Lösungen gültig ist.

### 1. Brauner's Messungen des geoelektrischen Effektes.

In den Jahren 1926—30 hat der Pflanzenphysiolog BRAUNER<sup>1</sup> eine Reihe von Arbeiten veröffentlicht, in denen er sich mit einem von ihm entdeckten Effekt, dem sogenannten geoelektrischen Effekt, beschäftigt. Der Effekt zeigt sich dadurch, dass die *EMK* eines Elements von dem Typus:



davon abhängig ist, wie das Element zur Richtung der Schwerkraft orientiert ist. Ist die Richtung der Dialyse durch beide Membranen horizontal, so wird die *EMK* ungefähr gleich Null sein. Ist dagegen die Dialyse durch die eine Membran aufwärts-, durch die andere abwärtsgerichtet, so erhält man eine *EMK*, deren Grösse von der Art der Membran und der Lösungen abhängt. Den Unterschied zwischen den Potentialwerten bei senkrechter und bei wagerechter Dialyse bezeichnet BRAUNER als das geoelektrische Potential (*GEP*).

Wenn eine *EMK* überhaupt in einem System wie dem obengenannten entstehen kann, muss es daran liegen, dass die beiden Membranpotentiale der Kette verschieden sind. BRAUNER und AMLONG<sup>2</sup> zeigten, dass dies wirklich der

<sup>1</sup> BRAUNER, L., Kolloidchem. Beih. **23** (1926) 143; Jahrb. f. wiss. Bot. **66** (1927) 381; **68** (1928) 711; **73** (1930) 513.

<sup>2</sup> BRAUNER, L., und AMLONG, H. U., Protoplasma **20** (1933) 279.

Fall ist. Sie bestimmten u. a. das Membranpotential des Systems:



sowohl für wagerecht- als für aufwärts- und abwärtsgerichtete Dialyse. Sie fanden dadurch drei verschiedene Werte (mit  $E(\rightarrow)$ ,  $E(\uparrow)$  und  $E(\downarrow)$  bezeichnet), die alle drei von dem Werte abwichen, den das Diffusionspotential annimmt, wenn die beiden Lösungen ohne Zwischenschaltung einer Membran sich berühren. Ausserdem fanden sie, wie man erwarten sollte, dass  $E(\uparrow) + E(\downarrow)$  für das System II gleich dem *GEP* für das System:



war.

Dass die Zwischenschaltung einer Membran im allgemeinen das Diffusionspotential zwischen zwei Lösungen ändert, war schon bekannt.<sup>1</sup> Durch Untersuchungen mit sehr undurchlässigen Membranen war es MICHAELIS und seinen Mitarbeitern<sup>2</sup> gelungen zu zeigen, dass diese Wirkung einer Membran folgenderweise erklärt werden kann:

Wenn zwei Lösungen sich direkt berühren, wird die Grösse des Diffusionspotentials wie bekannt durch die Konzentrationen, Wertigkeiten und Überführungszahlen aller anwesenden Ionen bestimmt. Für den einfachsten Fall: zwei verschieden konzentrierte Lösungen eines und des-

<sup>1</sup> CYBULSKI, N. und DUNIN-BORKOWSKI, J., Bull. int. de l'Acad. des Sciences de Cracovie **12** (1909) 660; MICHAELIS, L. und FUJITA, A., Biochem. Zs. **158** (1925) 28; **161** (1925) 47; **164** (1925) 23; FUJITA, A., ibid. **158** (1925) 11; **159** (1925) 370; **162** (1925) 245; MICHAELIS, L. und DOKAN, SH., ibid. **162** (1925) 258.

<sup>2</sup> Journ. Gen. Physiol. **10** (1927) 575; 671; 685.

selben mono-monovalenten Elektrolyten, ist das Diffusionspotential,  $E_D$ , durch

$$E_D = \frac{RT}{F \log e} (1 - 2 n_-) \log \frac{c_1}{c_2} \quad (1)$$

gegeben, wo  $n_-$  die Überführungszahl des Anions,  $c_1$  und  $c_2$  die Konzentrationen der Lösungen angeben. Die Formel setzt voraus, dass es zulässig ist mit Konzentrationen statt mit Aktivitäten zu rechnen, sowie auch dass die Überführungszahl unabhängig von der Konzentration ist. Dagegen ist keine Voraussetzung über die Art der Zwischenschicht gemacht (ob Diffusions- oder Mischungsschicht).

Wenn nun die Zwischenschaltung einer Membran eine Änderung des Potentials bewirkt, kann die Ursache die sein, dass die Überführungszahlen innerhalb der Membran anders sind als in membranfreier Lösung. Da das Membranmaterial gewöhnlich die beiden Ionen in verschiedenem Grade adsorbiert, ist es auch recht wahrscheinlich, dass die Wanderungsgeschwindigkeiten der beiden Ionen von der Nähe des Membranmaterials verschieden beeinflusst werden können. Die Ionenart, die am stärksten adsorbiert wird, wird auch am meisten in ihrer Bewegung gehemmt, und die Überführungszahl dieser Ionenart wird dadurch vermindert. Der für das Potential gefundene Wert,  $E_M$ , gestattet nach MICHAELIS die Berechnung eines Mittelwertes der Überführungszahl des Anions in der Membran,  $n_-'$ , nach der Formel:

$$E_M = \frac{RT}{F \log e} (1 - 2 n_-') \log \frac{c_1}{c_2} \quad (2)$$

Die Gültigkeit dieser Formel für die genannten sehr undurchlässigen Membranen wurde von MICHAELIS durch Ver-

gleich des nach (2) berechneten Wertes von  $n_-'$  mit direkt bestimmten Werten gezeigt.

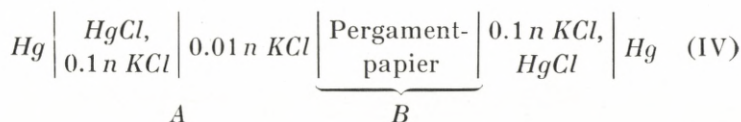
BRAUNER und AMLONG versuchten nun ihre Beobachtungen mittels ähnlicher Überlegungen zu erklären. Sie erhielten dadurch das merkwürdige Resultat, dass die Überführungszahl des Chloridions in *KCl* innerhalb einer Pergamentpapiermembran, davon beträchtlich abhängig sein sollte, ob das *KCl* wagerecht, aufwärts oder abwärts dialysiere. Dieses Ergebnis ist so unwahrscheinlich, dass es von Interesse schien, diesen Effekt näher zu untersuchen.

## 2. Eigene Messungen des geoelektrischen Effektes.

Einige orientierende Messungen mit einem System vom Typus II ergaben, dass das Potential unmittelbar nach einer Drehung des Systems eine schnelle Änderung zeigte. Die Änderung pro Minute wurde nach und nach kleiner, und es sah aus, als ob für jede Stellung der *E*-Wert einem für diese Stellung charakteristischen Grenzwert zustrebte. Dieses Ergebnis ist in völliger Übereinstimmung mit den Beobachtungen von BRAUNER und AMLONG. Sie sahen die Werte, die nach 20—30 Minuten erreicht worden waren, für die Grenzwerte an, und hielten den Fehler, den sie durch diese Annahme machten, für so unbedeutend, dass sie diese Werte als Grundlage ihrer Überlegungen benutzten. Wurden die Messungen aber über wesentlich längere Zeiten (mehrere Stunden bis einen Tag) fortgesetzt, so stellte es sich heraus, dass die *E*-Werte der drei verschiedenen Stellungen des Systems sich nicht je einem Grenzwert näherten, sondern dass in allen drei Fällen ein gemeinsamer Grenzwert erreicht wurde, und zwar der Wert des Diffusionspotentials eines membranfreien Systems. Es wurde auch beobachtet, dass der Unterschied zwischen  $E(\downarrow)$  und  $E(\uparrow)$  immer kleiner

wurde, wenn man ohne Erneuerung der Lösungen abwechselnd diese beiden  $E$ -Werte bestimmte. Das Verschwinden aller Unterschiede bei längeren Versuchszeiten muss unzweifelhaft der konzentrationsausgleichenden Wirksamkeit der Dialyse zugeschrieben werden. Um die Versuchsverhältnisse genau zu kennen, muss man deshalb seine Apparatur derart konstruieren, dass man während der Messung beide Seiten der Membran bespülen kann.

Die Messungen beschränkten sich im wesentlichen auf das System:



Diese Kette enthält zwei Grenzpotentiale: bei  $A$  das Diffusionspotential bei direkter Berührung der Lösungen ( $E_A$ ), und bei  $B$  das Membranpotential bei zwischengeschalteter Membran ( $E_B$ ). Die gemessene  $EMK$  wird durch  $E_{AB} = E_B - E_A$  gegeben.

Die Versuchsanordnung zeigt Fig. 1. Die Zwischenschicht bei  $A$  wird in der Weise hergestellt, dass man die  $0.1 n \text{ KCl}$ -Lösung in das Rohr  $A'A''$  aufsteigen lässt bis etwas oberhalb des Hahns. Der Hahn wird nun geschlossen, das Rohr oberhalb des Hahns wird gereinigt, mit der  $0.01 n \text{ KCl}$ -Lösung gefüllt, und zuletzt mit der übrigen Apparatur verbunden. Am Anfang einer Messung wird der Hahn geöffnet und während der ganzen Messserie offen gelassen. Durch Diffusion entsteht zwischen den beiden Lösungen eine Diffusionsschicht, die sich immer weiter verbreitet. Der Wert des Diffusionspotentials ist von der Ausdehnung dieser Diffusionsschicht unabhängig, und das Potential nimmt daher beinahe momentan seinen Endwert an. Für  $0.1 n \text{ KCl}$

und  $0.01\ n\ KCl$  ist dieser etwa  $0.3\ mV$ . Der Wert von  $E_B$  berechnet sich demnach zu  $E_B = E_{AB} + 0.3\ mV$ .

Der Teil der Apparatur, in welchem die Grenze bei  $B$  gebildet wird, ist in grösserem Massstabe in Fig. 1a dargestellt. Die Membran  $M$  ist zwischen den beiden gleichen Hälften des Hartgummirohres  $E'E''$  festgeklemmt. Der innere Durchmesser des Rohres war  $10\ mm$ , der Abstand der Membran

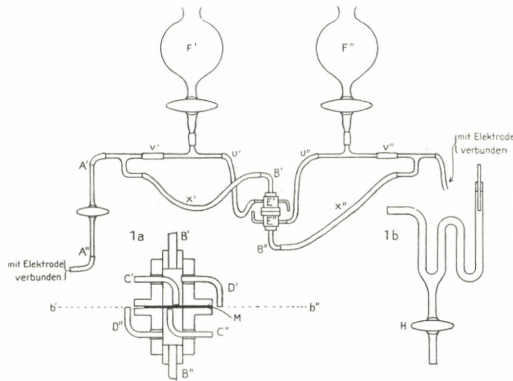


Fig. 1. Apparatur für Messung von Membranpotentialen.

von beiden Enden des Rohres war etwa  $25\ mm$ . Die Lösungen konnten bei  $C'$ , resp.  $C''$ , ein-, bei  $D'$ , resp.  $D''$ , abgeleitet werden. Die Flüssigkeit in dem Rohr war mit den Elektroden durch  $B'x'A'A''$ , bzw.  $B''x''$ , leitend verbunden. Das Hartgummirohr konnte um die wagerechte Achse  $b'b''$  gedreht werden.

Es stellte sich heraus, dass das gemessene Potential davon abhängig war, wie stark die Membranseiten gespült wurden. Bei langsamer Spülung nahm das Potential während der Messung ab, bei kräftiger Spülung wurden konstante Potentiale erhalten, die jedoch von den Strömungsgeschwindigkeiten beider Lösungen abhängig waren. Die Ergebnisse einer Serie Untersuchungen über die Bedeutung der Strömungsgeschwindigkeiten sind in der Tabelle 1 zu-



Tabelle 1.

Einfluss der Strömungsgeschwindigkeit auf das Diffusionspotential.

System: 0.1 n KCl | Pergamentpapier | 0.01 n KCl

	cm <sup>3</sup> pro Minute		$E_B =$ $E_{AB} + 0.3 \text{ mV}$
	0.01 n KCl	0.1 n KCl	
0.1 n KCl unten	39	46	8.2
	40	0	8.2
	14.7	12.3	5.9
	14.3	0	6.3
	11.4	0	5.5
	7.4	0	4.9
0.1 n KCl oben	43	30	7.8
	43	0	9.2
	33	20	8.0
	34	0	8.2
	22	15	6.7
	11	12	6.0
	5	3	5.5

sammengestellt. Man bemerkt, dass namentlich die Geschwindigkeit der 0.01 n KCl-Lösung von Bedeutung ist. Bei genügend kräftiger Spülung ist der  $E$ -Wert davon unabhängig, ob 0.1 n KCl oben oder unten ist. Geschwindigkeiten von mehr als 40 cm<sup>3</sup> pro Minute bewirken keine Zunahme des Potentials. Es liegt dann nahe anzunehmen, dass diese Strömungsgeschwindigkeit genügt, um die ursprünglichen Konzentrationen auf beiden Seiten der Membran aufrechtzuerhalten, so dass das ganze Konzentrationsgefälle innerhalb der Membran liegt. Der gefundene  $E$ -Wert sollte danach gleich  $E_M$  sein (siehe doch Seite 28).

Wird die Spülung gleichzeitig an beiden Seiten der Membran abgestellt, so nimmt das Potential sofort ab. Fig. 2 zeigt die Änderung von  $E_B$  für alle drei Stellungen des untersuchten Systems (Kurve  $A_1$  und  $A_2$ ). In allen drei Fällen

ist die Abnahme während der ersten halben Minute sehr stark. Die Wirkung der Dialyse zeigt sich also sofort. Die Kurven für wagerecht- und für abwärtsgerichtete Dialyse fallen zusammen. In diesen beiden Stellungen des Systems dauert die starke Abnahme nur kurz. Schon 1 Minute nach der Abstellung der Spülung wird die Abnahme sehr langsam (etwa  $1\text{ mV}$  in 30 Minuten). Für aufwärtsgerichtete Dialyse wird die Abnahme ebenfalls nach einer halben Minute

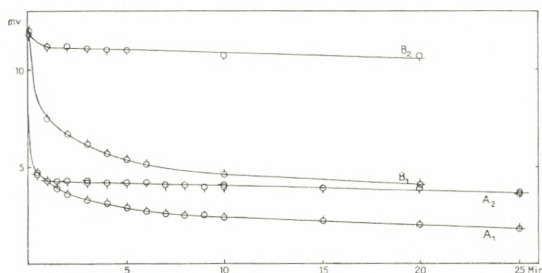


Fig. 2. Die zeitliche Änderung des Potentials im Systeme:  
0.01  $n$   $KCl$ -Pergamentpapier-0.1  $n$   $KCl$

○: 0.1  $n$   $KCl$  unten; ◻: 0.1  $n$   $KCl$  oben; ○ wagerechte Dialyse.

langsamer, behält jedoch zunächst einen wesentlich grösseren Wert als in den beiden anderen Fällen. Erst nach 15 Minuten ist die Abnahme so langsam geworden, dass die beiden Kurven ungefähr parallel verlaufen.

Zum Vergleich sind die Resultate vom BRAUNER und AMLONG für dasselbe System eingezeichnet (Kurve  $B_1$  und  $B_2$ ). Der Abstand zwischen den beiden Kurven ist hier beträchtlich grösser<sup>1</sup>, aber der Charakter der Kurven ist der-

<sup>1</sup> Die Ursache ist jedenfalls zum Teil in einem Unterschied der Permeabilität des Pergamentpapiers zu suchen. Wird die Permeabilität in  $\text{cm}^3\text{ H}_2\text{O}$  pro  $\text{cm}^2$  pro Sekunde bei einem Überdruck von 1 cm Wassersäule angegeben (siehe BJERRUM, N. und MANEGOLD, E. Kolloid-Zs. **42** (1927) 97), ist die Permeabilität von dem hier benutzten Papier etwa  $1.4 \times 10^{-7}$ , während BRAUNER'S Angaben (1928) auf diese Einheit umgerechnet  $0.12 \times 10^{-7}$  ergeben.

selbe. Für  $E(\downarrow)$  fehlt jedoch der sehr starke Abfall während der ersten Minute, weil der Anfangszustand nicht durch Spülung festgelegt ist. Als Anfangswert beider Kurven nehmen BRAUNER und AMLONG den Wert an, den sie für  $E(\rightarrow)$  nach Erreichung annähernder Konstanz finden.

### 3. Erklärung der Entstehung des geoelektrischen Effektes.

Aus den Messungen ergibt sich, dass die  $E_B$ - $t$ -Kurven für alle drei Stellungen des Systems in einem gemeinsamen Punkt anfangen. Da alle drei Kurven auch in einem gemeinsamen Punkt enden, ist es klar, dass nur diesen beiden Potentialwerten ein besonderes Interesse zukommt. Der Anfangswert, der durch kräftige Spülung der Membranseiten erreicht wird, muss dem Zustand entsprechen, wo praktisch das ganze Konzentrationsgefälle innerhalb der Membran liegt. Der Endwert dagegen, der erst nach längerer Zeit erreicht wird, muss dem Zustand entsprechen, wo praktisch das ganze Gefälle in der Lösung liegt. Der Anfangswert von  $E_B$  ist deshalb durch Formel (2), der Endwert durch Formel (1) gegeben. Die dazwischenliegenden  $E_B$ -Werte sind von der jeweiligen Konzentrationsverteilung abhängig. Liegt das Konzentrationsgefälle von  $c_1'$  bis  $c_2'$  innerhalb der Membran, während das Gefälle von  $c_1$  bis  $c_1'$  auf der einen, das Gefälle von  $c_2'$  bis  $c_2$  auf der anderen Seite der Membran liegt, so ist  $E_B$  durch

$$E_B = \frac{RT}{F \log e} \left[ (1 - 2n_-) \log \frac{c_1}{c_1'} \cdot \frac{c_2'}{c_2} + (1 - 2n'_-) \log \frac{c_1'}{c_2'} \right] \quad (3)$$

gegeben. Wird (3) mit (2), resp. (1), kombiniert, so erhält man:

$$E_B = E_M - \frac{RT}{F \log e} 2(n_- - n'_-) \log \frac{c_1}{c_2} \cdot \frac{c_2'}{c_1'} \quad (4)$$

und

$$E_B = E_D + \frac{RT}{F \log e} 2(n_- - n'_-) \log \frac{c_1'}{c_2'} \quad (5)$$

Die letzten zwei Formeln zeigen deutlich, wie sich  $E_B$  von  $E_M$  entfernt und  $E_D$  zustrebt, je nachdem  $c_1':c_2'$  sich dem Werte 1 nähert. Keiner der zwischen  $E_M$  und  $E_D$  liegenden Werte hat besondere Bedeutung. Die Überlegungen von BRAUNER und AMLONG, die auf der fehlerhaften Annahme beruhen, dass die nach etwa einer halben Stunde gemessenen Potentialwerte  $E_M$ -Werte seien, haben somit kein Interesse. Von grosser Bedeutung ist dagegen der von ihnen gegebene experimentelle Nachweis, dass das Membranpotential innerhalb mässiger Versuchszeiten nicht nur von der Art der Membran und der Lösungen abhängt, sondern auch von der Orientierung des Systems zu der Richtung der Schwerkraft. Wie dieser Einfluss der Schwerkraft auf die zeitliche Änderung des Potentials entsteht, bedarf noch einer Erklärung.

Überlegt man, was in einem der Wirkung der Dialyse überlassenen System geschehen wird, so ist es leicht einzusehen, dass die Dialyserichtung eine Rolle für die Änderungsgeschwindigkeit des Verhältnisses  $c_1':c_2'$  und dadurch für die des Potentials spielen kann.

Gleichgültig wie das System zu der Richtung der Schwerkraft orientiert ist, wird die primäre Wirkung der Dialyse die sein, dass an der Seite, wo sich die 0.01 *n* *KCl*-Lösung befindet, eine Grenzschicht entsteht, in der die *KCl*-Konzentration grösser als 0.01 *n* ist. An der anderen Seite der Membran wird eine Grenzschicht entstehen, in welcher  $c < 0.1n$  ist. Für *KCl* sowie für die meisten anderen Salze

nimmt das spezifische Gewicht der Lösungen mit steigender Konzentration zu. Die Grenzschichten werden deshalb schwerer als die  $0.01 n$ , bzw. leichter als die  $0.1 n$  *KCl*-Lösung. Ist nun ursprünglich die  $0.1 n$  Lösung die obere, so werden zu beiden Seiten der Membran Konvektionsströme entstehen, und der hindurchdialysierende Stoff wird dadurch stetig von den Membranseiten fortgeführt. Ist aber ursprünglich die  $0.1 n$  Lösung die untere, so werden solche Ströme nicht entstehen. Die Stoffmenge, die die Membran passiert hat, wird in diesem Fall nur mittels der in den beiden Lösungen stattfindenden Diffusion von den Membranseiten entfernt. Man muss demnach erwarten, dass der ursprüngliche Konzentrationsunterschied zwischen den beiden Membranseiten am besten aufrechterhalten wird, wenn das *KCl* durch die Membran abwärts diffundiert.

Die Verhältnisse bei horizontaler Dialyse sind schwerer zu übersehen. Es müssen hier Konvektionsströme den Membranseiten entlang entstehen, und man wird deshalb erwarten eine grössere Ähnlichkeit mit abwärtsgerichteter als mit aufwärtsgerichteter Dialyse zu finden.

Das Resultat dieser Überlegungen ist in guter Übereinstimmung mit den in Fig. 2 gegebenen Messungen. Die Potentialwerte für wagrecht- und abwärtsgerichtete Dialyse fallen hier zusammen und liegen beträchtlich näher dem Anfangswert als es die Werte für aufwärtsgerichtete Dialyse tun.

#### 4. Nachweis eines Schwerkrafteffektes auf die Dialysegeschwindigkeit.

Ist die gegebene Erklärung richtig, so muss die Dialyse-richtung nicht nur einen Einfluss auf die Änderung des Membranpotentials haben, sondern die Wirkung auf der

Dialysegeschwindigkeit muss auch direkt nachweisbar sein. Diese Wirkung muss derart sein, dass die Dialyse am meisten gehemmt wird, wenn die 0.1 *n* *KCl*-Lösung die untere ist.

Versuche über den Einfluss der Orientierung des Systems auf die Dialysegeschwindigkeit wurden mit der in Fig. 1 gezeigten Apparatur ausgeführt. Die leitende Verbindung mit den Elektroden wurde abgebrochen, und die Gummischläuche bei *B'* und *B''* mit Glasstopfen verschlossen. Ganz wie bei den Potentialmessungen wurden Messungen mit und ohne Bepflügelung der Membran ausgeführt. Die ersteren gestatten eine Berechnung des Diffusionskoeffizienten von *KCl* in der Membran, die letzteren zeigen den Einfluss der Orientierung des Systems auf die Dialysegeschwindigkeit.

Bei den Versuchen mit Spülung traten gewisse Schwierigkeiten auf. Um messbare Konzentrationsänderungen zu erreichen, war es notwendig eine und dieselbe Flüssigkeitsmenge (etwa 40 cm<sup>3</sup>) mehrmals über die Membran strömen zu lassen. Die aus *D'*, resp. *D''*, ausströmende Lösung wurde deshalb in einem Messzylinder aufgefangen, in den Scheidetrichter *F'*, resp. *F''*, gegossen, und von neuem an die Membran geführt. Bei dieser wiederholten Benutzung der Lösung war ein Verdampfungsverlust nicht zu vermeiden. Am Ende des Versuches wurde an beiden Membranseiten die ganze Flüssigkeitsmenge abgelassen und zusammen mit der zur Nachspülung gebrauchten Menge (etwa 10 cm<sup>3</sup>) gewogen. Der Konzentrationsunterschied zwischen dieser Lösung und der ursprünglichen wurde mittels eines Zeiss-Interferometers bestimmt. Die Kammer des Interferometers war 8 cm lang. Eine Einheit auf der Trommel (gleich eine hundertstel Umdrehung) entsprach einem Konzentrationsunterschied von  $2.5 \times 10^{-5}$  Molen *KCl* pro Liter. Die Ein-

stellungsgenauigkeit betrug etwa 2 Trommeleinheiten. Für 0.01 *n* *KCl* konnte somit eine Konzentrationsänderung von etwa 5 ‰, für 0.1 *n* *KCl* eine Änderung von etwa 0.5 ‰ bestimmt werden.

Aus dem Gewicht der Lösung, der Konzentrationsänderung und der Versuchsdauer wurde die Anzahl Mole *KCl* berechnet, die pro Sekunde durch die Membran diffundiert war. Wegen des unvermeidlichen Verdampfungsverlustes wurde die Änderung für die 0.01 *n* Lösung ein wenig zu gross, für die 0.1 *n* Lösung ein wenig zu klein gefunden. Der Unterschied der beiden Werte war jedoch nicht gross, so dass der Mittelwert als genau genug für die vorliegende Untersuchung angesehen werden kann.

Die Versuche ohne Spülung wurden derart ausgeführt, dass zunächst während einer kurzen Periode beide Seiten der Membran sehr kräftig gespült wurden. Dann wurde gleichzeitig die Spülung an beiden Seiten abgestellt, und das System eine gewisse Zeit in Ruhe gelassen. Nach Verlauf einer passenden Zeit wurde beiderseits die Lösung völlig abgelassen. Gewicht und Konzentrationsänderung wurden wie oben bestimmt, und daraus die Anzahl Mole *KCl* berechnet, die durchschnittlich pro Sekunde die Membran passiert hatte. Die Bestimmungen sind für kurze Versuchszeiten mit einer gewissen Ungenauigkeit behaftet, teils weil die Konzentrationsänderungen hier klein sind, teils weil die Zeit, die für Entleeren und Ausspülen gebraucht wurde, im Verhältnis zu der eigentlichen Versuchsdauer nicht klein ist. In den in Tabelle 2 gegebenen Zeiten ist diese Zeit (etwa 0.5—1 Minute) zu der eigentlichen Versuchsdauer addiert.

Aus Tabelle 2 sieht man, dass die *KCl*-Menge, die pro Sekunde die Membran passiert hat, stets am grössten ist,

Tabelle 2.

Einfluss der Dialyserichtung auf die Dialysegeschwindigkeit.

System: 0.1 n KCl | Pergamentpapier | 0.01 n KCl

cm <sup>2</sup> pro Minute		0.01 n KCl oben		0.1 n KCl oben	
0.01 n KCl	0.1 n KCl	Versuchs- dauer	Mole/Sek. 10 <sup>8</sup>	Versuchs- dauer	Mole/Sek. 10 <sup>8</sup>
Pergamentpapier:					
14.7	7.0	31'40''	2.6*		
14.3	7.0			31'10''	3.4*
0	0	5'20''	1.9	5'30''	2.2
0	0	10'30''	1.5	10'30''	2.1
0	0	15'30''	1.5	15'30''	2.1
0	0	20'30''	1.5	20'30''	2.1
$q = 3.14; 2\delta = 0.0135 \text{ cm}$		$D_M = 0.11$		$D_M = 0.14$	
Kollodiummembran:					
5	5			35'20''	3.4*
8	8	36'05''	1.2		
33	33	31'19''	3.2*		
0	0	6'30''	3.3	6'00''	2.4
0	0	10'30''	2.1	11'00''	2.5
0	0	34'00''	1.2	31'00''	2.2
0	0	63'00''	0.7	61'00''	1.8
$q = 3.14; 2\delta = 0.0045 \text{ cm}$		$D_M = 0.044$		$D_M = 0.047$	

Die gegebenen  $D_M$ -Werte sind aus den mit \*) versehenen Zahlen berechnet.

wenn gespült wird. Der Einfluss der Dialyserichtung zeigt sich sowohl in Versuchen mit, als in Versuchen ohne Spülung. Bei den ersteren ist, wenn die Dialyse aufwärtsgerichtet ist, eine stärkere Spülung notwendig um die maximale Dialysegeschwindigkeit zu erreichen, als wenn sie abwärtsgerichtet ist. Bei den letzteren Versuchen wird die Dialyse am meisten gehemmt, wenn sich das KCl aufwärts bewegt. Die Bedeutung der Dialyserichtung für die Dialysegeschwindigkeit ist somit gezeigt. Da aber die Dialysege-



schwindigkeit von dem Konzentrationsunterschied zwischen den Membranseiten abhängt, ist auch dadurch ihre Bedeutung für das Festhalten der ursprünglichen Konzentrationsverteilung und für  $E_B$  gezeigt.

Aus dem Wert der Dialysegeschwindigkeit, der aus den Versuchen mit Spülung gefunden wurde, kann man den scheinbaren Diffusionskoeffizienten des Kaliumchlorids in der Membran berechnen. Die Berechnung wird ausgeführt als ob die Membran eine homogene Phase sei, nicht ein porenhaltiges Gerüst. Der Diffusionskoeffizient,  $D_M$ , ist durch

$$\frac{dS}{dt} = D_M q \frac{dc}{dx} \quad (6)$$

definiert, wo  $dS/dt$  die Anzahl Mole ist, die durch die Fläche  $q$  pro Zeiteinheit passiert, wenn  $dc/dx$  die Konzentrationsänderung pro Längeneinheit angibt. Gewöhnlich wird als Zeiteinheit 24 Stunden, als Längeneinheit 1 cm gebraucht, während die Konzentration in Mole pro  $\text{cm}^3$  angegeben wird. Ist  $N$  die Anzahl Mole  $KCl$ , die aus  $0.1 n$   $KCl$  in  $0.01 n$   $KCl$  durch eine Membran von der Dicke  $2\delta$ , und der Fläche  $q$  pro Sekunde diffundiert, so ist:

$$D_M = N \times 60^2 \times 24 \frac{2\delta}{q \times 9 \times 10^{-5}} = 9.6 \times 10^8 \frac{N \times 2\delta}{q} \text{ cm}^2 \text{ pro Tag.}$$

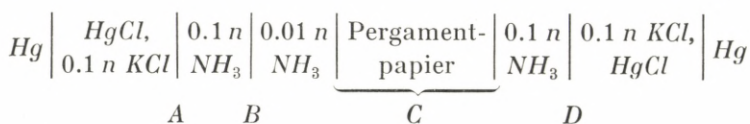
Aus den in Tabelle 2 angeführten Messungen mit einfachem Pergamentpapier ergibt sich  $D_M = 0.14 \text{ cm}^2$  pro Tag. Zum Vergleich sei genannt, dass ÖHOLM<sup>1</sup> für  $KCl$ , das bei  $18^\circ$  aus  $0.1 n$   $KCl$  in reines Wasser diffundiert, den Wert  $1.4 \text{ cm}^2$  pro Tag angibt. Ein Versuch mit 4-fachem Pergamentpapier ergab  $D_M = 0.29 \text{ cm}^2$  pro Tag. Der letzte Wert muss als der bessere angesehen werden (vergl. Seite 28). Das Ver-

<sup>1</sup> ÖHOLM, L. W., Zs. physikal. Chem. **50** (1905) 309.

hältnis zwischen dem scheinbaren Diffusionskoeffizienten der Membran und dem Diffusionskoeffizienten der Lösung wird somit gleich  $0.29 : 1.4 = 0.21$  gefunden.

### 5. Zusammenhang zwischen dem geelektrischen Effekt und dem spezifischen Gewicht der Lösungen.

Um eine weitere Bestätigung der ergebnen Erklärung zu bekommen, wurden Versuche mit  $NH_3$ -Lösungen, deren spezifisches Gewicht mit steigender Konzentration abnimmt, gemacht. Die gemessene Kette hatte die Zusammensetzung:



Die drei Übergangsschichten ohne Membran (A, B, D) wurden, wie auf Seite 9 beschrieben, hergestellt, indem in jedem Falle die spezifisch schwerste Lösung die untere war. Wegen der geringen Leitfähigkeit des  $NH_3$ -Wassers konnte nur mit einer Genauigkeit von etwa 1 *mV* gemessen werden. Die folgenden Werte wurden für die *EMK* der Kette gefunden:

Zeit nach Abstellung der Spülung: 0 1 2 5 10 20 Min.  
*EMK*, wenn 0.1 *n*  $NH_3$  unten ist: (20) 7 6 5 6 6 *mV*.  
*EMK*, wenn 0.1 *n*  $NH_3$  oben ist: (25) 11 7 4 1 0 *mV*.

Im Gegenteil zu dem, was für Ketten mit *KCl* gefunden wurde, wird hier, wie erwartet, der ursprüngliche *EMK*-Wert am besten festgehalten, wenn die 0.1 *n* Lösung die untere ist.

Schon BRAUNER und AMLONG haben den Gedanken gehabt, dass man aus einem Vergleich von *KCl*-Ketten mit Ketten mit  $NH_3$ -Wasser ersehen könne, ob Dichteunterschiede eine Rolle

bei dem Entstehen des geoelektrischen Effektes spielen. Leider haben sie für ihre Untersuchung nicht das System:  $0.01 n RX$ —Membran— $0.1 n RX$ , sondern das Doppelsystem:  $0.01 n RX$ —Membran— $0.1 n RX$ —Membran— $0.01 n RX$  gewählt. Aus der Tatsache, dass das untere Ende des Doppelsystems sowohl für  $RX = KCl$ , als für  $RX = NH_4OH$  positiv wurde, haben sie gefolgert, dass Dichteunterschiede für das Entstehen des Effektes belanglos seien. Dieser Schluss ist aber unberechtigt. Hat nämlich  $E_M$  für die beiden Systeme entgegengesetztes Vorzeichen, so wird in beiden Fällen das untere Ende des Doppelsystems positiv werden.

Über das Vorzeichen von  $E_M$  in der  $NH_3$ -Wasser-Kette geben weder die Messungen von BRAUNER und AMLONG noch meine Messungen Auskunft. Da die ausgeführten Messungen mit aller gewünschten Deutlichkeit die Bedeutung von Dichteunterschieden zeigen, war es für vorliegende Arbeit nicht von Interesse den  $E_M$ -Wert des  $NH_3$ -Wasser-Systems zu bestimmen.<sup>1</sup>

## 6. Versuch einer Kontrolle des gemessenen Membranpotentials durch Bestimmung der Überführungszahl des Anions.

Da der maximale Wert des Membranpotentials,  $E_M$ , eine für jedes System charakteristische Grösse ist, wäre es von Interesse, die aus dem  $E_M$ -Wert nach (2) berechnete Über-

<sup>1</sup> Es ist wahrscheinlich, dass  $E_M$  für die beiden untersuchten Ketten verschiedene Vorzeichen hat. Das Vorzeichen ist dadurch bestimmt, ob  $n_-'$  kleiner oder grösser als 0.5 ist. Für  $KCl$  ist  $n_'$  etwa 0.5, und eine geringfügige Änderung durch Adsorption des Anions bringt  $n_-'$  unter diesen Wert. Für  $NH_3$ -Wasser dagegen ist die Beweglichkeit des  $OH^-$ -Ions ungefähr dreimal so gross wie die des  $NH_4^+$ -Ions, und selbst eine ziemlich ausgeprägte selektive Adsorption wird nicht ausreichen, um  $n_-'$  kleiner als 0.5 zu machen.

führungszahl durch eine direkte Bestimmung zu kontrollieren. Leider zeigte es sich unmöglich in dieser Weise eine scharfe Kontrolle zu erhalten, weil die Überführungszahlen in  $0.1n$  und  $0.01n$  Lösung verschieden waren. Man muss sich deshalb darauf beschränken zu konstatieren, ob der berechnete Wert zwischen den beiden direkt gefundenen liegt.

Eine Abhängigkeit der Überführungszahl von der Konzentration ist auch von MICHAELIS<sup>1</sup> für Kollodiummembranen gefunden. Je verdünnter die Lösung ist, um so grösser wird der Einfluss der Membran auf die Überführungszahl. Dieses Ergebnis ist durchaus verständlich. Wenn das Diffusionspotential durch Zwischenschaltung einer Membran geändert wird, ist die Hauptursache die, dass eine Anzahl der einen Ionenart von dem Membranmaterial adsorbiert wird und deshalb nicht an dem Elektrizitätstransport teilnehmen kann. Wenn auch die Anzahl der adsorbierten Ionen nicht von der Konzentration der Lösung unabhängig ist, darf man jedoch annehmen, dass die Anzahl hauptsächlich durch die Grösse der Oberfläche des Membranmaterials bestimmt wird. In einer verdünnten Lösung wird deshalb der Bruchteil, der von der betreffenden Ionenart adsorbiert wird, grösser sein als in einer konzentrierteren Lösung.<sup>2</sup>

Für die Bestimmung der Überführungszahlen wurde die gewöhnliche Apparatur (Fig. 1) benutzt, nur wurden die Rohre  $B'$  und  $B''$  mit Rohren von der in Fig. 1 b gezeigten Form ersetzt. Das Hartgummirohr wurde so gedreht, dass

<sup>1</sup> MICHAELIS, L., WEECH, A. A. und YAMATORI, A. Journ. Gen. Physiol. **10** (1927) 685.

<sup>2</sup> Vergleiche BETHE, A. und TOROPOFF, T., Zs. physikal. Chem. **88** (1914) 686.

die Membran senkrecht stand. Die Anode war aus blankem Silber, die Kathode aus Silber auf dem eine reichliche Menge Silberchlorid elektrolytisch ausgefällt war. Die Apparatur wurde mit der zu untersuchenden Lösung ganz gefüllt. Die angelegte Spannung wurde so gewählt, dass die Stromstärke 0.1 bis 0.2 mAmp. betrug. Für 0.01 *n* KCl war eine Spannung von 220 Volt nötig. Dass diese Behandlung die Membran nicht schädigte, wurde durch Messung von  $E_M$  vor und nach der Bestimmung der Überführungszahl kontrolliert. Genau wie bei den Diffusionsversuchen strömte dieselbe Flüssigkeit mehrmals über die Membran. Am Ende des Versuches wurde an jeder Seite der Membran nicht nur die Lösung, die zirkuliert hatte, sondern auch die Lösung, die sich zwischen der Membran und dem Hahn befand, durch *H* abgelassen und zusammen mit der zur Nachspülung benutzten Lösung gewogen. Die Konzentrationsänderung wurde interferometrisch bestimmt. Ein Stromdurchgang wird an beiden Seiten der Membran eine Konzentrationsänderung bewirken, die von dem Unterschied der Überführungszahlen in der Lösung und in der Membran abhängt. Bei einer Elektrizitätsmenge von 1 Faraday werden der Seite der Membran, die der Kathode am nächsten ist,  $n_- \text{ Äq. Cl}^-$  durch die Lösung,  $(1 - n_-')$  Äq.  $K^+$  durch die Membran zugeführt werden. Gleichzeitig werden  $n_-'$  Äq.  $\text{Cl}^-$  durch die Membran,  $(1 - n_-)$  Äq.  $K^+$  durch die Lösung fortgeführt. Das Resultat wird eine Vergrößerung der KCl-Menge von  $(n_- - n_-')$  Molen sein. An der anderen Seite der Membran muss eine ebenso grosse Abnahme der Anzahl Mole KCl stattfinden.

Wie bei den Diffusionsversuchen kann ein Verdampfungsverlust nicht vermieden werden. Es wurde gezeigt, dass die Verdampfungsverluste gleich gross waren, wenn

an beiden Membranseiten mit derselben Geschwindigkeit gespült wurde. Wird die Konzentrationsänderung für gleich grosse Volumina und bei gleich grosser Strömungsgeschwindigkeit an beiden Seiten bestimmt, so wird deshalb die Verdampfung für die Bestimmung des Konzentrationsunterschiedes zwischen den beiden Lösungen keine Rolle spielen. Für 1 Faraday bekommt man unabhängig von der Grösse des Verdampfungsverlustes einen Konzentrationsunterschied von  $2(n_- - n_-') 1000/V$ , wenn  $V$  das auf jeder Seite benutzte Volumen ist.

Die während eines Versuches gebrauchte Elektrizitätsmenge betrug  $2-5 \times 10^{-5}$  Faraday. Die Menge wurde sowohl durch Ablesung der Stromstärke, sowie durch die hervorgerufene Konzentrationsänderung der Kathodelösung bestimmt. Die Übereinstimmung beider Bestimmungen war gut.

Aus der Elektrizitätsmenge,  $Q$ , und dem Konzentrationsunterschied,  $\Delta c$ , zwischen den Lösungen wurde  $n_- - n_-'$  aus der Formel

$$\Delta c = 2 Q (n_- - n_-') 1000/V \quad (7)$$

berechnet. Für  $0.1 n KCl$  wurde  $n_- - n_-'$  zu etwa 0.09, für  $0.01 n KCl$  zu etwa 0.24 gefunden. Da  $n_-$  gleich 0.50 gesetzt werden darf, ergibt sich für  $0.1 n KCl$ :  $n_-' = 0.41$ , für  $0.01 n KCl$ :  $n_-' = 0.26$ .

Wird  $n_-'$  nach Formel (2) aus dem Wert  $E_M \approx 8.5 mV$  (siehe Tabelle 1) berechnet, so erhält man  $n_-' = 0.43$ . Dieser Wert liegt nicht zwischen den beiden direkt bestimmten Werten. Der Wert  $8.5 mV$  scheint somit zu klein zu sein. Einen zu kleinen  $E_M$ -Wert wird man finden, wenn es nicht durch die Spülung gelingt das ganze Konzentrationsgefälle in der Membran zu halten. Wie Seite 28 näher

erörtert werden soll, ist es wahrscheinlich unmöglich, diese gewünschte Konzentrationsverteilung zu erreichen, wenn die Membran sowohl gut durchlässig, als dünn ist. Will man den  $E_M$ -Wert für Pergamentpapier bestimmen, soll man deshalb am liebsten mehrere Schichten anwenden. Eine Bestimmung mit 4-fachem Pergamentpapier ausgeführt, ergab den höheren  $E_M$ -Wert 13.8 mV. Wird dieser Wert für die Berechnung benutzt, so findet man  $n_- = 0.38$ . Dieser Wert liegt, wie erforderlich, zwischen den beiden direkt bestimmten Werten. Vielleicht ist jedoch auch dieser Wert noch ein wenig zu gross.

## 7. Die Anwendung der Diffusionsgesetze auf membranhaltige Systeme.

Die obigen, qualitativen Überlegungen stehen in guter Übereinstimmung mit den Versuchsergebnissen. In Fällen, wo keine Konvektionsströme entstehen, können sie durch eine mathematische Behandlung ergänzt werden.

Die Untersuchungen fielen in zwei Gruppen: Die Versuche bei denen ein stationärer Zustand durch Spülung der Membranseiten erzeugt wurde, und die, bei denen das System dem Einfluss der Diffusion frei überlassen wurde.

a. Versuche mit Spülung. Die Konzentrationsverteilung ist eine stationäre.

Der Zweck dieser Versuche ist eine Bestimmung von  $E_M$ . Mit der Spülung wird erstrebt, die Konzentration unterhalb der Membran gleich  $c_1$ , oberhalb der Membran gleich  $c_2$  zu erhalten ( $c_1 > c_2$ ). Gelingt es, diese Konzentrationen überall ausserhalb der Membran zu bewahren, so wird sich die Konzentration in der Membran linear mit dem Abstand ändern. Ist die Dicke der Membran  $2\delta$ , und wird  $c_1 = c_0 + \Delta$ ,  $c_2 =$

$c_0 - \Delta$  gesetzt, so wird man im Abstände  $x$  von der Mitte der Membran eine Konzentration haben, die durch

$$c = c_0 + \Delta x/\delta \quad (8)$$

gegeben ist.

Gelingt es nicht, diese Konzentrationsverteilung zu erzwingen, sondern ist das Resultat der Spülung dies, dass die Konzentrationen nicht ganz bis an die Membranseiten, sondern nur bis auf einen gewissen Abstand von der Membran ( $s_1$  resp.  $s_2$ ) konstant gehalten werden, so bekommt man anstatt  $E_M$  einen durch (3) bestimmten  $E_B$ -Wert. Wie gross die Abweichung von  $E_M$  wird, hängt teils von der Konzentrationsverteilung des erreichten stationären Zustandes, teils von dem Wert  $n_- - n_-'$  ab.

Die Konzentrationsverteilung in dem stationären Zustand ist durch

$$Dq \frac{c_1 - c_1'}{s_1} dt = kDq \frac{c_1' - c_2'}{2\delta} dt = Dq \frac{c_2' - c_2}{s_2} dt \quad (9)$$

gegeben, wo  $D$  und  $kD$  der Diffusionskoeffizient für  $KCl$  in der Lösung, bezw. Membran ist. Ferner ist  $q$  die Fläche und  $2\delta$  die Dicke der Membran, während  $c_1'$ , bezw.  $c_2'$ , die Konzentrationen an beiden Membranseiten angibt. Aus (9) findet man:

$$\frac{c_1'}{c_2'} = \frac{c_1 + k(c_1 s_2 + c_2 s_1)/2\delta}{c_2 + k(c_1 s_2 + c_2 s_1)/2\delta} \quad (10)$$

Der Ausdruck (10) zeigt, dass wenn  $k$  sehr klein ist, d. h. die Membran sehr undurchlässig ist,  $c_1' : c_2'$  annähernd gleich  $c_1 : c_2$  sein wird, auch wenn  $s_1$  und  $s_2$  nicht klein sind. Die Wirksamkeit der Spülung ist in diesem Falle belanglos. Für besser durchlässige Membranen wird dagegen die Art der Spülung von entscheidender Bedeutung sein.



Nur wenn  $s_1$  und  $s_2$  beide klein gegen  $2\delta$  sind, wird  $c_1' : c_2'$  annähernd gleich  $c_1 : c_2$ . Man ersieht ferner aus (10), dass  $s_2$  einen grösseren Einfluss auf  $c_1' : c_2'$  hat als  $s_1$ . Es ist somit wichtiger die Seite mit der verdünnteren Lösung gut zu spülen als die mit der konzentrierteren. Dieses Resultat stimmt mit den auf Seite 11 besprochenen Erfahrungen völlig überein.

Wird (10) mit (4) kombiniert, so bekommt man einen Ausdruck, aus welchem man ersehen kann, wie klein  $s_1$  und  $s_2$  sein müssen, um die Abweichung von  $E_M$  unter einem gewissen Werte zu halten. Wird der Einfachheit halber  $s_1 = s_2 = s$  angenommen, so ist die Abweichung durch

$$E_M - E_B = \frac{RT}{F \log e} 2(n_- - n'_-) \log \frac{c_1 [c_2 + ks(c_1 + c_2)/2\delta]}{c_2 [c_1 + ks(c_1 + c_2)/2\delta]} \quad (11)$$

gegeben. Für das System: 0.01 n KCl—Pergamentpapier—0.1 n KCl kann  $n_- - n'_- =$  etwa 0.12 gesetzt werden; (siehe unten). Wünscht man  $E_M$  mit einer Genauigkeit von 1 mV zu bestimmen, so muss nach (11)  $s$  so klein gehalten werden, dass

$$\frac{c_1 + ks(c_1 + c_2)/2\delta}{c_2 + ks(c_1 + c_2)/2\delta} \geq 8.47$$

wird. Für  $k$  wurde  $0.29/1.4 = 0.21$  gefunden (Seite 20). Die Dicke des Pergamentpapiers,  $2\delta$ , war 0.135 mm. Für das betrachtete System muss man deshalb

$$s \leq 0.089 \times 2\delta \approx 0.012 \text{ mm}$$

haben, um  $E_M - E_B \leq 1 \text{ mV}$  zu erreichen. Dieser für  $s$  gefundene Wert ist von derselben Grössenordnung wie die Dicke, die für die festhaftende Schicht bei sehr kräftiger

Rührung aus Versuchen über Lösungsgeschwindigkeit gefunden worden ist<sup>1</sup>. Es scheint demnach zweifelhaft, ob überhaupt selbst bei kräftiger Spülung ein so kleiner oder noch kleinerer  $s$ -Wert erreicht werden kann. Eine genaue Bestimmung von  $E_M$  darf deshalb nur erwartet werden, wenn  $2\delta$  durch Zwischenschaltung mehrerer Schichten vergrößert wird. Hierdurch wird keine andere der in (11) eingehenden Grössen geändert. Benutzt man 4-faches Pergamentpapier so muss  $s \leq 0.048$  mm sein, damit  $E_M - E_B \leq 1$  mV wird, was als erreichbar angesehen werden darf. Eine Messung mit 4-fachem Pergamentpapier ausgeführt, ergab wirklich einen höheren Wert für  $E_M$ . Bei einer Strömungsgeschwindigkeit von etwa  $40 \text{ cm}^3$  pro Minute wurde mit 1 Schicht  $8.5$  mV, mit 4 Schichten  $13.8$  mV gefunden. Der letzte Wert muss wegen der besseren Versuchsbedingungen der richtigere sein. Wird er der Berechnung von  $n_-'$  zugrunde gelegt, so findet man  $n_-' = 0.38$ . Für  $n_- - n_-'$  ergibt sich dann der oben benutzte Wert  $0.12$ .

In analoger Weise kann man berechnen wie gross  $s$  sein muss, damit  $E_B - E_D \leq 1$  mV ist. Wird (10) mit (5) kombiniert, bekommt man

$$E_B - E_D = \frac{RT}{F \log e} 2(n_- - n_-') \log \frac{c_1 + ks(c_1 + c_2)/2\delta}{c_2 + ks(c_1 + c_2)/2\delta}. \quad (12)$$

Für 1 Schicht Pergamentpapier muss  $s$  demnach grösser als etwa 3 mm sein.

b. Versuche ohne Spülung. Die Konzentrationsverteilung ändert sich während des Versuches.

Der Zweck dieser Versuche ist der, die zeitliche Änderung von  $E(\uparrow)$  zu bestimmen. Die Versuche wurden auf die Weise ausgeführt, dass die Membranseiten zunächst gespült

<sup>1</sup> BRUNNER, E. Zs. physikal. Chem. **47** (1904) 56.

wurden, so dass sich ein stationärer Zustand einstellte. In dem Augenblick, wo die Spülung abgestellt wird, fängt die Wirkung der Diffusion ausserhalb der Membran an. In beiden Lösungen folgt die Diffusion dem Gesetze:

$$\frac{dc}{dt} = D \frac{d^2c}{dx^2} \quad (13a)$$

während in der Membran das Gesetz:

$$\frac{dc}{dt} = D_M \frac{d^2c}{dx^2} = kD \frac{d^2c}{dx^2} \quad (13b)$$

gilt. Unter der Annahme, dass die Spülung so wirksam gewesen ist, dass das Konzentrationsgefälle anfangs ganz in der Membran liegt, soll die Lösung dieser beiden Gleichungen so beschaffen sein, dass die folgenden Bedingungen erfüllt werden:

für  $t = 0$ :  $c = c_0 + \Delta$  für  $x > +\delta$ ;  $c = c_0 - \Delta$  für  $x < -\delta$ ;

$$c = c_0 + x \frac{\Delta}{\delta} \text{ für } -\delta < x < +\delta$$

für jedes  $t$ :  $c = c_0 + \Delta$  für  $x = +\infty$ ;  $c = c_0 - \Delta$  für  $x = -\infty$ .

Nur wenn  $c_1'$  und  $c_2'$ , d. h. die Konzentrationen an beiden Membranseiten sich zu beliebigen Zeiten aus diesen Gleichungen berechnen lassen, ist ein Vergleich zwischen Messresultaten und Theorie möglich. Leider ist die exakte mathematische Behandlung eines Systems, wo für verschiedene Gebiete verschiedene Diffusionskoeffizienten gelten, sehr schwierig<sup>1</sup>, und für das hier vorliegende Problem wird das Resultat zu kompliziert, um eine numerische Auswertung zu erlauben. Auf einen solchen Vergleich muss deshalb hier verzichtet werden.

<sup>1</sup> Vergl. LOWAN, A. N., Heat conduction in a semi-infinite solid of two different materials. Duke Math. Journ. 1 (1935) 94.

Für grosse Versuchszeiten, wo die Dicke der Membran klein ist im Verhältnis zu dem Abstand zwischen den Ebenen, wo die Konzentrationen  $c_0 + \Delta$ , resp.  $c_0 - \Delta$ , sind, liefert die in FRANK und v. MISES, Die Differential- und Integralgleichungen der Mechanik und Physik (Bd. II S. 582 Ausg. 1936), gegebene Lösung für »Diffusion durch ein Diaphragma« eine Annäherung. Das dort betrachtete System ist folgendes: Zwei Lösungen von verschiedenen Konzentrationen sind an der Ebene  $x = 0$  durch eine unendlich dünne Membran voneinander getrennt. In beiden Lösungen ist der Diffusionskoeffizient gleich  $D$ . Die Diffusionsgeschwindigkeit durch die Membran ist dadurch gegeben, dass infolge der Stetigkeit des Diffusionsflusses in der Ebene  $x = 0$  folgende Gleichungen bestehen müssen:

$$D \frac{dc_1}{dx} = D' (c_1 - c_2) = D \frac{dc_2}{dx}. \quad (14)$$

Für eine Membran von endlicher Dicke wird  $D'$  eine Grösse sein, die sowohl von der Dicke als von dem Diffusionskoeffizienten der Membran abhängt. Mit den hier benutzten Bezeichnungen wird  $D' = D_M/2\delta = kD/2\delta$ .

In das in dieser Arbeit betrachtete System ist für  $t = 0$ :  $c_1 = c_0 + \Delta$  und  $c_2 = c_0 - \Delta$ . Werden diese Werte in die FRANK-v. MISES'sche Lösung eingesetzt, so erhält man für  $x = +0$ :  $c_1' = c_0 + \Delta \cdot a$ , für  $x = -0$ :  $c_2' = c_0 - \Delta \cdot a$ , wo  $a$  durch

$$a = e^{\frac{4D'^2 t}{D}} \left[ 1 - \psi \left( \frac{2D'}{\sqrt{D}} t \right) \right] \quad (15)$$

gegeben ist. Mit  $\psi$  wird das GAUSS'sche Fehlerintegral bezeichnet, d. h.

$$\psi(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-\omega^2} d\omega.$$

In Fig. 3 sind die experimentell gefundenen Punkte der  $E(\uparrow)$ - $t$ -Kurve für 4-faches Pergamentpapier und für eine Kollodiummembran eingezeichnet. Zum Vergleich sind die Kurven eingezeichnet, die man durch Kombination von (4) mit (15) berechnet. Für  $t \geq 10$  Min. gibt die Berechnung in beiden Fällen eine gute Übereinstimmung mit den experimentellen Kurven, wenn man die folgenden Werte für  $n_-'$

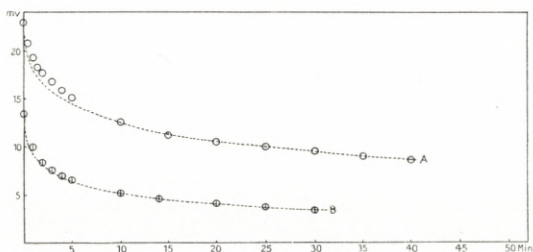


Fig. 3. Die zeitliche Änderung des Potentials im Systeme  $0.01 n KCl$  — Membran —  $0.1 n KCl$  bei aufwärtsgerichteter Dialyse.

Kurve A berechnet für  $n_- - n_-' = 0.20$ ;  $(k\sqrt{D}/\delta)^2 = 0.017$ .  $\circ$  gemessene Werte für Kollodiummembran II.

Kurve B berechnet für  $n_- - n_-' = 0.13$ ;  $(k\sqrt{D}/\delta)^2 = 0.095$ .  $\odot$  gemessene Werte für 4-faches Pergamentpapier.

und für  $4 D'^2/D$  benutzt: für 4-faches Pergamentpapier:  $n_- - n_-' = 0.13$  und  $4 D'^2/D = 0.095$ , für die Kollodiummembran:  $n_- - n_-' = 0.20$  und  $4 D'^2/D = 0.017$ .  $D$  ist nach ÖHOLM<sup>1</sup> gleich  $0.14 \text{ cm}^2/\text{Tag}$  oder etwa  $10^{-3} \text{ cm}^2/\text{Min}$ . Durch Einsetzung dieses Wertes und der Wert der Membrandicken in dem Ausdruck:  $4 D'^2/D = k^2 D/\delta^2$  lassen sich die  $k$ -Werte der beiden Membranen berechnen. Die Dicke ist für 4-faches Pergamentpapier mit  $0.0630 \text{ cm}$  und für die Kollodiummembran mit  $0.0045 \text{ cm}$  gemessen worden. Die entsprechenden  $k$ -Werte sind: für das Pergamentpapier  $0.31$ , für die Kollodiummembran  $0.009$ . Die direkte Bestimmung ergab für 1-faches

<sup>1</sup> loc. cit.

Pergamentpapier 0.10, für 4-faches Pergamentpapier 0.21. Wahrscheinlich ist der wirkliche  $k$ -Wert dann noch etwas höher. Für die Kollodiummembran wurde etwa 0.010 gefunden. Die Übereinstimmung ist in beiden Fällen befriedigend.

Für abwärts gerichtete Dialyse ist eine mathematische Behandlung undurchführbar. Werden die ursprünglichen Konzentrationen durch Spülung an beiden Membranseiten aufrechterhalten, muss man zwar annehmen, dass ein stationärer Zustand entsteht, in welchem die durch die Membran pro Zeiteinheit passierende Stoffmenge gleich der ist, die mittels Konvektion und Diffusion entfernt wird. Die Geschwindigkeit, mit der die Entfernung durch Konvektion stattfindet, ist wahrscheinlich eine Funktion des spezifischen Gewichtes in Abhängigkeit von der Konzentration. Die Art dieser Funktion ist aber nicht bekannt, und eine Berechnung deshalb nicht durchführbar. Man kann jedoch einsehen, dass man  $E_M$  nur dann messen kann, wenn die Gebiete, in welchen sich die Konzentrationen  $c_1$  und  $c_2$  einstellen, bis nahe an die Membran reichen, da sonst ein beträchtlicher Teil des Konzentrationsgefälles in der Lösung liegt.

Für Versuche ohne Spülung scheint es möglich, dass der Verlauf der  $E_B$ -Kurve von den Dimensionen der Apparatur abhängig ist. Die Messungen zeigen, dass ein gewisses Konzentrationsgefälle in der Membran längere Zeit aufrechterhalten wird. Die Dauer dieses Festhaltens wird wahrscheinlich um so grösser sein, je weiter der Abstand der Membran von den beiden Enden des Rohres ist. Der festgehaltene  $E(\downarrow)$ -Wert ist immer beträchtlich kleiner als der  $E_M$ -Wert. Beispielsweise ist für das System  $0.1 n KCl$ -

Pergamentpapier-0.01 *n* *KCl* der festgehaltene Wert etwa 4 *mV*, während der Anfangswert 8.5 *mV*, und der wirkliche  $E_M$ -Wert mindestens 13.8 *mV* betrug. Der Wert 4 *mV* entspricht einer Konzentrationsverteilung mit  $c_1':c_2'$  etwa gleich 2.

Die früher besprochenen Messungen von MICHAELIS und FUJITA sind alle derart ausgeführt, dass die von ihnen gemessenen *E*-Werte nicht, wie sie selber annehmen,  $E_M$ -Werte sind, sondern eher diesen längere Zeit konstanten  $E(\downarrow)$ -Werten entsprechen. Eine Ausnahme bilden jedoch ihre Messungen mit sehr undurchlässigen Membranen, bei welchen  $E(\downarrow)$  wahrscheinlich gleich  $E_M$  ist.

### 8. Über die Grösse der Effekte, die durch Zwischenschaltung einer Membran entstehen.

Die Wirkung der Zwischenschaltung einer Membran muss von der Art der Membran sowohl als von der Art der Lösungen abhängen.

Betreffend der Abhängigkeit der Art des gelösten Salzes sei nur auf die Arbeiten von MICHAELIS und FUJITA und von BRAUNER verwiesen, durch welche gezeigt ist, dass die am wenigsten adsorbierte der beiden Ionenarten den grösseren Einfluss auf das Potential hat.

Über die Bedeutung der Membranart und der Konzentration der Lösungen kann für ein gegebenes Salz folgendes gesagt werden: Für jedes System ist das charakteristische Membranpotential,  $E_M$ , durch (3) gegeben. Aus dieser Formel folgt, dass für einen gegebenen Wert von  $c_1:c_2$  das Potential um so grösser wird, je kleiner  $n'_$  ist.  $n'_$ , die Überführungszahl des Anions in der Membran, hängt nicht nur von der Art des Salzes und von der Membran ab, sondern auch von den absoluten Werten der Kon-

zentrationen. Je verdünnter die Lösungen sind, und je selektiver das Adsorptionsvermögen der Membran für die eine Ionenart des betreffenden Elektrolyten ist, um so grösser wird der numerische Wert von  $E_M$ . Für eine gegebene Membran steigt also  $E_M$  mit wachsender Verdünnung der Lösungen. Werden mehrere Membranen von demselben Materiale verglichen, so muss ein selektives Adsorptionsvermögen sich um so stärker geltend machen, je enger die Poren sind, d. h. je undurchlässiger die Membran ist. Für gut durchlässige Membranen ist gewöhnlich  $n_-'$  ungefähr gleich  $n_-$ , für sehr undurchlässige Membranen wird dagegen entweder  $n_-'$  oder  $n_+'$  annähernd gleich Null sein. Die grössten  $E_M$ -Werte muss man deshalb bei Messungen mit undurchlässigen Membranen erwarten.

Die Dicke der Membran spielt für den  $E_M$ -Wert theoretisch keine Rolle, praktisch ist bei gegebenen Versuchsbedingungen  $E_M$  um so genauer zu messen, je dicker die Membran ist.

Die Grösse des geoelektrischen Effekts ist dadurch begrenzt, dass er kleiner als  $E_M - E_D$  sein muss, weil er ja durch den verschiedenen Verlauf des Überganges von  $E_M$  zu  $E_D$  entsteht. Für gut durchlässige Membranen, wo der genannte Unterschied klein ist, muss der geoelektrische Effekt deshalb auch klein sein. Wenn für ein System  $E_M - E_D$  gross ist, ist es jedoch nicht sicher, dass  $GEP$  messbar ist. Für sehr undurchlässige Membranen ist die Stoffmenge, die selbst während langer Zeiten die Membran passiert, so gering, dass ein Unterschied in der Geschwindigkeit, mit welcher sie entfernt wird, belanglos ist. In einem solchen Fall wird  $GEP$  trotz einem grossen Wert von  $E_M - E_D$  verschwindend klein. Während somit die grössten  $E_M$ -Werte für undurchlässige Membranen zu erwarten sind,



muss man die grössten *GEP*-Werte bei einer gewissen mittelgrossen Durchlässigkeit der Membranen erwarten.

Dieser Zusammenhang zwischen Permeabilität und *GEP* ist schon von BRAUNER<sup>1</sup> experimentell nachgewiesen worden.

Rein schematisch können die Membranen in drei Gruppen eingeordnet werden.

1. Sehr undurchlässige Membranen:

$$kD \infty 0; n_- ' \text{ oder } n_+ ' \infty 0.$$

$E_M - E_D$  ist gross.  $E_M$  ist leicht messbar (siehe Formel (11)) und geht sehr langsam in  $E_D$  über. *GEP* ist klein.

2. Membranen von mittlerer Durchlässigkeit:

$$0 < kd < D; 0 < n_- - n_- ' < n_- \text{ oder } 0 < n_+ - n_+ ' < n_+.$$

$E_M - E_D$  ist gross.  $E_M$  ist nur bei gewissen Vorsichtsmassregeln messbar (Spülung, mehrfacher Membran). Die Änderung von  $E_M$  bis  $E_D$  dauert einige Stunden bis einen Tag. *GEP* ist gross, weil die Änderung von  $E(\uparrow)$  und  $E(\downarrow)$  mit verschiedener Geschwindigkeit stattfindet.

3. Leicht durchlässige Membranen:

$$kD \infty D; n_- - n_- ' \infty 0.$$

$E_M - E_D$  ist klein.  $E_M$  ist praktisch unmessbar.  $E_D$  wird sowohl für  $E(\uparrow)$  als für  $E(\downarrow)$  schnell erreicht. *GEP* ist klein.

Fig. 4 gibt die  $E_B$ - $t$ -Kurven für drei Kollodiummembranen, eine von jeder der genannten Gruppen an.  $kD$  für *KCl* ist nicht bestimmt worden. Für die Charakterisierung der Membranen sind die drei von BJERRUM und MANEGOLD<sup>2</sup> empfohlenen Grössen benutzt. Es sind diese:

<sup>1</sup> loc. cit. (1928).

<sup>2</sup> BJERRUM, N. und MANEGOLD, E., Kolloidchem. Zs. **42** (1927), 97.

$D_{H_2O}$ , die Durchlässigkeit für Wasser in  $\text{cm}^3$  pro  $\text{cm}^2$  Fläche pro Sekunde bei einem Überdruck von 1 cm Wassersäule.

$d_M$ , die Dicke der Membran in cm angegeben.

$W$ , das Wassergehalt in g  $H_2O$  pro  $\text{cm}^3$  der wassergefüllten Membran.

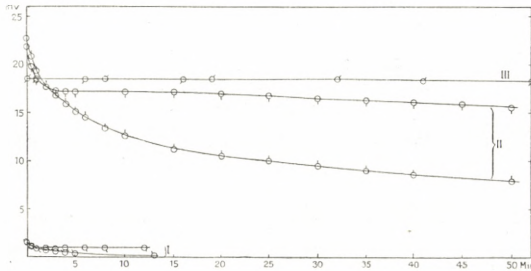


Fig. 4. Die zeitliche Änderung des Potentials im Systeme:  
0.01 n KCl — Kollodiummembran — 0.1 n KCl.

Membran I: .....  $\circ$  0.1 n KCl unten;  $\ominus$  0.1 n KCl oben.

Membran II: ....  $\circ$  0.1 n KCl unten;  $\ominus$  0.1 n KCl oben.

Membran III: ...  $\circ$  0.1 n KCl unten;  $\ominus$  0.1 n KCl oben.

Für die drei Kollodiummembranen hatten diese Größen die Werte:

	$D_{H_2O} \cdot 10^7$	$d_M \cdot 10^3$	$W$
Membran I . . . . .	1.72	6.7	0.70
Membran II . . . . .	0.14	4.5	0.40
Membran III . . . . .	0.04	5.5	0.098

## 9. Die Bedeutung der Membraneffekte bei der Messung galvanischer Ketten mit zwischengeschalteter Membran.

Dass Potentialunterschiede in membranhaltigen Systemen entstehen können, und dass diese in einigen Fällen stundenlang beinahe konstant bleiben, ist eine Tatsache, auf die man achten muss, wenn man mit solchen Systeme

men arbeitet. Die Bedeutung der Membraneffekte soll deshalb für drei Beispiele diskutiert werden.

### 1. Messung von DONNAN-Potentialen.

Das DONNAN'sche Membranpotential ist der Potentialunterschied, der sich zwischen den beiden Seiten einer Membran einstellt, wenn diese zwei Lösungen trennt, von denen die eine ein Ion enthält, für welche die Membran undurchlässig ist<sup>1</sup>.

Da man bei solchen Messungen die Einstellung des Gleichgewichtes zwischen den beiden Lösungen abwartet, wird es keine Rolle spielen, ob  $E_M$  mehr oder weniger von  $E_D$  abweicht. Höchstens mag davon die Rede sein, dass der schnellere oder langsamere Übergang des einen Wertes in den anderen die Einstellung des endlichen Potentials verzögert. Die Membran darf deshalb nicht undurchlässiger als notwendig gewählt werden, und muss jedenfalls für alle einfachen Ionen gut durchlässig sein.

### 2. Messung von galvanischen Ketten mit zwischen-geschaltetem Agar-Heber.

FUJITA<sup>2</sup> hat gezeigt, dass die Zwischenschaltung einer Agarmembran das Diffusionspotential zwischen zwei Lösungen desselben Salzes ändert. Wenn auch seine Versuchsanordnung sehr von den Verhältnissen abwich, die herrschen, wenn zwei Lösungen mittels eines Agar-Hebers verbunden sind, muss man doch erwägen, ob nicht durch den Gebrauch eines solchen, unerwünschte Potentialunterschiede entstehen können.

<sup>1</sup> Die Membranpotentiale, die von MICHAELIS gemessen sind mit Membranen, die für einfache anorganische Anionen undurchlässig waren, sind somit auch als Donnan-Potentialen anzusehen. Siehe: MICHAELIS, L. Journ. Gen. Physiol. **10** (1927) 579.

<sup>2</sup> FUJITA, A. Biochem. Zs. **162** (1925) 245.

Durch die Zwischenschaltung des Hebers beabsichtigt man, dasselbe Diffusionspotential zu erhalten, wie bei Zwischenschaltung einer konzentrierten *KCl*-Lösung. Für diesen Zweck ist die übliche Arbeitsmethode nicht geeignet. Man arbeitet gewöhnlich so, dass das Glasrohr, das mit dem *KCl*-haltigen Agar ganz gefüllt ist, in den beiden zu verbindenden Lösungen hineintaucht. Bei dieser Anordnung wird das *KCl*, das in der Lösung hinausdiffundiert, durch Konvektionsströme entfernt werden. Bewirkt nun das Agar-Heber einen messbaren Membraneffekt, so wird dieser mittels der Konvektionsströme aufrechterhalten. Die Anordnung zeigt sich somit als unzweckmässig. Eine zweckmässigere Anordnung erhält man, wenn das Rohr, dessen Öffnung aufwärtsgerichtet ist, nicht ganz mit dem *KCl*-haltigen Agar gefüllt wird. In dem agarfreien Ende des Heberrohrs wird sich dann eine Schicht konzentrierter *KCl*-Lösung allmählich ausbilden. Noch besser ist es, den Agar mit einer dünnen Schicht gesättigter *KCl*-Lösung zu bedecken. Diese Anordnung ist der von MICHAELIS und FUJITA<sup>1</sup> vorgeschlagenen sehr ähnlich. Sie tauchen das Agar-Heber in eine Elektrodelösung, die an *KCl* gesättigt ist. Die in vorliegender Arbeit empfohlene Methode hat den Vorteil, dass man durch sie eine zylindrische Symmetrie der Diffusionsschicht erhält, was nach GUGGENHEIM<sup>2</sup> für die Erreichung konstanter und reproduzierbarer Werte wichtig ist. Diese Symmetrie fehlt sowohl bei der üblichen Arbeitsweise als bei der MICHAELIS'schen Anordnung.

3. Versuche einer experimentellen Realisation der Verhältnisse, die der PLANCK'schen Formel für das Diffusionspotential entsprechen.

<sup>1</sup> MICHAELIS, L. und FUJITA, A., Biochem. Zs. **142** (1923) 398.

<sup>2</sup> GUGGENHEIM, E. A., Journ. Am. Chem. Soc. **52** (1930) 1315.

PLANCK<sup>1</sup> hat eine Formel zur Berechnung des Diffusionspotentials zwischen zwei Lösungen gegeben. Die Formel setzt voraus, dass ein stationärer Zustand zwischen den beiden Ebenen *A* und *B* durch Diffusion aufrechterhalten wird, während sich ausserhalb *A* die Lösung I, ausserhalb *B* die Lösung II befindet. Dieser Fall ist experimentell kaum realisierbar. BJERRUM<sup>2</sup> hat darauf aufmerksam gemacht, dass man die gewünschte Situation durch die Bespülung beider Seiten eines Diaphragmas erreichen könne, und GUGGENHEIM<sup>3</sup> hat Messungen für *HCl* gegen *KCl* mit einer Kollodiummembran als Diaphragma ausgeführt. Das Potential zeigte sich bei diesen Messungen etwas abhängig von den Strömungsgeschwindigkeiten, und der gefundene Wert stimmte nicht ganz mit dem berechneten überein.

Nach den obenerwähnten Ergebnissen ist dies Resultat nicht erstaunlich. Obwohl die von GUGGENHEIM benutzte Membran gut durchlässig war<sup>4</sup>, wird doch vielleicht  $n'_$  etwas verschieden von  $n_-$  sein. Der unter Spülung gemessene *E*-Wert wird dann nicht  $E_D$ , sondern ein Wert zwischen  $E_M$  und  $E_D$  sein.

Das PLANCK'sche Potential lässt sich demnach nur messen, wenn man ein Diaphragma hat, das die Überföhrungszahlen der anwesenden Ionen nicht ändert. Die Dicke des Diaphragmas ist dann theoretisch belanglos. Man darf jedoch nicht vergessen, dass die dünne Schicht ausserhalb des Diaphragmas, in welcher durch die Spülung eine Mischungsschicht entsteht, so klein wie möglich im Ver-

<sup>1</sup> PLANCK, M., Wied. Ann. **39** (1890) 161.

<sup>2</sup> BJERRUM, N., Zs. Elektrochem. **17** (1911) 58.

<sup>3</sup> loc. cit.

<sup>4</sup> Die Membran war eine der Marke »fein« von der Vereinigung Göttinger Werke. Für diese Marke liegt  $D_{H_2O} \cdot 10^7$  zwischen 75 und 400.

hältnis zu dem Gebiet, wo nur Diffusion stattfindet, sein muss. Darum ist es zweckmässig ein nicht zu dünnes Diaphragma zu benutzen.

Es ist mir eine Freude die Gelegenheit zu benutzen, dem Vorstand des Laboratoriums, Professor Dr. NIELS BJERRUM, für die Anregung zu dieser Arbeit, sowie für sein freundliches Interesse während der Untersuchung herzlichst zu danken. Für bereitwillige Hilfe bei der Diskussion der mathematischen Probleme bringe ich den Herren Dr. RICH. PETERSEN und Dr. W. FENCHEL meinen besten Dank.

Das chemische Laboratorium der königlich tierärztlichen und landwirtschaftlichen Hochschule. Kopenhagen. Juni 1937.

Det Kgl. Danske Videnskabernes Selskab.

Mathematisk-fysiske Meddelelser. **XV**, 6.

---

# Hg-DYNAMICS I

THEORY OF THE LAMINAR FLOW OF AN ELECTRICALLY CONDUCTIVE LIQUID IN A HOMOGENEOUS MAGNETIC FIELD

BY

JUL. HARTMANN



KØBENHAVN

LEVIN & MUNKSGAARD

EJNAR MUNKSGAARD

1937

Printed in Denmark.  
Bianco Lunos Bogtrykkeri A/S.



## Introduction.

**Hg-Dynamics.** In 1918 the author designed a new pump, the electromagnetic pump, for use in connection with the so-called jet-wave rectifier. The pump is shown in fig. 1. Between the pole pieces  $NS$  of a strong electromagnet a gap is formed. The surfaces of the pole pieces, which form the walls of the gap, are covered with an insulating coating, and the gap is closed above and below by two electrodes  $E_1 E_2$ . We will assume that the channel thus formed is inserted at such a point in a hydrodynamic circuit with mercury, that it is always kept filled with this liquid. If, then, an electric current is passed through the gap from one electrode to the other the interaction between this current and the field will produce a "hydromotive force" in the direction of the axis of the channel, i. e. the apparatus will act as a pump. It is, as a matter of fact, at the same time an electromotor and a pump, the armature of the motor being the liquid to be transported by the pump.

The invention is, as will be seen, no very ingenious one, the principle utilized being borrowed directly from a well known apparatus for measuring strong magnetic fields. Neither does the device represent a particularly effective pump, the efficiency being extremely low due mainly to the

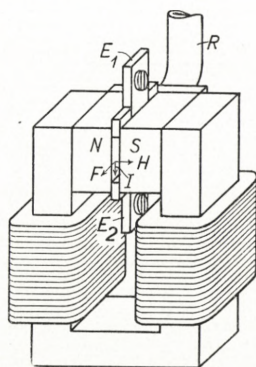


Fig. 1. Electromagnetic Pump.

large resistivity of mercury and still more to the contact resistance between the electrodes and the mercury. In spite hereof considerable interest was in the course of time bestowed on the apparatus, firstly because of a good many practical applications in cases where the efficiency is of small moment and then, during later years, owing to its inspiring nature. As a matter of fact the study of the pump revealed to the author what he considered a new field of

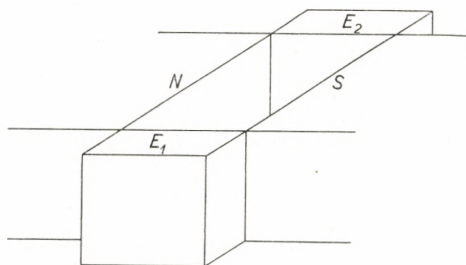


Fig. 2. Perpendicular Combination of homogeneous hydrodynamic, electric and magnetic Fields of Flow.

investigation, that of the flow of a conductive liquid in a magnetic field, a field for which the name Hg-dynamics was suggested<sup>1</sup>.

In the electromagnetic pump of the simple form described above we have to do with the system indicated in fig. 2. In the space confined between the pole pieces  $NS$  and the electrodes  $E_1E_2$  two homogeneous fields perpendicular to each other may be set up: a magnetic field and an electric current field. If the space forms part of a hydrodynamic circuit with mercury a third homogeneous — or nearly homogeneous — field perpendicular to the two others may be created, viz. a liquid current field. The theory of the electromagnetic pump is the theory of the

<sup>1</sup> Fourth International Congress for Applied Mechanics. Cambridge 1934. Abstracts of Papers p. 40.

interaction of these three fields and Hg-Dynamics in its general form is the theory of the same three fields in every possible combination, each of the fields being free to vary as well in space as in time.

It will at once be gathered that the complete analytical solution of the problems of Hg-dynamics is as a rule precluded. The distribution over the liquid medium of velocity, pressure, electromotive force, electric current, magnetic field intensity etc. is given by the combination of the ordinary hydrodynamical equations: — the equations of momentum and the equation of continuity — with the general equations of electrodynamics: — Faraday's Law of induction, Biot and Savart's Law, Ohm's Law, the Induction Law for moving conductors — the distribution being furthermore governed by the various mechanical, electric and magnetic boundary conditions. In spite of the tremendous intricacy of the analysis certain problems can be solved — exactly or approximately — because of their comparatively great simplicity. Other problems, while not susceptible of any theoretical analysis, nevertheless demand solution — owing to the interest attaching to them — and consequently attain solution through experimental analysis.

One of the problems which is capable of a fairly complete theoretical analysis under certain simple conditions is the laminar flow of a conductive liquid in a homogeneous magnetic field. An example of the other class of problems not accessible to such analysis is the turbulent flow of a conductive liquid in a homogeneous magnetic field<sup>1</sup>. Both problems have a bearing on the electromagne-

<sup>1</sup> It may here be stated that it was precisely the search for a means of influencing the turbulence of a liquid flow that led to the realisation of the existence of the new field of research, the field of Hg-Dynamics.

tic pump. The former is dealt with in the present theoretical paper. The experimental verification of the theory will be given in a following publication.

The author is indebted to several of his collaborators for suggestions in connection with the discussion of the problem dealt with in the present paper and other related problems. He particularly desires to thank Mr. LÖGSTRUP JENSEN for valuable assistance. He acknowledges with gratitude the receipt of financial aid granted by the Trustees of the Carlsberg Foundation in connection with the completion of the present work.

---

### 1. Differential Equation for the Velocity Distribution across a rectangular Channel with conductive Top and Bottom.

A channel with a rectangular cross-section, fig. 3, is considered. The width  $2a$  of the section is assumed small compared to the height  $h$ . In accordance herewith the

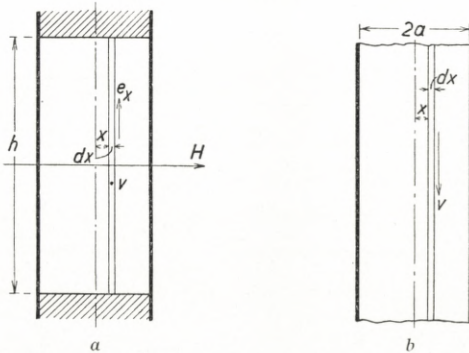


Fig. 3 a—b. Channel with rectangular Cross-Section.

effect of the upper and lower walls on the flow is neglected. These walls are formed of a highly conductive material, while the liquid of the flow is mercury. The magnetic field  $H$  is supposed to be perpendicular to the side walls of the channel.

The theory of the laminar flow with no field put on is well known. One  $\text{cm.}^2$  of a layer of thickness  $dx$  will be acted on by a resultant frictional force  $\eta \frac{d^2 v}{dx^2} \cdot dx$  in the direction of the flow,  $v$  being the velocity of the liquid and  $\eta$  the viscosity of the same. This force is compensated by a pressure drop  $\frac{P}{L} dx$ ,  $L$  being the length of the channel

and  $p$  the total pressure required to drive the liquid through the channel. So

$$(1) \quad \eta \frac{d^2 v}{dx^2} = -\frac{p}{L}$$

from which

$$(2) \quad v = \frac{1}{2} \frac{p}{\eta L} (a^2 - x^2), \quad (2a) \quad V = \int_{-a}^{+a} v h dx = \frac{2}{3} \frac{p}{\eta L} h a^3,$$

$V$  being the volume flow of the fluid. Hence there would be a parabolic velocity distribution across the channel.

Now, with the field on, electromotive forces are induced in the various layers, the force in an arbitrary layer being

$$(3) \quad e_x = 10^{-8} v_x h H \text{ Volt.}$$

The electromotive forces will give rise to a current distribution determined as follows. Let the voltage difference between the bottom and the top of the channel be  $E$ , then for an arbitrary layer the current  $I_x dx$ ,  $I_x$  being the current density, is determined by

$$(4) \quad E + e_x = I_x dx \cdot L \cdot \kappa \cdot \frac{h}{L dx} = I_x \kappa h,$$

$\kappa$  being the specific resistance of the liquid.

Again the total electric current flow from the bottom to the top is zero, i. e.

$$(5) \quad \int_{-a}^{+a} I_x dx = 0.$$

Introducing the expression for  $I_x$  taken from (4) into (5) we get

$$(6) \quad \int_{-a}^{+a} (E + e_x) dx = 0$$

or

$$(7) \quad E = -\frac{1}{2a} \int_{-a}^{+a} e_x dx = -e_m,$$

$e_m$  being the mean value of the electromotive force. So

$$(8) \quad I_x = \frac{1}{zh} (e_x - e_m) = 10^{-8} \frac{H}{z} (v_x - v_m) \text{ Amp./cm}^2$$

where

$$(9) \quad v_m = \frac{1}{2a} \int_{-a}^{+a} v_x dx.$$

Now the pressure gradient to which the interaction of the current and the field gives rise is

$$(10) \quad \dot{p}_e = \frac{1}{10} I_x H = 10^{-9} \frac{H^2}{z} (v_x - v_m)$$

and so the total pressure drop in the layer is  $\dot{p}_e L$  i. e.

$$(11) \quad p_e = 10^{-9} \frac{LH^2}{z} (v_x - v_m).$$

It is tending to hamper the motion of the flow while the frictional force tends (formally) to speed up the motion. So it will be gathered that the pressure required to maintain the flow is now determined by

$$(12) \quad p = 10^{-9} \frac{LH^2}{z} (v_x - v_m) - \eta L \frac{d^2 v_x}{dx^2}$$

or by

$$(13) \quad p = 10^{-9} \frac{LH^2}{z} \left( v - \frac{1}{2a} \int_{-a}^{+a} v dx \right) - \eta L \frac{d^2 v}{dx^2},$$

$v$  being written instead of  $v_x$ . This equation determines the steady velocity distribution after the magnetic field is put on.

We will assume that with no field put on there is a steady flow characterised by

$$(14) \quad v = \frac{1}{2} \frac{P}{\eta L} (a^2 - x^2).$$

Then, suddenly, the magnetic field is set on. We ask for the current distribution at the first moment when the velocity distribution is still unaltered. The current distribution is determined by

$$(15) \quad I = 10^{-8} \frac{H}{z} (v - v_m)$$

where

$$(16) \quad v_m = \frac{1}{2a} \int_{-a}^{+a} v dx = p \cdot \frac{a^2}{3 \eta L}.$$

Introducing  $v$  and  $v_m$  taken from the first and the third formula into (15), we find

$$(17) \quad I = 10^{-8} \frac{H}{2z\eta} \cdot \frac{p}{L} \left( \frac{a^2}{3} - x^2 \right).$$

Finally the electromagnetic pressure gradient is determined by

$$(18) \quad \dot{p}_e = 10^{-9} \frac{H^2}{2z\eta} \cdot \frac{p}{L} \left( \frac{a^2}{3} - x^2 \right).$$

We may consider an arrangement for which  $a = 0.125$  cm.,  $z = 10^{-4}$  Ohm·cm.,  $\eta = 0.0159$  (c. g. s.),  $H = 10^4$  Gauss (Liquid: Mercury). With these values the formulae (14), (15) and (18) become

$$v = 31.4 \frac{P}{L} \cdot (1.56 \cdot 10^{-2} - x^2) \text{ cm./sec.}$$

$$I = 31.4 \frac{P}{L} \cdot (0.520 \cdot 10^{-2} - x^2) \text{ Amp./cm.}^2$$

$$\dot{p}_e = 31.4 \cdot 10^3 \frac{P}{L} \cdot (0.520 \cdot 10^{-2} - x^2) \text{ dyne/cm.}^2.$$



In the middle plane  $x = 0$  all the three quantities are maximum. Their values are

$$v_{\max} = 0.490 \frac{P}{L} \text{ cm./sec.},$$

$$I_{\max} = 0.163 \frac{P}{L} \text{ Amp./cm.}^2,$$

$$\dot{p}_{e, \max} = 163 \frac{P}{L} \text{ dyne/cm.}^2.$$

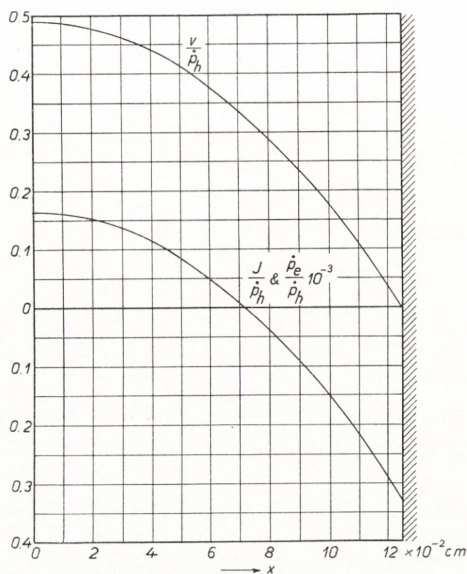


Fig. 4. Distribution of Velocity, Current and electromagnetic Pressure Gradient at the Moment the Field is set on.

Thus, it is seen, that at the first moment the electromagnetic pressure gradient is 163 times greater than the hydrodynamic pressure gradient required to maintain the viscous flow. Fig. 4 shows graphically the distribution of the quantities  $v/\dot{p}_h$ ,  $I/\dot{p}_h$  and  $\dot{p}_e/\dot{p}_h$ , where  $\dot{p}_h = p/L$ .

## 2. Solution of the Differential Equation for the Velocity Distribution.

The differential equation determining the velocity distribution across the channel was

$$(1) \quad p = 10^{-9} \frac{LH^2}{z} \left( v - \frac{1}{2a} \int_{-a}^{+a} v \, dx \right) - \eta L \frac{d^2 v}{dx^2},$$

or

$$(2) \quad \frac{d^2 v}{dx^2} - 10^{-9} \frac{H^2}{\eta z} \cdot v = -\frac{p}{\eta L} - 10^{-9} \frac{H^2}{\eta z} \cdot \frac{1}{2a} \int_{-a}^{+a} v \, dx.$$

Substituting

$$(3) \quad 10^{-9} \frac{H^2}{\eta z} = A^2,$$

$$(4) \quad \frac{p}{\eta L} = B,$$

$$(5) \quad 10^{-9} \frac{H^2}{\eta z} \cdot \frac{1}{2a} \int_{-a}^{+a} v \, dx = \frac{A^2}{2a} \int_{-a}^{+a} v \, dx = C,$$

where  $C$  is an unknown constant, (2) may be written

$$(6) \quad \frac{d^2 v}{dx^2} - A^2 v = -B - C,$$

the solution of which is

$$(7) \quad v = \frac{B+C}{A^2} + C_1 e^{Ax} + C_2 e^{-Ax}.$$

From the two conditions

$$(8) \quad v = 0, \quad \text{when } x = \pm a$$

it is found that

$$(9) \quad C_1 = C_2 = -\frac{B+C}{A^2} \cdot \frac{1}{e^{Aa} + e^{-Aa}},$$

hence

$$(10) \quad v = \frac{B+C}{A^2} \left( 1 - \frac{e^{Ax} + e^{-Ax}}{e^{Aa} + e^{-Aa}} \right).$$

Inserting this expression into (5), the following equation for the determination of  $C$  is derived

$$(11) \quad \frac{A^2}{2a} \int_{-a}^{+a} \frac{B+C}{A^2} \left( 1 - \frac{e^{Ax} + e^{-Ax}}{e^{Aa} + e^{-Aa}} \right) dx = C,$$

from which

$$(12) \quad C = B \left( Aa \cdot \frac{e^{Aa} + e^{-Aa}}{e^{Aa} - e^{-Aa}} - 1 \right).$$

Introducing into (10) we obtain the final formula for  $v$ :

$$(13) \quad v = \frac{Ba}{A} \cdot \frac{e^{Aa} + e^{-Aa}}{e^{Aa} - e^{-Aa}} \left( 1 - \frac{e^{Ax} + e^{-Ax}}{e^{Aa} + e^{-Aa}} \right),$$

or in hyperbolic functions

$$(14) \quad v = \frac{Ba}{A} \cdot \frac{\cosh Aa - \cosh Ax}{\sinh Aa}.$$

### 3. Influence of the magnetic Field on the Velocity Distribution.

We may now, first of all, discuss the influence of the magnetic field on the velocity distribution across the rectangular duct. Assuming in the first instance that  $Aa = z_0$  is markedly smaller than unity we may in the formula (13) of the preceding paragraph develop the exponential terms in series. Neglecting terms of a higher power than the fourth we find

$$(1) \quad v = \frac{Ba^2}{z_0} \cdot \frac{z_0^2 - z^2 + \frac{1}{12}(z_0^4 - z^4)}{2z_0 + \frac{z_0^3}{3}}$$

where  $z$  stands for  $Ax$ . This formula may be written

$$(2) \quad v = \frac{p}{2\eta L} \cdot (a^2 - x^2) \left( 1 - \frac{A^2}{12} (a^2 - x^2) \right)$$

from which the average velocity over the cross-section is found to be

$$(3) \quad v_m = \frac{P}{3\eta L} \cdot a^2 \left[ 1 - \frac{1}{15} (Aa)^2 \right].$$

On comparing formula (2) to formula (2) of paragraph 1 it is seen that the effect of the magnetic field, when  $Aa \ll 1$ , is to reduce the velocity of each layer of the flow slightly (and so also the average velocity). Let  $Aa = 0.5$ , then the reduction of the central velocity ( $x = 0$ ) is, as will be seen, about 2 p. c. and that of the average velocity about 1.7 p. c. Now  $A^2 a^2 = 10^{-9} \frac{H^2}{\eta z} \cdot a^2$ . So, with  $Aa = 0.5$ ,  $H^2 a^2 = \frac{\eta z}{4} \cdot 10^9$ . With mercury  $z = 10^{-4}$  Ohm · cm. approximately and  $\eta = 0.0160$  c. g. s. In case of  $a = 0.125$  cm. we find that  $H$  must be equal to 160 Gauss to make  $Aa = 0.5$ . With a 10 p. c. solution of NaCl  $\eta = 0.01$  c. g. s. and  $z = 8.25$  Ohm · cm. approximately from which  $Ha = 4.55 \cdot 10^3$  with  $Aa = 0.5$ . So, if  $H$  is taken to be 1000 Gauss  $a$  must be chosen equal to 0.445 cm. in order to make  $Aa = 0.5$ .

Now, if on the other hand  $Aa$  is very large so that  $e^{-Aa}$  is quite negligible compared to  $e^{Aa}$ , a great change in the average velocity of the flow and at the same time in the velocity distribution may ensue. The average velocity  $v_{m \cdot 2}$  is equal to  $\frac{C}{A^2}$  (paragraph 2, equation (5)). So from the formula (12) of paragraph 2 it follows that, under the conditions referred to, we have

$$(4) \quad v_{m \cdot 2} = \frac{B}{A^2} (Aa - 1)$$

or approximately

$$(5) \quad v_{m \cdot 2} = \frac{Ba}{A} = \frac{p}{\eta L} \cdot \frac{\sqrt{z\eta}}{H\sqrt{10^{-9}}} \cdot a.$$

The average velocity with no field on is

$$(6) \quad v_{m.1} = \frac{P}{3\eta L} \cdot a^2.$$

Hence

$$(7) \quad \frac{v_{m.1}}{v_{m.2}} = \frac{a}{3\sqrt{\eta z}} \cdot \frac{H}{10^4\sqrt{10}}.$$

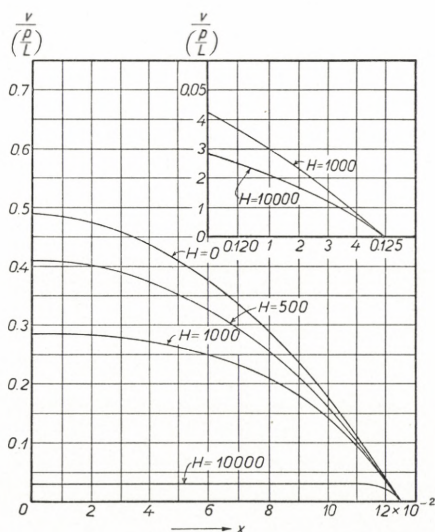


Fig. 5. Velocity Distributions when the Flow has become steady.

With mercury, with  $a = 0.125$  cm. and  $H = 10000$  Gauss  $Aa = 31.3$  and  $e^{Aa} = 38.0 \cdot 10^{12}$  while  $e^{-Aa}$  is extremely small. So the condition for the approximation is amply fulfilled and we find  $\frac{v_{m.1}}{v_{m.2}} = 10.45$ . Thus in the case considered the average velocity is reduced to  $\frac{1}{10}$  by the field, it being assumed that the pressure gradient is kept unchanged. At the same flow of liquid  $\frac{p_1/L_1}{p_2/L_2} = \frac{v_{m.2}}{v_{m.1}}$ . Thus in our example the pressure gradient must with the field on be 10 times what it is with no field on in order to produce the same flow.

Finally fig. 5 represents the velocity distribution over one half of a 0.250 cm. wide channel corresponding to four values of the magnetic field. It is assumed that the liquid is mercury and that the flow is laminar. The velocities correspond to unit pressure gradient. It is seen that the effect of the field is to reduce the average velocity and to flatten out the velocity distribution curve. At a field  $H = 10000$  Gauss the velocity has become constant over most of the section.

#### 4. The electromagnetic Viscosity.

Returning to the general expression for  $v$ , equation (13) paragraph 2, we may now derive the corresponding formula for the average velocity over the cross-section. It is found to be

$$\begin{aligned}
 v_m &= \frac{1}{2a} \int_{-a}^{+a} v \, dx = \frac{Ba}{A} \cdot \frac{e^{Aa} + e^{-Aa}}{e^{Aa} - e^{-Aa}} \left[ 1 - \frac{1}{Aa} \cdot \frac{e^{Aa} - e^{-Aa}}{e^{Aa} + e^{-Aa}} \right] \\
 (1) \qquad &= \frac{p}{\eta L} a^2 \cdot \frac{1 - \frac{\tanh z_0}{z_0}}{\frac{\tanh z_0}{z_0}} \cdot \frac{1}{z_0^2}.
 \end{aligned}$$

The volume flow through the duct is

$$(2) \qquad V = v_m \cdot 2ah = \frac{2p}{\eta} \cdot \frac{a^3 h}{L} \cdot \frac{1 - \frac{\tanh z_0}{z_0}}{\frac{\tanh z_0}{z_0}} \cdot \frac{1}{z_0^2}.$$

Comparing (2) with formula (2a) paragraph 1 it is seen that the volume flow with the magnetic field on is the same that would ensue without the field, but with the same pressure drop, if the viscosity were

$$(3) \quad \eta'_e = \eta \cdot \frac{z_0^2}{3} \cdot \frac{\tanh z_0}{1 - \frac{\tanh z_0}{z_0}}$$

For  $\eta'_e$  the term apparent or virtual viscosity may be used. Assuming  $z_0 = Aa$  sufficiently small compared to 1 we may use the approximation

$$\frac{\tanh z_0}{z_0} = 1 - \frac{z_0^2}{3} + \frac{2}{15} z_0^4, \quad (z_0 \ll 1)$$

giving

$$(4) \quad \eta'_e = \eta + \frac{1}{15} 10^{-9} \frac{H^2 a^2}{\varkappa} = \eta + \eta_e$$

where

$$(5) \quad \eta_e = \frac{1}{15} 10^{-9} \frac{H^2 a^2}{\varkappa}$$

may be termed the electromagnetic viscosity. Formula (4) obviously represents the variation of the apparent viscosity with the field intensity  $H$ , when  $H$  is small (or the channel very narrow).

If, on the other hand,  $z_0$  is large compared to 1 (strong fields), then we may write

$$(6) \quad \eta'_e = \eta \cdot \frac{z_0^2}{3} \cdot \frac{1}{z_0 - 1} = \eta \frac{z_0}{3} \left( 1 + \frac{1}{z_0} \right)$$

or

$$(6') \quad \eta'_e = \frac{1}{3} \cdot 10^{-5} \sqrt{10} \cdot \frac{Ha}{\sqrt{\varkappa}} \cdot \sqrt{\eta} + \frac{\eta}{3} = \frac{\sqrt{15}}{3} \sqrt{\eta \eta_e} + \frac{\eta}{3}$$

So while with feeble fields the apparent viscosity — or rather the electromagnetic viscosity — varies in a parabolic manner with the field intensity  $H$ , the apparent viscosity varies linearly with  $H$  when  $H$  is large. These features of the electromagnetic viscosity are clearly elucidated by a graph of the function

$$f(z_0) = z_0^2 \cdot \frac{\frac{\tanh z_0}{z_0}}{1 - \frac{\tanh z_0}{z_0}}$$

In fig. 6 such a graph is shown.

Finally we may draw attention to a peculiar interpretation which may be given of the electromagnetic viscosity

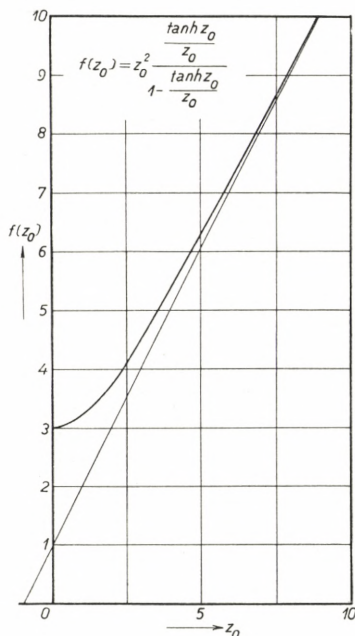


Fig. 6. Graph of the Function  $f(z_0) = z_0^2 \frac{\frac{\tanh z_0}{z_0}}{1 - \frac{\tanh z_0}{z_0}}$ .

$\eta_e$  (equation (5)). This quantity is  $\frac{H^2 a^2}{z}$  multiplied by a dimensionless factor. Let us assume a beam cut out of the liquid across the channel in the way illustrated in fig. 7. Let the width and the height of the beam both be 1 cm. and let these two dimensions be in the direction of



the axis of the channel and perpendicular to the same respectively. The length of the beam is  $2a$ . Now the volume of the beam is obviously  $2a$  so  $2H^2a$  is proportional to the magnetic energy stored up in the space occupied by the beam. The resistance met by the induced current in the beam is  $\frac{\alpha}{2a}$ . So the virtual viscosity is simply proportional to the ratio of the magnetic energy in the beam considered and the electric resistance of the fluid in the

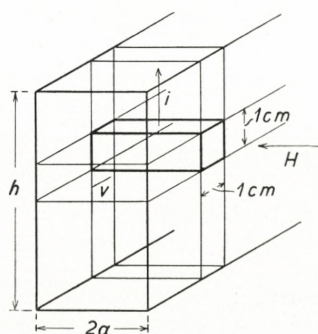


Fig. 7. Diagram illustrating the Nature of the electromagnetic Viscosity.

same beam perpendicularly to the magnetic field and to the flow. Replacing  $\frac{H^2 a^2}{\alpha}$  by  $\frac{H^2 a^2 h}{\alpha h}$  it is seen that the virtual viscosity is also proportional to the ratio of the magnetic energy in a slice of 1 cm. thickness cut out of the flow and the electric resistance of the fluid in this slice taken perpendicularly to the motion and to the field.

### 5. The Channel short-circuited through a Conductor containing Resistance and electromotive Force.

We shall now imagine the top and the bottom of the channel in fig. 3 to be connected through a conductor containing a resistance  $R$  and an electromotive force  $E$ . Then

the current density in an arbitrary layer of the flow is determined by

$$(1) \quad e + E = Izh + RL \int_{-a}^{+a} Idx,$$

where

$$(2) \quad e = 10^{-8} h H v.$$

Again it is assumed that a pressure gradient  $\dot{p}$  is maintained in the flow. Hence we have

$$(3) \quad \dot{p} = \frac{1}{10} IH - \eta \frac{d^2 v}{dx^2}.$$

Introducing  $e$  and  $I$  taken from (2) and (3) respectively into equation (1) this equation assumes the shape

$$(4) \quad \begin{aligned} 10^{-8} h H v + E &= \frac{10}{H} \dot{p} z h + \frac{10}{H} \eta z h \frac{d^2 v}{dx^2} \\ &+ \frac{10}{H} \dot{p} RL \cdot 2a + \frac{10}{H} \eta RL \left[ \left( \frac{dv}{dx} \right)_a - \left( \frac{dv}{dx} \right)_{-a} \right]. \end{aligned}$$

Noting that  $\left( \frac{dv}{dx} \right)_{-a} = - \left( \frac{dv}{dx} \right)_a$  we may write (4) in the form

$$(5) \quad \frac{d^2 v}{dx^2} - \frac{10^{-9} H^2}{z \eta} v + \frac{\dot{p}}{\eta} + \frac{\dot{p}}{\eta} \cdot \frac{R}{R_c} - \frac{EH}{10 \eta z h} + 2 \frac{RL}{z h} \cdot \left( \frac{dv}{dx} \right)_a = 0$$

where  $R_c = \frac{z h}{2 a L}$  is the electric resistance in the channel between top and bottom. Substituting

$$(6) \quad 10^{-9} \frac{H^2}{\eta z} = A^2,$$

$$(7) \quad \frac{\dot{p}}{\eta} = B,$$

$$(8) \quad \frac{\dot{p}}{\eta} \cdot \frac{R}{R_c} - \frac{EH}{10 \eta z h} = D,$$

$$(9) \quad \frac{2RL}{zh} \cdot \left(\frac{dv}{dx}\right)_a = C$$

where  $A$  and  $B$  thus stand for the same qualities as in paragraph 2, equation (5) may be written

$$(10) \quad \frac{d^2 v}{dx^2} - A^2 v + B + D + C = 0.$$

Noting that  $v = 0$  for  $x = a$  and  $x = -a$  the solution is found to be

$$(11) \quad v = \frac{B + D + C}{A^2} \left(1 - \frac{e^{Ax} + e^{-Ax}}{e^{Aa} + e^{-Aa}}\right).$$

$C$  is now found by differentiating equation (11):

$$(12) \quad \frac{dv}{dx} = -\frac{B + D + C}{A} \cdot \frac{e^{Ax} - e^{-Ax}}{e^{Aa} + e^{-Aa}}.$$

Introducing  $x = a$  we find

$$(13) \quad \left(\frac{dv}{dx}\right)_a = -\frac{B + D + C}{A} \cdot \frac{e^{Aa} - e^{-Aa}}{e^{Aa} + e^{-Aa}} = \frac{C}{2RL} zh$$

from which

$$(14) \quad C = -\frac{\frac{B + D}{A} \cdot \frac{e^{Aa} - e^{-Aa}}{e^{Aa} + e^{-Aa}}}{A^2 \left(\frac{zh}{2RL} + \frac{1}{A} \cdot \frac{e^{Aa} - e^{-Aa}}{e^{Aa} + e^{-Aa}}\right)}.$$

Introducing finally into (11) we arrive at the formula:

$$(15) \quad v = \frac{B + D}{A^2} \frac{\frac{zh}{2RL} A}{\frac{zh}{2RL} \cdot A + \frac{e^{Aa} - e^{-Aa}}{e^{Aa} + e^{-Aa}}} \left(1 - \frac{e^{Ax} + e^{-Ax}}{e^{Aa} + e^{-Aa}}\right).$$

Comparing with formula (13) in paragraph 2 it is readily seen that the shape of the velocity distribution

curve is not altered by connecting the top and the bottom by an external conductor. The shape is determined solely by the value of  $Aa$ . The external current branch and what is in it influence the intensity of the flow, i. e. the average value of  $v$ , only.

### 6. Discussion of the Theory for the electrically short-circuited Channel.

For a fuller discussion of formula (15) in the preceding paragraph we may introduce the values of  $B$  and  $D$ . Further it should be noted that  $\frac{zh}{2RL} \cdot A = \frac{R_c}{R} Aa$ . Finally, in expression (8) paragraph 5 for  $D$ , the quality  $\frac{E}{zh}$  is obviously the current density  $I_0$  which the electromotive force  $E$  would create in the channel if the latter were filled with liquid at rest and if the external conductor had no resistance. So the last term in (8) is simply  $\frac{\dot{p}_e}{\eta}$  where  $\dot{p}_e$  is the pressure gradient to which  $I_0$  would give rise. Taking these facts into account it is readily seen that expression (15) of paragraph 5 can be written:

$$(1) \quad v = \frac{\frac{\dot{p}}{\eta} \left(1 + \frac{R}{R_c}\right) - \frac{\dot{p}_e}{\eta}}{A^2} \cdot \frac{\frac{R_c}{R} Aa}{\frac{R_c}{R} Aa + \frac{e^{Aa} - e^{-Aa}}{e^{Aa} + e^{-Aa}}} \cdot \left(1 - \frac{e^{Ax} + e^{-Ax}}{e^{Aa} + e^{-Aa}}\right).$$

If  $R = \infty$  and  $E = 0$  we have returned to the case considered in paragraphs (1)—(4) and it is readily seen that (1) becomes identical with formula (13) of paragraph (2). Again if  $E = 0$ ,  $R = 0$  we simply have the channel short-circuited by an external conductor without resistance. Then (1) assumes the shape

$$(2) \quad v = \frac{B}{A^2} \left( 1 - \frac{e^{Ax} + e^{-Ax}}{e^{Aa} + e^{-Aa}} \right).$$

The effect of the short-circuiting is a reduction of the velocity. Comparing with formula (13) of paragraph 2 it is seen that the ratio of the velocities with and without short-circuiting is

$$(3) \quad \frac{v_s}{v_0} = \frac{1}{Aa} \cdot \frac{e^{Aa} - e^{-Aa}}{e^{Aa} + e^{-Aa}}.$$

In the case of a channel of width  $2a = 0.250$  cm. and with a flow of mercury in a field of 10000 Gauss  $Aa = 31.3$ . As with this value the last factor in (3) is 1,  $\frac{v_s}{v_0} = \frac{1}{Aa} = \frac{1}{31.3}$ . So with short-circuiting the velocity is reduced to about 3 p. c. of what it would be without short-circuiting. From this it will be gathered that the short-circuited channel may represent an extremely effective "brake" to the motion of the fluid.

The "braking" effect may be controlled by a resistance in the external conductor. The effect of the resistance is clearly seen from equation (1). Assuming  $\dot{p}_e = 0$ , i. e.  $E = 0$ , we find for the ratio of the velocities with and without resistance in the external conductor:

$$(4) \quad \frac{v_R}{v_S} = \left( 1 + \frac{R}{R_c} \right) \frac{\frac{R_c}{R} Aa}{\frac{R_c}{R} Aa + \frac{e^{Aa} - e^{-Aa}}{e^{Aa} + e^{-Aa}}}.$$

If  $e^{Aa}$  is large compared to  $e^{-Aa}$  (4) may be written

$$(5) \quad \frac{v_R}{v_S} = \frac{1 + \frac{R}{R_c}}{1 + \frac{1}{Aa} \cdot \frac{R}{R_c}}.$$

In the example above  $Aa = 31.1$ . In this case, if  $R = R_c$ ,  $\frac{v_R}{v_S} = 2$  approximately and if  $R = nR_c$ ,  $\frac{v_R}{v_S}$  increases as  $n$ , approximately, as long as  $n$  is a smaller number than 1, 2 or 3. With increasing  $n$   $\frac{v_R}{v_S}$  approaches the limit  $Aa$ .

Finally we may consider the general case in which there is both resistance and electromotive force in the external branch of the electric circuit. Then it is seen from (1) that the flow may be stopped altogether by applying such an electromotive force (in the direction of that induced by the flow) that

$$(6) \quad \dot{p}_e = \dot{p} \left( 1 + \frac{R}{R_c} \right)$$

or equal to the impressed pressure gradient proper if  $R$  is small compared to  $R_c$ . If, now,  $\dot{p}_e$  is increased beyond the value given by (6) then the flow will change its direction and the channel with its circuit and field will act as an electromagnetic pump. So the expression (1) forms part of the theory of this apparatus and shows that the velocity distribution in the channel is quite independent of the velocity of the flow and only dependent on the quantity

$$(7) \quad Aa = \frac{H}{10^4 \sqrt{10 \pi \eta}} \cdot a.$$

With large values of  $Aa$  the velocity is as seen from fig. 5 almost constant over the whole width of the channel. This information was derived on the supposition of a laminar flow. The very effective smoothing out caused by a strong magnetic field suggests that the results of our discussion may also prove applicable to a flow which without the magnetic field would be turbulent, while with the field it is made laminar.

### 7. Characteristics of the electromagnetic Pump.

The volume flow  $V$  through an electromagnetic pump may be derived from the expression (1) in paragraph 6. Seeing that the average value of the last factor of this expression is determined by

$$(1) \quad \frac{1}{2a} \int_{-a}^{+a} \left( 1 - \frac{e^{Aa} + e^{-Aa}}{e^{Aa} + e^{-Aa}} \right) dx = 1 - \frac{1}{Aa} \frac{e^{Aa} - e^{-Aa}}{e^{Aa} + e^{-Aa}}$$

we find for the volume flow

$$(2) \quad V = 2ha \cdot \frac{\dot{p}}{\eta} \left( 1 + \frac{R}{R_c} \right) - \frac{\dot{p}_e}{\eta} \cdot \frac{\frac{R_c}{R} \cdot Aa}{\frac{R_c}{R} \cdot Aa + \frac{e^{Aa} - e^{-Aa}}{e^{Aa} + e^{-Aa}}} \left( 1 - \frac{1}{Aa} \frac{e^{Aa} - e^{-Aa}}{e^{Aa} + e^{-Aa}} \right) \text{ cm.}^3/\text{sec.}$$

If here  $R$  is small compared to  $R_c$  as it ought to be and if further  $Aa$  is large, we may replace (2) by

$$(3) \quad V = 2ha \cdot \frac{\dot{p} - \dot{p}_e}{A^2} \cdot \left( 1 - \frac{1}{Aa} \right) \text{ cm.}^3/\text{sec.}$$

or if  $Aa$  is large compared to 1

$$(4) \quad V = \frac{2ha}{\eta A^2} (\dot{p} - \dot{p}_e) \text{ cm.}^3/\text{sec.}$$

from which

$$(5) \quad \dot{p} = \dot{p}_e + V \cdot \frac{\eta A^2}{2ha}.$$

The total pressure difference  $p$  between the inlet and outlet of the pump is  $\dot{p}L$  (apart from the losses at the inlet and outlet). Hence

$$\begin{aligned}
 (6) \quad p &= \dot{p}_e L + V \cdot \frac{\eta A^2}{2 ha} \cdot L \\
 &= \dot{p}_e L + 10^{-9} \frac{H^2}{2 ha z} \cdot LV.
 \end{aligned}$$

Seeing now that with the pump  $V$  is to be taken as negative and that  $\dot{p}_e L$  is "the manometric pressure"  $p_0$  of the pump we have

$$(7) \quad p = p_0 - 10^{-9} \frac{H^2}{2 ha z} \cdot LV$$

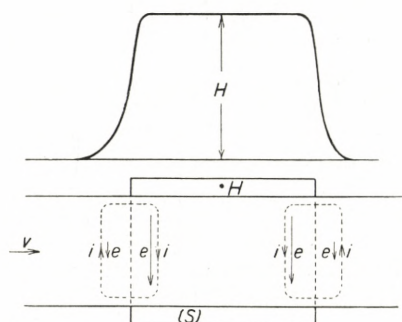


Fig. 8. Origin of the Pressure Loss at the Boundaries of the magnetic Field.

$V$  being the numerical value of the flow. So the  $p$ - $V$  characteristic should be a straight line or the pressure drop  $p_{h+e} = p_0 - p$  should be so:

$$(8) \quad p_{h+e} = p_0 - p = 10^{-9} \frac{H^2}{2 ha z} \cdot LV.$$

With the actual pump there is an additional pressure loss at the boundaries of the magnetic field. This will readily be understood from fig. 8. Here the channel is seen located within and outside a magnetic field produced between two pole pieces  $HS$ , one ( $S$ ) in front of and the other behind the channel. The field distribution curve too



is shown, it extends somewhat beyond the pole-pieces. The flow gives rise to electromotive forces which are smaller outside the pole-pieces than within. So circulating currents  $i$  must arise at the boundaries of the magnetic field having the directions indicated in the figure. Obviously the currents will cause hydromotive forces to be set up which at both boundaries will have a resultant opposing the flow of the liquid.

Now in order to develop an expression for the pressure loss at the two boundaries we may consider a greatly simplified model, fig. 9.

Here  $C$  is the channel,  $B$  is one of the boundary lines of the pole-pieces. To the left and right of  $B$  are two zones of width  $z$  in the direction of the flow. Within the left-hand zone the magnetic field has a constant value  $H_1$ , within the right-hand zone the field is  $H_2$ . We shall imagine three walls  $W_1$ ,  $W_2$  and  $W$  to be arranged. Of these  $W_1$  and  $W_2$  cover the full area of cross-section of the flow while  $W$  leaves two openings above and below, the width of these openings being  $z$ . The three walls possess the rather peculiar quality of permitting the hydraulic flow to pass without obstruction, while forming an absolute hindrance to the electric currents. To the left of  $W_1$  the field is constant, equal to  $H_1$ , to the right of  $W_2$  it is zero.

With this imaginary system a circulating current is set up the density  $I$  of which is obviously determined by

$$(9) \quad I = \frac{e_1 - e_2}{2hz} = \frac{10^{-8}v}{2z} \cdot (H_1 - H_2) \text{ Amp./cm.}^2.$$

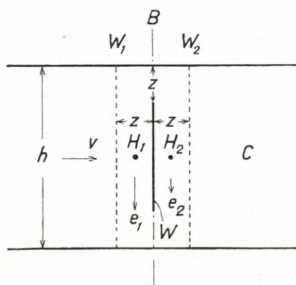


Fig. 9. To the Theory of the Pressure Loss at the Boundaries of the magnetic Field.

The corresponding pressure gradients within the left and right hand zone are respectively  $-\frac{1}{10}H_1I$  and  $\frac{1}{10}H_2I$ . So the pressure loss is

$$(10) \quad \Delta p = \frac{1}{10}zI(H_1 - H_2) = 10^{-9} \frac{z}{2z} \cdot (H_1 - H_2)^2 v \frac{\text{dyne}}{\text{cm.}^2}.$$

Assuming now that  $H_2 = 0$  and introducing the volume flow determined by  $V = 2ah \cdot v$ , (10) may be written

$$(11) \quad \Delta p = 10^{-9} \frac{zH^2}{2ahz} \cdot V.$$

Twice this quantity — corresponding to the two boundaries — should now be added to the pressure drop occurring inside the field and given by formula (8). So the total electromagnetic-hydraulic pressure drop in the pump is represented by

$$(12) \quad p_0 - p + \Delta p = 10^{-9} \frac{H^2}{2ahz} (L + z) V.$$

Thus it will be seen that the boundary effect may probably be taken into account by simply adding a certain length  $z$  to the actual length  $L$  of the channel or by replacing  $L$  by an effective length  $L_e$  somewhat greater than  $L$ . The correction  $z$  is likely to be proportional to the height  $h$  of the channel. Otherwise  $z$  also depends on the shape of the magnetic field curve and on the design of the outlet and inlet end of the pump when these ends are within the domain of the field or close to its boundaries.

Provisional Laboratory of Technical Physics. Royal Technical College, Copenhagen, May 1937.

Det Kgl. Danske Videnskabernes Selskab.  
Mathematisk-fysiske Meddelelser. **XV**, 7.

---

# Hg-DYNAMICS II

EXPERIMENTAL INVESTIGATIONS  
ON THE FLOW OF MERCURY IN A HOMOGENEOUS  
MAGNETIC FIELD

BY

JUL. HARTMANN AND FREIMUT LAZARUS



KØBENHAVN  
LEVIN & MUNKSGAARD  
EJNAR MUNKSGAARD  
1937

Printed in Denmark.  
Bianco Lunos Bogtrykkeri A/S.

### Introduction.

The main purpose of the investigations described in the present paper has been to study the influence of a homogeneous magnetic field on the flow of a conductive liquid — mercury — in pipes of circular or rectangular section. In a previous paper<sup>1</sup> this influence was examined theoretically on the assumption of a laminar flow in a flat channel of rectangular section, the field being perpendicular to the channel-sides of largest extension and the two other sides being formed by electrodes of highly conductive material. The experiments to be considered in the following were planned chiefly with a view to testing the main results of the theoretical discussion. They have, however, in addition thrown light upon phenomena not readily open to such discussion, in particular upon the influence of a homogeneous magnetic field on a turbulent flow and the transition of the turbulent form of flow into the laminar. In this respect the present work may be regarded as an extension of an investigation performed several years ago in the same laboratory and having as subject the comparison of the flow of water and mercury in pipes<sup>2</sup>. The

<sup>1</sup> Theory of the laminar Flow of an electrically conductive Liquid in a homogeneous magnetic Field. Det kgl. Danske Vidensk. Selsk. math.-fys. Medd. XV, 6, 1937.

<sup>2</sup> A Comparison between the Flow of Water and Mercury in Pipes etc. Mémoires de l'Académie Royale des Sciences de Danemark, Section des Sciences 8<sup>me</sup> Série, t. X, n<sup>o</sup> 5. 1926.

latter investigation showed that the REYNOLDS' Law of Similarity holds good for mercury even in cases in which the walls of the pipe are not wetted by this liquid. The investigations here reported amply confirm this fundamental experience.

Professor HARTMANN desires in this place to express his gratitude to the Trustees of the Carlsberg Foundation and of The H. C. Ørsted's Fund for having rendered possible by financial aid the completion of the present research work.

---

## I. The experimental Arrangements.

### The general Arrangement.

In fig. 1 a diagram is given of the general experimental arrangement. A hydrodynamic circuit is used in which the flow is maintained by an electromagnetic pump  $P$ .<sup>1</sup>  $R$  is

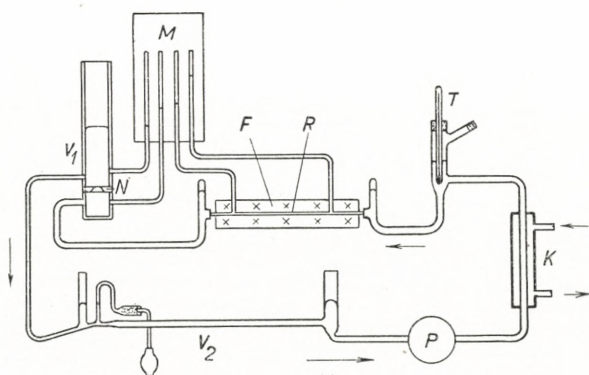


Fig. 1. General experimental Arrangement.

the duct — a glass tube or a rectangular channel in which the flow is examined. It is arranged in a homogeneous magnetic field  $F$  between the polepieces of a special electromagnet. The pressure drop in a certain length of the duct is observed by means of two manometer tubes connected to sockets on the duct. The volume of mercury passing each section of the circuit pr. sec. is measured by a simple flow meter of the calibrated nozzle pattern  $V_1$ . For rapid control of the volume flow a special flow meter  $V_2$  was

<sup>1</sup> The principle of this pump is described in the paper referred to in note 1, p. 3.

furnished. As the mercury was heated by the current passed through the electromagnetic pump a cooling device  $K$  had to be introduced, the temperature being controlled by the thermometer  $T$  and kept at  $20^{\circ}\text{C}$ . A photograph showing part of the circuit and particularly the electro-

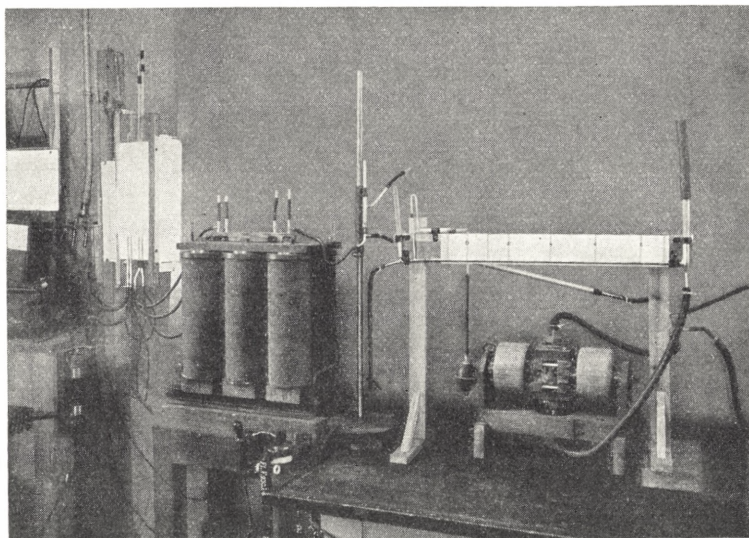


Fig. 2. Photograph of experimental Arrangement.

magnetic pump (to the right) and the electromagnet (to the left) is reproduced in fig. 2.

The various members of the experimental circuit will be described in detail below. Here some words may be said about the preparation of the ducts, especially the glass tubes. These were all comparatively narrow. Now, when a capillary has not undergone a special cleaning process the results of the flow experiments are quite indeterminate. This is due to a thin film of impurities adhering to the interior walls. In order to remove this film or layer the glass capillaries were for hours treated with concentrated



sulphuric acid with potassium bi-chromate  $K_2Cr_2O_7$ . Hereafter they were washed first with water and then with absolute alcohol. If the capillaries were then dried by drawing a flow of dry air through them the results of the experiments were reproducible with an exactitude of a few tenths of one p. c. and remained so for a very long time, provided the capillaries were not emptied and the mercury was thoroughly cleaned and dried before being introduced into the circuit.

With the rectangular ducts the use of the cleaning method here indicated was practically precluded. Instead the walls were treated with benzol in a way described below.

### Measurement of the Flow of Mercury.

The flow of mercury, i. e. the volume  $V$  passing each section of the circuit per. sec., was, as indicated, normally measured by means of a nozzle  $N$  through which the flow had to pass as indicated in fig. 1. The nozzle was arranged in a vertical tube and could readily be interchanged. Seven nozzles were used. In Tab. I the diameters of the bores are stated in the first column.

Tab. I.

Constants of Nozzles used in Measurement of Flow.

Orifice mm.	Constant $k$
2.950	1.960
2.043	1.000
1.551	0.584
1.106	0.294
0.832	0.179
0.651	0.111
0.412	0.0463

In the second column the constant  $k$  of the apparatus defined by

$$(1) \quad V = k\sqrt{h} \text{ cm.}^3/\text{sec.}$$

is entered. In the formula (1)  $h$  is the pressure drop in the nozzle measured in cm. *Hg.*

For the observation of  $h$  the manometer tubes shown in fig. 1 were employed. The tubes were mounted on a

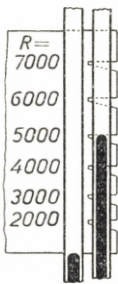


Fig. 3. Card-board with Notches for rapid Control of the Volume Flow.

transparent mm.-scale illuminated from behind. During each of the original experiments the magnetic field intensity  $H$  was kept constant and the flow adjusted for a series of values of the Reynolds' number  $R$ . With each experimental pipe or channel a set of experiments corresponding to a certain number of  $H$ -values was performed. In order to facilitate the work a card-board with notches as indicated in fig. 3 was produced corresponding to each experimental pipe or channel. When the pressure drop was equal to the distance

from the lower edge of the card-board to the horizontal edge of a certain notch the Reynolds' number of the flow in the experimental duct had the value written sideways to the notch in question. From this it will be gathered how the card-board was used for rapid setting of the flow during an experiment.

It was found expedient to supplement the above described flow meter by another apparatus intended particularly for rapid control of the nozzle gauge. The apparatus referred to is indicated by  $V_2$  in fig. 1. In fig. 4 it is shown on a larger scale. It consists mainly of a glass pipe  $BCD$  of known calibre. Just in front of the entrance to this pipe an air bubble may be pressed into the flow of mercury.

This bubble is carried with the flow and what is now measured is the time of its passage over the known distance between two marks *C* and *D* on the pipe. From this measurement and the calibre of the pipe the volume flow in the circuit is readily derived.

There are several precautions to be taken in the design and use of the flow gauge here considered. The bubble must fill out the whole section of the pipe. Therefore the

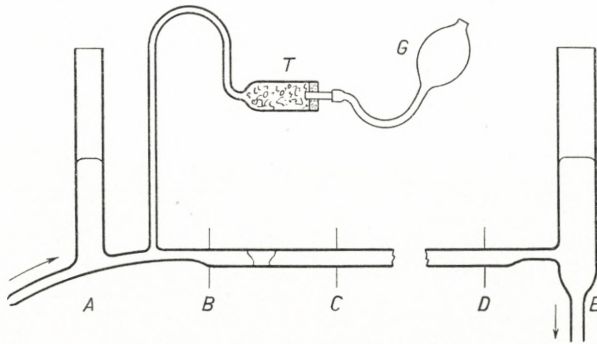


Fig. 4. Air Bubble Flow Meter.

aperture of the pipe must not be too large; in the apparatus in question it was 5.25 mm. At the moment at which the bubble is introduced slight fluctuations of the velocity will occur. So the bubble has to pass a certain length (9 cm.) of the pipe before entering the range *CD* which in the apparatus in question was 37 cm. It is absolutely essential that the walls of the pipe should be clean and dry; they were made so by the same method as was employed with the experimental capillaries. The air introduced into the pipe must also be dry and was therefore led through the tube *T* with a water absorbing material. In front of the inlet end of the apparatus a reservoir was arranged to stamp out the vibrations caused by the introduction of the bubble. At the exit end of the pipe another reservoir

was mounted, through which the bubble escaped after having passed the stretch *CD*.

The apparatus here described was, as stated, mainly used for quick control of the nozzle meter. In such a check the manometers of the latter gauge were always read during the passage of a bubble, because the bubble gave rise to a slight reduction of the velocity of the flow. The accuracy

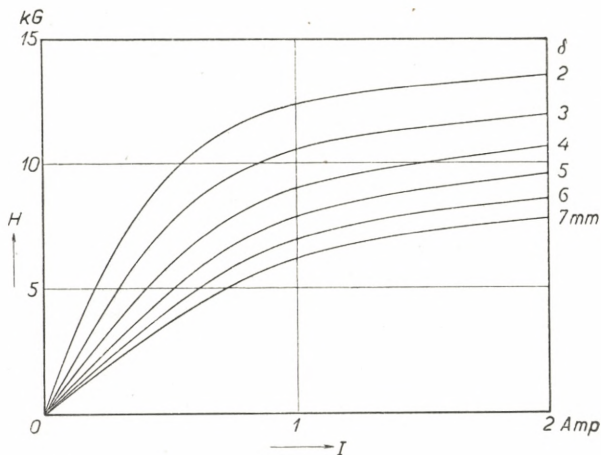


Fig. 5. Field-Intensity Curves at various Air Gaps.

of the measurement with the bubble apparatus was normally about 0.5 p. c.

### The Magnet.

The construction of the magnet will appear from fig. 2. The length of the field was 36 cm., the height of it 1.67 cm. Field intensity curves (with the exciting current as abscissa) were plotted for the values 2, 3, 4, 5, 6, 7 mm. of the width  $\delta$  of the air gap. These curves are reproduced in fig. 5. It should be noted that with the air gap 2 mm. and a field intensity  $H = 8000$  Gauss the intensity at points 3 cm. from the ends of the pole-pieces was but 1.5 p. c. smaller

than at the midpoint of the field. This lack of homogeneity was eventually reduced to 0.7 p. c. by an artificial straying applied to the centre of the field. Everything goes to show that this degree of homogeneity amply suffices for the purpose in question. It is a well-known fact that the field intensity is not a definite function of the exciting current unless certain precautions are taken. In the case considered a definite value of the field intensity was obtained

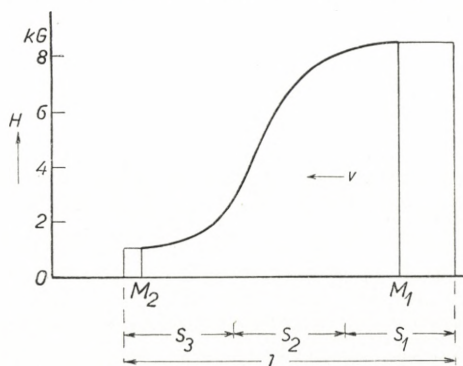


Fig. 6. Non-homogeneous magnetic Field produced by making one Set of Magnet Coils currentless.

by approaching the value of the magnetising current through a series of slow periodic current variations with an amplitude gradually going down to zero. This method was also used for demagnetising the magnet when the pressure drop was to be read corresponding to zero field intensity.

In connection with the magnet some observations on the effect of a non-homogeneous field on the pressure drop in the flow should be mentioned. Of the three sets of magnet coils (fig. 2) the set next to the exit end of the experimental duct was made currentless. In this way a field distribution curve like that indicated in fig. 6, where  $M_1$  and  $M_2$  are the positions of the manometer sockets, was obtained. The direction of the flow in the duct is indicated by the arrow. The flow thus first passes the stronger part of the field before entering the more feeble part.

With a volume velocity  $or$ , which is the same, a Reynold's number  $R$  well within the domain of laminar flow the observed pressure drop was not that corresponding to a homogeneous field of an intensity equal to the average value of the field — the average intensity was 4800 Gauss — but that corresponding to the higher field intensity 5200 Gauss. (Compare curves 9—18 below). On the other hand, if the set of coils at the inlet end was made currentless so that the flow took place in the direction from a lower field intensity to a higher one the observed pressure drop corresponded to a lower field-intensity — 5700 Gauss — than the average — 5900 Gauss. These observations find their explanation in the well-known fact that it takes time to change the velocity distribution — on which the pressure drop depends. The distribution corresponding to the field at the inlet end of the duct persists in some degree after the flow has left this field.

It was concluded from the experiments here indicated that the additional pressure drop to which the non-homogeneity illustrated in fig. 6 gave rise was otherwise practically imperceptible — in spite of the extremely pronounced character of the non-homogeneity. Thus homogeneity is required less in order to avoid this additional pressure drop than on account of the source of error illustrated by the experiments referred to above.

### The experimental Tubes and Channels.

Five cylindrical glass tubes, indicated in Tab. II below by the numbers 11—15, were used in the experiments here considered. They were selected from a large stock and were carefully calibrated. In no tube did the diameter vary more than about 0.5 p. c. within the part in which the pressure drop was measured. This part has a length of 28 cm. It was of course arranged entirely within the magnetic field. The whole length of the tube was 43 cm. the extensions on either side of the stretch of observation being 7.5 cm. At the inlet end of the tube 6 cm. of the 7.5 cm. were inside the magnetic field, at the exit end 2 cm. As stated elsewhere, the edge of the experimental tube or pipe was

kept sharp in order to secure transition from laminar to turbulent flow at a definite value of the velocity or of the Reynolds' number. The thickness of the wall of the tube was taken as small as possible, about 0.8 mm., with a view to reducing the air gap of the magnet, and so to be able to produce magnetic fields of the greatest possible intensity.

Connections to the manometer tubes were obtained

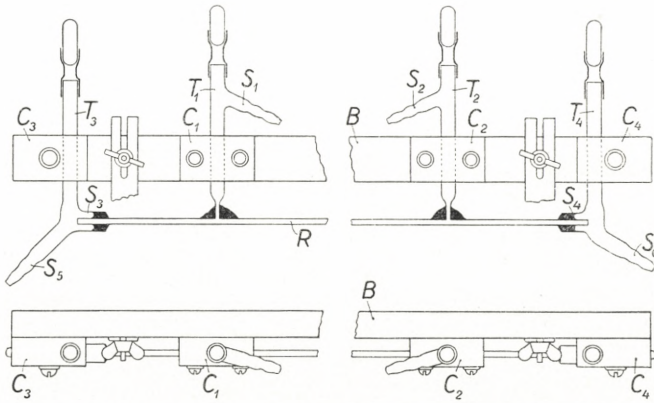


Fig. 7 a. Mounting of a cylindrical experimental Glass Pipe.

through bores at either end of the 28 cm. stretch of observation. The production of the bores by means of a swiftly rotating copper plug covered at the flat end with oil and carborundum required some practice. Slight annealing of the tube by drawing it through a luminous gas flame would seem a wise precaution before the process of boring in order to render breaking of the tube less likely.

The way in which the tube is mounted will appear from fig. 7 a. *B* is a light bar of wood to which four wooden blocks are fastened. The two extreme blocks  $C_3C_4$  carry two glass tubes or "shafts"  $T_3T_4$ , each with a rather wide socket,  $S_3S_4$ . Into these sockets the ends of the experimental

tube  $R$  are "cemented" by means of "picein". Above the bores in the experimental tube are placed two shafts  $T_1$  and  $T_2$ , carried by the blocks  $C_1C_2$  and cemented to the experimental tube likewise by picein. On  $T_1$  and  $T_2$  are the two sockets  $S_1S_2$  for the rubber tubes to the manometer. The flow of mercury is passed into and out of the experimental pipe by the sockets  $S_5S_6$ . The shafts  $T_1T_2T_3T_4$  serve as traps for dust particles and minute air

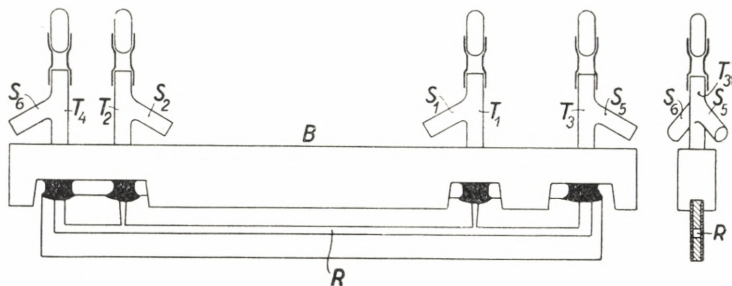


Fig. 7 b. Mounting of rectangular Channel.

bubbles. They are closed above by stoppers made from glass tubes.

We pass on to the rectangular ducts. Of these 18 were produced. They are indicated as  $K 21-K 38$  in Tab. II. In  $K 21$  the distance between the manometer sockets was 28 cm., in all the other ducts 14 cm. The total length of the ducts was in the latter cases 20 cm., the whole duct being mounted within the magnetic field. Fig. 7 b illustrates the mounting of a duct. The figure will be understood from the explanation given in connection with fig. 7 a. It may just be noted that the distance between  $T_1$  and  $T_3$  was 4 cm., that between  $T_2$  and  $T_4$  2 cm., finally that the flow was in the direction from  $T_1$  to  $T_2$ .

In fig. 8 a section of a duct is shown. The rectangular channel is formed between the plates  $BB$  and the pieces



Tab. II.  
List of experimental Ducts.

No.	$r$ cm.	$L$ cm.	$\nu_T \cdot 10^5$ at 20° C.	$\nu_L \cdot 10^5$ at 20° C.
11	0.0345	28.02	113	115
12	0.0583	27.93	121	121
13	0.0923	28.00	114	127
14	0.1147	28.02	123	—
15	0.1647	28.00	117	—

No.	$a$ cm.	$b$ cm.	$\frac{a}{b}$	$\nu_L \cdot 10^5$ at 20° C.
<i>K</i> 21	0.030	0.186	0.161	137
<i>K</i> 22	0.014	0.0625	0.224	99
<i>K</i> 23	0.0145	0.254	0.057	112
<i>K</i> 24	0.090	0.060	1.50	—
<i>K</i> 25	0.090	0.035	2.57	130
<i>K</i> 26	0.091	0.1075	0.846	167
<i>K</i> 27	0.091	0.081	1.12	170
<i>K</i> 28	0.091	0.0915	0.995	164
<i>K</i> 29	0.154	0.0415	3.71	145
<i>K</i> 30	0.155	0.060	2.58	159
<i>K</i> 31	0.155	0.1345	1.15	250
<i>K</i> 32	0.056	0.0935	0.60	130
<i>K</i> 33	0.057	0.058	0.99	130
<i>K</i> 34	0.155	0.030	5.16	178
<i>K</i> 35	0.155	0.0525	2.95	75
<i>K</i> 36	0.157	0.0338	4.65	95
<i>K</i> 37	0.157	0.042	3.75	135
<i>K</i> 38	0.157	0.055	2.88	158

AA. With the ducts *K* 34 and *K* 35, AA were made of copper, the surfaces forming the top and the floor of the channel being in these cases amalgamated. These ducts were made to agree exactly with the duct for which the theory was developed. With the other ducts AA were made of wood or "fiber". In the ducts *K* 21 to 28 *BB* were made

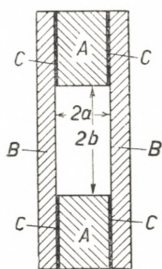


Fig. 8. Cross-section of experimental Channel.

of glass, in the others of celluloid. *C* is the cement between *A* and *B*; generally "asphalt lac" was employed. After the cementing together of the pieces *A* and *B*, lac would as a rule have entered the duct and had to be carefully removed by means of small pieces of cotton, wetted with benzol, which were pushed or drawn through the duct. The dimension  $2b$  was measured by means of a microscope, while  $2a$  was calculated from the total thickness of the duct and from the thicknesses of the *B*-sheets.

## II. The Results and their Discussion.

### Review of the Results.

For each tube or channel the variation with the magnetic field intensity of the pressure drop between the two sockets was determined for a number of values of the volume velocity  $V$  or of the REYNOLDS' number  $R$ . The latter is, in the case of a cylindrical pipe, defined by

$$(1) \quad R = \frac{v \cdot r}{\nu} = \frac{V}{\pi r \nu}$$

where  $v$  is the average velocity in cm./sec. over the section of the pipe,  $r$  the radius of the circular section in cm. and  $\nu = \frac{\eta}{\rho}$  the dynamical viscosity, i. e. the ratio of the viscosity and the density in c. g. s. units. In the case of a rectangular section,  $2a \cdot 2b$ ,  $r$  means the hydraulic radius defined as

$$(2) \quad r = \frac{2F}{O} = \frac{2ab}{a+b}$$

where  $F$  is the area and  $O$  the circumference of the cross-section.

The results of the experiments with each pipe or channel were represented in the shape of a series of curves having as abscissae the intensity  $H$  of the magnetic field, as ordinates the pressure drop  $h$  between the sockets measured in cm. Hg. Complete, representative sets of experiments are given in figs. 9—18. In each figure the dimensions of the pipe or channel are stated together with the length  $l$  ( $L$  in Tab. II) in which the pressure drop was measured.

It is well known that the turbulent flow in a pipe is changed into a laminar flow if the velocity or the REYNOLDS' number is reduced below a certain value. This value may be taken to be  $R = 1160$  (c. g. s.). A laminar flow may, on the other hand, be maintained even if the REYNOLDS' number is raised considerably above the critical value, provided the inlet end of the pipe is smooth and all disturbances of the flow are otherwise avoided. In the experiments here considered the edges of the pipe or channel were deliberately kept sharp in order to secure in every case a well defined transition from a laminar to a turbulent flow.

In the diagrams reproduced in figs. 9—18 this transition is clearly seen in all cases where a turbulent flow is observed. A curve is drawn through all the points of transition. Above and to the left of this curve, which is of a parabolic character, we have the domain of turbulent flow, below and to the right that of laminar flow. Within the first region the pressure drop and so the apparent viscosity decreases with increasing intensity of the magnetic field. This is due to the damping effect of the field on the vortices in the flow. Within the domain of laminar

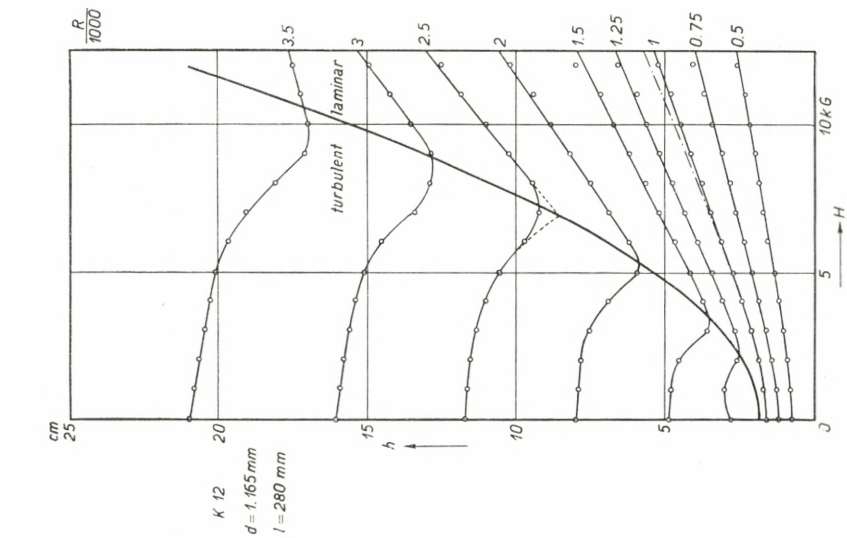


Fig. 10.

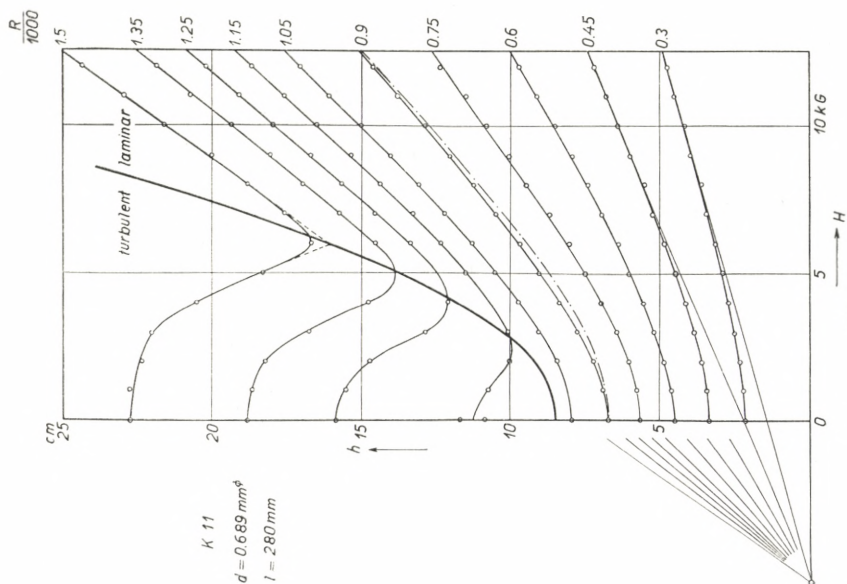


Fig. 9.

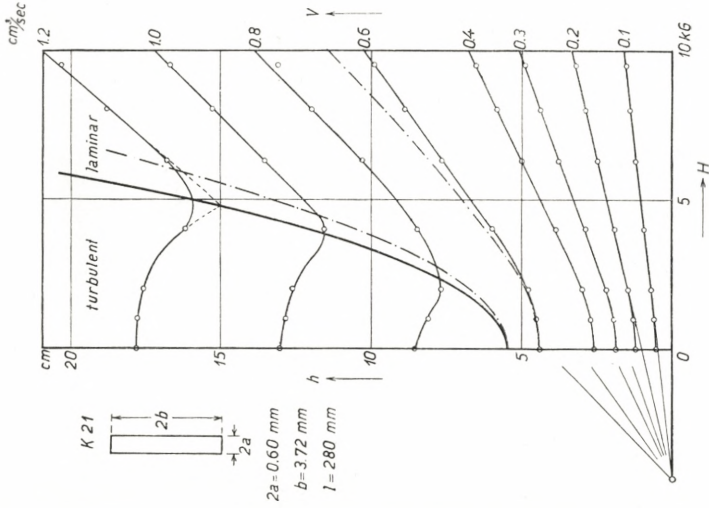


Fig. 11.

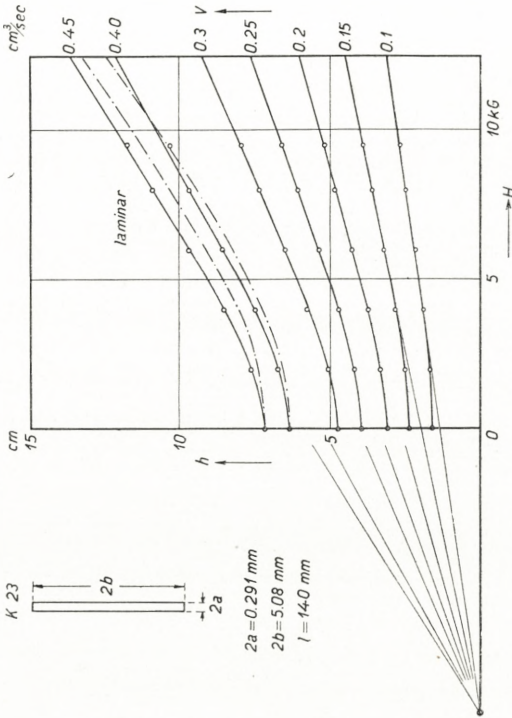


Fig. 12.

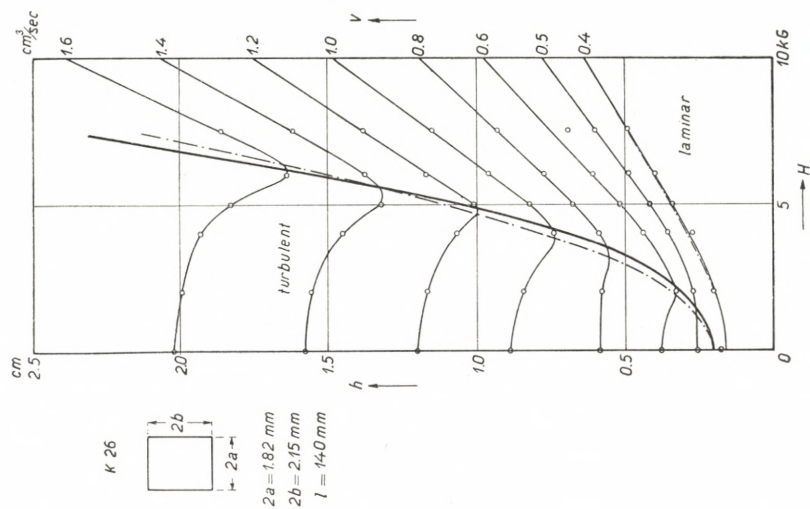


Fig. 13.

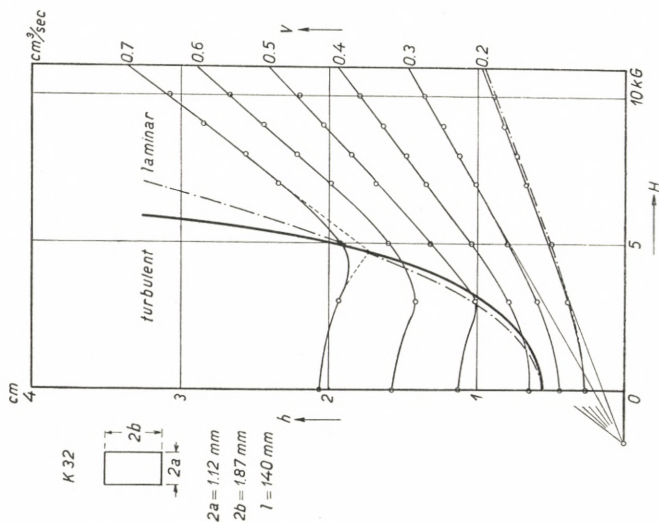


Fig. 14.

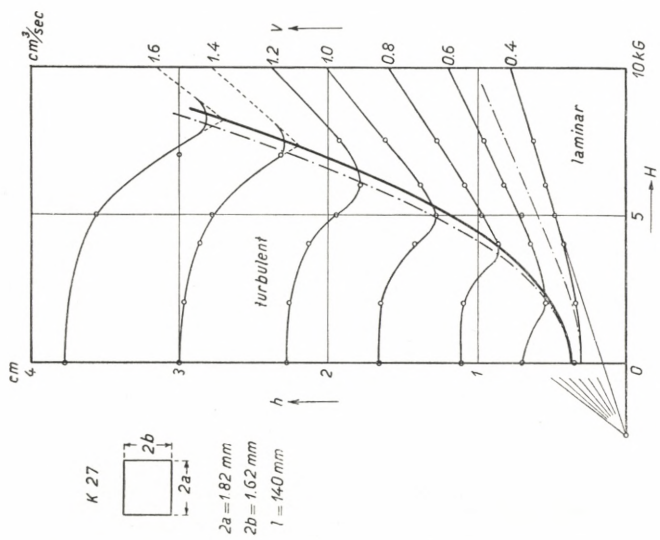


Fig. 16.

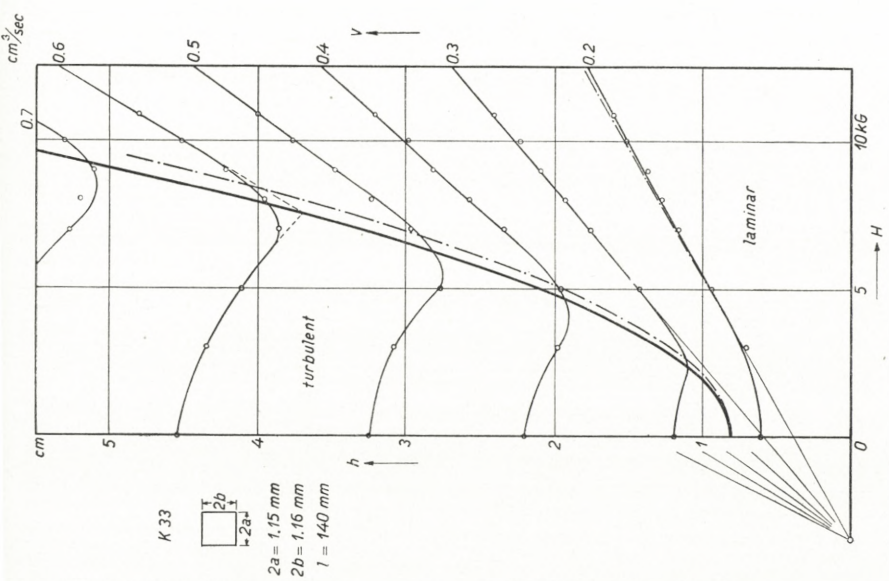


Fig. 15.

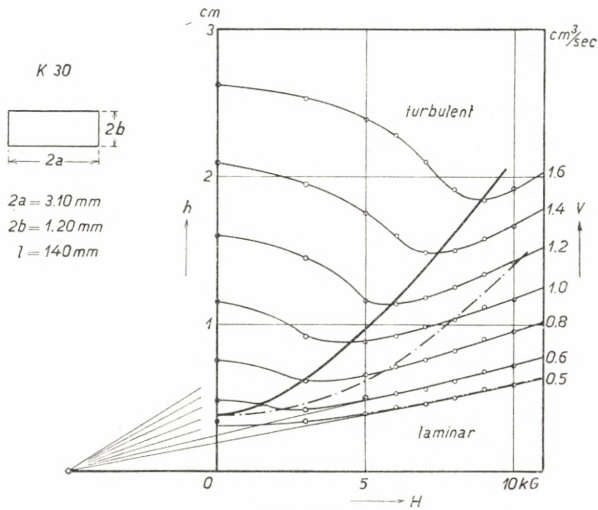


Fig. 17.

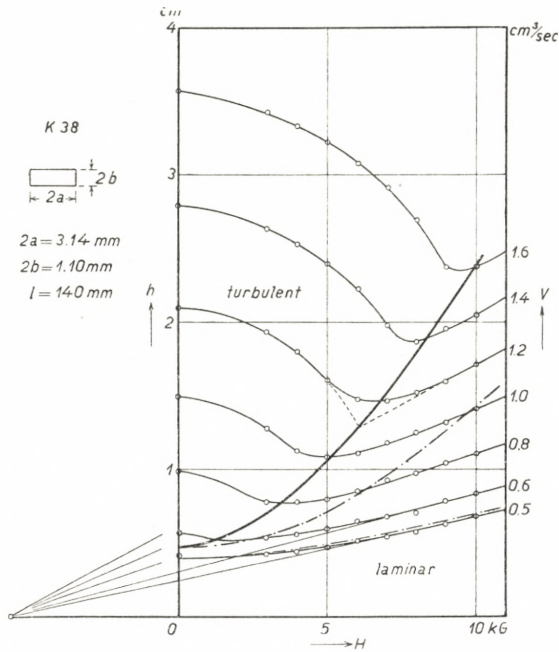


Fig. 18.

Fig. 9—18. Variation of Pressure Drop with Intensity of the magnetic Field.



flow the pressure drop increases rapidly with the field intensity. In cases where the flow is still laminar with no magnetic field on, the pressure drop may readily be raised to values twice the value corresponding to  $H = 0$  by putting on a field of quite moderate intensity. As a matter of fact the curves indicate that the pressure drop may be raised to any value by increasing the field since obviously the drop increases approximately as the field intensity when this is not too small. That is to say: the effect of the field on the laminar flow is to increase the apparent viscosity approximately proportionally to the field intensity. With rather small values of the field the apparent viscosity varies within the domain of laminar flow in a parabolic manner. The facts here stated may now be compared with the predictions of the theory referred to in the introduction to the paper. — It should, however, be borne in mind that the theory is based on certain simplifying assumptions and can only be expected to hold good in cases where these assumptions are fulfilled.

### Comparison with the Theory.

The main predictions of the theory may be thus stated.

In a narrow channel of rectangular section  $2a \cdot 2b$  cm.<sup>2</sup> ( $b \gg a$ ) placed in a homogeneous magnetic field of intensity  $H$  Gauss perpendicular to the side of the largest extension ( $2b$ ) there will, when the channel is passed by a laminar flow  $V \frac{\text{cm.}^3}{\text{sec.}}$  of an electrically conductive liquid, be a pressure drop  $p$  dyne/cm. determined by the formula:

$$(1) \quad p = \frac{3 VL}{4ba^3} \eta_e \frac{\text{dyne}}{\text{cm.}}$$

where

$$(2) \quad \eta'_e = \frac{\eta}{3} \cdot \frac{z_0^2 \tanh z_0}{z_0 - \tanh z_0} = \frac{\eta}{3} f(z_0)$$

where, again,

$$(3) \quad z_0^2 = 10^{-9} \frac{H^2 a^2}{\eta \varkappa}.$$

Here  $L$  is the length of that part of the channel in which the pressure drop is measured,  $\varkappa$  the specific resistivity of the liquid while  $\eta'_e$  is the apparent or virtual viscosity of the liquid in the magnetic field under the prevailing conditions. If  $z_0$  is small compared to 1 the apparent viscosity  $\eta'_e$  may be expressed by

$$(2a) \quad \eta'_e = \eta + \eta_e = \eta + \frac{1}{15} 10^{-9} \frac{H^2 a^2}{\varkappa}$$

showing that  $\eta'_e$  and so the pressure drop is that corresponding to zero field increased by an amount, the electromagnetic viscosity resp. the electro-magnetic pressure drop, which increases proportionally to the square of the field intensity. If  $z_0$  is large compared to 1 (strong fields) we derive the expression:

$$(2b) \quad \eta'_e = \frac{1}{3} \sqrt{10^{-9} H a} \cdot \sqrt{\frac{\eta}{\varkappa} + \frac{\eta}{3}}$$

from which it is seen that the apparent viscosity  $\eta'_e$  and so the pressure drop now increases linearly with  $H$ . The description of the conditions with a laminar flow in a homogeneous magnetic field thus given by the theory obviously fits in qualitatively with the observations. Quantitatively the agreement can only be expected to be tolerably close with flat channels, i. e. with small values of  $a$ . Fig. 11 corresponds to such a channel, and just in this case a comparatively very close agreement was found, as will be seen from the direct comparison between the ob-

served and the theoretical curve made with the curve  $V = 0.4 \text{ cm}^3/\text{sec}$ . It should be noted that the channel considered was not closed above and below by walls of highly conductive material as assumed in the theory. This, however, is obviously of small importance if only the channel is very high compared to its width, seeing that in this case the electric current lines remain practically rectilinear over most of the height of the section, while the conductive walls are replaced by the layers of mercury close to the top and the bottom of the channel.

Now the experiments not only cover cases in which the assumptions of the theory are fairly well fulfilled but also such in which the duct differs very much from a flat channel placed with its largest side perpendicular to the magnetic field. They even comprise investigations on the flow in cylindrical pipes. Obviously in such cases the theory as given by the equations (1)—(3) cannot be expected to hold good directly. It must be modified in some way or other and it is with this modification or adjustment we are concerned in the following. We may divide our problem into two. The pressure drop in cm. Hg, the quantity directly observed, may according to (1) and (2) be written

$$(4) \quad h = \frac{3VL\eta}{4ba^3\varrho g} \cdot \frac{f(z_0)}{3} \text{ cm. Hg.}$$

The coefficient to  $\frac{f(z_0)}{3}$  is simply the expression for the pressure drop  $h_0$  in a narrow channel when not placed in a magnetic field. If the channel is not narrow or if the duct is a cylindrical pipe the coefficient to  $\frac{f(z_0)}{3}$  in (4) should obviously be replaced by an appropriate expression corresponding to the duct in question. The function  $\frac{f(z_0)}{3}$  ex-

presses the change of the pressure drop to which the magnetic field gives rise (in the case of a laminar flow). In accordance herewith it is 1 for  $H = 0$ . It will be noted that the general character of the  $h$ - $H$ -curves is in all cases much the same. This suggests that the function  $\frac{f(z_0)}{3}$ , developed for the special case of a flat channel, may be made to cover other cases by applying a suitable reduction factor to the variable  $z_0$  or  $H$ .

The points of view here set forth are tested in the following paragraphs.

### The Flow at Zero Field-Intensity.

The expression to replace the coefficient to  $\frac{f(z_0)}{3}$  in equation (4) of the preceding paragraph is with a cylindrical pipe of radius  $r$

$$(1) \quad h_0 = \frac{8}{\pi} \cdot \frac{LV\nu}{r^4 g} = \frac{8L\nu^2}{r^3 g} \cdot R,$$

$\nu$  being the dynamical viscosity and  $R$  being the REYNOLDS' number defined by

$$(2) \quad R = \frac{v \cdot r}{\nu}.$$

With channels of rectangular section — sides  $2a$  and  $2b$  — a method for the calculation of the pressure drop at zero field-intensity is arrived at in the following way.

The Poiseuille Law for the laminar flow in a cylindrical pipe, i. e. (1), may be written in the form

$$(3) \quad \frac{2hr g}{Lv^2} \cdot \frac{v \cdot r}{\nu} = \psi \cdot R = 16,$$

where both  $\psi$  and  $R$  are dimensionless qualities. In case of a pipe of rectangular section  $2a \cdot 2b$  one may use the same form for the law, writing

$$(4) \quad \psi R = K$$

and replacing  $r$  by the "hydraulic radius" defined as the ratio of 2 times area of section and circumference.  $K$ , then, is a function of the ratio  $\frac{a}{b}$  and for this function LEA and TADROS have given a curve reproduced in fig. 19<sup>1</sup>. In case

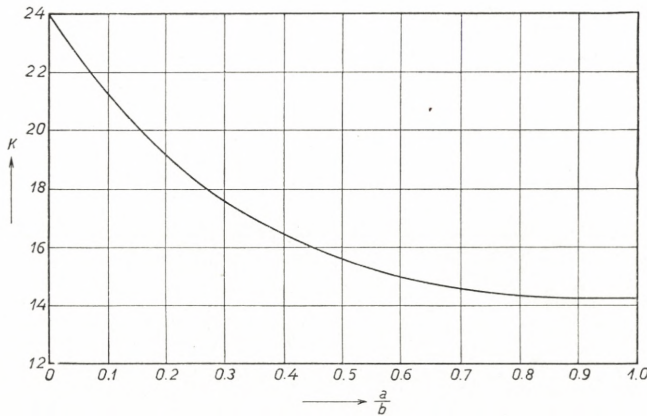


Fig. 19. Variation of  $K = \psi R$  with  $\frac{a}{b}$  for rectangular Pipes.  
(LEA and TADROS).

of a very flat channel i. e. with  $b = \infty$  the hydraulic radius is  $2a$  and the formula (4) together with the value  $K = 24$  taken from fig. 19 leads to the formula

$$(5) \quad h = \frac{3}{4} \frac{L\nu}{ba^3g} \cdot V \text{ (Flat channel, } a \ll b),$$

which may be directly derived. With a channel of quadratic section,  $a = b$ , the hydraulic radius is just equal to  $a$ . In this case the curve fig. 19 gives  $K = 14.22$  and (4) may be written

<sup>1</sup> F. C. LEA and A. G. TADROS. Phil. Mag. (7), 11, 1235, 1931.

$$(6) \quad h = 1.78 \cdot \frac{L\nu}{a^4 g} \cdot V \text{ (Quadratic channel, } a = b),$$

an expression which is a consequence of a more general formula derived by BOUSSINESQ and confirmed experimentally by SCHILLER<sup>1</sup>. In the general case, the formula for the pressure drop becomes:

$$(7) \quad h = \frac{K}{32} \cdot \frac{L\nu}{g} \cdot \frac{(a+b)^2}{a^3 b^3} \cdot V.$$

Now plotting the pressure drops, observed for  $H = 0$  within the domain of laminar flow, figs. 9—18, against  $R$  or  $V$ , straight lines are found from the slope of which the dynamical viscosity  $\nu$  may be calculated on the basis of expressions (1) and (7). In Tab. II the values thus determined are stated under  $\nu_L$ . With cylindrical pipes values are found of much the same size, viz.  $117 \cdot 10^{-5}$ , the value which is generally accepted for mercury at 20°C. With the rectangular ducts the results are rather fluctuating. This is thought to be due to difficulties in the production of the ducts and in the cleaning of them.

Values for  $\nu$  may also be calculated from the observations of the pressure drop in cylindrical pipes, at  $H = 0$ , within the domain of turbulent flow, seeing that an empirical formula has been derived by BLASIUS for this domain. The formula is

$$(8) \quad h = 0.06652 \nu^{7/4} \cdot r^{-5/4} \cdot \nu^{-1/4} \cdot g^{-1}.$$

By means of (8) the values of  $\nu$  entered in Tab. II under  $\nu_T$  were found. As will be seen, they agree well with the corresponding values from the laminar domain.

Fig. 20 illustrates the variation of the pressure drop  $h$

<sup>1</sup> Comp. Handbuch der Experimentalphysik IV. 4 Teil 1932, p. 146.

with the REYNOLDS' number  $R$ . The curves are derived from the diagram in fig. 9 corresponding to a cylindrical pipe. It will be seen that  $h$  is, within the laminar domain, ex-

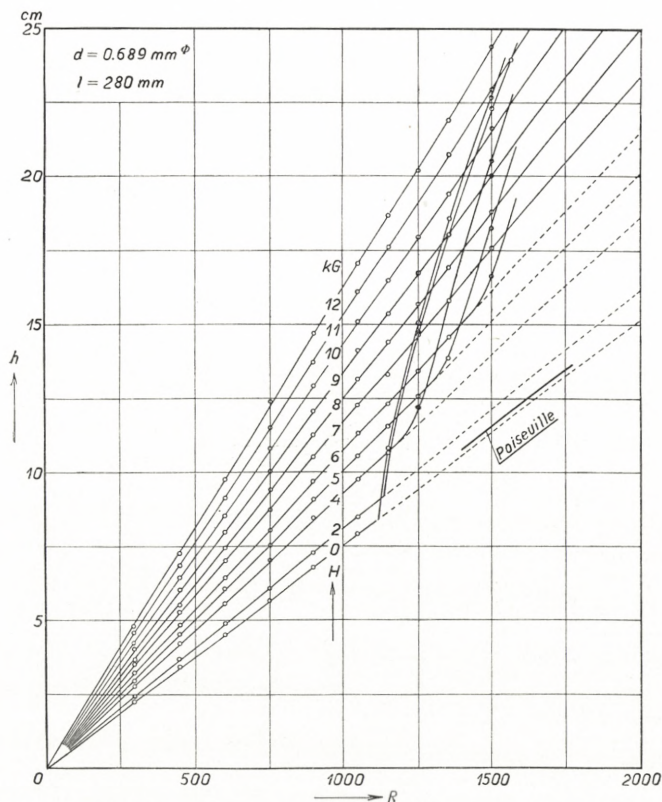


Fig. 20. Variation of Pressure Drop with REYNOLDS' Number  $R$ .

actly proportional to  $R$  as predicted by POISEUILLE'S expression (1) for  $H = 0$ .

It is, however, not only so for  $H = 0$  but for all values of  $H$ . In the diagram, fig. 20, the variation of  $h$  with  $R$  is also given for the turbulent flow. The transition, with  $H = 0$ , takes place at  $R = 1120$ , i. e. not too far from the generally accepted value  $R = 1160$ .

### Adjustment of the Function $f(z_0)$ to fit the Observations.

The theory of the pressure drop with a flat channel can now be given in the general form

$$(1) \quad h = h_0(V) \cdot \frac{f(z_0)}{3}$$

where  $h_0(V)$  is the pressure drop corresponding to zero field-intensity. The latter is proportional to the volume flow (or to the REYNOLDS' number).

We shall make an attempt at adjusting this theory for other ducts than a flat rectangular channel. In so doing we shall first try whether the pressure drop cannot be represented by the formula

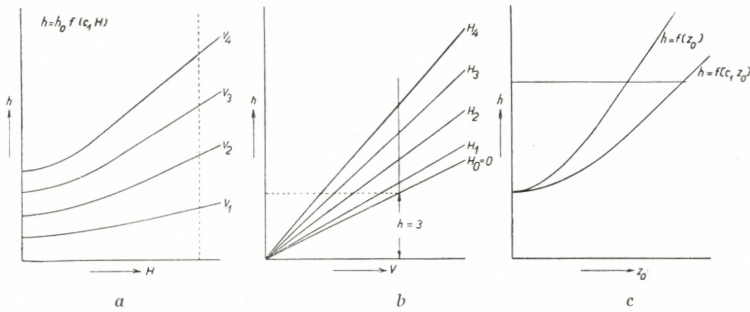
$$(2) \quad h = h_0(V) \cdot \frac{f(c_1 z_0)}{3}$$

where  $c_1$  is a number depending, with rectangular channels, on the value of  $\frac{a}{b}$  only.

The way in which this attempt was performed may be thus explained. We start with a set of observed curves of the type of figs. 9—18, fig. 21 *a*. From this set we may derive another set, fig. 21 *b*, of the type of fig. 20. By means of the latter diagram we may construct the  $h$ - $H$ -curve corresponding to  $h_0(V) = 3$ . The way it is done will be understood from the figure. The curve found in this way is  $h = f(c_1 z_0)$  with  $H$  as abscissa instead of  $z_0$ . Now  $z_0^2 = 10^{-9} \frac{H^2 a^2}{\eta z}$  from which with  $\eta = 0.0159$  c. g. s. and  $z = 10^{-4}$  Ohm·cm. (20° C.) we derive  $z_0 = 0.0250 aH$ . By means of the latter formula we transform the  $h$ - $H$ -curve into the  $h$ - $z_0$ -curve and thus have the curve which we try to express by  $h = f(c_1 z_0)$ . It is shown in fig. 21 *c*. In the same



figure the theoretical curve  $h = f(z_0)$  is drawn. For a number of values of  $h$  the ratio of the corresponding ab-



Figs. 21 a—c. Diagrams to explain the Adjustment of the Theory to fit the Observations.

scissae of  $h = f(z_0)$  and  $h = f(c_1 z_0)$  is calculated. This ratio  $c_1$  is found to be tolerably constant independent of

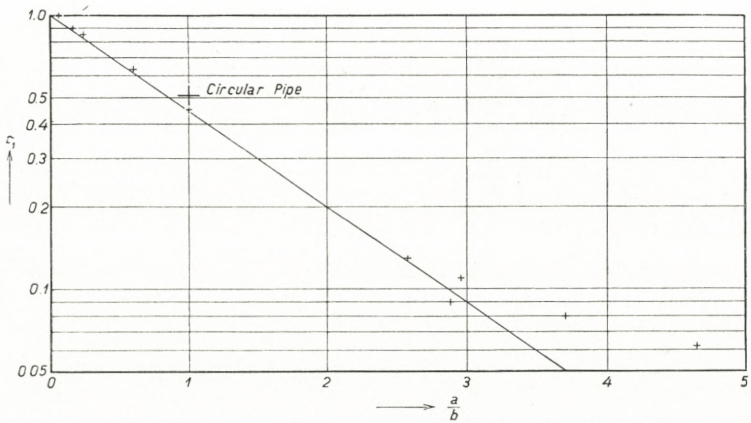


Fig. 22. Variation of the Reduction Factor  $c_1$  with  $\frac{a}{b}$ .

$h$  or  $z_0$ . So by multiplying the abscissae of the "observed"  $h-z_0$ -curve by a constant factor the curve is reduced to that corresponding to a flat channel (of which  $c_1 = 1$ ). Or we may, in general, express the pressure drop in a

rectangular channel placed in a homogeneous magnetic field by the formula (2).

The value of  $c_1$  of course depends on the ratio  $\frac{a}{b}$  of the two sides of the channel cross-section ( $a$  parallel to the magnetic force). In fig. 22 the variation of  $c_1$  with  $\frac{a}{b}$  is shown. Up to  $\frac{a}{b} = 3$  the variation may, in the semi-logarithmic system of coordinates, be represented by a straight line corresponding to the dependency

$$(3) \quad c_1 = 10^{-0.35 \frac{a}{b}}.$$

So, finally, the formula for the pressure drop in a rectangular channel may, with a laminar flow, be written

$$(4) \quad h = h_0 \cdot \frac{f\left(z_0 \cdot 10^{-0.25 \frac{a}{b}}\right)}{3} \text{ cm. Hg, } \frac{a}{b} < 3,$$

where the  $f$ -function is defined by

$$(5) \quad f(x) = x^2 \frac{\frac{\tanh x}{x}}{1 - \frac{\tanh x}{x}}.$$

With cylindrical pipes also the pressure drop with a laminar flow may be expressed by formula (2). The value which must here be ascribed to  $c_1$  was found to be 0.51 independently of the radius  $r$  of the pipe, at any rate between  $r = 0.35$  mm. and 0.92 mm. Thus with cylindrical pipes the pressure drop with a laminar flow is determined by

$$(6) \quad h = h_0 \cdot \frac{f(0.51 z_0)}{3} \text{ cm. Hg,}$$

where it should be noted that  $z_0$  is calculated from the field-intensity  $H$  by the expression

$$(7) \quad z_0 = 0.0250 rH$$

$r$  being the radius of the cross-section of the pipe.

In order to illustrate the agreement between the  $h$ - $H$ -curves calculated by means of expressions (4) and (6) and the corresponding observed curves a calculated curve, the dot and dash curve, is plotted in each of the diagrams figs. 9—18<sup>1</sup>. It will be seen that in most cases the agreement is fairly good. It should be noted that the calculated curve is drawn with the same  $h_0$  as the observed curve. No account has thus been taken of a possible error in the value of  $h_0$ .

In connection with the discussion here given attention may be called to a particular feature of the experimental curves of figs. 9—18. If, as indicated in the diagrams, tangents are drawn to the several curves of each set corresponding to the same abscissa (in the diagrams  $H = 8000$  Gauss) it is found that all the tangents intersect in the same point of the axis of abscissae or nearly so. This feature is a direct consequence of the theory, whether in the original or in the modified form. For the theory may be written

$$(8) \quad h = CV f(c_1 z_0), \quad z_0 = 0.0250 aH.$$

From this expression it follows that

$$(9) \quad \frac{dh}{dH} = \frac{dh}{dz_0} \cdot \frac{dz_0}{dH} = CV \cdot c_1 \cdot 0.0250 a f'(c_1 z_0) = C_1 V.$$

The equation of the tangent to one of the  $H$ - $h$ -curves at the point  $(H_1, h_1)$  is

$$(10) \quad \frac{h-h_1}{H-H_1} = C_1 V.$$

<sup>1</sup> In fig. 11 the test is made with  $V = 0.45$  cm.<sup>3</sup>/sec.

Hence the abscissa of the point of intersection with the axis of abscissae is determined by

$$(11) \quad H' = H_1 - \frac{h_1}{C_1 V} = H_1 - \frac{1}{0.0250 ac_1} \frac{f(c_1 z_0)}{f'(c_1 z_0)}$$

i. e.  $H'$  is independent of  $V$ . From the formula  $f(z_0) = \frac{z_0^2 \tanh z_0}{z_0 - \tanh z_0}$  the following expression for  $f'(z_0)$  may readily be derived:

$$(12) \quad f'(z_0) = \frac{z_0^3 \operatorname{sech}^2 z_0 + z_0^2 \tanh z_0 - 2 z_0 \tanh^2 z_0}{(z_0 - \tanh z_0)^2}.$$

By means of (11), (12) and the expression for  $f(z_0)$  the experimental values found for  $H'$  could be checked. A test of this description would, however, seem superfluous after the discussion given in the first part of the present paragraph.

### The Boundary Curve between the Domains of laminar and turbulent Flow.

In the experiments considered measures were taken to secure transition from a laminar to a turbulent flow at a definite value of the REYNOLDS' number  $R$ . The critical value may be denoted by  $R_c$ . It is determined by

$$(1) \quad R_c = \frac{v_c \cdot r_h}{\nu}.$$

Here  $r_h$  is the hydraulic radius which with a channel of rectangular section  $2a \cdot 2b$  is equal to  $\frac{2ab}{a+b}$ . Again  $v_c$ , the critical velocity, is equal to the volume-velocity  $V_c$  divided by  $4ab$ . Introducing in (1) this formula may be written

$$(2) \quad R_c = \frac{V_c}{2(a+b)\nu} = \frac{V_c \cdot \varrho}{2(a+b)\eta}.$$

Now, if the flow takes place in a homogeneous magnetic field the transition is displaced towards larger values of the volume flow i. e.  $V_c$  increases. The explanation is most likely to be found in the apparent increase of the viscosity  $\eta$ . On this assumption the variation of the critical volume-velocity  $V_c$  with  $H$  should be

$$(3) \quad V_c = \frac{2(a+b)}{\rho} R_c \cdot \eta'_e = \frac{2}{3} \frac{a+b}{\rho} R_c \eta f(z_0).$$

The corresponding value of the pressure drop  $h_c$  in the channel is determined by

$$(4) \quad h_c = \frac{3}{4} \cdot \frac{V_c L}{ba^3} \eta'_e = \frac{1}{6} R_c \frac{a+b}{ba^3} \cdot \frac{L}{G} \cdot \nu^2 f^2(z_0).$$

Instead of the critical REYNOLDS' number we may here introduce the critical volume-velocity  $V_{c.0}$  corresponding to  $H = 0$ . The latter is determined by  $R_c = \frac{V_{c.0}}{2(a+b)\nu}$  giving

$$(5) \quad h_c = \frac{1}{12} \cdot \frac{\eta}{\rho g} \cdot \frac{V_{c.0} L}{ba^3} \cdot f^2(z_0)$$

where it will be remembered that  $f(z_0) = \frac{z_0^2 \tanh z_0}{z_0 - \tanh z_0}$  and  $z_0 = Ha \sqrt{10^{-9} \kappa^{-1} \eta^{-1}}$ . Equation (5) should represent the boundary curve between the domains of the laminar and the turbulent flow in the case of a flat channel  $a \ll b$ . In figs. 9–18 the actual boundary curves are drawn in all cases where the observations include the turbulent domain. The boundary curve is of a parabolic character. It is not, of course, to be expected that it can be represented, with all ducts, by the formula (5). In the first instance the product with which  $f^2(z_0)$  is multiplied will generally not coincide exactly with the ordinate to the boundary curve

for  $H = 0$  — owing to shortcomings in the measurement of the dimensions of the ducts. We will therefore direct our attention to the shape of the boundary curve only and we shall make the natural assumption that  $f(z_0)$  in (5) must be replaced by  $f(c_1 z_0)$  as found in the previous paragraph. That is to say we will assume that, with rectangular ducts, the boundary curve may be represented by

$$(6) \quad h = h_{0.B} \cdot \frac{1}{9} \left[ f \left( z_0 \cdot 10^{-0.35 \frac{a}{b}} \right) \right]^2 \text{ cm. Hg}$$

and with cylindrical pipes by

$$(7) \quad h = h_{0.B} \cdot \frac{1}{9} \left[ f(0.51 z_0) \right]^2 \text{ cm. Hg.}$$

These assumptions are put to the test in figs. 9—18 (except fig. 11) where the full drawn curves are the observed curves while the dot-and-dash curves are calculated from (6) or (7). With the cylindrical pipes the agreement between the two curves is perfect so that only the observed curve is drawn. With the rectangular ducts it is still fairly good as long as  $a < b$ . With  $a > b$  larger discrepancies occur, which was of course to be expected.

## Appendix I.

### Check on the Reynolds' Law of Similarity.

The material of observations of the present research may be utilised for a control of the REYNOLDS' Law of Similarity in the case of a flow of mercury through cylindrical or rectangular pipes. For from each of the diagrams of which samples are given in figs. 9—18 the variation — at zero field-intensity — of the pressure drop with the

flow, characterised either by the volume-velocity  $V$  or by the REYNOLDS' number  $R$ , may be derived. We shall here confine ourselves to the observations from cylindrical pipes. Ten years ago the flow of mercury in such pipes was compared to that of water and it was found that the REYNOLDS' Law of Similarity holds good for mercury also<sup>1</sup>. The law was tested in the shape of a curve, having as abscissa the REYNOLDS' number defined by  $R_1 = \frac{vd}{\nu}$ ,  $d$  being the internal diameter of the pipe. The ordinate was the quantity  $\psi_1 = \frac{h}{L \left( \frac{4v^2}{gd} \right)}$  where  $h$  is the pressure drop in

the length  $L$  of the pipe. Now it has become customary to define  $R$  by  $R = \frac{vr}{\nu}$ , where  $r$  is the radius of the pipe, and to plot the quantity  $\psi = \frac{h}{L \frac{v^2}{gd}} = 4\psi_1$  against  $R$ . This

we shall do in the following. The earlier test with mercury was performed mainly with wider pipes because it was then found difficult to obtain reproduceable results with narrower pipes. So in all essentials, the test was confined to the part of the  $\psi_1 - R_1$  curve corresponding to values of  $R_1$  above 10000. With the new experiments the flow in rather narrow pipes could be studied without any difficulty owing to the introduction of an effective method of cleaning the pipes. The results of these experiments, therefore, supplement the older ones in a very happy way, rendering possible the checking of the law down to very low values of  $R$ .

The following pipes were used in the test:

$$K 11, \quad d = 0.0689 \text{ cm.}$$

$$K 12, \quad d = 0.1165 \text{ cm.}$$

<sup>1</sup> See note p. 3 of Introduction.

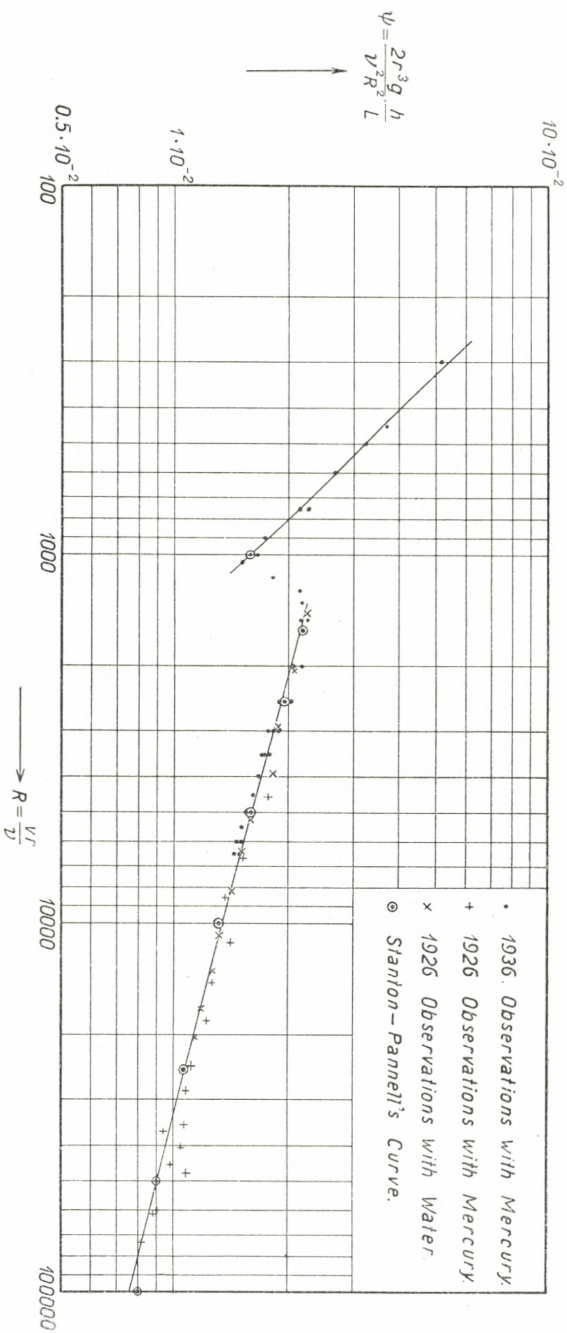


Fig. 23. Check on the Reynolds' Law of Similarity.



$K 13, d = 0.1845 \text{ cm.}$

$K 15, d = 0.3293 \text{ cm.}$

In all cases the pressure drop was measured over a length of 28 cm. of the pipe. The value of  $\psi$  was calculated from the formula:

$$(1) \quad \psi = \frac{2r^3 g \cdot h}{L\nu^2 \cdot R^2}$$

a form obtained from the formula given above by introducing  $R$  instead of  $\nu$ . In Tab. III all the values for  $\psi$  are

Tab. III.

$R \cdot 10^{-3}$	$\log_{10} R$	$K 11$ $\psi \cdot 10^3$	$K 12$ $\psi \cdot 10^3$	$K 13$ $\psi \cdot 10^3$	$K 15$ $\psi \cdot 10^3$
0.30	2.477	51.8	—	—	—
0.45	2.653	35.4	—	—	—
0.50	2.699	—	32.40	—	—
0.60	2.778	26.3	—	—	—
0.75	2.875	21.2	22.48	—	—
0.90	2.954	17.48	—	—	—
1.00	3.000	—	16.68	—	—
1.05	3.021	15.08	—	—	—
1.15	3.061	18.20	—	—	—
1.25	3.097	21.40	19.40	—	—
1.35	3.130	21.75	—	—	—
1.50	3.176	21.30	21.76	21.36	22.24
2.00	3.301	—	20.20	20.12	21.68
2.50	3.398	—	18.92	19.00	20.04
3.00	3.477	—	17.92	18.08	18.76
3.50	3.544	—	17.20	17.40	17.88
4.00	3.602	—	—	16.72	16.76
4.50	3.653	—	—	16.24	16.32
5.00	3.699	—	—	15.68	15.92
5.50	3.740	—	—	15.20	15.40
6.00	3.778	—	—	14.84	15.16
6.50	3.813	—	—	14.56	14.88

stated and in the diagram, fig. 23, they are plotted against  $R$ . In the same diagram values of  $\psi$  calculated from the

earlier test are entered, and so the  $\psi - R$  curve is extended up to values of  $R$  of about 70000, covering nearly the same interval as the well-known investigations by STANTON and PANNELL<sup>1</sup>. Points from the curve obtained in this latter investigation are shown in the diagram. Quite obviously the same law holds good for mercury as for the fluids examined by STANTON and PANNELL. It may be noted that the curve which may be drawn on the basis of the mercury experiments exhibits the same faint upward bend as the STANTON-PANNELL curve. If this curvature is neglected and a straight line drawn evenly among the points, the slope of this line is found to be almost exactly 4, in agreement with the formula given by BLASIUS. Again, it is found that the straight line representing the observations within the laminar domain corresponds, as it should, to the equation  $\psi R = 16$ .

## Appendix II.

### The Influence of the magnetic Field on the turbulent Flow.

From the diagrams figs. 9—18 it is seen that within the domain of turbulent flow the pressure drop decreases when the intensity of the magnetic field increases. This of course is due to a damping of the turbulence, but what is observed is not the sole effect of this damping. Together with the smoothing out of the vortices which manifests itself in a smaller pressure drop there is undoubtedly also the other effect of the field known from a laminar flow, thus an effect which tends to increase the pressure drop. The actual pressure drop is the resultant of these two

<sup>1</sup> Phil. Trans. Royal Soc. A. 214.

effects counteracting each other. In the following the two effects are termed the damping effect and the viscosity effect respectively.

We may make an attempt to isolate the effect of the field on the turbulence i. e. the damping effect. The reasoning on which this attempt is based may be stated as follows. In fig. 24  $h_T b_2 c$  is the observed curve for the variation of the pressure drop in a given length of the tube with a

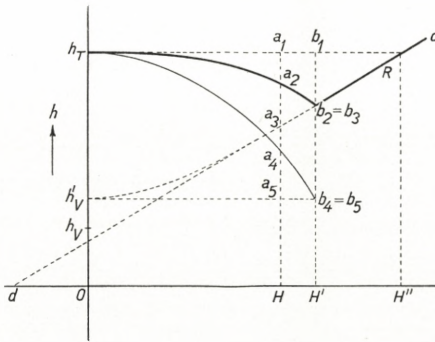


Fig 24. Diagram illustrating attempt at separating the viscosity effect and the damping effect within the domain of turbulent flow.

given flow or a given REYNOLDS' number. Now it would seem very likely that the damping effect i. e. the reduction of the pressure drop due to the damping of the vortices is proportional to the square of the field-intensity. That is to say, it may be anticipated that the curve for the damping effect is a simple parabola. The question then arises: How must the curve for the viscosity effect be in order to make the curve for the damping effect a parabola. Let us assume  $h'_V a_3 b_2$  to be the curve for the viscosity effect within the domain of turbulence. Then a point of the curve for the damping effect would be obtained by lowering a corresponding point  $a_2$  of the observed curve

by the amount  $a_3a_5$ . In this way the point  $a_4$  is arrived at and the whole curve for the damping effect would be  $h_T a_4 b_4$ . The problem is to choose  $h'_V a_3 b_3$  in such a way that  $h_T a_4 b_4$  becomes a parabola. In drawing  $h'_V a_3 b_3$  we know one thing for certain, namely that the curve is to pass the point  $b_3$ . But in addition we may reasonably make the following assumptions: 1) that the curve in the point  $b_3$  has its tangent in common with the known curve branch  $b_3c$ ; 2) that the curve has a smooth more or less parabolic shape approximately as indicated; 3) that close to  $H = 0$  the tangent is horizontal; and 4) that the curve is higher than the curve corresponding to a laminar flow, i. e. that its ordinate at  $H = 0$  is higher than the value of  $h$  determined by

$$(1) \quad h_V = \frac{8L\nu^2}{r^3g} \cdot R.$$

The last assumption requires some explanation. With laminar flow and no field the distribution of velocity across the pipe is parabolic. With turbulent flow, i. e. with the flow which actually obtains in the pipe, the distribution is uniform across most of the diameter, dropping rather abruptly to zero within a zone close to the wall. The latter type of velocity distribution is just that produced by a strong magnetic field acting on a laminar flow and manifesting itself in an increased pressure drop.

This is the reason why we conclude that the curve representing the viscosity effect must at  $H = 0$  be drawn through a point  $h'_V$  higher than that corresponding to  $h_V$  calculated from (1).

The experiments with cylindrical pipes were now dealt with in the way indicated. The most reliable of these experiments with regard to the pressure drop within the

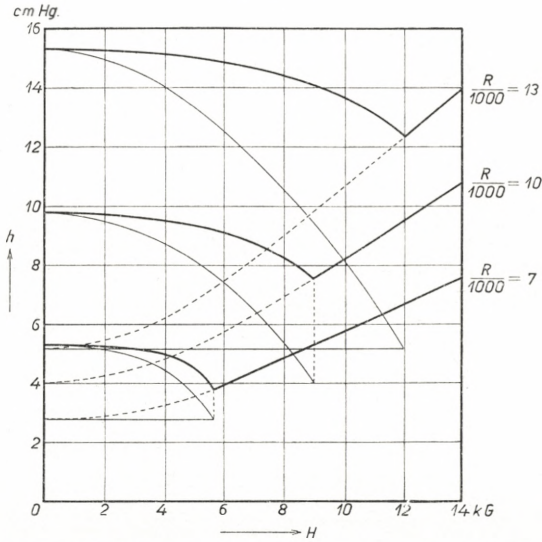


Fig. 25. Construction of Curves for the Damping Effect with Experiments performed with Pipe K 13.

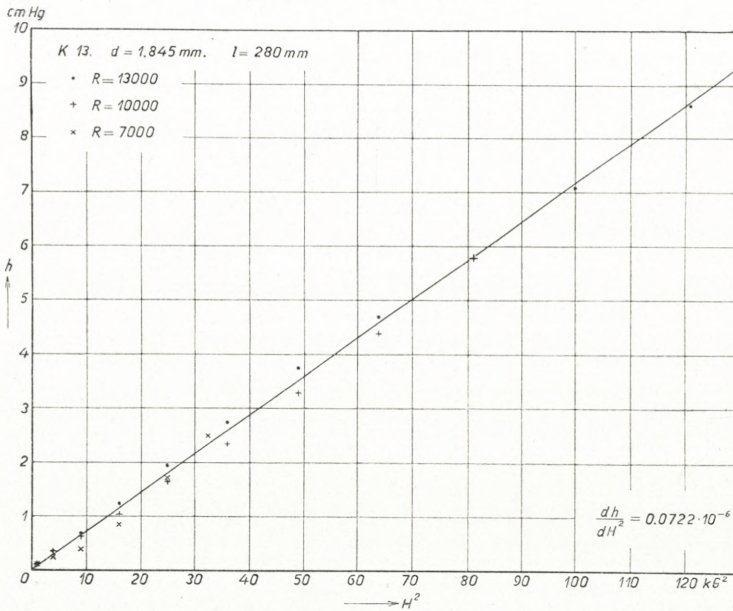


Fig. 26. Test on parabolic Character of the Damping Effect Curve in the Case of Pipe K 13.

turbulent domain were those performed with the widest pipes: *K* 13, 14, and 15, while the pressure drop with the narrower pipes *K* 11 and *K* 12 was rather uncertain within the said domain due to instability of the flow in the boundary region between the turbulent and the laminar domain. With the pipes *K* 13, 14, and 15, however, rather characteristic results were arrived at. If the curve  $h'_v a_3 b_3$  was drawn in such a way that the ordinate at  $H = 0$ ,  $h'_v$ , was twice the height  $h_v$  corresponding to a laminar flow ( $h_v = \frac{8L\nu^2}{r^3g} R$ ) then the curve for the damping effect became a parabola and this parabola was found to be independent of the intensity of the flow. In fig. 25 the construction of curves for the damping effect with the pipe *K* 13 is shown and in fig. 26 the test of the parabolic character and of the independency of the intensity of the flow, i. e. of  $R$ , is illustrated. From the latter diagrams the values of  $\frac{dh}{dH^2}$  entered in the following table were found.

Pipe	$d$ cm.	$\frac{dh}{dH^2} \cdot 10^6$	$d \cdot \frac{dh}{dH^2} \cdot 10^6$
<i>K</i> 12	0.1165	0.0960	0.0112
<i>K</i> 13	0.1845	0.0722	0.0133
<i>K</i> 14	0.2294	0.0571	0.0131
<i>K</i> 15	0.3298	0.0422	0.0139

From the results with *K* 13, 14, and 15 we tentatively draw the conclusion that the damping effect may be expressed by the simple formula

$$h = 0.0134 \cdot 10^{-6} \frac{H^2}{d} \text{ (Gauss, cm., cm. } Hg)$$

independently of the intensity of the flow (volume-velocity). It should be borne in mind that  $h$  is the reduction of the pressure drop due to the damping effect of the field on the

turbulence. It is thus the difference between the ordinates of the curve for the damping effect at  $H = 0$  and at the field-intensity in question. In the table the values found with  $K 12$  are also given. They are, however, less reliable.

One may seek a confirmation of the simple relation for the damping effect by a dimensional consideration. If it is justifiable to assume that the change in the pressure drop pr. cm. due to this effect can depend only on 1) the field intensity  $H$ , 2) the diameter  $d$  of the pipe, and 3) the velocity  $v$  of the flow then we may write down the equation

$$\frac{h \rho g}{L} = H^r d^s v^q.$$

Introducing the dimensions for the various qualities we find for the determination of  $r$ ,  $s$  and  $q$ :  $-\frac{r}{2} + s + q = -2$ ,  $\frac{r}{2} = 1$ ,  $-r - q = -2$  from which  $s = -1$ ,  $r = 2$ ,  $q = 0$  and so

$$\frac{h \rho g}{L} = c_1 \cdot \frac{H^2}{d} \text{ (independently of } v, V \text{ or } R)$$

or (with a constant  $L$  and with a given liquid)

$$h = c_2 \cdot \frac{H^2}{d}$$

as found above.

It is quite obvious that the attempt at separating the viscosity effect and the damping effect just explained is to be considered only as a provisional step in the investigation of this domain. So too much weight should not be attached to the results which may on a closer examination prove more or less false. It is contemplated to make the influence of the magnetic field on a turbulent flow the subject of a subsequent investigation.

Provisional Laboratory of Technical Physics. Royal Technical College.  
Copenhagen. May 1937.





Det Kgl. Danske Videnskabernes Selskab.  
Mathematisk-fysiske Meddelelser. **XV**, 8.

---

EINE DOPPELT-FASTPERIODISCHE  
GANZE TRANSZENDENTE  
FUNKTION

VON

RICHARD PETERSEN



KØBENHAVN  
LEVIN & MUNKSGAARD  
EJNAR MUNKSGAARD  
1938

Printed in Denmark.  
Bianco Lunos Bogtrykkeri A/S.

In natürlicher Fortsetzung früherer Untersuchungen werde ich jetzt zeigen, wie es möglich ist, durch Anwendung des Runge'schen Verfahrens für Polverschiebung eine doppelt-fastperiodische ganze Transzendente zu konstruieren.

Es sei unsere Ausgangsfunktion eine doppelt-periodische Funktion mit Polen dritter Ordnung in den Punkten eines Quadratnetzes, nämlich in  $2p+1+i(2q+1)$ , wo  $(p, q)$  alle Zahlenpaare mit ganzzahligen Koordinaten durchläuft.

Eine solche Funktion ist

$$f_0(s) = \sum_{(p,q)} \frac{1}{(s-(2p+1)-i(2q+1))^3}, \quad (1)$$

wo  $s = \sigma + it$ .

Es ist aber zweckmässig, diese Reihe folgendermassen umzuschreiben:

$$f_0(s) = \sum \left\{ \frac{1}{(s-(2p-1)-i(2q-1))^3} + \frac{1}{(s-(2p+1)-i(2q-1))^3} + \frac{1}{(s-(2p+1)-i(2q+1))^3} + \frac{1}{(s-(2p-1)-i(2q+1))^3} \right\};$$

das allgemeine Glied der Reihe enthält die 4 Glieder der ursprünglichen Darstellung, welche den Polen, die Eckpunkte eines Quadrates mit dem Mittelpunkt  $2p+i2q$  sind, entsprechen, und die Summation soll über alle Zahlenpaare  $(p, q)$  erstreckt werden, deren Koordinaten ganze ungerade Zahlen sind.

Ersetzen wir nun  $p$  durch  $2p+1$  und  $q$  durch  $2q+1$ , so entsteht die Reihe

$$f_0(s) = \sum \left\{ \frac{1}{(s-(4p+1)-i(4q+1))^3} + \frac{1}{(s-(4p+3)-i(4q+1))^3} \right. \\ \left. + \frac{1}{(s-(4p+3)-i(4q+3))^3} + \frac{1}{(s-(4p+1)-i(4q+3))^3} \right\},$$

wo  $(p, q)$  alle Zahlenpaare mit ganzzahligen Koordinaten durchläuft; das allgemeine Glied entspricht den Polen, welche Eckpunkte eines Quadrates mit dem Mittelpunkte  $4p+2+i(4q+2)$  sind.

Setzt man

$$\varphi_0(s-2-i2) = \left. \frac{1}{(s-1-i)^3} + \frac{1}{(s-3-i)^3} + \frac{1}{(s-3-i3)^3} + \frac{1}{(s-1-i3)^3} \right\} (2)$$

so lässt die gegebene Funktion sich in der Form

$$f_0(s) = \sum_{(p,q)} \varphi_0(s-(4p+2)-i(4q+2)) \quad (1)$$

schreiben. Die Bezeichnungen haben wir hier so gewählt, dass man für gegebene  $p$  und  $q$  unmittelbar sieht, zu welchem Quadrate das Glied gehört, z. B. erhält man für  $p = q = 0$  das Glied  $\varphi_0(s-2-i2)$ , das gerade den 4 Polen entspricht, die Eckpunkte des Quadrates mit dem Mittelpunkte  $2+i2$  sind.

Die Funktion  $f_0(s)$  ist doppelt-periodisch mit den Perioden  $2$  und  $i2$ .

Die Aufgabe besteht vor allen Dingen aus einer Verschiebung der Pole

$$4p+1+i(4q+1), \quad (4p+3)+i(4q+1), \\ (4p+3)+i(4q+3), \quad (4p+1)+i(4q+3)$$

nach dem Mittelpunkt  $4p + 2 + i(4q + 2)$  des Quadrates; da  $f_0(s)$  aber die oben erwähnten Perioden hat, werden wir uns darauf beschränken nur die Pole  $1 + i$ ,  $3 + i$ ,  $3 + i3$ ,  $1 + i3$  zu betrachten, und diese sollen nach dem Punkte  $2 + i2$  verschoben werden. (Fig. 1).

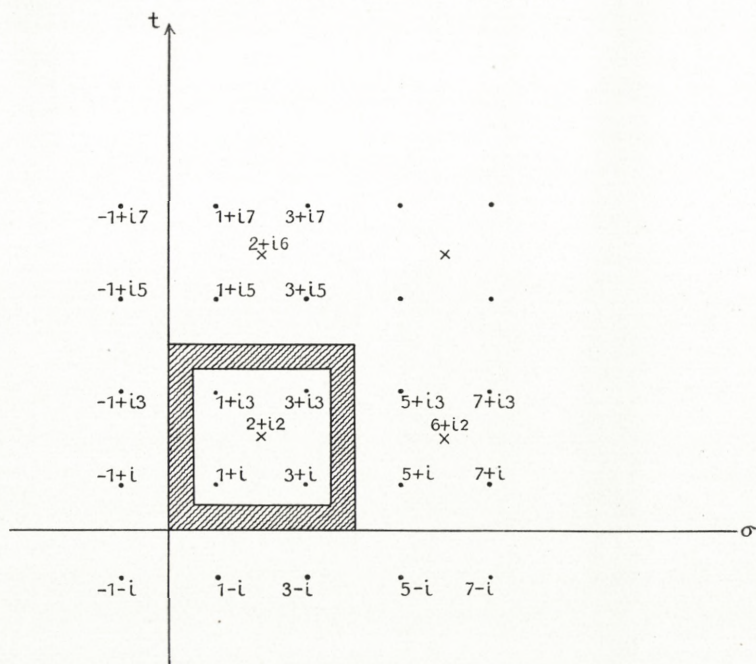


Fig. 1.

Zunächst verschieben wir den Pol  $1 + i$  nach  $2 + i2$ , dies erreicht man durch Entwicklung des Bruches  $\frac{1}{(s-1-i)^3}$  vom Punkte  $2 + i2$  aus.

Aus

$$\begin{aligned} \frac{1}{(s-1-i)^3} &= \frac{1}{(s-2-i2+1+i)^3} \\ &= \frac{1}{(s-2-i2)^3} \cdot \left(1 + \frac{1+i}{s-2-i2}\right)^{-3} \end{aligned}$$

erhält man durch Anwendung der Binomialreihe

$$\frac{1}{(s-1-i)^3} = \frac{1}{(s-2-i2)^3} + \frac{\binom{-3}{1}(1+i)}{(s-2-i2)^4} + \frac{\binom{-3}{2}(1+i)^2}{(s-2-i2)^5} + \dots,$$

diese Entwicklung ist gültig für

$$\left| \frac{1+i}{s-2-i2} \right| < 1 \quad \text{oder} \quad |s-2-i2| > |1+i| = \sqrt{2}.$$

Da der Kreis  $|s-2-i2| \leq \sqrt{2}$  ganz innerhalb des Quadrates mit den Eckpunkten  $\frac{1}{2} + i\frac{1}{2}$ ,  $\frac{7}{2} + i\frac{1}{2}$ ,  $\frac{7}{2} + i\frac{7}{2}$ ,  $\frac{1}{2} + i\frac{7}{2}$  gelegen ist, ist die obige Reihe absolut und gleichmässig konvergent in dem Gebiete der komplexen Ebene, das ausserhalb dieses Quadrates liegt (der Rand soll zum Gebiete mitgerechnet werden). Das hier erwähnte Quadrat werden wir im folgenden mit  $Q_{\frac{3}{2}}(2+i2)$  bezeichnen, wo  $2+i2$  der Mittelpunkt ist, während der Index  $\frac{3}{2}$  seinen Abstand von den Seiten angibt.

In genau derselben Weise verschiebt man dann die übrigen Pole  $3+i$ ,  $3+i3$ ,  $1+i3$  nach dem Punkte  $2+i2$ .

Jetzt haben wir also die vier Entwicklungen

$$\frac{1}{(s-1-i)^3} = \frac{1}{(s-2-i2)^3} + \frac{\binom{-3}{1}(1+i)}{(s-2-i2)^4} + \frac{\binom{-3}{2}(1+i)^2}{(s-2-i2)^5} + \dots$$

$$\frac{1}{(s-3-i)^3} = \frac{1}{(s-2-i2)^3} + \frac{\binom{-3}{1}(-1+i)}{(s-2-i2)^4} + \frac{\binom{-3}{2}(-1+i)^2}{(s-2-i2)^5} + \dots$$

$$\frac{1}{(s-3-i3)^3} = \frac{1}{(s-2-i2)^3} + \frac{\binom{-3}{1}(-1-i)}{(s-2-i2)^4} + \frac{\binom{-3}{2}(-1-i)^2}{(s-2-i2)^5} + \dots$$

$$\frac{1}{(s-1-i3)^3} = \frac{1}{(s-2-i2)^3} + \frac{\binom{-3}{1}(1-i)}{(s-2-i2)^4} + \frac{\binom{-3}{2}(1-i)^2}{(s-2-i2)^5} + \dots,$$

welche alle absolut und gleichmässig konvergent sind in dem Gebiete der komplexen Ebene, das ausserhalb des Quadrates  $Q_{\frac{3}{2}}(2+i2)$  gelegen ist (der Rand soll zum Gebiete mitgerechnet werden).

Durch Addition ergibt sich eine Entwicklung der Form

$$\varphi_0(s-2-i2) = \frac{4}{(s-2-i2)^3} + \frac{k_{0,1}}{(s-2-i2)^4} + \frac{k_{0,2}}{(s-2-i2)^5} + \dots, \quad (3)$$

die sowohl ausserhalb des Quadrates  $Q_{\frac{3}{2}}(2+i2)$  als auch auf seinem Rande absolut und gleichmässig konvergent ist. Zu bemerken ist aber, dass die Koeffizienten  $k_{0,m}$  den Wert 0 annehmen, wenn der Index  $m$  eine Zahl der Form  $4r+1$ ,  $4r+2$  oder  $4r+3$  ist.

Man erhält somit eine dem Pole  $2+i2$  entsprechende approximierende Funktion

$$\psi_1(s-2-i2) = \frac{4}{(s-2-i2)^3} + \frac{k_{0,1}}{(s-2-i2)^4} + \dots + \frac{k_{0,n_1-3}}{(s-2-i2)^{n_1}}, \quad (4)$$

wo  $n_1$  später festgelegt wird.

Wir wollen uns nun einen Überblick über die Güte der Approximation für Werte von  $s$  verschaffen, welche fern vom Pole  $2+i2$  liegen. Zu diesem Zwecke bestimmen wir die positive ganze ungerade Zahl  $N_1$  so, dass

$$\left| \frac{k_{0,1}}{(s-2-i2)^4} \right| + \dots + \left| \frac{k_{0,n-3}}{(s-2-i2)^n} \right| + \dots \leq \left| \frac{1}{(s-2-i2)^3} \right|$$

gültig ist für alle  $s$  ausserhalb oder auf dem Rande des Quadrates  $Q_{2N_1}(2+i2)$ , dessen Seiten die Entfernung  $2N_1$  vom Mittelpunkte  $2+i2$  haben, und dessen Eckpunkte

$$\begin{aligned} &2(1-N_1)+i2(1-N_1), \quad 2(1+N_1)+i2(1-N_1), \\ &2(1+N_1)+i2(1+N_1), \quad 2(1-N_1)+i2(1+N_1) \end{aligned}$$

sind. Für diese Werte von  $s$  besteht also — ganz unabhängig vom Werte von  $n_1$  — die Ungleichung

$$|\varphi_0(s-2-i2) - \psi_1(s-2-i2)| \leq \left| \frac{1}{(s-2-i2)^3} \right|. \quad (5)$$

Innerhalb des Quadrates  $Q_{2N_1}(2+i2)$  liegen  $N_1^2$  Quadrate, deren Eckpunkte Pole der ursprünglichen Funktion  $f_0(s)$  sind. (Fig. 2).

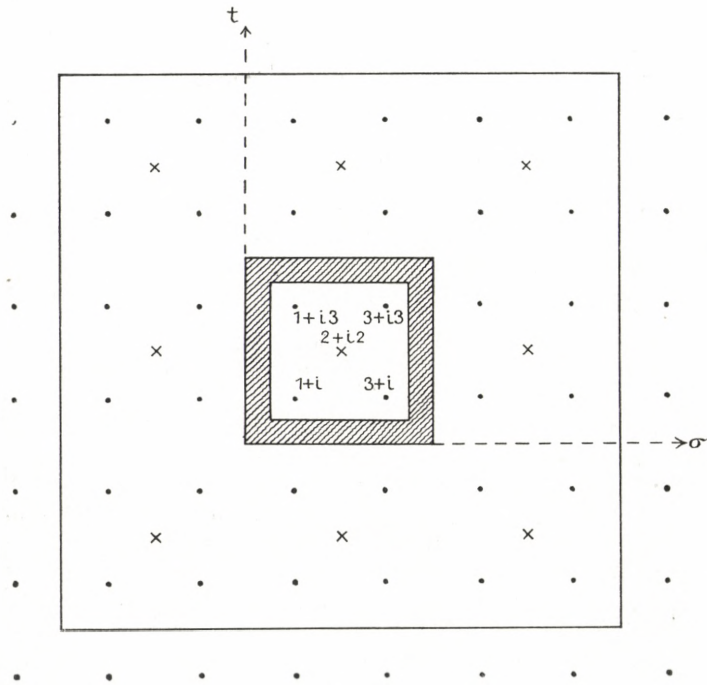


Fig. 2.

Wir gehen jetzt dazu über, die Grösse des die approximierende Funktion (4) charakterisierenden Index  $n_1$  zu bestimmen.

Nachdem wir eine Folge von positiven Zahlen  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n, \dots$  gewählt haben, welche nur der Bedingung  $\sum_1^{\infty} \varepsilon_n$  konvergent unterworfen sein sollen, wird  $n_1$  derart festgelegt, dass die Ungleichung

$$|\varphi_0(s-2-i2) - \psi_1(s-2-i2)| < \frac{\varepsilon_1}{2N_1^2} \quad (6)$$



bei jedem  $s$  ausserhalb oder auf dem Rande von  $Q_{\frac{3}{2}}(2+i2)$  befriedigt ist.

Mittels der durch die Bedingungen (5) und (6) definierten Funktion  $\psi_1(s-2-i2)$  bildet man nun

$$f_1(s) = \sum_{(p,q)} \psi_1(s-(4p+2)-i(4q+2)); \quad (7)$$

diese Funktion ist doppelt-periodisch mit den Perioden 4 und  $i4$ , und sie hat Pole höchstens  $n_1$ ter Ordnung in den Punkten  $4p+2+i(4q+2)$ , wo  $(p, q)$  alle Zahlenpaare mit ganzzahligen Koordinaten durchläuft.

Als Verschiebungsgebiet  $\Gamma_1(2+i2)$  dem ersten Schritte der Polverschiebung entsprechend wählt man das geschlossene Gebiet, das im Inneren des Quadrates  $Q_2(2+i2)$ , aber ausserhalb des Quadrates  $Q_{\frac{3}{2}}(2+i2)$ , liegt, und wir setzen ohne weiteres

$$\Gamma_1(2+i2) = Q_2(2+i2) - Q_{\frac{3}{2}}(2+i2)$$

(den Rand mitgerechnet).

Nun wollen wir die Differenz  $f_0(s) - f_1(s)$  für die dem Gebiete  $\Gamma_1(2+i2)$  angehörigen Werte von  $s$  abschätzen.

Aus (1) und (7) folgt

$$\sum_{(p,q)} \{ \varphi_0(s-(4p+2)-i(4q+2)) - \psi_1(s-(4p+2)-i(4q+2)) \}. \quad (8)$$

Durch die Ungleichungen (5) und (6) erhält man für Werte von  $s$ , die zum Gebiete  $\Gamma_1(2+i2)$  gehören:

$$\begin{aligned} & | \varphi_0(s-(4p+2)-i(4q+2)) - \psi_1(s-(4p+2)-i(4q+2)) | \\ & \leq \left| \frac{1}{(s-(4p+2)-i(4q+2))^3} \right|, \end{aligned}$$

welche gültig ist, bei allen Indices  $p$  und  $q$ , die Polen ausserhalb des Quadrates  $Q_{2N_1}(2+i2)$  entsprechen, und ferner die Ungleichung

$$|\varphi_0(s-(4p+2)-i(4q+2))-\psi_1(s-(4p+2)-i(4q+2))| < \frac{\varepsilon_1}{2N_1^2},$$

die bei allen den Polen im Inneren des Quadrates  $Q_{2N_1}(2+i2)$  entsprechenden Indices  $p$  und  $q$  gültig ist.

Aus (8) ergibt sich somit — indem man sich daran erinnert, dass  $N_1^2$  Quadrate im Inneren des Quadrates  $Q_{2N_1}(2+i2)$  liegen, deren Eckpunkte Pole der ursprünglichen Funktion  $f_0(s)$  sind — die Ungleichung

$$|f_0(s)-f_1(s)| < N_1^2 \cdot \frac{\varepsilon_1}{2N_1^2} + \sum' \left| \frac{1}{(s-(4p+2)-i(4q+2))^3} \right|, \quad (9)$$

wo die letzte Summation nur über die Pole ausserhalb  $Q_{2N_1}(2+i2)$  erstreckt werden soll.

Jetzt nehmen wir an, dass  $N_1$  schon so gross gewählt ist, dass ausser der Bedingung (5) auch die Ungleichung

$$\sum' \left| \frac{1}{(s-(4p+2)-i(4q+2))^3} \right| < \frac{\varepsilon_1}{2} \quad (10)$$

für alle  $s$  im Gebiete  $T_1(2+i2)$  befriedigt ist, wenn die Summation über alle Pole  $4p+2+i(4q+2)$  ausserhalb des Quadrates  $Q_{2N_1}(2+i2)$  erstreckt wird. Diese Bedingung ist natürlich wie auch die frühere unabhängig von der Wahl von  $n_1$ .

Aus (9) erhalten wir nun

$$|f_0(s)-f_1(s)| < N_1^2 \frac{\varepsilon_1}{2N_1^2} + \frac{\varepsilon_1}{2} = \varepsilon_1,$$

also

$$|f_0(s)-f_1(s)| < \varepsilon_1$$

für alle Punkte  $s$  im Gebiete  $T_1(2+i2)$ . Da aber die Funktion  $f_0(s) - f_1(s)$  die primitiven Perioden 4 und  $i4$  hat, gilt diese Ungleichung auch in jedem der Gebiete

$$T_1(4p+2+i(4q+2)) \\ = Q_2(4p+2+i(4q+2)) - Q_{\frac{3}{2}}(4p+2+i(4q+2)),$$

oder mit anderen Worten: sie gilt auch in einem »Kreuz«, das die beiden Achsen umschliesst. Genauer ausgedrückt:

$$|f_0(s) - f_1(s)| < \varepsilon_1 \quad (11)$$

ist gültig für jedes  $s$  in den Gebieten

$$-\frac{1}{2} \leq \sigma \leq \frac{1}{2} \quad \text{und} \quad -\frac{1}{2} \leq t \leq \frac{1}{2}.$$

Damit ist der erste Schritt der Polverschiebung beendet, und wir haben dadurch eine Funktion  $f_1(s)$  erreicht, welche doppelt-periodisch mit den Perioden 4 und  $i4$  ist und Pole höchstens  $n_1$ ter Ordnung in den Punkten  $4p+2+i(4q+2)$  hat, wo  $(p, q)$  alle Zahlenpaare mit ganzzahligen Koordinaten durchläuft.

Als Ausgangsfunktion bei dem zweiten Schritte der Polverschiebung benutzen wir nun diese Funktion

$$f_1(s) = \sum_{(p,q)} \psi_1(s - (4p+2) - i(4q+2)); \quad (7)$$

auch an dieser Stelle ist es zweckmässig, die Reihe umzuschreiben und zwar folgendermassen

$$f_1(s) = \\ \sum \{ \psi_1(s - (4p-2) - i(4q-2)) + \psi_1(s - (4p+2) - i(4q-2)) \\ + \psi_1(s - (4p+2) - i(4q+2)) + \psi_1(s - (4p-2) - i(4q+2)) \};$$

das allgemeine Glied dieser Reihe enthält die 4 Glieder der ursprünglichen Darstellung, welche den Polen entsprechen, die Eckpunkte eines Quadrates mit dem Mittelpunkte  $4p + i4q$  sind, und die Summation soll über alle Zahlenpaare  $(p, q)$ , deren Koordinaten ganze ungerade Zahlen sind, erstreckt werden.

Ersetzt man dann  $p$  durch  $2p + 1$  und  $q$  durch  $2q + 1$ , so entsteht die Reihe

$$f_1(s) = \sum_{(p,q)} \left\{ \psi_1(s - (8p + 2) - i(8q + 2)) + \psi_1(s - (8p + 6) - i(8q + 2)) \right. \\ \left. + \psi_1(s - (8p + 6) - i(8q + 6)) + \psi_1(s - (8p + 2) - i(8q + 6)) \right\},$$

wo  $(p, q)$  alle Zahlenpaare mit ganzzahligen Koordinaten durchläuft; das allgemeine Glied entspricht den Polen, welche Eckpunkte eines Quadrates mit dem Mittelpunkte  $8p + 4 + i(8q + 4)$  sind.

Setzt man

$$\varphi_1(s - 4 - i4) = \psi_1(s - 2 - i2) + \psi_1(s - 6 - i2) + \psi_1(s - 6 - i6) + \psi_1(s - 2 - i6),$$

so lässt die gegebene Funktion sich auf die Form

$$f_1(s) = \sum_{(p,q)} \varphi_1(s - (8p + 4) - i(8q + 4)) \quad (7)$$

bringen. Die Bezeichnungen sind derart gewählt, dass man für gegebene  $p$  und  $q$  unmittelbar sieht, zu welchem Quadrate das Glied gehört, z. B. erhält man für  $p = q = 0$  das Glied  $\varphi_1(s - 4 - i4)$ , das gerade den 4 Polen entspricht, die Eckpunkte eines Quadrates mit dem Mittelpunkte  $4 + i4$  sind.

Die Aufgabe besteht nun darin, die Pole

$$(8p+2) + i(8q+2), \quad (8p+6) + i(8q+2), \\ (8p+6) + i(8q+6), \quad (8p+2) + i(8q+6)$$

nach dem Mittelpunkt  $(8p+4) + i(8q+4)$  des Quadrates zu verschieben; da aber  $f_1(s)$  die früher erwähnten Perioden hat, können wir uns darauf beschränken, nur die Pole  $2+i2$ ,  $6+i2$ ,  $6+i6$ ,  $2+i6$  zu betrachten, und diese sollen dann nach dem Punkte  $4+i4$  verschoben werden.

Das Verfahren bleibt stets in genauer Analogie mit dem bei dem ersten Schritte der Polverschiebung benutzten. Durch Anwendung der Binomialreihe auf die einzelnen Glieder von  $\varphi_1(s-4-i4)$ , wobei  $4+i4$  der Entwicklungspunkt ist, erhält man die folgende Darstellung

$$\varphi_1(s-4-i4) = \frac{16}{(s-4-i4)^3} + \frac{k_{1,1}}{(s-4-i4)^4} + \frac{k_{1,2}}{(s-4-i4)^5} + \dots;$$

diese Reihe ist absolut und gleichmässig konvergent ausserhalb und auf dem Rande des Quadrates  $Q_3(4+i4)$ .

Wir finden hieraus eine dem Pole  $4+i4$  entsprechende approximierende Funktion

$$\psi_2(s-4-i4) = \frac{16}{(s-4-i4)^3} + \frac{k_{1,1}}{(s-4-i4)^4} + \dots + \frac{k_{1,n_2-3}}{(s-4-i4)^{n_2}}, \quad (12)$$

wo  $n_2$  doch erst später festgelegt wird.

Wir wollen uns jetzt einen Überblick über die Güte der Approximation für Werte von  $s$  verschaffen, welche fern vom Pole  $4+i4$  liegen. Zu diesem Zwecke bestimmen wir die positive ganze ungerade Zahl  $N_2$  so, dass

$$\left| \frac{k_{1,1}}{(s-4-i4)^4} \right| + \dots + \left| \frac{k_{1,n_2-3}}{(s-4-i4)^{n_2}} \right| + \dots \leq \left| \frac{1}{(s-4-i4)^3} \right|$$

gültig ist für alle  $s$  ausserhalb oder auf dem Rande des Quadrates  $Q_{4N_2}(4+i4)$ , dessen Seiten den Abstand  $4N_2$  vom Mittelpunkte  $4+i4$  haben, und dessen Eckpunkte

$$\begin{aligned} &4(1-N_2)+i4(1-N_2), \quad 4(1+N_2)+i4(1-N_2), \\ &4(1+N_2)+i4(1+N_2), \quad 4(1-N_2)+i4(1+N_2) \end{aligned}$$

sind. Für diese Werte von  $s$  besteht also — ganz unabhängig vom Werte von  $n_2$  — die Ungleichung

$$|\varphi_1(s-4-i4)-\psi_2(s-4-i4)| \leq \left| \frac{1}{(s-4-i4)^3} \right|. \quad (13)$$

Im Inneren des Quadrates  $Q_{4N_2}(4+i4)$  liegen  $N_2^2$  Quadrate, deren Eckpunkte Pole der ursprünglichen Funktion  $f_1(s)$  sind.

Wir gehen jetzt dazu über, die Grösse des die approximierende Funktion (12) charakterisierenden Index  $n_2$  zu bestimmen.

Wir wählen  $n_2$  derart, dass die Ungleichung

$$|\varphi_1(s-4-i4)-\psi_2(s-4-i4)| < \frac{\varepsilon_2}{2N_2^2} \quad (14)$$

bei jedem  $s$  ausserhalb oder auf dem Rande von  $Q_3(4+i4)$  befriedigt ist. Mittels der durch die beiden Bedingungen (13) und (14) festgelegten Funktion bildet man nun die Funktion

$$f_2(s) = \sum_{(p,q)} \psi_2(s-(8p+4)-i(8q+4)); \quad (15)$$

diese ist doppelt-periodisch mit den Perioden 8 und  $i8$ , und sie hat Pole höchstens  $n_2$ ter Ordnung in den Punkten  $8p+4+i(8q+4)$ , wo  $(p,q)$  alle Zahlenpaare mit ganzzahligen Koordinaten durchläuft.

Als das dem zweiten Schritte der Polverschiebung entsprechende Verschiebungsgebiet  $T_2(4+i4)$  wählen wir das geschlossene Gebiet, das im Inneren des Quadrates  $Q_4(4+i4)$ , aber ausserhalb des Quadrates  $Q_3(4+i4)$ , liegt, und wir setzen ohne weiteres

$$T_2(4+i4) = Q_4(4+i4) - Q_3(4+i4)$$

(den Rand mitgerechnet).

Nun wollen wir die Differenz  $f_1(s) - f_2(s)$  für die dem Gebiete  $T_2(4+i4)$  angehörigen Werte von  $s$  abschätzen.

Aus (7) und (15) folgt

$$f_1(s) - f_2(s) = \sum_{(p,q)} \{ \varphi_1(s - (8p+4) - i(8q+4)) - \psi_2(s - (8p+4) - i(8q+4)) \}. \quad (16)$$

Durch die Ungleichungen (13) und (14) erhält man für Werte von  $s$ , die zum Gebiete  $T_2(4+i4)$  gehören:

$$\begin{aligned} & \left| \varphi_1(s - (8p+4) - i(8q+4)) - \psi_2(s - (8p+4) - i(8q+4)) \right| \\ & \leq \left| \frac{1}{(s - (8p+4) - i(8q+4))^3} \right|, \end{aligned}$$

welche gültig ist bei allen Indices  $p$  und  $q$ , die Polen ausserhalb des Quadrates  $Q_{4N_2}(4+i4)$  entsprechen, und ferner die Ungleichung

$$\left| \varphi_1(s - (8p+4) - i(8q+4)) - \psi_2(s - (8p+4) - i(8q+4)) \right| < \frac{\varepsilon_2}{2N_2^2},$$

die bei allen den Polen im Inneren des Quadrates  $Q_{4N_2}(4+i4)$  entsprechenden Indices  $p$  und  $q$  gültig ist.

Aus (16) ergibt sich somit — indem man sich daran erinnert, dass  $N_2^2$  Quadrate im Inneren des Quadrates  $Q_{4N_2}(4+i4)$  liegen, deren Eckpunkte Pole der ursprünglichen Funktion  $f_1(s)$  sind — die Ungleichung

$$|f_1(s) - f_2(s)| < N_2^2 \frac{\epsilon_2}{2N_2^2} + \sum' \left| \frac{1}{(s - (8p+4) - i(8q+4))^3} \right|, \quad (17)$$

wo die letzte Summation nur über die Pole ausserhalb  $Q_{4N_2}(4+i4)$  erstreckt werden soll.

Jetzt nehmen wir an, dass  $N_2$  im voraus so gross gewählt ist, dass ausser der Bedingung (13) auch die Ungleichung

$$\sum' \left| \frac{1}{(s - (8p+4) - i(8q+4))^3} \right| < \frac{\epsilon_2}{2}$$

für alle  $s$  im Gebiete  $\Gamma_2(4+i4)$  befriedigt ist, wenn die Summation über alle Pole  $8p+4+i(8q+4)$  ausserhalb des Quadrates  $Q_{4N_2}(4+i4)$  erstreckt wird. Diese Bedingung ist, wie auch die frühere, unabhängig von der Wahl von  $n_2$ .

Aus (17) erhalten wir nun

$$|f_1(s) - f_2(s)| < N_2^2 \cdot \frac{\epsilon_2}{2N_2^2} + \frac{\epsilon_2}{2} = \epsilon_2,$$

also

$$|f_1(s) - f_2(s)| < \epsilon_2 \quad (18)$$

für alle Punkte  $s$  im Gebiete  $\Gamma_2(4+i4)$ . Da aber die Funktion  $f_1(s) - f_2(s)$  die primitiven Perioden 8 und  $i8$  hat, gilt diese Ungleichung auch in jedem der Gebiete  $\Gamma_2(8p+4+i(8q+4))$ , oder mit anderen Worten: sie gilt in einem »Kreuz«, das die beiden Achsen umschliesst. Genauer ausgedrückt:

$$|f_1(s) - f_2(s)| < \epsilon_2 \quad (18)$$

ist gültig für jedes  $s$  in den Gebieten

$$-1 \leq \sigma \leq 1 \quad \text{und} \quad -1 \leq t \leq 1.$$

Es wird nun angenommen, dass wir durch den  $(m-1)$ ten Schritt eine Funktion  $f_{m-1}(s)$  erreicht haben, welche dop-



pelt-periodisch mit den Perioden  $2^m$  und  $i2^m$  ist und Pole höchstens  $n_{m-1}$ ter Ordnung in den Punkten eines Quadratnetzes  $2^m p + 2^{m-1} + i(2^m q + 2^{m-1})$  hat, wo  $(p, q)$  alle Zahlenpaare mit ganzzahligen Koordinaten durchläuft.

In Analogie mit den beiden ersten Schritten wird  $f_{m-1}(s)$  auf die Form

$$f_{m-1}(s) = \sum_{(p,q)} \psi_{m-1}(s - (2^m p + 2^{m-1}) - i(2^m q + 2^{m-1}))$$

gebracht, die man dann in

$$\begin{aligned} f_{m-1}(s) = \sum_{(p,q)} \{ & \psi_{m-1}(s - (2^m p - 2^{m-1}) - i(2^m q - 2^{m-1})) \\ & + \psi_{m-1}(s - (2^m p + 2^{m-1}) - i(2^m q - 2^{m-1})) \\ & + \psi_{m-1}(s - (2^m p + 2^{m-1}) - i(2^m q + 2^{m-1})) \\ & + \psi_{m-1}(s - (2^m p - 2^{m-1}) - i(2^m q + 2^{m-1})) \} \end{aligned}$$

überführt. Das allgemeine Glied dieser Reihe enthält die 4 Glieder der ursprünglichen Darstellung, welche den Polen entsprechen, die Eckpunkte eines Quadrates mit dem Mittelpunkt  $2^m p + i2^m q$  sind, und die Summation soll über alle Zahlenpaare  $(p, q)$ , deren Koordinaten ganze ungerade Zahlen sind, erstreckt werden.

Ersetzt man  $p$  durch  $2p + 1$  und  $q$  durch  $2q + 1$ , so entsteht die Reihe

$$\begin{aligned} f_{m-1}(s) = \sum_{(p,q)} \{ & \psi_{m-1}(s - (2^{m+1} p + 2^{m-1}) - i(2^{m+1} q + 2^{m-1})) \\ & + \psi_{m-1}(s - (2^{m+1} p + 3 \cdot 2^{m-1}) - i(2^{m+1} q + 2^{m-1})) \\ & + \psi_{m-1}(s - (2^{m+1} p + 3 \cdot 2^{m-1}) - i(2^{m+1} q + 3 \cdot 2^{m-1})) \\ & + \psi_{m-1}(s - (2^{m+1} p + 2^{m-1}) - i(2^{m+1} q + 3 \cdot 2^{m-1})) \} \end{aligned}$$

wo  $(p, q)$  alle Zahlenpaare mit ganzzahligen Koordinaten durchläuft; das allgemeine Glied entspricht den Polen, welche Eckpunkte eines Quadrates mit dem Mittelpunkte  $2^{m+1}p + 2^m + i(2^{m+1}q + 2^m)$  sind.

Setzt man

$$\begin{aligned} & \varphi_{m-1}(s - 2^m - i2^m) = \\ & \psi_{m-1}(s - 2^{m-1} - i2^{m-1}) + \psi_{m-1}(s - 3 \cdot 2^{m-1} - i2^{m-1}) \\ & + \psi_{m-1}(s - 3 \cdot 2^{m-1} - i3 \cdot 2^{m-1}) + \psi_{m-1}(s - 2^{m-1} - i \cdot 3 \cdot 2^{m-1}), \end{aligned}$$

so lässt die gegebene Funktion sich auf die Form

$$f_{m-1}(s) = \sum_{(p,q)} \varphi_{m-1}(s - (2^{m+1}p + 2^m) - i(2^{m+1}q + 2^m)) \quad (19)$$

bringen. Die Bezeichnungen sind so gewählt, dass man für gegebene  $p$  und  $q$  unmittelbar sieht, zu welchem Quadrate das Glied gehört, z. B. erhält man für  $p = q = 0$  das Glied  $\varphi_{m-1}(s - 2^m - i2^m)$ , das gerade den vier Polen entspricht, die Eckpunkte eines Quadrates mit dem Mittelpunkte  $2^m + i2^m$  sind.

Die Aufgabe besteht nun darin, die Pole

$$\begin{aligned} & 2^{m+1}p + 2^{m-1} + i(2^{m+1}q + 2^{m-1}), \quad 2^{m+1}p + 3 \cdot 2^{m-1} + i(2^{m+1}q + 2^{m-1}), \\ & 2^{m+1}p + 3 \cdot 2^{m-1} + i(2^{m+1}q + 3 \cdot 2^{m-1}), \quad 2^{m+1}p + 2^{m-1} + i(2^{m+1}q + 3 \cdot 2^{m-1}) \end{aligned}$$

nach dem Mittelpunkte  $2^{m+1}p + 2^m + i(2^{m+1}q + 2^m)$  des Quadrates zu verschieben; da aber  $f_{m-1}(s)$  die früher erwähnten Perioden hat, können wir uns darauf beschränken, nur die Pole

$$2^{m-1} + i2^{m-1}, \quad 3 \cdot 2^{m-1} + i2^{m-1}, \quad 3 \cdot 2^{m-1} + i \cdot 3 \cdot 2^{m-1}, \quad 2^{m-1} + i \cdot 3 \cdot 2^{m-1}$$

zu betrachten, und diese sollen dann nach dem Punkte  $2^m + i2^m$  verschoben werden.

Wir fahren fort in genauer Analogie mit der bei dem ersten und zweiten Schritte der Polverschiebung benutzten Methode. Durch Anwendung der Binomialreihe auf die einzelnen Glieder von  $\varphi_{m-1}(s-2^m-i2^m)$ , wobei  $2^m+i2^m$  der Entwicklungspunkt ist, erhält man die folgende Darstellung

$$\varphi_{m-1}(s-2^m-i2^m) = \frac{4^m}{(s-2^m-i2^m)^3} + \frac{k_{m-1,1}}{(s-2^m-i \cdot 2^m)^4} + \dots,$$

diese Reihe ist absolut und gleichmässig konvergent ausserhalb und auf dem Rande des Quadrates  $Q_{3 \cdot 2^{m-2}}(2^m+i2^m)$ .

Wir finden hieraus eine dem Pole  $2^m+i2^m$  entsprechende approximierende Funktion

$$\psi_m(s-2^m-i2^m) = \frac{4^m}{(s-2^m-i2^m)^3} + \dots + \frac{k_{m-1, n_m-3}}{(s-2^m-i2^m)^{n_m}}, \quad (20)$$

wo  $n_m$  doch erst später festgelegt wird.

Die ungerade ganze Zahl  $N_m$  bestimmen wir nun so, dass

$$\left| \frac{k_{m-1,1}}{(s-2^m-i2^m)^4} \right| + \dots + \left| \frac{k_{m-1, n_m-3}}{(s-2^m-i2^m)^{n_m}} \right| + \dots \leq \left| \frac{1}{(s-2^m-i2^m)^3} \right|$$

gültig ist für alle  $s$  ausserhalb oder auf dem Rande des Quadrates  $Q_{2^{N_m}}(2^m+i2^m)$ , dessen Seiten den Abstand  $2^m N_m$  vom Mittelpunkte  $2^m+i2^m$  haben, und dessen Eckpunkte

$$\begin{aligned} &4(1-N_m) + i4(1-N_m), & 4(1+N_m) + i4(1-N_m), \\ &4(1+N_m) + i4(1+N_m), & 4(1-N_m) + i4(1+N_m) \end{aligned}$$

sind. Für diese Werte von  $s$  besteht also — ganz unabhängig vom Werte von  $n_m$  — die Ungleichung

$$\left| \varphi_{m-1}(s-2^m-i2^m) - \psi_m(s-2^m-i2^m) \right| \leq \left| \frac{1}{(s-2^m-i2^m)^3} \right|. \quad (21)$$

Im Inneren des Quadrates  $Q_{2^m N_m} (2^m + i2^m)$  liegen  $N_m^2$  Quadrate, deren Eckpunkte Pole der ursprünglichen Funktion  $f_{m-1}(s)$  sind.

Wir gehen nun dazu über, die Grösse des Index  $n_m$ , der die approximierende Funktion (20) charakterisiert, zu bestimmen;  $n_m$  wird so gewählt, dass die Ungleichung

$$\left| \varphi_{m-1}(s - 2^m - i2^m) - \psi_m(s - 2^m - i2^m) \right| < \frac{\epsilon_m}{2 N_m^2}$$

bei jedem  $s$  ausserhalb oder auf dem Rande von  $Q_{3 \cdot 2^{m-2}} (2^m + i2^m)$  befriedigt ist.

Mittels der durch die obenerwähnten Bedingungen festgelegten Funktion  $\psi_m(s - 2^m - i2^m)$  bildet man die Funktion

$$f_m(s) = \sum_{(p,q)} \psi_m(s - (2^{m+1}p + 2^m) - i(2^{m+1}q + 2^m)); \quad (22)$$

diese Funktion ist doppelt-periodisch mit den Perioden  $2^{m+1}$  und  $i2^{m+1}$ , und sie hat Pole höchstens  $n_m$  ter Ordnung in den Punkten  $2^{m+1}p + 2^m + i(2^{m+1}q + 2^m)$ , wo  $(p, q)$  alle Zahlenpaare mit ganzzahligen Koordinaten durchläuft.

Als Verschiebungsgebiet  $\Gamma_m(2^m + i2^m)$  dem  $m$ ten Schritte der Polverschiebung entsprechend wählt man das geschlossene Gebiet, das im Inneren des Quadrates  $Q_{2^m}(2^m + i2^m)$ , aber ausserhalb des Quadrates  $Q_{3 \cdot 2^{m-2}}(2^m + i2^m)$ , liegt, und wir setzen ohne weiteres

$$\Gamma_m(2^m + i2^m) = Q_{2^m}(2^m + i2^m) - Q_{3 \cdot 2^{m-2}}(2^m + i2^m)$$

(den Rand mitgerechnet).

Nun wollen wir die Differenz  $f_{m-1}(s) - f_m(s)$  für die dem Gebiete  $\Gamma_m(2^m + i2^m)$  angehörigen Werte von  $s$  abschätzen.

Aus (19) und (22) folgt

$$f_{m-1}(s) - f_m(s) = \sum_{(p,q)} \left\{ \varphi_{m-1}(s - (2^{m+1}p + 2^m) - i(2^{m+1}q + 2^m)) - \psi_m(s - (2^{m+1}p + 2^m) - i(2^{m+1}q + 2^m)) \right\}. \quad (23)$$

Durch die oben gefundenen Ungleichungen erhält man für Werte von  $s$ , die zum Gebiete  $\Gamma_m(2^m + i2^m)$  gehören:

$$\begin{aligned} & \left| \varphi_{m-1}(s - (2^{m+1}p + 2^m) - i(2^{m+1}q + 2^m)) \right. \\ & \left. - \psi_m(s - (2^{m+1}p + 2^m) - i(2^{m+1}q + 2^m)) \right| \\ & \leq \left| \frac{1}{(s - (2^{m+1}p + 2^m) - i(2^{m+1}q + 2^m))^3} \right|, \end{aligned}$$

welche gültig ist bei allen Indices  $p$  und  $q$ , die Polen ausserhalb des Quadrates  $Q_{2^m N_m}(2^m + i2^m)$  entsprechen, und ferner die Ungleichung

$$\begin{aligned} & \left| \varphi_{m-1}(s - (2^{m+1}p + 2^m) - i(2^{m+1}q + 2^m)) \right. \\ & \left. - \psi_m(s - (2^{m+1}p + 2^m) - i(2^{m+1}q + 2^m)) \right| < \frac{\varepsilon_m}{2 N_m^2}, \end{aligned}$$

die bei allen den Polen im Inneren des Quadrates  $Q_{2^m N_m}(2^m + i2^m)$  entsprechenden Indices  $p$  und  $q$  gültig ist.

Aus (23) ergibt sich somit — indem man sich daran erinnert, dass  $N_m^2$  Quadrate im Inneren des Quadrates  $Q_{2^m N_m}(2^m + i2^m)$  liegen, deren Eckpunkte Pole der ursprünglichen Funktion  $f_{m-1}(s)$  sind — die Ungleichung

$$\left. \begin{aligned} & |f_{m-1}(s) - f_m(s)| < N_m^2 \cdot \frac{\varepsilon_m}{2 N_m^2} \\ & + \sum' \left| \frac{1}{(s - (2^{m+1}p + 2^m) - i(2^{m+1}q + 2^m))^3} \right|, \end{aligned} \right\} \quad (24)$$

wo die letzte Summation nur über die Pole ausserhalb des Quadrates  $Q_{2^m N_m}(2^m + i2^m)$  erstreckt werden soll.

Jetzt nehmen wir an, dass  $N_m$  im voraus so gross gewählt ist, dass ausser der Bedingung (21) auch die Ungleichung

$$\sum' \left| \frac{1}{(s - (2^{m+1}p + 2^m) - i(2^{m+1}q + 2^m))^3} \right| < \frac{\epsilon_m}{2},$$

für alle  $s$  im Gebiete  $\Gamma_m(2^m + i2^m)$  befriedigt ist, wenn die Summation über alle Pole ausserhalb des Quadrates  $Q_{2^m N_m}(2^m + i2^m)$  erstreckt wird. Diese Bedingung ist unabhängig von der Wahl von  $n_m$ .

Aus (24) erhält man

$$|f_{m-1}(s) - f_m(s)| < N_m^2 \cdot \frac{\epsilon_m}{2 N_m^2} + \frac{\epsilon_m}{2} = \epsilon_m,$$

also

$$|f_{m-1}(s) - f_m(s)| < \epsilon_m$$

für alle Punkte  $s$  im Gebiete  $\Gamma_m(2^m + i2^m)$ . Da aber die Funktion  $f_{m-1}(s) - f_m(s)$  die primitiven Perioden  $2^{m+1}$  und  $i2^{m+1}$  hat, gilt diese Ungleichung auch in jedem der Gebiete  $\Gamma_m(2^{m+1}p + 2^m + i(2^{m+1}q + 2^m))$ , oder mit anderen Worten: sie gilt in einem »Kreuze«, das die beiden Achsen umschliesst. Genauer ausgedrückt:

$$|f_{m-1}(s) - f_m(s)| < \epsilon_m \quad (25)$$

ist gültig für jedes  $s$  in den Gebieten

$$-2^{m-2} \leq \sigma \leq 2^{m-2} \quad \text{und} \quad -2^{m-2} \leq t \leq 2^{m-2}.$$

Nachdem wir nun die Konstruktion der Funktionen  $f_1(s), f_2(s), \dots, f_m(s), \dots$  durchgeführt haben, müssen wir vor allem zeigen, dass die Folge

$$f_0(s), f_1(s), \dots, f_m(s), \dots$$

eine Grenzfunktion  $f(s)$  hat, die eine ganze Transzendente ist.

Zu diesem Zwecke betrachtet man, wie gewöhnlich, das Verhalten im Inneren des Kreises  $|s| < \varrho$ .

Wird  $M$  so gross gewählt, dass  $2^{M-2} > \varrho$ , ergibt sich aus (25) die Ungleichung

$$|f_{m-1}(s) - f_m(s)| < \varepsilon_m,$$

die für  $m \geq M$  bei allen Punkten  $s$  im Inneren des Kreises gültig ist.

Wir gehen jetzt dazu über, die Konvergenz der Reihe

$$f_M(s) + \sum_{M+1}^{\infty} (f_m(s) - f_{m-1}(s))$$

innerhalb des Kreises  $|s| < \varrho$  zu untersuchen.

Da die Pole der einzelnen Glieder dieser Reihe alle im Gebiete  $|s| \geq 2^M > 2^{M-2} > \varrho$  liegen, sieht man unmittelbar, dass jedes der Glieder eine analytische Funktion im Gebiete  $|s| < \varrho$  ist. Aus der Ungleichung  $|f_m(s) - f_{m-1}(s)| < \varepsilon_m$  in Verbindung mit unserer Voraussetzung, dass  $\sum_1^{\infty} \varepsilon_m$  konvergent ist, ergibt sich nunmehr durch eine Majorantenbetrachtung, dass die Reihe gleichmässig konvergent für  $|s| < \varrho$  ist, und dass ihre Summe  $f(s)$  eine analytische Funktion ist.

Dies ist aber damit gleichbedeutend, dass die Folge von Funktionen  $f_0(s), f_1(s), \dots$  eine Grenzfunktion  $f(s)$  hat. Da  $\varrho$  beliebig gross gewählt werden kann, haben wir damit bewiesen, dass  $f(s)$  eine ganze Transzendente ist.

Um den trivialen Fall, wo  $f(s)$  auf eine Konstante reduziert wird, auszuschliessen, wollen wir die Zahlen  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_m, \dots$  noch einer Bedingung unterwerfen.

Aus

$$f(s) = f_0(s) + \sum_1^{\infty} (f_m(s) - f_{m-1}(s))$$

erhält man für  $\sigma = 0$

$$f(it) = f_0(it) + \sum_1^{\infty} (f_m(it) - f_{m-1}(it)),$$

wo

$$|f_m(it) - f_{m-1}(it)| < \varepsilon_m.$$

Wird jetzt ein solcher Wert  $t_1$  von  $t$  gewählt, für welchen  $f_0(it_1)$  von Null verschieden ist, und wird dann die Zahlenfolge  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_m, \dots$  derart festgelegt, dass die Summe der Reihe  $\sum_1^{\infty} \varepsilon_m$  kleiner als  $\frac{1}{2}|f_0(it_1)|$  wird, so erhält man — indem man sich daran erinnert, dass  $f_0(0) = 0$  — die folgenden Ungleichungen

$$|f(it_1)| > |f_0(it_1)| - \sum_1^{\infty} \varepsilon_m > \frac{1}{2}|f_0(it_1)|$$

$$|f(0)| < |f_0(0)| + \sum_1^{\infty} \varepsilon_m < \frac{1}{2}|f_0(it_1)|.$$

Falls man den Zahlen  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_m, \dots$  diese verschärfte Bedingung auferlegt, wird  $f(s)$  also nicht auf eine Konstante reduziert.

Es bleibt noch übrig, die fastperiodische Struktur der Grenzfunktion zu untersuchen.

$c$  bezeichne eine beliebige positive Zahl; dann bestimmt man  $M$  so, dass  $2^{M-2} > c$ ; aus der Darstellung



$$f(s) = f_M(s) + \sum_{M+1}^{\infty} (f_m(s) - f_{m-1}(s))$$

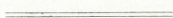
lässt sich nun sofort schliessen, dass  $f(s)$  im Streifen  $(-c, c)$  fastperiodisch ist, denn die Reihe ist gleichmässig konvergent, und ihre einzelnen Glieder sind in diesem Streifen fastperiodische (sogar reinperiodische) Funktionen. Da aber diese Überlegung für jede feste positive Zahl  $c$  gültig ist, haben wir damit bewiesen, dass die Grenzfunktion  $f(s)$  jedenfalls in  $[\sigma] = [-\infty, +\infty]$  fastperiodisch ist. Als Grenzfunktion reinperiodischer Funktionen ist  $f(s)$  übrigens grenzperiodisch.

In genau derselben Weise studiert man das Verhalten im Streifen  $-c < t < c$ ; man zeigt, dass  $f(s)$  in diesem Streifen fastperiodisch ist,  $f(s)$  ist also in jedem festen Streifen  $-c < t < c$  eine fastperiodische Funktion von  $s$ , d. h.  $f(s)$  ist fastperiodisch in  $[t] = [-\infty, +\infty]$ .

Es ist uns somit gelungen, eine Funktion  $f(s) = f(\sigma + it)$  aufzubauen, welche die folgenden Eigenschaften hat:

$f(\sigma + it)$  ist sowohl eine fastperiodische Funktion von  $t$  für jedes feste  $\sigma = \sigma_0$  als auch eine fastperiodische Funktion von  $\sigma$  für jedes feste  $t = t_0$ , und ferner ist  $f(\sigma + it)$  eine fastperiodische Funktion von  $s = \sigma + it$  in  $[\sigma] = [-\infty, +\infty]$  und  $[t] = [-\infty, +\infty]$ , d. h.

Wir haben die Existenz ganzer transzendenter Funktionen von doppelt-fastperiodischer Struktur bewiesen.





Det Kgl. Danske Videnskabernes Selskab.

Mathematisk-fysiske Meddelelser. **XV**, 9.

---

# ÜBER DIE INNERE PAAR- ERZEUGUNG BEIM $\beta$ -ZERFALL

VON

N. ARLEY UND C. MØLLER



KØBENHAVN  
LEVIN & MUNKSGAARD  
EJNAR MUNKSGAARD  
1938

Printed in Denmark.  
Bianco Lunos Bogtrykkeri A/S.

In einer früheren Note hat der eine von uns<sup>1</sup> darauf hingewiesen, dass beim  $\beta$ -Zerfall ausser den gewöhnlichen Elektronen stets auch Positronen zu erwarten sind. Eine vorläufige grobe Abschätzung der Wahrscheinlichkeit dieser Prozesse ergab für die relative Anzahl  $\frac{N^+}{N^-}$  der emittierten Positronen und Elektronen die Grössenordnung  $10^{-4}$  in Übereinstimmung mit den Experimenten von ALICHANOW, ALICHANIAN und KOSODAEW<sup>2</sup>. Eine genauere numerische Auswertung von  $\frac{N^+}{N^-}$  für die in diesen Experimenten auftretenden  $\beta$ -Energien zeigte jedoch, dass die theoretisch zu erwartenden Positronenanzahlen viel zu klein sind, um diese Experimente erklären zu können.

Nach der Fermischen Theorie<sup>3</sup> besteht ein  $\beta$ -Zerfall darin, dass sich ein Neutron in ein Proton verwandelt unter gleichzeitiger Emission eines Elektrons und eines Neutrinos. Weiter ist die Wahrscheinlichkeit dafür, dass ein solcher Prozess stattfindet, in erster Näherung proportional dem Quadrat der universellen Konstante  $g$  der Fermischen Theorie. Wegen der elektrischen Ladung des Elektrons und des Protons stehen diese Teilchen in elektromagnetischer Wechselwirkung mit den virtuellen Elektronen, welche nach der DIRAC'schen Theorie<sup>4</sup> des Positrons alle

<sup>1</sup> C. MÖLLER, Nature, 137, 314 (1936).

<sup>2</sup> A. I. ALICHANOW, A. I. ALICHANIAN, M. S. KOSODAEW, Journ. de Phys. 7, 163 (1936).

<sup>3</sup> E. FERMI, Zs. f. Phys. 88, 161 (1934).

<sup>4</sup> DIRAC, Principles of Quantum Mechanics, 238 (1930).

Zustände negativer Energie auffüllen. Diese Wechselwirkung bewirkt, dass das  $\beta$ -Elektron während seiner Erzeugung einen Teil seiner Energie an ein Elektron in einem negativen Zustand abgeben kann, wobei das letztere Elektron in einen Zustand positiver Energie übergeht. Es entsteht dabei in der Verteilung der Elektronen in den negativen Zuständen ein »Loch«, welches in der DIRAC'schen Theorie als ein Positron interpretiert wird. Man erhält also in dieser Weise in höherer Näherung die Möglichkeit eines Prozesses bei dem sich ein Neutron in ein Proton verwandelt unter gleichzeitiger Emission eines Neutrinos, zweier Elektronen und eines Positrons.

Da die Wechselwirkungsenergie zwischen zwei Teilchen der Ladung plus oder minus  $e$  in erster Näherung proportional  $e^2$  ist, wird die Wahrscheinlichkeit des erwähnten Prozesses proportional  $(ge^2)^2$ , oder aus dimensionellen Gründen proportional  $(g\alpha)^2$ , wo  $\alpha = \frac{e^2}{\hbar c} = \frac{1}{137}$  die Feinstrukturkonstante bedeutet. Für  $\frac{N^+}{N^-}$  ergibt sich also

$$\frac{N^+}{N^-} = k \cdot \alpha^2,$$

wo  $k$  eine Funktion der oberen Grenze des  $\beta$ -Spektrums ist, die wir in § 1 und § 3 genauer bestimmen werden (vgl. Fig. 2).

Wie aus § 1 hervorgeht, haben wir den Prozess bei dem die zwei Elektronen und das Positron entstehen als einen Elementarprozess höherer Näherung behandelt. Man kann jedoch formal denselben Prozess auch als einen zusammengesetzten Prozess behandeln, indem man erst das emittierte  $\beta$ -Teilchen durch eine ausgehende Kugelwelle mit bestimmter Energie und mit einer Singularität im Nullpunkt

darstellt, und dann die Wahrscheinlichkeit einer Paarerzeugung durch diese Kugelwelle berechnet<sup>1</sup>. Trotzdem die beiden Methoden in ihren Ausgangspunkten sehr verschieden sind, liefern sie doch, wie wir in § 2 zeigen werden, dieselbe Endformel. Dieses Resultat ist in Übereinstimmung mit dem Ergebnis von KNIPP und UHLENBECK<sup>2</sup>, welche gezeigt haben, dass die beiden Methoden auch bei der Berechnung der Ausstrahlung von Lichtquanten beim  $\beta$ -Zerfall dasselbe Resultat liefern. Es muss jedoch bemerkt werden, dass man bei unserem Prozess im Gegensatz zu dem Prozess von KNIPP und UHLENBECK die Ladung des Protons mitberücksichtigen muss. In unserem Fall erhält man nämlich nur Übereinstimmung zwischen der von uns und der von TISZA benutzten Methode, wenn man in der letzteren der Ladung des entstehenden Protons in einer von ihm angegebenen Weise Rechnung trägt (vergl. § 2).

### § 1.

Wie in der Einleitung erwähnt, kann sich ein Neutron nach der Fermischen Theorie in ein Proton verwandeln unter gleichzeitiger Emission eines Elektrons und eines Neutrinos. Die Wahrscheinlichkeit dieses Prozesses ist in erster Näherung bestimmt durch das Matrixelement der Wechselwirkungsenergie zwischen den schweren und den leichten Teilchen (Neutronen, Protonen beziehungsweise Elektronen, Neutrinos).

Geht das Neutron vom Zustand  $N$  über in ein Proton im Zustand  $P$  unter Emission eines Elektrons im Zustand  $s$  und eines Neutrinos im Zustand  $\sigma$ , hat dieses Matrixelement die folgende Form

<sup>1</sup> Vergl. L. TISZA, Sow. Phys. 11, 425 (1937).

<sup>2</sup> J. K. KNIPP und G. UHLENBECK, Physica, III, 425 (1936).

$$(P|c_{s\sigma}|N) = g \int \Phi_N(\mathbf{X}) \bar{\psi}_p(\mathbf{X}) \cdot \bar{\psi}_s(\mathbf{X}) \delta \bar{q}_\sigma(\mathbf{X}) d\mathbf{X}, \quad (1)$$

wobei  $\Phi_N(\mathbf{X})$ ,  $\psi_p(\mathbf{X})$ ,  $\psi_s(\mathbf{X})$  und  $q_\sigma(\mathbf{X})$  die Eigenfunktionen des Neutrons, Protons, Elektrons und Neutrinos in den betreffenden Zuständen sind, und die Integration über den ganzen Raum zu erstrecken ist. Ferner ist  $g$  die universelle Konstante der Fermischen Theorie und

$$\delta = \begin{pmatrix} 0 \div 1 & 0 & 0 \\ +1 & 0 & 0 & 0 \\ 0 & 0 & 0 & +1 \\ 0 & 0 \div 1 & 0 \end{pmatrix}.$$

Wenn man den Einfluss der Kernladung auf die Eigenfunktion des Elektrons vernachlässigt, so dass man für  $\psi_s$  und  $q_\sigma$  ebene Wellen benutzen kann, wird die Wahrscheinlichkeit dafür, dass ein  $\beta$ -Elektron mit einer Energie zwischen  $E$  und  $E + dE$  emittiert wird für »erlaubte« Übergänge nach dieser Theorie gleich

$$W(E) dE = g^2 \frac{2\pi}{\hbar} \frac{(4\pi)^2}{(hc)^6} |(P|1|N)|^2 \cdot (E^2 \div m^2 c^4)^{1/2} \cdot E \cdot (D \div E)^2 \cdot dE, \quad (2)$$

wo  $D$  die Differenz der Energie des Kernes vor und nach dem Prozess ist.  $D$  ist also gleich der oberen Grenze des  $\beta$ -Spektrums, wenn man die Ruheenergie des Elektrons miteinbezieht. Ferner ist

$$(P|1|N) = \int \Phi_N(\mathbf{X}) \bar{\psi}_p(\mathbf{X}) \cdot d\mathbf{X}, \quad (3)$$

was man aus (1) erhält unter der Voraussetzung, dass die DE BROGLIE-Wellenlängen des Elektrons und des Neutrinos gross sind im Vergleich mit den Kerndimensionen.



Um nun den in der Einleitung erwähnten Prozess zu erhalten, bei dem ein Positron entsteht, müssen wir ausser der Fermischen Wechselwirkung zwischen schweren und leichten Teilchen auch die elektromagnetische Wechselwirkung zwischen Elektronen unter sich und zwischen den Elektronen und Protonen berücksichtigen, und zwar genügt es nicht einfach die Coulombwechselwirkung zu nehmen, da die  $\beta$ -Teilchen Geschwindigkeiten haben, welche mit der Lichtgeschwindigkeit vergleichbar sind. Die elektromagnetische Wechselwirkung zwischen den Elektronen gestatten einen Übergang bei dem ein Elektron übergeht vom Zustand  $1^\circ$  in den Zustand 1, während gleichzeitig ein zweites Elektron übergeht vom Zustand  $s$  in den Zustand 2. Solange man sich nun auf die Näherung beschränkt, in der die Wahrscheinlichkeitsamplitude proportional dem Quadrat der elektrischen Ladung ist, kann man bekanntlich<sup>1</sup> die Wechselwirkung beschreiben durch Einführung eines Operators, dessen Matrixelement für den betrachteten Übergang  $(1^\circ, s) \rightarrow (1, 2)$  die folgende Form hat

$$(1, 2 | U | 1^\circ, s) = (2 | \div e \Phi_{1, 1^\circ}(\boldsymbol{x}) \div e(\alpha, \boldsymbol{A}_{1, 1^\circ}(\boldsymbol{x})) | s). \quad (4)$$

$\Phi_{1, 1^\circ}(\boldsymbol{x})$  und  $\boldsymbol{A}_{1, 1^\circ}(\boldsymbol{x})$  bilden zusammen die zur Zeit  $t = 0$  und an der Stelle  $\boldsymbol{x}$  genommenen Komponenten des Viererpotentials, welches der dem Übergang  $1^\circ \rightarrow 1$  zugeordneten Ladungs- und Stromverteilung entspricht.

Für den Fall dass  $1^\circ$  und 1 beide freie Zustände sind mit den Eigenfunktionen

<sup>1</sup> C. MÖLLER, I: Zs. f. Phys. 70, 786 (1931); II: Ann. d. Phys. 14, 531 (1932). L. ROSENFELD, Zs. f. Phys. 73, 253 (1931). H. BETHE und E. FERMI, Zs. f. Phys. 77, 296 (1932). V. A. FOCK und B. PODOLSKY, Sow. Phys. 2, 275 (1932).

$$\psi_{1^0} = \frac{u_{1^0}}{V^{1/2}} \cdot e^{\frac{i}{\hbar c}(\mathbf{p}_{1^0}, \mathbf{x})}, \quad \psi_1 = \frac{u_1}{V^{1/2}} \cdot e^{\frac{i}{\hbar c}(\mathbf{p}_1, \mathbf{x})} \quad (5)$$

hat man<sup>1</sup>

$$\left. \begin{aligned} \Phi_{1, 1^0} &= \frac{h^2 c^2 e}{\pi V} \frac{\bar{u}_1 u_{1^0} \cdot e^{\frac{i}{\hbar c}(\mathbf{p}_{1^0} \div \mathbf{p}_1, \mathbf{x})}}{|\mathbf{p}_{1^0} \div \mathbf{p}_1|^2 \div (E_{1^0} \div E_1)^2} \\ \mathbf{A}_{1, 1^0} &= \frac{h^2 c^2 e}{\pi V} \frac{\bar{u}_1 \alpha u_{1^0} \cdot e^{\frac{i}{\hbar c}(\mathbf{p}_{1^0} \div \mathbf{p}_1, \mathbf{x})}}{|\mathbf{p}_{1^0} \div \mathbf{p}_1|^2 \div (E_{1^0} \div E_1)^2} \end{aligned} \right\} \quad (6)$$

Hier bedeuten  $\frac{\mathbf{p}_{1^0}}{c}$ ,  $E_{1^0}$  und  $\frac{\mathbf{p}_1}{c}$ ,  $E_1$  die Impulse und Energien der Elektronen in den Zuständen  $1^0$  und 1, während  $u_{1^0}$  und  $u_1$  die entsprechenden vierkomponentigen auf eins normierten DIRAC'schen Wellenamplituden darstellen. Endlich ist  $V$  wie gewöhnlich das Volumen eines grossen Kastens, in dem wir uns das ganze System eingesperrt denken.

Entsprechend ist die Wechselwirkung zwischen Proton und Elektron beschrieben durch einen Operator mit dem Matrixelement

$$(1, P | U_p | 1^0, P') = (P | + e \Phi_{1, 1^0} + e(\alpha^P, \mathbf{A}_{1, 1^0}) | P'), \quad (7)$$

wo  $P$  und  $P'$  zwei Protonenzustände sind, und  $\alpha^P$  der DIRAC'sche Matrixvektor für das Proton darstellt.

Es ist nun klar, dass man in höherer Näherung einen Prozess erhalten kann, bei dem das  $\beta$ -Teilchen im Moment seiner Entstehung einen Teil seiner Energie an ein Elektron in einem Zustand negativer Energie abgibt, so dass sich ein Neutron im Zustand  $N$  in ein Proton im Zustand  $P$  verwandelt unter Emission eines Neutrinos im Zustand  $\sigma$ , zweier Elektronen in den Zuständen 1 und 2 und eines

<sup>1</sup> loc. cit. C. MØLLER, I Seite 789, Formel (4).

Positrons, das dem freigewordenen negativen Elektronenzustand  $1^\circ$  entspricht.

Die Wahrscheinlichkeit pro Zeiteinheit für diesen Prozess ist dann nach der gewöhnlichen Störungsrechnung gegeben durch

$$\frac{2\pi}{\hbar} \delta(E_N + E_{1^\circ} + E_P + E_\sigma + E_1 + E_2) \cdot \left| \sum_s (1, 2 | U | 1^\circ, s) \frac{(P | c_{s\sigma} | N)}{E_1 + E_2 + E_{1^\circ} + E_s} + \sum_{P'} (1, P | U_P | 1^\circ, P') \frac{(P' | c_{2\sigma} | N)}{E_1 + E_P + E_{1^\circ} + E_{P'}} \right. \\ \left. + (\text{dasselbe mit 1 und 2 vertauscht}) \right|^2, \quad (8)$$

wo die Summationen über  $s$  und  $P'$  über alle Elektronen- bzw. alle Protonenzustände zu erstrecken sind. Das letzte Glied (dasselbe mit 1 und 2 vertauscht) ist ein Austauschglied, das der Gleichartigkeit der beiden Elektronen Rechnung trägt. (8) verschwindet also wie das Pauliprinzip verlangt, wenn die Zustände 1 und 2 identisch sind.

Wir setzen jetzt für die Eigenfunktionen der Elektronenzustände 2 und  $s$  die ebenen Wellen

$$\psi_2 = \frac{u_2}{V^{1/2}} \cdot e^{\frac{i}{\hbar c}(\mathbf{p}_2, \mathbf{x})}, \quad \psi_s = \frac{u_s}{V^{1/2}} \cdot e^{\frac{i}{\hbar c}(\mathbf{p}_s, \mathbf{x})}, \quad (9)$$

indem wir hier überall den Einfluss der Kernladung vernachlässigen.

Einsetzen von (6) und (9) in (4) ergibt dann

$$(1, 2 | U | 1^\circ, s) = \frac{\hbar^2 c^2 e^2}{\pi V^2} \frac{\bar{u}_1 u_{1^\circ} \cdot \bar{u}_2 u_s + (\bar{u}_1 \alpha u_{1^\circ}, \bar{u}_2 \alpha u_s)}{|\mathbf{p}_{1^\circ} + \mathbf{p}_1|^2 + (E_{1^\circ} + E_1)^2} \cdot \int e^{\frac{i}{\hbar c}(\mathbf{p}_{1^\circ} + \mathbf{p}_1 + \mathbf{p}_2 + \mathbf{p}_s, \mathbf{x})} d\mathbf{x}. \quad (10)$$

Das in (10) auftretende Integral über das Volumen des Kastens ist wegen der Orthogonalität der Eigenfunktionen gleich Null, ausser wenn  $\mathbf{p}_s = \mathbf{p}'$

$$\mathbf{p}' = \mathbf{p}_1 + \mathbf{p}_2 \div \mathbf{p}_1^\circ \quad (11)$$

und ist in diesem Fall gleich  $V$ , so dass der Impuls im Gegensatz zur Energie beim Übergang in einen Zwischenzustand erhalten bleibt. Daher reduziert sich die Summation über  $s$  in (8) auf eine Summe über die vier zu dem durch (11) gegebenen Impuls  $\mathbf{p}_s = \mathbf{p}'$  gehörigen Zustände. Diese letzte Summation lässt sich mit Hilfe der von CASIMIR<sup>1</sup> stammenden Methode ausführen; zum Beispiel wird unter Benutzung von (1) und (3)

$$\begin{aligned} \sum_{\mathbf{p}_s = \mathbf{p}'} \frac{\bar{u}_2 u_s \cdot (P|c_{s\sigma}|N)}{E_1 + E_2 \div E_1^\circ \div E_s} &= \frac{g}{V} \sum_{\mathbf{p}_s = \mathbf{p}'} \frac{\bar{u}_2 (E + H') u_s \cdot \bar{u}_s \delta \bar{u}_\sigma}{E^2 \div E_s^2} (P|1|N) = \\ &= \frac{g}{V} \frac{(P|1|N)}{E^2 \div \mathbf{p}'^2 \div m^2 c^4} \cdot \bar{u}_2 (E + H') \delta \bar{u}_\sigma, \end{aligned} \quad (12)$$

wo

$$\left. \begin{aligned} H' &= \div (\alpha, \mathbf{p}') \div \beta mc^2 \\ E &= E_1 + E_2 \div E_1^\circ \end{aligned} \right\} \quad (13)$$

gesetzt ist.

Für die Summe über  $s$  in (8) erhält man so, indem man

$$I = E^2 \div \mathbf{p}'^2 \div m^2 c^4, \quad I_1 = (E_1^\circ \div E_1)^2 \div |\mathbf{p}_1^\circ \div \mathbf{p}_1|^2 \quad (14)$$

gesetzt hat,

$$\begin{aligned} \div \frac{gh^2 c^2 e^2}{\pi V^2} (P|1|N) \frac{[\bar{u}_1 u_1^\circ \cdot \bar{u}_2 \div (\bar{u}_1 \alpha u_1^\circ, \bar{u}_2 \alpha)] (E + H') \delta \bar{u}_\sigma}{I \cdot I_1} &= \\ = \div \frac{gh^2 c^2 e^2}{\pi V^2} (P|1|N) \cdot X_{12}^e, \end{aligned} \quad (15)$$

wo  $X_{12}^e$  als Abkürzung eingeführt ist.

<sup>1</sup> H. CASIMIR, Helv. Phys. Act. 6, 287 (1933).

Bei der Auswertung der Summe über  $P'$  in (8) kann man das zweite Glied in (7) vernachlässigen, da es proportional ist dem Verhältnis zwischen Protonengeschwindigkeit und Lichtgeschwindigkeit. Benutzt man ferner, dass unter den früher erwähnten Voraussetzungen über die DE BROGLIE-Wellenlängen der Elektronen

$$\int \overline{\psi}_{P'} \psi_{P'} \cdot e^{\frac{i}{\hbar c}(\mathbf{p}_1 \cdot \mathbf{p}_1, X)} d\mathbf{X} = \int \overline{\psi}_P \psi_{P'} \cdot d\mathbf{X} = \begin{cases} 0 & \text{für } P' \neq P \\ 1 & \text{für } P' = P \end{cases}$$

erhält man für die Summe über  $P'$  unter Benutzung von (6)

$$+ \frac{gh^2 c^2 e^2}{\pi V^2} (P|1|N) \frac{\bar{u}_1 u_1 \cdot \bar{u}_2 \delta \bar{u}_\sigma}{I_1 \cdot (E_1 \div E_1^\circ)} = + \frac{gh^2 c^2 e^2}{\pi V^2} (P|1|N) \cdot X_{12}^P \quad (16)$$

wo  $X_{12}^P$  wieder als Abkürzung steht.

Wir fragen jetzt nach der Wahrscheinlichkeit eines Prozesses, bei dem  $\mathbf{p}_1$  in dem Intervall  $[d\Omega_1; (E_1, E_1 + dE_1)]$ ,  $\mathbf{p}_2$  in  $[d\Omega_2; (E_2, E_2 + dE_2)]$ ,  $\mathbf{p}_\sigma$  in  $[d\Omega_\sigma; (E_\sigma, E_\sigma + dE_\sigma)]$  und schliesslich  $\div \mathbf{p}_1^\circ$  in  $[d\Omega_1^\circ; (\div E_1^\circ, \div E_1^\circ + dE^+)]$  liegt, wobei  $d\Omega_1$  usw. die Raumelemente der respektiven Emissionsrichtungen bedeuten. Diese Wahrscheinlichkeit erhält man aus (8) durch Multiplikation mit den Faktoren

$$\left. \begin{aligned} & \frac{V d\Omega_1 p_1 E_1 dE_1}{(hc)^3}, & \frac{V d\Omega_2 p_2 E_2 dE_2}{(hc)^3}, & \frac{V d\Omega_\sigma E_\sigma^2 dE_\sigma}{(hc)^3}, \\ & & & \frac{V d\Omega_1^\circ p_1^\circ |E_1^\circ| dE^+}{(hc)^3}, \end{aligned} \right\} (17)$$

welche die Anzahl der Zustände in den betreffenden Intervallen angeben.

Wir summieren jetzt diese Wahrscheinlichkeit über alle Zustände des Neutrinos und über alle Spinzustände der Elektronen und des Positrons. Ferner integrieren wir über

alle Raumwinkel und erhalten dadurch die Wahrscheinlichkeit dafür, dass die beiden Elektronen und das Positron mit Energien in den Intervallen  $(E_1, E_1 + dE_1)$ ,  $(E_2, E_2 + dE_2)$  und  $(\div E_{1^0}, \div E_{1^0} + dE^+)$  emittiert werden. Indem wir die Energie  $\div E_{1^0}$  des Positrons auch mit  $E^+$  bezeichnen,  $E_N \div E_P = D$  setzen und (13), (15) und (16) benützen, erhalten wir für diese Wahrscheinlichkeit den Ausdruck

$$P(E_1, E_2, E^+) dE_1 dE_2 dE^+ = \frac{2\pi}{\hbar} \frac{V^4 p_1 E_1 p_2 E_2 (D \div E)^2 p_{1^0} E^+}{(hc)^{12}} g^2 (hc)^4 e^4 | \langle P | 1 | N \rangle |^2 (4\pi)^4 \cdot f(E_1, E_2, E_{1^0}) dE_1 dE_2 dE^+, \quad (18)$$

$$f(E_1, E_2, E_{1^0}) = \frac{1}{(4\pi)^4} \int d\Omega_1 \int d\Omega_2 \int d\Omega_{1^0} \int d\Omega_{\sigma} \sum | X_{12}^e \div X_{12}^P \div (X_{21}^e \div X_{21}^P) |^2. \quad (19)$$

Hier bedeutet  $\sum$  Summation über die zwei möglichen Spin-zustände der beiden Elektronen, des Positrons und des Neutrinos, während  $X_{21}^e$  und  $X_{21}^P$  aus  $X_{12}^e$  und  $X_{12}^P$  durch Vertauschung von 1 und 2 hervorgehen.

Mit Hilfe von (2) kann man schliesslich (18) in der folgenden Form schreiben

$$P(E_1, E_2, E^+) = \left( \frac{e^2}{\hbar c} \right)^2 \frac{4}{\pi^2} p_1 E_1 p_2 E_2 p_{1^0} E^+ \frac{W(E)}{E \cdot (E^2 \div m^2 c^4)^{1/2}} \cdot f(E_1, E_2, E_{1^0}). \quad (20)$$

Die Formel (20) gilt unverändert, wenn man statt des Fermischen Wechselwirkungsansatzes den Ansatz von KONOPINSKY und UHLENBECK<sup>1</sup> benutzt, nur hat man dann für

<sup>1</sup> E. J. KONOPINSKY und G. E. UHLENBECK, Phys. Rev. 48, 7 (1935).

$W(E)$  den entsprechenden Ausdruck einzusetzen. Vielleicht gilt (20) sogar unabhängig von speziellen Annahmen über den Mechanismus des  $\beta$ -Zerfalls.

## § 2.

Bevor wir zu einer näheren Auswertung von (20) übergehen, wollen wir jetzt zeigen, dass diese Formel sich auch mit Hilfe der von TISZA benutzten Methode ableiten lässt. Da in der Arbeit von TISZA die allgemeine Formel nicht angegeben ist, indem er schon bei der Herleitung seiner Endformeln gewisse Approximationen einführt, müssen wir hier erst kurz die Ableitung der allgemeinen Formel mit Hilfe dieser Methode skizzieren.

Nach TISZA beschreibt man das ausgehende  $\beta$ -Teilchen mit der Energie  $E$  durch eine passende Mischung der folgenden vier singulären Lösungen der DIRAC-Gleichung, welche alle auslaufende Kugelwellen darstellen.

$$\psi_{a_1}^+ = N_{a_1} \cdot \begin{cases} \frac{i\hbar c}{E + mc^2} \left( \frac{\partial}{\partial x} \div i \frac{\partial}{\partial y} \right) U \\ \div \frac{i\hbar c}{E + mc^2} \frac{\partial}{\partial z} U \\ 0 \\ U \end{cases} \quad \psi_{a_1}^- = N_{a_1} \cdot \begin{cases} \frac{i\hbar c}{E + mc^2} \frac{\partial}{\partial z} U \\ \frac{i\hbar c}{E + mc^2} \left( \frac{\partial}{\partial x} + i \frac{\partial}{\partial y} \right) U \\ U \\ 0 \end{cases} \quad (21)$$

$$\psi_{a_2}^+ = N_{a_2} \cdot \begin{cases} 0 \\ U \\ \frac{i\hbar c}{E \div mc^2} \left( \frac{\partial}{\partial x} \div i \frac{\partial}{\partial y} \right) U \\ \div \frac{i\hbar c}{E \div mc^2} \frac{\partial}{\partial z} U \end{cases} \quad \psi_{a_2}^- = N_{a_2} \cdot \begin{cases} U \\ 0 \\ \frac{i\hbar c}{E \div mc^2} \frac{\partial}{\partial z} U \\ \frac{i\hbar c}{E \div mc^2} \left( \frac{\partial}{\partial x} + i \frac{\partial}{\partial y} \right) U \end{cases} \quad (22)$$

$$U = \frac{e^{\frac{i}{\hbar c} p r}}{r} \quad N_a = \left( \frac{E + \varepsilon_a m c^2}{8 \pi p c} \right)^{1/2} \quad p = (E^2 - m^2 c^4)^{1/2}$$

$$\varepsilon_a = \begin{cases} +1 & \text{für } a = a_1 \\ -1 & \text{für } a = a_2 \end{cases}$$

Die Lösungen  $a_1$  und  $a_2$  entsprechen den Fällen  $l = 0$ ,  $j = \frac{1}{2}$  bzw.  $l = 1$ ,  $j = \frac{1}{2}$  und sind so normiert, dass der gesamte Strom durch eine grosse Kugel gleich eins ist<sup>1</sup>.

Wir berechnen jetzt die Wahrscheinlichkeit dafür, dass das  $\beta$ -Teilchen von einem der vier Zustände  $a$  in einen Zustand 2 während gleichzeitig ein Elektron in dem negativen Zustand 1<sup>o</sup> in den positiven Zustand 1 übergeht. Um diese Übergangswahrscheinlichkeit zu erhalten, bilden wir die dem Übergang  $a \rightarrow 2$  entsprechende Strom- und Ladungsdichte

$$\div e \cdot \bar{\psi}_2 \psi_a \cdot e^{\frac{i}{\hbar c} (E - E_2) t}, \quad ce \cdot \bar{\psi}_2 \alpha \psi_a \cdot e^{\frac{i}{\hbar c} (E - E_2) t}, \quad (23)$$

wo  $\psi_2$  durch (9) gegeben ist. Wie TISZA bemerkt hat, erfüllen die Ausdrücke (23) die Kontinuitätsgleichung nur ausserhalb des singulären Nullpunktes, was damit zusammenhängt, dass  $\psi_a$  nur das auslaufende  $\beta$ -Teilchen und nicht das im Kern entstehende Proton beschreibt. Um die Kontinuitätsgleichung zu erfüllen, müssen wir daher in (23) ein Glied hinzufügen, das nur im Nullpunkt von Null verschieden ist und der entstehenden Protonenladung Rechnung trägt. Wie in § 1 können wir auch hier die Bewegung des Protons vernachlässigen, so dass in (23) nur die La-

<sup>1</sup> Vergl. KNIPP und UHLENBECK, loc. cit. p. 427. (21) stimmen, abgesehen von triviellen Phasenfaktoren, mit den Lösungen von K. und U. überein.



dungsdichte zu modifizieren ist. Nennen wir die zusätzliche Ladungsdichte  $\varrho'_{2,a} \cdot e \frac{i}{\hbar c} (E \div E_2)t$ , so lautet die Kontinuitätsgleichung

$$\frac{i}{\hbar c} (E \div E_2) [\div e \cdot \bar{\psi}_2 \psi_a + \varrho'_{2,a}] + \operatorname{div} ce \cdot \bar{\psi}_2 \alpha \psi_a = 0. \quad (24)$$

Wenn man diese Gleichung über eine Kugel mit Radius  $r$  und Zentrum im singulären Punkt integriert und dann  $r$  gegen Null gehen lässt, erhält man mit Hilfe des GAUSS'schen Satzes

$$\varrho'_{2,a} = \frac{\hbar c \delta(\boldsymbol{x})}{i(E \div E_2)} \lim_{r \rightarrow 0} \int ce \cdot \bar{\psi}_2 \alpha_n \psi_a \cdot d\sigma, \quad (25)$$

wo  $\delta(\boldsymbol{x}) = \delta(x) \cdot \delta(y) \cdot \delta(z)$  und  $\delta(x)$  die DIRAC'sche Deltafunktion ist. Das Integral in (25) ist über die ganze Oberfläche der Kugel zu erstrecken, und  $\alpha_n$  bedeutet die Komponente von  $\alpha$  nach der äusseren Normale. Durch Einführung der Ausdrücke (9) und (21) für  $\psi_2$  und  $\psi_a$  in (25) erhält man durch elementare Rechnung

$$\varrho'_{2,a} = \div e \frac{(hc)^2 N_a \delta(\boldsymbol{x})}{\pi V^{1/2} (E \div E_2) (E + \varepsilon_a mc^2)} \cdot \bar{u}_2 u_a, \quad (26)$$

wo die Grössen  $u_{a_1}^+$ ,  $u_{a_1}^\div$  bzw.  $u_{a_2}^+$ ,  $u_{a_2}^\div$  die Amplituden der vier verschiedenen ebenen Wellen mit dem Impuls Null und der Energie  $+mc^2$  bzw.  $\div mc^2$  darstellen, also

$$u_{a_1}^+ = \begin{Bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{Bmatrix} \quad u_{a_1}^\div = \begin{Bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{Bmatrix} \quad u_{a_2}^+ = \begin{Bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{Bmatrix} \quad u_{a_2}^\div = \begin{Bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{Bmatrix} \quad (27)$$

Die erwähnte Übergangswahrscheinlichkeit ist nun gleich

$$\frac{2\pi}{\hbar} \delta(E \div E_2 + E_1^\circ \div E_1) \left| (1 | \div e \Phi_{2,a} \div e(\alpha, \mathbf{A}_{2,a}) | 1^\circ) + \right. \\ \left. \div (\text{dasselbe mit 1 und 2 vertauscht}) \right|^2. \quad (28)$$

Hier sind  $\Phi_{2,a}$  und  $\mathbf{A}_{2,a}$  die zur  $t = 0$  genommenen retardierten Potentiale, welche der Ladungsdichte  $\rho_{2,a} = \div e \cdot \bar{\psi}_2 \psi_a + \dot{\rho}'_{2,a}$  und der Stromdichte  $ec \cdot \bar{\psi}_2 \alpha \psi_a$  entsprechen.

Zum Beispiel erhält man unter Benutzung von (5)

$$(1 | \Phi_{2,a} | 1^\circ) = \frac{\bar{u}_1 u_1^\circ}{V} \cdot \int d\mathbf{x} \rho_{2,a}(\mathbf{x}) \int d\mathbf{x}' \frac{e^{\frac{i}{\hbar c}(E \div E_2) | \mathbf{x} \div \mathbf{x}' |}}{|\mathbf{x} \div \mathbf{x}'|} \cdot \\ \cdot e^{\frac{i}{\hbar c}(\mathbf{p}_1^\circ \div \mathbf{p}_1, \mathbf{x}')} = \frac{\bar{u}_1 u_1^\circ}{V} \cdot \int d\mathbf{x} \left[ \frac{\div e \cdot \bar{u}_2 \psi_a \cdot e^{\frac{i}{\hbar c}(\mathbf{p}_2, \mathbf{x})}}{V^{1/2}} + \right. \\ \left. + \frac{\div e h^2 c^2 N_a \delta(\mathbf{x}) \cdot \bar{u}_2 u_a}{\pi V^{1/2} (E \div E_2) (E + \varepsilon_a mc^2)} \right] \frac{(hc)^2}{\pi} \frac{e^{\frac{i}{\hbar c}(\mathbf{p}_1^\circ \div \mathbf{p}_1, \mathbf{x})}}{|\mathbf{p}_1^\circ \div \mathbf{p}_1|^2 \div (E \div E_2)^2}. \quad (29)$$

In derselben Weise erhält man

$$(1 | \mathbf{A}_{2,a} | 1^\circ) = \frac{\bar{u}_1 u_1^\circ}{V} \cdot \int d\mathbf{x} ce \cdot \bar{u}_2 \alpha \psi_a \cdot e^{\frac{i}{\hbar c}(\mathbf{p}_2, \mathbf{x})} \cdot \\ \cdot \frac{(hc)^2}{\pi} \frac{e^{\frac{i}{\hbar c}(\mathbf{p}_1^\circ \div \mathbf{p}_1, \mathbf{x})}}{|\mathbf{p}_1^\circ \div \mathbf{p}_1|^2 \div (E \div E_2)^2}. \quad (30)$$

Weiter erhält man leicht mit Hilfe von (11), (14) und (21)

$$\int \psi_a e^{\frac{i}{\hbar c}(\mathbf{p}_1^\circ \div \mathbf{p}_1 \div \mathbf{p}_2, \mathbf{x})} d\mathbf{x} = \div \frac{(hc)^2}{\pi I} \cdot N_a w_a, \quad (31)$$

wo die Spinoren  $w_a$  gegeben sind durch

$$\left. \begin{aligned}
 w_{a_1}^+ &= \begin{pmatrix} \frac{p'_x \div ip'_y}{E + mc^2} \\ \frac{p'_z}{E + mc^2} \\ 0 \\ 1 \end{pmatrix} & w_{a_1}^- &= \begin{pmatrix} \frac{p'_z}{E + mc^2} \\ \frac{p'_x + ip'_y}{E + mc^2} \\ 1 \\ 0 \end{pmatrix} \\
 w_{a_2}^+ &= \begin{pmatrix} 0 \\ 1 \\ \frac{p'_x \div ip'_y}{E \div mc^2} \\ \frac{p'_z}{E \div mc^2} \end{pmatrix} & w_{a_2}^- &= \begin{pmatrix} 1 \\ 0 \\ \frac{p'_z}{E \div mc^2} \\ \frac{p'_x + ip'_y}{E \div mc^2} \end{pmatrix}
 \end{aligned} \right\} (32)$$

Also erhält man unter Benutzung von (14), (22), (29), (30), (31) und des Energiesatzes  $E \div E_2 = E_1 \div E_1^\circ$ , der durch die Deltafunktion in (28) gesichert ist,

$$(1 \div e \Phi_{2,a} \div e(\alpha, A_{2,a}) | 1^\circ) = \frac{(hc)^4 e^2}{\pi^2 V^{3/2}} \cdot \frac{1}{(8\pi pc)^{1/2} (E + \varepsilon_a mc^2)^{1/2}} \cdot \left\{ \frac{E + \varepsilon_a mc^2}{I_1 \cdot I} [\bar{u}_1 u_{1^\circ} \cdot \bar{u}_2 w_a \div (\bar{u}_1 \alpha u_{1^\circ}, \bar{u}_2 \alpha w_a)] \div \frac{\bar{u}_1 u_{1^\circ} \cdot \bar{u}_2 u_a}{I_1 \cdot (E_1 \div E_1^\circ)} \right\}. \quad (33)$$

Durch direkte Ausrechnung findet man leicht, dass

$$(E + \varepsilon_a mc^2) w_a = (E + H') u_a \quad (34)$$

wo  $H'$  und  $u_a$  durch (13) und (27) gegeben sind. Wir erhalten jetzt die Wahrscheinlichkeit,  $\Phi_a$ , einer Paarerzeugung bei dem die Energie des Elektrons bzw. des Positrons zwischen  $E_1$  und  $E_1 + dE_1$  bzw.  $E^+ = \div E_1^\circ$  und  $E^+ + dE^+$  liegt durch Multiplikation von (28) mit den Zustandsdichten

$$\frac{V p_1 E_1 dE_1 d\Omega_1}{(hc)^3}, \quad \frac{V p_2 E_2 dE_2 d\Omega_2}{(hc)^3}, \quad \frac{V p_{1^0} E^+ dE^+ d\Omega_{1^0}}{(hc)^3}$$

und Integration über  $dE_2$ ,  $d\Omega_1$ ,  $d\Omega_2$  und  $d\Omega_{1^0}$  sowie Summation über die beiden Spinrichtungen der beteiligten Teilchen. Es ergibt sich mit Hilfe von (34)

$$\Phi_a dE_1 dE^+ = \left(\frac{e^2}{\hbar c}\right)^2 \frac{4}{\pi^2} \frac{p_1 E_1 p_2 E_2 p_{1^0} E^+ dE_1 dE^+}{p(E + \varepsilon_a mc^2)} \frac{2}{(4\pi)^3} \cdot \int d\Omega_1 \int d\Omega_2 \int d\Omega_{1^0} \sum_{s_1, s_2, s_{1^0}} |A u_a|^2, \quad (35)$$

wo  $E_2$  hier überall gleich  $E \div E_1 \div E^+$  zu setzen ist, und der Spinor  $A$  gegeben ist durch

$$A = \frac{[\bar{u}_1 u_1 \cdot \bar{u}_2 \div (\bar{u}_1 \alpha u_{1^0}, \bar{u}_2 \alpha)] (E + H')}{I_1 \cdot I} \div \frac{\bar{u}_1 u_1 \cdot \bar{u}_2}{I_1 \cdot (E_1 \div E_{1^0})} \quad (36)$$

$\div$  (dasselbe mit 1 und 2 vertauscht).

Wir müssen jetzt die  $\Phi_a$ 's in denjenigen Verhältnissen kombinieren, in welchen die  $\psi_a$ 's in den  $\beta$ -Strahlen vertreten sind. Wir bilden also

$$\Phi = \sum_a g_a \Phi_a, \quad \sum_a g_a = 1 \quad (37)$$

und entnehmen die Gewichte  $g_a$  aus der Fermi'schen Theorie. Nach KNIPP und UHLENBECH<sup>1</sup> ist

$$\left. \begin{aligned} g_{a_1}^+ &= g_{a_1}^- = \frac{E + mc^2}{4E}, & g_{a_2}^+ &= g_{a_2}^- = \frac{E \div mc^2}{4E} \\ \text{oder} & & g_a &= \frac{E + \varepsilon_a mc^2}{4E} \end{aligned} \right\} \quad (38)$$

<sup>1</sup> loc. cit., p. 430.

Setzt man (35) und (38) in (37) ein, hebt sich der Faktor  $E + \varepsilon_a mc^2$  weg, so dass die Summation über  $a$  in (37) sich auf eine Summation des Ausdrucks  $|Au_a|^2$  über die vier verschiedenen Zustände  $u_a$  mit dem Impuls Null reduziert. Mit Hilfe der Vollständigkeitsrelation erhält man

$$\sum_a |Au_a|^2 = \bar{A}A. \quad (39)$$

Multiplizieren wir endlich  $\Phi$  in (37) mit dem Energiespektrum  $W(E)$  der  $\beta$ -Strahlen, erhalten wir schliesslich unter Benutzung von (22) die Grösse

$$P(E_1, E_2, E^+) = \left(\frac{e^2}{\hbar c}\right)^2 \frac{4}{\pi^2} \frac{p_1 E_1 p_2 E_2 p_1^+ E^+}{E (E^2 \div m^2 c^4)^{1/2}} \frac{W(E)}{(4\pi)^3} \cdot \int d\Omega_1 \int d\Omega_2 \int d\Omega_{1^0} \sum_{s_1, s_2, s_{1^0}} \frac{1}{2} \bar{A}A. \quad (40)$$

was, wie man leicht sieht, mit (20) identisch ist. Wenn man nämlich die Ausdrücke für  $X_{12}^e$  usw. in (19) einsetzt, bemerkt man sofort, dass sich  $f$  in (20) in der folgenden Form schreiben lässt

$$f = \frac{1}{(4\pi)^3} \int d\Omega_1 \int d\Omega_2 \int d\Omega_{1^0} \sum_{s_1, s_2, s_{1^0}} \frac{1}{4\pi} \int d\Omega_\sigma \sum_{s_\sigma} |A \delta \bar{u}_\sigma|^2.$$

Führt man jetzt die Summation über die beiden Spinrichtungen und die Mittelung über alle Raumrichtungen des Neutrinos aus, erhält man

$$f = \frac{1}{(4\pi)^3} \int d\Omega_1 \int d\Omega_2 \int d\Omega_{1^0} \sum_{s_1, s_2, s_{1^0}} \frac{1}{2} \bar{A}A,$$

wodurch (20) und (40) identisch werden.

## § 3.

In den erwähnten Experimenten von ALICHANOW usw. misst man direkt die Energieverteilung der Positronen sowie das Verhältnis  $\frac{N^+}{N^-}$  der totalen Anzahlen der emittierten Positronen und Elektronen. Die aus unserer Theorie folgende Energieverteilung  $W^+(E^+)$  der Positronen erhält man aus (20) durch Integration über die möglichen Werte von  $E_1$  und  $E_2$ , d. h.

$$W^+(E^+) = \int_{mc^2}^{D \div E^+ + mc^2} dE_2 \int_{mc^2}^{D \div E^+ + E_2} dE_1 P(E_1, E_2, E^+). \quad (41)$$

Die Integrationsgrenzen folgen aus der Bedingung, dass die vom Kern abgegebene Energie  $D$  zwischen den vier emittierten Teilchen zu verteilen ist; hierbei ist, wie überall, die Neutrinomasse gleich Null gesetzt. Man sieht, dass das Positronenspektrum sich über das Intervall

$$mc^2 \leq E^+ \leq D \div 2 mc^2 \quad (42)$$

erstreckt.

Für  $\frac{N^+}{N^-}$  erhält man also

$$\frac{N^+}{N^-} = \frac{\int_{mc^2}^{D \div 2 mc^2} W^+(E^+) dE^+}{\int_{mc^2}^D W(E) dE}. \quad (43)$$

Bevor wir aber (41) und (43) numerisch auswerten können, müssen wir erst die Spinsumationen in (19) ausführen. Durch Ausquadrieren von  $|X_{12}^e \div X_{12}^p \div X_{21}^e + X_{21}^p|^2$  erhält man offenbar sechzehn Glieder. Die entsprechenden sech-

zehn Spinsummationen können wieder mit Hilfe der CASIMIR'schen<sup>1</sup> Methode ausgeführt werden. Die Rechnungen sind elementar aber langwierig. Wir haben in den Schlussformeln die folgenden Symbole für die auftretenden Vierervektoren und deren Verknüpfungen eingeführt:

$$\left. \begin{aligned} P_1 &= \{ \boldsymbol{p}_1, E_1 \}, P_2 = \{ \boldsymbol{p}_2, E_2 \}, P_{1^\circ} = \{ \boldsymbol{p}_{1^\circ}, E_{1^\circ} \}, \\ P' &= \{ \boldsymbol{p}', E \} = P_1 + P_2 \div P_{1^\circ} \\ (P_1, P_2) &= E_1 E_2 \div (\boldsymbol{p}_1, \boldsymbol{p}_2), (P', P_{1^\circ}) = E E_{1^\circ} \div (\boldsymbol{p}', \boldsymbol{p}_{1^\circ}) \text{ u. s. w.} \\ [P_1, P_2] &= E_1 E_2 + (\boldsymbol{p}_1, \boldsymbol{p}_2), [P', P_{1^\circ}] = E E_{1^\circ} + (\boldsymbol{p}', \boldsymbol{p}_{1^\circ}) \text{ u. s. w.} \end{aligned} \right\} (44)$$

Es gilt offenbar

$$\begin{aligned} (P_1, P_1) &= E_1^2 \div \boldsymbol{p}_1^2 = m^2 c^4 = (P_2, P_2) = (P_{1^\circ}, P_{1^\circ}) \\ [P_1, P_1] &= E_1^2 + \boldsymbol{p}_1^2 = 2 E_1^2 \div m^2 c^4 \text{ u. s. w.} \end{aligned}$$

Dagegen ist

$$(P', P') \geq 9 m^2 c^4.$$

Die runden Klammern sind relativistisch invariant im Gegensatz zu den eckigen Klammern. Die letzteren treten nur in denjenigen Gliedern auf, welche von der Protonenwechselwirkung herrühren, was damit zusammenhängt, dass die schweren Teilchen von Anfang an unrelativistisch behandelt worden sind. Unter Benutzung von (44) erhält man dann für die sechzehn erwähnten Glieder, nachdem schon die Integration über die Raumrichtungen des Neutrinos ausgeführt ist:

<sup>1</sup> loc. cit.

$$\begin{aligned}
& \frac{1}{4\pi} \int d\Omega_\sigma \sum |X_{12}^e|^2 = \\
& \quad \frac{4}{I^2 \cdot I_1^2 \cdot E_1 E_2 E_{1^\circ}} \left\{ [m^2 c^4 + (P_1, P_2) \div (P_1, P_{1^\circ}) \div (P_2, P_{1^\circ})] \cdot \right. \\
& \quad \cdot [m^2 c^4 E_2 \div E_{1^\circ} (P_1, P_2) \div E_1 (P_{1^\circ}, P_2)] + \\
& \quad \div E [2 m^2 c^4 [(P_1, P_2) \div (P_2, P_{1^\circ})] + m^2 c^4 [(P_1, P_{1^\circ}) \div m^2 c^4] + \\
& \quad \left. \div (P_1, P_2) [(P_{1^\circ}, P_1) + (P_{1^\circ}, P_2)] \div (P_{1^\circ}, P_2) [(P_1, P_2) \div (P_1, P_{1^\circ})] \right\} \\
& \frac{1}{4\pi} \int d\Omega_\sigma \sum |X_{21}^e|^2 = \text{(dasselbe mit 1 und 2 vertauscht).} \\
& \frac{1}{4\pi} \int d\Omega_\sigma \sum \overline{X_{21}^e} X_{12}^e = \frac{1}{4\pi} \int d\Omega_\sigma \sum \overline{X_{12}^e} X_{21}^e = \\
& \quad \frac{1}{I^2 \cdot I_1 \cdot I_2 \cdot E_1 E_2 E_{1^\circ}} \cdot \left\{ 4 m^2 c^4 E (P_1, P_2) + \right. \\
& \quad \div 4 E (P', P_{1^\circ}) [(P_1, P_2) \div m^2 c^4] + \\
& \quad \left. + 4 E_{1^\circ} [m^2 c^4 + (P_1, P_2) \div (P_1, P_{1^\circ}) \div (P_2, P_{1^\circ})] [(P_1, P_2) \div m^2 c^4] \right\}.
\end{aligned} \tag{45}$$

Diese vier Glieder enthalten nur die Wechselwirkung zwischen den Elektronen im Gegensatz zu den jetzt folgenden zwölf Gliedern, welche auch von der Wechselwirkung zwischen Proton und Elektron abhängen. Vernachlässigung der letzteren Wechselwirkung bedeutet also, dass nur die Glieder (45) in (19) übrigbleiben.

$$\begin{aligned}
& \frac{1}{4\pi} \int d\Omega_\sigma \sum |X_{12}^P|^2 = \\
& \quad \frac{1}{I_1^2 (E_1 \div E_{1^\circ})^2 E_1 E_2 E_{1^\circ}} \cdot \left\{ E_2 [(P_1, P_{1^\circ}) + m^2 c^4] \right\} \\
& \frac{1}{4\pi} \int d\Omega_\sigma \sum |X_{21}^P|^2 = \text{(dasselbe mit 1 und 2 vertauscht).}
\end{aligned} \tag{46}$$



$$\left. \begin{aligned} \frac{1}{4\pi} \int d\Omega_\sigma \sum \bar{X}_{12}^P X_{21}^P &= \frac{1}{4\pi} \int d\Omega_\sigma \sum \bar{X}_{21}^P X_{12}^P = \\ & \frac{1}{4 I_1 I_2 (E_1 \div E_1^\circ) (E_2 \div E_1^\circ) E_1 E_2 E_1^\circ} \cdot \left\{ E_1^\circ ([P_1, P_2] + m^2 c^4) + \right. \\ & \left. + E_1 ([P_1^\circ, P_2] + m^2 c^4) + E_2 ([P_1^\circ, P_1] + m^2 c^4) \div 2 E_1 E_2 E_1^\circ \right\} \end{aligned} \right\} (46)$$

$$\left. \begin{aligned} \frac{1}{4\pi} \int d\Omega_\sigma \sum \bar{X}_{12}^e X_{12}^P &= \frac{1}{4\pi} \int d\Omega_\sigma \sum \bar{X}_{12}^P X_{12}^e = \\ & \frac{1}{I \cdot I_1^2 \cdot (E_1 \div E_1^\circ) E_1 E_2 E_1^\circ} \cdot \left\{ ([P_1, P_1^\circ] + m^2 c^4) ([P', P_2] + m^2 c^4) + \right. \\ & \left. \div E_1^\circ E_2 [P', P_1] \div E_1 E_2 [P', P_1^\circ] + E E_1^\circ (P_1, P_2) + E E_1 (P_2, P_1^\circ) \right\} \\ \\ \frac{1}{4\pi} \int d\Omega_\sigma \sum \bar{X}_{21}^e X_{21}^P &= \frac{1}{4\pi} \int d\Omega_\sigma \sum \bar{X}_{21}^P X_{21}^e = \\ & \text{(dasselbe mit 1 und 2 vertauscht)} \\ \\ \frac{1}{4\pi} \int d\Omega_\sigma \sum \bar{X}_{21}^e X_{12}^P &= \frac{1}{4\pi} \int d\Omega_\sigma \sum \bar{X}_{12}^P X_{21}^e = \\ & \frac{1}{2 I \cdot I_1 \cdot I_2 \cdot (E_1 \div E_1^\circ) E_1 E_2 E_1^\circ} \cdot \left\{ \div ([P', P_2] + m^2 c^4) \cdot \right. \\ & \quad \cdot ([P_1^\circ, P_1] + m^2 c^4) + [P_2, P_1^\circ] [P', P_1] + \\ & \quad \left. \div (P_1, P_2) (P', P_1^\circ) + 2 m^2 c^4 (E E_1 + E_1^\circ E + E_2 E_1^\circ + E_1 E_2) \right\} \\ \\ \frac{1}{4\pi} \int d\Omega_\sigma \sum \bar{X}_{12}^e X_{21}^P &= \frac{1}{4\pi} \int d\Omega_\sigma \sum \bar{X}_{21}^P X_{12}^e = \\ & \text{(dasselbe mit 1 und 2 vertauscht)}. \end{aligned} \right\} (47)$$

Um jetzt  $f(E_1, E_2, E_1^\circ)$  zu erhalten, müssen wir noch die Ausdrücke (45) bis (47) über  $d\Omega_1$ ,  $d\Omega_2$  und  $d\Omega_1^\circ$  integrieren. Diese Integrationen sind jedoch analytisch schwer durchführbar, und da uns ausserdem die zwei Energie-

integrationen in (41) fehlen, sind wir bei der Durchführung der Rechnungen auf numerische Methoden angewiesen. Da aber eine mehrfache numerische Integration eine sehr erhebliche Rechenarbeit erfordert, und da es hier auf grössere Genauigkeit nicht ankommt, machen wir die folgende Approximation, deren Berechtigung wir später diskutieren werden (vergl. p. 27 ff.). Wir setzen in dem Ausdruck  $E_1 E_2 E^+ \cdot f(E_1, E_2, E_1^\circ)$

$$E_1 = E_2 = \div E_1^\circ = mc^2,$$

wodurch der Integrand in (19) eine Konstante wird. Die Raumintegrationen ergeben dann einfach einen Faktor  $(4\pi)^3$ , und man erhält

$$E_1 E_2 E^+ \cdot f = \frac{9}{128} (mc^2)^{\div 3}. \quad (48)$$

Indem wir von jetzt an alle Energien in Einheiten  $mc^2$  ausdrücken, erhalten wir für (41) unter Berücksichtigung von (2), (20), (48) und  $\frac{e^2}{\hbar c} = \frac{1}{137}$

$$\begin{aligned} W^+(E^+) &= 2.16 \times 10^{\div 5} \cdot K \cdot (E^{+2} \div 1)^{1/2} \int_1^{D \div E^+ \div 1} dE_2 \\ &\int_1^{D \div E^+ \div E_2} dE_1 (E_2^2 \div 1)^{1/2} (E_1^2 \div 1)^{1/2} (D \div E_1 \div E_2 \div E^+)^2 \frac{9}{128} = \quad (49) \\ &= 2.16 \times 10^{\div 5} \cdot K \cdot (E^{+2} \div 1)^{1/2} \cdot G_1(E^+) \end{aligned}$$

mit

$$K = g^2 \frac{2\pi}{\hbar} \frac{(4\pi)^2}{(hc)^6} |(P|1|N)|^2 (mc^2)^5. \quad (50)$$

Ebenso wird (2)

$$W(E) = K \cdot (E^2 \div 1)^{1/2} \cdot E \cdot (D \div E)^2. \quad (51)$$

Die erste Integration in (49) wurde analytisch ausgeführt, die zweite dagegen numerisch. Man erhält z. B. für  $D = 7.2$  (Ra C, obere Grenze des  $\beta$ -Spektrums bei 3.16 M.V.) die Kurve in Fig. 1.

Für andere Werte von  $D$  erhält man Kurven mit genau derselben Form. Vor allem bemerkt man, dass das Maxi-

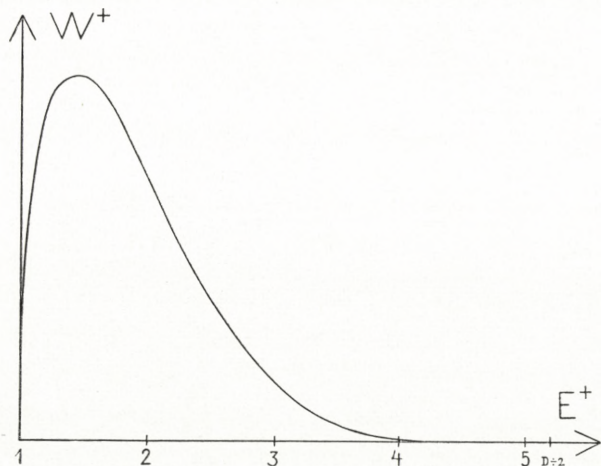


Fig. 1.

um viel weiter links liegt als bei den gewöhnlichen  $\beta$ -Spektren, indem die Kurve nach rechts einen sehr flachen Schwanz hat. Wie wir später sehen werden, ist dies Verhalten nicht durch die angewandten Approximationen vorgetauscht, indem auch die exakte Kurve dieselbe Form zeigt (siehe p. 29).

Jetzt können wir die Integrale in (43) erhalten, indem sich  $N^+$  durch numerische Integration der berechneten Positronenspektren ergibt. In Fig. 2 ist  $\frac{N^+}{N^-}$  als Funktion von  $D$  graphisch dargestellt.

In derselben Figur beträgt der entsprechende Wert für den Fall der KONOPINSKY-UHLENBECK Wechselwirkung 0.23 für  $D = 7.2$ .

Man sieht also, dass der theoretische Wert von  $\frac{N^+}{N^\pm}$  auch im besten Fall ( $D = 7.2$ , RaC) um einen Faktor  $10^3$  kleiner ist als die von ALICHANOW usw. gefundenen Werte (ca.  $2-3 \times 10^{\pm 4}$ ). Ferner bemerkt man, dass  $\frac{N^+}{N^\pm}$  theoretisch sehr

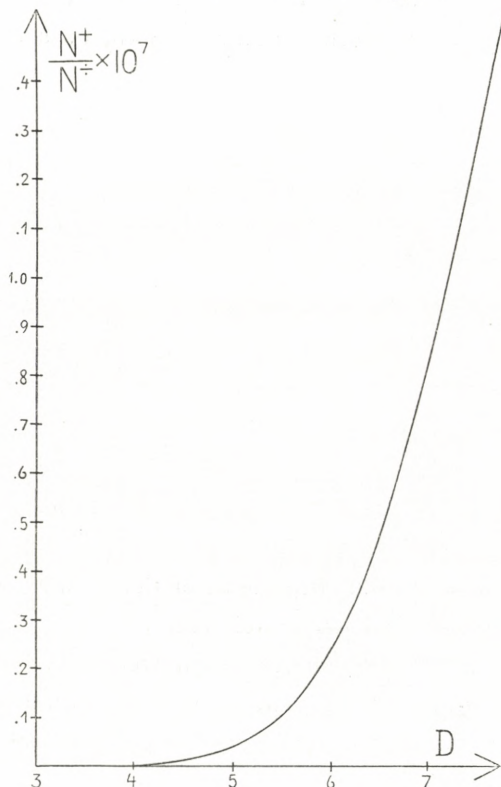


Fig. 2.

schnell mit  $D$  anwächst, während die Experimente für die Elemente RaC,  $D = 7.2$ , und Th ( $C + C''$ ),  $D = 5.4$ , ungefähr denselben Wert liefern. Die gefundenen Positronen haben daher vermutlich einen ganz anderen Ursprung.<sup>1</sup>

<sup>1</sup> Vergl. neue Versuche von A. I. ALICHANOW und P. E. SPIWAK. Sow. Phys. 11, 351 (1937).

Da jedoch  $\frac{N^+}{N^\div}$ , wie Fig. 2 zeigt, theoretisch sehr schnell mit  $D$  anwächst, wäre es interessant, andere  $\beta$ -Strahlen mit sehr hohen oberen Grenzen des Spektrums nach Positronen zu untersuchen. Zum Beispiel sollte der hier betrachtete Effekt bei  $Li^8$  ( $D = 20 mc^2 \approx 10$  M.V.) sehr wohl beobachtbar sein, indem  $\frac{N^+}{N^\div}$  hier von der Grössenordnung  $10^{\div 4}$  ist.

Man könnte vielleicht glauben, dass der kleine Wert von  $\frac{N^+}{N^\div}$  durch die in (48) gemachte Approximation verursacht ist. Das ist jedoch nicht der Fall. Wenn man nämlich die Funktion (48) nach oben ziemlich grob abschätzt, findet man, dass  $E_1 E_2 E^+ \cdot f$  für  $D = 7.2$  bzw.  $D = 5.0$  sicher für alle Werte von  $E_1, E_2, E^+$ , mit  $3 \leq E_1 + E_2 + E^+ \leq D$ , bedeutend kleiner ist als  $12.8 (mc^2)^{\div 3}$  bzw.  $3.03 (mc^2)^{\div 3}$ . Benutzt man diese Werte statt (48) werden die entsprechenden Werte von  $\frac{N^+}{N^\div}$  gleich  $1.8 \times 10^{\div 5}$  bzw.  $2.0 \times 10^{\div 7}$ , was noch immer wenigstens um einen Faktor 10 zu klein ist.

Zum Schluss müssen wir noch kurz die Berechtigung der in (48) gemachten Approximation diskutieren. Wir haben also die angenäherte Funktion (49) mit der folgenden exakten zu vergleichen

$$\begin{aligned}
 W^+(E^+) &= 2.16 \times 10^{\div 5} \cdot K \cdot (E^{+\div 2} \div 1)^{1/2} \int_1^{D \div E^+ \div 1} dE_2 \int_1^{D \div E^+ \div E_2} dE_1 \cdot \\
 (E_2^2 \div 1)^{1/2} (E_1^2 \div 1)^{1/2} (D \div E_1 \div E_2 \div E^+)^2 \cdot E_1 E_2 E^+ \cdot f(E_1, E_2, E_1^+) &= \\
 &= 2.16 \times 10^{\div 5} \cdot K \cdot (E^{+\div 2} \div 1)^{1/2} \cdot G(E^+),
 \end{aligned} \tag{52}$$

die man aus (41) erhält, wenn man (50) benutzt und wieder  $mc^2$  als Energieeinheit einführt. Die Approximationsfunktion  $G_1(E^+)$  in (49) unterscheidet sich von der exakten Funktion  $G(E^+)$  in (52) dadurch, dass man in dieser den Faktor

$E_1 E_2 E^+ \cdot f$  konstant gleich seinem Wert für  $E_1 = E_2 = E^+ = 1$  gesetzt hat. Das ist natürlich für nicht zu kleine  $D$  eine sehr grobe Approximation. Eine bedeutend bessere Approximationsfunktion, wir nennen sie  $G_2(E^+)$ , würde man erhalten, wenn man in  $E_1 E_2 E^+ \cdot f$  nur  $E^+$  von vorn herein gleich 1 setzt. Dadurch wird der Integrand in (19) soviel vereinfacht, dass die Raumintegrationen sich verhältnismässig einfach analytisch durchführen lassen.  $G_2(E^+)$  stimmt offenbar für  $E^+ = 1$  mit  $G(E^+)$  überein und bleibt eine sehr gute Annäherung, solange  $E^+ \div 1 < 1$  ist. Für  $D < 4$  ist ferner  $G_1 \approx G_2 \approx G$ . Um aber zu sehen, wieviel sich  $G_1(1)$  für grössere Werte von  $D$  von  $G_2(1) = G(1)$  unterscheidet, haben wir den Wert der letzteren für  $D = 7.2$  berechnet. Man findet  $G_2(1) = G(1) = 4.8$  während  $G_1(1)$  gleich 3.6 war. Die Abweichung beträgt also sogar für den höchsten in den Experimenten von Alichanow usw. vorkommenden Wert von  $D$  nur 25 0/0.

In dem Gebiet  $E^+ > D \div 3$  erhält man eine andere sehr gute Annäherungsfunktion,  $G_3(E^+)$ , indem man den Faktor  $E_1 E_2 E^+ \cdot f$  konstant gleich seinem Wert an der oberen Grenze des Spektrums setzt. Da dort  $E^+ = D \div 2$  und somit  $E_1 = E_2 = 1$ , erhält man leicht

$$E_1 E_2 E^+ \cdot f = \frac{D^2 (D \div 2)}{8 (D \div 1)^4}.$$

Es ist also

$$G_3(E^+) = \frac{D^2 (D \div 2)}{8 (D \div 1)^4} \cdot \frac{128}{9} \cdot G_1(E^+),$$

wo der Faktor gegen 1 geht für  $D \rightarrow 3$ .

Weiter erkennt man durch elementare Differentiation, dass die vier genannten  $G$ -Funktionen sowie ihre vier ersten Ableitungen sämtlich an der oberen Grenze  $E^+ = D \div 2$

verschwinden, sodass sie alle und damit auch die entsprechenden  $W^+$ -Kurven den p. 25 besprochenen Schwanz aufweisen. Eben aus diesem Grunde ist das Gültigkeitsgebiet der  $G_3$ -Kurve für die Berechnung der totalen Positronenanzahl  $N^+$  unwesentlich.

Aus allen diesen Gründen haben wir uns wegen der hier erforderlichen Genauigkeit mit der Berechnung der im Verhältnis zu  $G$  und  $G_2$  viel einfacheren Funktion  $G_1(E^+)$  begnügt.

### Zusammenfassung.

In der vorliegenden Arbeit wurde die innere Paarerzeugung beim  $\beta$ -Zerfall theoretisch behandelt. § 1 gibt eine Darstellung der Theorie sowie die allgemeine Formel für die Wahrscheinlichkeit dieses Prozesses. In § 2 wird gezeigt, dass die von TISZA für die Behandlung desselben Problems herangezogene Methode trotz des gänzlich verschiedenen Ausgangspunktes dieselbe Endformel liefert wie die in § 1 entwickelte Theorie. Schliesslich folgt in § 3 auf Grund der allgemeinen Formeln des § 1 die numerische Berechnung des Positronenspektrums sowie der relativen Anzahl  $\frac{N^+}{N^-}$  der emittierten Positronen und Elektronen als Funktion der oberen Grenze des  $\beta$ -Spektrums. Es zeigt sich, dass die theoretischen Werte von  $\frac{N^+}{N^-}$  viel zu klein sind um die Experimente von ALICHANOW, ALICHANIAN und KOSODAEW zu erklären. Da jedoch  $\frac{N^+}{N^-}$  sehr schnell mit der oberen Grenze des  $\beta$ -Spektrums anwächst, würde die innere Paarerzeugung bei Elementen mit höheren Werten der oberen Grenze des  $\beta$ -Spektrums

sehr wohl beobachtbar sein. Zum Beispiel ist bei  $\text{Li}^8$  mit der oberen Grenze bei 10 M.V.  $\frac{N^+}{N^-}$  von der Grössenordnung  $10^{\pm 4}$ .

Zum Schluss möchten wir Herrn Professor N. BOHR für sein freundliches Interesse an dieser Arbeit herzlichst danken.





Det Kgl. Danske Videnskabernes Selskab.

Mathematisk-fysiske Meddelelser. **XV**, 10.

---

ON THE SLOWING  
DOWN AND CAPTURE OF NEUTRONS IN  
HYDROGENOUS SUBSTANCES

BY

O. R. FRISCH, H. v. HALBAN JUN.  
AND JØRGEN KOCH



KØBENHAVN

LEVIN & MUNKSGAARD

EJNAR MUNKSGAARD

1938

Printed in Denmark.  
Bianco Lunos Rogtrykkeri A/S.

## 1. Introduction.

FERMI and his coworkers (1) have discovered that neutrons can be slowed down by surrounding their source with substances containing hydrogen such as water or paraffin wax. This discovery has been followed by a number of investigations (2) which have thrown much light on the phenomenon so that we have now a fairly complete picture of the processes involved.

Fast neutrons in a large container filled with water will collide with the protons of the water losing, on an average, half of their kinetic energy at each collision (3); they will also lose a (smaller) fraction of their energy in collisions with the oxygen nuclei (3). When the neutrons have come down to about 1 eV., energy transfer to the protons is hampered by the chemical bond forces (4); nevertheless, most of the neutrons get into thermal equilibrium ( $kT = 0.026$  eV. at room temperature) and diffuse like atoms of a dissolved substance until finally, after another 100—200 collisions, they are captured by some of the nuclei present.

We have investigated the capture of neutrons in different liquids containing hydrogen. By varying the chemical composition of the liquid we have been able to separate the effects due to the different elementary constituents and to determine the capture cross-section of the nuclei of *H*, *D*, *C*, *N*, and *O* relative to the capture cross-section of *Li*

and  $B$  which are known from absorption experiments with slow neutron beams (5). Some information on the interaction of fast neutrons with these nuclei and on the disintegration of  $D$  and  $Be$  by gamma-rays (6, 7) has also been obtained.

Furthermore, we have tried to obtain information of a qualitative character on the number of collisions required to bring a neutron through the "chemical bond region" and into thermal equilibrium. For this purpose we have studied the absorption, in different substances, of neutrons slowed down at different temperatures (8, 9, 10) and under various geometrical conditions.

## 2. Capture of neutrons in different liquids.

If a source of neutrons is placed in a large container filled with water (11), a spherically symmetrical density distribution of neutrons around the source will be established, the density  $\varrho(r)$  decreasing with increasing distance  $r$  from the source. The exact shape of the density distribution depends on a number of parameters, such as the mean free path at all the neutron energies involved, the directional distribution of scattered neutrons, and so on. The total number  $n = 4\pi \int_0^{\infty} \varrho(r) r^2 dr$  of neutrons present at any instant, however, depends only on their mean life  $\tau$  and the number  $q$  of neutrons produced per unit time:

$$n = q \cdot \tau = 4\pi \int_0^{\infty} \varrho(r) r^2 dr \quad (1)$$

Taking  $\tau$  from diffusion experiments, AMALDI and FERMI(11) and recently AMALDI, TUVE and HAFSTAD(12) have used this method for determining the number  $q$  of neutrons emitted by different sources per unit time. For

exploring the density distribution they have made use of the activity induced in rhodium.

In our experiments we have made a somewhat different use of equation (1). By comparing  $n$  with different liquids but the same source ( $q$  constant), we have obtained relative values for  $\tau$  in the liquids investigated. For exploring the density distribution we used a small boron-lined ionisation chamber which was immersed in the liquid. The number of counts  $J$  is then proportional to the neutron density  $q$ ; in the equation  $J = s \cdot q$  the factor  $s$  (the sensitivity of the chamber) is independent of the velocity of the neutrons, provided that the absorption of boron follows the " $1/v$ -law" (13, 14, 4) for all the neutron velocities involved (i. e. if its cross-section is inversely proportional to the velocity of the neutrons). As a matter of fact, we found that the number of counts is reduced by a factor of about 100 (in water) on surrounding the chamber with cadmium; this shows that about 99 per cent of the neutron density is made up of neutrons with energies below 1 eV. (13, 15), in which region we can safely adopt the  $1/v$ -law for boron. The chamber sensitivity  $s$  need not be known since we are only interested in relative measurements. A rough estimate of  $s$  can, however, be obtained from the area of the boron layer and the known range of the alpha-particles emitted under the action of neutrons.

Assuming the capture cross-sections  $\sigma_c$  of all nuclei present in the liquid to be inversely proportional to the velocity of the neutrons, the chance per unit time of their getting captured does not depend on the velocity of the neutrons. (This corresponds to the assumption that resonance capture is of no importance (4).) On introducing the "capture probability per nucleus"  $p_k = v \cdot \sigma_{ck}(v)$  (the index

$k$  referring to the different kinds of nuclei), the chance per unit time of being captured becomes  $\chi = \sum_k P_k N_k$  ( $N =$  number of nuclei per cm.<sup>3</sup>). The mean life is then

$$\tau = \frac{1}{\chi} = \frac{1}{\sum_k P_k N_k} \quad (2)$$

and the total number of neutrons present at any instant is

$$n = q \cdot \tau = \frac{q}{\sum_k P_k N_k}. \quad (3)$$

By comparing the values obtained for  $n$  in two different liquids we find

$$\frac{n_1}{n_2} = \frac{\sum_k P_k N_{k2}}{\sum_k P_k N_{k1}}. \quad (4)$$

If the two liquids are both composed of the same two constituents  $A$  and  $B$ , the relative proportion only being different, then the equation can evidently be used to determine  $\frac{P_A}{P_B} = \frac{\sigma_{cA}}{\sigma_{cB}}$  = the ratio of the capture cross-sections of the two constituents.

In this way we have determined the ratio between the capture cross-sections of  $H$  and  $C$ , by comparing measurements on benzene and liquid paraffin. We have also made measurements on water and aqueous solutions of boric acid (16, 17), lithium hydroxide, deuterium oxide, and ammonium nitrate; this enabled us to find the capture cross-sections of oxygen, boron, lithium, deuterium and nitrogen, with respect to hydrogen. Since boron and lithium show hardly any scattering (18), their capture cross-sections may be identified with their total cross-sections obtained from

absorption experiments on beams of thermal neutrons; in this way we obtained absolute values for the capture cross-sections of the light elements we have investigated.

### 3. Experimental arrangement for the capture experiments.

As a source of neutrons we used a 100 mCurie of  $\text{RaSO}_4$  mixed with about 1 gm. of thoroughly powdered beryllium contained in a platinum tube, 25 mm. long and 8 mm. wide,



Fig. 1. Small boron chamber. *a*, outer brass tube; *b*, soft glass tube providing insulation between *a* (earthed) and *c*; *c*, the high tension (100 Volt) electrode with the boron lining *d*; *e*, collecting electrode, connected to the amplifier by the thin wire *k*; *f*, grounded shield, insulated from *c* by the glass tube *j*, which is punched (*h*) to keep the insulating glass bead *g* in position.

with 0.3 mm. walls. The platinum tube was enclosed in a brass container which in turn was put in a soft glass tube (of 13 mm. outer diameter) to protect it from contact with the liquid.

The boron-lined ionisation chamber was made quite small so as to disturb the neutron distribution as little as possible, the construction of the chamber being shown by the scale drawing fig. 1. The chamber was connected to a Wynn-Williams linear amplifier and scale-of-eight counter system. The entire arrangement is shown in fig. 2.

Several times a day the sensitivity of the counting arrangement was checked with a "standard neutron density": the chamber was placed at the center of a tin can

(20 cm. high, 20 cm. diameter) filled with water and the source was placed at a distance of 6.0 cm. from the chamber. The number of counts per minute under these conditions varied between 1000 and 800, during the course of

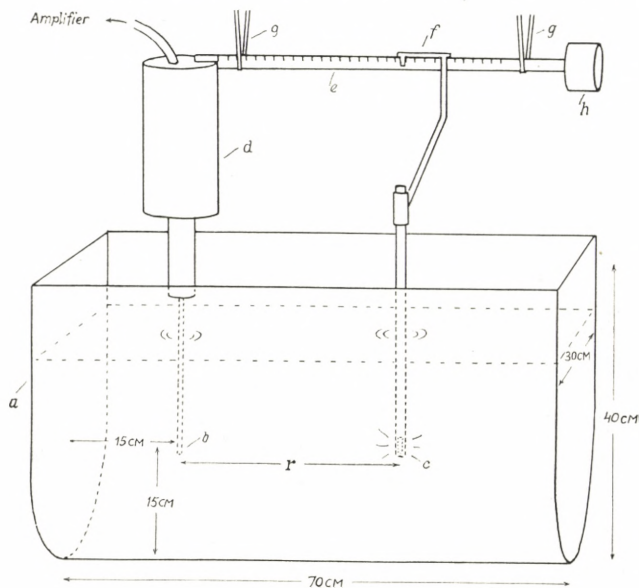


Fig. 2. Arrangement for capture experiments. *a*, container filled with 60 ltrs. liquid (galvanised iron, painted with paraffin wax inside when used for aqueous liquids); *b*, boron chamber (see fig. 1); *c*, neutron source; *d*, first amplifier stage; *e*, brass rod with scale; *f*, slider supporting source; *g*, rubber suspension; *h*, counterweight.

the investigation, but was in general constant to within a few per cent during one day.

#### 4. Results of the capture experiments.

In fig. 3, 4, 5 the function  $\rho(r)r^2$  is plotted for all the liquids investigated. The neutron densities  $\rho$  are measured in units of the "standard density" and all lengths are measured in cms.



In order to determine the areas below the curves, which represent the total number  $n$  of neutrons present at any instant in the liquid (see formula (1)), two alternative methods were used. In the "numerical method" the area of the measured part of the curve was calculated by the "trapezoidal rule"

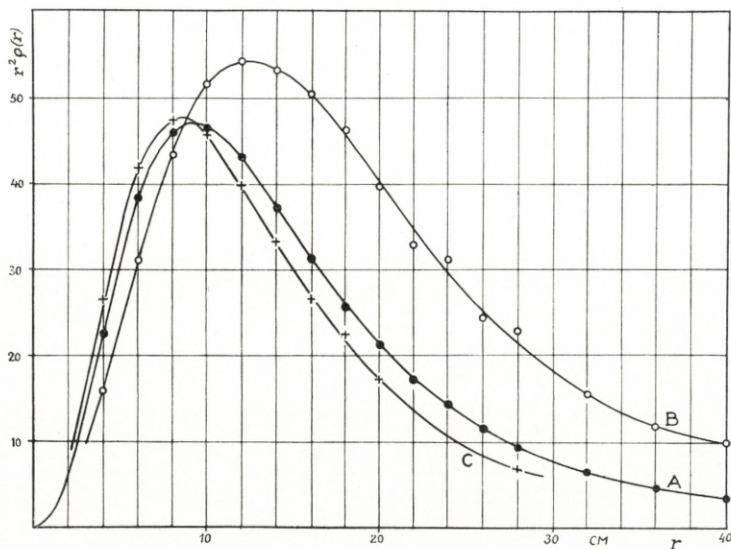


Fig. 3. Neutron density times squared distance from source plotted with respect to distance. *A*, water (distilled); *B*, benzene; *C*, liquid paraffin.

using as ordinates the measured (not smoothed) values of  $\rho(r)r^2$ . To obtain the (small) area under the tail of the curve beyond  $r = 40$  cm. an exponential form was assumed.

The "geometrical method" was based on the striking similarity in the shapes of the curves. This similarity is borne out by plotting the curves on a double logarithmic scale: it is then possible to bring the different curves into coincidence by shifting the origin; fig. 6 shows how good a fit can be obtained. This means that, if we return to the original curves (fig. 3, 4, 5), any of them can be made to

coincide with the water curve A by changing the scale of the abscissae and ordinates by suitable factors  $F_x$  and  $F_y$ . By such a change of scale, any area will be changed by a factor  $Q = F_x \cdot F_y$  which is, therefore, the area of the given curve relative to the area of the water curve.

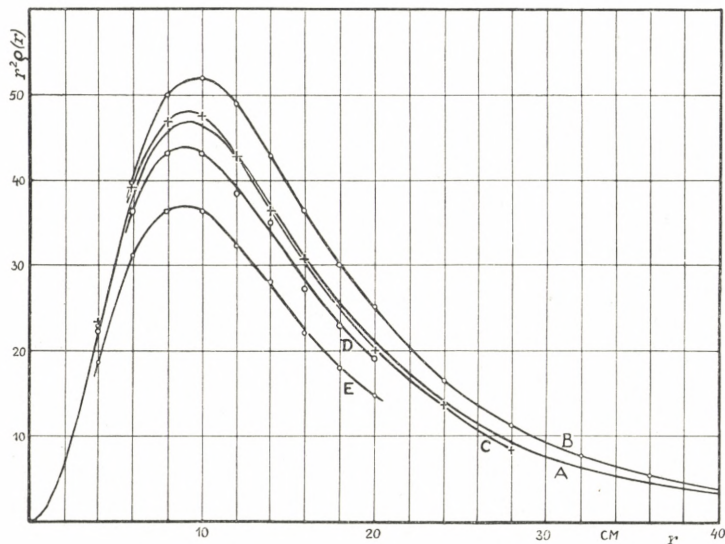


Fig. 4. Neutron density times squared distance from source; A, water; B, 10% heavy water; C, 25% sugar solution; D, 7.6% ammonium nitrate solution; E, 0.41% lithium hydroxide solution.

In table 1 all the figures required for the later discussions are collected. It is seen that the areas determined numerically and geometrically agree quite well in a number of cases. Where there was a small difference, as in the case of the heavy water, the arithmetical mean has been adopted.

With the boric acid solutions, however, the deviations are large and obviously systematic. In fact we cannot expect the boric acid curves to be quite similar in shape to the water curve since the addition of the boron consider-



ably shortens the diffusion path of the slow neutrons, thereby shifting the maximum towards the origin, and at the same time has practically no influence on the slowing down of the fast neutrons which is the determining factor for the

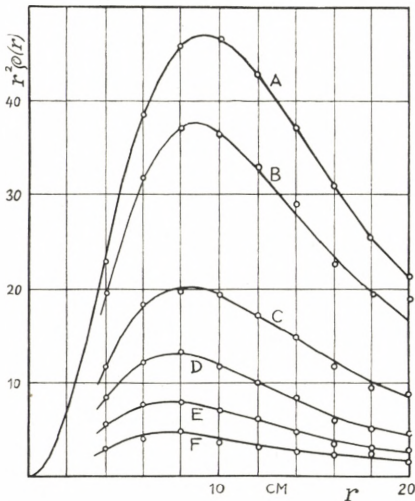


Fig. 5. Neutron density times squared distance from source; Boric acid solutions: A, 0%; B, 0.1%; C, 0.5%; D, 1%; E, 2%; F, 4%.

slope of the curve at large distances from the source. Applying the "geometrical method" to the boric acid curves means extrapolating with a slope which is too steep and thereby obtaining values for the area which are too small. In the case of the boric acid solutions we have therefore based the further calculations on the numerically calculated values only; the part of the curves between 20 and 40

cm. which had not been measured was extrapolated assuming the same logarithmic slope as in the water curve.

The area of the LiOH-curve was determined geometrically and then corrected assuming the same difference between geometrical and numerical method as in the case of the 0.1 per cent  $H_3BO_3$  solution; in this way the figure in brackets (0.76) was obtained.

## 5. Corrections.

The error introduced by the limited size of the container (in theory it should be infinite) was found, by re-

measuring the water curve with a much larger container (about 80 cm. in all directions), to be of the order of one per cent or less. Hence it would appear that the error is equally small, and therefore negligible, in all the other liquids, with the exception of benzene. In benzene the hydrogen density is much smaller than in the other liquids investigated; the mean free path for neutrons is larger and consequently the whole curve is more stretched ( $F_x$  is considerably larger than unity). A considerable increase in  $n$  might therefore be anticipated for measurements carried out in a much larger volume of benzene.

In order to avoid this costly and dangerous experiment (a fire is bad enough with 60 ltrs. of benzene) we tried to determine the necessary correction in an indirect way, based on a similarity argument. We have remeasured the water curve using a container which was similar in shape to the old one but smaller by a factor of 1.32 (the value of  $F_x$  for benzene) in all directions. We believe then that we commit the same error as we committed in the benzene measurement<sup>1</sup>. The

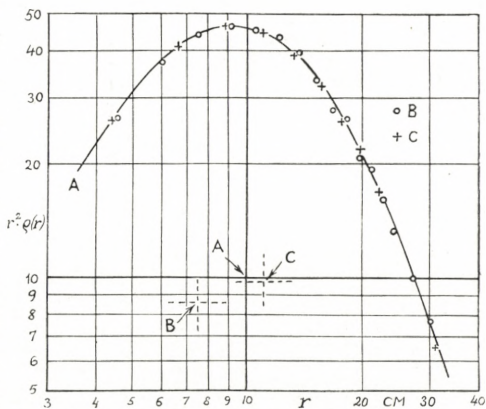


Fig. 6. Same as fig. 3 (with corresponding letters) but on double logarithmic plot. In plotting B and C the frame has been shifted in such a way that the point (10, 10) lies for the three sets of points at the positions indicated by the arrows A, B, C. It is possible in this way to bring the curves into coincidence.

<sup>1</sup> This would be exactly correct if the slowing-down process in benzene were exactly the same as in water with the sole difference that the

value for  $n$  obtained with the small water container was found 5 per cent smaller than with the large container, and consequently we have increased the value of  $Q$  for benzene by 5 per cent, from 1.55 to 1.63.

The error introduced by the limited size of the ionisation chamber (in theory it should be a point) was estimated in the following way. The chamber was immersed in water and a thin-walled brass tube of twice the chamber's diameter was placed around it. If the space between the chamber and the tube was filled with air (instead of water), the number of counts was increased by 4 per cent, with the source at a distance of 10 cm. Since the disturbance introduced by a cylindrical "air bubble" should be roughly proportional to the square of its diameter, it was concluded that the disturbance introduced by the ionisation chamber was of the order of one per cent only; no correction has been applied for this. The disturbance caused by the absorption in the chamber walls and the boron layer was estimated to be negligible.

### 6. Capture cross-sections.

From the figures given in table 1, relative capture cross-sections have been calculated using equation (4). First we compared benzene and liquid paraffin; this gives the equation

$$\frac{n_1}{n_2} = \frac{1.63}{0.92} = 1.77 = \frac{(3.82 p_C + 7.15 p_H)}{(4.05 p_C + 4.05 p_H)}$$

and therefrom

$$\frac{p_C}{p_H} = \frac{\sigma_{cC}}{\sigma_{cH}} = 0.005 \pm 0.02.$$

mean free path for every neutron velocity is longer by a factor 1.32. This is certainly not exactly true, but the good fit of fig. 6 indicates that it is at least roughly true.

This means that the absorption cross-section of carbon is negligible with respect to that of hydrogen, in fact probably smaller than one fiftieth. Qualitatively, the vanishing capture cross-section of carbon is evident from the fact that the ratio of the neutron numbers (1.77) is practically identical with the inverse ratio of the hydrogen densities  $\left(\frac{7.15}{4.05} = 1.765\right)$  in the two liquids.

Assuming  $\sigma_{cC} = 0$  we can proceed to determine  $\frac{\sigma_{cO}}{\sigma_{cH}}$  by comparing water and benzene. From the equation

$$\frac{n_1}{n_2} = \frac{1.63}{1.00} = 1.63 = \frac{(6.70 p_H + 3.35 p_O)}{4.05 p_H}$$

we get

$$\frac{p_O}{p_H} = \frac{\sigma_{cO}}{\sigma_{cH}} = -0.03 \pm 0.02$$

which means that the capture cross-section of oxygen, like that of carbon, is negligible, at any rate probably smaller than  $0.02 \cdot \sigma_{cH}$ . Again, the ratio of the neutron numbers (1.63) is practically identical with the inverse ratio of the hydrogen densities  $\left(\frac{6.70}{4.05} = 1.65\right)$ .

The sugar experiment does not give additional information because the amount of carbon contained in the sugar solution is too small. The result is, however, in agreement with the foregoing for the neutron number  $n$  as well as the hydrogen density are not considerably changed by dissolving the sugar; only the shape of the curve is slightly modified ( $F_x = 0.98$ ) because the additional carbon shortens the mean free path of the neutrons (especially of the fast ones, which are decisive for the shape).

On dissolving ammonium nitrate in water, however, the

neutron number is considerably reduced ( $Q = 0.93$ ), while the slight decrease (one per cent) in hydrogen density would cause in itself a corresponding increase in neutron number. From the equation

$$\frac{n_1}{n_2} = 0.93 = \frac{6.70 p_H}{(6.63 p_H + 0.118 p_N)}$$

we get

$$\frac{p_N}{p_H} = \frac{\sigma_{cN}}{\sigma_{cH}} = 5 \pm 1.$$

The heavy water experiment will be discussed later in connection with the experiments with photo-neutrons.

The boric acid and lithium hydroxide experiments are of special interest because they enable one to place all the measured cross-sections on an absolute scale. On comparing the five boric acid solutions (0.1, 0.5, 1, 2, 4 per cent) with water, the following values for  $\frac{P_B}{P_{H_2O}} = \frac{P_B}{2 P_H}$  are obtained: 935, 970, 1010, 945, 940. The agreement is satisfactory and as a weighted mean the value of  $970 \pm 50$  was accepted. On comparing the lithium hydroxide solution with water we get  $\frac{P_{H_2O}}{P_{Li}} = 102 \pm 10$ .

It is known that thermal neutrons are very little scattered in boron and lithium (18) and we may therefore identify their capture cross-sections with their total cross-sections for thermal neutrons as it is determined by absorption experiments on beams of thermal neutrons. Assuming a value of  $(500 \pm 30) \cdot 10^{-24}$  cm.<sup>2</sup> for  $\sigma_B$  (see § 10) we find  $\sigma_{cH_2O} = (0.51 \pm 0.04) \cdot 10^{-24}$  cm.<sup>2</sup> and consequently  $\sigma_{cH} = (0.26 \pm 0.02) \cdot 10^{-24}$  cm.<sup>2</sup>. Taking, furthermore,  $\sigma_{Li} = (63 \pm 4) \cdot 10^{-24}$  cm.<sup>2</sup> (see § 10) we get  $\sigma_{cH} = (0.30 \pm 0.04) \cdot$



$\cdot 10^{-24}$  cm.<sup>2</sup>. As the most probable value we suggest the weighted mean  $\sigma_{cH} = (0.27 \pm 0.02) \cdot 10^{-24}$  cm.<sup>2</sup>.

From this the absolute capture cross-sections of C, N, and O can be calculated; the results are collected in table 2 (including D, see § 7).

Our value of  $\sigma_{cH}$  is somewhat smaller than the value given by AMALDI and FERMI ( $0.31 \cdot 10^{-24}$  cm.<sup>2</sup>), which fur-

Table 2.

Capture cross-sections of light elements for neutrons of  $\sqrt{2kT/m} = 2.2 \cdot 10^5$  cm./sec. (see § 10).

Element	H	D	C	N	O
$\sigma_c \cdot 10^{24}$ cm <sup>-2</sup>	$0.27 \pm 0.02$	$< 0.03$	$< 0.01$	$1.3 \pm 0.3$	$< 0.01$

thermore corresponds (4) to a velocity of  $2.5 \cdot 10^5$  cm.sec.<sup>-1</sup> and should therefore be increased by a factor of 1.13 to be compared with our value. Our figure for  $\sigma_{cN}$  can be compared with the cross-section of nitrogen for disintegration by thermal neutrons which has been determined by BURCHAM and GOLDHABER (19) by counting the protons emitted in the process; the agreement between their value ( $\approx 1 \cdot 10^{-24}$  cm.<sup>2</sup>) and our figure indicates that most of the slow neutrons captured by a nitrogen nucleus give rise to the emission of a proton rather than of a gamma-quantum, just as one would expect it to be on general theoretical reasons, for such a light nucleus.

## 7. Experiments with photo-neutrons.

In order to discuss the heavy water experiment described in the last section we have to take account of the

neutrons which are created throughout the liquid under the action of the radium gamma-rays (6). The possibility of creation of neutrons when fast neutrons from the source collide with deuterons has also to be kept in mind. In order to get information on both these effects we have made some measurements where the Ra-Be source was replaced by a neutron-free gamma-ray source; furthermore by surrounding this source with beryllium, low energy neutrons were obtained (7) which are certainly not able to

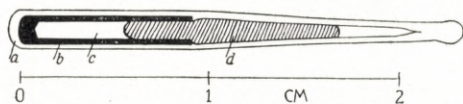


Fig. 7. Neutron-free gamma-ray source. *a*, glass tube; *b*, chrom-nickel tube; *c*, radon; *d*, mercury.

disintegrate deuterons by their impact.

The neutron-free gamma-ray source (see fig. 7) consisted of a small tube of a chrom-nickel alloy sealed into glass. The tube was filled with about 300 mCurie radon; a mercury seal prevented the radon from coming in contact with the glass (which would have caused neutron emission); the glass was sealed off on top of the mercury, for safety. The source was found to emit one thousandth of the neutrons emitted by a source in which the same amount of radon was mixed with beryllium.

The source was placed at the center of either of two identical thin-walled brass containers (cylindrical with 1.8 cm. height and 1.9 cm. diameter); one of them was filled with 2.7 gr. metallic beryllium, the other with 3.0 gr. carbon which has scattering properties for neutrons and gamma-rays similar to those of beryllium but emits no photo-neutrons. Measurements were carried out in the same way as described above, with ordinary and "heavy" water; the

decay of the radon was taken into account and the results were always expressed per 100 mCurie.

The results are shown in fig. 8. Because of the smaller energy of the neutrons, curves *A* and *B* have a much steeper slope and the maximum lies at smaller distances from the origin than in the curves shown in fig. 3, 4 and 5. Curve *D*, however, is especially flat; this is due to the fact that the neutrons are produced throughout the liquid and that it is not their absorption but that of the gamma-rays which is responsible for the slope. The half-value thickness of water for gamma-rays of 2.2 MeV is<sup>1</sup> 15 cm., which is in good agreement with the final slope of curve *D* (see curve *G*). In curve *B* at small distances from the source the neutrons come mostly from the beryllium while for large distances the curve flattens out, most of the neutrons being then due to photo-disintegration of the deuterons.

The difference of *A* and *C* (curve *E*) corresponds to the photo-neutrons from the beryllium only and curve *E* is therefore very similar in shape to the corresponding curve  $F = B - D$ . The areas of curve *F* and *E* have been determined numerically and their ratio is found to be 1.11. From this ratio we obtain from  $1.11 = \frac{6.70 p_H}{(6.10 p_H + 0.60 p_D)}$  the value  $\frac{p_D}{p_H} = \frac{\sigma_{cD}}{\sigma_{cH}} = -0.1 \pm 0.1$ .

In the experiments with the Ra + Be source a slightly higher value for the ratio of the areas of the heavy and ordinary water curves was found (1.13; see table 1). If we, however, deduct from the heavy water curve the photo-

<sup>1</sup> Calculated from figures given by L. MEITNER and H. H. HUPFELD, ZS. f. Physik, **67**, 147 (1931). It should be noted that  $\gamma$ -rays which have been scattered once have almost certainly not enough energy left to disintegrate deuterons.

neutrons produced in the deuterium, that is, the area of  $D-1.11 \cdot C$  (fig. 8), then the ratio becomes 1.11 in agreement with the ratio found when using the low energy beryllium photo-neutrons. We conclude from this that the number of neutrons created by collisions between fast neutrons and deuterons is too small to be detected in this way. This is not surprising since only about one out of ten fast neutrons will happen to hit a deuteron before having lost part of its energy and because only a small fraction of the primary neutrons have sufficient energy to disintegrate the deuteron ( $\approx 3.2$  MeV).

It is assumed in the calculation of the capture cross-sections that all the neutrons emitted by the source reach low energies where the  $1/v$ -law can be applied while only a negligible fraction vanishes before getting slow. Although for example nitrogen is known to offer a considerable cross-section for disintegration by fast neutrons (22), still the number of neutrons which get lost, in this way, in our  $\text{NH}_4\text{NO}_3$ -solution can be estimated to be very small, just as the number of neutrons created in collisions with deuterons in our heavy water was found experimentally to be negligible. Furthermore, our result that the total number  $n$  of neutrons found in water, benzene and paraffin oil is inversely proportional to the hydrogen densities in the three liquids excludes any considerable loss of neutrons on account of the presence of oxygen and carbon.

### 8. Yields.

The relative neutron yields of the sources investigated by us can be deduced fairly accurately from the experiments. From the ratio of the areas of curve  $E$  (fig. 8) and

curve *A* (fig. 4) which is 0.039 we see that the  $Rn\gamma + Be$  source emits 3.9 per cent of the neutrons emitted by the  $Ra + Be$  source. The total area of curve *D*— $1.11 \cdot C$  (fig. 8) (indicating the production of photo-neutrons in our heavy water) can be estimated

to be 0.020 of that of curve *B* (fig. 4). Furthermore, we have found that our  $Ra + Be$  source emitted about 65 times as many neutrons as 1 mCurie radon mixed with beryllium. The total number of neutrons emitted by such a source has been measured (21, 11) and values ranging from 7.000 to 27.000 per second have been obtained. Assuming a value of 20.000 we can calculate the cross-sections of deuterium and beryllium for photo-disintegration by the gamma-rays of  $Ra$

$B + C$ . The intensities of the two hardest lines (2.2 and 1.8 MeV.) have been found (22) to correspond to the emission of 0.074 and 0.26 quanta per disintegration, respectively. With deuterium the hardest line only can be effective and the cross-section is found equal to  $7 \cdot 10^{-28}$  cm.<sup>2</sup> (see however, the note at the end of the paper). For beryllium we have to introduce two cross-sections  $\sigma_{1.8}$  and  $\sigma_{2.2}$  corresponding to the two lines of 1.8 and 2.2 MeV.; from the

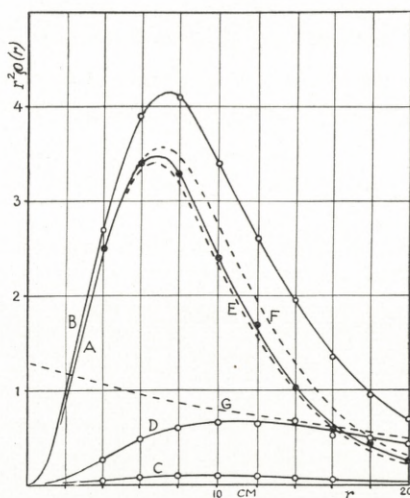


Fig. 8. Neutron density times squared distance from source. *A* and *B*: gamma-ray source + beryllium; *C* and *D*: without beryllium; *A* and *C* in water; *B* and *D* in 10% heavy water. *E* = *A* - *C*; *F* = *B* - *D*. *G* indicates the calculated absorption of the hardest gamma-ray line of radium *C* (2.2 MeV.) in water.

observed total number of neutrons we then get the equation  $\sigma_{1.8} + 0.3 \cdot \sigma_{2.2} = 2 \cdot 10^{-27}$  cm.<sup>2</sup>. Should the number of neutrons emitted by 1 mCurie of Rn + Be turn out to be larger than 20.000 per second then the cross-sections given have to be increased in the same proportion.

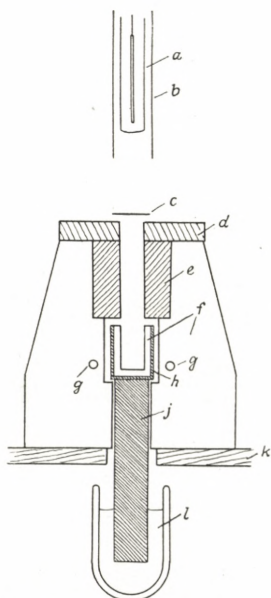


Fig. 9. Arrangement for neutron beam experiments. *a*, boron trifluoride chamber (see fig. 10); *b*, cadmium shield; *c*, absorber; *d*, boron shield (about 1 gr./cm.<sup>2</sup>); *e*, lead block (the chamber is fairly sensitive for gamma-rays); *f*, paraffin wax; *g*, sources; *h*, copper beaker soldered to copper bar *j*; *k*, wooden support; *l*, liquid air.

## 9. Neutron beam experiments.

In addition to the experiments on the capture of neutrons described above we have carried out some experiments on the absorption (including scattering) of slow neutron beams in a number of substances. On the one hand we wanted to redetermine the cross-sections of boron and lithium, the values published so far disagreeing considerably among themselves; on the other hand we tried to get additional information on the influence of the temperature of the paraffin on the absorbability in different elements of the neutrons stopped by cadmium ("C-neutrons") (11).

A first set of measurement was carried out using, as neutron source, a paraffin howitzer the inner part of which could be cooled down to the temperature of liquid air (fig. 9). The copper beaker covered, on the inside, with a one cm. layer of paraffin wax was soldered on a copper bar of

4 cm. diameter, the lower end of which was immersed in liquid air. The temperature of the beaker was within a few degrees that of the liquid air, varying slightly with the liquid air level; about 2 ltrs. of liquid air per hour were required.

As detector for the slow neutrons an ionisation chamber filled with boron trifluoride was used. The construction of the chamber may be taken from fig. 10. The chamber was

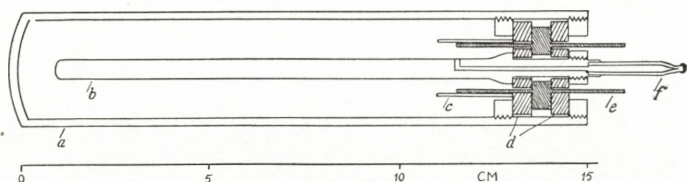


Fig. 10. Boron trifluoride chamber. *a*, steel cylinder; *b*, steel collecting electrode; *c*, glass tube (may be unnecessary); *d*, ebonite rings; *e*, grounded ring; *f*, copper capillary (through which the gas is introduced), squeezed and sealed with solder. The chamber was made gas-tight by apiezon wax *W*, which also covers the inside surface of the ebonite and protects it from contact with the  $\text{BF}_3$ .

at first filled to about 10 atm. pressure; excessive potentials were, however, required to collect the ions in a sufficiently short time and the pressure was therefore lowered to about  $1\frac{1}{2}$  atm., about 800 Volt still being required for satisfactory operation. The chamber has proved very stable and reliable during nearly a year's work and the apiezon wax *W* used for insulation does not seem to deteriorate in contact with the boron trifluoride.

A cadmium screen ( $0.4 \text{ gm./cm.}^2$ ) was used as usually to discriminate between fast and slow neutrons (C-neutrons). The cadmium was always placed below (instead of above) the absorbing sample; this is important in the case of paraffin where the slowing down of neutrons in the sample may considerably lower the apparent absorption if the Cd is

placed on top of the paraffin. All the substances were investigated as pure elements, except for Cl where  $\text{CCl}_4$  was used.

### 10. Neutron beam experiments: Results.

The results are collected in table 3. The absorber thickness was always chosen such as to reduce the beam intensity to about one half; from the absorber thickness and the intensity reduction the values of  $\mu$  were calculated.

Table 3.

Mass absorption coefficients (of the compounds used) in  $\text{cm}^2/\text{g}$ . and total cross-sections (of the elements) for beams of thermal neutrons. The subscripts give the temperature of the paraffin used in slowing down the neutrons.

Absorber	$\mu_{290}$	$\mu_{80}$	$\frac{\mu_{80}}{\mu_{290}}$	$\sigma_{290} \cdot 10^{24} \text{ cm}^{-2}$
H (paraffin wax)	4.3 $\pm$ 0.2	4.7 $\pm$ 0.1	1.09 $\pm$ 0.05	48 $\pm$ 3
Li . . . . .	5.5 $\pm$ 0.2	8.3 $\pm$ 0.3	1.51 $\pm$ 0.10	63 $\pm$ 4
B . . . . .	27.7 $\pm$ 0.9	42.3 $\pm$ 1.2	1.53 $\pm$ 0.07	500 $\pm$ 30
Al . . . . .	0.035 $\pm$ 0.002	0.039 $\pm$ 0.002	1.10 $\pm$ 0.10	1.6 $\pm$ 0.1
Cl ( $\text{CCl}_4$ )	0.82 $\pm$ 0.02	1.32 $\pm$ 0.04	1.61 $\pm$ 0.06	51 $\pm$ 2
Fe . . . . .	0.115 $\pm$ 0.007	0.12 $\pm$ 0.007	1.05 $\pm$ 0.10	12 $\pm$ 0.7
As . . . . .	0.070 $\pm$ 0.006	0.094 $\pm$ 0.004	1.34 $\pm$ 0.13	8.8 $\pm$ 0.8
Ag . . . . .	0.34 $\pm$ 0.015	0.52 $\pm$ 0.02	1.53 $\pm$ 0.09	60 $\pm$ 4
Cd . . . . .	16.6 $\pm$ 0.7	16.6 $\pm$ 0.5	1.00 $\pm$ 0.05	3100 $\pm$ 150
Au . . . . .	0.29 $\pm$ 0.015	0.43 $\pm$ 0.03	1.48 $\pm$ 0.15	90 $\pm$ 5

It must, however, be kept in mind that the C-neutrons are inhomogeneous and that their velocity distribution is changed when they pass through an absorber the cross-section of which depends on their velocity. If the neutrons originally have a MAXWELLIAN distribution and if the absorber



follows the "1/v-law" (i. e. if its cross-section is inversely proportional to the velocity of the neutrons), then the cross-section obtained when using a very thin absorber corresponds to a neutron velocity of  $\sqrt{\pi kT/2m} = 1.95 \cdot 10^5$  cm.sec.<sup>-1</sup>; with thicker absorbers, however, a hardening occurs such that the value which is found using a "half-value" absorber corresponds (23) to a velocity which is about 1.13 times larger and is, in fact, very nearly  $= \sqrt{2kT/m} = 2.2 \cdot 10^5$  cm.sec.<sup>-1</sup>.

It is seen that some of the substances (Li, B, Cl, Ag, and Au) show a "temperature effect" (8, 9, 10) (increase of absorption coefficient when the paraffin wax from which the neutrons come is cooled to liquid air temperature) of about 1.5 while some other substances show practically no temperature effect (paraffin wax, Al, Fe); with As an intermediate value is obtained.

In fact, no temperature effect should be expected if the observed cross-section is all due to scattering; the strong scattering of slow neutrons by aluminium and iron is well known (24, 18). In paraffin wax, however, although absorption is negligible compared with scattering, a small temperature effect should be expected on account of the chemical bond action (4); calculations carried out by Mr. N. ARLEY (to appear shortly in Kgl. Danske Vid. Selsk. Math. Phys. Medd.) give a factor of about 1.3. Such an effect has also been observed by FINK (9).

In the case of lithium and boron, there are strong theoretical reasons (4) to expect their cross-sections to be inversely proportional to the neutron velocity, over a considerable velocity range; also in the case of silver and gold one would expect the 1/v-law to be valid for neutrons of thermal energies. It is then very satisfactory that all these elements show the same amount of temperature effect. That chlor-

ine shows it as well is in agreement with the fact (25) that also in chlorine most of the cross-section is due to capture. In the case of arsenic, the value of 1.34 seems to indicate that both capture and scattering contribute to the observed cross-section; indeed, unpublished experiments of one of us (O. R. F.) on the activation of As by C-neutrons

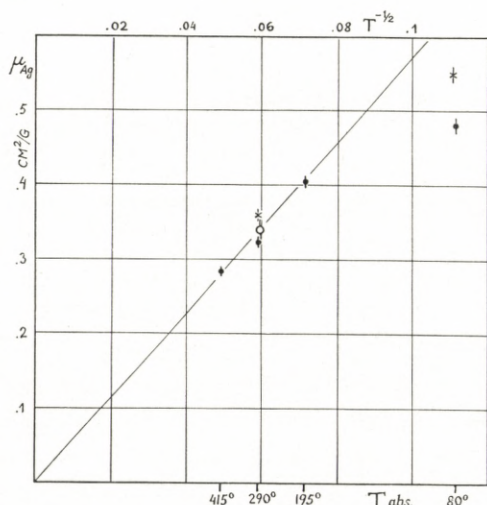


Fig. 11. Absorption coefficient of silver plotted with respect to  $T^{-\frac{1}{2}}$  ( $T$  = absolute temperature of the paraffin). Dots, "flat" paraffin block; crosses, thick-walled howitzer; open circle, boric acid difference experiment. (The straight line would correspond to the " $T^{-\frac{1}{2}}$ -law").

indicated a capture cross-section of about one half the total cross-section.

However, the amount of the temperature effect observed (1.5) is not equal to  $\left(\frac{T_1}{T_2}\right)^{-\frac{1}{2}} = \left(\frac{80}{290}\right)^{-\frac{1}{2}} = 1.9$ , the inverse ratio of the mean velocities of neutrons in temperature equilibrium (28). Since this discrepancy might be due to incomplete temperature equilibrium of the neutrons cooled in a thin layer of paraffin only (see fig. 9), we have made some

experiments (27) with a flat block of paraffin of 5—6 cm. thickness which could be kept at four different temperatures: 80° (liquid air), 195° (solid carbon dioxide), 290° (room temperature), and 415° (the paraffin being molten, in this case). The absorption coefficient of silver  $\mu_{Ag}$  for C-neutrons was measured for each temperature. If the neutrons were in thermal equilibrium one would expect  $\mu_{Ag}$  to be proportional to  $T^{-\frac{1}{2}}$ . From fig. 11 this relation is seen to hold fairly well for the three highest temperatures while the lowest point strongly deviates from the " $T^{-\frac{1}{2}}$ -law", the ratio being again about 1.5 only, instead of 1.9.

This discrepancy has been observed before. FOMIN and his colleagues (10) found a temperature effect of 2.2 when comparing the absorption of silver for neutrons of 20° (liquid hydrogen) and 290° while  $\left(\frac{T_1}{T_2}\right)^{-\frac{1}{2}} = 3.8$  in this case. Furthermore for the absorption of boron the factor was only  $1.65 \pm 0.2$ . In our experiments, no significant difference between silver and boron was found, comparing neutrons at liquid air temperature (80°) and 290°. FOMIN et al. do not explicitly give the temperature effect for the step from 80° to 290°, but from their absorption figure (on silver) a value of about 1.3 seems to follow. FINK (9) using an arrangement quite similar to ours, obtains for Li, B, and Ag values of 1.20, 1.32, and 1.20, respectively, which are all lower than our figures; we do not know the reason for this difference.

The discrepancy between the observed temperature effects and the value of  $\left(\frac{T_1}{T_2}\right)^{-\frac{1}{2}}$  has sometimes been taken as an argument against the validity of the  $1/v$ -law. We believe, however, that the explanation lies in the fact that the neutrons defined by their strong absorption in cadmium

(C-neutrons) have not thermal energies throughout but include a "tail" of neutrons with energies up to the cadmium absorption limit (about 1 volt.) which is more intense than would follow from MAXWELL'S distribution law.

In order to calculate the number of "tail neutrons" relative to the thermal ones a detailed knowledge is required of the slowing down mechanism which is not available at present. A very rough estimate, however, indicating the order of magnitude may be obtained as follows. We know that neutrons with energies below 0.1 eV. lose their energy at a much lower rate than fast ones because the protons in the paraffin, on account of their chemical binding, cannot be regarded as free when offered energy amounts less than one quantum of oscillation energy; they should rather be regarded as rigidly connected to the carbon atom to which they belong. The carbon atoms having much lower oscillation frequencies may be regarded as free, in first approximation. One may then get a rough picture of the slowing down mechanism by assuming (2) that the neutrons collide with particles of mass 14, (considering the paraffin as a structure of  $\text{CH}_2$ -molecules).

In collisions with such particles, neutrons will lose one eighth of their energy at every impact, at an average, and about ten collisions will be required to bring a neutron from 0.1 Volt down to 0.03 Volt (mean energy at room temperature). On the other hand a neutron, on leaving the paraffin, has, on the average, suffered about 100 collisions (see the next section). There is therefore, roughly speaking, a 10 per cent chance of a neutron emerging from the paraffin before having reached thermal equilibrium, and consequently 10 per cent of the neutrons which get into the boron chamber will be "tail neutrons". Taking their greater velocity into account

(which is of importance not only as regards the penetrating of the absorber but also the sensitivity of the boron chamber) we get the result that the mean absorption coefficient of C-neutrons of room temperature is about 5 per cent less than for neutrons in perfect thermal equilibrium.

At liquid air temperature, however, a much larger difference should result. Firstly, about twice as many collisions should be required (about 20) for the neutrons to get into thermal equilibrium, starting from 0.1 Volt; secondly, the number of collisions which they survive should be considerably smaller (perhaps 60) on account of the increased capture cross-section of the protons for slower neutrons.

It has been suggested (28) that the capture cross-section of the protons at low temperature may be decreased on account of their zero-point oscillations. However, a detailed discussion shows that quite generally the mean life of slow neutrons in any substance which has no resonances in the energy range concerned should be entirely independent of temperature, depending only on the density of absorbing nuclei (in the way discussed in the first chapter); and there are certainly no resonances near the thermal range, in the case of paraffin.

Furthermore, the slowing down of neutrons below room temperature is additionally hampered by the fact that the carbon atoms in the paraffin can be no more regarded as free to take up such small energy amounts (one eighth of 0.03 volt or less). This effect should be expected to cause rapidly increasing deviations from the  $T^{-\frac{1}{2}}$ -law as temperatures goes down, in agreement with our experiments.

If this explanation is true, one would expect the absorbability of the neutrons to be higher when the arrang-

ement is such that the mean life of the neutrons in the cold paraffin is increased. We have therefore measured  $\mu_{Ag}$ , using a thick-walled ( $\approx 4$  cm.) paraffin howitzer which could be cooled as a whole; the neutron sources were placed outside the Dewar flask, below the howitzer, with more paraffin wax underneath. The values obtained (see the crosses fig. 11) are in fact larger than with the flat block, especially at the low temperature.

We have also tried to measure the absorption of thermal neutrons at room temperature without using cadmium to distinguish them from the faster ones. We have used two equal howitzers made of water and a 0.5 per cent boric acid solution, respectively; the liquids were contained in thin-walled metal forms. The boric acid should have very little influence on the intensity of the fast and "tail" neutrons emitted from the howitzer since those neutrons stay only very short time in the liquid; the difference in intensity (which was about 50 per cent) should be almost entirely due to thermal neutrons. The absorption of these "difference" neutrons in silver was determined by measuring the intensity difference between the two howitzers with and without a silver absorber in front of the boron chamber. The absorption coefficient obtained (see the open circle in fig. 11) was practically equal to the value obtained for C-neutrons using a howitzer; this value may therefore be regarded as the value for true thermal neutrons, with an error less than 5 per cent. This is of some importance because all the neutron energy determinations based on the  $1/v$ -absorption of boron (13, 14, 28, 23, 29, 15) depend on the assumption of thermal equilibrium of C-neutrons at room temperature, an assumption which gains strong support by our measurements.

### 11. Mean life-time of neutrons in water.

From the measurements reported in § 6 and § 9 we can calculate the mean life-time  $\tau_0$  of neutrons in water. The life-time in a boric acid solution is given by

$$\tau = \frac{1}{\left(\frac{1}{\tau_0} + p_B N_B\right)} \quad (5)$$

(this follows from equation (2)). In the graphical representation given in fig. 12 which is, of course, essentially identical with the calculation given in § 6 the intersection of the straight line indicates the boron density  $N'_B$  which by itself would permit the neutrons to live just as long as they do in pure water; the life-time is then found  $\tau_0 = 1/p_B N'_B$ . The value for  $p_B$  is taken from § 10:  $p_B = \sigma_B \cdot v = 500 \cdot 10^{-24} \text{ cm.}^2 \cdot 2.2 \cdot 10^5 \text{ cm. sec.}^{-1} = 1.1 \cdot 10^{-16} \text{ cm.}^3 \text{ sec.}^{-1}$ . From this (and a corresponding calculation making use of the LiOH experiment, see § 6) we get  $\tau_0 = (2.7 \pm 0.2) \cdot 10^{-4} \text{ sec.}$ , which is somewhat larger than the value  $1.7 \cdot 10^{-4} \text{ sec.}$  given by AMALDI and FERMI.

The average number of collisions suffered by a neutron is then equal to  $N = \frac{\bar{v} \tau_0}{\lambda}$ , where  $\bar{v} = \sqrt{\frac{8 kT}{\pi m}} = 2.48 \cdot 10^5 \text{ cm. sec.}^{-1}$ . Taking the mean free path  $\lambda = 0.31 \pm 0.02 \text{ cm.}$  according to our value of the hydrogen scattering cross-section (table 3) we find  $N = 205 \pm 20$ , which is again somewhat larger than the value ( $N = 140$ ) obtained by AMALDI and FERMI.

We have also made similar measurements where a boron chamber was placed outside a can (28 cm. high, 19 cm. diameter) filled with water or boric acid solution and with

the source placed in the center. The results are also shown in fig. 12 (open circles). The life is seen to be shorter just as one should expect since the life of the neutrons is limited not only by capture but also by escaping through the surface. A more detailed discussion of similar experi-

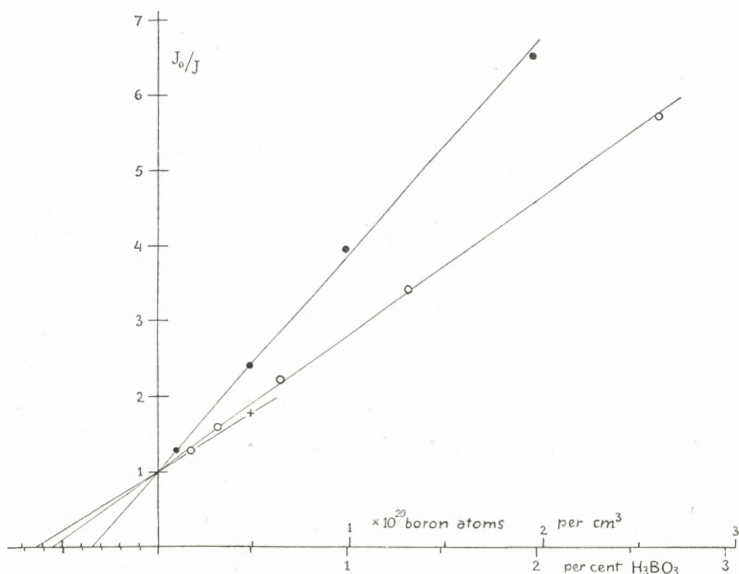


Fig. 12. Abscissa, Amount of boron added to water; ordinate, Reciprocal of the intensity reduction caused by adding the boron. Filled circles, Intensity integrated over large volume (cf. table 1); Open circles, Intensity observed outside water can; Cross, Water howitzer (§ 10).

ments, taking into account the diffusion of the neutrons, has been given by WESTCOTT (30).

In a howitzer the surface should but little shorten the life of the neutrons as only those which pass the bottom almost normally will escape for ever. However from the intensity reduction caused by the boron in the water-howitzer experiment described at the end of § 10 a rather



short life-time would follow (see the cross in fig. 12). This probably comes from the fact that the source was placed very near the bottom of the bore in this experiment, so that there was a strong neutron density gradient; such a gradient would reduce the life in a similar way as a free surface, or in other words, the neutrons observed near the source are preferably those which have suffered only few collisions.

It is a pleasure to the authors to express their gratitude to Prof. N. BOHR for the hospitality extended to them at the Institute of Theoretical Physics of Copenhagen, and for his kind interest in the work. We are greatly indebted to Prof. G. PLACZEK for stimulating discussions. One of us (O.R.F.) wishes to thank the Rask-Ørsted Foundation for a grant. Our thanks are furthermore due to the Norsk Hydro-Elektrisk Kvælstofaktieselskab, Oslo, for the generous loan of sixty litres of heavy water of ten per cent deuterium oxide content.

### SUMMARY

1. The density distribution of slow neutrons around a source was studied, the source being placed in large volumes of different liquids. From the relative total numbers of neutrons their relative life in the liquids could be obtained. Assuming the life to be limited only by the capture of the neutrons while slow, the capture cross-sections of a number of light elements were obtained, relative to the capture cross-sections of boron and lithium, which are known absolutely from neutron beam absorption experiments. The results are collected in table 2.

2. The yield of photo-neutrons from beryllium and deuterium irradiated with radium gamma-rays has been compared with the yield of neutrons from beryllium mixed with radon. Assuming the latter yield as 20 000 neutrons per second per mCurie, the cross-section of deuterium for dissociation by the gamma quantum of 2.2 MeV. is found to be  $7 \cdot 10^{-28}$  cm<sup>2</sup>. In the case of beryllium the two hardest lines of radium with 1.8 and 2.2 MeV. are energetic enough to cause dissociation; in this case we find  $\sigma_{1.8} + 0.3 \cdot \sigma_{2.2} = 2 \cdot 10^{-27}$  cm<sup>2</sup> where  $\sigma_{1.8}$  and  $\sigma_{2.2}$  correspond to the 1.8 and to the 2.2 MeV. line respectively.

3. The total cross-sections for C-neutrons of a number of elements have been measured by the neutron beam method, the neutron source being a paraffin howitzer the inner part of which could be cooled to liquid air temperature. The cross-section was found to be larger by a factor of about 1.5 for "cold" neutrons, in some of the elements, including those (Li, B, Ag, Au) which are known to show strong absorption and little scattering. In the elements where scattering is known to be strong (H, Fe, Al) and in Cd little or no influence of temperature was found. Arsenic gave an intermediate value.

4. The absorption coefficient of silver has been measured, the neutrons being slowed down in paraffin at different temperatures between 80° and 415°, and under different geometrical conditions. The results (which are shown graphically in fig. 11) support the view that the C-neutrons though containing mostly neutrons with thermal energies include a "tail" of faster neutrons extending to the cadmium absorption limit of about 1 volt. This tail does not change the absorption coefficient of thermal neutrons by

more than a few per cent, at room temperature, but becomes very important at lower temperatures.

5. From our measurements the mean life-time of thermal neutrons in water is found to be  $(2.7 \pm 0.2) \cdot 10^{-4}$  sec. and the mean number of collisions is found equal to  $205 \pm 20$ .

*Institute of Theoretical Physics,  
University, Copenhagen.*

#### NOTE ADDED IN PROOF

In a recent paper of CHADWICK, FEATHER and BRETSCHER (Proc. Roy. Soc. A 163, 366, 1937) the binding energy of the deuteron has been redetermined and found to be  $2.25 \pm 0.05$  MeV (instead of the formerly accepted value of 2.14 MeV). Hence it appears that the neutrons obtained (see p. 19—21) on irradiating heavy water with the gamma-rays of RaB + C cannot be due to the gamma-line of 2.198 MeV but must be produced by some harder gamma-rays. Some evidence for the existence of such gamma-rays has been given in papers of SKOBELZYN (ZS. f. Physik 43, 354, 1927), ELLIS (Proc. Roy. Soc. A 143, 350, 1934), and GRAY (Proc. Roy. Soc. A 159, 263, 1937). Too little, however, is known about their energies and intensities to give a basis for a modified calculation of the cross-sections of the deuteron and of the beryllium nucleus.

## REFERENCES

- 1) E. AMALDI, O. D'AGOSTINO, E. FERMI, B. PONTECORVO, F. RASETTI and E. SEGRÈ, *Ricerca Scient.* **2**, 282, 380 (1934). *Proc. Roy. Soc. A* **149**, 522 (1935).
- 2) See H. A. BETHE, *Rev. Mod. Phys.* **9**, 245 (1937).
- 3) E. U. CONDON and G. BREIT, *Phys. Rev.* **49**, 229 (1936).
- 4) H. A. BETHE and G. PLACZEK, *Phys. Rev.* **51**, 450 (1937).
- 5) J. R. DUNNING, G. B. PEGRAM, G. A. FINK and D. P. MITCHELL, *Phys. Rev.* **48**, 295 (1935).
- 6) J. CHADWICK and M. GOLDBABER, *Proc. Roy. Soc. A* **151**, 479 (1935).
- 7) L. SZILARD and T. A. CHALMERS, *Nature* **134**, 494 (1934).
- 8) C. H. WESTCOTT and H. NIEWODNICZAŃSKI, *Proc. Camb. Phil. Soc.* **31**, 617 (1935).
- 9) G. A. FINK, *Phys. Rev.* **50**, 738 (1936).
- 10) V. FOMIN, F. G. HOUTERMANS, I. W. KURTSHATOV, A. I. LEIPUNSKI, L. SHUBNIKOV and G. SHTSHEPKIN, *Nature* **138**, 326 (1936); V. FOMIN, F. G. HOUTERMANS, A. I. LEIPUNSKI, L. B. RUSINOV and L. W. SHUBNIKOV, *Nature* **138**, 505 (1936).
- 11) E. AMALDI and E. FERMI, *Phys. Rev.* **50**, 899 (1936).
- 12) E. AMALDI, L. R. HAFSTAD and M. A. TUVE, *Phys. Rev.* **51**, 896, (1937).
- 13) O. R. FRISCH and G. PLACZEK, *Nature* **137**, 357 (1936).
- 14) D. F. WEEKES, M. S. LIVINGSTON and H. A. BETHE, *Phys. Rev.* **49**, 471 (1936).
- 15) O. R. FRISCH, *Kgl. Danske Vid. Selsk. math. phys. Medd.* **14**, Nr. 12 (1937).
- 16) T. BJERGE and C. H. WESTCOTT, *Proc. Roy. Soc. A* **150**, 709 (1935).
- 17) E. SEGRÈ, *Ricerca Scient. II (VII)* **1**, 9—10.
- 18) M. GOLDBABER and G. A. BRIGGS, *Proc. Roy. Soc. A* **162**, 127 (1937).
- 19) W. E. BURCHAM and M. GOLDBABER, *Proc. Camb. Phil. Soc.* **32**, 632 (1936).

- 20) N. FEATHER, Proc. Roy. Soc. A **136**, 709 (1932).
  - 21) F. A. PANETH, E. GLÜCKAUF and H. LOLEIT, Proc. Roy. Soc. A **157**, 412 (1936).
  - 22) C. D. ELLIS and G. H. ASTON, Proc. Roy. Soc. A **129**, 180 (1930).
  - 23) H. H. GOLDSMITH and F. RASETTI, Phys. Rev. **50**, 328 (1936).
  - 24) A. C. G. MITCHELL, E. J. MURPHY and M. D. WHITAKER, Phys. Rev. **50**, 133 (1936).
  - 25) J. H. E. GRIFFITHS and L. SZILARD, Nature **139**, 323 (1937).
  - 26) P. PREISWERK and H. v. HALBAN, Nature **136**, 1027 (1935).
  - 27) O. R. FRISCH, H. v. HALBAN and J. KOCH, Nature **139**, 922 (1937).
  - 28) H. v. HALBAN and P. PREISWERK, Nature **137**, 905 (1936).
  - 29) R. JAECKEL, ZS. f. Phys. **104**, 762 (1937).
  - 30) C. H. WESTCOTT, Proc. Camb. Phil. Soc. **33**, 122 (1937).
-



MATHEMATISK-FYSISKE  
MEDDELELSER

UDGIVNE AF

DET KGL. DANSKE VIDENSKABERNES SELSKAB

BIND XV



KØBENHAVN

LEVIN & MUNKSGAARD

EJNAR MUNKSGAARD

1937—38





## INDHOLD

---

1. Die Struktur periodischer Transformationen von Flächen. Von JAKOB NIELSEN.
  2. Flächen zweiter Ordnung und gescharte Kollineationen. Von DAVID FOG.
  3. The Dissociation Constant of the Ortho-Chloro-Anilinium Ion. By KAI JULIUS PEDERSEN.
  4. Om Relationen mellem Varme og Arbejde. Af J. N. BRØNSTED.
  5. Über Potentialunterschiede in membranhaltigen Systemen. Von AUGUSTA UNMACK.
  6. Hg-Dynamics I. Theory of the Laminar Flow of an electrically conductive liquid in a homogeneous magnetic field. By JUL. HARTMANN.
  7. Hg-Dynamics II. Experimental Investigations on the Flow of Mercury in a homogeneous magnetic field. By JUL. HARTMANN and FREIMUT LAZARUS.
  8. Eine Doppelt-fastperiodische ganze transzendente Funktion. Von RICHARD PETERSEN.
  9. Über die innere Paarerzeugung beim  $\beta$ -Zerfall. Von N. ARLEY und C. MÖLLER.
  10. On the Slowing down and Capture of Neutrons in Hydrogenous Substances. By O. R. FRISCH, H. v. HALBAN jun. and JØRGEN KOCH.
  11. Artificial Activity of Hafnium and some other elements. By G. HEVESY and HILDE LEVI.
  12. Algebraic equations with almost-periodic coefficients. By HARALD BOHR and DONALD A. FLANDERS.
  13. A General Chemical Method for the Preparation of the Deuterated Benzenes. By A. LANGSETH and A. KLIT.
- 
-



Det Kgl. Danske Videnskabernes Selskab.

Mathematisk-fysiske Meddelelser. **XV**, 11.

---

ARTIFICIAL  
ACTIVITY OF HAFNIUM AND  
SOME OTHER ELEMENTS

BY

G. HEVESY AND HILDE LEVI



KØBENHAVN  
LEVIN & MUNKSGAARD

EJNAR MUNKSGAARD

1938

Printed in Denmark.  
Bianco Lunos Bogtrykkeri A/S.

### Artificial Radioactivity of Hafnium.

Some time ago we found that under the action of neutron bombardment a radioactive isotope of hafnium is produced, the activity decaying with a period of a few months(4). To determine the period of decay more exactly, we activated 280 mgm. of hafnium oxide prepared by one of us(3) by placing it in a paraffin block together with radium-beryllium sources, containing 600 mgm. of radium element as sulphate and twenty times as much metallic beryllium powder. After irradiation for three months the hafnium oxide was removed from the paraffin block and put into an aluminium dish having a surface of 1.2 cm<sup>2</sup> and a height of 2 mm. The dish was placed directly below the aluminium window of our counter, the window having a thickness of about 20  $\mu$ . We followed the decay of the hafnium

Table 1.  
Decay-measurement of Hafnium.

Date	Nr. of Days	Kicks/min.
17. VIII. 36.....	0	16.1
18. VIII. 36.....	1	15.3
24. VIII. 36.....	7	13.7
12. IX. 36.....	25	12.0
29. X. 36.....	73	7.8
7. I. 37.....	143	2.3
22. I. 37.....	158	1.8
25. II. 37.....	192	1.0

preparation for 200 days by comparing its activity with that of an uranium standard. The decay curve obtained is seen in fig. 1 and table 1.

From the latter we can conclude that the half life of hafnium is  $55 \pm 10$  days (standard mean square deviation).

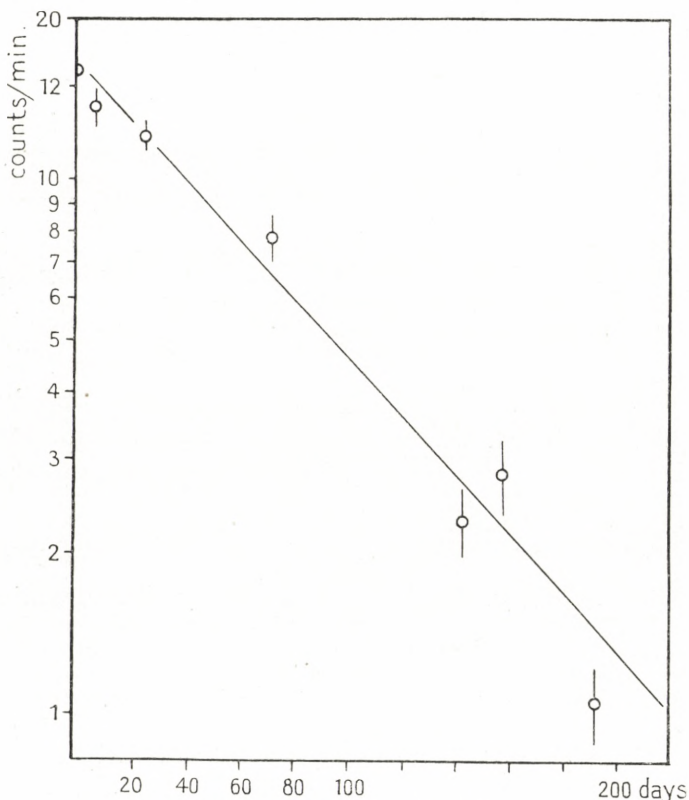


Fig. 1. Decay Curve of Hafnium.

Our initial activity was 20 kicks per min., the natural effect being about 4 kicks per min. We followed the decay curve until we had a net activity of 1 kick per min. From the fact that, in spite of the long activation, such a modest activity was obtained we can conclude that hafnium does

not belong to the elements showing a strong artificial radio-activity. This is partly due to the fact that the capture of neutrons by most of the hafnium isotopes leads, as explained later, to the formation of a heavier stable isotope; as stable isotopes 176, 177, 178, 179, and 180 are known and only the absorption of neutrons by the last mentioned isotope can lead to the formation of an active product. The relative abundance of the isotopes in the naturally occurring element hafnium, as determined by ASTON, is seen from table 2.

Table 2.  
Relative Abundance of the Hafnium Isotopes.

Mass number	Abundance
176	5 %
177	19 %
178	28 %
179	18 %
180	30 %

We measured also the absorption in aluminium of the  $\beta$ -rays emitted by hafnium. The values obtained are seen from table 3.

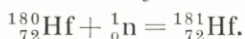
Table 3.  
Absorption in Aluminium of the  $\beta$ -rays Emitted by Hafnium.

Thickness of the Al-foil	Kicks/min.
0	15.8
11 mgm./cm <sup>2</sup>	10.2
16.5 mgm/cm <sup>2</sup>	7.4

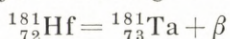
(half value thickness:  $16 \pm 1$  mgm./cm<sup>2</sup>).

From the figures in table 3 follows that an aluminium layer of 16 mgm. per cm<sup>2</sup> reduces the intensity of the  $\beta$ -rays

emitted by a hafnium oxide layer of 230 mgm/cm<sup>2</sup> to one half of its initial value. The comparison of the absorbing power of aluminium for the  $\beta$ -rays of hafnium and scandium, decaying with periods of 55 and 90 days respectively, shows no great difference; the ratio of the two half-value thicknesses being 1. 2. The softness of the hafnium radiation is partly responsible for the low activities obtained after long exposure of hafnium oxide with radium-beryllium sources of a few hundred milliCurie, the  $\beta$ -radiation emitted being absorbed to an appreciable extent in the hafnium oxide sample itself. In the case of hafnium, as already mentioned, every place between the mass numbers 176 and 180 is occupied by a known stable isotope; the formation of the active hafnium isotope can, then, only be due to the process



On emitting  $\beta$ -rays according to the equation



the active hafnium isotope becomes the only stable isotope of tantalum known. Hafnium is thus partly converted into tantalum under the action of neutron bombardment, while, as shown by us previously, hafnium is formed under the action of neutrons on lutecium. It is quite possible that, under bombardment with a powerful stream of deuterium or of neutrons, further decay periods of hafnium will be discovered.

### The Effect of Neutron Bombardment on Scandium.

A few years ago we embarked on the investigation of the effect of neutron bombardment on scandium (4), (5), (6), chiefly in the hope of being able to prepare an artificial radioactive isotope of potassium and to get some information on the then not entirely elucidated nature of the natural radioactivity of potassium. We bombarded few grams of very



pure scandium oxide prepared by Prof. STERBA-BÖHM and used by Prof. HÖNIGSCHMIDT in his work on the atomic weight of scandium. After neutron bombardment the scandium oxide was dissolved in dilute hydrochloric acid and 100—150 mgm. of sodium chloride and the same amount of calcium oxide was added. The filtrate obtained after precipitation with carbonate-free ammonia was treated with oxalic acid and the calcium oxalate formed was removed. The sodium chloride which had been added to the solution of the scandium chloride compound was recovered after the removal of the ammonium chloride content of the last filtrate by evaporation. The activities of the three fractions, namely scandium oxide, sodium chloride, and calcium oxalate, were then determined. Only the two first preparations mentioned were found to be active. The activity of the scandium oxide decayed very slowly while the various sodium chloride fractions obtained in different experiments lost half of their slight activity within 10 and 18 hours. We had just finished the experiment mentioned when a note was published by FERMI and his collaborators(1) concerning the action of neutrons on potassium. They found that potassium captured neutrons by giving birth to a potassium isotope decaying with a half life of 16 hours. The values found by us for the period of the slight activity of different potassium preparations obtained from scandium showed a half life between 10 and 18 hours; we thought it justifiable, therefore, to identify the element found by us with that found by FERMI and his collaborators. The initial activities measured amounted usually to about 10 kicks/min. In one case, through the courtesy of the Medical Radium Station and Dr. J. C. JACOBSEN, we were loaned an unusually strong neutron source containing 600 milliCuries

radium-emanation. The decay curve obtained for potassium 42 in this experiment is seen from figure 2.

The activity found by us in the filtrate of scandium precipitate could only be that of  $^{42}\text{K}$ , as the presence of active impurities was excluded by the fact that the above mentioned

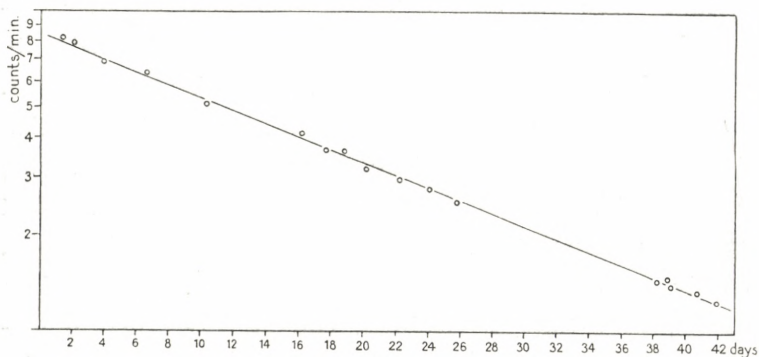
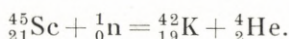


Fig. 2. Decay Curve of the Potassium Precipitate.

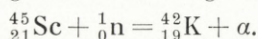
very pure scandium sample was used. The possibility that we measured the half life of radioactive sodium of 15 hours can be excluded with certainty not only for the reason mentioned above but also for the following reasons: Sodium,  $^{24}_{11}\text{Na}$ , can be prepared either from  $^{23}\text{Na}$  by simple neutron capture, or from magnesium if the capture is followed by emission of a proton, or from aluminium if the capture is followed by emission of an  $\alpha$ -particle. From the first mentioned process with the neutron sources at our disposal only very weak activities can be obtained even when starting with pure sodium. To prepare measurable amounts of radio-sodium from a few grams of impure scandium oxide an appreciable amount of magnesium or aluminium would have had to be present in the preparation. 15 mgm. of aluminium mixed with 150 mgm. of ammonium nitrate, for example, gave after activation

to saturation less than 0.5 kicks/min. and the activity obtained by similar amounts of magnesium with the same sources as used when activating scandium was still smaller. The amount of radio-sodium obtained from magnesium is less, and that of sodium by neutron capture very appreciably less than that obtained from aluminium. Scandium having just one stable isotope  ${}^{45}_{21}\text{Sc}$ , only the potassium isotope  ${}^{42}\text{K}$  can be produced under neutron bombardment according to the equation

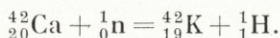


In the case of neutron capture by potassium, on the other hand, both reactions  ${}^{39}\text{K} + \text{n}$  and  ${}^{41}\text{K} + \text{n}$  can occur. While Fermi and his collaborators left it open which of the two last named potassium isotopes were produced, we could conclude from our experiments that the process witnessed by FERMI was  ${}^{41}\text{K} + \text{n} = {}^{42}\text{K}$ , and also that the process  ${}^{39}\text{K}$  very probably leads to the formation of the potassium isotope  ${}^{40}\text{K}$  which is responsible for the natural radioactivity of potassium. Recently WALKE (9), by making use of LAWRENCE'S powerful cyclotron, which supplies a many thousand times stronger neutron beam as obtained from our radium-beryllium sources, was able to follow the decay of  ${}^{42}\text{K}$  through ten periods and determined its half life period to be  $12.4 \pm 0.2$  hours, i. e., a somewhat lower value than that following from our figure 2.

Besides preparing  ${}^{42}\text{K}$  according to the equation



we succeeded (4) also in preparing this isotope by the process



Walke (10), while reproducing FERMI'S results and also ours as to the preparation of  ${}^{42}\text{K}$  from scandium, was

unable to reproduce our experiments in which  $^{42}\text{K}$  was prepared from calcium. This negative result induced us to repeat our experiments, this time by bombarding with fast neutrons as much as 1 kg of calcium carbonate. These were dissolved in a minimum amount of HCl, precipitated by a minimum amount of ammonium oxalate, which sufficed to precipitate all calcium after dissolving 100 mgm. of sodium chloride. Before we finished these experiments, a second paper of WALKE(12), (13), was published in which he described successful experiments in producing  $^{42}\text{K}$  from calcium, thus corroborating our statement.

### Activity of Scandium.

After the removal of the radio-potassium produced, the scandium was still showing a weak activity which could not be removed by chemical operations and which we, therefore, had to ascribe to a radioactive isotope of scandium. The decay of the weak activity of scandium observed for 240 days is seen from table 4, which shows the presence

Table 4.

Activity of a Scandium Sample after Separation of the Active Potassium.

Date	Nr. of Days	Kicks/min.
5. III. 36.....	0	$7.3 \pm 0.6$
21. III. 36.....	16	$6.0 \pm 0.4$
24. IV. 36.....	50	$7.2 \pm 0.4$
20. V. 36.....	76	$5.1 \pm 0.4$
5. VI. 36.....	92	$5.7 \pm 0.4$
1. VII. 36.....	117	$4.9 \pm 0.3$
13.VIII.36.....	161	$5.3 \pm 0.4$
24. X. 36.....	233	$4.4 \pm 0.3$

of a very weak activity decaying with a period longer than a year. We could not follow up this very weak activity further but concentrated our interest on another period ob-

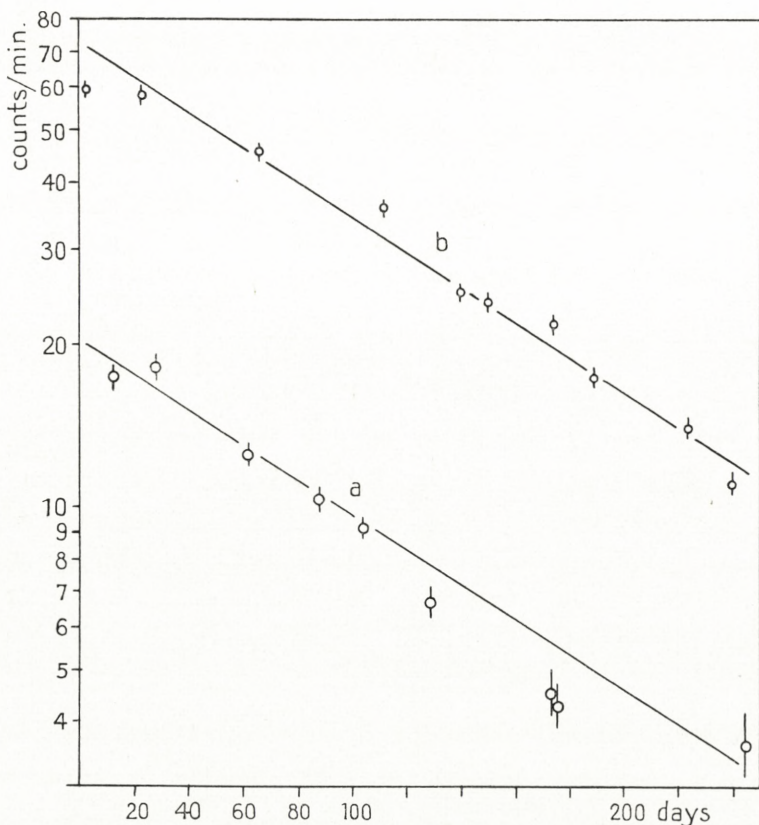


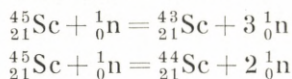
Fig. 3. Decay Curve of Scandium Irradiated in a Paraffin Block for a) 24 Days b) 50 Days.

tained after activating for 24 days in a paraffin block which contained emanation-beryllium sources of an average strength of 50 milliCuries. The result obtained is seen from figure 3a; the half life works out to be  $90 \pm 5$  days.

In the next set of experiments we activated simultane-

ously three scandium preparations for 50 days with radium-beryllium sources of a strength of about 200 milliCuries: one in the usual way inside the paraffin block, the second one in a paraffin block but with the preparation surrounded by a shield of cadmium, which absorbed nearly 100 % of the C-neutrons, and the third one with fast neutrons. The result of the activation of the first named sample is seen from figure 3b. The investigation of the second sample led to the result that in the presence of cadmium the artificial radioactivity of scandium is reduced to 2 % of the value obtained in the absence of cadmium. From fig. 3b there follows for the half life period of scandium the value  $90 \pm 5$  days. Quite recently Walke(13), by making use of LAWRENCE'S powerful cyclotron, bombarded scandium with deuterons and obtained a period of  $85 \pm 2$  days.

We want furthermore to mention an early experiment in which we bombarded scandium with fast neutrons emitted by a mixture of 600 milliCuries emanation and beryllium powder; we observed a period of decay of about 50 hours. As shown by POOL, CORK, and THORNTON(8), and by WALKE(13), under bombardment with fast neutrons the following two reactions occur as well:



the scandium isotopes obtained emit positrons and have half lives of 4 and 43 hours respectively; it was presumably the last mentioned reaction which we observed.

Scandium 44 was also produced(13) by the action of  $\alpha$ -particles on potassium 41 and(11) by the action of deuterons on calcium 43, while scandium 43 was produced (11) by the action of  $\alpha$ -particles on calcium 40, and(2) by the action of deuterons on calcium 42.

WALKE was furthermore successful in producing scandium 42 under the action of  $\alpha$ -particles on potassium 39, and of scandium 41 under the action of deuterons on calcium. The list of the known radioactive isotopes of scandium is therefore as follows from table 5.

Table 5.  
Active Isotopes of Scandium.  
(According to Walke(11)).

Active Isotope	Particle Emitted	Half life
$^{41}\text{Sc}$	positron	53 min.
$^{42}\text{Sc}$	positron	4.1 h
$^{43}\text{Sc}$	positron	4.0 h
$^{44}\text{Sc}$	positron	52 h
$^{46}\text{Sc}$	electron	85 days and possibly also a period of about 1 year.

We measured the reduction of counts when covering an active scandium preparation decaying with a period of 90 days with aluminium foils of varying thickness. The result obtained can be seen from table 6.

Table 6.  
Absorption in Aluminium of the  $\beta$ -rays Emitted  
by Scandium.

Thickness of the Al-foil	Kicks/min.
1st experiment { 0	12.5
5.5 mgm./cm <sup>2</sup>	8.8
11.0 mgm./cm <sup>2</sup>	7.1
2nd experiment { 0	61.8
11.0 mgm./cm <sup>2</sup>	35.3
16.5 mgm./cm <sup>2</sup>	25.8

(half value thickness:  $13 \pm 1$  mgm./cm<sup>2</sup>)

In view of the softness of the  $\beta$ -rays emitted we used very thin scandium oxide layers; about 50 mgm./cm<sup>2</sup>. In spite of the thin layers used the soft components were absorbed in the preparation to an appreciably greater extent than were the hard ones; accordingly we have to reckon with the possibility that the radiation emitted by scandium is still softer than indicated by the figures of table 6.

### The Radioactivity of Europium and its Analytical Application.

In their fundamental research on the action of neutrons Fermi and his collaborators(1) investigated also the activity of a gadolinium preparation bombarded by neutrons and found an activity decaying with a period of 8 hours. A few years later, SUGDEN(7), investigating the radioactivity of europium, discovered a very strong activity decaying with a period of 9.2 h. and, at that time, interpreted the above mentioned period of decay of gadolinium to be due to the presence of some europium in the sample investigated. Investigations carried out by us, in which we made use of different gadolinium samples prepared by Prof. ROLLA and partly by Prof. PRANDTL and the late Baron AUER v. WELSBACH, confirmed completely the conclusion arrived at by SUGDEN, and this induced us to make use of the radioactivity of europium produced under the action of neutrons to determine the amount of europium present in gadolinium preparations. Prof. ROLLA, being engaged in the preparation of large amounts of pure gadolinium compounds, sent us several samples the europium content of which he wished ascertained. We describe in the following the analytical procedure used by us.



Thin layers of the gadolinium oxide samples to be investigated were fixed between two glass plates and placed within a paraffin block. Usually we investigated simultaneously the activation of 4 symmetrically placed preparations. It is of importance to bombard layers having the same thickness and to bombard them with neutrons in such a

Table 7.

Activity of Different Gadolinium Preparations. (The sample labelled „Standard” is the  $Gd_2O_3$  to which 2%  $Eu_2O_3$  was added: samples 1—4 represent progressive stages in the purification process carried out by Prof. Rolla).

Samples	kick/min.
“Standard”	125
1	60
2	60
3	30
4	25

way that each preparation is hit by the same number of neutrons; the latter was achieved by arranging the sources in the block circularly. We used in these experiments radium-beryllium sources containing 600 mgm. of radium, the neutron emission of which corresponds to that from about 400 milliCuries of radium emanation; in addition a beryllium-emanation mixture containing 300 milliCuries emanation was also present. After the lapse of 3 days an average mixture of each sample was placed in a small aluminium dish having a surface of  $1.2 \text{ cm}^2$  and was put below the window of a Geiger-counter. The intensity of the activity of the different gadolinium samples investigated is proportional to the europium content. In order to arrive at a

figure stating the absolute concentration, we mixed 2 % of europium oxide with one gadolinium preparation denoted as sample 2 in table 7 and compared the activity of the latter with that of the gadolinium preparations of unknown europium content. The results are seen from table 7.

That, in spite of the large amount of radium and emanation used, the activities measured were not stronger is due to the high absorbing power of gadolinium, which reduces the density of thermal neutrons in the surroundings of the preparation. This effect is especially marked on account of the fact that the thermal neutrons diffuse and are likely to pass through the preparation several times. The latter effect can be best estimated by comparing the activity of pure europium oxide with that obtained when this material is embedded in gadolinium oxide. We activated simultaneously 200 mgm. of europium oxide and 200 mgm. of gadolinium oxide containing 2 % of europium. If gadolinium absorbed to the same extent as europium, the first named preparation should be 50 times more active than the last mentioned one. Actually we find the ratio to be 200 from which it follows that the presence of gadolinium in our preparations reduced the activity of europium to  $\frac{1}{4}$  of the value which would have been obtained if the same amount of pure europium oxide had been subjected to the irradiation.

---

### Summary.

---

The irradiation of hafnium with neutrons has been shown to produce a radioactivity with a half life of  $55 \pm 7$  days and which may be ascribed to  $^{181}_{72}\text{Hf}$ . The intensity of the  $\beta$ -rays emitted is reduced to half of its initial value by an aluminium foil having a weight of 16 mgm./cm<sup>2</sup>.

Scandium,  $^{46}_{21}\text{Sc}$ , was found to decay with a half life of  $90 \pm 5$  days. The half value thickness for the absorption in aluminium of the  $\beta$ -rays from this element was found to be 13 mgm./cm<sup>2</sup>.

The europium content of gadolinium oxide samples prepared by Professor LUIGI ROLLA was determined by making use of the artificial radioactivity produced under the action of neutrons on the europium present in the various samples.

We would like to express our best thanks to Professor NIELS BOHR for the kind interest he has taken in this work and to Professor STERBA-BÖHM for kindly lending us the scandium oxide used in the experiments.

Institute of theoretical Physics, Copenhagen.

---

## REFERENCES

---

- (1) E. AMALDI, E. FERMI and others: Proc. Roy. Soc. A. **149**, 522, 1935.
  - (2) O. R. FRISCH: Nature **136**, 220, 1935.
  - (3) G. HEVESY: Kgl. Danske Vid. Selsk. Math.-fys. Medd. VI, 7, S. 91, 1925.
  - (4) G. HEVESY and HILDE LEVI, Nature **135**, 580, 1935.
  - (5) G. HEVESY: Kgl. Danske Vid. Selsk. Math.-fys. Medd. XIII, 3, 1935.
  - (6) G. HEVESY and HILDE LEVI: Kgl. Danske Vid. Selsk. Math.-fys. Medd. XIV, 5, 1936.
  - (7) S. SUGDEN: Nature **135**, 469, 1935.
  - (8) M. L. POOL, J. M. CORK and R. L. THORNTON: Phys. Rev. **52**, 41, 1937.
  - (9) H. WALKE: Phys. Rev. **51**, 439, 1937.
  - (10) D. G. HURST and H. WALKE: Phys. Rev. **51**, 1033, 1937.
  - (11) H. WALKE: Phys. Rev. **52**, 400, 1937.
  - (12) H. WALKE: Phys. Rev. **52**, 663—669, 1937.
  - (13) M. ZYW: Nature **134**, 64, 1934.
-

Det Kgl. Danske Videnskabernes Selskab.

Mathematisk-fysiske Meddelelser. **XV**, 12.

---

ALGEBRAIC EQUATIONS WITH  
ALMOST-PERIODIC COEFFICIENTS

BY

HARALD BOHR AND DONALD A. FLANDERS



KØBENHAVN

LEVIN & MUNKSGAARD

EJNAR MUNKSGAARD

1937

Printed in Denmark.  
Bianco Lunos Bogtrykkeri A/S.

DEDICATED TO  
PROFESSOR RICHARD COURANT  
ON HIS FIFTIETH BIRTHDAY

## INTRODUCTION

It is well known that, generally speaking, continuous one-valued functions of one or more almost-periodic functions lead again to almost-periodic functions. Also it has been shown that in various situations where problems with almost-periodic data give rise to many-valued solutions, these solutions are themselves almost-periodic, or at any rate show interesting almost-periodic features. The case of algebraic functions was first treated by WALTHER<sup>1)</sup>, who proved that the solutions of an algebraic equation whose coefficients are complex almost-periodic functions of a real variable  $t$  are always almost-periodic functions, provided the discriminant  $D(t)$  of the equation not merely is different from zero for all  $t$  but actually has a positive number as the greatest lower bound of its absolute value. By means of well-established relations between the translation numbers and the Fourier exponents of an almost-periodic function one can easily obtain from WALTHER's proof some first results on the connection between the Fourier exponents of the solutions of the equation and the Fourier exponents of the coefficients.

The latter result was sharpened by CAMERON<sup>2)</sup> in an interesting paper dealing with the general question of implicitly given almost-periodic functions. CAMERON also answered in

1) A. WALTHER, Algebraische Funktionen von fastperiodischen Funktionen. Monatshefte für Mathematik und Physik. Bd. 40, 1933, p. 444—457.

2) R. H. CAMERON, Implicit Functions of Almost-periodic Functions. Bulletin of the American Mathematical Society. 1934, p. 895—904.

the negative the question left open by WALTHER, whether the restriction on the discriminant that  $GLB |D(t)| > 0$  could be replaced by the weaker condition  $D(t) \neq 0$ . In fact he stated that there exists an almost-periodic function  $f(t)$  which, while  $f(t) \neq 0$ , has  $GLB |f(t)| = 0$ , and which has the property that the two continuous roots  $\pm \sqrt{f(t)}$  of the equation  $y^2 - f(t) = 0$  (with discriminant  $4f(t)$ ) are not almost-periodic<sup>1)</sup>.

---

In the present paper we investigate systematically the problem of the almost-periodic solutions of an algebraic equation whose coefficients are almost-periodic functions of a real variable and whose discriminant is not near zero. We shall not make use of the previous investigations quoted above, but start afresh; the former results will naturally present themselves in the course of our investigation. The problem will be studied both from an analytical and from an algebraic point of view. That the latter to some degree predominates is simply due to the fact that a certain Abelian substitution group, the "almost-translation group" of the roots of the equation, presents itself as a natural basis for any thorough-going discussion of the problem and turns out to have a fundamental influence on the structure of the solutions. A principal result of our paper is the fact that any arbitrarily given transitive Abelian substitution group can occur as the almost-translation group of the roots of an algebraic equation with almost-periodic coefficients. Thus the solutions of an algebraic equation present a much more rich and varied aspect when the coefficients are general al-

<sup>1)</sup> Since CAMERON did not indicate explicitly the construction of such a function, the present authors have constructed an example which presumably follows the same lines.



most-periodic functions than in the classical case when the coefficients are pure periodic functions with a common period, which leads to only cyclic groups.

The paper is divided into seven sections. In section I we set down, in such form as will be most convenient for our later applications, some familiar facts concerning Abelian substitution groups and their character groups. In section II we introduce the notion of almost-translation substitution and almost-translation group, basing our considerations on an arbitrary finite set of complex functions. In section III we apply the notions and results of section II to such sets of functions when they form the roots of an algebraic equation with almost-periodic coefficients. Section IV (like section I) is of auxiliary character, and assembles some well known important relations between the translation numbers and the Fourier exponents of one or more almost-periodic functions. In section V we apply these relations to our present problem. The principal contribution of the paper is found in section VI, where we deduce necessary and sufficient conditions that a given finite set of almost-periodic functions shall have as its almost-translation group an arbitrarily given transitive Abelian substitution group. In these conditions the characters of the group play a predominant role; and by means of the characters we are enabled to give a certain canonical representation of the functions considered. We also give general examples — which from various aspects may be said to be the simplest, as well — of sets of functions with arbitrary transitive almost-translation groups. The paper is concluded in section VII by the application of the results of section VI to the original algebraic problem.

---

## I. Preliminary Remarks on Abelian Substitution Groups.

We shall be concerned throughout with substitutions which permute the  $m$  objects in a finite set. If the objects are distinguished by attaching the indices  $1, \dots, m$  to a fixed symbol, we shall denote the set by enclosing the fixed symbol in  $[\ ]$ 's: e. g.  $[a]$  denotes the set composed of  $a_1, \dots, a_m$ . Where no ambiguity arises we shall not distinguish between a substitution operating on the elements of  $[a]$ , and the corresponding substitution operating on the indices. Thus we shall customarily denote the substitution  $S$  which takes  $a_1$  into  $a_{g_1}, \dots, a_m$  into  $a_{g_m}$ , by  $\begin{pmatrix} 1, \dots, m \\ g_1, \dots, g_m \end{pmatrix}$ . We shall indicate the effect of  $S$  on any  $a_h$  by writing  $a_{g_h} = Sa_h$ . The substitution resulting from operating first with  $S_1$  and then with  $S_2$  will be denoted by  $S_2S_1$ . Finally, a relation which holds between each  $a_h$  and its corresponding  $Sa_h$  will frequently be denoted by enclosing a specimen relation in  $[\ ]$ 's. Thus if, as often in the later sections, the  $a$ 's represent numbers such that  $|a_h - Sa_h| \leq \varepsilon$  for  $h = 1, \dots, m$ , we shall write simply  $[|a_h - Sa_h| \leq \varepsilon]$ .

Let  $\Gamma$  be an Abelian group of substitutions on  $[a]$  (or on the indices of  $[a]$ ). A generating system of  $\Gamma$  is a set of elements of  $\Gamma$ , say  $S'_1, \dots, S'_q$ , in terms of which every element  $S$  of  $\Gamma$  can be expressed as a power product, as  $S = (S'_1)^{e_1} \dots (S'_q)^{e_q}$ . (If this representation is unique in the sense that each factor  $(S'_i)^{e_i}$  is uniquely determined

by  $S$ , the generating system is called a basis of  $\Gamma$ ). A generating system which contains the smallest possible number, say  $\mu$ , of elements  $S$  is called a minimal generating system. This number  $\mu$ , which is characteristic of the group, has a property which is particularly important for our purposes, namely that out of every generating system  $S'_1, \dots, S'_q$  there can always be chosen  $\mu$  elements, say  $S'_1, \dots, S'_\mu$ , such that ( $E$  denoting the identity element) for every relation  $(S'_1)^{e_1} \dots (S'_\mu)^{e_\mu} = E$  between these elements, we have  $GCD(e_1, \dots, e_\mu) > 1$ .

Of special importance for our investigation is the case where the Abelian substitution group  $\Gamma$  is transitive. Then  $\Gamma$  is of order  $m$ , and each substitution of  $\Gamma$  is uniquely determined by specifying its effect on any one element of  $[a]$ . Hence we can always assign the indices  $1, \dots, m$  to the objects  $a$  and the substitutions  $S$  in such a way that  $S_h a_1 = a_h$ , so that  $S_1 = \begin{pmatrix} 1, \dots \\ 1, \dots \end{pmatrix}, \dots, S_m = \begin{pmatrix} 1, \dots \\ m, \dots \end{pmatrix}$ . Such a concordant indexing of  $[a]$  and  $\Gamma$  has the readily verifiable advantage that the effect of multiplying every element of  $\Gamma$  by any fixed element  $S_h$  is to perform on the elements of  $\Gamma$  a substitution whose symbol (in terms of the indices) is identical with that of  $S_h$ . Thus, if  $S_h S_1 = S_{h_1}, \dots, S_h S_m = S_{h_m}$ , then  $S_{h_1}, \dots, S_{h_m}$  is a permutation of  $S_1, \dots, S_m$ , and  $\begin{pmatrix} 1, \dots, m \\ h_1, \dots, h_m \end{pmatrix}$ , interpreted as a substitution on  $[a]$ , is precisely  $S_h$ .

Since  $\Gamma$  is Abelian we know that a complete set of characters of  $\Gamma$  forms (with respect to ordinary multiplication) a group  $\Gamma'$  isomorphic with  $\Gamma$ . We may then denote the elements of the character group by  $\chi_1(S), \dots, \chi_m(S)$ . (We shall denote the principal character by its value, 1).

The notions of generating system and minimal generating system apply to  $T^*$  as they did to  $T$ , but with certain additional features due to the connection between the two groups. Thus, we shall find useful for our later applications the following well known simple criterion for a generating system of  $T^*$ : the characters  $\chi'_1, \dots, \chi'_q$  form a generating system of  $T^*$  if and only if  $\chi'_1(S_h) = \chi'_1(S_g), \dots, \chi'_q(S_h) = \chi'_q(S_g)$  implies  $S_h = S_g$ . Also, since  $T^*$  is isomorphic with  $T$ , a minimal generating system of  $T^*$  has the same number  $\mu$  of elements as a minimal generating system of  $T$ . Further, out of any generating system  $\chi'_1, \dots, \chi'_q$  can be chosen  $\mu$  characters, say  $\chi'_1, \dots, \chi'_\mu$ , such that every relation  $(\chi'_1)^{e_1} \dots (\chi'_\mu)^{e_\mu} = 1$  implies that  $GCD(e_1, \dots, e_\mu) > 1$ . Finally, since  $T$  is transitive, and since by definition  $\chi(S_h)\chi(S_g) = \chi(S_h S_g)$ , it follows from the previously noted result of indexing the elements of  $T$  and  $[a]$  concordantly: If  $\chi(S)$  is any character and  $S_{h_1}, \dots, S_{h_m}$  is just that permutation of  $S_1, \dots, S_m$  produced by applying to the indices of the  $S$ 's the substitution which denotes  $S_h$ , then

$$\chi(S_h)\chi(S_1) = \chi(S_{h_1}), \dots, \chi(S_h)\chi(S_m) = \chi(S_{h_m}).$$

We conclude this section with some remarks on those powers of the substitutions of a (not necessarily transitive) group  $T$  which leave a given element of  $[a]$  unaltered. For fixed element  $a_h$  we may regard the relation  $S^e a_h = a_h$  as an equation in the variable substitution  $S$ , whose range is the group  $T$ . We denote by  $\nu_h$  the least positive integer  $e$  for which this relation is an identity in  $S$ . Trivially,  $\nu_h$  is not greater than the order  $g$  of the group, since then  $S^g a_h = E a_h = a_h$ , not only for every  $S$  in  $T$ , but for every  $a_h$  in  $[a]$ . Equally trivially,  $\nu_h \leq m!$ , since the set of all

substitutions on  $[a]$  forms the symmetric group  $\Sigma_m$  of order  $m!$ . If  $T$  is Abelian and transitive, and hence of order  $m$ , we have the sharper result that  $\nu_h$  is  $\leq m$ ; and in fact  $\nu_h$  is independent of  $h$  and equal to  $m'$ , the least common multiple of the orders of the elements of  $T$ . If the Abelian group  $T$  is intransitive, each  $a_h$  belongs to a transitivity set containing, say,  $m_h$  elements; and  $T$  is homomorphic with a transitive Abelian group  $T'$  operating on this subset and hence of order  $m_h$ . In this case  $\nu_h = m'_h \leq m_h$ , where  $m'_h$  is the least common multiple of the orders of the substitutions in  $T'$ .

## II. Almost-Translation Substitutions and the Almost-Translation Group.

In this section we suppose that the  $m$  elements of the finite set  $[a]$  upon which substitutions are to be performed, are distinct one-valued complex functions,  $f_1(t), \dots, f_m(t)$ , of a real variable, defined for  $-\infty < t < +\infty$ . In accordance with our previous notation the set is then denoted by  $[f(t)]$ .

We first assume only that each function is continuous for all values of  $t$ .

*Definition.* For given  $\varepsilon > 0$  we shall say of a real number  $\tau$  that it  $\varepsilon$ -performs the substitution  $S$  on  $[f(t)]$  if

$$|f_h(t+\tau) - Sf_h(t)| \leq \varepsilon \text{ for } -\infty < t < +\infty.$$

We shall denote by  $\{\tau_{(S)}(\varepsilon)\}$  the set of all real numbers  $\tau$  each of which  $\varepsilon$ -performs the fixed substitution  $S$  on  $[f(t)]$ ; and by  $\{\tau_{[f]}(\varepsilon)\}$  the set of all real numbers  $\tau$  each of which  $\varepsilon$ -performs some substitution on  $[f(t)]$ .

Remark. As the  $m$  functions are distinct from each other, it is clear (quite apart from any consideration of the continuity of the functions) that for sufficiently small positive  $\varepsilon$ , no  $\tau$  can  $\varepsilon$ -perform more than one substitution on  $[f(t)]$ . In other words, there exists  $\varepsilon^* > 0$  such that if  $0 < \varepsilon \leq \varepsilon^*$  and  $\tau$  belongs to both  $\{\tau_{(S_1)}(\varepsilon)\}$  and  $\{\tau_{(S_2)}(\varepsilon)\}$ , then  $S_1 = S_2$ .

Lemma. If  $\tau_1 \in \{\tau_{(S_1)}(\varepsilon_1)\}^1$  and  $\tau_2 \in \{\tau_{(S_2)}(\varepsilon_2)\}$ , then  $(\tau_1 + \tau_2) \in \{\tau_{(S_2 S_1)}(\varepsilon_1 + \varepsilon_2)\}$  and  $(\tau_1 + \tau_2) \in \{\tau_{(S_1 S_2)}(\varepsilon_1 + \varepsilon_2)\}$ .

Proof. By assumption  $[|f_h(t + \tau_1) - S_1 f_h(t)| \leq \varepsilon_1]$ . Replacing  $t$  by  $t + \tau_2$ ,

$$(1) \quad [ |f_h(t + \tau_2 + \tau_1) - S_1 f_h(t + \tau_2)| \leq \varepsilon_1 ].$$

Also, by assumption  $[|f_h(t + \tau_2) - S_2 f_h(t)| \leq \varepsilon_2]$ , whence, replacing  $f_h$  by  $S_1 f_h$ ,

$$(2) \quad [ |S_1 f_h(t + \tau_2) - S_2 S_1 f_h(t)| \leq \varepsilon_2 ].$$

From (1) and (2) follows

$$[ |f_h(t + \tau_2 + \tau_1) - S_2 S_1 f_h(t)| \leq \varepsilon_1 + \varepsilon_2 ],$$

which says that  $(\tau_2 + \tau_1) \in \{\tau_{(S_2 S_1)}(\varepsilon_1 + \varepsilon_2)\}$ . The other half of the lemma is derived similarly (or simply by interchanging the indices 1 and 2).

Corollary. If  $\tau \in \{\tau_{(S)}(\varepsilon)\}$  and  $\nu$  is any positive integer, then  $\nu\tau \in \{\tau_{(S^\nu)}(\nu\varepsilon)\}$ , or<sup>2)</sup>

$$\nu \{\tau_{(S)}(\varepsilon)\} \subseteq \{\tau_{(S^\nu)}(\nu\varepsilon)\}.$$

1) The symbol  $\varepsilon$  denotes the relation "belonging to", while the symbol  $\varepsilon$  is used for positive numbers.

2) For fixed real  $\nu$  we shall denote by  $\nu\{\tau\}$  the set of all numbers  $\nu\tau$ .

**Definition.** If  $S$  is any substitution on  $[f(t)]$  such that  $\{\tau_{(S)}(\varepsilon)\}$  is relatively dense for every positive  $\varepsilon$ , we shall say that  $S$  is an almost-translation substitution on  $[f(t)]$ , or that  $f(t)$  admits  $S$  as an almost-translation substitution.

**Theorem 1.** The set  $\Gamma$  of all almost-translation substitutions admitted by  $[f(t)]$  is either vacuous or an Abelian substitution group (which we shall call the almost-translation group of  $[f(t)]$ ).

**Proof.** If  $\Gamma$  is not empty, let  $S_1$  and  $S_2$  be any substitutions in  $\Gamma$ . For any  $\varepsilon > 0$  take arbitrary  $\tau_1 \in \left\{ \tau_{(S_1)}\left(\frac{\varepsilon}{2}\right) \right\}$  and  $\tau_2 \in \left\{ \tau_{(S_2)}\left(\frac{\varepsilon}{2}\right) \right\}$ . If we put  $\varepsilon_1 = \varepsilon_2 = \frac{\varepsilon}{2}$  in the lemma above, we see that  $\tau_1 + \tau_2$  belongs to both  $\{\tau_{(S_2 S_1)}(\varepsilon)\}$  and  $\{\tau_{(S_1 S_2)}(\varepsilon)\}$ . But these relations hold for every positive  $\varepsilon$  and corresponding  $\tau_1, \tau_2$ ; in particular, when  $\varepsilon \leq \varepsilon^*$ , which requires that  $S_2 S_1 = S_1 S_2$ . Thus the product of any two substitutions in  $\Gamma$  is commutative. Furthermore, since the sets  $\left\{ \tau_{(S_1)}\left(\frac{\varepsilon}{2}\right) \right\}$  and  $\left\{ \tau_{(S_2)}\left(\frac{\varepsilon}{2}\right) \right\}$  are relatively dense for every  $\varepsilon > 0$  (actually we only need to have one set relatively dense and the other not empty), the set of all sums  $\tau_1 + \tau_2$ , and a fortiori  $\{\tau_{(S_2 S_1)}(\varepsilon)\}$ , is relatively dense for all positive  $\varepsilon$ 's. Thus  $S_2 S_1 = S_1 S_2$  belongs to  $\Gamma$  (which is a subset of the finite group  $\Sigma_m$ ), so  $\Gamma$  is an Abelian group.

**Corollary.** If  $[f(t)]$  has an almost-translation group  $\Gamma$ , then each function  $f_h(t)$  is almost-periodic.

**Proof.** The identity-substitution  $E$  must be in  $\Gamma$ ; i. e. every positive  $\varepsilon$  determines a relatively dense set  $\{\tau_{(E)}(\varepsilon)\}$  of real numbers  $\tau$  for each of which

$$[|f_h(t + \tau) - E f_h(t)| \leq \varepsilon].$$

Since  $E f_h = f_h$ , this simply says that for each fixed  $h$ , every  $\tau_{(E)}(\varepsilon)$  is a translation-number  $\tau_{f_h}(\varepsilon)$  of  $f_h(t)$ <sup>1)</sup>.

**Theorem 2.** *A necessary and sufficient condition that  $[f(t)]$  admit an almost-translation substitution (and therefore have an almost-translation group  $\Gamma$ ) is that the set  $\{\tau_{[f]}(\varepsilon)\}$  be relatively dense for each  $\varepsilon > 0$ .*

*Proof.* The necessity is immediate. To prove the sufficiency we show that the identity  $E$  is actually admitted by  $[f(t)]$  as an almost-translation substitution. By assumption the set  $\left\{\tau_{[f]}\left(\frac{\varepsilon}{m!}\right)\right\}$  is relatively dense, and to each of the numbers  $\tau = \tau_{[f]}\left(\frac{\varepsilon}{m!}\right)$  there corresponds a substitution  $S$  (depending on  $\tau$ ) such that  $\tau \in \left\{\tau_{(S)}\left(\frac{\varepsilon}{m!}\right)\right\}$ . Hence  $m! \tau$  belongs to  $\{\tau_{(S^{m!})}(\varepsilon)\}$ , i. e. to  $\{\tau_{(E)}(\varepsilon)\}$ . Thus the set  $\{\tau_{(E)}(\varepsilon)\}$  contains the relatively dense set  $m! \left\{\tau_{[f]}\left(\frac{\varepsilon}{m!}\right)\right\}$ , and is itself relatively dense.

For the remainder of this section we suppose that we have to do with a finite set  $[f(t)]$  of functions, each of which is almost-periodic. By the above corollary, this condition is automatically fulfilled when the set has an almost-translation group. We collect here some remarks concerning this case which will be useful in later sections;

<sup>1)</sup> In accordance with the usage prevailing in the literature we shall, throughout the paper, denote by  $\tau_f(\varepsilon)$  a translation-number (corresponding to  $\varepsilon$ ) of a single function  $f(t)$ . We have been careful to differentiate from this the two other symbols of similar type introduced here, viz.  $\tau_{(S)}(\varepsilon)$  and  $\tau_{[f]}(\varepsilon)$ , each of which denotes a number performing an office similar to that of the translation-number of a single function, but in connection with a finite set of functions.



in particular we shall see that the converse of the above corollary is also true, that is that the almost-periodicity of the separate functions  $f_h(t)$  implies the existence of an almost-translation group of the set  $[f(t)]$ .

1°. A familiar and important property of such a set is the fact that each  $\varepsilon > 0$  determines a relatively dense set of real numbers  $\tau$ , each of which is a translation number for every one of the functions  $f_h(t)$ , so that for  $1 \leq h \leq m$ ,

$$|f_h(t + \tau) - f_h(t)| \leq \varepsilon \text{ for } -\infty < t < +\infty.$$

Even more, as BOCHNER has shown, there exists an almost-periodic function  $F(t)$ , called a majorant of the set, such that for every positive  $\varepsilon$  the set  $\{\tau_F(\varepsilon)\}$  of its translation-numbers is identical with the set-theoretical product of the sets  $\{\tau_{f_1}(\varepsilon)\}, \dots, \{\tau_{f_m}(\varepsilon)\}$  of translation-numbers of the individual functions.

2°. Our set  $[f(t)]$  certainly admits the identity  $E$  as an almost-translation substitution, and hence has an almost-translation group  $\Gamma$  (which may consist of  $E$  only). In fact, for every  $\varepsilon > 0$ , the set  $\{\tau_F(\varepsilon)\}$  determined in 1° is precisely the set  $\{\tau_{(E)}(\varepsilon)\}$ .

3°. If  $S$  is such a substitution of the whole group  $\Sigma_m$  that for some  $\varepsilon > 0$  the set  $\{\tau_{(S)}(\varepsilon)\}$  is not empty, then the set  $\{\tau_{(S)}(2\varepsilon)\}$  is relatively dense. For if  $\tau_0$  is some number in  $\{\tau_{(S)}(\varepsilon)\}$  and  $\tau \in \{\tau_{(E)}(\varepsilon)\}$ , then  $(\tau_0 + \tau) \in \{\tau_{(S)}(2\varepsilon)\}$ , and this set is therefore relatively dense since  $\{\tau_{(E)}(\varepsilon)\}$  is relatively dense.

4°. There exists a fixed number  $\varepsilon^{**} > 0$  with the property that any substitution  $S$  of  $\Sigma_m$  belongs to  $\Gamma$  provided only that for some positive  $\varepsilon \leq \varepsilon^{**}$ , the set  $\{\tau_{(S)}(\varepsilon)\}$  is not empty. For let us denote by  $\mathcal{A}(\varepsilon)$  the (possibly vacuous,

certainly finite) set of substitutions belonging to  $\Sigma_m$  but not to  $\Gamma$ , each of which is  $\varepsilon$ -performed by some real  $\tau$ . It is evident that  $0 < \varepsilon_1 < \varepsilon_2$  implies  $\mathcal{A}(\varepsilon_1) \subseteq \mathcal{A}(\varepsilon_2)$ , and that this, combined with the finiteness of every  $\mathcal{A}(\varepsilon)$ , implies that for some positive  $\varepsilon^{**}$ ,  $0 < \varepsilon_1 \leq \varepsilon^{**} < \varepsilon_2$  gives us  $\mathcal{A}(\varepsilon_1) \equiv \mathcal{A}(\varepsilon^{**}) \subseteq \mathcal{A}(\varepsilon_2)$ . If  $\mathcal{A}(\varepsilon^{**})$  actually contained a substitution  $S$ , for this  $S$  every  $\{\tau_{(S)}(\varepsilon)\}$  would be non-vacuous, and by 3° every  $\{\tau_{(S)}(2\varepsilon)\}$  would be relatively dense.  $S$  would then by definition belong to  $\Gamma$ , contrary to hypothesis. Thus, if for some positive  $\varepsilon \leq \varepsilon^{**}$ , a substitution  $S$  of  $\Sigma_m$  is  $\varepsilon$ -performed by even one  $\tau$ , then that substitution belongs to  $\Gamma$ .

5°. Let  $f_h(t)$  be any fixed member of  $[f(t)]$ , and  $\nu_h$  the corresponding positive integer (defined in section I) for which  $S^{\nu_h} f_h(t) \equiv f_h(t)$  for every  $S$  in  $\Gamma$ . Then for each sufficiently small  $\varepsilon$  (in fact for  $\varepsilon \leq \nu_h \varepsilon^{**}$ ),

$$\{\tau_{f_h}(\varepsilon)\} \supseteq \nu_h \left\{ \tau_{[f]} \left( \frac{\varepsilon}{\nu_h} \right) \right\}.$$

For, for each  $S$  in  $\Gamma$ ,  $\{\tau_{f_h}(\varepsilon)\} \supseteq \{\tau_{(S^{\nu_h})}(\varepsilon)\}$ , and  $\{\tau_{(S^{\nu_h})}(\varepsilon)\} \supseteq \nu_h \left\{ \tau_{(S)} \left( \frac{\varepsilon}{\nu_h} \right) \right\}$ , and hence

$$\{\tau_{f_h}(\varepsilon)\} \supseteq \nu_h \left\{ \tau_{(S)} \left( \frac{\varepsilon}{\nu_h} \right) \right\}.$$

But the last relation, just proved for each  $S$  in  $\Gamma$ , holds for trivial reasons for every  $S$  in  $\Sigma_m$  which does not belong to  $\Gamma$ , since for such an  $S$  the set  $\left\{ \tau_{(S)} \left( \frac{\varepsilon}{\nu_h} \right) \right\}$  is empty on account of 4° (as  $\frac{\varepsilon}{\nu_h} \leq \varepsilon^{**}$ ). Now the set-theoretical sum, taken over all the  $S$ 's in  $\Sigma_m$ , of the sets  $\left\{ \tau_{(S)} \left( \frac{\varepsilon}{\nu_h} \right) \right\}$ , is exactly the set  $\left\{ \tau_{[f]} \left( \frac{\varepsilon}{\nu_h} \right) \right\}$ ; so  $\{\tau_{f_h}(\varepsilon)\} \supseteq \nu_h \left\{ \tau_{[f]} \left( \frac{\varepsilon}{\nu_h} \right) \right\}$ , and the proof is completed.

### III. The Existence of an Almost-Translation Group for the Solutions of an Algebraic Equation with Almost-Periodic Coefficients.

Let

$$(3) \quad y^m + x_1(t)y^{m-1} + \dots + x_{m-1}(t)y + x_m(t) = 0$$

be an algebraic equation of degree  $m$  in the complex variable  $y$ , with leading coefficient 1 and remaining coefficients,  $x_1(t), \dots, x_m(t)$ , almost-periodic functions of the real variable  $t$ . We denote the discriminant of the equation (which, by the way, is also an almost-periodic function) by

$$D(t) = d[x_1(t), \dots, x_m(t)].$$

If  $D(t) \neq 0$  for  $-\infty < t < +\infty$ , the  $m$  roots of the equation are distinct for every value of  $t$ , and since the coefficients are continuous, these roots may be sorted in just one way (except for choice of notation) into  $m$  one-valued functions,  $y_1(t), \dots, y_m(t)$ , each of which is continuous for all values of  $t$ . Further, as the coefficients  $x_j(t)$  are bounded and uniformly continuous, the roots  $y_h(t)$  will also be bounded and uniformly continuous in  $-\infty < t < +\infty$ .

We now assume not merely that  $D(t) \neq 0$ , but that

$$|D(t)| > \alpha > 0.$$

Then there clearly exists a  $\beta > 0$  such that for every  $t$

$$(4) \quad |y_h(t) - y_g(t)| > \beta \text{ for } h \neq g.$$

Consequently, for any two values  $t_1$  and  $t_2$ , there can exist at most one substitution  $S$  (we may denote it by  $S(t_1, t_2)$  to indicate its dependence on  $t_1$  and  $t_2$ ) such that

$$\left[ |y_h(t_1) - S y_h(t_2)| \leq \frac{\beta}{2} \right].$$

We shall now prove the important theorem:

**Theorem 3.** *The set  $[y_h(t)]$  of continuous solutions of an equation (3) with almost-periodic coefficients  $x_j(t)$  and discriminant satisfying  $|D(t)| > \alpha > 0$ , has an almost-translation group  $T$ , and hence is composed of almost-periodic functions.*

The last fact was first given by WALTHER, as was mentioned in the introduction.

**Proof:** According to theorem 2 of section II it suffices to show that to every given  $\varepsilon > 0$  there corresponds a relatively dense set of real numbers  $\{\tau_{[y]}(\varepsilon)\}$ , i. e. of numbers  $\tau$  to each of which there corresponds some substitution  $S = S(\tau)$  of  $\Sigma_m$  such that  $[[y_h(t + \tau) - Sy_h(t)] \leq \varepsilon]$ . Naturally it suffices to consider "small" positive  $\varepsilon$ 's; we may therefore take the given  $\varepsilon$  to be  $< \frac{\beta}{2}$ , where  $\beta$  is the positive number occurring in the inequality (4). With this restriction on  $\varepsilon$ , if for some two values  $t_1$  and  $t_2$  we have found two substitutions  $S_1$  and  $S_2$  such that

$$[[y_h(t_1) - S_1 y_h(t_2)] \leq \varepsilon] \quad \text{and} \quad [[y_h(t_1) - S_2 y_h(t_2)] \leq \varepsilon],$$

we may conclude that  $S_1 = S_2$ .

Corresponding to the given  $\varepsilon$  we determine (as is possible because the coefficients  $x_j(t)$  are bounded) a  $\delta > 0$  such that for any two real numbers  $t'$  and  $t''$  satisfying

$$|x_j(t') - x_j(t'')| \leq \delta \quad (j = 1, \dots, m)$$

there is a substitution  $S = S(t', t'')$ , for which

$$\left[ |y_h(t') - Sy_h(t'')| \leq \frac{\varepsilon}{3} \right].$$

From the preceding remark this substitution is uniquely determined since  $\frac{\varepsilon}{3} < \varepsilon < \frac{\beta}{2}$ .

Consider the relatively dense set  $\{\tau_X(\delta)\}$  of translation numbers (corresponding to this  $\delta$ ) of a majorant  $X(t)$  of the coefficients  $x_j(t)$ . We shall prove the relative density of  $\{x_{[y]}(\varepsilon)\}$  by showing that  $\{\tau_{[y]}(\varepsilon)\} \supseteq \{\tau_X(\delta)\}$ . In other words, we shall show that to each fixed  $\tau$  satisfying for all  $t$  the inequalities

$$|x_j(t+\tau) - x_j(t)| \leq \delta \quad (j = 1, \dots, m)$$

there corresponds a substitution  $S = S(\tau)$ , such that

$$[|y_h(t+\tau) - Sy_h(t)| \leq \varepsilon] \quad (-\infty < t < +\infty).$$

Note first that, from the manner of choosing  $\delta$ , our fixed  $\tau$  and any fixed  $t$  certainly determine one, and only one, substitution  $S = S(t, \tau)$  which satisfies the relation

$$\left[ |y_h(t+\tau) - Sy_h(t)| \leq \frac{\varepsilon}{3} \right].$$

The proof will obviously be complete when we have shown that this substitution  $S(t, \tau)$  is independent of  $t$ , i. e. depends only on  $\tau$ . For this purpose we first determine an  $\eta > 0$  such that the inequality  $|t' - t''| \leq \eta$  implies the inequalities  $|y_h(t') - y_h(t'')| \leq \frac{\varepsilon}{3}$  for  $1 \leq h \leq m$ , i. e. implies

$$\left[ |y_h(t') - Ey_h(t'')| \leq \frac{\varepsilon}{3} \right].$$

Since the whole real axis can be covered by intervals of length  $\eta$ , in order to prove that  $S(t, \tau)$  is independent of  $t$  it suffices to prove that for two arbitrary numbers  $t_1$  and  $t_2$  satisfying  $|t_1 - t_2| \leq \eta$ , the two substitutions  $S_1 = S(t_1, \tau)$ , and  $S_2 = S(t_2, \tau)$  are identical.

From the definition of  $S_1$  and  $S_2$  we have

$$\left[ |y_h(t_1 + \tau) - S_1 y_h(t_1)| \leq \frac{\varepsilon}{3} \right] \text{ and } \left[ |y_h(t_2 + \tau) - S_2 y_h(t_2)| \leq \frac{\varepsilon}{3} \right].$$

Since  $|(t_1 + \tau) - (t_2 + \tau)| = |t_1 - t_2| \leq \eta$ , we have also

$$\left[ |y_h(t_1 + \tau) - E y_h(t_2 + \tau)| \leq \frac{\varepsilon}{3} \right].$$

Hence

$$\left[ |S_1 y_h(t_1) - S_2 y_h(t_2)| \leq \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon \right].$$

Replacing  $y_h$  by  $S_1^{-1} y_h$ , this becomes

$$(5) \quad \left[ |y_h(t_1) - S_2 S_1^{-1} y_h(t_2)| \leq \varepsilon \right].$$

On the other hand, from  $|t_1 - t_2| \leq \eta$  we get

$$(6) \quad \left[ |y_h(t_1) - E y_h(t_2)| \leq \frac{\varepsilon}{3} < \varepsilon \right].$$

But (as we emphasized above) relations (5) and (6) enable us to conclude that  $S_2 S_1^{-1}$  and  $E$  are identical. Thus  $S_1 = S_2$ , and the proof is completed.

We here introduce an abbreviation of our notation which will be useful later on. If  $f(t)$  and  $g(t)$  are almost-periodic functions we shall write

$$\{ \tau_f \} \subseteq \{ \tau_g \}$$

in order to indicate that, corresponding to any  $\varepsilon_1 > 0$  (or, equally well, to any sufficiently small  $\varepsilon_1 > 0$ ), there exists an  $\varepsilon_2 > 0$  such that  $\{ \tau_f(\varepsilon_2) \} \subseteq \{ \tau_g(\varepsilon_1) \}$ . Such symbolic abbreviations we shall use not only in connection with the translation numbers of single almost-periodic functions,

but also for translation numbers  $\tau_{[f]}(\varepsilon)$  and  $\tau_{(S)}(\varepsilon)$  referring to sets of functions. For instance, by

$$\{\tau_{[f]}\} \subseteq \{\tau_g\}$$

we shall mean that to any  $\varepsilon_1 > 0$  there corresponds an  $\varepsilon_2 > 0$  such that  $\{\tau_{[f]}(\varepsilon_2)\} \subseteq \{\tau_g(\varepsilon_1)\}$ .

In terms of this notation we see that, for our algebraic equation with almost-periodic coefficients  $x_j(t)$  and  $|D(t)| > \alpha > 0$ , the translation numbers of the majorant  $X(t)$  and the translation numbers belonging to the set of roots  $[y(t)]$  are in so far the same as

$$\{\tau_X\} \subseteq \{\tau_{[y]}\} \quad \text{and} \quad \{\tau_{[y]}\} \subseteq \{\tau_X\}.$$

The first of these two relations has been directly shown in the course of the proof above. The second relation holds because, corresponding to any given  $\varepsilon_1 > 0$ , since the roots  $y_h(t)$  are bounded, we can find a positive  $\varepsilon_2$  so small that every  $\tau$  which  $\varepsilon_2$ -performs some substitution on  $[y(t)]$  must satisfy, for all  $t$ , the inequalities

$$|x_j(t + \tau) - x_j(t)| \leq \varepsilon_1 \quad (j = 1, \dots, m).$$

#### IV. Auxiliary Remarks on Relations between Translation Numbers and Fourier Exponents of Almost-Periodic Functions.

With any almost-periodic function  $f(t)$  there is associated a certain series as its Fourier series,

$$f(t) \infty \sum a_n e^{i\lambda_n t}.$$

In the treatment of various questions concerning the Fourier exponents  $\lambda_n$  of  $f(t)$ , it is not just the set of exponents themselves which is of primary importance, but the larger

set containing all linear combinations with integral coefficients of a finite number of these exponents,

$$g_1\lambda_1 + g_2\lambda_2 + \dots + g_N\lambda_N.$$

This set is usually denoted by  $M_f$  and is called the module of the function  $f(t)$ ; evidently it is the smallest number-module which contains all the exponents  $\lambda_n$ .

There exist important relations between the translation numbers  $\tau_f(\varepsilon)$  and the exponents  $\lambda_n$  of an almost-periodic function  $f(t)$ . Especially we shall have to use a necessary and sufficient condition<sup>1)</sup> that the module  $M_g$  of an almost-periodic function  $g(t)$  be contained in the module  $M_f$  of an almost-periodic function  $f(t)$ , expressed in terms of the translation numbers of the two functions. By using the abbreviated notation introduced at the end of the previous section we can express this condition very simply as follows:

Lemma 1. Let  $f(t)$  and  $g(t)$  be two almost-periodic functions. In order that  $M_g \subseteq M_f$  it is necessary and sufficient that  $\{\tau_f\} \subseteq \{\tau_g\}$ .

Roughly speaking, the fewer translation numbers, the more exponents.

We shall also recall a relation between the translation numbers and the Fourier exponents of a single almost-periodic function:

Lemma 2. In order to show that the real number  $\theta$  belongs to the module  $M_f$  of an almost-periodic function  $f(t) \sim \sum a_n e^{i\lambda_n t}$ , it suffices to show that to any positive  $\varepsilon$

<sup>1)</sup> Compare H. BOHR, Ueber fastperiodische ebene Bewegungen. Commentarii Mathematici Helvetici. Vol. 4, 1932, p. 51—64.



there corresponds a positive  $\delta = \delta(\varepsilon)$ , such that each  $\tau$  belonging to  $\{x_f(\delta)\}$  satisfies the inequality

$$|e^{i\theta\tau} - 1| \leq \varepsilon.$$

We shall not enter upon the proof of these two lemmas, but content ourselves with pointing out that they depend essentially on a famous theorem of KRONECKER on Diophantine approximations. As we shall make direct use of KRONECKER'S theorem in section VI, but in a form slightly different from the usual one, we take this occasion to re-state his theorem in this form, using the exponential function  $e^{it}$ .

**KRONECKER'S Theorem.** Let  $\lambda_1, \dots, \lambda_N$  be arbitrarily given real numbers, and let  $\eta_1, \dots, \eta_N$  be given complex numbers of absolute value 1. In order that to each positive  $\varepsilon$  there correspond a real  $\tau$  satisfying the  $N$  inequalities

$$|e^{i\lambda_n\tau} - \eta_n| \leq \varepsilon \quad (n = 1, \dots, N),$$

it is necessary and sufficient that whenever a linear relation with integral coefficients

$$g_1\lambda_1 + \dots + g_N\lambda_N = 0$$

holds between the  $\lambda$ 's, the corresponding relation

$$\eta_1^{g_1} \dots \eta_N^{g_N} = 1$$

hold between the  $\eta$ 's.

One sees immediately that the condition is necessary; the real content of the theorem lies in the fact that it is also sufficient.

### V. Relations between the Fourier Exponents of the Roots and the Fourier Exponents of the Coefficients of an Algebraic Equation.

We return to the algebraic equation

$$y^m + x_1(t)y^{m-1} + \dots + x_{m-1}(t)y + x_m(t) = 0$$

of section III. We denote by  $X(t)$  some majorant of the almost-periodic coefficients  $x_1(t), \dots, x_m(t)$ . As is well known, the module  $M_X$  is independent of the choice of the majorant, and is not only the smallest number-module containing all the Fourier exponents of  $X(t)$ , but is also the smallest number-module containing all the Fourier exponents of all the functions  $x_j(t)$ .

As before, we assume that the discriminant of the equation satisfies the inequality

$$\frac{GLB}{-\infty < t < \infty} |D(t)| > 0,$$

so we know that the equation has as continuous roots a set of almost-periodic functions,

$$y_1(t), \dots, y_m(t).$$

Thus the roots have a majorant  $Y(t)$ , and the module  $M_Y$  is the smallest number-module containing all the Fourier exponents of the functions in the set  $[y(t)]$ .

In this section we shall discuss the connection between the Fourier exponents of the roots of the equation and the Fourier exponents of its coefficients. The connection will be exhibited by demonstrating some important relations between the modules  $M_X$ ,  $M_Y$ , and  $M_{y_h}$ , and their multiples.<sup>1)</sup>

<sup>1)</sup> By the multiple  $r \cdot M$  we shall understand the module arising from the module  $M$  by multiplying every number in  $M$  by  $r$ .

**Theorem 4.**  $M_X \subseteq M_Y$ .

*Proof.* This is an immediate consequence of the facts: that the coefficients  $x_j(t)$  are polynomials in the roots  $y_h(t)$ ; and that the Fourier series of a polynomial function of almost-periodic functions is determined simply by formal calculation from the Fourier series of these functions. Thus each exponent of each  $x_j(t)$  is a linear combination with integral coefficients of the exponents of the roots  $y_1(t), \dots, y_m(t)$ , and therefore belongs to  $M_Y$ . Hence the whole module  $M_X$  is contained in  $M_Y$ .

We now proceed to a theorem which is not at all trivial, and the proof of which uses essentially the connection between Fourier exponents and translation numbers mentioned in section IV. As in section II we denote by  $T$  the almost-translation group of the set  $[y(t)]$  of roots, and by  $\nu_h (\leq m)$  the least positive integer  $e$  such that  $S^e y_h = y_h$  for every  $S$  in  $T$ , as defined in section I.

**Theorem 5.** For each  $h$  among  $1, \dots, m$  we have the relation

$$(7) \quad M_{y_h} \subseteq \frac{1}{\nu_h} M_X.$$

The fact that to each of the roots  $y_h(t)$  there corresponds some number  $\nu_h \leq m$  for which  $M_{y_h} \subseteq \frac{1}{\nu_h} M_X$ , was first found by CAMERON in the paper quoted in the introduction.

*Proof.* In remark 5° of section II it was shown that for each sufficiently small  $\varepsilon$ ,

$$\nu_h \left\{ \tau_{[y]} \left( \frac{\varepsilon}{\nu_h} \right) \right\} \subseteq \{ \tau_{y_h}(\varepsilon) \}.$$

Hence, using the abbreviation introduced at the end of section III, we have

$$(8) \quad \nu_h \{ \tau_{[y]} \} \subseteq \{ \tau_{y_h} \}.$$

But, as shown in section III, the translation numbers  $\tau_{[y]}$  and  $\tau_X$  are the "same" in the sense that

$$\{ \tau_{[y]} \} \subseteq \{ \tau_X \} \quad \text{and} \quad \{ \tau_X \} \subseteq \{ \tau_{[y]} \}.$$

Hence the relation (8) is equivalent to the relation

$$\nu_h \{ \tau_X \} \subseteq \{ \tau_{y_h} \}.$$

Now it is evidently the same thing to say of a number  $\tau$  that it is a  $\nu_h \cdot \tau_X(\delta)$  as to say that it is a  $\tau_{X^*}(\delta)$ , where  $X^*(t)$  denotes the function  $X\left(\frac{t}{\nu_h}\right)$ . Thus we may write

$$\{ \tau_{X^*} \} \subseteq \{ \tau_{y_h} \},$$

and by lemma 1 of section IV we conclude that

$$M_{y_h} \subseteq M_{X^*}.$$

But the Fourier exponents of  $X^*(t) = X\left(\frac{t}{\nu_h}\right)$  are simply the exponents of  $X(t)$  itself, each divided by  $\nu_h$ . Hence  $M_{X^*} = \frac{1}{\nu_h} M_X$ , and we get the desired result,

$$M_{y_h} \subseteq \frac{1}{\nu_h} M_X.$$

Corollary 1.

$$M_Y \subseteq \frac{1}{m!} M_X.$$

For this merely contracts to one relation the  $m$  relations  $M_{y_h} \subseteq \frac{1}{m!} M_X$ , each of which must hold since every  $\nu_h$ , being  $\leq m$ , is a divisor of  $m!$ . This corollary, substantially

due to WALTHER, may also be proved readily without using the sharper relations (7).

Corollary 2. The modules  $M_X$  and  $M_Y$  have the same maximum number of rationally independent elements (i. e. elements linearly independent with respect to the rational domain).

This follows immediately from  $M_X \subseteq M_Y$  and  $M_Y \subseteq \frac{1}{m!} M_X$ .

In the remainder of the paper we shall confine our attention to the case where the almost-translation group  $\Gamma$  is transitive. In this case we get a rather complete survey of the relations connecting the Fourier series of the different roots. This restriction is a natural one since, when our given algebraic equation has as roots a set of functions with an intransitive group  $\Gamma$ , it may be split up into a number of equations of lower degree, each one having as its roots the functions of a transitivity system of the original set; at the same time, as we proceed to show, this process does not enlarge the modules of the coefficients, i. e. the module of the coefficients of each new equation is contained in the original module  $M_X$ .

For let  $y_1(t), \dots, y_\rho(t)$  be a transitivity system of the set  $[y(t)]$ , and let

$$(y - y_1(t)) \dots (y - y_\rho(t)) = y^\rho + \xi_1(t) y^{\rho-1} + \dots + \xi_\rho(t) = 0$$

be the corresponding new equation, with  $\Xi(t)$  as a majorant of its coefficients. Then we have to show that

$$M_\Xi \subseteq M_X.$$

Now, roughly speaking, every "fine" translation number of  $X(t)$  gives rise to one of the substitutions in  $\Gamma$ , and hence,

when applied to the transitivity system  $y_1(t), \dots, y_\rho(t)$ , merely effects some permutation of this subset. But this means that it is a "fine" translation number of  $\Xi(t)$ . Speaking precisely, we have

$$\{\tau_X\} \subseteq \{\tau_\Xi\},$$

and therefore

$$M_\Xi \subseteq M_X.$$

## VI. The Fourier Series of a finite Set of Almost-Periodic Functions with a Transitive Almost-Translation Group.

In section II we introduced the notion of the almost-translation group of a finite set  $[f(t)]$  of  $m$  distinct complex functions of  $t$ , each function defined and continuous for  $-\infty < t < +\infty$ . We there showed that a necessary and sufficient condition that such a set of functions have an almost-translation group  $T$  is that each function be almost-periodic. In section III we found that in order that an algebraic equation of degree  $m$  in the complex variable  $y$  with leading coefficient 1 have as roots a set  $[y(t)]$  of  $m$  distinct almost-periodic functions of  $t$  (i. e. a set of  $m$  functions having an almost-translation group), it is sufficient that (a) the coefficients be almost-periodic functions of  $t$ , and (b) the discriminant  $D(t)$  satisfy the inequality

$$(9) \quad \underset{-\infty < t < \infty}{GLB} |D(t)| > 0.$$

Condition (a) is evidently necessary as well; for the coefficients of the equation, being symmetric polynomials in the roots, must be almost-periodic themselves. As regards the condition (b), concerning the discriminant, the situation

is more complicated. In fact on the one hand, as mentioned in the introduction, condition (b) cannot be replaced by the weaker condition,  $D(t) \neq 0$  for all  $t$ ; on the other hand there exist algebraic equations of degree  $m$  with almost-periodic coefficients and satisfied by  $m$  distinct almost-periodic functions, for which not only  $GLB |D(t)| = 0$  but even  $D(t) = 0$  for some  $t^1$ . Thus the restriction (9) on the discriminant introduces an element extraneous to the notion of the almost-translation group. In the present section, therefore, where we deal with those properties stemming directly from the group, we shall not think of our functions as roots of an algebraic equation, but only as having an almost-translation group  $T$ . As pointed out at the close of the previous section, we consider only the case where  $T$  is transitive.

Let, then,  $[f(t)]$  be a finite set of  $m$  distinct almost-periodic functions having as almost-translation group a transitive Abelian group  $T$  of substitutions on the set. As we remarked in section I, the group has  $m$  elements which we may (and do) index concordantly with the functions, so that  $S_h f_1 = f_h$ . We begin by deducing a number of properties of the Fourier series of the functions  $f_h(t)$ .

1°. Every function  $f_h(t)$ ,  $h = 2, \dots, m$ , has exactly the

<sup>1)</sup> This may occur even in the case where  $T$  is transitive. A simple example is given by the two (periodic) functions

$$y_1(t) = \frac{1}{2} (e^{it} + e^{-it}) = \cos t, \quad y_2(t) = -\frac{1}{2} (e^{it} + e^{-it}) = -\cos t$$

with the translation group  $T = \Sigma_2$  of order 2, where the discriminant  $D(t) = 4 \cos^2 t$  of the corresponding equation

$$(y - y_1(t)) \cdot (y - y_2(t)) = y^2 - \cos^2 t = 0$$

vanishes at  $t = \frac{\pi}{2} + n\pi$ .

same Fourier exponents,  $\lambda_1, \lambda_2, \dots$ , as  $f_1(t)$ ; and the Fourier coefficients in  $f_h(t)$  and  $f_1(t)$  belonging to the same exponent  $\lambda_n$ , have the same numerical value. Thus we may write the Fourier series of the functions in the form

$$f_h(t) \sim \sum \eta_{h,n} a_n e^{i\lambda_n t} \quad (h = 1, \dots, m),$$

where each factor  $\eta_{h,n}$  has the numerical value 1. Within the restriction that their numerical values be 1, the factors  $\eta_{h,n}$  in the Fourier series of any one function may of course be selected quite arbitrarily. Thus, while keeping the notation above for convenience, we shall suppose that each  $\eta_{1,n}$  in the expansion of  $f_1(t)$  is equal to 1.

*Proof.* Corresponding to any  $h$  among  $2, \dots, m$ , and to an arbitrary sequence of positive numbers  $\varepsilon_\nu$  tending to zero, we can choose a sequence of translation numbers  $\tau_\nu \in \{\tau_{(S_h)}(\varepsilon_\nu)\}$ . Then the sequence of functions  $f_1(t + \tau_\nu)$  will tend to  $f_h(t)$  uniformly throughout  $-\infty < t < +\infty$ , that is the function  $f_h(t)$  belongs to the class of almost-periodic functions usually called the uniform closure of the set  $\{f_1(t+k)\}$  and denoted by  $C_U\{f_1(t+k)\}$ . It is a well-known property of such a set that every function in it has the properties ascribed above to  $f_h(t)$ .

2°. The factors  $\eta_{1,n}, \dots, \eta_{m,n}$  corresponding to any Fourier exponent  $\lambda_n$  are exactly the respective values,  $\chi_n(S_1), \dots, \chi_n(S_m)$ , of some character  $\chi_n$  of the group  $T$ .

It will be convenient to represent this assertion by the following scheme, where, fixing upon any  $\lambda_n$ , we have dropped the index  $n$  momentarily and have arranged in vertical columns the terms and the factors  $\eta$  corresponding to this exponent in the Fourier series of  $f_1(t), \dots, f_m(t)$ .



$\begin{aligned} & a\eta_1 e^{i\lambda t} \\ & a\eta_2 e^{i\lambda t} \\ & \vdots \\ & a\eta_m e^{i\lambda t} \end{aligned}$	$\begin{aligned} \eta_1 &= \chi(S_1) \\ \eta_2 &= \chi(S_2) \\ & \vdots \\ \eta_m &= \chi(S_m) \end{aligned}$
--	---

Proof. From the definition of a character it follows that we have to show that if  $S_{h_1}$  and  $S_{h_2}$  are any (not necessarily distinct) substitutions in  $T$ , and if  $S_{h_1} S_{h_2} = S_{h_3}$ , then

$$\eta_{h_1} \cdot \eta_{h_2} = \eta_{h_3}.$$

Take again an arbitrary sequence of positive numbers converging to 0, say  $\varepsilon_\nu \rightarrow 0$ , and for each  $\nu$  denote by  $\tau_{1\nu}$  and  $\tau_{2\nu}$  arbitrarily chosen translation numbers belonging to  $\{\tau_{(S_{h_1})}(\varepsilon_\nu)\}$  and  $\{\tau_{(S_{h_2})}(\varepsilon_\nu)\}$  respectively. Then by the lemma of section II the numbers  $\tau_{3\nu} = \tau_{1\nu} + \tau_{2\nu}$  belong to  $\{\tau_{(S_{h_3})}(2\varepsilon_\nu)\}$ . Hence, as  $\nu \rightarrow \infty$ , we have, uniformly throughout  $-\infty < t < +\infty$ ,

$$f_1(t + \tau_{1\nu}) \rightarrow f_{h_1}(t), \quad f_1(t + \tau_{2\nu}) \rightarrow f_{h_2}(t), \quad f_1(t + \tau_{3\nu}) \rightarrow f_{h_3}(t).$$

It is a trivial fact concerning almost-periodic functions that uniform convergence of a sequence of such functions implies ordinary convergence of the coefficients belonging to any exponent. Since the coefficient corresponding to the exponent  $\lambda$  in the Fourier series of  $f_1(t + \tau)$ , for arbitrary  $\tau$ , is given by  $a e^{i\lambda\tau}$ , and since  $a \neq 0$ , we conclude from the uniformly convergent sequences above that

$$e^{i\lambda\tau_{1\nu}} \rightarrow \eta_{h_1}, \quad e^{i\lambda\tau_{2\nu}} \rightarrow \eta_{h_2}, \quad e^{i\lambda\tau_{3\nu}} \rightarrow \eta_{h_3}.$$

Finally, since  $e^{i\lambda\tau_{1\nu}} \cdot e^{i\lambda\tau_{2\nu}} = e^{i\lambda\tau_{3\nu}}$  for every  $\nu$ , we have in the limit  $\eta_{h_1} \cdot \eta_{h_2} = \eta_{h_3}$ .

3°. Those characters of  $\Gamma$  whose values actually occur (in accordance with property 2°) as coefficients in the Fourier series of  $f_1(t), \dots, f_m(t)$ , form a generating system of the character group  $\Gamma^*$ .

Proof. Let us denote by  $\chi'_1, \chi'_2, \dots, \chi'_q$  the distinct characters in  $\Gamma^*$  which thus appear. On the analogy of the scheme in 2°, we write their values in vertical columns, thus:

$$\begin{array}{|c|} \hline \chi'_1(S_1) \\ \chi'_1(S_2) \\ \vdots \\ \chi'_1(S_m) \\ \hline \end{array}, \quad \begin{array}{|c|} \hline \chi'_2(S_1) \\ \chi'_2(S_2) \\ \vdots \\ \chi'_2(S_m) \\ \hline \end{array}, \dots, \quad \begin{array}{|c|} \hline \chi'_q(S_1) \\ \chi'_q(S_2) \\ \vdots \\ \chi'_q(S_m) \\ \hline \end{array}$$

If we now apply the criterion for a generating system of  $\Gamma^*$  which we gave in section I, we see that the denial of our assertion is equivalent to asserting that for some  $h \neq g$  the values in row  $h$  are identical with the corresponding values in row  $g$ . But this would make the Fourier series of  $f_h(t)$  and  $f_g(t)$  identical, which (from the uniqueness theorem in the theory of almost-periodic functions) would imply  $f_h(t) \equiv f_g(t)$ , contrary to hypothesis.

4°. If any linear relation with integral coefficients, say

$$(10) \quad g_1 \lambda_1 + \dots + g_N \lambda_N = 0,$$

connects a finite number of the Fourier exponents  $\lambda_n$ , then we have the multiplicative relation

$$(11) \quad \chi_1^{g_1} \dots \chi_N^{g_N} = 1$$

connecting the characters  $\chi_1, \dots, \chi_N$  corresponding (as in property 2<sup>o</sup>) to the exponents  $\lambda_1, \dots, \lambda_N$ , respectively.

The relation (11) may be expressed in terms of the values of the characters as a set of  $m$  relations

$$(12) \quad [\chi_1(S_h)]^{g_1} \dots [\chi_N(S_h)]^{g_N} = 1 \quad (1 \leq h \leq m).$$

*Proof.* Instead of proving the assertion for some arbitrarily chosen relation (10) it will be more convenient to prove it simultaneously for all linear relations with last index equal to an arbitrarily chosen fixed  $N$ . As before, for fixed  $h$ , arbitrary sequence of positive numbers  $\varepsilon_\nu \rightarrow 0$ , and corresponding sequence  $\tau_\nu$  of translation numbers with each  $\tau_\nu \in \{x_{(S_h)}(\varepsilon_\nu)\}$ , we have, uniformly throughout  $-\infty < t < +\infty$ , that  $f_1(t + \tau_\nu) \rightarrow f_h(t)$ , and hence for each  $n (= 1, 2, \dots)$

$$e^{i\lambda_n \tau_\nu} \rightarrow \eta_{h,n} = \chi_n(S_h).$$

Hence for any arbitrarily small  $\varepsilon$  the  $N$  inequalities

$$|e^{i\lambda_n \tau} - \chi_n(S_h)| \leq \varepsilon \quad (n = 1, \dots, N)$$

are satisfied by some  $\tau$  (in fact by every  $\tau_\nu$  with sufficiently great index  $\nu$ ). According to the "trivial" part of KRONECKER'S theorem stated in the end of section IV, this requires that, corresponding to each relation (10) with last index  $N$ , there must hold the corresponding relation (12). Thus (11) holds.

We now turn the problem about, i. e. start from a given transitive Abelian substitution group  $\Gamma$ , and ask what conditions a set of almost-periodic functions  $f_h(t)$  must fulfill in order that they be distinct and that  $[f(t)]$

have  $\Gamma$  as its almost-translation group. To avoid misunderstanding, however, we emphasize that so far we do not know at all whether every type of transitive Abelian substitution group can occur as the almost-translation group of some set  $[f(t)]$ . Only after having proved the theorem below shall we take up the main problem of determining whether the conditions stated in the theorem can really be fulfilled for any given group of this sort.

**Theorem 6.** *Let  $[f(t)]$  be a finite set of  $m$  almost-periodic functions,  $f_1(t), \dots, f_m(t)$ ; and let  $\Gamma$  be an arbitrary transitive Abelian group of  $m$  substitutions which we denote by  $S_1 = \begin{pmatrix} 1, \dots \\ 1, \dots \end{pmatrix}, \dots, S_m = \begin{pmatrix} 1, \dots \\ m, \dots \end{pmatrix}$ . Then in order that  $[f(t)]$  be composed of distinct functions and have  $\Gamma$  as its almost-translation group, it is not only necessary (as was shown above) but also sufficient that the following four conditions be fulfilled:*

1. All the functions  $f_h(t)$  have exactly the same Fourier exponents  $\lambda_n$ , and the absolute values of the corresponding Fourier coefficients are the same.

2. Further, the Fourier series of the functions  $f_h(t)$  have the form

$$(13) \quad f_h(t) \infty \sum \chi_n(S_h) a_n e^{i\lambda_n t},$$

where  $\chi_n(S)$ , for each  $n$ , is a character of the group  $\Gamma$ .

3. Those characters  $\chi_n(S)$  which actually occur in (13) form a generating system of the character-group  $\Gamma^*$ .

4. If any finite set,  $\lambda_1, \dots, \lambda_N$ , of the Fourier exponents are connected by a linear relation

$$g_1 \lambda_1 + \dots + g_N \lambda_N = 0$$

with integral coefficients, then the corresponding characters are connected by the relation

$$\chi_1^{g_1} \dots \chi_N^{g_N} = 1.$$

**Proof.** That  $m$  almost-periodic functions whose Fourier series have the form (13) must be distinct, is immediately to be seen from condition 3, which implies that for no two distinct indices,  $h$  and  $g$ , are the Fourier series of  $f_h(t)$  and  $f_g(t)$  formally the same. (Here we use only the trivial fact that two distinct Fourier series cannot belong to the same function, in contrast to the corresponding point in the necessity proof, where we had to use the uniqueness theorem, that two distinct functions cannot have the same Fourier series.) Since our functions are distinct, they have an almost-translation group  $I'$ ; and the main point of the proof is to show that conditions 1—4 insure that  $I' = I$ .

We first observe that it suffices to show that each of the substitutions of  $I$  belongs to  $I'$ . In fact, when this has been shown, it follows immediately that  $I'$  must be identical with  $I$ . For then  $I'$  must be transitive (since it contains the transitive sub-group  $I$ ) and, being Abelian, cannot contain more than  $m$  substitutions.

By way of preparation we make the preliminary remark that the set-theoretical sum  $\{g(t)\}$  of the  $m$  sets of functions

$$\{f_1(t+k)\}, \dots, \{f_m(t+k)\} \quad (-\infty < k < +\infty)$$

is a majorisable set of almost-periodic functions, in fact majorisable by any majorant  $F(t)$  of the set  $[f(t)]$ ; and that every function in  $\{g(t)\}$  has the same Fourier exponents  $\lambda_1, \lambda_2, \dots$  as each function in  $[f(t)]$ . Hence, according to a wellknown theorem on majorisable sets, due to BOCHNER, it holds, roughly speaking, that formal convergence of the Fourier series of a sequence of functions drawn from the set  $\{g(t)\}$  (i. e. actual convergence of the coefficients belonging to each fixed exponent) implies uniform convergence of the sequence of functions. Exactly speaking: to

each  $\varepsilon > 0$  there corresponds a positive integer  $N$  and a  $\delta > 0$  such that any two functions  $g'(t) \approx \sum b'_n e^{i\lambda_n t}$  and  $g''(t) \approx \sum b''_n e^{i\lambda_n t}$ , of the set  $\{g(t)\}$ , the coefficients of which fulfill the inequalities

$$|b'_n - b''_n| \leq \delta \quad (n = 1, \dots, N),$$

themselves satisfy the inequality

$$|g'(t) - g''(t)| \leq \varepsilon \quad (-\infty < t < +\infty).$$

Now let  $S_h$  be an arbitrarily chosen fixed element of  $\Gamma$ . We shall show that  $S_h$  also belongs to  $\Gamma'$ , i. e. that  $S_h$  is an almost-translation substitution of the set  $[f(t)]$ . From remark 4<sup>o</sup> in section II it suffices to prove that to some positive  $\varepsilon \leq \varepsilon^{**}$  there corresponds at least one real number  $\tau$  which satisfies for all values of  $t$  the  $m$  inequalities

$$(14) \quad |f_l(t + \tau) - S_h f_l(t)| \leq \varepsilon \quad (l = 1, \dots, m).$$

We first make clear, for each  $l$ , which of the functions  $f_1, \dots, f_m$  is denoted by  $S_h f_l$ . This can be seen at once by considering the Fourier series of the function:

$$f_l(t) \approx \sum \chi_n(S_l) a_n e^{i\lambda_n t}.$$

In this Fourier series the index  $l$  occurs only as index for that substitution  $S_l$  of which the characters  $\chi_n$  are taken. As we noted in section I, since  $\chi_n$  is a character of  $\Gamma$ , the substitution  $S_h$  performed on the index  $l$  simply results in replacing  $\chi_n(S_l)$  by  $\chi_n(S_h S_l)$ , that is by the product  $\chi_n(S_h) \chi_n(S_l)$ . Hence the function  $S_h f_l(t)$  is just that function among  $f_1(t), \dots, f_m(t)$  whose Fourier series is given by

$$\sum \chi_n(S_h) \chi_n(S_l) a_n e^{i\lambda_n t}.$$

We now proceed to prove that there exists a  $\tau$  which

satisfies the inequalities (14). As the Fourier series of  $f_l(t+\tau)$ , for an arbitrary fixed  $\tau$ , is given by

$$\sum e^{i\lambda_n \tau} \chi_n(S_l) a_n e^{i\lambda_n t},$$

and as, throughout, we deal only with functions of our majorisable set  $\{g(t)\}$ , the inequality

$$|f_l(t+\tau) - S_h f_l(t)| \leq \varepsilon$$

is (by the remark above) certainly fulfilled by any  $\tau$  for which the  $N$  inequalities

$$|e^{i\lambda_n \tau} \chi_n(S_l) a_n - \chi_n(S_h) \chi_n(S_l) a_n| \leq \delta \quad (n = 1, \dots, N),$$

i. e. (since  $|\chi_n(S_l)| = 1$ ) the  $N$  inequalities

$$(15) \quad |(e^{i\lambda_n \tau} - \chi_n(S_h)) a_n| \leq \delta \quad (n = 1, \dots, N),$$

are fulfilled. We see that the index  $l$  has disappeared, so that any  $\tau$  satisfying the inequalities (15) will certainly satisfy the  $m$  inequalities (14). Denoting by  $a$  the maximum of  $|a_1|, \dots, |a_N|$ , the inequalities (15) are certainly satisfied by any  $\tau$  which satisfies

$$(16) \quad |e^{i\lambda_n \tau} - \chi_n(S_h)| \leq \frac{\delta}{a} \quad (n = 1, \dots, N).$$

But according to the "non-trivial" part of KRONECKER'S theorem, the inequalities (16) certainly have a solution  $\tau$  since condition 4 assures us that to every linear relation with integral coefficients such as  $g_1 \lambda_1 + \dots + g_N \lambda_N = 0$  there corresponds the relation  $(\chi_1(S_h))^{g_1} \dots (\chi_N(S_h))^{g_N} = 1$ . Thus the proof of theorem 6 is completed.

We can now easily prove the following corollary, which is to be considered one of the main results of the paper.

**Corollary.** *Corresponding to each arbitrarily given transitive Abelian group  $T$  of substitutions on the indices  $1, \dots, m$ , there exists a set  $[f(t)]$  of  $m$  distinct almost-periodic functions which has  $T$  as its almost-translation group.*

*Proof.* We have only to show that to the given group  $T$  there correspond  $m$  almost-periodic functions  $f_1(t), \dots, f_m(t)$  whose Fourier series satisfy conditions 1—4 of theorem 6. In order to avoid any trouble arising from the somewhat intricate condition 4, we choose the exponents  $\lambda_n$  rationally independent, i. e. such that a relation  $g_1\lambda_1 + \dots + g_N\lambda_N = 0$  with integral coefficients can occur only if every  $g$  is zero. Condition 4 then falls away. Next, to be sure that the series we set up are the Fourier series of almost-periodic functions, we limit ourselves to only a finite number of terms. Now let  $\chi'_1, \dots, \chi'_q$  be any generating system of the character-group  $T^*$ , let  $\lambda_1, \dots, \lambda_q$  be arbitrarily chosen rationally independent real numbers, and let  $a_1, \dots, a_q$  be arbitrary complex numbers  $\neq 0$ . Then it is clear that the  $m$  almost-periodic functions

$$f_h(t) = \sum_{n=1}^q \chi'_n(S_h) a_n e^{i\lambda_n t} \quad (h = 1, \dots, m)$$

satisfy conditions 1—4, and so the set  $[f(t)]$  has  $T$  as its almost-translation group.

If we wish to have as few terms as possible in the example just constructed we shall take  $q = \mu$ , where  $\mu$  is the number of elements in a minimal generating system of  $T$  (or of  $T^*$ ), mentioned in section I; and for our characters  $\chi'_1, \dots, \chi'_\mu$  we shall take any minimal generating



system of  $T^*$ . We shall show that the example thus obtained of a set  $[f(t)]$  of functions belonging to the given group  $T$  is not only a simplest possible example, in that the Fourier series contains as few terms as possible, but at the same time is the most general example of a set  $[f(t)]$  belonging to  $T$ , for which the functions  $f_h(t)$  contain the minimum number  $\mu$  of terms. By this we mean that the only restriction (other than as to the number of terms) of a voluntary character which has been introduced so far, namely *that the exponents be rationally independent*, is in fact a necessary one. We shall prove, namely, the following general theorem:

**Theorem 7.** Let

$$f_h(t) \approx \sum \chi_n(S_h) a_n e^{i\lambda_n t} \quad (h = 1, \dots, m)$$

be any  $m$  almost-periodic functions with the transitive group  $T$  as almost-translation group of the set  $[f(t)]$ , and let  $\mu$  denote the number of elements in a minimal generating system of  $T$ . Then among the exponents  $\lambda_1, \lambda_2, \dots$  there occur at least  $\mu$  which are rationally independent.

*Proof.* Among the characters  $\chi_n$  involved in the Fourier series there certainly occurs some generating system,  $\chi'_1, \dots, \chi'_q$ , of  $T^*$ . By a remark in section I, out of this generating system there can always be chosen a set of just  $\mu$  characters, say  $\varphi_1, \dots, \varphi_\mu$ , such that any relation  $\varphi_1^{g_1} \dots \varphi_\mu^{g_\mu} = 1$  implies  $GCD(g_1, \dots, g_\mu) > 1$ . Now let  $n_1, \dots, n_\mu$  be arbitrarily chosen indices such that

$$\chi_{n_1} = \varphi_1, \dots, \chi_{n_\mu} = \varphi_\mu.$$

Then the exponents  $\lambda_{n_1}, \dots, \lambda_{n_\mu}$  must be rationally independent; otherwise there would exist a linear relation

$$g_1 \lambda_{n_1} + \dots + g_\mu \lambda_{n_\mu} = 0$$

with  $GCD(g_1, \dots, g_\mu) = 1$ . But by condition 4 this would imply

$$\chi_{n_1}^{g_1} \dots \chi_{n_\mu}^{g_\mu} = 1, \text{ i. e. } \varphi_1^{g_1} \dots \varphi_\mu^{g_\mu} = 1,$$

contrary to our hypothesis that no such relation can exist with  $GCD(g_1, \dots, g_\mu) = 1$ .

In the case considered, where the almost-translation group  $T$  is transitive, the characters of  $T$  enable us to give a certain canonical representation of the  $m$  functions in the set  $[f(t)]$ , in which all the functions  $f_h(t)$  are represented as linear combinations of a finite set of almost-periodic functions which do not depend on  $h$ .

Starting from the particular form of the Fourier series of the functions  $f_h(t)$  given in condition 2 of theorem 6, namely

$$(17) \quad f_h(t) \approx \sum \chi_n(S_h) a_n e^{i\lambda_n t} \quad (h = 1, \dots, m),$$

we obtain our representation formally by collecting those terms for which the characters are the same. Thus we may write

$$(18) \quad f_h(t) \approx \sum_{\chi \in T^*} \chi(S_h) \left( \sum_{\lambda_n = \chi} a_n e^{i\lambda_n t} \right) \quad (h = 1, \dots, m),$$

where, for the sake of uniformity of notation, we suppose that  $\sum_{\lambda_n = \chi} a_n e^{i\lambda_n t}$  is simply empty if  $\chi$  does not occur in the series in (17). Since the series in ( )'s on the right side of the relations (18) are independent of  $h$ , the principal step in establishing our canonical representation is to prove that these series are the actual Fourier series of some almost-periodic functions of  $t$ .

In general, a subseries of the Fourier series of an almost-periodic function is not itself the Fourier series of any almost-periodic function. In our case, however, each series

$$(19) \quad \sum_{\chi_n = \chi} a_n e^{i\lambda_n t},$$

is the Fourier series of an almost-periodic function.

For if  $\psi(S)$  is any fixed character and we multiply the relations (18) by the respective values  $\bar{\psi}(S_1), \dots, \bar{\psi}(S_m)$  of the conjugate character  $\bar{\psi}$  and add the resulting relations, we have

$$\sum_{h=1}^m \bar{\psi}(S_h) f_h(t) \propto \sum_{h=1}^m \bar{\psi}(S_h) \left\{ \sum_{\chi \in \Gamma^*} \chi(S_h) \left( \sum_{\chi_n = \chi} a_n e^{i\lambda_n t} \right) \right\}$$

or

$$(20) \quad \sum_{h=1}^m \bar{\psi}(S_h) f_h(t) \propto \sum_{\chi \in \Gamma^*} \left( \sum_{h=1}^m \bar{\psi}(S_h) \chi(S_h) \right) \left( \sum_{\chi_n = \chi} a_n e^{i\lambda_n t} \right).$$

Now, due to the orthogonality of the characters,

$\sum_{h=1}^m \bar{\psi}(S_h) \chi(S_h) = m$  or  $0$  according as  $\chi$  is or is not the chosen character  $\psi$ . Hence after dividing by  $m$  the relation (20) becomes

$$\frac{1}{m} \sum_{h=1}^m \bar{\psi}(S_h) f_h(t) \propto \sum_{\chi_n = \psi} a_n e^{i\lambda_n t}.$$

Replacing the letter  $\psi$  by  $\chi$ , we see that the series (19) is the Fourier series of the almost-periodic function

$$\Phi_\chi(t) = \frac{1}{m} \sum_{h=1}^m \bar{\chi}(S_h) f_h(t) \propto \sum_{\chi_n = \chi} a_n e^{i\lambda_n t}.$$

Hence, by elementary theorems on almost-periodic functions, the Fourier series of the function

$$\sum_{\chi \in \Gamma^*} \chi(S_h) \Phi_\chi(t)$$

is just the Fourier series of  $f_h(t)$ , as written in (18), so that

$$f_h(t) = \sum_{\chi \in \Gamma^*} \chi(S_h) \Phi_\chi(t) \quad (h = 1, \dots, m)$$

which is our desired canonical representation of the functions  $f_h(t)$  as linear combinations of a common set of  $m$  almost-periodic functions, namely the functions  $\Phi_\chi(t)$ .

In the next section, where we shall consider functions  $f_h(t)$  which are the roots of an algebraic equation with almost-periodic coefficients, we shall return to this canonical representation and prove an interesting property concerning the connection between the Fourier exponents of each function  $\Phi_\chi(t)$  and those of the coefficients of the equation.

### VII. Algebraic Equations with Almost-Periodic Coefficients whose Roots have a Transitive Almost-Translation Group.

In this last section we return to the consideration of the algebraic equation

$$(21) \quad y^m + x_1(t)y^{m-1} + \dots + x_{m-1}(t)y + x_m(t) = 0$$

with almost-periodic coefficients. In case the discriminant  $D(t)$  satisfies the condition

$$(22) \quad \begin{matrix} G L B \\ -\infty < t < +\infty \end{matrix} |D(t)| > 0,$$

we know that the roots  $y_h(t)$  are again almost-periodic functions. We put the following

**Problem:** To indicate necessary and sufficient conditions which  $m$  functions,  $y_1(t), \dots, y_m(t)$ , assumed to be almost-periodic, must fulfill in order that

(a) they may be the roots of an algebraic equation (21) of degree  $m$  with (eo ipso almost-periodic) coefficients  $x_j(t)$  satisfying (22), and

(b) the set  $[y(t)]$  may have a given transitive Abelian group  $T$  as its almost-translation group.

It is not to be expected that an answer to this problem should be as neat as the answer — given in theorem 6 of section VI — to the analogous problem bearing only on the Fourier series of the functions  $y_h(t)$ . For the condition that the functions be distinct has been replaced by the condition that the discriminant of the equation (21) whose roots they are shall satisfy (22); and it does not seem possible to transform this more complex condition into a simple condition on the Fourier series of the functions. However, for our purpose it suffices to remark that — as the  $y_h(t)$ 's are almost-periodic, and hence bounded — the demand that the  $y_h(t)$ 's satisfy (22) is obviously equivalent to requiring that

$$(23) \quad \underset{h \neq g}{\overset{GLB}{-\infty < t < +\infty}} |y_h(t) - y_g(t)| > 0.$$

Thus we may give the following

**Answer to the Problem:** In order that the almost-periodic functions  $y_h(t)$  shall satisfy the conditions (a) and (b) in question, it is necessary and sufficient that they satisfy the condition (23), and that their Fourier series satisfy conditions 1—4 of theorem 6.

With the help of this answer we now easily reach the following principal result, already alluded to in the introduction.

**Theorem 8.** *Corresponding to any arbitrarily given transitive Abelian group  $\Gamma$ , there exists an algebraic equation (21) with almost-periodic coefficients  $x_j(t)$  satisfying (22) and such that the set  $[y(t)]$  of its (eo ipso almost-periodic) roots has  $\Gamma$  as its almost-translation group.*

**Proof.** If as in section VI we denote by  $\chi'_1, \dots, \chi'_\mu$  an arbitrarily chosen minimal generating system of the character group  $\Gamma^*$ , by  $a_1, \dots, a_\mu$  arbitrary complex numbers (each  $\neq 0$ ), and by  $\lambda_1, \dots, \lambda_\mu$  arbitrary rationally independent real numbers, we know already that the  $m$  (distinct) almost-periodic functions

$$y_h(t) = \sum_{n=1}^{\mu} \chi'_n(S_h) a_n e^{i\lambda_n t} \quad (h = 1, \dots, m)$$

satisfy the conditions 1—4 of theorem 6. We complete the proof by showing that we can easily make these functions satisfy the additional condition (23), simply by choosing the (thus far arbitrary) coefficients  $a_n$  so that the sequence  $|a_1|, \dots, |a_\mu|$  of their absolute values decreases rather strongly. In order not to spoil too much the generality of our example by putting too great restrictions on the  $|a_n|$ 's, we base our limitation on the respective orders  $\gamma_n$  ( $1 < \gamma_n \leq m$ ) of the characters  $\chi'_n$  in our minimal generating system  $\chi'_1, \dots, \chi'_\mu$ . Since  $(\chi'_n)^{\gamma_n} = 1$ , so that each of the  $m$  values  $\chi'_n(S_1), \dots, \chi'_n(S_m)$  is a  $\gamma_n$ -th root of unity, for any two distinct substitutions  $S_h$  and  $S_g$  of  $\Gamma$  the difference

$\chi'_n(S_h) - \chi'_n(S_g)$  is either 0 or numerically  $\geq 2 \sin \frac{\pi}{\gamma_n}$ . Hence for any two distinct functions  $y_h(t)$  and  $y_g(t)$  and any real  $t$  it is obvious that

$$|y_h(t) - y_g(t)| = \left| \sum_{n=1}^{\mu} a_n (\chi'_n(S_h) - \chi'_n(S_g)) e^{i\chi_n t} \right| \geq 2|a_p| \sin \frac{\pi}{\gamma_p} - 2(|a_{p+1}| + \dots + |a_{\mu}|),$$

where  $p = p(h, g)$  denotes the smallest (certainly existing) index among  $1, \dots, \mu$  for which  $\chi'_p(S_h) \neq \chi'_p(S_g)$ . Thus we see that our condition (23) is certainly satisfied if we choose the coefficients  $a_1, \dots, a_{\mu}$  such that for each  $n = 1, \dots, \mu$ ,

$$(24) \quad A_n = |a_n| \sin \frac{\pi}{\gamma_n} - (|a_{n+1}| + \dots + |a_{\mu}|) > 0;$$

in fact these conditions imply

$$\inf_{\substack{-\infty < t < +\infty \\ h \neq g}} |y_h(t) - y_g(t)| \geq 2 \cdot \min(A_1, \dots, A_{\mu}) > 0.$$

Remark. If we wish to have conditions on the  $a_n$ 's which depend only on the group  $\Gamma$  rather than on the choice of the minimal generating system  $\chi'_1, \dots, \chi'_{\mu}$  of  $\Gamma^*$ , we may replace the conditions (24) by the somewhat stronger conditions

$$(25) \quad |a_n| \sin \frac{\pi}{\gamma} - (|a_{n+1}| + \dots + |a_{\mu}|) > 0 \quad (n = 1, \dots, \mu),$$

where  $\gamma$  ( $\geq$  each  $\gamma_n$ ) denotes the highest order of any element  $\chi$  in  $\Gamma^*$  (or, equivalently, of any element  $S$  in  $\Gamma$ ); or we may go further and replace  $\gamma$  by  $m$ .

**Example.** As the simplest possible example of a transitive but not cyclic Abelian substitution group, we consider the group  $\Gamma$  of the following four substitutions:

$$S_1 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \end{pmatrix} \quad S_2 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{pmatrix} \quad S_3 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 1 & 2 \end{pmatrix} \quad S_4 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 \end{pmatrix}$$

for which a complete table of character values is

	$\chi_1$	$\chi_2$	$\chi_3$	$\chi_4$
$S_1$	1	1	1	1
$S_2$	1	1	-1	-1
$S_3$	1	-1	1	-1
$S_4$	1	-1	-1	1

Here  $\mu = 2$ , and a minimal generating system of  $\Gamma^*$  is formed by any pair from  $\chi_2, \chi_3, \chi_4$ , say by  $\chi_2, \chi_3$ . Hence, if we let  $\lambda_1$  and  $\lambda_2$  be any two real numbers with an irrational ratio, and let  $a_1$  and  $a_2$  have any complex values other than 0, the four functions

$$y_1(t) = a_1 e^{i\lambda_1 t} + a_2 e^{i\lambda_2 t}$$

$$y_2(t) = a_1 e^{i\lambda_1 t} - a_2 e^{i\lambda_2 t}$$

$$y_3(t) = -a_1 e^{i\lambda_1 t} + a_2 e^{i\lambda_2 t}$$

$$y_4(t) = -a_1 e^{i\lambda_1 t} - a_2 e^{i\lambda_2 t}$$

will satisfy conditions 1—4 of theorem 6, and hence have  $\Gamma$  as almost-translation group; and will satisfy the algebraic equation

$$y^4 - 2y^2(a_1^2 e^{2i\lambda_1 t} + a_2^2 e^{2i\lambda_2 t}) + (a_1^2 e^{2i\lambda_1 t} - a_2^2 e^{2i\lambda_2 t})^2 = 0$$

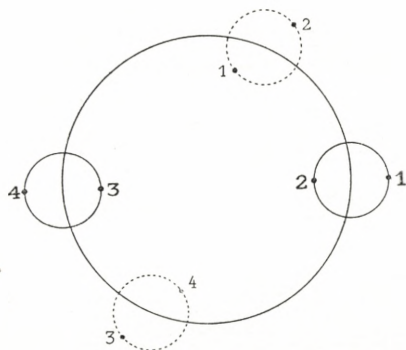
of type (21). If in addition (since in (25) we have  $\gamma = \gamma_1 = \gamma_2 = 2$ ) we take  $|a_1| > |a_2| > 0$ , the functions will satisfy (23), that is the discriminant of the equation will satisfy (22).

In the figure, where we have chosen  $a_1 > a_2 > 0$ , the positions in the complex plane of the four functions  $y_1(t), y_2(t), y_3(t), y_4(t)$  are denoted respectively by 1, 2, 3, 4 when  $t = 0$ , and by 1, 2, 3, 4 for some other value of  $t$ .

In order to show that this set of four almost-periodic functions actually has the non-cyclic group  $\Gamma$  as its almost-translation group — which simple observation was a starting-point



of this investigation — we need not, of course, refer to the general theory above. In fact, it is obvious from the very figure that the almost-translation group of the set can contain no other substitutions than those in  $T$ . That each of these four substi-



tutions actually is an almost-translation substitution of this set of functions, can be verified immediately with the help of KRONECKER'S theorem for the case  $N = 2$ .

Remark 1. In the special, classical, case where the coefficients  $x_j(t)$  of the equation (21) are continuous pure periodic functions with a common period, say with least positive common period  $p$ , (and where the condition  $GLB |D(t)| > 0$  is equivalent to the condition  $D(t) \neq 0$  for  $0 \leq t < p$ ), the case lies so clear and is so well known that we may leave to the reader the slight task of showing how the theory developed in this paper for the general case of almost-periodic coefficients may be applied to deriving the periodicity of the roots, and the essential features of their (ordinary) Fourier series. There is one point however which it may be worth while to emphasize: namely, that in this case the almost-translation group  $T$  (whether it is transitive or intransitive) becomes, as is to be expected, merely the usual "exact translation group" of the set of roots, and hence is simply the cyclic group

composed of all powers of that substitution  $S$  which is (exactly) performed by  $p$ . For it is clear on the one hand, that any integral multiple  $gp$  of  $p$  exactly performs  $S^g$  i. e.  $\varepsilon$ -performs  $S^g$  for every positive  $\varepsilon$ ; and on the other, that each "fine" translation number of the majorant  $X(t)$  must be "very near" some  $gp$ , and hence must  $\varepsilon$ -perform the same substitution  $S^g$  which  $gp$  exactly performs.

Remark 2. Of somewhat greater interest is the case where the coefficients  $x_j(t)$  in equation (21) are limit-periodic functions with a common period  $p$ ; that is, expressed in terms of the exponents, the case in which the module  $M_X$  contains only rational multiples of a single number, namely  $\alpha = \frac{2\pi}{p}$ . From the general relation

$$M_Y \subseteq \frac{1}{m!} M_X$$

it follows that the exponents of the roots  $y_h(t)$  likewise are rational multiples of  $\alpha$ , and hence *the roots themselves are limit-periodic functions with common period  $p$* . We may also remark that the really interesting case of a non-cyclic transitive group — interesting because it can never occur when the roots are pure periodic — is equally impossible in the limit-periodic case. That is to say, in the limit-periodic case *a transitive almost-translation group  $\Gamma$  must be a cyclic group*, i. e. the number  $\mu$  must be equal to 1. For, by theorem 7 of section VI, the assumption  $\mu > 1$  would imply that the module  $M_Y$  contained at least two rationally independent numbers, contrary to what we have just seen.

Finally we prove the following theorem concerning the canonical representation of the functions  $y_h(t)$  mentioned at the end of section VI. We suppose here that  $[y(t)]$  is the set of roots of an algebraic equation (21) with almost-periodic coefficients  $x_j(t)$  satisfying condition (22), and that its almost-translation group  $\Gamma$  is transitive.

We denote the canonical representation of the functions  $y_h(t)$  by

$$(26) \quad y_h(t) = \sum_{\chi \in \Gamma^*} \chi(S_h) \Phi_\chi(t) \quad (h = 1, \dots, m),$$

where the almost-periodic functions

$$\Phi_\chi(t) \infty \sum_{\lambda_n = \chi} a_n e^{i\lambda_n t}$$

are independent of  $h$ . As before,  $X(t)$  denotes a majorant of the  $x_j(t)$ 's, and  $M_X$  the module of their Fourier exponents.

**Theorem 9.** *If  $\chi$  is any fixed character in  $\Gamma^*$ , then the difference  $\lambda' - \lambda''$  of any two Fourier exponents of  $\Phi_\chi(t)$  is contained in the module  $M_X$ .*

*Proof.* Let  $\lambda'$  and  $\lambda''$  be two distinct Fourier exponents of the function  $\Phi_\chi(t)$ . (The proof either "goes by default" or is trifling in case  $\Phi_\chi(t)$  is identically 0, or has just one Fourier exponent). Let  $a' e^{i\lambda' t}$  and  $a'' e^{i\lambda'' t}$  be the corresponding terms in the Fourier series of  $\Phi_\chi(t)$  (and hence also in the Fourier series of  $f_1(t)$ ).

In order to prove that  $\lambda' - \lambda''$  belongs to  $M_X$  it suffices, from lemma 2 in section IV, to prove that to any  $\varepsilon > 0$  there corresponds a  $\delta = \delta(\varepsilon) > 0$  such that the inequality

$$|e^{i(\lambda' - \lambda'')\tau} - 1| \leq \varepsilon,$$

i. e. the inequality

$$(27) \quad |e^{i\lambda'\tau} - e^{i\lambda''\tau}| \leq \varepsilon,$$

is satisfied by every  $\tau$  belonging to  $\{\tau_X(\delta)\}$ . Now, since

$$\{\tau_X\} \subseteq \{\tau_{[y]}\} \quad \text{and} \quad \{\tau_{[y]}\} \subseteq \{\tau_X\},$$

this is equivalent to proving that to the given  $\varepsilon$  there corresponds a  $\delta' = \delta'(\varepsilon) > 0$  such that (27) is fulfilled by every  $\tau$  belonging to  $\{\tau_{[y]}(\delta')\}$ . But for  $\delta$  sufficiently small (in fact for  $\delta \leq \varepsilon^{**}$ ) we know from remark 4° in section II that  $\{\tau_{[y]}(\delta')\}$  is equal to the set-theoretical sum of the  $m$  sets  $\{\tau_{(S)}(\delta')\}$ , where  $S$  belongs to  $T$ . Hence it suffices to prove that to the given  $\varepsilon$  there corresponds a  $\delta'' = \delta''(\varepsilon) > 0$  such that the inequality (27) is satisfied whenever  $\tau$  belongs to one of the sets  $\{\tau_{(S_h)}(\delta'')\}$ . We complete the proof by showing that  $\frac{\varepsilon}{2}a$ , where  $a$  denotes  $\min(|a'|, |a''|)$ , is a suitable value of  $\delta''$ .

In fact, let  $\tau$  be a fixed number belonging to any one of the  $m$  sets in question, say to  $\{\tau_{(S_h)}(\delta'')\}$ . Then by the very definition of the translation numbers  $\tau_{(S_h)}(\delta'')$  we have

$$(28) \quad |y_1(t+\tau) - y_h(t)| \leq \delta'' = \frac{\varepsilon}{2}a \quad (\text{for } -\infty < t < +\infty).$$

In the Fourier series of the almost-periodic function

$$y_1(t+\tau) - y_h(t)$$

the coefficients belonging to the exponents  $\lambda'$  and  $\lambda''$  are given by

$$a' e^{i\lambda'\tau} - a' \chi(S_h) \quad \text{and} \quad a'' e^{i\lambda''\tau} - a'' \chi(S_h)$$

respectively. As any Fourier coefficient of an almost-periodic function  $f(t)$  is numerically  $\leq LUB|f(t)|$ , we conclude from (28) that

$$|a' e^{i\lambda'\tau} - a' \chi(S_h)| \leq \frac{\varepsilon}{2} a \quad \text{and} \quad |a'' e^{i\lambda''\tau} - a'' \chi(S_h)| \leq \frac{\varepsilon}{2} a,$$

and hence (dividing by  $|a'|$  and  $|a''|$  respectively)

$$|e^{i\lambda'\tau} - \chi(S_h)| \leq \frac{\varepsilon}{2} \quad \text{and} \quad |e^{i\lambda''\tau} - \chi(S_h)| \leq \frac{\varepsilon}{2}.$$

Thus  $\tau$  satisfies the inequality  $|e^{i\lambda'\tau} - e^{i\lambda''\tau}| \leq \varepsilon$ , i. e. the inequality (27), and the proof is complete.





Det Kgl. Danske Videnskabernes Selskab.

Mathematisk-fysiske Meddelelser. **XV**, 13.

---

A GENERAL CHEMICAL METHOD  
FOR THE PREPARATION OF THE  
DEUTERATED BENZENES

BY

A. LANGSETH AND A. KLIT



KØBENHAVN

LEVIN & MUNKSGAARD

EJNAR MUNKSGAARD

1937

Printed in Denmark.  
Bianco Lunos Bogtrykkeri A/S.



## Survey of the Method.

The Raman spectra of the deuterium derivatives of benzene should furnish valuable information concerning the internal vibrations of the benzene molecule, and from these one should also be able to make direct inferences concerning the structure of benzene. We have therefore prepared each of the derivatives in pure form and have investigated their spectra. In this paper we shall report only the methods of preparation.

The introduction of deuterium into the benzene nucleus has usually been effected by exchange between benzene and some deuterium compound, with or without a catalyst. The exchange method is quite satisfactory for preparing hexadeuterobenzene. For the preparation of the intermediate deuterium derivatives in pure form, however, the method of direct exchange is not applicable, since one can obtain thereby only mixtures of the various compounds. It is therefore necessary to resort to chemical procedures which can be depended upon to introduce deuterium into definite places in the ring. Two such procedures which immediately suggest themselves are: (1) the decarboxylation of the various calcium benzene-carboxylates with  $\text{Ca}(\text{OD})_2$ , and (2) the replacement of halogen atoms with deuterium by means of a Grignard reaction.

The first of these methods was used by ERLÉNMEYER (Helv. chim. acta 18, 1464 (1935)) to prepare hexadeutero-

benzene. However, the attempts of REDLICH (Monatsh. 68, 374 (1936)) to prepare *p*-tetra deuterobenzene in this way led to a mixture of the three tetra compounds, a result which is not surprising in view of the high temperatures required for decarboxylation. The decarboxylation method therefore appears to us to be unsafe for the preparation of any pure deuterium derivatives except the hexa compound.

The first use of the Grignard method was that of REDLICH (Monatsh. 67, 213 (1936)) who prepared mono-deuterobenzene by hydrolyzing  $C_6H_5MgBr$  with  $D_2O$ . He also used the same procedure for making the para-di derivative since it is possible to prepare  $p-C_6H_4(MgBr)_2$ . The Grignard method has the strong disadvantage that save for one or two exceptions such as the preceding, only one of the halogens in a polyhalogenated benzene will react with magnesium to form a Grignard reagent. In order to replace more than one halogen with deuterium, it is necessary to carry out the reaction stepwise, preparing a new Grignard reagent after each hydrolysis. This entails the wasteful and troublesome purification and drying of each intermediate halogen-deuterium compound from which the succeeding Grignard reagent is to be prepared. A further disadvantage is that the hygroscopic nature of the basic magnesium halide produced by the hydrolysis necessitates the use of a large excess of  $D_2O$ . This excess, as well as the deuterium in the basic halide, is extremely troublesome to recover.

All of these difficulties are obviated by using  $DCl$  rather than  $D_2O$  for breaking up the Grignard complex. The decomposition of the complex with dry  $DCl$  gas may be carried out in the ether solution in which the complex has been formed. We have found that magnesium does not react with  $DCl$  in the dry ether solution and more-

over that it is still possible for the magnesium under these circumstances to form a new Grignard compound. Therefore if the introduction of the DCl is begun as soon as the appropriate polyhalogenated benzene has been transformed to the first Grignard complex, one will obtain straightaway a monodeutero-halogen-compound. The latter will react again with magnesium, forming a new Grignard complex which will in turn be decomposed by the DCl. The replacement of several halogen atoms by deuterium can thus proceed in an effectually continuous fashion. The fact that the reaction can be carried out in a single vessel without isolation of the intermediate compounds results in tremendous saving of time and materials.

The DCl required for the decomposition of the Grignard compound is generated by addition of  $D_2O$  to  $SOCl_2$ . In this way the gas can be produced in a high state of purity with little loss of deuterium (ca. 10–15%). This loss is quite small in comparison with the previously mentioned losses of deuterium attendant upon the hydrolysis of the Grignard compound, inasmuch as here the final magnesium compound is a normal rather than a basic salt.

A further convenience of the method is the ease with which the deuterated benzene may be obtained in purified form from the reaction mixture. It is only necessary to distill off the ether solution in vacuo from the crystalline magnesium salts, wash and dry it, and then remove the ether from the benzene by fractionation through an efficient column.

There is one feature of the DCl method which might lead to a rather considerable loss of deuterium. If the DCl be introduced into the reaction vessel so rapidly that an excess is present, some of it may react with magnesium

iodide to form DI. The DI in turn reacts with the diethyl ether to produce ethyl iodide (cf. the well-known ZEISEL method for the estimation of alkoxy groups); from the latter, mono-deuteroethane is formed by Grignard reaction. The amount of ethane given off depends only on the extent of the excess of DCl, and consequently the stream of gaseous ethane may serve as a rather sensitive indicator for the excess. With help of this indicator, the most effective rate of admission of DCl gas to the reaction vessel may be closely determined. Furthermore it is obvious that the replacement of halogen by deuterium in the benzene nucleus has been completed when it is no longer possible to cut off the evolution of ethane by slowing down the addition of DCl.

Following the program we have just outlined, we have prepared all of the deuterated benzenes in amounts varying from seven to fourteen grams. The success of the preparations depends largely on careful control of the rate of addition of DCl. Therefore it should be emphasized that close attention must be paid to the details of the addition which are described in the experimental part.

The purity of the products may be indicated briefly by saying that their boiling point ranges were quite narrow, and that their Raman spectra showed only the presence of small amounts of benzenes with lower deuterium content, but no traces of isomers or of extraneous impurities. It is quite important to appreciate the fact that the usual criteria are of little value in judging the purity of these preparations. Analyses of the deuterium content, determinations of density, refractive index and boiling point, and the other customary methods of organic chemistry can in general only confirm that the material under consideration has

the proper hydrogen-deuterium ratio. The boiling point range (not the absolute value of the boiling point, which is not known to us a priori) is a helpful indication of purity, but it must be realized that the great differences in the Raman spectra of different isomers furnish by far the most reliable criterion of the purity of the several preparations.

### Experimental.

#### Preparation of Deuterated Benzenes by Modified Grignard Reaction.

(1) Preparation of DCl by decomposition of  $\text{SOCl}_2$  with  $\text{D}_2\text{O}$ . We prepared DCl by adding  $\text{D}_2\text{O}$  very slowly to nearly-boiling  $\text{SOCl}_2$ . The dropping funnel *A* (Fig. 1) contained the  $\text{D}_2\text{O}$  (99.0—99.6 %). The funnel's outlet had been extended some fifty centimeters by means of a capillary tube (to give increased hydrostatic pressure), and the part of the outlet tube which dipped below the surface of the  $\text{SOCl}_2$  in the flask *B* was drawn to a fine nozzle to prevent the occurrence of any reaction in the tube itself. The oil bath *C* was kept at a temperature of 80—85° during the endothermic hydrolysis of the  $\text{SOCl}_2$ . The reflux condenser *D* returned to *B* any  $\text{SOCl}_2$  carried away by the DCl and  $\text{SO}_2$ . The traps *E* and *F* were cooled by a solid  $\text{CO}_2$ -acetone bath. *E* was maintained at a temperature of —65° to —55° to remove the  $\text{SO}_2$  without attendant condensation of any appreciable amount of DCl. For the same reason the height of the acetone bath was kept a centimeter or so below the end of the inlet tube. The last traces of  $\text{SO}_2$  in the DCl were eliminated in *F*, whose temperature was —78°. The rate of evolution of DCl could be closely regulated by the rate of addition of  $\text{D}_2\text{O}$ . The entire DCl

generator was set up ready for instantaneous use before any attempt was made to start the Grignard reaction.

(2) The modified Grignard reaction. The Grignard reaction was carried out in the three-necked flask *K*. The central neck was attached by means of a ground

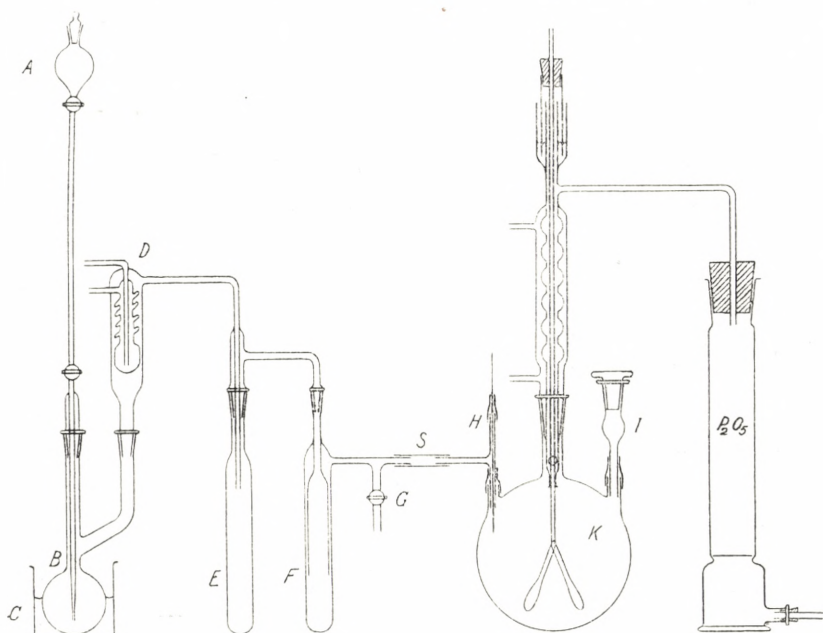


Figure 1.

joint to a reflux condenser through which a mercury seal stirrer extended into the flask. A  $P_2O_5$ -tower closed the exit tube of the condenser to insure the exclusion of atmospheric moisture. The halogen compounds were introduced through the side neck *I*, which was equipped as illustrated in Fig. 1 for the ready introduction of solid material. In case the halogen compound could be added in solution, the introduction was effected by means of a dropping funnel. The inlet tube for the  $DCl$  gas entered

the flask through the side neck *H*. The tube was fitted out with a glass rod with which cakes of the magnesium salts could be removed when they threatened to clog the side neck. To keep the ether vapor out of the trap *F*, a slow stream of dry nitrogen was passed into the system through the stopcock *G*; if in spite of this precaution a counter-distillation of ether took place, the tube could be closed for the time being by means of a pinch clamp at *S*.

The amount of magnesium (equivalent to the quantity of halogen compound) used in each preparation was freshly turned out on a lathe, and immediately covered with dry ether in the flask *K*. In some cases we tried the use of magnesium turnings activated by heating them with a little iodine, but the essential thing appears to be the use of freshly turned magnesium. About one fourth of the halogen compound was then added to the ether together with a small crystal of iodine. Stirring was begun, and continued throughout the entire course of the preparation. The reaction started automatically with the simpler halogen derivatives. With the other compounds, however, it was necessary to keep the ether solution gently boiling by steady electrical heating. A zone of condensation in the reflux condenser was noticed. When this zone began to rise, the reaction was known to have started. If this rise did not occur within half an hour, one c. c. of ethyl bromide and a little more iodine were added to etch the surface of the magnesium turnings. As a rule this procedure caused the reaction to start within a few minutes. If not, the procedure was repeated once or twice until the formation of the Grignard complex had definitely begun. After a short lapse of time, heating was no longer required and it was occasionally necessary to cool the flask. The remainder of the

halogen compound was then added as rapidly as possible without provoking too violent a reaction. Some time later the zone of condensation was observed to fall; this was taken as an indication that the first halogen had reacted with the magnesium.

At this point the introduction of DCl was started. For the quantities of material which we have used (cf. Table I), it was found that the most suitable flow of DCl was obtained by adding  $D_2O$  to the  $SOCl_2$  at the rate of ten grams an hour. Ordinarily this flow of DCl kept the reaction mixture boiling. The heat evolved was in certain cases so considerable that the reaction vessel had to be cooled to prevent too vigorous boiling. For a few of the reactions, however, it was found necessary to supply heat.

The supply of DCl was diminished or cut off altogether if more than a trace of iodine made its appearance in the ether in the reflux condenser. The free iodine originated in the oxidation of DI by a small amount of oxygen present in the stream of the nitrogen. The occurrence of the free iodine was an indication that the DCl was not being used completely for the decomposition of the Grignard complex. The liberation of iodine was always accompanied by the evolution of a gas through the  $P_2O_5$ -tower.

We have mentioned in the introduction that a side reaction something like a ZEISEL reaction occurs when an excess of DCl is present. Such a reaction should produce mono-deuteroethane, and in agreement with expectation, we have found that the gas evolved is inflammable and has about the same density as air. Furthermore if the introduction of DCl be continued after all the magnesium has been consumed, the presence of some ethyl iodide is to be expected in the ether solution. Because of the proximity of



the boiling points of ethyl iodide and benzene, it is difficult to separate these compounds by distillation only. In the first deuterated benzene which we prepared, we observed the presence of ethyl iodide which had failed to be removed by fractionation. This observation confirmed the nature of the side reaction. Henceforth the deuterated benzenes were freed from all ethyl iodide by distillation over freshly turned magnesium.

Although an excess of DCl is thus undesirable, it is not to be presumed that the DCl should be added so slowly as to avoid all possibility of this excess. If the Grignard compound is not decomposed by DCl soon after its formation, subsidiary reactions of the Grignard compound with the halogen atoms on other benzene nuclei may lead to the formation of considerable amounts of compounds of the type of diphenyl and thus reduce the yield of the deuterated benzene.

As we have remarked, the generation of DCl may be closely controlled by the rate of addition of  $D_2O$  to the  $SOCl_2$ . This close control is essential to the success of the whole preparation, and was maintained in the following way. As soon as the first Grignard compound had been formed, the introduction of DCl was begun as previously described. While the decomposition of the Grignard complex was taking place, the gas coming out of the reaction vessel was normally only the nitrogen which was being introduced at *G*. The flow of nitrogen was small, i. e. less than 1 c. c. of gas per second. When the reaction had continued for some time, however, a very considerable increase (ten fold or more) in the outflow of gas was noticed to which attention was directed by the appearance of free iodine in the reflux condenser. Since the increase was due

to the liberation of ethane, the evolution of DCl was cut off completely to allow the excess of DCl to be used up. Ten or fifteen minutes later, the generation of DCl at a somewhat slower rate was resumed and continued until the evolution of ethane again become noticeable. The supply of DCl was again cut off, and the entire procedure repeated. However, if the ethane began coming off with little delay (i. e. within a minute or two), the displacement of halogen by deuterium was known to be complete, and the addition of DCl was stopped. The boiling reaction mixture was allowed to cool, and the process of recovering the deuterated benzene began.

(3) Recovery and Purification of the Deuterated Benzenes. The flask *K* was removed from the apparatus, and the ether solution distilled off as follows: The flask was incorporated in a simple vacuum-tight still, the receiving vessel of which was immersed in an ice-salt mixture. The system was evacuated and then closed off. Distillation was hastened by heating the flask with a water bath, whose temperature was raised gradually to the boiling point. It was found that this procedure removed all of the benzene from the magnesium salts, and left behind only ether of crystallization and certain of the byproducts to which we have referred. Stronger cooling of the receiving vessel merely resulted in the undesirable distillation of these byproducts.

The distillate, yellow to orange in color, was shaken with dilute sodium hydroxide solution, which removed DCl and at the same time eliminated the color. The ether solution was then dried with calcium chloride, and after the addition of fresh magnesium turnings and a little iodine, the

greater part of the solvent was removed by distillation through an efficient column 1.4 meters long.

The remaining solution was transferred in vacuo from the large still to a small one. The latter was designed particularly for the efficient fractionating of small quantities. An indication of its efficiency may be given by remarking that the middle fraction between ether and benzene was in general only a few drops when the rate of distillation was a half dozen drops per minute. Little benzene was lost in the column of the still since the column had been wound with a thermal element by means of which it could be heated at the end of the distillation. The samples of the various deuterated benzenes as a rule were collected within a boiling point range of  $0.2^{\circ}$ , most of the material distilling within a range of  $0.05^{\circ}$ .

In table 1 the chief data for the syntheses of the various benzenes are given. Column II lists the halogen compounds which have been used as the starting point for the syntheses. The preparation of these compounds will be described in the Appendix.

The number of moles of the halogen compounds and of half-moles of  $D_2O$  employed are given respectively in the next two columns. The ratio of the theoretical number of gram-atoms of deuterium to be substituted for halogen to the number of half-moles of  $D_2O$  actually used is listed in column V to illustrate the economy of each preparation. The ratio can of course never exceed unity.

The boiling point range over which the compounds were collected is given in column VI. The temperature at which most of the material distilled is given as the boiling point in the next column. All temperatures have been reduced

Table 1.

Deuterated Benzene (I)	Halogen Compound (II)	Moles of Halogen Compound (III)	Halfmoles of D <sub>2</sub> O (IV)	G-atoms D Halfmoles D <sub>2</sub> O (V)	B. p. range (VI)	B. p. (VII)	Yield		Yield per step % (X)
							g. (VIII)	% (IX)	
C <sub>6</sub> H <sub>5</sub> D	C <sub>6</sub> H <sub>5</sub> Br	0.30	0.60	0.50	79.45-79.65	79.60	13.3	56.1	56.1
1,2-C <sub>6</sub> H <sub>4</sub> D <sub>2</sub>	1,2-C <sub>6</sub> H <sub>4</sub> BrI	0.37 <sub>5</sub>	1.23	0.61	79.85-80.00	79.95	10.6	35.3	59.4
1,3-C <sub>6</sub> H <sub>4</sub> D <sub>2</sub>	1,3-C <sub>6</sub> H <sub>4</sub> BrI	0.30	1.00	0.60	79.65-79.80	79.75	9.0	37.5	61.2
1,4-C <sub>6</sub> H <sub>4</sub> D <sub>2</sub>	1,4-C <sub>6</sub> H <sub>4</sub> Br <sub>2</sub>	0.30	1.03	0.58	79.65-79.75	79.70	13.6	56.6	75.2
1,2,3-C <sub>6</sub> H <sub>3</sub> D <sub>3</sub>	2,6-C <sub>6</sub> H <sub>3</sub> Br <sub>2</sub> I	0.38	2.00	0.57	79.60-79.85	79.80	7.2	23.4	61.6
1,2,4-C <sub>6</sub> H <sub>3</sub> D <sub>3</sub>	2,5-C <sub>6</sub> H <sub>3</sub> Br <sub>2</sub> I	0.40	2.00	0.60	79.60-79.95	79.85	7.3 <sub>5</sub>	22.8	61.1
1,3,5-C <sub>6</sub> H <sub>3</sub> D <sub>3</sub>	1,3,5-C <sub>6</sub> H <sub>3</sub> Br <sub>3</sub>	0.30	1.51	0.60	79.50-79.65	79.55	13.4	55.1	82.0
1,2,3,4-C <sub>6</sub> H <sub>2</sub> D <sub>4</sub>	2,3,6-C <sub>6</sub> H <sub>2</sub> Br <sub>3</sub> I	0.50	2.50	0.80	79.50-79.65	79.60	11.4	27.8	72.6
1,2,3,5-C <sub>6</sub> H <sub>2</sub> D <sub>4</sub>	2,4,6-C <sub>6</sub> H <sub>2</sub> Br <sub>3</sub> I	0.50	2.50	0.80	79.60-79.65	79.65	12.9	31.5	74.9
1,2,4,5-C <sub>6</sub> H <sub>2</sub> D <sub>4</sub>	2,4,5-C <sub>6</sub> H <sub>2</sub> Br <sub>3</sub> I	0.50	3.13	0.64	79.60-79.65	79.60	14.4	35.2	77.0
C <sub>6</sub> H D <sub>5</sub>	2,3,4,6-C <sub>6</sub> HBr <sub>4</sub> I	0.50	3.00	0.83	79.50-79.60	79.60	13.0	31.4	79.3

to standard barometric pressure with the help of the assumption that  $\frac{dT}{dp}$  has the same value as for benzene, namely 0.0043 °/mm at 80°. The boiling points may be seen to vary quite irregularly.

The yield in grams and percent based on the initial amount of halogen compound is given in columns VIII and IX. Since the entire preparation of benzene substituted by  $n$  D-atoms may be regarded as  $n$  consecutive reactions, the mean percentage yield per step may be computed by taking the  $n^{\text{th}}$  root of the entire percentage yield. This quantity is found in the last column. It is seen to lie between 60 % and 80 %.

It may be seen from the table that in the earlier preparations, in which the yield per step is comparatively small, a rather large amount of  $D_2O$  has been used. In these cases we had not yet learned to utilize the evolution of ethane as a guide to the most efficient introduction of  $DCl$ , by means of which the latter portion of each preparation can be regulated in the best way. The greater yield per step for the higher deuterated benzenes, in the preparation of which the detailed directions were followed strictly, shows clearly that the best results are obtained by using 1.2—1.4 halfmoles of  $D_2O$  per gram-atom of deuterium to be introduced.

## Appendix.

### The preparations of the intermediate halogen compounds.

In order to prepare a given deuterated benzene in the manner just described, it is first necessary to select an analogous halogen compound which will react readily with magnesium, give little by-products and be comparatively easy to synthesize. It is well-known that in general a chlorine atom attached to the benzene nucleus will not form a Grignard compound, while a bromine or iodine atom in the same place does react with magnesium. We may therefore eliminate the chlorine derivatives from consideration, and restrict our choice to bromo- and iodo-compounds. We have previously remarked that iodine compounds may produce through

side reactions considerable amounts of by-products which make the deuterated benzene difficult to purify and which reduce the final yield. For these reasons it is advisable to select a halogen compound containing as little iodine as is compatible with satisfactory reactivity.

In this connection we remark that we were not able to get 1.2.4.5-tetrabromobenzene to react with magnesium. By using instead a compound in which one of the bromine atoms had been replaced by iodine, namely 2.4.5-tribromo-iodobenzene, it was possible for us to carry through the synthesis of the corresponding deuterobenzene, although the formation of the first Grignard reagent in this instance required an unusually long time. The nature of the variation in reactivity of different halogen derivatives indicates that not only the kind and number of halogen atoms but also their relative positions in the benzene nucleus must surely play an important role in determining the reactivity of the compounds. Thus we have found that two isomers of 2.4.5-tribromo-iodobenzene, as well as a bromo derivative of the compound, react readily with magnesium. Perhaps this reactivity may be connected with the fact that these three compounds all have a vicinal Br-I-Br arrangement, in contradistinction to 2.4.5-tribromo-iodobenzene.

The importance of having as little iodine as possible in the halogen compounds is clearly evident from table 1, in which it may be observed that para-di- and sym-tri-deuterobenzene have been prepared in relatively high yield from pure bromo compounds. In these cases the iodides in the reaction mixture could be formed only from the small amount of iodine used for activating the magnesium. However, in several cases, even in the simpler preparations, we have used compounds containing one iodine atom purely because it is so much easier to introduce iodine than bromine into benzene by means of the diazo reaction.

In the following the preparation and properties of the intermediate halogen compounds are described. It is true that these compounds have been prepared by methods which are familiar to every chemist. We have described the preparative details, however, because they do not appear to be readily accessible in the literature, and because, in some cases, considerable improvements have been made in procedure. The order of description is determined by the partially deuterated benzenes obtained from the intermediates.

(1) Bromobenzene. The bromobenzene used for the Grignard reaction was the middle fraction obtained from a large amount of commercial bromobenzene by distillation through an efficient column.

(1.2) o-Bromo-iodobenzene. o-Bromo-nitrobenzene (commercial product of Fraenkel & Landau) was recrystallized once from alcohol (m. p. 40.8–41.2°) and then reduced with iron powder and a little sulfuric acid by the method of Holleman (*Recueil des trav. chim.* **25**, 186).

The o-Bromoaniline obtained from one mol of o-Bromonitrobenzene was suspended in water (1–1.5 liters) and concentrated hydrochloric acid (500 c. c.) added. The suspension was chilled in an icebath, and then diazotized by the addition, during the course of 10 to 15 minutes, of sufficient sodium nitrite (ca. 0.95 mol — as indicated by KI-paper). Aqueous potassium iodide (170 g. in 200 c. c.) was added, and after the reaction had been allowed to proceed for one hour, the water layer was decanted and the oil remaining washed with dilute sodium hydroxide. The product was distilled in vacuo (1–2 mms) over  $P_2O_5$ . First fraction: 16 g.; middle fraction: 165 g.; final fraction: 7 g. Total: 188 g. or 71 %; m. p. of middle fraction 9–10°. Holleman (*Recueil des trav. chim.* **34**, 228) gives 5°.

(1.3) m-Bromo-iodobenzene. A similar procedure was followed in preparing m-Bromo-iodobenzene from m-Bromo-nitrobenzene. It was found more efficient, however, to pour the diazotized bromoaniline slowly into a large volume of cold dilute potassium iodide solution. The crude product was distilled with steam. Yield: 165 g. or 57 % m. p. –12° to –10°. Holleman (*loc. cit.*) gives –18°.

(1.4) p-Dibromobenzene. Benzene (270 c. c., dried over  $CaCl_2$ ) and iron powder ("Ferrum reductum", 6 g.) were placed in a one liter two-necked flask equipped with reflux condenser (ground joint) and a dropping funnel. By means of the funnel, whose tip extended beneath the surface of the benzene, enough bromine (5 c. c.) to start the reaction was introduced. When the temperature of the reaction mixture rose and hydrogen bromide began to come off, the remainder of the bromine (290 c. c.) was added during the course of 40 to 50 minutes. When all the bromine had been added, the mixture was poured into a large mortar and macerated with water (ca. 1 liter). The water was drained off, and the crude crystals

were washed and dried with suction; in this way nearly all the ortho-compound may be removed. The material was then recrystallized from alcohol (1 liter) and dried between filter papers. Yield: 390 g. or 55 %/o, m. p. 88°.

(1.2.3) Potassium salt of 2.6-Dibromo-sulfanilic acid. To sulfanilic acid (191 g. — 1 mol.) suspended in water (3 liters), bromine (102 c. c.) dissolved in concentrated hydrobromic acid (200 c. c.) was added with constant stirring during the course of an hour. A small amount of crystalline tribromoaniline formed by a side reaction was filtered off, and potassium hydroxide (60 g.) added to the solution. The latter was concentrated over a flame until crystals appeared. The concentrate was then cooled, and the crude crystals filtered, washed with cold water and recrystallized once or twice from water (until free from bromide ion.) Yield: 65 %/o.

2.6-Dibromo-iodobenzene. The potassium salt (370 g. — 1 mol.) was refluxed with dilute sulfuric acid (340 c. c. conc. sulfuric acid in 160 c. c. water) until complete solution was effected (about 2 hours). The warm solution was poured slowly into a large amount of water, whereupon 2.6-dibromoaniline was precipitated in the form of very fine crystals. The precipitate was washed several times by decantation and then collected on a filter with suction. The moist dibromoaniline was immediately suspended in water (ca. 2 liters) and concentrated hydrochloric acid (400 c. c.) added. The mixture was then chilled and sodium nitrite solution (65—70 g. in sufficient water) added through a dropping funnel over a period of an hour, during which the temperature of the mixture never rose above 5°. A small amount of insoluble material was removed by filtration and the clear solution poured into a cold solution of potassium iodide (250 g.), iodine (250 g.) and concentrated sulfuric acid (250 c. c.) in water (6—8 liters). Steam was passed into the mixture until evolution of nitrogen had ceased, and the free iodine remaining was removed by the addition of a solution of sodium sulfite. The crystalline precipitate of 2.6-Dibromo-iodobenzene was collected on a filter, washed with water and then recrystallized from alcohol. Yield: 195 g. The alcoholic mother liquor could be made to yield 60 g. additional. Total yield: 255 g. or 70 %/o. m. p. 98—99°. WILLGERODT (J. für prakt. Chem. (2) 71,540 (1905)) gives 99°.

Note: It is highly important that the potassium salt be enti-



rely bromide-free, as the presence of a trace of bromide ion will cause the transformation of the dibromoaniline, during the liberation of the sulfonic acid group, into a mixture of mono- and tribromoaniline.

(1.2.4) 2,5-Dibromo-nitrobenzene. To fuming nitric acid (800 c. c.) in a two-liter three-necked flask p-dibromobenzene (400 g) was added, with constant stirring, during the course of 10–15 minutes. The temperature of the reaction mixture was kept below 15° during the nitration by cooling the flask with an ice-salt bath. The stirring was continued for an hour without further cooling, and the mixture was then poured into ice-water. The precipitated 2,5-dibromo-nitrobenzene was washed first by decantation, filtered and then washed free from acid on the filter. Yield: approximately theoretical.

2,5-dibromoaniline. The dibromo-nitrobenzene (281 g. — 1 mol.) was reduced by ordinary iron powder (250 g.) in a large flask containing five liters of water. Sulfuric acid (1 mol.) was added in small portions with continued stirring over a period of an hour. After three hours' further stirring, the mixture was made basic with sodium hydroxide (120 g.) and allowed to settle. After decantation of the supernatant solution the dibromoaniline was distilled with steam. Yield of crude material: 85%.

2,5-dibromo-iodobenzene. The crude dibromoaniline (126 g. — 0.5 mol.) was dissolved in a little alcohol (150 c. c.) and filtered to remove insoluble impurities (trace of iron powder carried over during steam distillation). The alcoholic solution was poured into dilute hydrochloric acid (250 c. c. conc. HCl in one liter water) with stirring. The temperature was kept below 5° with ice during the diazotization. The correct amount (ca. 35 g.) of sodium nitrite was determined by using potassium iodide paper, as before. The diazotized mixture was poured into aqueous potassium iodide (125 g. in 5 liters) and allowed to stand overnight. The precipitated 2,5-dibromo-iodobenzene was steam-distilled after addition of some dilute sodium hydroxide. Yield: 75% — The melting point was not sharp, however, so the material was recrystallized as follows. The crude product (300 g) was dissolved in petroleum ether (400 c. c.) at about 30°. After the solution was cooled to 10°, a large part (160 g.) of the solute crystallized out. The crystals were filtered off, washed with ice cold petroleum ether (100 c. c.) and dried, first between filter papers and then in vacuo over sulfuric acid. m. p.

37.5-37.9°. WILLGERODT (l. c.) gives 38°. Concentration of the mother liquor yielded 80 g additional. The melting point of this latter material extended over a somewhat wider range.

(1.4.5) sym-Tribromobenzene. The preparation of this compound followed Organic Syntheses vol. 13, p. 96.

(1.2.3.4) 2,3,6-Tribromo-iodobenzene. 2,6-Dibromo-iodobenzene (362 g. — 1 mol.) was dissolved in carbon tetrachloride (375 c. c.) in a one-liter two-necked flask, to which a reflux condenser was attached through a ground joint. To this solution were added a little bromine (4 c. c.) and aluminium chloride (2-3 g.). Reaction began at once, and thereupon more bromine (50 c. c.) was introduced without delay through a dropping funnel. Ten minutes after the start of the reaction, the whole mixture was poured into dilute hydrochloric acid. Any free halogen remaining was removed with a little sodium sulfite, and the clear pale yellow carbon tetrachloride was drawn off by means of a separatory funnel. To prevent any crystallization the temperature of the solution was not allowed to drop below 40°. The solution was boiled down to 250-300 c. c., and allowed to stand overnight in an icebox. The crystalline 2,3,6-Tribromo-iodobenzene was filtered off, washed twice with high-boiling (b. p. 60°-70°) petroleum ether and dried first in air and then in vacuo over phosphorus pentoxide. Yield: ca. 70% m. p. 78-79.5°.

(1.2.3.5) 2,4,6-Tribromoaniline. To a solution of aniline (93 g — 1 mol.) in methyl alcohol (1.5 liters) cooled by an ice-bath, bromine (153 c. c.) was added with constant stirring during the course of fifteen minutes. When continued stirring had reduced the temperature of the mixture to 15°, the crystals of the tribromoaniline were filtered off and washed with cold methyl alcohol. Further material was recovered by dropping an equal volume of water into the mother liquor. Total yield: 219 g + 108 g., or 96% m. p. 119°.

2,4,6-Tribromo-iodobenzene. The diazotization was carried out in glacial acetic acid, in the manner suggested by HANTZSCH (Ber. **34**, 3337 (1901)). Tribromoaniline (330 g — 1 mol.) was dissolved in boiling acetic acid (1 liter) and concentrated sulfuric acid (60 c. c.) added. The solution was poured into a further quantity (0.5 liter) of acetic acid and cooled. Amyl nitrite (150 c. c.) was dropped slowly (1-2 hours) with stirring into the mixture, the temperature being kept constant at 15°. The diazotized mixture

was stirred for half an hour longer, and then added to a cold solution of potassium iodide (250 g), iodine (250 g) and concentrated sulfuric acid (200 c. c.) in water (5 liters). The reaction mixture was warmed to 60° and decolorized with sodium sulfite (ca. 300 g.). The precipitate of 2.4.6-Tribromo-iodobenzene was collected on a filter, washed, dried and then recrystallized from alcohol in which decolorizing carbon was suspended. Yield: 260 g. or 59%. m. p. 104—105°. SILBERSTEIN (J. für prakt. Chem. **27**, 120 (1883)) gives 103.5°.

(1.2.4.5) 2.4.5-Tribromo-iodobenzene. The technique used in preparing this compound was the same as that for 2.3.6-Tribromoiodobenzene. Iodobenzene (306 g. — 1.5 mol.), prepared according to Organic Syntheses Coll. Vol. 1, p. 316, was dissolved in carbon tetrachloride (300 c. c.) and "ferrum reductum" (6 g.) added. Bromine (240 c. c.) was introduced through a dropping funnel, the addition requiring half an hour. The reaction mixture was kept at boiling temperature during the addition of the bromine and for ten minutes thereafter, and was then poured into carbon tetrachloride (100 c. c.). The whole volume of carbon tetrachloride solution was shaken with dilute sodium hydroxide, and cooled. The crystals of 2.4.5-Tribromo-iodobenzene were filtered off and recrystallized from carbon tetrachloride to which had been added some decolorizing carbon. Yield: 207 g or 31%. m. p. 165.0—166.7°. WILLGERODT (Chem. Centr. Blatt 1885, 836) gives 164—165°. The yield could be increased to 36% by working over the mother liquors.

(1.2.3.4.5) 2.3.4.6-Tetrabromoaniline. A solution of m-bromoaniline (69 g. — 0.4 mol.) in glacial acetic acid (750 c. c.) was mixed very rapidly with an acetic solution of bromine (64 c. c. in 250 c. c.). The mixture was stirred for two minutes and then poured into a large volume of water. The precipitated 2.3.4.6-Tetrabromoaniline was filtered off and dried with suction. It was recrystallized from alcohol (2 liters). Yield 142 g., + 17 g. from the solvent alcohol, or 96% m. p. 115—116°. CLAUS and WALLBAUM (J. für prakt. Chem. (2) **56**, 50 (1897)) give 115°.

2.3.4.6-Tetrabromo-iodobenzene. This compound was prepared in very nearly the same way as 2.4.6-Tribromo-iodobenzene, except that the amounts of the solvent acetic acid were doubled. The diazotization was judged to be complete when a test sample removed from the mixture was entirely soluble in a small volume

of water. If the sample was not soluble, more amyl nitrite was added. Twice the previous quantity of water was used in making up the iodide solution to which the diazotized mixture was added. The precipitated 2.3.4.6-Tetrabrome-iodobenzene was filtered, washed with alcohol and recrystallized from carbon tetrachloride containing decolorizing carbon. Yield from 1 mol. of tetrabromaniline (410 g.): 279 g. having m. p. 148—149°, and 120 g. recovered from the solvent. m. p. 144—148°. Total yield: 399 g. or 77%.

### SUMMARY

The preparation of the deuterated benzenes from analogous halogen compounds by a modified Grignard reaction is described. The essential feature of the method is that deuterium chloride rather than deuterium oxide is used for the decomposition of the Grignard compounds. The advantages of the method are: 1) the initial compounds are simple to prepare; 2) the replacement of halogen by deuterium is carried out continuously in a single chemical operation; 3) the deuterated benzenes are obtained in pure state and with good yields. Their purity has been ascertained by a study of their Raman spectra.

In an Appendix the preparation of the initial halogen compounds is described in detail.

*Universitetets kemiske Laboratorium,  
Copenhagen.*