

Det Kgl. Danske Videnskabernes Selskab.

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

L'HYDROLYSE
ACIDE ET FERMENTATIVE DES
GLUCOSIDES- β

PAR

STIG VEIBEL ET ERLING FREDERIKSEN



KØBENHAVN
EJNAR MUNKSGAARD

1941

Printed in Denmark.
Bianco Lunos Bogtrykkeri A/S.

Les glucosides- β s'hydrolysent, comme on sait, à la température ambiante sous l'influence des catalyseurs biologiques connus sous le nom d'émulsines ou, avec une détermination plus générale, de β -glucosidase. On sait encore que les glucosides s'hydrolysent aussi sous l'influence des ions hydrogène, mais seulement à des températures si élevées, que toute comparaison directe de l'hydrolyse »enzymatique« et de l'hydrolyse »acide« à une même température est rendue impossible, le catalyseur biologique étant détruit à la température où l'action catalytique des ions hydrogène peut être observée.

Aussi ne trouve-t-on dans la littérature que peu de travaux qui s'occupent d'une telle comparaison. L'examen cinétique de l'hydrolyse acide des glucosides reste, après tout, très incomplet, et les résultats obtenus ne donnent aucune idée générale du mécanisme du procès.

R. KUHN (1) est, autant que nous savons, le premier qui se soit occupé de cette question d'après des idées modernes. KUHN est d'avis qu'en mesurant la vitesse d'hydrolyse d'un glucoside on détermine la décomposition du complexe substrat-catalyseur, pour l'hydrolyse acide aussi bien que pour l'hydrolyse enzymatique, mais il n'a pas déterminé les constantes de dissociation des deux complexes, et par conséquent on ne sait si les concentrations des complexes sont identiques dans les deux cas, même si l'on se sert de solutions d'une même concentration en ce qui concerne le glucoside. KUHN a examiné l'hydrolyse de l'hélicoside, du salicoside, du β -phénylglucoside et du β -méthylglucoside, catalysée soit par l'acide chlorhydrique à 77° soit par l'émulsine à 30°, et

il n'a trouvé aucun parallélisme entre les deux séries de constantes. Il se borne à constater ce fait, en ajoutant seulement qu'il est incapable d'en donner une explication.

JOSEPHSON (2) rappelle des expériences sur l'hydrolyse acide ou enzymatique du saccharose qui ont montré qu'ici les constantes sont les mêmes pour les deux types différents de catalyseurs, mais lui non plus n'a trouvé un tel parallélisme en ce qui concerne les glucosides- β ; il ne s'est occupé cependant que superficiellement de la question.

MOELWYN-HUGHES (3) a examiné la question plus à fond. Il appelle l'attention sur le fait que l'énergie d'activation est plus grande pour l'hydrolyse catalysée par des ions hydrogène que pour l'hydrolyse enzymatique. Comme il faut examiner les deux types différents d'hydrolyse à des températures différentes, une comparaison des deux séries de constantes n'est possible que si les énergies d'activation sont connues pour toutes les substances en question. MOELWYN-HUGHES a lui-même déterminé l'énergie d'activation en ce qui concerne l'hydrolyse acide, en examinant la vitesse d'hydrolyse des solutions de glucosides- β 0,1 m avec de l'acide chlorhydrique 1 m à 80° et à 60°, et il a trouvé pour le β -méthylglucoside $E = 33730$ cal et, pour le salicoside et pour l'arbutoside, des valeurs un peu inférieures. Il est, comme KUHNE, d'avis que c'est la vitesse d'hydrolyse du complexe H^+ -substrat ou enzyme-substrat qu'on détermine, et qu'on est, par conséquent, en droit de comparer les deux séries de constantes, pourvu qu'on tienne bien compte des différences des énergies d'activation.

HELPERICH et ses collaborateurs (4) ont étudié l'hydrolyse fermentative d'une série de glucosides avec des aglucones aromatiques, et ils ont trouvé qu'un point remarquable pour l'hydrolyse fermentative est qu'un phénol-glucoside- β , dans lequel un atome d'hydrogène en position ortho vis-à-vis du groupement $-OH$ a été remplacé par un groupement $-CH_3$, est hydrolysé beaucoup plus vite que le glucoside non substitué. Cette différence ne se retrouve pas si l'hydrolyse est catalysée par des ions hydrogène, et HELPERICH en conclut qu'il y a une différence de mécanisme entre les deux types d'hydrolyse. Un des élèves de HELPERICH, M. SCHEIBER (5), exprime cette opinion dans les termes suivants:

» . . . Denn betrachtet man die Werte und Befunde der Ferment- und Säurehydrolyse vom kinetischen Standpunkt, so sind sich beide Wirkungen in ihrer sichtbaren Leistung

— der Spaltung — gleich. Aber unter dem Gesichtspunkt der Spezifitätslehre gesehen, muss man die Säurehydrolyse als ein unspezifisches absolutes Mass desjenigen Energiebetrages auffassen, der notwendig ist, um eine Trennung der Glykosidbindung zu bewerkstelligen. Die Fermenthydrolyse würde demgegenüber — in Bestätigung der klassischen Anschauungen der Ferment-Substratverbindung zugleich — als ein spezifisches, subjektives Verhältnis zwischen Ferment und Substrat erscheinen, mithin ein Vorgang sein, der durch äussere und innere Umstände beeinflussbar ist. Die Konsequenz einer solchen Annahme fordert die Differenz der gefundenen Spaltwerte der Fermenthydrolyse gegenüber den Säurezahlen nach oben wie nach unten, da sowohl eine bessere wie eine schlechtere Spaltbarkeit denkbar ist. Die Versuche bestätigen diese Forderung.«

Comme les expériences de HELFERICH et ses collaborateurs aussi bien que celles de SCHEIBER sont toutes faites à une seule température, et comme l'on n'a point essayé de déterminer les constantes de dissociation des complexes substrat- β -glucosidase, le passage cité ci-dessus semble être dénué de sens concret, en ce qui concerne l'énergie d'activation aussi bien qu'en ce qui concerne la spécificité, qui se manifeste dans les valeurs, et des constantes de dissociation des complexes, et des constantes de la vitesse d'hydrolyse des complexes actifs. Pour en tirer quelque sens il faut dire qu'il prouve les difficultés qui se présentent lorsqu'on compare l'hydrolyse acide avec l'hydrolyse fermentative, car dans le premier cas le catalyseur, les ions hydrogène, se trouve dans une concentration très grande par rapport à la concentration du substrat, tandis que dans le second cas c'est le contraire, la concentration du substrat étant très grande par rapport à celle du catalyseur; c'est dire qu'il faut dans les deux cas déterminer avec un très grand soin l'énergie d'activation et si possible les constantes de dissociation des complexes catalyseur-substrat avant qu'une telle comparaison soit possible.

L'un de nous (6, 7, 8), examinant l'hydrolyse fermentative d'une série des glucosides- β à aglucone alifatique, a indiqué les énergies d'activation et les constantes de dissociation des complexes β -glucosidase-substrat. Nous décrirons ci-dessous un examen de l'hydrolyse enzymatique de deux

amylglucosides- β non examinés jusqu'à présent et des expériences correspondantes pour toute une série de glucosides- β où l'hydrolyse a été catalysée par des ions hydrogène. La comparaison entre les deux types d'hydrolyse est donc rendue possible pour ces glucosides, et nous en discuterons plus bas les résultats.

Partie expérimentale.

1. Substrats. Nous avons examiné les glucosides- β des alcools suivants: méthyl-, propyl-, isopropyl-, diéthylcarbinol, triméthylcarbinol, diméthyl-éthylcarbinol (hydrate d'amylène) et butylcarbinol tertiaire (alcool pseudoamylique). La plupart de ces glucosides ont été décrits par VEIBEL et collaborateurs (9, 10). Nous n'avons pas trouvé mentionnés dans la littérature le diéthylcarbinol- β -d-glucoside et le pseudoamyl- β -d-glucoside; aussi décrirons-nous ici leur préparation et leurs propriétés.

Diéthylcarbinol- β -d-glucoside. 226 g d'acétobromoglucose furent surchargés avec 307 g de diéthylcarbinol. 125 g CO_3Ag_2 furent ajoutés, et pour faciliter la dissolution de l'acétobromoglucose on ajouta de plus 400–500 cm^3 d'éther éthylique anhydre. Le mélange fut ensuite agité à la machine pendant 3 heures, jusqu'à ce qu'un échantillon eût montré qu'il ne restait trace de l'acétobromoglucose. Le mélange fut alors filtré à la trompe, les sels d'argent furent lavés avec de l'éther éthylique anhydre et les filtrats réunis furent évaporés sous pression réduite. Le tétracétate du glucoside cherché ne tarda pas à cristalliser dans le ballon aussitôt que la plus grande partie du solvant fut éloignée. Il fut recristallisé dans l'alcool éthylique dilué. A l'état de pureté le tétracétate montre F. 109° – 110° et $[\alpha]_D^{20} = -26.9^\circ$ (alcool, $c = 1.002$, $\alpha^{20} = -0.540^\circ$, $l = 2$). Rendement 98 g = 43 %.

96 g du tétracétate furent dissous dans 500 cm^3 d'alcool méthylique, 0.1 g de sodium fut ajouté et la solution fut bouillie à reflux au bain-marie pendant 2 heures (ZEMPLEN et PASCU, (11)). Un échantillon étant alors parfaitement soluble dans de l'eau, la solution fut évaporée sous pression réduite, et le résidu fut recristallisé plusieurs fois dans l'acétate d'éthyle. A l'état de pureté le diéthylcarbinol- β -d-glucoside montre F. 130° – 131° et $[\alpha]_D^{20} = -38.0^\circ$ (eau, $c = 1.016$, $\alpha^{20} = -0.770^\circ$, $l = 2$). Rendement 46.5 g = 81 %.

Pseudoamyl- β -d-glucoside. L'alcool pseudoamylique a été préparé selon les indications de VEIBEL et ses collaborateurs (12). 180 g d'acétobromoglucose furent dissous dans un mélange de 257 g d'alcool pseudoamylique et de 500—600 cm³ d'éther éthylique anhydre. 86 g de CO₃Ag₂ furent ajoutés et le mélange fut agité à la machine pendant 4 heures et demie. Après filtration à la trompe le filtrat fut évaporé sous pression réduite, le résidu fut distillé avec des vapeurs d'eau, de sorte que l'alcool pseudoamylique fut enlevé aussi complètement que possible. Le tétracétate brut du glucoside cherché fut recristallisé dans de l'alcool dilué. A l'état de pureté on trouve F. 134°—135° et $[\alpha]_D^{20} = -26.9^\circ$ (alcool, $c = 1.041$, $\alpha^{20} = -0.560^\circ$, $l = 2$). Rendement 76 g = 41.5 %.

Le tétracétate (70 g) fut privé des groupes d'acétyle par le procédé ordinaire. Recristallisé dans l'acétate d'éthyle jusqu'au point de fusion constant, le glucoside montre F. 136°—137° et $[\alpha]_D^{20} = -36.0^\circ$ (eau, $c = 1.500$, $\alpha^{20} = -1.080^\circ$, $l = 2$). Rendement 19.7 g = 47 %.

Méthodique. Dans les expériences décrites ci-dessous la concentration du glucoside a été 0.1 m ou 0.2 m, la concentration de l'acide chlorhydrique a été 1.0 m ou 0.5 m. La température a été maintenue à l'aide d'un thermostat à huile, chauffé électriquement et réglé par un régulateur à mercure; la constance de la température a été $\pm 0.1^\circ$.

Pour une expérience on pèse 0.005 ou 0.010 mol du glucoside dans un matras calibré de 25 cm³. Le glucoside est dissous dans 20 cm³ d'eau et l'on place le matras dans le thermostat où se trouvent aussi un matras avec de l'eau, un matras avec de l'acide chlorhydrique 2.0 m ou 1.0 m et un tube de verre Pyrex, contenant environ 50 cm³. Après 20 à 30 min. le matras calibré est rempli d'eau jusqu'à la marque. On agite bien le matras pour assurer l'homogénéité du contenu et on le replace dans le thermostat. 20 cm³ de la solution sont transvasés dans un tube d'essai de verre ordinaire, qui est placé ensuite dans le tube de verre Pyrex (voir fig. 1) et entouré de 20 cm³ d'acide chlorhydrique. On

ferme le tube d'un bouchon de caoutchouc et on le laisse dans le thermostat encore 20 à 30 minutes pour assurer l'égalité de température dans les deux solutions. Pour commencer l'expérience le tube d'essai est percé avec un bâton de verre pointu et on mélange bien les deux solutions en se servant du tube troué comme agitateur. Le tube troué est ensuite enlevé, et à distances convenables on prend des échantillons de 5 cm³ du mélange, qui sont ajoutés à 2 cm³ d'une solution de carbonate de potasse, 20 % en cas d'acide chlorhydrique 1 m, 10 % en cas d'acide chlorhydrique 0.5 m. L'hydrolyse du glucoside est par cette mesure complètement suspendue. On laisse ensuite l'échantillon pendant 3 heures au moins, afin de laisser à la mutarotation le temps de s'effectuer, et on détermine la rotation de la solution à l'aide d'un polarimètre à champ visuel divisé en trois, permettant la lecture exacte de 0.01°, de sorte qu'on puisse juger 0.005°.

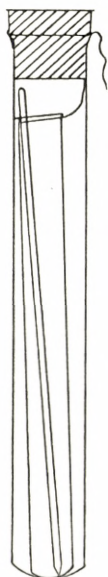


Fig. 1.

L'un de nous (13) a montré que la rotation spécifique des glucosides et du glucose dépend du p_H de la solution, de sorte que plus grand est le p_H de la solution, plus petit est $[\alpha]_D$ des substances dextrogyres, plus grand est la valeur numérique du $[\alpha]_D$ des substances lévogyres. C'est pourquoi nous avons examiné si, par le procédé suivi ici, il y a un déplacement de $[\alpha]_D$, et nous avons trouvé qu'un tel déplacement n'a pas lieu, le p_H des échantillons ajoutés à la solution de CO_3K_2 étant 6.6—6.7, c'est-à-dire à peu près neutre, et la rotation spécifique du glucose et des glucosides ne commençant à changer qu'à partir de p_H 8 environ.

Tableau I.

Rotation des solutions des différents glucosides et du glucose, 0.1 m à des températures différentes, déterminée à 20°.

Substance	Température	
Méthylglucoside	90°	- 1.010°
»	80°	- 0.990°
»	70°	- 0.970°
»	60°	- 0.950°
Propylglucoside	90°	- 1.320°
»	80°	- 1.300°
»	70°	- 1.280°
»	60°	- 1.260°
Isopropylglucoside	90°	- 1.400°
»	80°	- 1.370°
»	70°	- 1.340°
»	60°	- 1.320°
Diéthylcarbinolglucoside	80°	- 1.425°
»	70°	- 1.400°
»	60°	- 1.375°
»	50°	- 1.350°
Triméthylcarbinolglucoside	50°	- 0.670°
»	40°	- 0.665°
»	30°	- 0.660°
»	20°	- 0.660°
Diméthyl-éthylcarbinolglucoside	40°	- 0.645°
»	30°	- 0.640°
»	20°	- 0.640°
»	0°	- 0.640°
Pseudoamylglucoside	90°	- 1.375°
»	80°	- 1.350°
»	70°	- 1.325°
»	60°	- 1.305°
Glucose	90°	+ 1.410°
»	80°	+ 1.390°
»	70°	+ 1.375°
»	60°	+ 1.360°
»	50°	+ 1.350°
»	40°	+ 1.340°
»	30°	+ 1.335°
»	20°	+ 1.330°
»	0°	+ 1.330°

Pour tenir compte du fait que c'est à une température élevée que les solutions sont 0.1 m ou 0.2 m en regard du glucoside, tandis que c'est à la température ambiante qu'on détermine la rotation des échantillons, nous avons préparé, des différents glucosides et du glucose, des solutions aqueuses qui sont 0.1 m aux températures en question, et nous en avons déterminé les rotations en retirant des échantillons de 5 cm³ à la température élevée et en les ajoutant à 2 cm³ d'eau. Les valeurs de la rotation des échantillons trouvés ainsi sont les valeurs initiales (les solutions des glucosides) ou la valeur finale (la solution de glucose) de la rotation dans les différentes expériences; nous donnons ces valeurs dans le tableau I.

Dans les tableaux suivants les constantes de vitesse sont calculées du temps zéro jusqu'au temps t avec la minute comme unité de temps et avec des logarithmes à base 10.

Tableau II.

n-Propylglucoside 0.1 m. HCl 0.996 m. 80°.

$$\alpha_{\text{init.}} = -1.300^{\circ}. \quad \alpha_{\text{fin.}} = +1.390^{\circ}.$$

t min	α_{obs}	x	c - x	k · 10 ⁴	v	k · v
0	-1.300	—	2.690	—	—	—
15	-1.130	0.170	2.520	18.9	0.0252	0.475
30	-0.960	0.340	2.350	19.6	0.0447	0.874
60	-0.640	0.660	2.030	20.4	0.0696	1.417
120	-0.140	1.160	1.530	20.4	0.0793	1.619
180	+0.240	1.540	1.150	20.5	0.0674	1.381
240	+0.530	1.830	0.860	20.6	0.0505	1.042
$\frac{\sum k \cdot v}{\sum v} = 20.2.$				20.1	0.3367	6.808

En ce qui concerne les valeurs moyennes des constantes, nous nous sommes servis d'une méthode, proposée par M. J. A. CHRISTIANSEN (13 a), d'après laquelle le poids qu'il faut attribuer aux déterminations particulières des constantes est calculé d'après l'équation

$$v = -\text{const.} (1-x')^2 \log(1-x'), \quad [1]$$

x' étant $\frac{x_t}{(c-x)_0}$, c'est-à-dire la fraction du substrat qui a été hydrolysée au temps t de la détermination en question. La valeur moyenne de la constante k est alors

$$k = \frac{\sum k \cdot v}{\sum v}. \quad [2]$$

Voici quelques exemples.

Tableau III.

Isopropylglucoside 0.1 m. HCl 1.002 m. 70°.

$$\alpha_{\text{init.}} = +1.340^\circ. \quad \alpha_{\text{fin.}} = +1.375^\circ.$$

t min	α_{obs}	x	c - x	$k \cdot 10^4$	v	k · v
0	-1.340	—	2.715	—	—	—
30	-1.210	0.130	2.585	7.10	0.0193	0.1372
60	-1.060	0.280	2.435	7.88	0.0380	0.2998
120	-0.790	0.550	2.165	8.19	0.0625	0.5169
180	-0.540	0.800	1.915	8.42	0.0754	0.6351
240	-0.310	1.030	1.685	8.63	0.0798	0.6890
360	+0.060	1.400	1.315	8.75	0.0739	0.6460
480	+0.350	1.690	1.025	8.82	0.0603	0.5316
				8.26	0.4092	3.4556
				$\frac{\sum k \cdot v}{\sum v} = 8.44.$		

Tableau IV.

Triméthylcarbinolglucoside 0.1 m. HCl 0.5005 m. 40°.

$$\alpha_{\text{init.}} = -0.665^\circ. \quad \alpha_{\text{fin.}} = +1.340^\circ.$$

t min	α_{obs}	x	c - x	$k \cdot 10^4$	v	k · v
0	-0.665	—	2.005	—	—	—
30	-0.540	0.125	1.880	9.33	0.0246	0.2298
60	-0.440	0.225	1.780	8.62	0.0407	0.3511
90	-0.330	0.335	1.670	8.82	0.0551	0.4858
150	-0.090	0.575	1.430	9.79	0.0747	0.7307
210	+0.080	0.745	1.260	9.61	0.0797	0.7654
300	+0.310	0.975	1.030	9.64	0.0763	0.7360
420	+0.580	1.245	0.760	10.03	0.0605	0.6072
				9.41	0.4116	3.9060
				$\frac{\sum k \cdot v}{\sum v} = 9.49.$		

Tableau V.

Diméthyléthylcarbinolglucoside 0.1 m. HCl 0.996 m. 40°.

$$\alpha_{\text{init.}} = -0.645^\circ. \quad \alpha_{\text{fin.}} = +1.340^\circ.$$

t min	α_{obs}	x	c - x	$k \cdot 10^4$	v	k · v
0	-0.645	—	1.985	—	—	—
5	-0.410	0.235	1.750	119.6	0.0452	5.41
11	-1.180	0.465	1.520	110.6	0.0694	7.68
15	-0.040	0.605	1.380	109.6	0.0771	8.45
25	+0.280	0.925	1.060	109.8	0.0771	8.46
50	+0.760	1.405	0.580	109.9	0.0437	4.81
90	+1.120	1.755	0.230	108.4	0.0110	1.19
				111.3	0.3235	36.00

$$\frac{\sum k \cdot v}{\sum v} = 111.3.$$

Tableau VI.

Pseudoamylglucoside 0.1 m. HCl 1.005 m. 70°.

$$\alpha_{\text{init.}} = -1.325^\circ. \quad \alpha_{\text{fin.}} = +1.375^\circ.$$

t min	α_{obs}	x	c - x	$k \cdot 10^4$	v	k · v
0	-1.325	—	2.700	—	—	—
30	-1.240	0.085	2.615	4.67	0.0131	0.0613
60	-1.140	0.185	2.515	5.13	0.0267	0.1373
120	-0.900	0.425	2.275	6.20	0.0528	0.3275
180	-0.700	0.625	2.075	6.36	0.0675	0.4292
240	-0.490	0.835	1.865	6.70	0.0767	0.5133
360	-0.170	1.155	1.545	6.74	0.0795	0.5352
540	+0.260	1.585	1.115	7.11	0.0655	0.4658
				6.13	0.3818	2.4696

$$\frac{\sum k \cdot v}{\sum v} = 6.47.$$

Hydrolyse fermentative du diméthyléthylcarbinol- β -d-glucoside, du diéthylcarbinol- β -d-glucoside et du pseudoamyl- β -d-glucoside. L'hydrolyse fermentative de ces glucosides n'étant pas examinée du tout ou n'étant examinée que préliminairement, nous tenons à donner ici les données expérimentales permettant la détermination des constantes de la vitesse d'hydrolyse k_{obs} et des constantes d'affinité K_m pour les trois glucosides, ainsi que la détermination des constantes d'affinité K_{m_2} pour les alcools contenus dans les deux derniers glucosides,

Tableau VII.

Hydrolyse des glucosides- β 0.1 m avec HCl 1.0 m.

Glucoside	Valeur de $k \cdot 10^4$ à diverses températures.								
	0°	20°	30°	40°	50°	60°	70°	80°	90°
Méthyl						0.91	3.93	14.63	53.5
Propyl						1.20	5.27	20.30	69.0
Isopropyl						2.06	8.42	33.9	113.5
Diéthylcarbinol ...					0.74	3.08	13.37	48.1	
Triméthylcarbinol.	0.85	4.97	26.0	116.3					
Diméthyléthylcarb.	0.098	4.25	23.0	111.7					
Pseudoamyl						1.74	6.44	25.9	89.5

Tableau VIII.

Hydrolyse des glucosides- β 0.1 m avec HCl 0.5 m.

Glucoside	Valeur de $k \cdot 10^4$ à diverses températures.				
	30°	40°	50°	70°	80°
Méthyl				1.51	6.02
Propyl				1.86	7.77
Isopropyl				3.09	13.10
Diéthylcarbinol				5.05	19.03
Triméthylcarbinol		9.48	42.8		
Diméthyléthylcarbinol ...	8.13	39.6			

Tableau IX.

Hydrolyse des glucosides- β 0.1 m et 0.2 m avec HCl 1.0 m.

Glucoside	$k \cdot 10^4$	
	0.1 m	0.2 m
Propyl, 80°	20.3	20.1
Isopropyl, 80°	33.9	32.8
Diméthyléthylcarbinol, 40°	111.7	105.7

de sorte qu'on puisse calculer les valeurs correspondantes de k_3 , les constantes de la vitesse d'hydrolyse des complexes ferment-glucoside. Pour K_{m_1} , la constante d'affinité du glucose, la valeur 0.18 déterminée antérieurement (8) a été introduite.

La technique employée a été celle indiquée par VEIBEL et ERIKSEN (7), c'est-à-dire qu'on a dissous le glucoside dans une solution tampon d'acétate, 0.1 m, à p_H 4.4. A l'aide de cette solu-

Tableau X.

Diméthyléthylcarbinol-glucoside. Hydrolyse fermentative. 30°. Glucoside 0.0400 m. Émulsine 0.1324 g dans 50 cm³. sal. f. = 0.34. 1 cm³ de toluène ajouté à 50 cm³.

$$\alpha_{\text{init.}} = -0.345^\circ. \quad \alpha_{\text{fin.}} = +0.620^\circ. \quad \alpha_{\text{Émulsine}} = -0.195^\circ.$$

t min	α_{obs}	x	e - x	k · 10 ⁴
0	-0.540	—	0.945	—
180	-0.445	0.095	0.850	2.56
360	-0.360	0.180	0.765	2.54
540	-0.285	0.255	0.690	2.49
720	-0.215	0.325	0.620	2.58
			moyenne...	2.54

$$k_{\text{obs}}/e (\text{sal. f.}) = 0.56 \cdot 10^{-2}.$$

Tableau XI.

Diméthyléthylcarbinol-glucoside. Détermination de K_m .

e = 0.1324 g dans 50 cm³. sal. f. = 0.34.

1 cm³ de toluène ajouté à 50 cm³.

c	10 ⁴ · k'_{\text{obs}}	e (sal. f.) / k'_{\text{obs}}	10 ⁴ · k'_{\text{obs}} (K_m + e) = 10 ⁴ · k_3 · e (sal. f.)
0.0200	3.20	140.7	0.51
0.0400	2.55	176.5	0.46
0.0800	2.07	217.4	0.46
0.1200	1.83	246.0	0.48
0.1600	1.63	276.2	0.49
0.2400	1.28	351.7	0.49
			moyenne... 0.48

$$K_m = 0.14 \text{ (voir fig. 2).}$$

$$k_3 = 0.11 \cdot 10^{-2}.$$

tion et d'une solution aqueuse d'une préparation d'émulsine à force enzymatique (sal. f.) connue, les deux solutions étant placées dans un thermostat à 30°, on a fait des mélanges qui sont 0.04 m en regard du glucoside (ou pour la détermination de K_m une série de concentrations variées), et qui contiennent de plus une quantité connue du ferment. Pour la détermination de K_{m_2} on s'est servi d'une série de mélanges qui sont 0.04 m en regard

$$e(\text{sal. f.})/k'_{\text{obs}}$$

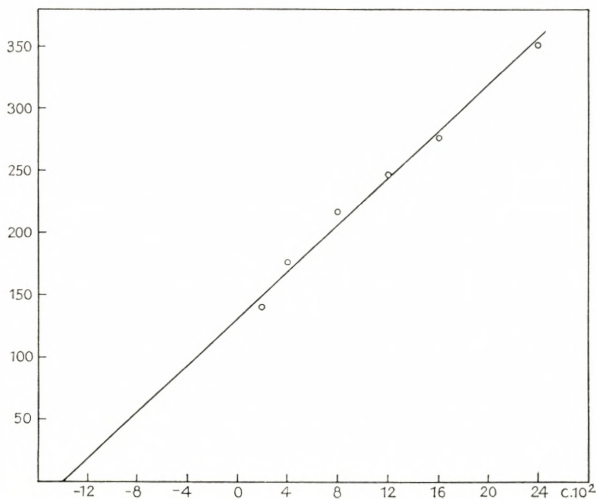
Fig. 2. Diméthyléthylcarbinol-glucoside. Détermination de K_m .

Tableau XII.

Diéthylcarbinol-glucoside. Détermination de K_m .

$$e = 0.0069. \text{ sal. f.} = 0.078.$$

c	$10^4 \cdot k'_{\text{obs}}$	$10^2 \cdot e \cdot \text{sal. f.}/k'_{\text{obs}}$
0.01	22.45	24.2
0.02	19.44	27.7
0.03	15.04	35.8
0.04	12.96	41.5
0.06	8.80	61.2
0.10	6.17	87.2

$$K_m = 0.020 \text{ (voir fig. 3).}$$

du glucoside, mais qui contiennent en plus l'alcool en question dans des concentrations croissantes.

Pour tous ces mélanges les constantes de la vitesse d'hydrolyse sont déterminées en retirant des échantillons de 5 cm³ à distances convenables et en les ajoutant à 1 cm³ d'une solution de CO₃K₂ de 20 %, suspendant ainsi toute hydrolyse fermentative. Le p_H des échantillons est ainsi porté de 4.4 à 10.6–10.7, et en déterminant la rotation des échantillons il faut considérer qu'à

Tableau XIII.

Diéthylcarbinol-glucoside. Détermination de K_{m_2} . $c = 0.0400$. $K_m = 0.020$. $e = 0.0128$. sal. f. = 0.078.

k_{obs}	c_{alcool}	$k_{\text{obs}}/k_{\text{obs,alc}}$	$(k_{\text{obs}}/k_{\text{obs,alc}}) - 1$
			c_{alcool}
24.33	0.00	—	—
21.04	0.01	1.157	15.7
18.58	0.02	1.309	15.5
15.63	0.04	1.556	14.4
11.21	0.08	2.170	14.6
9.13	0.12	2.665	13.9
moyenne...			14.8

$$K_{m_2} = \frac{0.020 \cdot 1}{0.060 \cdot 14.8} = 0.022.$$

Tableau XIV.

Diéthylcarbinol-glucoside. Hydrolyse fermentative. 30° .Glucoside 0.0400 m. Émulsine 0.0064 g dans 50 cm³.

sal. f. = 0.078.

 $\alpha_{\text{mit.}} = -0.640^\circ$. $\alpha_{\text{fin.}} = +0.620^\circ$. $\alpha_{\text{Émulsine}} = -0.010^\circ$. $K_m = 0.020$. $K_{m_1} = 0.18$. $K_{m_2} = 0.022$. $(K_m/K_{m_1} + K_m/K_{m_2} - 1) = 0.02$.

t min	α_{obs}	x	c - x	$10^4 \cdot k_{\text{obs}}$	$10^4 \cdot k_3 \cdot e$ (sal. f.)
0	-0.650	—	1.260	—	—
20	-0.570	0.080	1.180	14.25	0.856
40	-0.520	0.130	1.130	9.40	0.565
60	-0.450	0.200	1.060	13.90	0.835
90	-0.390	0.260	1.000	8.43	0.508
120	-0.320	0.330	0.930	10.50	0.632
150	-0.250	0.400	0.860	11.33	0.683
180	-0.190	0.460	0.800	10.47	0.632
240	-0.090	0.560	0.700	9.67	0.584
300	+0.010	0.660	0.600	11.15	0.678
moyenne...				11.0	0.66

$$10^2 \cdot k/e \text{ (sal. f.)} = 220.$$

$$10^2 \cdot k_3 = 13.3.$$

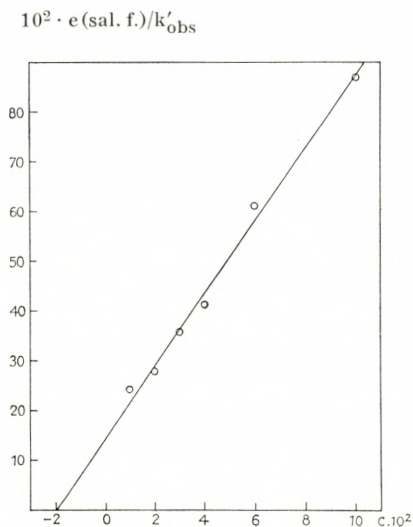
Fig. 3. Diéthylcarbinol-glucoside. Détermination de K_m .

Tableau XV.

Diéthylcarbinol-glucoside. Hydrolyse fermentative. 20° .Glucoside 0.0400 m. Émulsine 0.0128 g dans 50 cm^3 .

sal. f. = 0.078.

 $\alpha_{\text{init.}} = -0.640^\circ$. $\alpha_{\text{fin.}} = +0.620^\circ$. $\alpha_{\text{Émulsine}} = -0.020^\circ$. $K_m = 0.020$. $K_{m_1} = 0.18$. $K_{m_2} = 0.022$. $(K_m/K_{m_1} + K_m/K_{m_2} - 1) = 0.02$.

t min	α_{obs}	x	c - x	$10^4 \cdot k_{\text{obs}}$	$10^4 \cdot k_3 \cdot e$ (sal. f.)
0	-0.660	—	1.260	—	—
20	-0.590	0.070	1.190	12.46	0.748
40	-0.530	0.130	1.130	10.70	0.643
60	-0.480	0.180	1.080	9.85	0.592
90	-0.370	0.290	0.970	15.53	0.935
120	-0.300	0.360	0.900	10.86	0.654
150	-0.225	0.435	0.825	12.57	0.758
180	-0.160	0.500	0.760	11.91	0.718
240	-0.040	0.620	0.640	12.43	0.751
360	+0.060	0.720	0.540	12.30	0.744
moyenne . . .				12.1	0.73
$10^2 \cdot k_{\text{obs}}/e$ (sal. f.) = 121.			$10^2 \cdot k_3 = 7.3$.	$k_{\beta_{30}}/k_{\beta_{20}} = 1.82$.	

Tableau XVI.

Pseudoamyl-glucoside. Détermination de K_m .

$$c = 0.0475. \text{ sal. f.} = 0.23.$$

c	$10^4 \cdot k'_{\text{obs}}$	$e \cdot \text{sal. f.} / k'_{\text{obs}}$
0.02	7.94	13.8
0.04	5.20	21.0
0.08	3.80	28.8
0.12	2.84	38.5
0.16	2.28	47.9
0.24	1.76	62.1

$$K_m = 0.046 \text{ (voir fig. 4).}$$

Tableau XVII.

Pseudoamyl-glucoside. Détermination de K_{m_2} .

$$c = 0.0400. K_m = 0.046. e = 0.1171. \text{ sal. f.} = 0.23.$$

k_{obs}	c_{alcool}	$k_{\text{obs}}/k_{\text{obs}_{\text{alc}}}$	$\frac{(k_{\text{obs}}/k_{\text{obs}_{\text{alc}}}) - 1}{c_{\text{alcool}}}$
12.4	0.00	—	—
11.6	0.01	1.069	6.9
11.1	0.02	1.117	5.9
10.0	0.04	1.240	6.0
9.3	0.08	1.333	4.2
8.5	0.12	1.459	3.8
moyenne...			5.4

$$K_{m_2} = \frac{0.046 \cdot 1}{0.086 \cdot 5.4} = 0.099.$$

ce p_H les rotations spécifiques, et des glucosides et du glucose, sont tournées de 3—4 % à gauche, et que la rotation d'une solution de glucose, qui est maintenue à ce p_H , se modifie avec le temps, la modification étant, pour une solution 0.1 m, de 0.0033° par heure (VEIBEL et ERIKSEN (13)). Les constantes k_{obs} sont calculées de point en point (unité de temps: minute; logarithmes à base 10); en ce qui concerne les expériences destinées à la détermination de K_m , on calcule de plus les valeurs de k'_{obs} , les constantes de la vitesse d'hydrolyse calculées non pas de point en

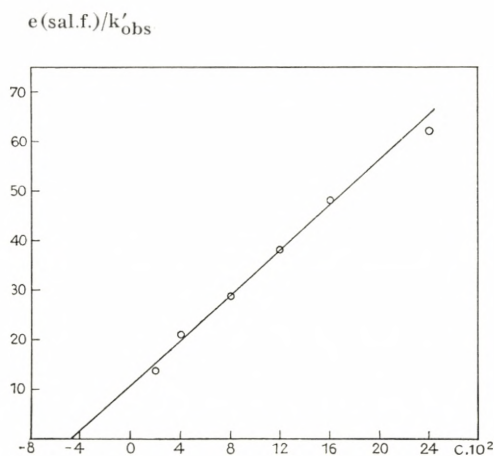
Fig. 4. Pseudoamyl-glucoside. Détermination de K_m .

Tableau XVIII.

Pseudoamylglucoside. Hydrolyse fermentative. 30° .
 Glucoside 0.0400 m. Émulsine 0.2415 g dans 50 cm^3 .
 sal. f. = 0.23.

$$\alpha_{\text{init.}} = -0.620^\circ. \quad \alpha_{\text{fin.}} = +0.620^\circ. \quad \alpha_{\text{émulsine}} = -0.420^\circ.$$

$$K_m = 0.046. \quad K_{m_1} = 0.18. \quad K_{m_2} = 0.099.$$

$$(K_m/K_{m_1} + K_m/K_{m_2} - 1) = -0.28.$$

t min	α_{obs}	x	c-x	$10^4 \cdot k_{\text{obs}}$	$10^4 \cdot k_3 \cdot e(\text{sal. f.})$
0	-1.040	—	1.240	—	—
20	-0.920	0.120	1.120	22.1	1.89
40	-0.805	0.235	1.005	23.0	1.94
60	-0.700	0.340	0.900	24.0	2.00
90	-0.560	0.480	0.760	24.5	2.02
120	-0.435	0.605	0.635	26.0	2.11
150	-0.325	0.715	0.525	27.5	2.20
180	-0.235	0.805	0.435	27.2	2.15
240	-0.100	0.940	0.300	26.9	2.10
300	+0.010	1.050	0.190	33.1	2.55
moyenne...				25.2	2.11

$$10^2 \cdot k_{\text{obs}}/e(\text{sal. f.}) = 4.54.$$

$$10^2 \cdot k_3 = 0.39.$$

Tableau XIX.

Pseudoamyl-glucoside. Hydrolyse fermentative. 20°.
 Glucoside 0.0400 m. Émulsine 0.4576 g dans 50 cm³.
 sal. f. = 0.23.

$$\alpha_{\text{init.}} = -0.620^\circ. \quad \alpha_{\text{fin.}} = +0.620^\circ. \quad \alpha_{\text{Émulsine}} = -0.84^\circ.$$

$$K_m = 0.046. \quad K_{m_1} = 0.18. \quad K_{m_2} = 0.099.$$

$$(K_m/K_{m_1} + K_m/K_{m_2} - 1) = -0.28.$$

t min	α_{obs}	x	c - x	$10^4 \cdot k_{\text{obs}}$	$10^4 \cdot k_3 \cdot e$ (sal. f.)
0	-1.460	—	1.240	—	—
20	-1.355	0.105	1.135	19.2	1.64
40	-1.260	0.200	1.040	18.6	1.57
60	-1.180	0.280	0.960	17.4	1.46
90	-1.070	0.390	0.850	18.0	1.49
120	-0.965	0.495	0.745	19.1	1.56
150	-0.875	0.585	0.655	18.6	1.51
180	-0.805	0.655	0.585	16.4	1.32
240	-0.680	0.780	0.460	17.4	1.38
300	-0.585	0.875	0.365	16.7	1.31
moyenne...				17.9	1.47

$$10^2 \cdot k_{\text{obs}}/e \text{ (sal. f.)} = 1.70. \quad 10^2 \cdot k_3 = 0.14. \quad k_{30}/k_{320} = 2.78.$$

point, mais du temps zéro jusqu'au temps t. Des valeurs de k_{obs} et k'_{obs} on peut calculer K_m , K_{m_2} et k_3 d'après les expressions suivantes (VEIBEL et LILLELUND (8)):

$$k'_{\text{obs}} (K_m + c) = \text{const.} \quad \text{ou} \quad 1/k'_{\text{obs}} \cdot \text{const.} = K_m + c \quad [3]$$

$$K_{m_2} = \frac{K_m \cdot c_{\text{alcohol}}}{(K_m + c) (k_{\text{obs}}/k_{\text{obsale}} - 1)} \quad [4]$$

$$k_3 = k_{\text{obs}} (K_m + c + (K_m/K_{m_1} + K_m/K_{m_2} - 1) x)/e \text{ (sal. f.)}, \quad [5]$$

e étant la quantité du ferment qui se trouve dans 50 cm³ de la solution.

L'équation [3] peut être appliquée pour la détermination graphique de K_m . Des valeurs correspondantes de $1/k'_{\text{obs}}$ et c déterminent une ligne droite, dont le point de section avec l'axe d'abscisse est $-K_m$. Par la détermination graphique les inexactitudes des déterminations particulières de k'_{obs} sont partiellement égalisées.

Tableau XX.

Hydrolyse fermentative des glucosides- β . 0.0400 m.

30° et 20°.

Alcool	$10^2 \cdot k_{\text{obs}}/c$ (sal. f.)		K_m	K_{m_2}	$10^2 \cdot k_3$	
	30°	20°			30°	20°
Méthyl	2.7	1.35	0.62	(∞)	1.89	0.94
Propyl	22.6	11.3	0.16	0.18	5.08	2.34
Isopropyl	16.9	7.9	0.40	0.41	7.51	3.62
Diéthylcarbinol	220	121	0.020	0.022	13.5	7.3
Triméthylcarbinol....	0.03	—	1.46	0.37	(0.05)	—
Diméthyléthylcarbinol	0.56	—	0.14	—	(0.11)	—
Pseudoamyl	4.54	1.70	0.046	0.099	0.39	0.14

Nous n'avons pas déterminé la valeur de K_{m_2} , la constante d'affinité de l'hydrate d'amylène; pour la détermination de k_3 en cas de glucoside de l'hydrate d'amylène il faut donc se servir de l'équation [3]: $k_3 = k'_{\text{obs}}(K_m + c)$, qui donne une valeur approximative de k_3 . En appliquant les valeurs de k'_{obs} indiquées dans le tableau XI, on obtient, tout de même, une approximation assez bonne, car dans ces expériences c'est seulement la phase initiale de l'hydrolyse qu'on a suivie; l'action inhibitive des produits d'hydrolyse n'est par conséquent que faible. Dans le tableau XI nous avons réuni les valeurs calculées pour $k_3 \cdot c$ (sal. f.) pour toutes les concentrations du glucoside examinées, ainsi que la valeur moyenne de k_3 .

Il faut se souvenir que dans toutes les expériences avec le glucoside de l'hydrate d'amylène on a ajouté aux solutions du toluène afin d'empêcher le développement des microorganismes. Les valeurs de k'_{obs} et de k_3 ne sont donc plus directement comparables avec les valeurs correspondantes trouvées pour d'autres glucosides sans application de toluène, car l'addition du toluène tend à augmenter la force enzymatique apparente du ferment (VEIBEL (14)).

Discussion.

Une comparaison des tableaux VII—IX d'une part et du tableau XX de l'autre va montrer qu'il y a une différence essentielle entre l'hydrolyse catalysée par des ions hydrogène et celle catalysée par le ferment β -glucosidase. En ce qui concerne l'hydrolyse acide les constantes de vitesse sont du même ordre de grandeur pour les glucosides des alcools primaires et secondaires, tandis que les constantes de vitesse des glucosides des alcools tertiaires sont supérieures aux autres de plusieurs ordres de grandeur. D'autre part, en ce qui concerne l'hydrolyse fermentative, bien que les constantes de vitesse soient du même ordre de grandeur pour les glucosides des alcools primaires et secondaires, celles des glucosides des alcools tertiaires sont inférieures aux autres de un ou de deux ordres de grandeur. Il ne s'agit pas seulement d'un empêchement stérique, car le diméthyl-éthyl-carbinol-glucoside est hydrolysé plus vite que le triméthyl-carbinol-glucoside, et notamment aussi bien quand l'hydrolyse est catalysée par des ions hydrogène que quand elle est catalysée par le β -glucosidase.

Une comparaison des tableaux VII et VIII montre qu'en cas d'hydrolyse acide il y a une dépendance de la concentration des ions hydrogène ou peut-être de l'activité de ces ions. L'activité des ions hydrogène dans des solutions d'acide chlorhydrique 1.0 m ou 0.5 m à 80° et à 70° n'étant pas connue, nous nous sommes servis des activités correspondantes à 25°, qui sont, d'après DUBOUX (15), 1.57 et 0.57. Le rapport entre ces deux activités est 2.75, tandis que LEININGER et KILPATRICK (16) ont calculé, pour le rapport entre les valeurs de $k_{\text{H}_3\text{O}^+}$ pour 1.0 m HCl et 0.5 m HCl, la valeur 1.317 ($k_{\text{H}_3\text{O}^+}$ étant $k_{\text{obs}}/c_{\text{H}_3\text{O}^+}$), ce qui donne

$k_{\text{obs } 1.0 \text{ m}}/k_{\text{obs } 0.5 \text{ m}} = 2.63$. Expérimentalement (hydrolyse du saccharose) ils ont trouvé 2.79. Ces chiffres sont valables à 0°. Le rapport calculé diminue avec une température croissante; à 40° on calcule 1.232 ou pour $k_{\text{obs } 1.0 \text{ m}}/k_{\text{obs } 0.5 \text{ m}}$ la valeur 2.46.

L'explication donnée par LEININGER et KILPATRICK étant considérée comme correcte, il faut en conclure que, à condition que l'hydrolyse acide des glucosides, comme celle du saccharose, soit catalysée par des ions hydrogène, on trouvera dans une solution 0.5 m en ce qui concerne HCl et en même temps 0.5 m en ce qui concerne KCl à peu près le même effet causé par la présence des électrolytes que celui qu'on trouve dans une solution 1.0 m en ce qui concerne HCl.

Des expériences préliminaires sur l'hydrolyse acide du glucoside de l'hydrate d'amylène ont montré que le rapport $k_{\text{obs } 0.5 \text{ m HCl} + 0.5 \text{ m KCl}}/k_{\text{obs } 0.5 \text{ m HCl}}$ est 1.31, tandis que $k_{\text{obs } 1.0 \text{ HCl}}/k_{\text{obs } 0.5 \text{ HCl}}$ est 2.70, c'est-à-dire que l'effet causé par la présence des électrolytes peut être exprimé dans le dernier cas par le facteur 1.35.

Le tableau IX montre que la vitesse d'hydrolyse est directement proportionnelle à la concentration du glucoside, les constantes de vitesse étant les mêmes pour des solutions 0.1 m et 0.2 m à l'égard du glucoside.

Dans le tableau XXI nous donnons les valeurs des constantes, recalculées avec des logarithmes naturelles et avec la seconde comme unité de temps. Dans le tableau sont indiqués en plus le rapport entre les constantes trouvées avec de l'acide chlorhydrique 1.0 m et 0.5 m ainsi que la valeur de ce rapport calculé d'après une extrapolation des données de LEININGER et KILPATRICK (16) mentionnées plus haut. L'extrapolation est faite graphiquement, voir fig. 5.

On voit que, bien que les valeurs trouvées par nous pour les glucosides ne soient pas coïncidentes avec les valeurs calculées pour le saccharose par LEININGER et KILPATRICK, on trouve ici la même variation du rapport avec la tem-

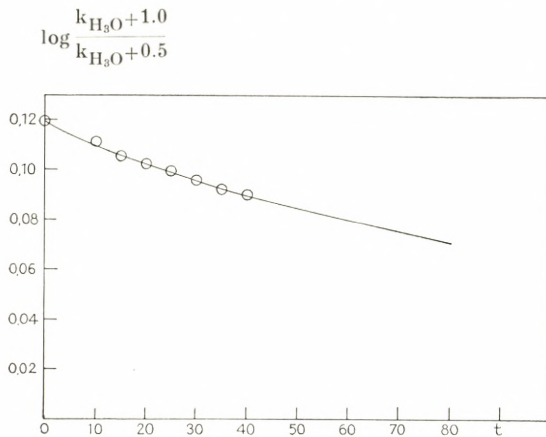


Fig. 5.

pérature que celle qu'on a calculée et trouvée pour l'hydrolyse du saccharose.

Dans le tableau XXII nous indiquons les valeurs de l'énergie d'activation trouvées pour l'hydrolyse des glucosides à des températures différentes, calculées d'après l'équation $Q = R \cdot \frac{T_1 \cdot T_2}{T_1 - T_2} \cdot \ln \frac{k_{T_1}}{k_{T_2}}$. On voit que l'exactitude des expériences ne suffit pas pour la constatation d'une variation éventuelle de l'énergie d'activation avec la température. Quant à l'influence de la concentration de l'acide chlorhydrique il semble, par contre, que l'exactitude soit suffisante pour démontrer une diminution de l'énergie d'activation avec une concentration croissante de l'acide, même effet que LEININGER et KILPATRICK ont trouvé pour l'hydrolyse du saccharose.

Tableau XXI.

Hydrolyse acide des glucosides- β à diverses températures.
Valeurs de $10^5 \cdot k$ (logarithmes naturelles, unité de temps :
seconde).

	Méthylglucoside 1.0 m HCl	Méthylglucoside 0.5 m HCl	Propylglucoside 1.0 m HCl	Propylglucoside 0.5 m HCl	Isopropylglucoside 1.0 m HCl	Isopropylglucoside 0.5 m HCl	Diéthylcarbinolglucoside 1.0 m HCl	Diéthylcarbinolglucoside 0.5 m HCl	Triméthylcarbinolglucoside 1.0 m HCl	Triméthylcarbinolglucoside 0.5 m HCl	Diméthyléthylcarbinolglucoside 1.0 m HCl	Diméthyléthylcarbinolglucoside 0.5 m HCl	Pseudoamylglucoside 1.0 m HCl
$k \cdot 10^5$													
0°	0.038
20°	0.33	..	1.63
30°	1.91	..	8.82	3.12	..
40°	9.97	3.64	42.9	15.2	..
50°	0.28	..	44.6	16.4
60°	0.35	..	0.46	..	0.79	..	1.18	0.67
70°	1.51	0.58	2.02	0.71	3.23	1.19	5.13	1.94	2.47
80°	5.61	2.31	7.79	2.98	13.0	5.03	18.3	7.30	9.92
90°	20.5	..	26.5	..	43.6	34.3
$k_{1.0}/k_{0.5}$													
30°	2.82	..
40°	2.74	..	2.82	..
50°	2.72
70°	..	2.60	..	2.83	..	2.73	..	2.65
80°	..	2.43	..	2.61	..	2.59	..	2.53

Valeurs calculées d'après l'extrapolation des données de
LEININGER et KILPATRICK (voir fig. 5).

t	30°	40°	50°	70°	80°
$k_{1.0}/k_{0.5}$	2.49	2.46	2.43	2.38	2.36
Saccharose					

Tableau XXII.

Énergie d'activation à diverses températures.

	Méthylglucoside	Propylglucoside	Isopropylglucoside	Diéthylcarbinol- glucoside	Triméthylcarbinol- glucoside	Diméthyléthylcarbinol- glucoside	Pseudoamylglucoside
1. Hydrolyse avec 1.0 m HCl.							
20°/ 0°	29950	..
30°/20°	31150	29770	..
40°/30°	31160	29780	..
40°/ 0°	29870	..
50°/40°	30150
50°/20°	30810
60°/50°	30470
70°/60°	33190	33570	31950	33310	29690
80°/70°	31620	32440	33500	30790	33430
80°/50°	31520
90°/80°	33010	31740	30720	31610
90°/60°	32610	32430	32090	31540
2. Hydrolyse avec 0.5 m HCl.							
40°/30°	29830	..
50°/40°	30250
80°/70°	33260	34390	34740	31910

Dans le tableau XXIII nous avons indiqué quelques données concernant l'hydrolyse fermentative des glucosides dont l'hydrolyse acide a été examinée dans ce mémoire. Ces données ont été tirées d'un mémoire de VEIBEL et LILLELUND (8) ou indiquées plus haut dans le présent mémoire.

Les valeurs moyennes de l'énergie d'activation de l'hydrolyse acide des glucosides des alcools primaires, secondaires

Tableau XXIII.

Énergie d'activation. Hydrolyse fermentative.

Glucoside	$10^2 \cdot k_{\text{obs}}$		$10^2 \cdot k_3$		Q
	30°	20°	30°	20°	
Méthyl.....	2.7	1.35	1.89	0.94	12200
Propyl.....	22.6	11.3	5.08	2.34	13500
Isopropyl.....	16.9	7.9	7.51	3.62	13100
Diéthylcarbinol.....	220	121	13.3	7.3	10600
Triméthylcarbinol....	0.03	—	0.05	—	—
Diméthyléthylcarbinol	0.56	—	0.11	—	—
Triéthylcarbinol*)....	2.0	0.73	0.16	0.05	19960
Pseudoamyl.....	4.54	1.70	0.39	0.14	18060

*) VEIBEL et LILLELUND (17).

et tertiaires sont respectivement 32520 cal, 31810 cal et 30340 cal. Q est donc à peu près le même pour les deux premières catégories de glucosides et un peu inférieur pour la troisième catégorie. Le pseudoamylglucoside se range et par la valeur de la constante de sa vitesse d'hydrolyse et par celle de son énergie d'activation parmi les glucosides des alcools primaires, si l'on ne considère que les valeurs trouvées par l'hydrolyse acide.

En regardant les valeurs correspondantes pour l'hydrolyse fermentative (tableau XXIII), on trouvera, premièrement, que l'effet du catalyseur biochimique est, comme d'habitude, de diminuer l'énergie d'activation nécessaire pour réaliser l'hydrolyse, et, deuxièmement, qu'ici aussi les valeurs de Q pour les glucosides des alcools primaires et secondaires sont à peu près égales. Quant à l'énergie d'activation de l'hydrolyse fermentative des glucosides des alcools tertiaires, il faut dire que les deux glucosides en question sont hydrolysés si lentement à 30° qu'une détermination des constantes de la vitesse d'hydrolyse à 20° serait très

inexacte, de sorte que le calcul de Q , basé sur de telles expériences, n'aurait pas grande valeur. Nous avons déterminé antérieurement (VEIBEL et LILLELUND (17)) la constante de la vitesse d'hydrolyse du triéthylcarbinol- β -d-glucoside à 30° et à 20°. Ce glucoside est hydrolysé considérablement plus vite que les deux glucosides d'alcools tertiaires mentionnés dans le présent mémoire, et les déterminations des constantes sont donc du même ordre d'exactitude que celle qui est valable pour les autres glucosides. Ces expériences ont donné pour Q la valeur 19960 cal, valeur considérablement supérieure aux valeurs trouvées pour les glucosides des alcools primaires et secondaires. On trouve donc ici une différence essentielle entre l'hydrolyse acide et l'hydrolyse fermentative.

Le glucoside de l'alcool pseudoamylique se range par la valeur de k_{obs} parmi les glucosides des alcools primaires, mais par la valeur de k_3 et surtout par la valeur de Q , il se rapproche des glucosides des alcools tertiaires.

Cela attire de nouveau l'attention sur les relations stériques, qui peuvent être telles que l'ion hydrogène n'est pas empêché d'exercer son action hydrolysante, tandis que le molécule du catalyseur biochimique, qui est beaucoup plus grand que l'ion hydrogène, en est empêché par les alcyls tertiaires. On peut espérer d'être renseigné sur ces relations en calculant la constante B de l'équation d'ARRHENIUS: $\ln k = \frac{-Q}{RT} + B$, B étant une expression du nombre des collisions effectives, de sorte que plus grande est la valeur de B , plus grand est le nombre des collisions qui ont fait naître une réaction. Il faut admettre que la détermination de B d'après les expériences décrites plus haut ne devient que très peu exacte, car l'intervalle de température accessible à l'expérience n'est forcément qu'une fraction minimale de l'intervalle nécessaire pour une détermination exacte de B , embrassant des valeurs de la température allant de 300° environ jusque vers ∞ (degrés absolus). L'inclinaison de la ligne déterminant B n'est donc connue qu'avec très peu d'exactitude.

Un calcul de B pour l'hydrolyse enzymatique n'est pas possible, car on ne connaît pas la concentration moléculaire du ferment. Il faut s'imaginer, néanmoins, que l'ordre des constantes B sera indépendant du facteur tenant compte du poids moléculaire du ferment, de sorte que les constantes B calculées d'après les valeurs de k_3 indiquées plus haut (tableau XXIII) en ce qui concerne la dépendance de la catégorie de l'alcool présent dans le glucoside, peuvent être comparées avec les constantes B pour l'hydrolyse acide.

Dans le tableau XXIV nous indiquons les valeurs de k, de Q et de B, et pour l'hydrolyse acide, et pour l'hydrolyse enzymatique, les valeurs de $B_{\text{enzymatique}}$ n'ayant, toutefois, aucune signification réelle, mais seulement comparative en ce qui concerne la variation de B avec la catégorie d'alcool.

Pour les calculs nous nous sommes servis, en ce qui concerne l'hydrolyse acide, de la valeur de Q déterminée à l'aide des expériences où la différence de température est la plus grande, soit 30°. La constance de B pour chaque glucoside prouve de nouveau que les expériences décrites dans le présent mémoire ne suffisent pas pour constater une variation éventuelle de Q avec la température.

En ce qui concerne l'hydrolyse enzymatique, la différence de température n'est que de 10°, et pour les glucosides des deux alcools tertiaires dont l'hydrolyse acide a été étudiée, nous n'avons examiné l'hydrolyse enzymatique qu'à une seule température. Pour le calcul de B dans ces cas nous nous sommes servis de la valeur $Q = 19960$ cal trouvée pour le triéthylcarbinol- β -d-glucoside; la détermination de cette valeur a été mentionnée plus haut (p. 28).

On verra au tableau XXIV que les valeurs trouvées pour la constante B ne donnent aucun appui à la théorie d'un empêchement stérique, causant la différence entre les glucosides des alcools tertiaires et ceux des autres alcools. En ce qui concerne l'hydrolyse acide, les valeurs des constantes B sont à peu près les mêmes pour les deux premières catégories de glucosides et supérieures pour ceux des alcools tertiaires, qui sont hydrolysés beaucoup plus vite que les autres. En ce qui concerne l'hydrolyse fermentative, les valeurs de B pour les glucosides des alcools tertiaires sont de beaucoup supérieures aux valeurs de B pour les autres glucosides, qui sont néanmoins hydrolysés beaucoup

Tableau XXIV.
Valeurs de k, Q et B.

Glucoside	1.0 m HCl				0.5 m HCl			Enzymatique				
	t	k · 10 ⁵	Q	B	k · 10 ⁵	Q	B	t	k ₃ · 10 ³	Q	B	
Méthyl	{	60°	0.35	32610	15.96	—	—	—	—	—	—	—
		70°	1.51	32610	15.96	0.58	33260	15.97	30°	0.725	12200	5.7
		80°	5.61	32610	15.96	2.31	33260	15.97	—	—	—	—
		90°	20.5	32610	15.96	—	—	—	—	—	—	—
						16.0			16.0			
Propyl	{	60°	0.46	32430	15.96	—	—	—	—	—	—	—
		70°	2.02	32430	15.98	0.71	34390	16.79	30°	1.95	13500	7.0
		80°	7.79	32430	15.98	2.98	34390	16.78	—	—	—	—
		90°	26.5	32430	15.96	—	—	—	—	—	—	—
						16.0			16.8			
Isopropyl . . .	{	60°	0.79	32090	15.97	—	—	—	—	—	—	—
		70°	3.23	32090	15.97	1.19	34740	17.23	30°	2.88	13100	6.9
		80°	13.0	32090	15.98	5.03	34740	17.22	—	—	—	—
		90°	43.6	32090	15.97	—	—	—	—	—	—	—
						16.2			17.2			
Diéthyl- carbinol . . .	{	50°	0.28	31520	15.80	—	—	—	—	—	—	—
		60°	1.18	31520	15.77	—	—	—	30°	5.10	10600	5.3
		70°	5.13	31520	15.79	1.94	31910	15.63	—	—	—	—
		80°	18.3	31520	15.77	7.30	31910	15.63	—	—	—	—
						15.8			15.6			
Triméthyl- carbinol . . .	{	20°	0.33	30810	17.51	—	—	—	—	—	—	—
		30°	1.91	30810	17.52	—	—	—	30°	0.02	19960	9.7
		40°	9.97	30810	17.52	3.64	30250	16.70	—	—	—	—
		50°	47.6	30810	17.51	16.4	30250	16.70	—	—	—	—
						17.5			16.7			
Diméthyl- éthyl- carbinol . . .	{	0°	0.04	29870	17.50	—	—	—	—	—	—	—
		20°	1.63	29870	17.50	—	—	—	—	—	—	—
		30°	8.82	29870	17.50	3.12	29830	17.02	30°	0.042	19960	10.0
		40°	42.9	29870	17.50	15.2	29830	17.03	—	—	—	—
						17.5			17.0			
Triéthyl- carbinol . . .	{	—	—	—	—	—	—	30°	0.061	19960	10.2	
Pseudoamyl	{	60°	0.67	31540	15.54	—	—	—	—	—	—	—
		70°	2.47	31540	15.50	—	—	—	30°	0.15	18060	9.2
		80°	9.92	31540	15.54	—	—	—	—	—	—	—
		90°	34.3	31540	15.54	—	—	—	—	—	—	—
						15.5			9.2			

plus vite que les glucosides avec les grandes valeurs de B. Le pseudoamylglucoside se range aussi par sa valeur de B parmi les glucosides des alcools tertiaires en ce qui concerne l'hydrolyse fermentative, tandis que, en ce qui concerne l'hydrolyse acide, il est plus rapproché des glucosides des alcools primaires et secondaires.

Les valeurs de la constante B n'autorisent donc pas l'extension de la théorie du rapport entre le volume moléculaire du substituant et la vitesse d'hydrolyse enzymatique, proposée par HELFERICH, GRÜNLER et GNÜCHTEL (18) pour des substituants introduits dans le molécule de glucose d'un glucoside, aux substituants introduits dans l'aglucone, car cette théorie demande que plus grand est le volume moléculaire, plus petite est la vitesse d'hydrolyse.

En ce qui concerne le mécanisme de l'hydrolyse des glucosides il semble utile de se rappeler que les glucosides sont des acétals mixtes. L'idée de comparer l'hydrolyse des acétals et l'hydrolyse des glucosides semble donc bien naturelle.

SKRABAL et ses collaborateurs (19) ont examiné l'hydrolyse acide de quelques formals. Ils ont trouvé les constantes de la vitesse d'hydrolyse suivantes:

Substance	$10^2 \cdot k$ (25°)
Diméthylformal	9.18
Diéthylformal	78
Dipropylformal	86
Diisopropylformal	433
Di-méthyléthylcarbinolformal ..	595

BRØNSTED et GROVE (20) ont trouvé pour le diméthylacétal $k_{H_3O^+} = 3.92$ (20°), et RIESCH et KILPATRICK (21) indiquent pour le diéthylacétal $k_{H_3O^+} = 2.43$ (25°). RIESCH et KILPATRICK ont de plus déterminé l'énergie d'activation. Ils ont trouvé $k_{35}/k_{25} = 3.395$, d'où ils calculent

$Q = 22323$ cal. Plus tard LEININGER et KILPATRICK (22) ont trouvé $k_{35}/k_{25} = 3.996$ ou $Q = 25290$ cal.

On voit donc que les constantes de la vitesse d'hydrolyse pour les acétals des alcools primaires sont relativement petites, celles des alcools secondaires beaucoup plus grandes. Les acétals des alcools tertiaires ne semblent pas avoir été examinés.

Nous avons trouvé que les constantes de la vitesse d'hydrolyse sont du même ordre de grandeur pour les glucosides des alcools primaires et secondaires, celles des alcools secondaires étant peut-être un peu plus grandes que celles des alcools primaires. Si l'on se souvient que dans les glucosides l'acétalisation de l'un des deux groupements $-OH$ a toujours eu lieu avec un alcool secondaire (l'atome de carbone no. 5 dans le glucoside), de sorte que la différence entre les catégories d'acétal, en ce qui concerne les glucosides, est restreinte à l'autre groupement $-OH$, qui peut être remplacé par un groupement $-OR$, où R est l'alcyde d'un alcool primaire, secondaire ou tertiaire, il semble naturel que la différence entre les vitesses d'hydrolyse des deux premières catégories de glucosides ne soit pas aussi grande que celle entre les acétals simples des alcools primaires et secondaires, tandis qu'une différence considérable se manifeste, quand un alcyde tertiaire est introduit dans le molécule du glucose. Il faut tout de même reconnaître que les valeurs indiquées par SKRABAL sont déterminées à 25° , tandis que les valeurs trouvées par nous, en ce qui concerne les glucosides des alcools primaires et secondaires, sont déterminées à une température plus élevée ($60^\circ-90^\circ$). Il y a donc une différence considérable entre les valeurs des constantes d'hydrolyse des acétals simples et celles des glucosides.

Tout en admettant que l'hydrolyse des glucosides puisse être comparée à l'hydrolyse des acétals, on n'est que très peu renseigné sur le mécanisme de l'hydrolyse. HEISIG (23) a montré qu'un glucoside, soumis à l'alcoololyse catalysée par des ions hydrogène, échange son aglucone contre l'alcyde de l'alcool servant comme dissolvant, en subissant en même temps une transposition Walden, de sorte que le α -phénylglucoside, alcoolysé dans l'alcool méthylique contenant de l'acide chlorhydrique, donne naissance au β -méthylglucoside. Si le même mécanisme est suivi pendant l'hydrolyse, le produit primaire de l'hydrolyse d'un β -glucoside sera du α -glucose, qui à son tour subira la mutarotation, et le changement de rotation pendant la mutarotation sera à gauche.

Quelques expériences préliminaires nous ont donné une vérification des résultats de HEISIG. Toutefois, l'alcoololyse n'est pas complète; même dans des solutions diluées du glucoside un équilibre s'établira, prouvant qu'il y a dans une solution alcoolique d'un glucoside, contenant de l'acide chlorhydrique, un échange mutuel d'alcyde entre le glucoside et l'alcool, de sorte que dans l'équilibre on trouvera, à côté de l'alcool servant comme dissolvant, l'alcool correspondant à l'aglucone du glucoside dissous et les deux glucosides. L'échange des alcyles est, dans les deux directions, accompagné d'une transposition Walden.

Dans une solution aqueuse, où l'alcyde de l'alcool a été remplacé par un atome d'hydrogène, un tel mécanisme peut amener l'hydrolyse complète, ou presque, du glucoside, même si l'équilibre entre glucoside et glucose primairement libéré ne correspond qu'à une hydrolyse partielle du glucoside, car le glucose subira à son tour la mutarotation, forçant ainsi l'hydrolyse progressive du glucoside. Cela reste valable, indépendamment d'une transposition Walden

accompagnant ou non l'hydrolyse, et l'hydrolyse acide peut donc être regardée comme un échange mutuel entre le glucoside et l'eau du groupement —OR et du groupement —OH, suivi de la mutarotation du glucose α ou β formé par l'échange.

Des expériences préliminaires semblent indiquer que l'hydrolyse, par opposition à l'alcoololyse, ne soit pas accompagnée d'une transposition Walden, car un échantillon d'une solution acide d'un glucoside- β , partiellement hydrolysée, semble mutaroter vers la droite, indiquant que c'est du glucose- β qui se trouve dans la solution.

Un mécanisme semblable a été proposé pour l'hydrolyse fermentative des glucosides par LETTRÉ (24), qui est d'avis que le groupement prostéthique du ferment consiste en un molécule d'un hydrate de carbone, renfermant un groupement aldéhydique hydratisé, —CH(OH)₂, rendant ainsi possible l'échange mutuel des groupements —OR et —OH entre glucoside et ferment. Le ferment alcylé peut à son tour réagir avec de l'eau sous régénération du ferment. La théorie de LETTRÉ ne peut pas être correcte dans la formulation proposée par lui (comp. VEIBEL et LILLELUND (17)), mais les expériences de HELFERICH et collaborateurs (24, 25) portent à croire que le ferment contient vraiment un ou plusieurs hydrates de carbone; des expériences ultérieures montreront le bien-fondé du principe de la théorie de LETTRÉ. Il faut se souvenir aussi des expériences de NEUBERG (27), qui ont montré que le diéthyl-acétal peut être hydrolysé sous l'influence de l'émulsine.

Résumé.

On a examiné l'hydrolyse acide de quelques glucosides- β des alcools primaires, secondaires et tertiaires. Les constantes de la vitesse d'hydrolyse ainsi que les énergies d'activation Q et les constantes B dans l'équation d'ARRHENIUS: $\ln k = \frac{-Q}{RT} + B$ sont indiquées dans les tableaux VII—IX, XXI—XXII et XXIV. L'hydrolyse fermentative des glucosides, dont l'hydrolyse acide a été examinée dans ce mémoire, a été examinée antérieurement par VEIBEL et collaborateurs sauf pour trois glucosides, dont l'hydrolyse fermentative est décrite dans ce mémoire. Les constantes trouvées pour tous les glucosides sont indiquées dans les tableaux XX, XXIII et XXIV. Il faut se rendre compte que les constantes B pour l'hydrolyse fermentative n'ont pas de signification physique concrète, la concentration moléculaire du ferment n'étant pas connue.

On a trouvé que les constantes de la vitesse d'hydrolyse, l'énergie d'activation et les constantes B , en ce qui concerne l'hydrolyse acide, sont assez rapprochées les unes des autres pour les glucosides des alcools primaires et secondaires, tandis que pour les glucosides des alcools tertiaires les constantes de la vitesse d'hydrolyse sont supérieures aux autres par 2 ou 3 ordres de grandeur, en même temps que Q est un peu inférieur et B un peu supérieur pour ces derniers glucosides que pour les autres.

En ce qui concerne l'hydrolyse fermentative, les constantes de la vitesse d'hydrolyse, les énergies d'activation et les constantes B sont également assez rapprochées les unes des autres pour les glucosides des alcools primaires et secondaires (mais différentes des constantes de l'hydrolyse acide). Pour les glucosides des alcools tertiaires, par contre, les

constantes de la vitesse d'hydrolyse sont inférieures aux constantes des autres glucosides de un ou de deux ordres de grandeur. L'énergie d'activation et les constantes B sont sensiblement plus grandes pour les derniers glucosides que pour les autres.

Le pseudoamylglucoside se range en ce qui concerne l'hydrolyse acide parmi les glucosides des alcools primaires, et en ce qui concerne l'hydrolyse fermentative plutôt parmi les glucosides des alcools tertiaires.

On a discuté un mécanisme possible de l'hydrolyse des glucosides, profitant des expériences publiées par HEISIG au sujet de l'alcoololyse des glucosides.

(Laboratoire de Chimie de l'Université, Copenhague).

Bibliographie.

1. R. KUHN et H. SOBOTKA, *Ztschr. physiol. Chem.* 109, 65 (1924).
2. K. JOSEPHSON, *Ztschr. physiol. Chem.* 147, 1 (1925).
3. E. A. MOELWYN-HUGHES, *Trans. Farad. Soc.* 25, 81 (1929).
4. B. HELFERICH, H. E. SCHEIBER, R. STREEK et F. VORSATZ, *Liebigs Ann.* 518, 211 (1935).
5. H. E. SCHEIBER, Dissertation, Leipzig (1935).
6. S. VEIBEL et F. ERIKSEN, *D. Kgl. Danske Vidensk. Selskab, Math.-fys. Medd.* XIII, 17 (1936).
7. S. VEIBEL et F. ERIKSEN, *D. Kgl. Danske Vidensk. Selskab, Math.-fys. Medd.* XIV, 15 (1937).
8. S. VEIBEL et H. LILLELUND, *D. Kgl. Danske Vidensk. Selskab, Math.-fys. Medd.* XVII, 6 (1940).
9. S. VEIBEL et F. ERIKSEN, *Bull. Soc. chim.* [5] 3, 277 (1936).
10. S. VEIBEL et H. LILLELUND, *Bull. Soc. chim.* [5] 5, 494 (1938).
11. G. ZEMPLEN et E. PASCU, *Ber. dtsh. chem. Ges.* 62, 1613 (1929).
12. S. VEIBEL, F. LUNDQVIST, F. ANDERSEN et E. FREDERIKSEN, *Bull. Soc. chim.* [5] 6, 990 (1939).
13. S. VEIBEL et F. ERIKSEN, *Biochem. Journ.* 30, 163 (1937).
- 13a. J. A. CHRISTIANSEN, Communication privée.
14. S. VEIBEL, *Enzymol.* 2, 367 (1938).
15. M. DUBOUX, *Bull. Soc. Vaud. Sc. nat.* 59, 49 (1937).
16. P. M. LEININGER et M. KILPATRICK, *Journ. Am. chem. Soc.* 60, 2891 (1938).
17. S. VEIBEL et H. LILLELUND, *Ztschr. physiol. Chem.* 253, 55 (1938).
18. B. HELFERICH, S. GRÜNLER et A. GNÜCHTEL, *Ztschr. physiol. Chem.* 248, 85 (1937).
19. A. SKRABAL et K. H. MINTL, *Ztschr. phys. Chem.* 111, 98, (1924).
20. J. N. BRØNSTED et C. GROVE, *Journ. Am. chem. Soc.* 52, 1394 (1930).
21. L. C. RIESCH et M. KILPATRICK, *Journ. phys. chem.* 39, 561 (1935).

- 38 S. VEIBEL et E. FREDERIKSEN: L'hydrolyse acide et fermentative etc.
22. P. M. LEININGER et M. KILPATRICK, Journ. Am. chem. Soc. 61, 2510 (1939).
23. H. HEISIG, Dissertation, Breslau 1937.
24. H. LETTRÉ, Ztschr. angew. Chem. 50, 581, (1937).
25. B. HELFERICH, W. RICHTER et S. GRÜNLER, Ber. Sächs. Akad. d. Wiss. math.-fys. Kl. 89, 385 (1938).
26. B. HELFERICH et W. W. PIGMANN, Ztschr. physiol. Chem. 259, 253 (1939).
27. C. NEUBERG et R. ZIFFER, Enzymol. 5, 389 (1939).
-

Indleveret til Selskabet den 30. Juli 1940.
Færdig fra Trykkeriet den 6. Maj 1941.

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

CONSTRUCTION OF A CYCLOTRON

(INSTITUTE FOR THEORETICAL PHYSICS, COPENHAGEN)

BY

J. C. JACOBSEN



KØBENHAVN

I KOMMISSION HOS EJNAR MUNKSGAARD

1941

CONTENTS

	Page
Introduction	3
Cyclotron Principles	5
The Magnet	11
The Oscillator	17
The Tank	19
The Dees	24
Working Experiences	27

Introduction.

The ingenious construction by LAWRENCE of an electromagnetic resonance-accelerator for high-speed ions, the so-called cyclotron, has, as well known, been of great importance for the recent development in nuclear physics. By means of the cyclotron one can in fact obtain intense beams of ions with energies considerably greater than has hitherto been possible with high tension installations, where the ions are accelerated in one step. Even though the cyclotron cannot yet produce ion-beams with energies as well defined as those obtainable from high tension apparatus, it thus constitutes a most important auxiliary for the extension of the field of nuclear researches. When a few years ago by grants to Professor BØHR from the Carlsberg Foundation and the Rockefeller Foundation it became possible to acquire the means for enlarging the scope of such researches in this Institute, it was therefore planned besides a one million volt high tension installation¹ to build a cyclotron of a size enabling us to work with deuteron beams having energies up to ten million volts.

The realisation of this plan was above all made possible by a gift from the Thrige Foundation of a suitable magnet

¹ T. BJERGE, K. J. BROSTRØM, J. KOCH and T. LAURITSEN, A high tension apparatus for nuclear research. D. Kgl. Danske Vidensk. Selskab, Math.-fys. Medd. XVIII. 1. (1940).

with an iron core of 35 tons, the successful construction of which we owe to director V. MEYER of the Ths. B. Thrige factories, Odense, Denmark. For our work it has further been of greatest importance that Professor LAWRENCE himself has taken a most kind interest in our plans and has not only provided us with detailed drawings of his own constructions, but also was helpful in securing the invaluable assistance of one of his own collaborators, Dr. L. J. LASLETT, who stayed here one and a half year as stipendiary of the Rockefeller Foundation and the Rask-Ørsted Foundation.

The first beam from the cyclotron was obtained already in 1938 but, to begin with, the intensity was small and energies of only a few million volts could be reached. After various improvements of the apparatus the intensity of the beam was gradually increased and the cyclotron has in the course of the last years been in extensive use as a source of radioactive isotopes for the physico-biological investigations carried out in this Institute by Professor G. HEVESY and his collaborators. In the meantime greater energies were also obtained and in the summer of 1940 deuteron beams of ten million volts were successfully applied to the fission of uranium and thorium.¹ Although various further improvements of the cyclotron are under preparation, the apparatus has now been so long in function that it may be of interest to give an account of the experience gained with the present construction which in certain details differs from the construction of other cyclotrons described hitherto.

¹ J. C. JACOBSEN and N. O. LASSEN, Phys. Rev. **58**, 867 (1940).

Cyclotron Principles.

The principles underlying the construction of the cyclotron have been described by LAWRENCE and his co-workers¹, but for the sake of completeness the general formulae may be repeated here.

A charged particle with mass m and charge e is supposed to move with velocity v in a magnetic field H perpendicular to the lines of force. The radius r of the circular path is determined by

$$\frac{mv^2}{r} = e \cdot v \cdot H.$$

The time required for a complete revolution is

$$\tau = \frac{2\pi r}{v} = 2\pi \frac{m}{eH} \quad (1)$$

and is, thus, independent of the velocity of the particle as long as m remains constant.

To accelerate the ions, two radio frequency electrodes are used, each semicircular in shape and hollow; they are mounted in a vacuum chamber between the poles of a magnet so that the lines of force are perpendicular to the plane of the electrodes.

Between the electrodes (in the following referred to as the dees) a high frequency voltage difference is maintained.

Ions moving in the interior of the dees are accelerated in passing the diametral gap between the dees; if the period of oscillation of the high frequency source is equal to the time of revolution given by (1) the ions will remain in the

¹ E. O. LAWRENCE and EDLEFSON, *Science* **72**, 376 (1930); E. O. LAWRENCE and M. S. LIVINGSTON, *Phys. Rev.* **40**, 19 (1932); E. O. LAWRENCE and M. S. LIVINGSTON, *Phys. Rev.* **45**, 608 (1934); E. O. LAWRENCE and COOKSEY, *Phys. Rev.* **50**, 1131 (1936).

same phase relative to the high frequency potential difference between the dees and, as a consequence of successive accelerations, will move in a series of ever widening semi-circular paths. The number of accelerations may be very large, so that the energy finally obtained may be many times greater than the potential difference between the dees.

If λ is the wavelength of the high frequency source, the condition for resonances is

$$\frac{\lambda}{c} = 2\pi \frac{m}{e \cdot H}. \quad (2)$$

For deuterons and a field of about 15 000 Ørstedes, this gives for λ a value of about 25 metres.

The energy finally attained by the ions depends upon the diameter of the dees. If r is the radius of the last semi-circle, the velocity of the ions is

$$v = \frac{e}{m} \cdot r \cdot H$$

and the energy

$$V = \frac{1}{2} \frac{e}{m} \cdot r^2 \cdot H^2 \cdot 10^{-8}. \quad (3)$$

Here, V is expressed in electron volts, $\frac{e}{m}$ and H in e. m. units. To obtain a high energy, both r and H must be made large. In practice, however, magnetic fields larger than about 20 000 Ørstedes are very expensive to maintain; a further restriction on the magnetic field is imposed by the circumstance that it has been found necessary to have a possibility for applying local corrections to the field, and these corrections can only be made if the steel of the pole pieces is not too near saturation. The result is that fields higher than about

17 000 Ørsteds can hardly be used. With this value of H , the value of r necessary to produce deuterons with 10 MeV energy is $r = 37$ cm. As a consequence of the restrictions imposed on the magnetic field, the only way of increasing the energy seems to be an increase of the radius of the dees.

Inherent in the resonance principle is a feature which gives a high intensity of the ion beam, viz. that the resonance condition is independent of the momentary value of the potential difference between the dees. Disregarding for the moment the focussing conditions which will be considered later, this means essentially that the whole of the half-cycle of the high frequency is utilized.

Targets to be irradiated may be inserted in the diametral gap between the dees. For non-volatile substances, this method yields good service in so far as the whole of the circulating ion beam is utilized but for substances which are likely to spoil the vacuum in the cyclotron it is necessary to remove the high energy ions from the space inside the dees into a separate bombarding chamber. For this purpose, an auxiliary electrode which is maintained at a constant negative potential is placed at a suitable distance from the cylindrical wall of one of the dees. The electrostatic force between the deflecting electrode and the wall of the dee will increase the radius of curvature of the beam and in this way make it possible to bring the beam outside the dee. Experience has shown that an increase of 15% in the radius of curvature is suitable; the electric field F , which is necessary, is calculated from

$$e \cdot F = 0.15 \cdot \frac{mv^2}{r} = 0.15 \cdot \frac{2eV}{r}, \text{ or}$$

$$F = 0.30 \cdot \frac{V}{r}.$$

For $V = 10$ MeV and $r = 38$ cm, this gives $F = 80\,000$ volts per cm.

The distance between successive paths of an ion depends on the voltage difference between the dees. If ΔV is the gain in energy per revolution, the distance Δr between successive paths is given by

$$\Delta V = 10^{-8} \cdot \frac{e}{m} \cdot H^2 \cdot r \Delta r;$$

for $\Delta V = 70\,000$ volts, corresponding to a voltage difference of 35 000 volts between the dees, $H = 17\,000$ Ørstedes and $r = 38$ cm, giving for deuterons $\Delta r = 0.12$ cm. The thickness of the cylindrical wall of the dee facing the deflecting electrode must be small compared to Δr in order to reduce as far as possible the loss of ions due to collisions with the wall.

The ions to be accelerated are produced near the centre of the cyclotron. In its simplest form, the ion source involves a beam of electrons projected vertically between the dees through hydrogen or deuterium gas at a low pressure. Although more efficient methods for the production of ions have been developed, this simple form of ion source has been used throughout in the present work. The number of ions produced depends on the gas pressure and the number and velocity of the electrons. The extension of the space between the dees inside which the ions are actually produced is determined partly by the final length of the filament used as a source of electrons and partly by the cycloidal motion carried out by the electrons under the combined action of the magnetic field and the electric field between the dees. In practice, the extension of the ion source is of the order

of a few centimetres; this feature is undesirable, because ions starting at different places will reach the deflecting field with different energies. Since, furthermore, the paths of different ions reaching the deflecting field are not parallel in space, a considerable loss of ions takes place during the passage through the deflecting field due to collisions with the deflecting electrode and the dee wall; in practice, the ion beam leaving the deflecting field amounts to roughly 10% of the beam inside the dees.

The complicated conditions arising when regard is taken to a possible asymmetry in the magnetic field are referred to later in connection with the measurements of the energy spread of the beam leaving the deflecting field.

In the simple resonance condition the magnetic field is supposed to be homogeneous. In practice, however, a magnetic field is used which is somewhat stronger at the centre than at the periphery, the difference in the field between the centre and the exit slit being of the order of 1%. This inhomogeneity involves the existence of a radial component of the field which, in turn, gives rise to a restoring force on the ions, the force being directed from both sides towards the plane of symmetry. If z is the distance from the plane of symmetry, and H_r is the radial component of the field, then

$$\frac{\delta H_z}{\delta r} = \frac{\delta H_r}{\delta z} \quad \text{or} \quad H_r = z \cdot \frac{\delta H_z}{\delta r} \propto z \cdot \frac{\delta H}{\delta r}.$$

The restoring force is, thus, proportional to the distance from the plane of symmetry and the ions, consequently, carry out oscillations on both sides of the plane of symmetry.

The electric field between the dees also gives rise to a focussing effect; the equipotential surfaces between the dees are located in such a way that the ions are subject to a force directed against the median plane before the passage of the gap and directed away from the median plane after the passage of the gap. Since the ions are accelerated, the latter force acts during a shorter time than the former and the result is a focussing effect. This sort of focussing is mainly effective near the centre of the cyclotron where the energy of the ions is small. For high energy ions, a different sort of electric focussing is effective, depending on the change in the voltage difference between the dees during the passage of the ions through the gap. A simple consideration will show that a focussing effect takes place if the voltage difference between the dees is decreasing while the ions pass through the gap. As a whole, the magnetic focussing is more important than the electric focussing, because the electric focussing is a differential effect whereas the magnetic forces act over the whole path of the ions. The focussing effects are extremely important in securing a high intensity of the ion beam. As a consequence of the focussing effects, a considerable fraction of the ions generated at the centre of the cyclotron is able to reach the edge of the dee without colliding with the top or bottom plates.

The desirability of employing a slightly inhomogeneous field involves that the actual field shows small deviations from the value H_{res} which fulfills the resonance condition (2). If it is assumed, as seems generally to be the case, that the actual field is stronger than H_{res} at the centre and weaker at the edge, the ions will gain in phase relative to the high frequency field between the dees during the first part of their path and lag behind in phase during the later

part¹. A phase lag is further introduced by the relativistic change in mass which becomes effective for ions with high energy.

The existence of a phase lag sets an upper limit to the energy which can be obtained because, if the number of accelerations is increased beyond a given limit, the ions will be retarded and consequently lost.

The electric and magnetic focussing effects have been treated in detail by ROSE² and by WILSON³. For the phase lag just mentioned, ROSE has shown that it is inversely proportional to the voltage difference between the dees. In principle, there is thus no definite upper limit to the energy obtainable, the limitation being caused by the amount of power, only, which is available for the oscillator.

The Magnet.

The magnet and the oscillator are placed in a room below ground level which is situated between the high tension hall and the main building of the Institute. The magnet is placed on a socket so that the gap is 1.5 m above the floor; this placement was originally done in order to have the gap nearly midway between the floor and the ceiling, so that in experiments with neutrons the scattering from floor and ceiling would be a minimum. In how far this precaution is justified remains to be seen;

¹ Here, it is assumed that the frequency of the oscillator is constant. For use in radio transmitters, methods have been developed for modulating the frequency of the oscillator and it might well be feasible to modulate the frequency of the oscillator in such a way that ions travelling through an inhomogeneous magnetic field would remain in exact resonance over their whole path.

² M. E. ROSE, *Phys. Rev.* **53**, 392 (1938).

³ R. R. WILSON, *Phys. Rev.* **53**, 408 (1938).

up till now, all experiments with neutrons have been carried out with the target placed inside the tank wall. The oscillator is placed at a distance of 3 m from the magnet;

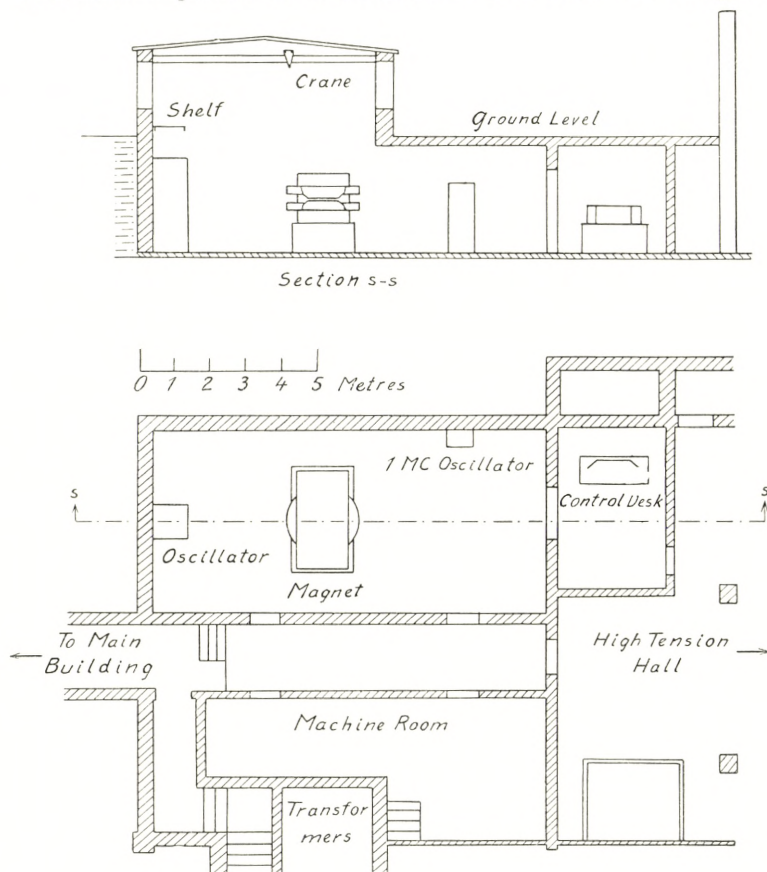


Fig. 1.

the anode voltage supply (a three-phase transformer with a six-phase rectifier) is placed on a shelf above the oscillator.

An adjacent room, which has been partitioned off from the high tension hall, contains the control desk with the main switchboard for the cyclotron.

The diameter of the pole faces is 90 cm, giving an area of 63.7 dm^2 . From the face, the pole pieces are cylindrical for a height of 1 cm and, from this point, are tapered at an angle of 45° with the axis up to a diameter of 116 cm. From this height, the pole pieces are more nearly cylindrical, the diameter at the base of the pole piece is 119 cm, giving an area of 110 dm^2 . The yoke

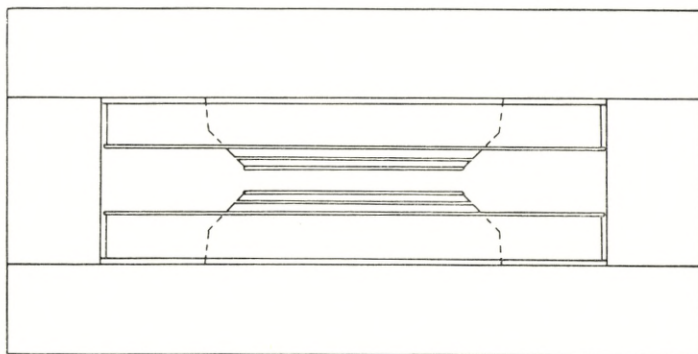


Fig. 2.

has a rectangular cross section with area 55 dm^2 . The strongly tapered shape of the pole piece makes it possible to obtain high fields in the gap without undue saturation at the base of the pole piece. The highest field obtainable is about 20 000 Ørsted. To make possible the application of local corrections to the field ("shimming"), slits of 6 mm height are provided in the pole pieces at a distance of 4 cm from the faces.

The coils are flat "pancake" coils with 300 turns each, placed in oiltight tanks in transformer oil. The oil is circulated and cooled by water in a heat exchanger. The rate of exchange of heat between the oil and the water is such that, with 70 kW input into the coils, corresponding to a field of 20 000 Ørsted, the temperature becomes stationary

at about 60°C . The total weight of the magnet is 37 tons, including 3 tons of copper for the coils.

The power supply is a 250 volts d. c. generator driven by a 380 a. c. motor. The field of the d. c. generator is excited from a small d. c. generator. Manually controlled

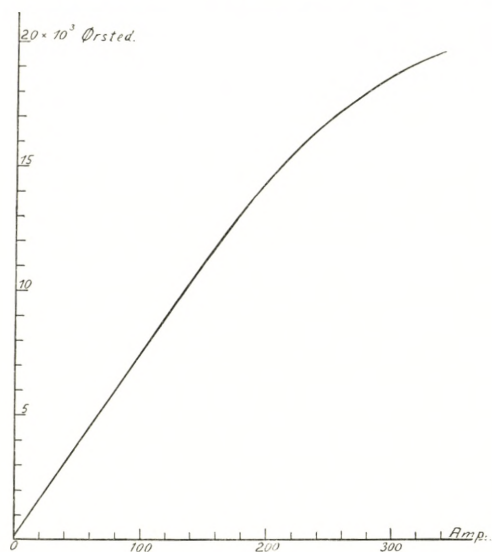


Fig. 3.

resistors serve for regulating the d. c. voltage. To reduce the effect of fluctuations in the supply from the a. c. mains, a device is used which is very similar to that published by PERRY¹. This device will only compensate for fluctuations in the voltage across the coils of the magnet, so that the change in the current due to the gradual warming up of the coils must be compensated for by manual regulation.

The symmetry of the field about the median plane was tested by a small steel needle with a mirror fixed to it in the manner described by WILSON². Small deviations

¹ T. PERRY, *Phys. Rev.* **53**, 943 (1938).

² R. R. WILSON, *l. c.*

were found, but it was estimated that symmetry could be established by shimming, a consequence which was verified when a beam had been obtained.

The radial gradient of the field was determined by a test coil (connected to a ballistic galvanometer) which could

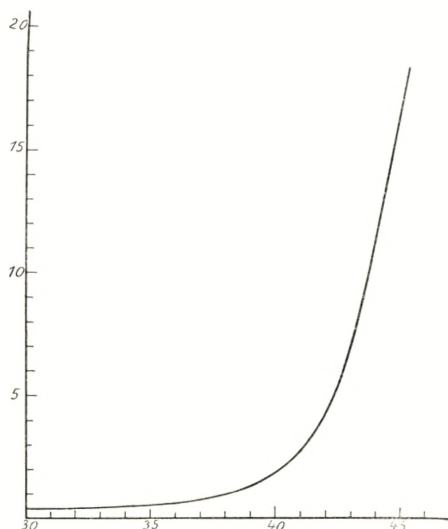


Fig. 4. Abscissa: distance from centre in cm. Ordinate: percentage deviation from field at centre.

be moved in steps along a diameter. The deflections were calibrated by a known mutual inductance.

In Fig. 3 is shown the variation of the field with the magnet current; Fig. 4 and Fig. 5 show at a somewhat different scale the field as a function of the distance from the centre. As ordinate, in Figs. 3 and 4, is plotted $100 \cdot \frac{H_c - H_r}{H_c}$, when H_c is the field at the centre and H_r the field at a distance r from the centre. When the tank is evacuated, the steel plates forming the top and the bottom of the tank are bent inwards; the displacement amounts

to 0.18 mm for each plate. When the magnet is excited, the plates are partly straightened out. Curve I in Fig. 4 shows the field with the tank evacuated and with air in the shimming gap, curve II was obtained with the shimming gap evacuated.

The results of Fig. 3 and Fig. 4 were obtained with a

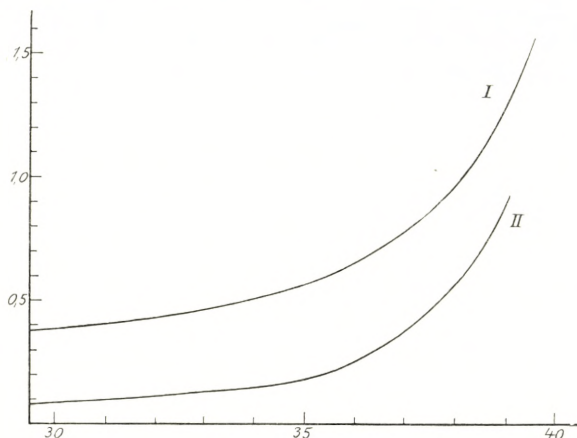


Fig. 5. Abscissa: distance from centre in cm. Ordinate: percentage deviation from field at centre.

field of 17 000 Ørsted. The quantity $\frac{H_c - H_r}{H_c}$ varies little with H_c , so that almost the same curves represent the gradient at a field of 9 000 Ørsted.

To reduce the high frequency losses, the pole faces are plated with electrolytically deposited copper in a thickness of about 0.1 mm. After more than a year's use, most of the copper plating is apparently removed by sputtering from a region above and below the gap between the dees. Under these circumstances, the high frequency losses in the top and bottom plates amount to less than 10 per cent of the total energy expended in the tank circuit, as can be

shown by observing the rise in temperature of the plates when the cyclotron is running. So far, it has not been possible to decide whether the copper plating is actually of any use in reducing the losses in the tank circuit as a whole.

The Oscillator.

The oscillator tubes are of the Sloan type and are continuously exhausted. The filament is a single hairpin loop of 1 mm tungsten wire with a total length of 47 cm, the anode cylinder is a $2\frac{1}{2}$ " copper tube. The grid is wound from 1 mm copper wire with a spacing of 5 mm. Up to the present, only a single hairpin loop has been used as a filament in each oscillator, but it is planned to increase the number of filaments. With an input of 25 kW into the oscillators, the anode losses amount to about 45 per cent of the input.

The voltage supply for the anode is a three-phase transformer connected to a six-phase rectifier with mercury rectifying valves. The a. c. mains are fed into the primary of the transformer through an intermediate transformer with variable ratio, the variation is carried out by remote control from the control desk. The rectifier supplies up to 5 amps at 10 000 volts, but with the single hairpin used at present as a filament in the oscillator, only about 2.5 amps can be utilized with good efficiency. In case of sparks passing over in the high tension circuit, the current through the rectifier is limited partly by wire wound resistors in the secondary of the high tension transformer and partly by an overload switch in the primary.

The coupling between the tank circuit and the oscillators is shown on Fig. 6. The tank circuit consists of two

copper tubes with 8 cm diameter and a spacing of 30 cm, made from sheet copper. A movable shorting bar allows tuning at different wavelengths. The cooling water leaving the dees is used to cool the copper tubes. The anode and grid circuits are both made from copper tubes with 2 cm diameter. Mycalex condensers with a capacity of about 1500 cm serve to keep the anode voltage off the tank circuit.

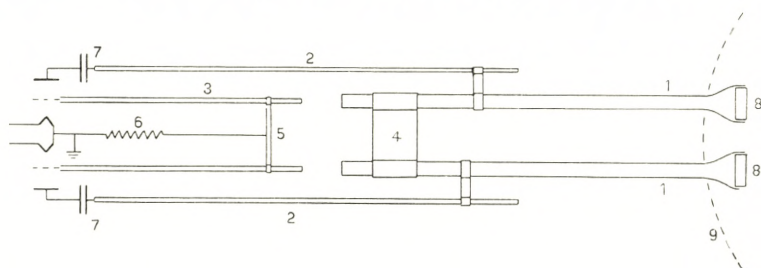


Fig. 6. 1 tank circuit, 2 anode circuit, 3 grid circuit, 4 and 5 shorting bars, 6 grid leak, 7 coupling condensers, 8 flange on dee insulator, 9 edge of magnet coil.

The voltage supply for the anodes is connected through choke coils. The grid leak of about 1500 ohms is made from a glass tube with flowing water and carbon electrodes.

The anode line is connected to the tank circuit at such a distance from the shorting bar that the proper excitation of the anodes is obtained. The voltage node of the tank circuit is not at the middle of the shorting bar, so that the connections from the tank circuit must be displaced somewhat along the tank line in order to obtain the same excitation of the anodes.

Disregarding for a moment the grid circuit, the arrangement contains two circuits, an anode circuit and a tank circuit, which are coupled tightly together. As a consequence of the tight coupling, the system has two modes of oscillations with different frequencies and different ampli-

tudes. By moving the shorting bar of the grid circuit it is possible to select an oscillation which gives a large amplitude in the tank circuit. Under given circumstances, for instance if a spark is drawn from the tank circuit, the system may "jump" into the other mode of oscillation, which gives a much smaller amplitude in the tank circuit and also a different wavelength. The "jumping" between two different modes of oscillation disturbs very little in the actual work with the cyclotron; in this connection it may be mentioned that, as a consequence of the existence of two resonance frequencies, the oscillations of the tank circuit cannot be represented exactly by a simple sin-wave.

While the plant was still under development, it was tried to have the tank circuit and the anode circuit as resonant circuits tuned to the same frequency and coupled through an untuned transmission line. It was soon found, however, that to obtain sufficient energy transfer to the tank circuit the coupling had to be made so tight that the resonance frequency of the anode and tank circuits would split into two frequencies with the same possibility of "jumping" as for the system described above. It was then preferred to use the present system which, apart from the grid circuit, does not contain any circuits which have to be tuned together.

The Tank.

The two steel plates of 4 cm thickness, which form the top and bottom of the tank, are bolted to the main pole piece by 16 bolts uniformly distributed near the edge. As a consequence, the seal between the cylindrical wall of the tank and the top and bottom plate must be made with

the tank placed between the magnet poles. The removable part of the tank is a brass ring with the same outside diameter as the cylindrical part of the pole pieces; the height of the ring is 88 mm, the wall thickness is 15 mm. The ring is of cast brass; several leaks due to pores in the casting were found and finally tightened by painting the whole ring with glyptal. The ring and the adjacent

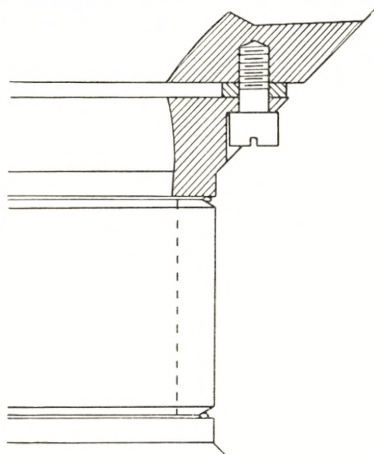


Fig. 7.

parts of the pole pieces are shown in section in Fig. 7.

The ring is kept in position by four brass pieces (not shown in the figure) which are bolted both to the ring and the pole pieces. The upper and lower end of the ring are slightly conical against the outside; into the slit thus formed between the brass ring and the pole piece a rubber string (ordinary rubber insulated flex) is pressed and the joint is painted with glyptal.

The joint is easily made and has never given rise to trouble. An obvious advantage of this method of assembling the tank is, that the magnetic field remains unchanged when the tank is withdrawn from the gap, so that the properties of the field as regards homogeneity and symmetry can be examined before the tank is brought into position. Further, if a suitable arrangement of shims has once been found, the shims can be left in the shimming gap, when a removal of the tank for repair becomes necessary. Only a prolonged experience and a comparison with the results of other workers can show whether this method of assem-

bling the tank is ultimately to be preferred; here it can only be stated that no difficulty has been encountered in making the seal between the ring and the top and bottom plates in the way described here.

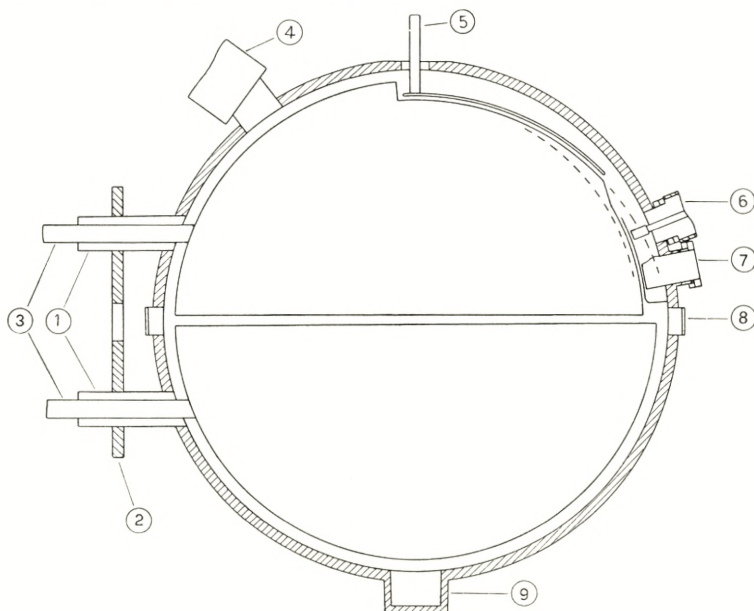


Fig. 8. 1 brass tubes, 2 brass plate, 3 dee supports, 4 pump line, 5 deflecting electrode, 6 vacuum lock, 7 irradiation pot, 8 window, 9 filament support.

The tank with the dees and the deflecting system is shown in Fig. 8. The dee insulators, which should be placed to the left of the tank in Fig. 8, are shown in detail in Fig. 10.

The studs for the dee supports, the pumping line, windows etc. are either brazed or tin-soldered in the ring. As a support for the dee insulators, two brass tubes (1) of 6 cm diameter are brazed to the tank wall, the brass tubes support a brass plate (2), into which bolts are screwed

which serve for the adjustment of the dees. The brass tubes extend 2 cm beyond the brass plate, the ends of the tubes being used as a seat for the flange carrying the insulator.

The pumping system is a 5-inch oil pump backed by a 2-inch oil pump and a mechanical pump. The pump line is a copper tube with 13 cm diameter and a length of 90 cm. Between the pump line and the pump a large valve is placed which allows of separating the pumps from the tank. The pumping speed of the pumping system itself is about 100 litres of air per second, but the pump line does not allow a speed of more than 50 litres per second. At the joint between the pump line and the brass ring, the diameter of the pump line cannot exceed 7.5 cm, and this measure sets a limit to the pumping speed. The pressure is read on a McLeod gauge or an ionization gauge. For locating leaks, a hot wire gauge made from a small carbon filament lamp is generally used.

The arrangement used for deflecting the ions out from the space inside the dees follows the device worked out at Berkeley. The deflecting field increases the radius of the path of the ion beam by 15 per cent; for deuterium ions with 10 MeV. energy, a field of 80 000 volts per cm is thus necessary. Much trouble has been caused by sparks between the rod carrying the deflecting electrode and the pole piece, the sparks passing over in the direction of the magnetic field. It has never been possible to apply a potential of more than 50 000 volts to the deflecting electrode.

The beam enters the deflecting field at 38 cm distance from the centre. The outside of the tank wall being at a distance of 45 cm from the centre, the shift in the path of the beam caused by the deflecting field is insufficient to bring the beam outside the tank wall. To irradiate

samples in vacuum, a vacuum lock is used which is shown in detail in Fig. 9. For irradiation in air, an irradiation pot is used which has an aluminium window with 3 cm stopping power.

The ion source is of the usual design, where electrons from a tungsten filament ionize the gas at the centre of the cyclotron. The filament is a straight tungsten wire of 0.6 mm diameter and 4 cm length. The current for heating

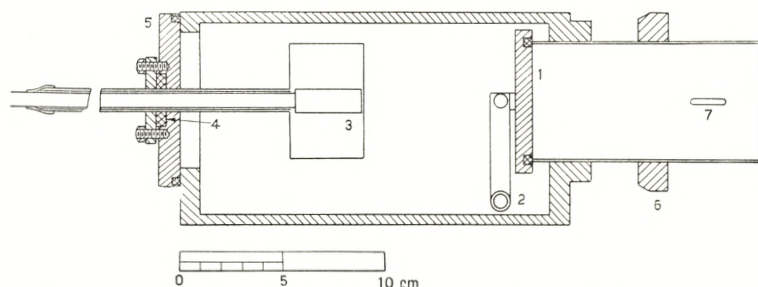


Fig. 9. 1 brass lid, 2 ground joint, 3 target holder, 4 rubber gasket, 5 external lid, 6 tank wall, 7 approximate cross section of beam.

the filament is supplied by an oscillator employing two Philips 7A/700 valves in push-pull operating at a wavelength of 300 metres. The filament forms part of a resonant circuit which by a transmission line is coupled to the oscillator. The emission drawn from the filament is usually about 200 milliamps. To accelerate the electrons, the filament is kept at a negative potential relative to ground, which can be varied continually up to 1000 volts.

The vacuum lock (Fig. 9) which is used for irradiating samples in vacuum consists of a rectangular brass box with dimensions $7.5 \cdot 9.5 \cdot 17$ cm; the box is connected to the tank wall by a brass tube with 6 cm diameter. The separation between the box and the tank is a brass lid (1) which is moved by a ground joint (2); the lid is tightened

by a rubber gasket seated against the end of the brass tube, which projects somewhat inside the box. The sample (3) is carried by a water-cooled brass tube which slides through a rubber gasket (4), so that the sample can be moved in and out and rotated. As a lubricant for the gasket (4), castor oil is used. The gasket (4) is carried by an external lid (5) tightened by a rubber gasket. Suppose now that the lid (1) is closed and the box filled with air. To introduce a sample, the lid (5) is placed in position, and the inner end of the sample holder is pressed against the lid (1). The box is now evacuated by an auxiliary pump. When this is done, the brass tube with the target holder is pulled outwards, so that the lid (1) can be opened and the target is finally slid into position. Targets with an area of $5 \cdot 5$ cm can be introduced through this lock.

In certain experiments, it is of importance to irradiate different samples with beams carrying exactly the same current. Experiments have been carried out with uranium, where four different samples were placed on the target holder and irradiated with deuteron beams carrying the same current, but with different energies, by rotating the target holder continuously during the irradiation. The energy of the deuterons was varied by absorbing foils placed in front of the separate targets.

The Dees.

The dees have an outside radius of 41 cm. The top and bottom plates are spun copper plates of 1 mm thickness; along the outside the copper plates are fastened by screws to a brass ring of height 25 mm and thickness 3 mm; along the diameter a cooling pipe of 6 mm copper

tubing forms the mechanical support. The internal height at the centre is 45 mm, the height decreases nearly regularly from the centre to the outside. In view of the determinations of the movements of the ions in the cyclotron carried out by WILSON¹, it would probably have been of advantage to keep the maximum height out to a distance of some 20 cm from the centre. The support for the dee is a 30 mm rod of solid copper. Four grooves are cut in the copper rod into which the cooling pipes for the dee are soldered.

The insulator supporting the dee is shown in Fig. 10. The insulator is of fused quartz ("Rotosil"); its length is 30 cm, the diameter of the end flanges is 14 cm. Brass flanges are cemented at the ends of the insulator by ordinary red sealing wax. The brass flange (1) carrying the inner end of the insulator rests with a conical seat on a brass tube (2) projecting from the tank wall and is kept in position by four bolts (3) which are fixed to a brass plate (4). Tightness between the brass flange (1) and the

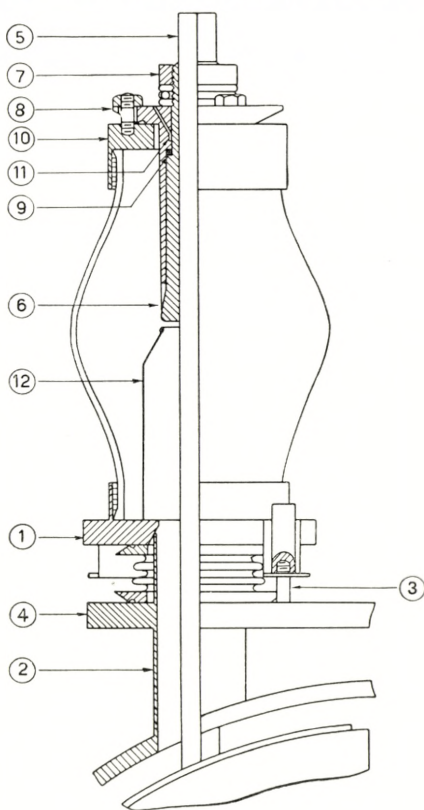


Fig. 10.

¹ R. R. WILSON, Phys. Rev. **53**, 408 (1938).

brass plate (4) is afforded by a bellow. By means of the bolts, the dee can be moved up and down and sideways within certain limits.

To the dee rod (5) is soldered a brass sleeve (6) which by the nut (7) is screwed together with the flange (8);

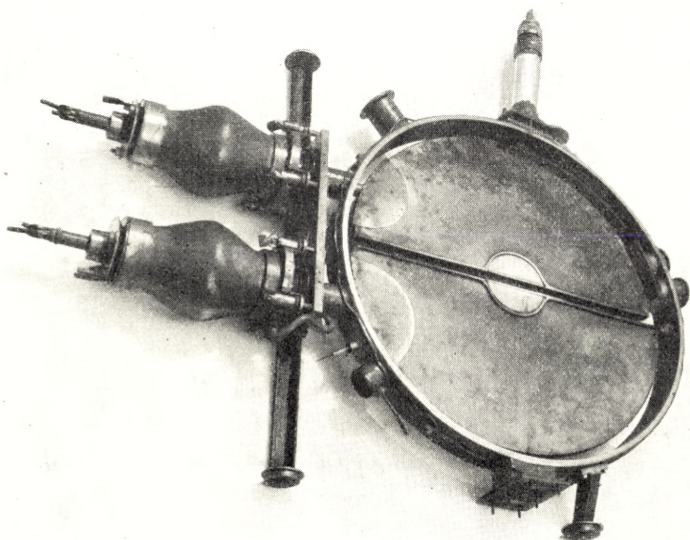


Fig. 11. The ring with the dees and insulators removed from the gap. The ring rests on four wheels, of which three are seen in the figure. The filament has been removed from its holder (lower part of the figure).

lightness between (6) and (8) is afforded by an internal rubber gaskett (9). The flange (8) is fixed to the flange (10) by three bolts in such a way that some movement of the dee rod relative to the flange is possible. A groove (11) is cut in the inner side of (8) so that a gas stream can be floated round the gaskett to control its tightness. A copper tube (12) which has its rounded end about 1 cm from the end of (6) is fixed to the flange (1) in order to shield the insulator from electric fields.

As a whole, the present construction of the dee support is hardly satisfactory. The use of four bolts for the adjustment of the dee is unsatisfactory since, to obtain a movement of the dee, all four bolts have to be moved. The sealing wax used for the connection between the insulator and the flanges is quite satisfactory as regards mechanical stability, but the cooling of the flange must of course be effective. The main reason for cementing the brass flanges to the insulator by sealing wax is that, in this way, the outside dimensions of the flange can be kept small; this is highly desirable in view of the small distance between the magnet coils. The present clearance of 5 cm between the flange and the magnet coil is sufficient to allow about 25 kW. input into the oscillator at a wavelength of 23 m without sparking.

Working Experiences.

Most of the observations described in this section were made with deuterons of energy from 8 to 9.5 MeV.; for a short time, protons of 5 MeV. energy were used. Without presenting any essential new features, the observations may be of sufficient interest to justify a short description. With 8 MeV deuterons a beam of $2 \mu\text{A}$ could be obtained through the deflecting field, for higher energies the beam decreased rapidly; with 9.5 MeV deuterons $0.1 \mu\text{A}$ could be obtained. The homogeneity of the beam was examined by means of the range curves of LIVINGSTON and BETHE¹.

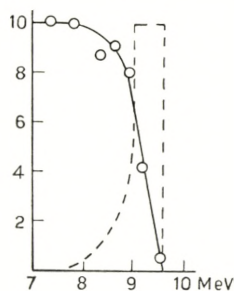


Fig. 12.

¹ M. S. LIVINGSTON and H. A. BETHE, Rev. Mod. Phys. 9, 226 (1937).

The result for a deuteron beam with 9.5 MeV maximum energy is shown in Fig. 12. The full curve shows the directly measured absorption of the deuterons, the dotted curve is a differential curve, showing the relative number of deuterons with a given energy.

The height of the beam after leaving the deflecting field is about 3 mm; at an early stage of the work, when the cyclotron

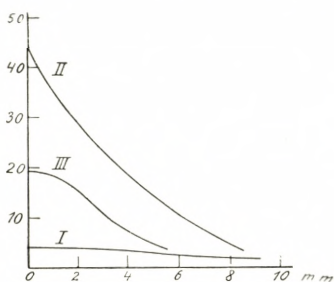


Fig. 13. Abscissa: distance from median plane. Ordinate: activity in arbitrary units. Curve I, distance from centre 20 cm. Curve II, distance from centre 30 cm. Curve III, distance from centre 35 cm.

was running with deuterons of about 5 MeV. energy, a few measurements were made of the height of the beam inside the dees. A copper wire with 3 mm diameter was placed vertically between the dees; after a short run, the distribution of activity along the wire was determined by measuring the activity through a lead slit which was placed in front of a counter. The results are shown in Fig. 13 for three

different distances between the wire and the centre of the cyclotron. The results demonstrate that for the region investigated nearly the whole activity on the wire is found within 5 mm distance from the median plane. The curves in Fig. 13 give very nearly a true representation of the amplitudes of the ions about the median plane, because the ions will spend most of the time at the maximum distance from the median plane.

From Fig. 5 it appears that a considerable change in the magnetic field can be produced by evacuating the shimming gap. Curves I and II of Fig. 5 represent roughly the limits between which the gradient of the field has been

varied. Some increase of the gradient could be produced by pyramid shims in the shimming gap. Several such arrangements were tried, but the result was that, although the beam was extremely sensitive to an asymmetry of the field about the median plane, only very slight changes of the beam could be produced by changing the gradient of the field. At the same time, it was found that a considerable increase of the beam could be produced by a sort of azimuthal shimming, where rectangular slabs of iron were placed mainly in the outer region of the upper shimming gap. By a suitable arrangement of three or four such slabs, the beam current could usually be increased roughly a hundred times. The action of this sort of shimming is probably to direct the beam through the deflection channel or, in other words, to shift the paths of the high energy ions in such a way as to bring the centre of the circulating ion beam to coincide with the geometrical centre.

To control the dee voltage, a circular disk with 1 cm diameter is inserted in the tank wall, insulated by a pyrex tube; the disk is connected to a rectifier in the manner described by WHITMER¹. This voltmeter gives only relative values of the dee voltage, and its reading further depends somewhat on the position of the dee; it has been very useful, however, in connection with the adjustment of the oscillator.

To obtain an absolute measure of the dee voltage, the dee circuit is disconnected from ground and the position of the filament is changed, so that it is below one of the dees. With the filament in this position, the dee circuit collects a negative charge until its potential is equal to the peak voltage, and this potential is measured with a voltmeter

¹ R. M. WHITMER, *Rev. Sc. Instr.* 10, 165, 1939.

with a high resistance. Corresponding to an input of 22 kW into the oscillators, a voltage difference of 35 kV peak value between the dees was found. This result may seem rather low in comparison with the value given by other workers¹ employing similar dee circuits. An explanation of the difference may be found in the present construction of the dee support which, probably, offers a relatively high resistance to the high frequency current. The smallness of the height of the beam mentioned above may be brought into connection with the relatively low value of the dee voltage. The reason for the insensitivity of the beam against changes in the gradient of the magnetic field then remains somewhat obscure. Unless the gradient of the field, which is necessary to give a reasonable increase in the beam current, must fulfill very special conditions, which can only be found out by pure chance, the beam current is only slightly dependent on the gradient of the field, as stated above.

The screening against neutrons has been accomplished in a simple way by placing paraffin blocks in the gap between the magnet coils. The paraffin blocks together with the oil tanks containing the coils constitute an enclosure of about 45 cm thickness, which is sufficient for the purpose, at least with the neutron intensities hitherto obtained.

On several occasions, it was noticed that the beam, which was obtained momentarily by letting the magnetic field sweep through the resonance value, was much stronger with decreasing than with increasing field. A similar observation has been made by HENDERSON and WHITE² who

¹ ALLEN, SAMPSON and FRANKLIN, *Journ. Franklin Inst.* **228**, 543 (1939).
HENDERSON, KING, RISSER, YEARIAN and HOWE, *Journ. Franklin Inst.* **228**, 563, (1939).

² M. C. HENDERSON and M. G. WHITE, *Rev. Sc. Instr.* **9**, 19 (1938).

explain the effect as being due to eddy currents in the pole pieces, induced by the changing flux. Besides this effect which undoubtedly is present, another effect has been noticed which may be of even more influence on the gradient of the field, viz. relatively large movements of the top and

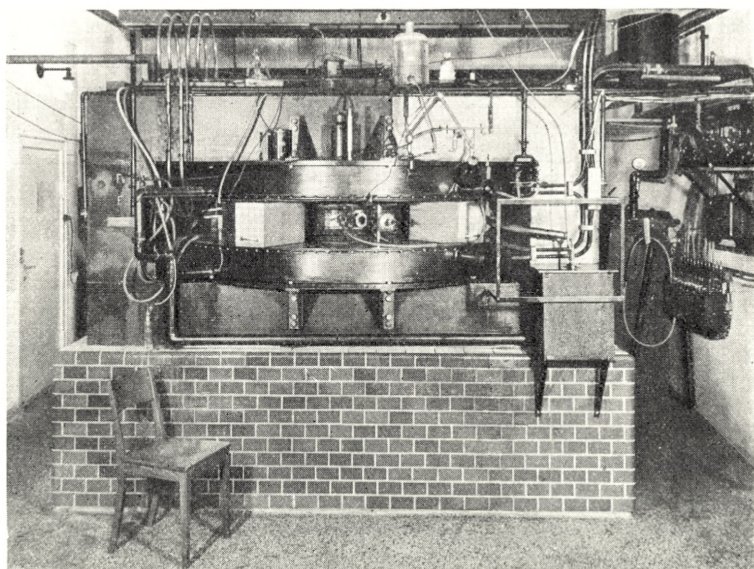


Fig. 14. The magnet seen from the side opposite the oscillator. Near the centre of the figure the vacuum lock and the irradiation pot can be distinguished. The paraffin blocks used for screening against neutrons have been partly removed.

bottom plates of the tank when the field is changing. This effect was noticed during some measurements of the bending of the top plate of the tank due to atmospheric pressure. For these measurements, a simple mechanical device was used, which transferred the movements of the top plate to a mirror, so that the width of the shimming gap could be read on a scale. When the field was changing, movements of the top plate were observed which, in some cases,

were larger than the shift of 0.18 mm caused by the air pressure of 6 tons. No detailed study of this effect was carried out but, from the observed magnitude of the displacements, it was estimated that the effect on the gradient of the field was comparable with the effect due to the atmospheric pressure on the top and bottom plates.

In concluding this account I wish to thank the director of the Institute, Professor N. BOHR, for his continual interest and helpful advice. Special thanks are also due Director V. MEYER from the Thrige factories, who planned the magnet and several parts of the electrical equipment. During the construction of the cyclotron much help was given by Dr. O. R. FRISCH, Mr. S. HØFFER-JENSEN, Mr. N. O. LASSEN and, especially, by Dr. L. J. LASLETT, whose skill and experience were of invaluable help for our work during his stay in Copenhagen. Acknowledgements are further due Mr. H. W. OLSEN and his collaborators in the institute workshop for the excellent elaboration of the many often complicated pieces of the apparatus.

Det Kgl. Danske Videnskabernes Selskab.

Mathematisk-fysiske Meddelelser. **XIX**, 3.

THE MOTION OF COMET 1929 I
(SCHWASSMANN-WACHMANN 2) IN THE
YEARS 1936-1942

BY

HANS Q. RASMUSEN



KØBENHAVN

I KOMMISSION HOS EJNAR MUNKSGAARD

1941

Printed in Denmark.
Bianco Lunos Bogtrykkeri A/S.

In a previous publication¹ I have given the results of a calculation of the elements of periodic comet Schwassmann-Wachmann 2 from observations in the apparitions 1929-1934 as well as the perturbations during this period. The elements are:

Epoch and osculation 1929 February 18.0 U. T.

$$\begin{array}{r}
 M = 354^{\circ}.91062 \\
 \omega = 357^{\circ}.71203 \\
 \Omega = 126^{\circ}.32728 \\
 i = 3^{\circ}.72839 \\
 e = 0.3945470 \\
 a = 3.452821 \\
 \mu = 0^{\circ}.1536183
 \end{array}
 \left. \vphantom{\begin{array}{l} M \\ \omega \\ \Omega \\ i \\ e \\ a \\ \mu \end{array}} \right\} 1950.0$$

Comparison of the computed orbit with selected observations showed so small residuals that it will hardly be worth while to revise the elements until observations from the opposition 1941-1942 are available. The elements are accurate enough for the purpose of calculating a search ephemeris for the opposition just mentioned. I have, therefore, continued the integration of the perturbed co-ordinates from 1936 to 1942. As before, account was taken of the attractions by Jupiter and Saturn.

¹) D. Kgl. Danske Vidensk. Selskab, Math.-fys. Medd. XIII, 16 and Publikationer og mindre Meddelelser fra Københavns Observatorium Nr. 106.

The calculated co-ordinates, referred to the equinox 1950, are given in the following table.

0 ^h U. T.		x		y		z	
1936	Jan.	13 ¹	- 2.303 459	+ 0.202 413	+ 0.208 371		
	Feb.	2	2.380 985	- 0.027 971	0.123 206		
	—	22	2.437 737	0.258 138	+ 0.036 955		
	March	13	2.474 915	0.486 261	- 0.049 603		
	April	2	2.493 850	0.710 823	0.135 805		
	—	22	2.495 924	0.930 603	0.221 098		
	May	12	2.482 521	1.144 635	0.305 029		
	June	1	2.454 992	1.352 175	0.387 232		
	—	21	2.414 624	1.552 662	0.467 419		
	July	11	2.362 630	1.745 693	0.545 362		
	—	31	2.300 140	1.930 988	0.620 890		
	Aug.	20	2.228 200	2.108 367	0.693 875		
	Sept.	9	2.147 769	2.277 734	0.764 225		
	—	29	2.059 732	2.439 052	0.831 875		
	Oct.	19	1.964 891	2.592 335	0.896 787		
	Nov.	8	1.863 985	2.737 638	0.958 939		
	—	28	1.757 684	2.875 041	1.018 326		
	Dec.	18	1.646 602	3.004 649	1.074 953		
1937	Jan.	7	1.531 297	3.126 581	1.128 835		
	—	27	1.412 282	3.240 969	1.179 995		
	March	8	1.164 951	3.447 672	1.274 264		
	April	17	0.907 907	3.625 915	1.358 022		
	May	27	0.643 936	3.776 874	1.431 570		
	July	6	0.375 418	3.901 702	1.495 223		
	Aug.	15	- 0.104 401	4.001 506	1.549 294		
	Sept.	24	+ 0.167 333	4.077 342	1.594 095		
	Nov.	3	0.438 216	4.130 203	1.629 920		
	Dec.	13	0.706 851	4.161 023	1.657 055		
1938	Jan.	22	0.971 980	4.170 678	1.675 769		
	March	3	1.232 457	4.159 990	1.686 315		
	April	12	+ 1.487 221	- 4.129 731	- 1.688 934		

¹ In the previous publication the x -co-ordinate for this date was erroneously printed as x : - 2.303859, instead of - 2.303459.

0 ^h U. T.		x	y	z
1938	May 22	+ 1.735 284	- 4.080 629	- 1.683 852
	July 1	1.975 706	4.013 372	1.671 282
	Aug. 10	2.207 590	3.928 613	1.651 428
	Sept. 19	2.430 065	3.826 976	1.624 483
	Oct. 29	2.642 274	3.709 060	1.590 630
	Dec. 8	2.843 370	3.575 444	1.550 048
1939	Jan. 17	3.032 499	3.426 696	1.502 907
	Feb. 26	3.208 797	3.263 376	1.449 377
	April 7	3.371 378	3.086 040	1.389 623
	May 17	3.519 327	2.895 252	1.323 814
	June 26	3.651 690	2.691 593	1.252 120
	Aug. 5	3.767 466	2.475 662	1.174 720
	Sept. 14	3.865 598	2.248 095	1.091 800
	Oct. 24	3.944 959	2.009 576	1.003 561
	Dec. 3	4.004 346	1.760 850	0.910 226
1940	Jan. 12	4.042 468	1.502 743	0.812 043
	Feb. 21	4.057 929	1.236 187	0.709 295
	April 1	4.049 226	0.962 247	0.602 313
	— 21	4.035 308	0.822 885	0.547 351
	May 11	4.014 726	0.682 157	0.491 481
	— 31	3.987 256	0.540 249	0.434 765
	June 20	3.952 663	0.397 368	0.377 267
	July 10	3.910 704	0.253 739	0.319 063
	— 30	3.861 128	- 0.109 607	0.260 233
	Aug. 19	3.803 672	+ 0.034 757	0.200 867
	Sept. 8	3.738 067	0.179 055	0.141 067
	— 28	3.664 036	0.322 957	0.080 945
	Oct. 18	3.581 293	0.466 099	- 0.020 627
	Nov. 7	3.489 550	0.608 077	+ 0.039 748
	— 27	3.388 512	0.748 444	0.100 021
	Dec. 17	3.277 888	0.886 700	0.160 014
1941	Jan. 6	3.157 386	1.022 291	0.219 530
	— 26	3.026 727	1.154 600	0.278 343
	Feb. 15	2.885 647	1.282 938	0.336 202
	March 7	+ 2.733 905	+ 1.406 542	+ 0.392 821

0 ^h U. T.		<i>x</i>	<i>y</i>	<i>z</i>
1941	March 27	+ 2.571 300	+1.524 560	+ 0.447 882
	April 16	2.397 681	1.636 048	0.501 027
	May 6	2.212 966	1.739 964	0.551 855
	— 26	2.017 165	1.835 159	0.599 918
	June 15	1.810 414	1.920 382	0.644 722
	July 5	1.592 995	1.994 275	0.685 723
	— 25	1.365 384	2.055 387	0.722 330
	Aug. 14	1.128 292	2.102 194	0.753 909
	Sept. 3	0.882 706	2.133 127	0.779 796
	— 23	0.629 933	2.146 625	0.799 310
	Oct. 13	0.371 636	2.141 193	0.811 773
	Nov. 2	+ 0.109 852	2.115 490	0.816 551
	— 22	- 0.153 010	2.068 424	0.813 083
	Dec. 12	0.414 207	1.999 258	0.800 925
1942	Jan. 1	0.670 740	1.907 711	0.779 801
	— 21	0.919 472	1.794 036	0.749 631
	Feb. 10	- 1.157 278	+ 1.659 070	+ 0.710 564

From the co-ordinates and velocities the following osculating elements have been derived:

Epoch and osculation 1942 January 1.0 U. T.

$$\begin{array}{l}
 M = 353^{\circ}.38019 \\
 \omega = 358^{\circ}.00835 \\
 \Omega = 126^{\circ}.04333 \\
 i = 3^{\circ}.72521 \\
 e = 0.3854489 \\
 a = 3.488424 \\
 \mu = 0^{\circ}.1512726
 \end{array}
 \left. \vphantom{\begin{array}{l} M \\ \omega \\ \Omega \\ i \\ e \\ a \\ \mu \end{array}} \right\} 1950.0$$

According to this calculation the comet will pass through perihelion 1942 February 13.76080 U.T. The comet will be well placed for observation from November 1941.

The following ephemeris has been calculated with the elements just given.

		0 ^h U. T.					
1941		$\alpha_{1941.0}$		$\delta_{1941.0}$	r	Δ	
		h	m				
Oct.	13	7	8.3	+ 19° 29'	2.320	2.040	
	17		13.6	19 20	2.309	1.984	
	21		18.7	19 10	2.299	1.929	
	25		23.4	19 1	2.289	1.874	
	29		27.9	18 52	2.280	1.820	
Nov.	2		32.0	18 43	2.270	1.767	
	6		35.9	18 35	2.261	1.715	
	10		39.3	18 28	2.252	1.664	
	14		42.4	18 21	2.244	1.615	
	18		45.1	18 16	2.236	1.568	
	22		47.3	18 12	2.228	1.522	
	26		49.1	18 10	2.220	1.477	
	30		50.4	18 9	2.213	1.436	
	Dec.	4		51.3	18 10	2.206	1.396
		8		51.6	18 13	2.199	1.358
12			51.4	18 18	2.193	1.324	
16			50.8	18 25	2.187	1.293	
20			49.7	18 34	2.182	1.265	
24			48.1	18 45	2.177	1.240	
28			46.1	18 58	2.172	1.219	
32		7	43.8	+ 19 12	2.167	1.202	

Corrections for T 1 day later:

Oct. 13.	-1.5^m	+1'
Nov. 22.	-2.1	+3
Dec. 32.	-2.7	+4

I am indebted to the Carlsbergfond for a grant which supported me in carrying out the present investigation.

HANS Q. RASMUSEN.

Det Kgl. Danske Videnskabernes Selskab.

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

UNE INÉGALITÉ
DE KOLMOGOROFF ET LES FONC-
TIONS PRESQUE-PÉRIODIQUES

PAR

THØGER BANG



KØBENHAVN

I KOMMISSION HOS EJNAR MUNKSGAARD

1941

Printed in Denmark.
Bianco Lunos Bogtrykkeri A/S.

INTRODUCTION

Dans ce travail nous allons démontrer quelques inégalités des bornes supérieures des intégrales successives d'ordres entières ou fractionnaires et les appliquer aux fonctions presque-périodiques⁽¹⁾.

Soit $f_0(x)$ une fonction intégrable au sens de M. LEBESGUE, définie sur tout l'axe réel; une suite des intégrales successives de $f_0(x)$ d'ordre jusqu'à n sera désignée $f_1(x)$, $f_2(x)$, \dots $f_n(x)$. Posons

$$M_k = \overline{\text{borne}}_{-\infty < x < \infty} |f_k(x)| \quad (\text{fini ou infini}).$$

Entre les quantités M_0, M_1, \dots, M_n il y a des relations; M. HADAMARD a démontré⁽²⁾ l'inégalité

$$M_1 \leq \sqrt{2 M_0 M_2}$$

et on voit facilement que cette inégalité ne peut pas être améliorée. Et récemment on a démontré: Pour $0 < k < n$ il existe une limitation de M_k dépendant seulement de k, n, M_0 , et M_n . Dans les cas $n = 3$ et $n = 4$ de telles limitations, et justement les plus favorables, ont été trouvées par M. G. ŠILOV; pour toute valeur de n une limitation a été trouvée par M. A. GORNY, mais pas la plus favorable, laquelle M. A. KOLMOGOROFF a indiquée le premier⁽³⁾; dans la suite cette limitation sera donc nommée »l'inégalité de KOLMOGOROFF«. Autant que je sais, la démonstration de cette

(1) Ici comme partout dans ce travail: Au sens de M. BOHR.

(2) Comptes rendus des séances de la Société mathématique de France 1914, p. 68—72.

(3) Voir Comptes rendus de l'Académie des Sciences, 206 (1938), p. 1245—1247 (Gorny), et 207 (1938), p. 764—765 (KOLMOGOROFF).

inégalité n'est pas encore publiée; un des buts du présent travail sera donc d'en communiquer une⁽¹⁾.

L'inégalité de KOLMOGOROFF est de la forme

$$M_k \leq M_0^{1-\frac{k}{n}} \cdot M_n^{\frac{k}{n}} \cdot K_n^k$$

où K_n^k est une constante, dépendant seulement de k et n , et cette constante a une telle valeur qu'on a l'égalité, si pour $f_0(x)$ on choisit la fonction

$$\varphi_0(x) = \text{sign} \sin x = \frac{4}{\pi} \sum_{\nu=0}^{\infty} \frac{\sin(2\nu+1)x}{2\nu+1}.$$

Soient M_0 et M_n finis (le théorème est sans intérêt en dehors de ce cas), l'inégalité nous apprend que M_k est également fini, et nous donne une borne supérieure de cette quantité; comme cette borne est atteinte pour la fonction $\varphi_0(x)$, cette fonction sera nommée fonction extrémale.

Nous donnerons de plus un résumé du contenu des pages suivantes:

Dans la section I l'inégalité de KOLMOGOROFF sera démontrée pour le cas où $f_n(x)$ est une fonction presque-périodique, et l'inégalité sera utilisée à donner une démonstration simple des inégalités de MM. BOHR et FAVARD concernant les bornes supérieures des intégrales successives d'une somme finie trigonométrique sans terme constant, et nous allons indiquer comment il est possible de déduire d'une façon tout à fait analogue la limitation bien connue de M. BERNSTEIN pour la dérivée d'une somme finie trigonométrique⁽²⁾.

(1) La démonstration se trouve dans Uchenye Zapiski Moskov. Gos. Univ. Matematika 30 (1939), p. 3—16 (ce renseignement m'a été donné par M. B. JESSEN); malheureusement ce périodique n'existe pas en Danemark.

(2) Ce n'est que pour ces applications que l'inégalité de KOLMOGOROFF est énoncée pour les fonctions presque-périodiques; pour la démonstration

Dans la section II nous introduisons la notion d'intégrale d'ordre fractionnaire d'une fonction presque-périodique par une définition qui s'appuie sur la représentation par séries de FOURIER: S'il existe deux fonctions presque-périodiques $f_0(x)$ et $f_\alpha(x)$, dont les séries de FOURIER sont respectivement

$$\sum a_\lambda e^{i\lambda x} \quad \text{et} \quad \sum \frac{a_\lambda}{(i\lambda)^\alpha} e^{i\lambda x}$$

nous dirons que $f_\alpha(x)$ est une intégrale presque-périodique d'ordre α de $f_0(x)$ (α réel ou complexe). Les sommes trigonométriques finies possèdent des intégrales presque-périodiques d'ordre quelconque; au contraire pour une fonction presque-périodique arbitraire se présente la question de la nature de l'ensemble des nombres α pour lesquels la fonction a une intégrale presque-périodique d'ordre α .

Entre les bornes supérieures absolues des intégrales presque-périodiques d'ordre fractionnaire il y a aussi quelques inégalités; d'abord nous nous bornerons aux sommes finies trigonométriques, pour lesquelles nous démontrerons que pour $0 < \bar{\alpha} < \bar{\beta}^{(1)}$ on a

$$M_\alpha \leq M_0^{1-\frac{\bar{\alpha}}{\bar{\beta}}} \cdot M_\beta^{\frac{\bar{\alpha}}{\bar{\beta}}} \cdot K_\beta^\alpha$$

où K_β^α est une constante ne dépendant que de α et β . Grâce à cette inégalité nous pourrions résoudre le problème cité, et comme solution on obtiendrait que l'ensemble des valeurs de α en question sera une bande (verticale) dans le plan complexe, et on verra aussi que l'inégalité est valable pour une fonction presque-périodique quelconque, dans le cas général (section III), la validité pour les fonctions périodiques est suffisante, et nous n'utilisons pas dans la preuve des résultats profonds de la théorie des fonctions presque-périodiques.

(1) Dans ce travail $\bar{\alpha}$ signifie la partie réelle de α ; pour éviter des malentendus nous n'emploierons pas le nombre complexe conjugué de α .

pourvu que les intégrales d'ordre α et β existent. La détermination des constantes K_{β}^{α} les plus favorables ne sera pas accomplie, il sera seulement montré que dans le cas général la fonction $\varphi_0(x)$ n'est plus une fonction extrémale, puisqu'il n'en existe pas qui soit indépendante de α et de β . A la fin de cette section nous généraliserons les inégalités de MM. BOHR, FAVARD et BERNSTEIN au cas de l'intégration (ou »dérivation«) d'ordre fractionnaire.

Dans la section III nous démontrerons que l'inégalité de KOLMOGOROFF est valable, non seulement pour les fonctions presque-périodiques, mais aussi pour toute fonction d'une variable réelle; nous ferons la preuve en montrant qu'à une telle fonction on peut toujours construire une fonction périodique de telle sorte que, pour un nombre fini d'ordres d'intégration, la différence entre les bornes supérieures des intégrales du même ordre de ces deux fonctions n'excède pas ε , où ε est un nombre positif quelconque, donné d'avance.

Pour les intégrales d'ordre fractionnaire des fonctions presque-périodiques on a la formule

$$f_{\alpha}(x) = \frac{1}{\Gamma(\alpha)} \int_0^{\infty} f_0(x-t) \cdot t^{\alpha-1} dt \quad \left\{ \begin{array}{l} \text{(pour} \\ 0 < \alpha < 1). \end{array} \right.$$

M. E. R. LOVE a utilisé cette formule pour définir l'intégration d'ordre fractionnaire aussi pour des fonctions non presque-périodiques⁽¹⁾. Il se pourrait qu'entre les bornes supérieures de telles intégrales on ait les mêmes inégalités que pour les fonctions presque-périodiques, et probablement on pourrait vérifier cette proposition par une »approximation« avec des fonctions périodiques analogue à la

⁽¹⁾ Proceedings of the London Mathematical Society 44(1938), p. 363—397.

citée, mais la théorie de l'intégration d'ordre fractionnaire est encore incomplète, et nous n'insisterons donc pas sur ce point.

I. L'inégalité de Kolmogoroff pour les fonctions presque-périodiques.

Dans cette partie démontrons d'abord le théorème:

Théorème I (l'inégalité de KOLMOGOROFF pour les fonctions presque-périodiques): Soit $f_0(x)$ une fonction réelle et bornée, définie sur tout l'axe réel et avec une intégrale d'ordre n presque-périodique, $f_n(x)$. Désignons par $f_k(x)$ l'intégrale d'ordre k de $f_0(x)$ (une dérivée de $f_n(x)$). Toutes les fonctions $f_k(x)$ ($k = 1, 2, \dots, n \div 1$) sont bornées⁽¹⁾, et pour la borne supérieure de $f_k(x)$, désignée M_k , on a

$$M_k \leq M_0^{1-\frac{k}{n}} \cdot M_n^{\frac{k}{n}} \cdot \frac{t_k}{t_n^{\frac{k}{n}}}$$

où

$$t_h = \frac{4}{\pi} \left[1 + \frac{1}{(-3)^{h+1}} + \frac{1}{5^{h+1}} + \frac{1}{(-7)^{h+1}} + \dots \right]$$

et cette inégalité est la plus favorable.

Nous faisons remarquer que l'inégalité peut s'écrire sous la forme plus élégante

$$\left(\frac{M_k}{M_0 \cdot t_k} \right)^{\frac{1}{k}} \leq \left(\frac{M_n}{M_0 \cdot t_n} \right)^{\frac{1}{n}} \quad (0 < k < n).$$

Que l'inégalité ne puisse pas être améliorée on le voit en considérant la fonction

⁽¹⁾ D'où il résulte qu'ils sont uniformément continues, donc aussi presque-périodiques.

$$\varphi_0(x) = \text{sign} \sin x = \frac{4}{\pi} \sum_{\nu=0}^{\infty} \frac{\sin(2\nu+1)x}{2\nu+1}$$

pour laquelle M_k est égale à t_k et M_0 est 1, donc

$$\left(\frac{M_k}{M_0 \cdot t_k}\right)^{\frac{1}{k}} = \left(\frac{M_n}{M_0 \cdot t_n}\right)^{\frac{1}{n}} = 1.$$

Il reste à démontrer que l'inégalité est vérifiée pour une fonction presque-périodique quelconque $f_n(x)$, et pour cela il suffit de démontrer la relation

$$\left(\frac{M_{n-1}}{M_0 \cdot t_{n-1}}\right)^{\frac{1}{n-1}} \leq \left(\frac{M_n}{M_0 \cdot t_n}\right)^{\frac{1}{n}}.$$

Nous raisonnerons par l'absurde, en supposant qu'il existe une fonction $f_0(x)$ et un nombre α , tels que

$$\left(\frac{|f_{n-1}(\alpha)|}{M_0 \cdot t_{n-1}}\right)^{\frac{1}{n-1}} > \left(\frac{M_n}{M_0 \cdot t_n}\right)^{\frac{1}{n}}$$

et en montrant qu'une comparaison de $f_0(x)$ avec $\varphi_0(x)$ aboutit à une contradiction. D'abord, considérons de près $\varphi_0(x)$ et ses intégrales:

$\varphi_0(x)$ a la période 2π ; $\varphi_0(x) = 1$ pour $0 < x < \pi$, et $\varphi_0(x) = -1$ pour $\pi < x < 2\pi$; $\varphi_n(x)$ est également périodique; dans la suite »un intervalle de période« signifiera un intervalle $x_1 \leq x < x_1 + 2\pi$, où $\varphi_n(x_1) = -t_n$; dans chaque intervalle de période, $\varphi_n(x)$ s'accroît de $-t_n$ à t_n , puis décroît à $-t_n$.

Si

$$\left(\frac{|f_{n-1}(\alpha)|}{M_0 \cdot t_{n-1}}\right)^{\frac{1}{n-1}} > \left(\frac{M_n}{M_0 \cdot t_n}\right)^{\frac{1}{n}}$$

il serait possible d'obtenir une fonction $g_0(x) = a \cdot f_0(bx + c)$, de sorte que

$$\overline{\text{borne}} |g_0(x)| = |a| \cdot M_0 < 1 = \overline{\text{borne}} |\varphi_0(x)|$$

et

$$\overline{\text{borne}} |g_n(x)| = \frac{|a|}{|b|^n} \cdot M_n < t_n = \overline{\text{borne}} |\varphi_n(x)|$$

tandis que

$$g_{n-1}\left(\frac{\alpha - c}{b}\right) = \frac{a}{b^{n-1}} \cdot f_{n-1}(\alpha) > t_{n-1} = \overline{\text{borne}} |\varphi_{n-1}(x)|$$

et où on a choisi c de telle façon que $g_n\left(\frac{\alpha - c}{b}\right) = \varphi_n\left(\frac{\alpha - c}{b}\right)$, avec $x_1 < \frac{\alpha - c}{b} < x_2$, $\varphi_n(x_1) = -t_n$, $\varphi_n(x_2) = +t_n$, x_1 et x_2 dans le même intervalle de période.

Comme la différence $d_0(x) = g_0(x) - \varphi_0(x)$ est négative pour $2h\pi < x < (2h+1)\pi$ (h entier) et positive pour $(2h+1)\pi < x < (2h+2)\pi$, cette fonction n'admet que deux changements de signe dans chaque intervalle de période.

Sa n -ième intégrale $d_n(x) = g_n(x) - \varphi_n(x)$ possède une valeur positive pour $x = x_1 + 2h\pi$ et une valeur négative pour $x = x_2 + 2h\pi$; cette fonction a donc au moins deux zéros dans chaque intervalle de période. Pour $x = \frac{\alpha - c}{b}$ elle a un zéro avec une dérivée positive (car $g_{n-1}\left(\frac{\alpha - c}{b}\right) > t_{n-1}$ et $\varphi_{n-1}\left(\frac{\alpha - c}{b}\right) \leq t_{n-1}$), et puisque $d_n(x_1) > 0$ et $d_n(x_2) < 0$, elle possède au moins trois zéros entre x_1 et x_2 , donc au moins quatre dans l'intervalle de période $x_1 \leq x < x_1 + 2\pi$; or $d_n(x)$ est presque-périodique (la différence entre une fonction périodique et une fonction presque-périodique) et il y a donc un nombre infini d'intervalles de période qui contiennent au moins quatre zéros.

Dès lors on peut obtenir un intervalle d'une telle longueur que, dans cet intervalle, le nombre de zéros de $d_n(x)$ sur-

passé le nombre de zéros (ou changements de signe) de $d_0(x)$ de plus de n . Mais chaque dérivation ne peut pas diminuer le nombre de zéros d'un intervalle de plus de un; nous arrivons là à une contradiction, et notre proposition est donc établie.

Dans le théorème nous pouvons supprimer le mot réel. En effet, soient $f_0(x)$, $f_k(x)$ et $f_n(x)$ des fonctions complexes d'une variable réelle, on peut voir que l'inégalité est vérifiée, rien qu'en considérant les fonctions réelles $e^{i\theta} \cdot f_0(x)$, $e^{i\theta} \cdot f_k(x)$ et $e^{i\theta} \cdot f_n(x)$, dont les dernières sont des intégrales d'ordre k et n de la première car on a (θ parcourant les nombres réels de 0 à 2π):

$$\begin{aligned} \overline{\text{borne}}_x |f_k(x)| &= \overline{\text{borne}}_{\theta} \overline{\text{borne}}_x |e^{i\theta} \cdot f_k(x)| \leq \\ \overline{\text{borne}}_{\theta} \left[\overline{\text{borne}}_x |e^{i\theta} \cdot f_0(x)|^{1-\frac{k}{n}} \cdot \overline{\text{borne}}_x |e^{i\theta} \cdot f_n(x)|^{\frac{k}{n}} \cdot \frac{t_k}{t_n^{\frac{k}{n}}} \right] &\leq \\ \left[\overline{\text{borne}}_{\theta} \overline{\text{borne}}_x |e^{i\theta} \cdot f_0(x)| \right]^{1-\frac{k}{n}} \cdot \left[\overline{\text{borne}}_{\theta} \overline{\text{borne}}_x |e^{i\theta} \cdot f_n(x)| \right]^{\frac{k}{n}} \cdot \frac{t_k}{t_n^{\frac{k}{n}}} & \\ = \overline{\text{borne}}_x |f_0(x)|^{1-\frac{k}{n}} \cdot \overline{\text{borne}}_x |f_n(x)|^{\frac{k}{n}} \cdot \frac{t_k}{t_n^{\frac{k}{n}}} & \end{aligned}$$

Comme une application simple de l'inégalité de KOLMOGOROFF nous démontrons:

Théorème II (l'inégalité de BOHR-FAVARD): Soit $p_0(x)$ une somme finie trigonométrique sans terme constant

$$p_0(x) = \sum_{h=1}^k a_h e^{i\lambda_h x}, \quad \overline{\text{borne}} |p_0(x)| = M_0$$

et par Λ désignons $\min |\lambda_h| (> 0)$, on a alors pour l'intégrale d'ordre n

$$p_n(x) = \sum_{h=1}^k \frac{a_h}{(i\lambda_h)^n} e^{i\lambda_h x}$$

la limitation

$$\overline{\text{borne}} |p_n(x)| \leq M_0 \cdot \frac{t_n}{\Lambda^n}.$$

Pour $n = 1$ cette inégalité a été démontrée par H. BOHR⁽¹⁾, pour toute valeur de n par J. FAVARD⁽²⁾.

D'après le théorème I on a

$$\left(\frac{M_n}{M_0 \cdot t_n}\right)^{\frac{1}{n}} \leq \lim_{m \rightarrow \infty} \left(\frac{M_m}{M_0 \cdot t_m}\right)^{\frac{1}{m}} = \lim M_m^{\frac{1}{m}} \cdot \lim \frac{1}{(M_0 \cdot t_m)^{\frac{1}{m}}}$$

les deux dernières limites existent, la première est égale à $\frac{1}{\Lambda}$ (car les intégrations successives font ressortir les termes dans lesquels $|\lambda_h| = \Lambda$; M_m est donc asymptotiquement égal à $\frac{c}{\Lambda^m}$), la seconde est égale à 1 (M_0 est une constante, et t_m tend vers $\frac{4}{\pi}$); donc

$$\left(\frac{M_n}{M_0 \cdot t_n}\right)^{\frac{1}{n}} \leq \frac{1}{\Lambda} \quad \text{ou} \quad M_n \leq M_0 \cdot \frac{t_n}{\Lambda^n}.$$

Notre théorème est ainsi démontré.

Pour la n -ième dérivée de $p_0(x)$

$$p_0^{(n)}(x) = \sum_{h=1}^k a_h (i\lambda_h)^n \cdot e^{i\lambda_h x}$$

on a :

Théorème III (l'inégalité de BERNSTEIN):

$$\overline{\text{borne}} |p_0^{(n)}(x)| \leq M_0 \cdot \max |\lambda_h|^n;$$

(1) Prace Matematyczno-Fizyczne 43 (1935), p. 273—288.

(2) Matematisk Tidsskrift B 1936, p. 81—94.

la démonstration de ce théorème peut être effectuée d'une façon tout à fait analogue, seulement en considérant des dérivées au lieu des intégrales, de manière que $\max |\lambda_h|$ se produit naturellement.

Il faut remarquer que dans des démonstrations antérieures de cette inégalité, dues à M. RIESZ et C. DE LA VALLÉE-POUSSIN, on a employé une énumération de zéros, donc une méthode qui a quelque ressemblance avec celle de ce travail⁽¹⁾.

II. Inégalités entre les bornes supérieures des intégrales d'ordre fractionnaire d'une fonction presque-périodique.

Définition: Soit $f_0(x)$ une fonction presque-périodique, dont la série de FOURIER s'écrit $\sum a_\lambda e^{i\lambda x}$; si

$$\sum \frac{a_\lambda}{(i\lambda)^\alpha} \cdot e^{i\lambda x}$$

est la série de FOURIER d'une fonction presque-périodique, nous dirons que celle-ci est une intégrale presque-périodique d'ordre α de $f_0(x)$, et nous la désignons $f_\alpha(x)$. α peut être réel ou complexe, et par $(i\lambda)^\alpha$ nous entendons

$$e^{\left(\frac{\pi}{2} i \operatorname{sign} \lambda + \log |\lambda|\right) \cdot \alpha}.$$

Si $f_1(x)$ existe, on sait, selon des théorèmes bien connus, que cette fonction est une intégrale ordinaire de $f_0(x)$; comme une conséquence immédiate de cette section il s'ensuit que la proposition analogue est vraie pour toute valeur entière de α .

⁽¹⁾ Voir par exemple C. DE LA VALLÉE-POUSSIN: Leçons sur l'approximation p. 39—42.

M_α signifie comme toujours $\overline{|f_\alpha(x)|}$, et $\bar{\alpha}$ est la partie réelle de α .

Une somme finie trigonométrique sans terme constant a une intégrale presque-périodique d'ordre α pour tout nombre complexe α , et pour ces intégrales nous démontrons le théorème suivant:

Théorème IV: Soit $\bar{\gamma} < \bar{\alpha} < \bar{\beta}$; alors on a

$$M_\alpha \leq M_\gamma^{\frac{\bar{\beta}-\bar{\alpha}}{\bar{\beta}-\bar{\gamma}}} \cdot M_\beta^{\frac{\bar{\alpha}-\bar{\gamma}}{\bar{\beta}-\bar{\gamma}}} \cdot K_{\beta-\gamma}^{\alpha-\gamma}$$

où $K_{\beta-\gamma}^{\alpha-\gamma}$ est une constante finie, ne dépendant que de $\alpha-\gamma$ et $\beta-\gamma$.

Supposons que $K_{\beta-\gamma}^{\alpha-\gamma}$ est la constante la plus favorable, c'est-à-dire

$$K_{\beta-\gamma}^{\alpha-\gamma} = \text{borne}_f \frac{M_\alpha}{M_\gamma^{\frac{\bar{\beta}-\bar{\alpha}}{\bar{\beta}-\bar{\gamma}}} \cdot M_\beta^{\frac{\bar{\alpha}-\bar{\gamma}}{\bar{\beta}-\bar{\gamma}}}}$$

où $f_0(x)$ parcourt toutes les sommes finies trigonométriques; il faut démontrer alors que cette borne supérieure est finie.

Que $K_{\beta-\gamma}^{\alpha-\gamma}$ ne dépende que de $\alpha-\gamma$ et $\beta-\gamma$, on le voit immédiatement, en observant que d'après la définition $f_{\alpha+\delta}(x)$ est une intégrale d'ordre δ de $f_\alpha(x)$, et nous pouvons donc supposer que $\gamma = 0$.

Posons $K_{\beta-\gamma}^0 = K_{\beta-\gamma}^{\beta-\gamma} = 1$.

Le théorème s'énonce aussi

$$\log M_\alpha - \log K_{\beta-\gamma}^{\alpha-\gamma} \leq \frac{\bar{\beta}-\bar{\alpha}}{\bar{\beta}-\bar{\gamma}} \log M_\gamma + \frac{\bar{\alpha}-\bar{\gamma}}{\bar{\beta}-\bar{\gamma}} \log M_\beta$$

et on voit qu'il exprime une sorte de »convexité« des nombres $\log M_\alpha$; dans la langue géométrique: β et γ étant

fixes, il existe une fonction finie $L(\alpha)$, par exemple $\log K_{\bar{\beta}-\gamma}^{\alpha-\gamma}$, telle que tout point dont les coordonnées sont $(\bar{\alpha}, \log M_{\alpha} - L(\alpha))$ est situé au-dessous de ou sur le segment qui joint les points $(\bar{\gamma}, \log M_{\gamma} - L(\gamma))$ et $(\bar{\beta}, \log M_{\beta} - L(\beta))$.

Au moyen de cette qualité on peut voir qu'il suffit de faire la démonstration dans le cas $\bar{\beta} - \bar{\gamma} < 1$, une telle »convexité« locale pouvant être étendue à une valeur de $\bar{\beta} - \bar{\gamma}$ quelconque.

En effet, supposons que le théorème soit vrai dans ce cas, et soit $\bar{\gamma} < \bar{\delta} < \bar{\varepsilon} < \bar{\beta}$ avec $\bar{\varepsilon} - \bar{\gamma} < 1$ et $\bar{\beta} - \bar{\delta} < 1$; alors il existe deux fonctions $L_1(\alpha)$ et $L_2(\alpha)$, telles que le point $(\bar{\alpha}, \log M_{\alpha} - L_1(\alpha))$ où $\bar{\gamma} < \bar{\alpha} < \bar{\varepsilon}$ (donc aussi dans le cas $\alpha = \delta$) n'est pas situé au-dessus du segment de $(\bar{\gamma}, \log M_{\gamma} - L_1(\gamma))$ à $(\bar{\varepsilon}, \log M_{\varepsilon} - L_1(\varepsilon))$ et le point $(\bar{\alpha}, \log M_{\alpha} - L_2(\alpha))$, où $\bar{\delta} < \bar{\alpha} < \bar{\beta}$ (donc aussi dans le cas où $\alpha = \varepsilon$) n'est pas au-dessus du segment de $(\bar{\delta}, \log M_{\delta} - L_2(\delta))$ à $(\bar{\beta}, \log M_{\beta} - L_2(\beta))$.

Par addition d'un terme linéaire à $L_2(\alpha)$ nous pouvons obtenir que $L_2(\delta) = L_1(\delta)$ et $L_2(\varepsilon) = L_1(\varepsilon)$; alors les deux points $(\bar{\delta}, \log M_{\delta} - L_1(\delta))$ et $(\bar{\varepsilon}, \log M_{\varepsilon} - L_2(\varepsilon))$ sont situés au-dessous de ou sur le segment qui joint $(\bar{\gamma}, \log M_{\gamma} - L_1(\gamma))$ à $(\bar{\beta}, \log M_{\beta} - L_2(\beta))$, et si nous posons

$$L(\alpha) = \begin{cases} L_1(\alpha) & \text{pour } \bar{\gamma} < \bar{\alpha} < \bar{\varepsilon} \\ L_2(\alpha) & \text{pour } \bar{\varepsilon} \leq \bar{\alpha} < \bar{\beta} \end{cases}$$

le point $(\bar{\alpha}, \log M_{\alpha} - L(\alpha))$ est aussi situé au-dessous de ou sur ce segment, et notre théorème est démontré dans la bande plus large $\bar{\gamma} < \bar{\alpha} < \bar{\beta}$.

En continuant de cette manière, on peut vérifier le théorème pour une valeur de $\bar{\beta} - \bar{\gamma}$, aussi grande que l'on veut.

Donc, le problème se ramène au cas $\gamma = 0 < \bar{\alpha} < \bar{\beta} < 1$.

Les intégrales presque-périodiques d'un ordre δ , dont la partie réelle est un nombre entre 0 et 1, peuvent être représentées par des intégrales définies; c'est pourquoi nous avons fait ces considérations préliminaires; on a

$$f_{\delta}(x) = \frac{1}{\Gamma(\delta)} \int_0^{\infty} f_0(x-t) \cdot t^{\delta-1} dt \quad (1)$$

et par cette formule on voit facilement qu'on peut inversement exprimer $f_0(x)$ au moyen de $f_{\delta}(x)$:

$$f_0(x) = \frac{1}{\Gamma(-\delta)} \int_0^{\infty} [f_{\delta}(x-t) - f_{\delta}(x)] \cdot t^{-\delta-1} dt.$$

De l'expression de $f_{\delta}(x)$ on déduit

$$|f_{\delta}(x) - f_{\delta}(x-a)|$$

$$\begin{aligned} &= \left| \frac{1}{\Gamma(\delta)} \int_0^a f_0(x-t) \cdot t^{\delta-1} dt + \frac{1}{\Gamma(\delta)} \int_a^{\infty} f_0(x-t) \cdot [t^{\delta-1} - (t-a)^{\delta-1}] dt \right| \\ &\leq \frac{1}{|\Gamma(\delta)|} \left[\int_0^a M_0 \cdot |t^{\delta-1}| dt + \int_a^{\infty} M_0 \cdot |t^{\delta-1} - (t-a)^{\delta-1}| dt \right] \\ &= \frac{M_0}{|\Gamma(\delta)|} \cdot a^{\bar{\delta}} \cdot \left[\int_0^1 u^{\bar{\delta}-1} du + \int_1^{\infty} |u^{\delta-1} - (u-1)^{\delta-1}| du \right] = M_0 \cdot a^{\bar{\delta}} \cdot C(\delta) \end{aligned}$$

où $C(\delta)$ est une constante, ne dépendant ni de a ni de $f_0(x)$, car les dernières intégrales sont toutes deux convergentes.

Toutes ces considérations préliminaires faites, nous pouvons accomplir la démonstration du théorème:

(1) Voir E. R. LOVE l. c., p. 394. D'ailleurs, dans ce cas, où les fonctions considérées sont les sommes finies trigonométriques, cette relation et la suivante sont assez faciles à vérifier directement.

On a, puisque $0 < \bar{\beta} - \bar{\alpha} < 1$

$$f_{\alpha}(x) = \frac{1}{\Gamma(\alpha - \beta)} \int_0^{\infty} [f_{\beta}(x-t) - f_{\beta}(x)] \cdot t^{\alpha - \beta - 1} dt;$$

la différence entre crochets est bornée par $M_0 \cdot t^{\bar{\beta}} \cdot C(\beta)$ et aussi par $2M_{\beta}$, donc, η étant un nombre positif quelconque

$$\begin{aligned} |f_{\alpha}(x)| &\leq \frac{1}{|\Gamma(\alpha - \beta)|} \left[\int_0^{\eta} M_0 \cdot t^{\bar{\beta}} \cdot C(\beta) \cdot |t^{\alpha - \beta - 1}| dt + \int_{\eta}^{\infty} 2M_{\beta} \cdot |t^{\alpha - \beta - 1}| dt \right] \\ &= \frac{1}{|\Gamma(\alpha - \beta)|} \left[M_0 \cdot C(\beta) \int_0^{\eta} t^{\bar{\alpha} - 1} dt + 2M_{\beta} \int_{\eta}^{\infty} t^{\bar{\alpha} - \bar{\beta} - 1} dt \right] \\ &= \frac{1}{|\Gamma(\alpha - \beta)|} \left[M_0 \cdot C(\beta) \cdot \frac{\eta^{\bar{\alpha}}}{\bar{\alpha}} + 2M_{\beta} \cdot \frac{\eta^{\bar{\alpha} - \bar{\beta}}}{\bar{\beta} - \bar{\alpha}} \right], \end{aligned}$$

et en posant $\eta = \left(\frac{M_{\beta}}{M_0}\right)^{\frac{1}{\bar{\beta}}}$

$$\begin{aligned} |f_{\alpha}(x)| &\leq \frac{1}{|\Gamma(\alpha - \beta)|} \left[M_0 \cdot C(\beta) \cdot \frac{1}{\bar{\alpha}} \left(\frac{M_{\beta}}{M_0}\right)^{\frac{\bar{\alpha}}{\bar{\beta}}} + 2M_{\beta} \cdot \frac{1}{\bar{\beta} - \bar{\alpha}} \cdot \left(\frac{M_{\beta}}{M_0}\right)^{\frac{\bar{\alpha}}{\bar{\beta}} - 1} \right] \\ &= M_0^{1 - \frac{\bar{\alpha}}{\bar{\beta}}} \cdot M_{\beta}^{\frac{\bar{\alpha}}{\bar{\beta}}} \cdot \frac{1}{|\Gamma(\alpha - \beta)|} \cdot \left[\frac{C(\beta)}{\bar{\alpha}} + \frac{2}{\bar{\beta} - \bar{\alpha}} \right] = M_0^{\frac{\bar{\beta} - \bar{\alpha}}{\bar{\beta} - 0}} \cdot M_{\beta}^{\frac{\bar{\alpha} - 0}{\bar{\beta} - 0}} \cdot D(\alpha, \beta). \end{aligned}$$

Notre théorème est ainsi démontré, et nous avons $K_{\beta}^{\alpha} \leq D(\alpha, \beta)$.

Nous sommes maintenant en mesure de démontrer un résultat concernant les intégrales presque-périodiques d'ordre fractionnaire d'une fonction presque-périodique quelconque:

Théorème V: S'il existe deux fonctions presque-périodiques $f_{\gamma}(x)$ et $f_{\beta}(x)$, dont les séries de FOURIER sont respectivement

$$\sum \frac{a_\lambda}{(i\lambda)^\gamma} e^{i\lambda x} \quad \text{et} \quad \sum \frac{a_\lambda}{(i\lambda)^\beta} e^{i\lambda x}$$

($\bar{\gamma} < \bar{\beta}$), une série de la forme

$$\sum \frac{a_\lambda}{(i\lambda)^\alpha} e^{i\lambda x}$$

avec $\bar{\gamma} < \bar{\alpha} < \bar{\beta}$ est la série de FOURIER d'une fonction presque-périodique $f_\alpha(x)$, et cette fonction est bornée par

$$M_\alpha \leq M_\gamma \frac{\bar{\beta} - \bar{\alpha}}{\bar{\beta} - \bar{\gamma}} \cdot M_\beta \frac{\bar{\alpha} - \bar{\gamma}}{\bar{\beta} - \bar{\gamma}} \cdot K_{\beta - \gamma}^{\alpha - \gamma}.$$

Nous pouvons déterminer une suite de sommes finies trigonométriques $s_{1,\gamma}(x)$, $s_{2,\gamma}(x)$, \dots qui converge uniformément vers $f_\gamma(x)$, et telle que leurs intégrales presque-périodiques d'ordre $\beta - \gamma$ $s_{1,\beta}(x)$, $s_{2,\beta}(x)$, \dots convergent uniformément vers $f_\beta(x)$ (on peut choisir par exemple les sommes d'approximation FEJÉR-BOCHNER).

Alors la suite $s_{1,\alpha}(x)$, $s_{2,\alpha}(x)$, \dots des intégrales d'ordre $\alpha - \gamma$ de $s_{1,\gamma}(x)$, $s_{2,\gamma}(x)$, \dots est convergente aussi, car supposant

$$|s_{p,\gamma}(x) - s_{q,\gamma}(x)| < \varepsilon \quad \text{et} \quad |s_{p,\beta}(x) - s_{q,\beta}(x)| < \varepsilon$$

il en résulte, en vertu du théorème IV, que

$$|s_{p,\alpha}(x) - s_{q,\alpha}(x)| < \varepsilon \cdot K_{\beta - \gamma}^{\alpha - \gamma};$$

la convergence est donc uniforme, la fonction limite est presque-périodique, et on voit immédiatement que sa série de FOURIER s'écrit

$$\sum \frac{a_\lambda}{(i\lambda)^\alpha} e^{i\lambda x}.$$

Enfin, comme l'inégalité

$$M_\alpha \leq M_\gamma \frac{\bar{\beta} - \bar{\alpha}}{\bar{\beta} - \bar{\gamma}} \cdot M_\beta \frac{\bar{\alpha} - \bar{\gamma}}{\bar{\beta} - \bar{\gamma}} \cdot K_{\beta - \gamma}^{\alpha - \gamma}$$

est vraie pour toutes les sommes finies trigonométriques, elle est vraie encore pour la fonction limite $f_\alpha(x)$.

Notre théorème est ainsi complètement démontré.

Comme une conséquence immédiate de ce théorème nous avons:

Théorème V bis: L'ensemble des nombres α , pour lesquels la série formelle

$$\sum \frac{a_\lambda}{(i\lambda)^\alpha} e^{i\lambda x}$$

est la série de FOURIER d'une fonction presque-périodique, est une bande verticale dans le plan complexe.

Par le mot »bande« nous entendons aussi une droite ou un demi-plan ou le plan entier; et il doit être souligné que nous n'avons pas examiné les circonstances sur les droites qui limitent la bande.

Pour les valeurs entières de α et β la fonction $\varphi_0(x) = \text{sign} \sin x$ est une fonction extrémale; cette propriété ne peut pas être généralisée aux valeurs quelconques de α et β ⁽¹⁾. En effet, nous démontrerons que dans le cas général il n'existe pas du tout une fonction extrémale indépendante de α et β . A cet effet nous déterminons K_1^α pour les valeurs réelles de α ($0 < \alpha < 1$):

(1) On voit (seulement en modifiant un peu la démonstration) que le théorème V est vrai, même si la fonction $f_\gamma(x)$ n'est que bornée et presque-périodique au sens de STEPANOFF, et on peut donc aussi appliquer ce théorème à la fonction discontinue $\varphi_0(x)$.

M_0 et M_1 étant donnés, il faut déterminer

$$\overline{\text{borne}}_f M_\alpha = \overline{\text{borne}}_f \overline{\text{borne}}_x \left| \frac{1}{\Gamma(\alpha)} \int_0^\infty f_0(x-t) \cdot t^{\alpha-1} dt \right|$$

pour les fonctions presque-périodiques $f_\alpha(x)$, pour lesquelles

$$\overline{\text{borne}}_x |f_0(x)| = M_0 \quad \text{et} \quad \overline{\text{borne}}_x |f_1(x)| = M_1.$$

Comme au théorème I, il nous suffit de considérer les fonctions réelles $f_0(x)$.

On a

$$f_\alpha(x) = \frac{1}{\Gamma(\alpha)} \int_0^\infty f_0(x-t) \cdot t^{\alpha-1} dt$$

$$= \frac{1}{\Gamma(\alpha)} \int_0^{\frac{2M_1}{M_0}} M_0 \cdot t^{\alpha-1} dt + \frac{1}{\Gamma(\alpha)} \left\{ \int_0^{\frac{2M_1}{M_0}} [f_0(x-t) - M_0] \cdot t^{\alpha-1} dt + \int_{\frac{2M_1}{M_0}}^\infty f_0(x-t) \cdot t^{\alpha-1} dt \right\}$$

et en calculant le premier terme et en intégrant par parties chacune des dernières intégrales on obtient (avec un calcul facile)

$$f_\alpha(x) = \frac{2^\alpha \cdot M_0^{1-\alpha} \cdot M_1^\alpha}{\Gamma(\alpha + 1)}$$

$$+ \frac{1-\alpha}{\Gamma(\alpha)} \left\{ \int_0^{\frac{2M_1}{M_0}} [f_1(x) - f_1(x-t) - M_0 t] \cdot t^{\alpha-2} dt + \int_{\frac{2M_1}{M_0}}^\infty [f_1(x) - f_1(x-t) - 2M_1] \cdot t^{\alpha-2} dt \right\};$$

les fonctions sous les signes d'intégration sont non-positives, donc

$$f_\alpha(x) \leq \frac{2^\alpha \cdot M_0^{1-\alpha} \cdot M_1^\alpha}{\Gamma(1 + \alpha)}.$$

De même, nous aurions

$$-f_\alpha(x) \leq \frac{2^\alpha \cdot M_0^{1-\alpha} \cdot M_1^\alpha}{\Gamma(1+\alpha)}$$

d'où

$$M_\alpha \leq \frac{2^\alpha \cdot M_0^{1-\alpha} \cdot M_1^\alpha}{\Gamma(1+\alpha)}$$

et comme d'autre part, nous pouvons choisir $f_0(x)$, tel que les deux dernières intégrales deviennent aussi petites que l'on veut pour quelques valeurs de x , nous avons

$$K_1^\alpha = \frac{2^\alpha}{\Gamma(1+\alpha)}. \quad (\alpha \text{ réel})$$

Pour toute fonction fixe, la borne supérieure M_α est une fonction continue de α (α étant réel), d'où

$$\lim_{\alpha \rightarrow 1} \frac{M_\alpha}{M_0^{1-\alpha} \cdot M_1^\alpha} = 1.$$

Mais

$$\lim_{\alpha \rightarrow 1} K_1^\alpha = \lim \frac{2^\alpha}{\Gamma(1+\alpha)} = 2$$

et ceci nous montre que, dans le cas général, il n'existe pas de fonction extrémale, et notre proposition est ainsi établie.

Les nombres K_1^α que nous venons de considérer, et K_n^k du théorème I sont donc d'une nature différente, ce qui semble indiquer qu'une détermination de K_β^α dans le cas général serait assez difficile; c'est pourquoi ce problème ne sera pas approfondi ici.

Enfin, nous allons démontrer deux théorèmes concernant les sommes finies trigonométriques (mais qui d'ailleurs se généralisent immédiatement aux fonctions presque-périodiques quelconques, exactement comme les théorèmes II et III):

Théorème II bis (l'inégalité de BOHR-FAVARD généralisée): Soit $p_0(x)$ une somme finie trigonométrique

$$p_0(x) = \sum_{h=1}^k a_h e^{i\lambda_h x}$$

et soit $\Lambda = \min |\lambda_h|$ (> 0), on a pour l'intégrale $p_\alpha(x)$ d'ordre α , $\bar{\alpha}$ étant positif, en posant

$$M_\alpha = \overline{\text{borne}}_x |p_\alpha(x)| = \overline{\text{borne}}_x \left| \sum_{h=1}^k \frac{a_h}{(i\lambda_h)^\alpha} e^{i\lambda_h x} \right|$$

la limitation

$$M_\alpha \leq M_0 \cdot \frac{C_\alpha}{\Lambda^{\bar{\alpha}}}$$

où C_α est une constante, ne dépendant que de α .

Le théorème se démontre facilement au moyen des théorèmes II et IV:

Le théorème II donne

$$M_n \leq M_0 \cdot \frac{t_n}{\Lambda^n}$$

d'où, en choisissant $n > \bar{\alpha}$

$$M_\alpha \leq K_n^\alpha \cdot M_0^{1 - \frac{\bar{\alpha}}{n}} \cdot \left(M_0 \cdot \frac{t_n}{\Lambda^n} \right)^{\frac{\bar{\alpha}}{n}} = M_0 \cdot \frac{\left(K_n^\alpha \cdot t_n^{\frac{\bar{\alpha}}{n}} \right)}{\Lambda^{\bar{\alpha}}}$$

et notre théorème est démontré.

De même, on peut généraliser l'inégalité de BERNSTEIN:

Théorème III bis: Soit $p_0(x)$ une somme finie trigonométrique, on a pour l'intégrale d'ordre $-\alpha$ (la « dérivée d'ordre α »), $\bar{\alpha}$ étant positif, la limitation

$$M_{-\alpha} \leq M_0 \cdot D_\alpha \cdot \max |\lambda_h|^{\bar{\alpha}}$$

où D_α est une constante, ne dépendant que de α .

Pour les nombres α réels, le théorème II bis a été démontré par BELA V. SZ. NAGY⁽¹⁾, et le théorème III bis a été énoncé par P. CIVIN⁽²⁾, et ces auteurs communiquent aussi quelques valeurs numériques des constantes C_α et D_α (mais non les plus favorables). Ici, nous pouvons aussi donner des constantes numériques, en employant la formule $K_1^\alpha = \frac{2^\alpha}{\Gamma(1+\alpha)}$, et voici comment nous pouvons améliorer le résultat de CIVIN: Soit $0 < \alpha < 1$, on a (d'après l'inégalité de BERNSTEIN)

$$M_{-\alpha} \leq K_1^{1-\alpha} \cdot M_0^{1-\alpha} \cdot M_{-1}^\alpha \leq \frac{2^{1-\alpha}}{\Gamma(2-\alpha)} \cdot M_0 \cdot \max |\lambda_h|^\alpha$$

c'est-à-dire pour D_α ($0 < \alpha < 1$), nous avons $\frac{2^{1-\alpha}}{\Gamma(2-\alpha)}$ (la valeur donnée par CIVIN est $2 + \frac{4}{\alpha}$).

III. L'inégalité de Kolmogoroff pour une fonction arbitraire d'une variable réelle.

Dans cette partie nous allons démontrer que l'inégalité de KOLMOGOROFF est valable, non seulement pour les fonctions presque-périodiques, mais aussi pour toute fonction d'une variable réelle; c'est-à-dire le théorème suivant:

Théorème I bis (l'inégalité de KOLMOGOROFF générale): Soit $f_0(x)$ une fonction bornée (borne $|f_0(x)| = M_0 < \infty$) définie sur l'axe réel, et supposons qu'elle possède une intégrale bornée d'ordre n , $f_n(x)$ (borne $|f_n(x)| = M_n < \infty$), les intégrales de $f_0(x)$ d'un ordre inférieur à n (les dérivées de $f_n(x)$) sont aussi bornées, et l'intégrale d'ordre k , $f_k(x)$ est bornée par

⁽¹⁾ Berichte über die Verhandlungen der Sächsischen Akademie, 91 (1939), p. 3.

⁽²⁾ Bulletin of the American Mathematical Society, 46 (1940), p. 410.

$$M_k \leq M_0^{1-\frac{k}{n}} \cdot M_n^{\frac{k}{n}} \cdot \frac{t_k}{t_n^{\frac{k}{n}}}$$

avec

$$t_h = \frac{4}{\pi} \left\{ 1 + \frac{1}{(-3)^{h+1}} + \frac{1}{5^{h+1}} + \frac{1}{(-7)^{h+1}} + \dots \right\}$$

et cette inégalité ne peut pas être améliorée.

Le théorème se déduit immédiatement du théorème I (comme nous l'avons dit, ce théorème est vrai aussi pour les fonctions complexes d'une variable réelle), au moyen de la proposition suivante:

Théorème VI: A toute fonction bornée $f_0(x)$ ($-\infty < x < \infty$), qui possède une intégrale bornée d'ordre n , $f_n(x)$, on peut toujours construire une fonction périodique $g_n(x)$, telle que

$$\left| \overline{\text{borne}} \left| f_n^{(h)}(x) \right| - \overline{\text{borne}} \left| g_n^{(h)}(x) \right| \right| < \varepsilon$$

$$(h = 0, 1, 2, \dots, n) \quad (1)$$

où ε est un nombre positif quelconque, donné d'avance.

Nous abordons la démonstration de ce théorème en nous assurant que les fonctions $f_n^{(n-h)}(x) = f_h(x)$ sont bornées, c'est-à-dire que les bornes supérieures mentionnées dans le théorème sont finies; nous emploierons une méthode tout à fait analogue à celle utilisée dans la démonstration du théorème I; aussi ne ferons-nous que l'indiquer sommairement.

Il suffit de considérer les fonctions réelles $f_0(x)$.

Posons

$$q(x) = c \cdot \left(x - \frac{1}{2}\right) \left(x - \frac{3}{2}\right) \dots \left(x - \frac{2n-1}{2}\right)$$

où c est choisi assez grand pour que

(1) par $g_n^{(n)}(x)$ nous entendons une fonction dont l'intégrale est $g_n^{(n-1)}(x)$.

$$c \cdot \left(\frac{1}{2}\right)^n > M_n \quad \text{et} \quad c \cdot n! > M_0.$$

Alors, dans les points $x = 0, 1, \dots, n$, le polynôme $q(x)$ prend des valeurs de signes alternés, et les valeurs absolues sont plus grandes que M_n ; par conséquent, la différence $q(x) - f_n(x+a)$ prend aussi des valeurs de signes alternés, et nous voyons que cette différence a au moins un zéro entre deux consécutifs de ces nombres, et cela quel que soit le nombre a . Le nombre total de zéros ne peut pas excéder n (car la n -ième dérivée $c \cdot n! - f_0'(x+a)$ est toujours positive) et la différence a donc exactement un zéro entre deux consécutifs des entiers $0, 1, \dots, n$.

De cette propriété nous concluons que

$$M_{n-1} \leq \overline{\text{borne}} \left| q'(x) \right|_{0 < x < n}$$

puisque, sans cela, il serait possible de choisir le nombre a de telle manière que $q(x) - f_n(x+a)$ aurait trois zéros entre deux entiers consécutifs.

Alors, en employant cette limitation, nous pouvons trouver une borne de M_{n-2} , et ainsi de suite. Donc, il existe une constante finie $A = A(n, M_0, M_n)$, telle que

$$\left| f_n^{(k)}(x) \right| < A \quad (-\infty < x < \infty; k = 0, 1, \dots, n).^{(1)}$$

Nous allons maintenant donner la démonstration du théorème, en construisant une fonction périodique $g_{n,L}(x)$, telle que

⁽¹⁾ L'évaluation de A est sans intérêt pour nous; il faut mentionner que, par cette méthode, on peut améliorer les inégalités dues à A. GORNY (Acta mathematica 71 (1939), p. 317) et à H. CARTAN (Comptes rendus de l'Académie des Sciences 208 (1939), p. 414) entre les bornes supérieures des dérivées successives d'une fonction définie sur un intervalle fini; au lieu de $q(x)$ il faut prendre un polynôme de TCHEBYSCHEFF convenablement normé: $a \cos(n \arccos bx)$.

$$\lim_{L \rightarrow \infty} \left| \overline{\text{borne}} |g_{h,L}(x)| - \overline{\text{borne}} |f_h(x)| \right| = 0$$

$$(h = 0, 1, \dots, n; g_{h,L}(x) = g_{n,L}^{(n-h)}(x))$$

et nous démontrerons que pour $g_{n,L}(x)$ nous pouvons prendre la fonction n fois dérivable, qui possède la période $2L$, et qui dans l'intervalle $-L < x < L$ ne s'écarte de $e^{-\frac{x^2}{L}} \cdot f_n(x)$ que d'un polynôme de degré n .

Posons $e^{-\frac{x^2}{L}} \cdot f_n(x) = g_{n,L}^*(x)$, et supposons $L > 1$.

La dérivée d'ordre $n-h$ s'écrit

$$g_{h,L}^*(x) = e^{-\frac{x^2}{L}} \cdot f_h(x) + e^{-\frac{x^2}{L}} \cdot \frac{1}{L} \cdot \sum_{j,p,q,r} c_j \cdot \frac{x^p}{L^q} \cdot f_r(x)$$

où Σ signifie une somme finie, c_j est une constante numérique, $0 \leq q$, $0 \leq p \leq q+1$ et $h < r \leq n$; la justesse se voit immédiatement par induction.

Pour $|x| \leq L$ on a $\left| \frac{x^p}{L^q} \right| \leq \max\{1, |x|\}$, donc

$$\left| g_{h,L}^*(x) - e^{-\frac{x^2}{L}} \cdot f_h(x) \right| < e^{-\frac{x^2}{L}} \cdot \frac{1}{L} \cdot \max\{1, |x|\} \cdot A \cdot \Sigma |c_j|$$

d'où, premièrement

$$\begin{aligned} & \left| g_{h,L}^*(x) - e^{-\frac{x^2}{L}} \cdot f_h(x) \right| < \\ & < \frac{A \cdot \Sigma |c_j|}{L} \cdot \overline{\text{borne}} \left| e^{-\frac{x^2}{L}} \cdot \max\{1, |x|\} \right| < \frac{A \cdot \Sigma |c_j|}{L} \cdot \sqrt{L} < \frac{B}{\sqrt{L}} \end{aligned}$$

où $B = B(n, M_0, M_n)$ est une constante (comme h ne peut prendre que $(n+1)$ valeurs différentes, on peut choisir B indépendant de h), et secondement, pour $x = \pm L$

$$\left| g_{h,L}^*(\pm L) - e^{-L} \cdot f_h(\pm L) \right| < e^{-L} \cdot A \cdot \Sigma |c_j| < \frac{B}{e}$$

donc

$$\left| g_{h,L}^*(\pm L) \right| < \frac{A+B}{e}.$$

Dans l'intervalle $-L < x < L$, nous posons $g_{0,L}(x) = g_{0,L}^*(x) - a_0$ avec $a_0 = \frac{1}{2L} \int_{-L}^L g_{0,L}^*(x) dx = \frac{1}{2L} [g_{1,L}^*(L) - g_{1,L}^*(-L)]$ par quoi nous obtenons que $\int_{-L}^L g_{0,L}(x) dx = 0$, c'est-à-dire que $g_{0,L}(x)$ prolongé périodiquement a une intégrale périodique. Pour $g_{1,L}(x)$ nous prenons $\int g_{0,L}(x) dx$, où la constante d'intégration est choisie de manière que $\int_{-L}^L g_{1,L} dx = 0$; $g_{2,L}(x) = \int g_{1,L} dx$ avec $\int_{-L}^L g_{2,L} dx = 0$, et ainsi de suite, jusqu'à $g_{n,L}(x)$, qui doit être simplement une intégrale de $g_{n-1,L}(x)$.

Nous avons, pour $|x| < L$ et $h = 0, 1, \dots (n-1)$

$$g_{h,L}(x) = g_{h,L}^*(x) - \frac{a_0}{h!} \cdot x^h - \frac{a_1}{(h-1)!} \cdot x^{h-1} - \dots - a_h$$

avec

$$\begin{aligned} a_h &= \frac{1}{2L} \int_{-L}^L \left[g_{h,L}^*(x) - \frac{a_0}{h!} \cdot x^h - \frac{a_1}{(h-1)!} \cdot x^{h-1} - \dots - \frac{a_{h-1}}{1!} \cdot x \right] dx \\ &= \frac{1}{2L} [g_{h+1,L}^*(L) - g_{h+1,L}^*(-L)] - \frac{a_{h-2}}{2!} \cdot \frac{L^2}{3} - \frac{a_{h-4}}{4!} \cdot \frac{L^4}{5} - \dots \end{aligned}$$

De cette formule de récurrence on déduit

$$a_k = \sum_{j,q} d_j \cdot \frac{1}{2L} \cdot [g_{q,L}^*(L) - g_{q,L}^*(-L)] \cdot L^{k+1-q}$$

où Σ est une somme finie, d_j une constante numérique et $1 \leq q \leq k+1$.

Posons enfin $a_n = 0$.

De l'évaluation de $|g_{q,L}^*(\pm L)|$ on obtient

$$|a_k| < \sum_{j,q} |d_j| \cdot \frac{1}{2L} \cdot 2 \cdot \frac{A+B}{e^L} \cdot L^{k+1} = \frac{L^k}{e} \cdot (A+B) \cdot \Sigma |d_j|$$

et alors, dans l'intervalle $|x| < L$

$$\begin{aligned} |g_{h,L}(x) - g_{h,L}^*(x)| &= \left| \frac{a_0}{h!} \cdot x^h + \frac{a_1}{(h-1)!} \cdot x^{h-1} + \dots + a_h \right| \\ &\leq (h+1) \cdot \frac{L^h}{e} \cdot (A+B) \cdot \max_h \{ \sum |d_j| \} \end{aligned}$$

ce qui nous montre qu'il existe une constante $C = C(n, M_0, M_n)$, telle que

$$|g_{h,L}(x) - g_{h,L}^*(x)| < \frac{L^n}{e} \cdot C$$

uniformément pour $L > 1$, $h = 0, 1, \dots, n$ et $|x| < L$.

Conformément à la méthode de construction $g_{n,L}(x)$ (prolongé périodiquement) est une intégrale périodique d'ordre n d'une fonction périodique $g_{0,L}(x)$, et il ne nous reste qu'à vérifier la proposition

$$\lim_{L \rightarrow \infty} \left| \overline{\text{borne}} |g_{h,L}(x)| - \overline{\text{borne}} |f_h(x)| \right| = 0.$$

Dans l'intervalle $-L < x < L$ on a

$$\begin{aligned} &\left| g_{h,L}(x) - e^{-\frac{x^2}{L}} \cdot f_h(x) \right| \leq \\ &\leq \left| g_{h,L} - g_{h,L}^* \right| + \left| g_{h,L}^* - e^{-\frac{x^2}{L}} \cdot f_h \right| < \frac{L^n}{e} \cdot C + \frac{1}{\sqrt{L}} \cdot B \end{aligned}$$

et comme $g_{h,L}(x)$ a la période $2L$ on en déduit

$$\lim_{L \rightarrow \infty} \left| \overline{\text{borne}} \left| g_{h,L}(x) \right| - \overline{\text{borne}}_{|x| < L} \left| e^{-\frac{x^2}{L}} \cdot f_h(x) \right| \right| = 0.$$

Enfin nous voyons que

$$\lim_{L \rightarrow \infty} \left| \overline{\text{borne}}_{|x| < L} \left| e^{-\frac{x^2}{L}} \cdot f_h(x) \right| - \overline{\text{borne}} |f_h(x)| \right| = 0$$

car, $\left| e^{-\frac{x^2}{L}} \cdot f_h(x) \right| < |f_h(x)|$ et, x étant fixe, tout point x va entrer dans l'intervalle $|x| < L$ et $e^{-\frac{x^2}{L}}$ convergera vers 1.

La démonstration de notre théorème est ainsi complètement achevée.



Det Kgl. Danske Videnskabernes Selskab.

Mathematisk-fysiske Meddelelser. **XIX**, 5.

PYRENSTUDIEN

II. REDUKTION DES NITROPYRENS

VON

HAKON LUND UND ARNE BERG



KØBENHAVN

I KOMMISSION HOS EJNAR MUNKSGAARD

1941

Printed in Denmark.
Bianco Lunos Bogtrykkeri A/S.

Einleitung.

In dieser zweiten Mitteilung¹⁾ über Pyrenderivate soll über die Ergebnisse berichtet werden, die bei der Reduktion des Nitropyrens erhalten wurden.

3-Nitropyren wird in einfachster Weise durch Nitrierung von Pyren in Eisessig mit der berechneten Menge konz. Salpetersäure dargestellt, und die Reduktion von Nitropyren zu Aminopyren nach den bekannten Methoden gestaltet sich auch sehr leicht. Dagegen sind die aus der Benzolchemie bekannten Zwischenstufen der Reduktion: Nitroso-, Azoxy-, Hydroxylamino, Azo- und Hydrazo-Verbindungen in der Pyrenreihe bisher nicht beschrieben worden. Das Ziel der vorliegenden Untersuchungen war, diese Stoffe darzustellen und ihre Eigenschaften zu studieren. Es ist uns bisher jedoch nicht gelungen, sämtliche Zwischenstufen zu isolieren; trotz vieler Bemühungen fehlen in der genannten Reihe noch die Azo- und die Hydrazoverbindungen.

Ein solches Ergebnis kann nicht überraschen. Die Nitrosubstitute mehrkerniger Kohlenwasserstoffe weichen in ihrem Verhalten gegenüber Reduktionsmitteln beträchtlich von den Nitro-benzolkohlenwasserstoffen ab. Die Anzahl der Literaturangaben über partielle Reduktion solcher Verbindungen ist verhältnismässig gering; es geht jedoch aus den vorliegenden Arbeiten klar hervor, dass die Reduktions-

¹⁾ Erste Mitteilung: D. Kgl. Danske Vidensk. Selskab, Math.-fys. Medd. XVIII, 9 (1941).

verfahren, die sich in der Benzolchemie bewährt haben, oft andersartige Reaktionen hervorrufen, wenn sie auf mehrkernige Nitrokohlenwasserstoffe angewendet werden. Zur Erläuterung soll hier eine kurze Übersicht über solche Untersuchungen gegeben werden, und zwar sollen dabei die Ergebnisse der von uns vorgenommenen, entsprechenden Versuche mit Pyren hinzugefügt werden.

1. Kaliumhydroxyd und Alkohol.

Bekanntlich wird Nitrobenzol von alkoholischem Kali in Azoxybenzol überführt. α -Nitronaphthalin liefert, mit alkoholischem Kali gekocht, einen Teer, aus dem kein kristallisiertes Produkt gewonnen werden kann. Mit kaltem methylalkoholischem Kali behandelt, gibt α -Nitronaphthalin wasserlösliche Substanzen, die nicht durch Reduktion, sondern durch Umlagerung entstehen (Kaliumsalz des Naphtochinonmonoxims¹). Dagegen wird β -Nitronaphthalin von kaltem methylalkoholischem Kali überhaupt nicht angegriffen, während beim Kochen etwas — ungefähr 30 % d. Theor. — Azoxynaphthalin neben grösseren Mengen wasserlöslicher Produkte entsteht²). 9-Nitroanthracen wird von methylalkoholischem Kali in Anthrachinonmonoxim umgelagert. Es konnte keine Reduktion festgestellt werden³). Auch 9-Nitrophenanthren erleidet hauptsächlich Umlagerung. Es entstehen jedoch kleine Mengen von Azo- und Hydrazophenanthren⁴).

Nitropyren, in alkoholischer Lösung einige Zeit mit Kaliumhydroxyd gekocht, liefert hauptsächlich Azoxy-

¹) MEISENHEIMER, *Ann.* **355**, 299 (1907).

²) MEISENHEIMER & WITTE, *Ber.* **36**, 4164 (1903).

³) MEISENHEIMER, *Ann.* **323**, 205 (1902).

⁴) MEISENHEIMER, *Ann.* **355**, 307 (1907).

pyren. Es ist jedoch schwierig, in dieser Weise reines Azoxyphenylpyren zu erhalten. Die ausgeschiedenen, sehr schwer löslichen Körper, welche unter variierenden Versuchsbedingungen isoliert wurden, waren zwar einander sehr ähnlich — ihr Stickstoffgehalt betrug immer 5,9 bis 6,1 % — konnten aber nicht reines Azoxyphenylpyren sein, da dessen Formel 6,28 % N verlangt. Aller Wahrscheinlichkeit nach bestanden die Niederschläge aus Molekülverbindungen von Azoxyphenylpyren und Nitrophenylpyren: Eine Verbindung aus 1 Mol. Azoxyphenylpyren und 1 Mol. Nitrophenylpyren enthält 6,06 % N. Bei zweckmässiger Durchführung der Reaktion konnte jedoch schliesslich Azoxyphenylpyren in reinem Zustand und mit befriedigender Ausbeute erhalten werden.

2. Zinkstaub und alkoholisches Kaliumhydroxyd.

Nitrobenzol liefert Azo- und Hydrazobenzol. α -Nitronaphthalin wird in einen Körper von unbekannter Konstitution verwandelt. Daneben entsteht etwas α -Naphthylamin¹⁾. β -Nitronaphthalin erzeugt eine gelbe Lösung, die, mit Luft geschüttelt, einen roten Niederschlag liefert, der aus mehreren sauerstofffreien Verbindungen besteht. β -Azonaphthalin konnte aus dem Gemisch isoliert werden. Es muss also angenommen werden, dass primär u. a. Hydrazonaphthalin entstanden ist²⁾. 9-Nitrophenanthren wird mit geringer Ausbeute zur entsprechenden Azoverbindung reduziert³⁾.

Nitrophenylpyren wird beim Kochen mit Zinkstaub und Kaliumhydroxyd zum grössten Teil erst als Azoxyphenylpyren ausgefällt und darauf langsam — vermutlich auf Grund

¹⁾ WACKER, *Ann.* **317**, 375 (1901); CUMMING & STEEL, *J. Chem. Soc.* **123**, 2464 (1923).

²⁾ MEISENHEIMER & WITTE, *Ber.* **36**, 4153 (1903).

³⁾ SCHMIDT & STROBEL, *Ber.* **36**, 2508 (1903).

der Schwerlöslichkeit dieser Substanz — weiter reduziert. Es bildet sich hauptsächlich Aminopyren. Daneben entstehen mehrere Körper, die nicht isoliert wurden. Die nur schwach gefärbte Lösung wird in Berührung mit der Luft dunkel grün; Hydrazopyren kann somit kaum als die sauerstoffempfindliche Komponente der Lösung in Frage kommen.

3. Zinkstaub und Ammoniumchlorid in Alkohol (+ Wasser).

Aus Nitrobenzol entsteht Phenylhydroxylamin. α -Nitronaphthalin wird sukzessiv in Azoxy-, Azo- und Hydrazonaphthalin überführt, wenn die Reaktionstemperatur zwischen 70 und 75° gehalten wird. Beim Kochen entsteht α -Naphthylamin¹⁾.

Nitropyren wird bei anhaltendem Kochen zum Amin reduziert. Es ist jedoch augenfällig, dass die Reaktion über Zwischenstufen verläuft, und man kann den Prozess derart leiten, dass die entstandene Lösung hauptsächlich Hydroxylaminopyren enthält. Es ist aber bisher nicht gelungen, diese Substanz aus der Lösung zu isolieren, da sie sich während der Aufarbeitung in andere Stoffe verwandelt. Führt man dagegen die Reaktion so aus, dass man zu einem Gemisch von Nitropyren und Zinkstaub Äther und eine wässrige Lösung von Ammoniumchlorid gibt, die etwas freies Ammoniak enthält, kann man eine ätherische Lösung erhalten, die beim Eingiessen in Benzin kristallines, ziemlich reines Hydroxylaminopyren liefert.

4. Alkalisulphhydrate.

Nitrobenzol wird in der Kälte zu Phenylhydroxylamin reduziert; in der Siedehitze entsteht Anilin. α -Nitronaph-

¹⁾ CUMMING & STEEL, loc. cit.

thalin¹⁾ und β -Nitronaphthalin²⁾ werden bei 0° von alkoholischem Ammoniumhydrosulfid in die Hydroxylamine überführt; beim Erhitzen entstehen die Amine.

Nitropyren wird beim Sieden von Natriumhydrosulfid in wässrigem Alkohol zu Aminopyren reduziert (zweckmässige Methode zur Darstellung des Aminopyrens, Ann. **531**, 109 (1937)). Jedem, der diesen Prozess ausführt, muss es jedoch klar werden, dass die Reduktion in zwei Stufen verläuft: Wenn man zur heissen Suspension von Nitropyren in Alkohol die Sulfidlösung gibt, geht das Nitropyren schnell und unter bedeutender Wärmeentwicklung in Lösung; aber erst nach etwa zweistündigem Sieden ist die Aminbildung beendet. Das Zwischenprodukt ist zweifelsohne Hydroxylaminopyren. Es wurde indessen festgestellt, dass Natrium- oder Ammoniumhydrosulfid auch bei niedriger Temperatur das Nitropyren zum Amin reduziert; es gelang somit nicht, die Reduktion an der Hydroxylaminstufe zum Stillstand zu bringen.

Aus dieser Übersicht geht hervor, dass wir als direkte Reduktionsprodukte des Nitropyrens ausser dem Amin nur Azoxy- und Hydroxylaminopyren haben nachweisen können. Viele andere Reduktionsmittel wurden versucht, jedoch ohne Erfolg. Das Amin konnte mit den verschiedensten Reduktionsmitteln erzeugt werden: Katalytische Reduktion, Reduktion mit Stannochlorid (auch wenn im Unterschuss zugesetzt), Natriumamalgam u. v. a. Elektrolytische Reduktion war erfolglos; es entstand ein schwarzer, kristallisationsunfähiger Teer.

Nach unseren Erfahrungen erzeugen alle Reduktions-

¹⁾ WILLSTÄTTER & KUBLI, Ber. **41**, 1936 (1908).

²⁾ BAUDISCH & FÜRST, Ber. **50**, 324 (1917).

mittel, die Nitropyren überhaupt angreifen, immer etwas Aminopyren, das an der intensiv blauen Fluoreszenz seiner Lösungen leicht erkennbar ist.

Es gelang uns, ausser den oben erwähnten Reduktionsprodukten noch das Nitrosopyren darzustellen, allerdings auf eine ganz unerwartete Weise. Die gewöhnliche Methode zur Darstellung der aromatischen Nitrosokohlenwasserstoffe: Oxydation des entsprechenden Arylhydroxylamins, versagte in diesem Falle trotz zahlreicher Versuche. Als wir aber die Eigenschaften des Hydroxylaminopyrens nach anderen Richtungen hin untersuchten, stiessen wir auf das Nitrosopyren.

Es war unsere Absicht zu untersuchen, ob die Umlagerungsreaktion des Phenylhydroxylamins in p-Aminophenol auf Hydroxylaminopyren übertragen werden kann. Diese Umlagerung wird durch Mineralsäuren bewirkt, am besten durch Schwefelsäure, die die wenigsten Nebenreaktionen hervorruft. Als wir nun eine Hydroxylaminopyren-lösung mit Schwefelsäure vermischten, entstanden augenblicklich zwei Substanzen, von denen die eine sich leicht als das Sulfat des Aminopyrens identifizieren liess, während die andere, ein roter, sehr schön kristallisierender Körper, sich als Nitrosopyren erwies.

Statt einer Umlagerung hat also die Säure eine Disproportionierung bewirkt:



Eine solche Disproportionierung eines Arylhydroxylamins ist unsers Wissens nicht früher beobachtet worden. Doch ist die Beobachtung BAMBERGERS¹⁾, dass Phenyl-

¹⁾ BAMBERGER & BRADY, Ber. **33**, 274 (1900).

hydroxylamin in luftfreier, wässriger Lösung in Gegenwart von Natriumhydroxyd Azoxybenzol und Anilin liefert, wohl so zu deuten, dass zuerst zwei Mol. Hydroxylamin ein Mol. Anilin und ein Mol. Nitrosobenzol ergeben, woraufhin sich das Nitrosobenzol sofort mit Phenylhydroxylamin unter Bildung von Azoxybenzol umsetzt. Diese letzte Reaktion tritt nicht in saurer Lösung ein. Der Unterschied zwischen dem Phenyl- und dem Pyrenylhydroxylamin besteht also darin, dass die Disproportionierung des letztgenannten auch in saurer Lösung stattfindet. Die Umlagerung Hydroxylamin \rightarrow Aminophenol konnte beim Pyrenabkömmling nicht nachgewiesen werden, sei es nun, weil sie nicht stattfinden kann, oder aber weil die Disproportionierung so schnell verläuft, dass die gebildete Oxyaminmenge minimal wird.

Die fehlgeschlagenen Versuche, Nitrosopyren bei Oxydation des Hydroxylamins zu gewinnen, finden durch die Disproportionierung ihre Erklärung: das sofort gebildete Amin wird zu einem Anilinschwarz-ähnlichen Körper oxydiert, und Nitrosopyren wird — wie die Versuche zeigen — von kalter, schwefelsaurer Chromsäurelösung rasch weiter oxydiert.

Weitere Versuche, um die noch fehlenden Verbindungen Azo- und Hydrazopyren zu erhalten, sind zurzeit im Gang.

Beschreibung der Versuche.

I. Nitropyren und alkoholisches Kali. Darstellung von Azoxypyren.

Beim Erhitzen einer Suspension oder einer Lösung von Nitropyren in Alkohol, die Kaliumhydroxyd enthält, wird die anfangs gelbe Lösung bald dunkel; nach einiger Zeit

setzt sich ein brauner, sehr schwerlöslicher Niederschlag in Flocken ab, und das Nitropyren verschwindet. Obgleich die Versuchsbedingungen zunächst in vielfacher Weise variiert wurden — verschiedene Alkohole und verschiedene Kalikonzentrationen wurden verwendet — wurden Produkte erhalten, deren Stickstoffgehalt ziemlich konstant war, aber keiner einfachen Formel entsprach. Die folgenden Ergebnisse wurden ermittelt:

Stickstoffgehalt der ausgeschiedenen, umkristallisierten Produkte: In Äthylalkohol bei variierender Kalikonzentration 5,88, 5,95, 6,04, 6,06, 5,94 %; in n-Propylalkohol 5,84, 5,99 %; in n-Butylalkohol 5,90, 5,99 %; in Isoamylalkohol 5,90, 5,98 %.

Die Formel des Azoxypyrens verlangt 6,28 % N. Es dürfte somit ein Gemisch vorliegen. Wiederholte Umkristallisierungen führten nicht zu einer Trennung der Produkte. Aller Wahrscheinlichkeit nach hatte das gebildete Azoxypyren mit noch vorhandenem Nitropyren eine Molekülverbindung der zwei Komponenten gebildet, die wegen ihrer Schwerlöslichkeit nicht weiter reagieren konnte. Ein direkter Beweis hierfür konnte jedoch nicht erbracht werden, da auch das reine Azoxypyren in Alkohol so schwer löslich ist, dass das Zusammengiessen gesättigter Lösungen der beiden Stoffe keinen Niederschlag hervorruft.

Schliesslich stellte sich heraus, dass annäherungsweise reines Azoxypyren erhalten werden kann, wenn eine Lösung von Nitropyren in siedendem Alkohol mit hinreichend viel Kaliumhydroxyd versetzt wird. Die Darstellung des Azoxypyrens gelang in befriedigender Weise nach der folgenden Vorschrift:

Eine Lösung von 10 g Nitropyren in 1250 ml siedendem Alkohol wurde mit einer Lösung von 55 g Kaliumhydroxyd

in 125 ml heissem Alkohol versetzt und auf dem Wasserbade unter Rückfluss gekocht. Nach etwa 15 Minuten begann die Ausscheidung des Azoxypyrens, und nach einer Stunde war die Reaktion beendet. Der Niederschlag wurde auf dem Filter zuerst mit Alkohol, dann mit Wasser und schliesslich wieder mit Alkohol gewaschen. Ausbeute 6,5 g. Das Produkt wurde zur Reinigung mit 100 ml Äther gründlich ausgekocht und dann aus etwa 1 l Xylol umkristallisiert. Es wurden auf diese Weise 4,0 g braune Kristalle erhalten, die bei 264° korr. schmolzen und sich bei Erhöhung der Temperatur um wenige Grade unter Luftentwicklung zersetzten.

Die alkoholische Mutterlauge des Rohproduktes fluoreszierte intensiv blau (Anwesenheit von Aminopyren).

Analyse.

Gef. 85,9 % C; 4,15 % H; 6,37 % N.

Ber. für $C_{32}H_{18}ON_2$: 86,1 % C; 4,07 % H; 6,28 % N.

Die Reduktion des Nitropyrens zum Azoxypyren verlief etwas schneller mittels einer alkalisch-alkoholischen Lösung von Glukose:

5 g Nitropyren wurden in 700 ml heissem Alkohol gelöst und auf etwa 70° abgekühlt; danach wurden 5,5 g Traubenzucker, in wenig Wasser gelöst, hinzugefügt, und endlich wurde mit einer Lösung von 5 g Kaliumhydroxyd in 5 ml Wasser versetzt. Nach wenigen Minuten schied sich das Azoxypyren als gelbroter Niederschlag mit einer Ausbeute von 3,0 g ab; Schmp. nach Umkristallisieren aus Xylol 263° korr.

Analyse.

Gef. 85,6 % C; 4,17 % H; 6,22 % N.

Azoxypyren ist in allen gewöhnlichen Lösungsmitteln sehr wenig löslich und wird deshalb von den verschiedenen Reduktionsmitteln sehr langsam angegriffen. Am leichtesten wird es zum Amin reduziert, wenn das flockige Rohprodukt mit alkoholischem Ammoniumhydrosulfid erhitzt wird. Wenn die Substanz in Form kompakterer Kristalle vorliegt, ist es praktisch unmöglich, sie in Reaktion zu bringen.

Ausser bei den erwähnten Prozessen kann das Azoxypyren auf folgende Weisen erhalten werden: durch Reduktion von Nitrosopyren mittels alkoholischen Kalis, durch Oxydation von Hydroxylaminopyren mit Sauerstoff und durch Zusammengiessen der Lösungen von Nitrosopyren und Hydroxylaminopyren.

II. Nitropyren und Zinkstaub + Ammoniumchlorid.

Die Klarlegung der Vorgänge bei der Einwirkung von Zinkstaub auf Nitropyren in alkoholischer Lösung erwies sich als sehr schwierig; einige hierbei gewonnene Erfahrungen sollen im folgenden nicht unerwähnt bleiben.

Eine Lösung oder Suspension von Nitropyren in Alkohol, dem etwas Ammoniumchlorid oder Ammoniakwasser zugesetzt ist, wird von Zinkstaub in der Kälte langsam, beim Erwärmen schneller reduziert. Wenn die Luft nicht ausgeschlossen wird, entsteht dabei Azoxypyren, das nur sehr langsam weiter reduziert werden kann. Wird der Zutritt von Luft verhindert, erhält man schliesslich eine hellgelbe Lösung, die hauptsächlich Hydroxylaminopyren neben einer kleinen Menge Aminopyren enthält. Wird die Lösung sehr lange gekocht, erhält man das Amin als Hauptprodukt. Die Reduktion tritt schneller ein, wenn etwas freies Ammoniak neben dem Ammoniumchlorid zugegen ist.

Die alkoholische Lösung des Hydroxylaminopyrens wird vom Luftsauerstoff ausserordentlich schnell zu Azoxyphenylpyren oxydiert. Die beinahe farblose Lösung wird in Berührung mit der Luft augenblicklich rot, und das flockige Azoxyphenylpyren scheidet sich an der Oberfläche der Lösung ab. Da es nicht gelang, das Hydroxylaminopyren aus der alkoholischen Lösung zu isolieren, versuchten wir, es mittels Benzaldehyds abzufangen. (Bekanntlich gibt Phenylhydroxylamin mit Benzaldehyd eine Verbindung



Wir erhielten dabei auch einen Niederschlag, der sich aber als die Schiff-Base aus Benzaldehyd und Aminopyren erwies. (Schmp. 122°; Analyse: Gef. 90,7 % C; 4,98 % H; 4,75 % N. Ber. für C₂₃H₁₅N: 90,5 % C; 4,96 % H; 4,59 % N.) Zum Vergleich wurde diese Verbindung aus Aminopyren hergestellt, und die Identität der beiden Substanzen wurde sichergestellt.

Es wurde versucht, das Hydroxylaminopyren beim Fällen der alkoholischen Lösung mit Wasser zu isolieren. Es fiel ein nur schwach gefärbter Niederschlag aus, der sich aber nach wenigen Minuten in ein Gemisch von Azoxyphenylpyren und Aminopyren verwandelt hatte. Der von BAMBERGER beobachtete Zerfall des Phenylhydroxylamins in Azoxybenzol und Anilin in Gegenwart von wässrigem Kaliumhydroxyd tritt hier offenbar schon mit kleinen Mengen Ammoniak schnell ein.

In ätherischer Lösung verläuft die Luftoxydation des Hydroxylaminopyrens beträchtlich langsamer als in alkoholischer Lösung. Zwar wird die ätherische Lösung schnell rot; aber es dauert 15—30 Minuten, ehe die Ausscheidung des auch in Äther schwer löslichen Azoxyphenylpyrens anfängt. Zur

Darstellung des Hydroxylaminopyrens ist es aus diesem und noch anderen Gründen zweckmässig, die Reaktion in Äther durchzuführen. Um das sichere Gelingen dieses Verfahrens zu gewährleisten, ist es aber notwendig, die unten beschriebenen Bedingungen ziemlich genau zu erfüllen. Das betrifft im besonderen die zur Reaktion notwendige Menge der wässrigen Lösung von Ammoniumchlorid und Ammoniak: Eine wesentlich grössere oder kleinere Menge hindert oder verzögert die Reaktion in hohem Grade.

Wenn die Isolierung des reinen Hydroxylamins erwünscht ist, verfährt man in folgender Weise:

Drei SCHLENK-Röhren¹⁾, I, II und III, werden so verbunden, dass zwischen I und II ein Wattefilter und zwischen II und III ein Jena-Glasfilter eingeschaltet ist. In I werden 2 g feingepulvertes Nitropyren, 3 g Zinkstaub (von guter Qualität), 1 ml einer wässrigen Lösung, die aus 2 Vol. 25 % Ammoniumchlorid und 1 Vol. konz. Ammoniakwasser besteht, und endlich 30—40 ml Äther angebracht; Rohr II wird mit 75 ml Leichtbenzin (Sp. 60—70°) beschickt, und die ganze Apparatur wird mit sauerstofffreiem Stickstoff gefüllt. Die Reaktion tritt von selbst oder bei leichtem Schütteln des Rohres ein und wird durch zeitweises Schütteln gefördert. Nach 15—30 Minuten ist das Nitropyren verschwunden und die rote Farbe der Lösung in eine gelbliche verwandelt. Dann wird diese Lösung in das Benzin filtriert. Es scheidet sich sofort ein gelber Niederschlag ab, der bald hellbraun wird. Das Produkt wird filtriert, mit etwas in Rohr II unter Stickstoff eingeführtem

¹⁾ SCHLENK & THAL, Ber. **46**, 2843 (1913); HOUBEN-WEYL, Bd. IV, 960. (Leipzig 1924).

Benzin gewaschen und in einem Stickstoffstrom auf dem Filter getrocknet. Ausbeute 1,2 g. — Die Mutterlauge scheidet beim Stehen an der Luft etwa 0,2 g Azoxypyren ab, das bei Oxydation des in der Mutterlauge verbliebenen Hydroxylamins entstanden ist.

Das Hydroxylaminopyren zersetzt sich sehr unscharf bei etwa 200°.

Analyse.

Gef. 81,8 % C; 4,81 % H; 6,09 % N.

Ber. für $C_{16}H_{11}ON$: 82,4 % C; 4,72 % H; 6,01 % N.

Benötigt man nur eine ätherische Lösung von Hydroxylaminopyren, so gestaltet sich die Darstellung wesentlich einfacher:

In einem grossen Reagenzrohr werden 5 g feinpulverisiertes Nitropyren, 8 g Zinkstaub und 3 ml der oben erwähnten Ammoniumchlorid-Ammoniaklösung mit 75–100 ml Äther übergossen. Nach kurzem Rühren mit einem Spatel fängt der Äther an zu sieden. Lässt die Reaktion nach, rührt man wieder kurze Zeit; wird sie zu lebhaft, muss mit Wasser gekühlt werden. Wenn die rote Farbe verschwunden ist, ist die Reaktion beendet, und die Lösung kann leicht vom Zinkschlamm abgossen werden. Sie kann zu verschiedenen Reaktionen verwendet werden:

a) »Pyren-cupferron«. Beim Eingiessen einer ätherischen, ammoniakhaltigen Lösung von Hydroxylaminopyren in eine ätherische Lösung von Amylnitrit fällt augenblicklich ein gelber, kristallinischer Körper aus. Aus 5 g Nitropyren konnten auf diese Weise 3,5 g erhalten werden. Es handelt sich um das »Pyren-cupferron«, $C_{16}H_9 \cdot N(NO) \cdot ONH_4$.

Die Analyse musste mit dem Rohprodukt ausgeführt werden, weil es uns nicht gelang, ein zum Umkristallisieren geeignetes Mittel zu finden.

Analyse.

Gef. 69,1 % C; 4,97 % H; 14,5 % N.

Ber. für $C_{16}H_{13}O_2N_3$: 68,8 % C; 4,66 % H; 15,0 % N.

Die Verbindung ist in Wasser sehr schwer löslich. Die Fällung der Eisen- und Kupfersalze konnte folgendermassen ausgeführt werden: In Aceton, das ein wenig Eisessig enthält, löst sich die Substanz mit gelber Farbe, und die Lösung hält sich beim Zusatz von 1 Vol. Wasser klar. Ein Tropfen Ferrichlorid erzeugt darin sofort einen voluminösen, dunklen Niederschlag, ein Tropfen Kupfersulfatlösung dagegen einen hellbraunen.

b) Lässt man eine Lösung von Hydroxylaminopyren in Äther ruhig an der Luft stehen, scheidet sich langsam und in quantitativer Ausbeute Azoxypyren in reinem Zustand aus.

c) Die Lösung kann zur Darstellung des Nitrosopyrens durch Reaktion mit verdünnter Schwefelsäure verwendet werden.

Die Darstellung des Nitrosopyrens gelingt am besten mit kleinen Substanzmengen. Eine ätherische Lösung von Hydroxylaminopyren, aus 5 g Nitropyren wie oben beschrieben dargestellt, wird schnell unter kräftigem Rühren in ein Becherglas gegossen, das 100 ml mit 100 ml Äther überschichteter etwa 6-n Schwefelsäure enthält und mit Kohlendioxyd gefüllt ist. Es entsteht eine wässrige Suspension von Aminopyrensulfat und eine orangefarbene ätherische Schicht. Nach der Trennung der Schichten schüttelt man die Wasserphase nochmals mit Äther aus. Die ver-

einigten Ätherauszüge werden mit Wasser gewaschen, getrocknet und eingedampft. Der Rückstand wird aus Alkohol umkristallisiert, wobei man das Nitrosopyren in prachtvollen, roten Nadeln erhält. Ausbeute 1,2—1,3g, Schmp. 150°.

Aus 1 g Nitropyren konnte 0,4 g (Rohprodukt) erhalten werden, während man aus 10 g nur 1,5 g gewann.

Analyse.

Gef. 82,8 % C; 3,96 % H; 5,92 % N.

Ber. für $C_{16}H_9ON$: 83,1 % C; 3,90 % H; 6,06 % N.

0,202 g Nitrosopyren, in 25,1 g Benzol gelöst, erniedrigte den Schmelzpunkt des Benzols um 0,175°. Daraus ergibt sich das Molekulargewicht des gelösten Nitrosopyrens zu 231; aus der Formel wird 231 berechnet. Das Nitrosopyren ist somit in dieser Lösung völlig monomer.

Die Farbe der Lösungen ist rotorange, ungefähr wie die einer Chromsäurelösung. In ganz dünner Schicht hat die gelbe Farbe einen deutlich grünen Stich.

Reaktionen des Nitrosopyrens.

Eine alkoholische Lösung von Nitrosopyren, mit wässrigem Ammoniumhydrosulfid kurze Zeit erwärmt, liefert Aminopyren.

Eine alkoholische Lösung, mit ein wenig Kaliumhydroxyd versetzt, gibt nach kurzer Zeit bei etwa 50° Azoxy-pyren. Versetzt man die alkoholische Lösung mit Diäthylamin und kocht kurze Zeit, wird auch Azoxy-pyren erhalten.

Eine ätherische Lösung von Nitrosopyren mit der äquivalenten Menge einer ätherischen Lösung von Hydroxylaminopyren vermischt, liefert in quantitativer Ausbeute Azoxy-pyren.

Beim Kochen einer alkoholischen Lösung von Nitrosopyren mit Cyanessigeste wurde ein Kondensationsprodukt gewonnen, wahrscheinlich $C_{16}H_9 \cdot N:C(CN) \cdot COOC_2H_5$, das jedoch nicht näher untersucht worden ist.

In der Hoffnung, Azopyren zu erhalten, haben wir Nitrosopyren und Aminopyren in Eisessig erhitzt. Es entstand eine schwarze Masse, aus der kein kristallisierbarer Körper isoliert werden konnte.



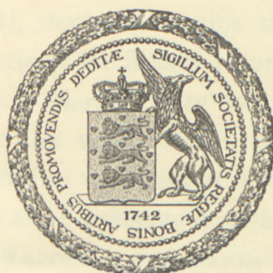
Det Kgl. Danske Videnskabernes Selskab.

Mathematisk-fysiske Meddelelser. **XIX**, 6.

DEUTERON INDUCED FISSION OF
URANIUM AND THORIUM

BY

J. C. JACOBSEN AND N. O. LASSEN



KØBENHAVN

I KOMMISSION HOS EJNAR MUNKSGAARD

1941

DEUTERON INDUCED FISSION OF
URANIUM AND THORIUM

CONTENTS

	Page
1. Introduction	3
2. Experimental method	4
3. Yield of fission as a function of deuteron energy	7
4. Comparison between uranium and thorium	12
5. Absolute value of fission cross-section	14
6. Comparison of the results with the theory	18
References	24



1. Introduction.

The first account of experiments with the object of producing nuclear fission with artificially accelerated charged particles was published by GANT [1]. Using deuteron beams of 8 MeV from the cyclotron in the Cavendish Laboratory in Cambridge, GANT obtained detectable fission effects in uranium, but he did not attempt a quantitative determination of the cross-section for the process. In view of the desirability of a closer study of the problem and especially of testing the theoretical considerations developed by BOHR and WHEELER [2], further experiments were carried out by the writers by means of the cyclotron constructed in the Institute of Theoretical Physics at Copenhagen and recently described by one of us [3]. In these experiments, of which a preliminary account has already been published [4], deuteron beams up to 9,5 MeV were used and cross-sections for fission in uranium and thorium for varying deuteron energies were measured. In particular, it was found that the value of the relative fission cross-section for thorium and uranium at 9 MeV deuteron energy was about 0,7, in close agreement with the value obtained independently by KRISHNAN and BANKS [5] in the Cavendish Laboratory. In the present paper, a fuller account of the method used in our experiments and of the results obtained is given together with a closer comparison with the theoretical expectations.

2. Experimental method.

The irradiation of the thorium and uranium samples was carried out within the vacuum tank of the cyclotron since the deflecting field was not sufficient to bring the deuteron beam outside the tank wall. The targets were inserted and removed from the cyclotron by means of a vacuum lock which, besides avoiding air penetration into the tank, had the advantage that the targets could be brought close to the deflecting electrode, where the cross-section of the beam is smallest. This part of the arrangement was described in the earlier paper [3].

In most of the experiments with uranium, the target was a thin layer of metallic uranium supported on a sheet of aluminium. The target was prepared by evaporation of metallic uranium in the form of powder from a tungsten ribbon in vacuum. The thickness of the uranium layer was determined by weighing one of the targets; the amount of uranium in the other targets was then determined relatively to this one by counting the β -particles emitted by the samples. The thickness of the uranium layers used was about 0,5 mg per cm^2 . For the experiments with thorium, a quantity of thorium in the form of powder was available. Still, attempts to obtain a layer of metallic thorium by evaporation were unsuccessful because, what was soon found, the powder consisted mainly of thorium oxide. The thorium targets finally used in the experiments were in all cases "thick" targets obtained by compressing thorium oxide.

The target holder is shown in Fig. 1. The targets are fastened by screws in the positions a_1 , a_2 , a_3 and a_4 on the four sides of a brass block supported by a brass tube.

The latter is insulated by a pyrex tube from an outer brass tube which slides through a rubber gasket (not shown in the figure) so that the target holder can be moved in and out and rotated. The outer brass tube supports a cylindrical brass case with four openings, c_1 , c_2 , c_3 and c_4 , facing the four sides of the brass block; the energy of

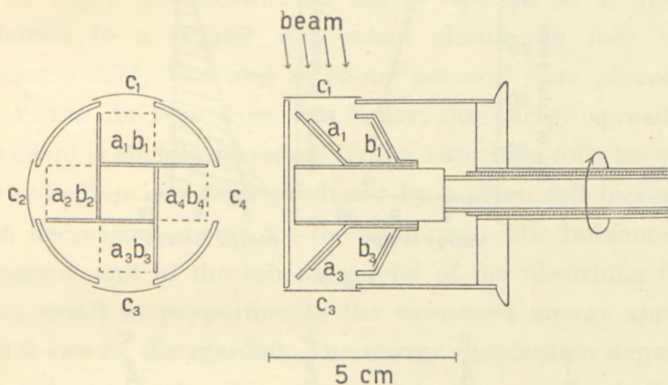


Fig. 1.

the deuterons hitting the target can be varied by covering the openings in the case with aluminium foils. The space inside the brass case is divided into four partitions by copper sheets. Facing the targets, aluminium foils of thickness $6,5 \text{ mg per cm}^2$ used for collecting the fission products could be placed in the positions b_1 , b_2 , b_3 and b_4 . The β -ray activity of the fission products was measured by a counter connected to a scale-of-eight amplifier. The stopping power of the window of the counter corresponded to 4 cm of air, the solid angle occupied by the window, as seen from the radioactive source measured, was about $\frac{1}{10} \cdot 4 \pi$.

The deuterons in the beams used had energies up to 9,5 MeV. The total beam current was 0,1–0,2 microamp.,

which was sufficient for the experiments; with larger currents the cooling of the targets would also have caused difficulties with the present arrangement. The energy distribution of the deuteron beam was determined by

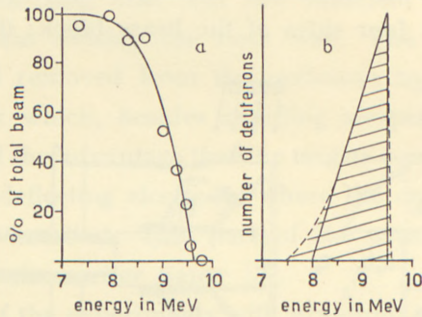


Fig. 2.

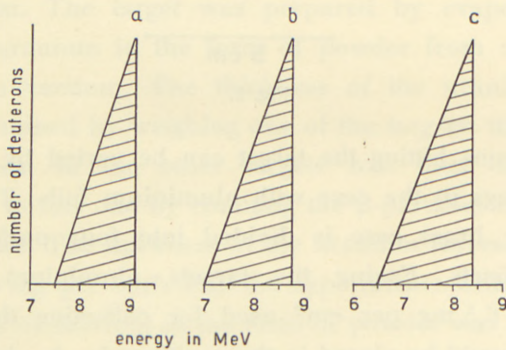


Fig. 3.

placing aluminium foils in the beam and measuring the transmitted current; the connection between range and energy was obtained by use of the range curves given by LIVINGSTON and BETHE [6]. The ordinates of the curve *a* in Fig. 2 indicate, for one of the deuteron beams used, the part of the beam current with energy greater than the values plotted as abscissae. The corresponding energy

distribution *b*, shown in Fig. 2, is obtained by differentiation of the absorption curve *a*. The distribution thus obtained is given by the dotted line. For the evaluation of the results, however, the curve shown by the full line was used. It will appear from the later discussion of the results that the difference between the two curves is insignificant.

In Fig. 3 are shown the energy spectra *a*, *b* and *c* obtained in a similar way when aluminium foils with thickness 6,75, 13,5 and 20,25 mg per cm² were placed in the beam. As seen from this figure, the energy spread of the beam is slightly increased by the absorbing foils because the reduction in energy produced by a given foil increases with decreasing energy for the deuterons. The influence of straggling and of the inhomogeneity of the absorbing foils is so small in proportion to the measured energy spread that it can be disregarded. The energy distribution depends upon the particular shims inserted in the gaps in the pole pieces of the cyclotron magnet (see [3] p. 13) and, each time when a change had been made in the arrangement of shims, a new determination of the energy distribution had, therefore, to be made.

3. Yield of fission as a function of deuteron energy.

The procedure used in the determination of the yield of fission products as a function of the deuteron energy was to irradiate two targets of uranium or thorium with deuterons with different energies but with the same beam current and to measure the ratio of the amounts of fission products collected. For a definite length of the time of irradiation, the decay curves obtained with different deuteron energies run very nearly parallel. In relative determ-

inations, it was thus sufficient to measure the activity of the fission products at some definite time after the end of the irradiation. In the actual experiments, this time was 90 minutes.

During the irradiation, which lasted for 40 minutes, the target holder was kept in constant rotation in order to eliminate the influence of the fluctuations in the beam current. The rate of rotation was about 1 revolution in 5 seconds, which was sufficient to ensure that each of the four sides of the target holder received the same number of deuterons. The two uranium targets used were "thin" targets made by evaporation of uranium on aluminium and measured in the way mentioned above. It was found that the uranium amount on the targets, which we shall refer to as uranium I and uranium II, were in the ratio I:II = 1:1,08. They were placed in the two positions in the target holder denoted by a_1 and a_2 in Fig. 1. In the position a_3 was placed an aluminium sheet without uranium and, facing all three targets, were placed the collecting foils b_1 , b_2 and b_3 . The activity of b_3 gave a control on the activity of the collecting foils produced by neutrons and by scattered deuterons; this activity was found to have a period of about 10 minutes due to the formation of ^{27}Mg ($^{27}\text{Al}(n, p)^{27}\text{Mg}$), but it was not strong and, 90 minutes after the end of the irradiation, it had completely disappeared. The activity of b_1 and b_2 at that time was then a direct measurement of the number of fission particles collected. In the position a_4 of the target holder was placed a copper sheet which was used as a current integrator to obtain a relative measurement of the total number of deuterons for comparison with other experiments; the amount of ^{64}Cu (period 13 hours) produced during the

irradiation is proportional to the total number of deuterons hitting the target.

Table I gives the data from one of the experiments.

Table I.

Target	Thickness of absorbing foil	Activity of collecting foil (90 minutes after the irradiation)
1. Uranium I	13,5 mg/cm ²	14,0 counts per min.
2. — II	20,25 —	6,4 — — —
3. Aluminium	6,75 —	0 — — —
4. Copper	0 —	{ no collecting foil; activity of the copper target itself in relative measure: 620

At first we consider the case of the target 2 which was hit by deuterons with the upper limit 8,4 MeV (Fig. 3 c); here the conditions are simple, because almost the whole fission effect can be ascribed to deuterons with energy near the upper limit. In fact, in other experiments with deuteron energies below 8 MeV a fission effect was found, which was small compared with that found with 8,4 MeV deuterons. This shows, firstly, that the effect due to fission by neutrons is small, and, secondly, that the effect found with the target 2 can be ascribed to deuterons in the energy interval 8,0 to 8,4 MeV. The fission effect for deuterons in this energy interval is overestimated, since part of the effect is due to deuterons with smaller energy; but the error is without influence on the results obtained for higher deuteron energies. From the energy distribution curve (Fig. 3 c) it is found that 38 % of the beam lies in the region between 8,0 and 8,4 MeV; the activity observed

was 6,4; now, if the whole beam had been lying in this region, then the effect would have been $6,4:0,38 = 17$.

We then consider the case of the target 1, which was hit by deuterons with the upper limit 8,75 MeV (Fig. 3 b). The effect 14,0 is at first corrected to $14,0 \cdot 1,08 = 15,1$ due to the difference in the thicknesses of the targets. In the same way as before, we can neglect the effect due to deuterons with energies smaller than 8,0 MeV and ascribe the whole effect to deuterons in the two regions 8,0—8,4 MeV and 8,4—8,75 MeV; from the energy distribution curve (Fig. 3 b) it is found that 31 % and 36 % of the beam lie in these two regions, respectively. The deuterons in the region 8,0—8,4 MeV must give the effect $17 \cdot 0,31 = 5,3$ and the deuterons in the other region 8,4—8,75 must then give the rest of the effect $15,1 - 5,3 = 9,8$; if now the whole beam had been lying in the last region, the effect would have been $9,8:0,36 = 27$.

The results of this experiment can be summarized as follows:

Mean deuteron energy	8,2 MeV	8,58 MeV
Relative fission yield	17	27

In another experiment, the thicknesses of the foils c_1 and c_2 were 0 and 6,75 mg per cm^2 , corresponding to the energy distribution curves in Fig. 2 b and Fig. 3 a, respectively. Before comparing with the experiment above we must apply a correction to the activities found in this experiment, because the activity of the copper target was now only 550, measured in the same units as in the former experiment. As before, we divide the effect into parts due to deuterons in smaller energy regions. The effect of the deuterons in the regions 8,0—8,4 MeV and 8,4—8,75 MeV may be cal-

culated from the figures given above and the energy distribution curves. In quite the same manner as just described, we get further figures corresponding to the regions 8,75—9,1 MeV and 9,1—9,5 MeV.

Table II gives the results of these two experiments.

Table II.

Mean deuteron energy	Relative fission yield of uranium
8,2 MeV	17
8,58 —	27
8,92 —	48
9,3 —	81

The stopping power of the targets was about 5 mm of air so that the energy loss of the deuterons in passing through the targets was negligible. The results are, therefore, an approximation to those which would be obtained by bombarding an infinitely thin target with a homogeneous beam.

In the experiments with thorium, thick targets of ThO_2 were used, but the method was the same. The results are given in Table III.

Table III.

Mean deuteron energy	Relative fission yield of thorium
7,6 MeV	7
8,12 —	12
8,55 —	52
3,92 —	85
9,3 —	175

It may be stressed that the figures in Tables II and III which refer to "thin" and "thick" targets, respectively, give no information about the ratio between the fission yields in uranium and thorium under similar conditions. Since, however, the fission products are only escaping from a thin upper layer of the thick targets and the deuterons lose only a small part of their energy in passing through this layer, the results are in relative measure also for thorium approximately the same which would be obtained by using a thin target and a homogeneous beam.

4. Comparison between uranium and thorium.

In some of the experiments with thorium, thick targets of U_3O_8 were placed on the target holder together with the targets of ThO_2 ; in this way a comparison between the yield of fission from uranium and thorium was obtained. In Fig. 4 are shown the decay curves of the collecting foils from such an experiment; the data of this experiment are given in Table IV.

Table IV.

Target	Max. deuteron energy	Corresp. curve in fig. 4
1. ThO_2	9,5 MeV	1
2. U_3O_8	9,5 —	2
3. ThO_2	8,75 —	3
4. Copper	9,5 —	4

Curve 1 gives the activity of the collecting foil facing the target 1 (ThO_2), corrected for the neutron effect given by curve 4. The vertical line in the figure shows the time at which the relative activities of the fission products were

measured; this was chosen so that the correction due to neutrons could be omitted. The mean value of the yield from thorium relative to that from uranium of two such experiments, which were in very good agreement, was

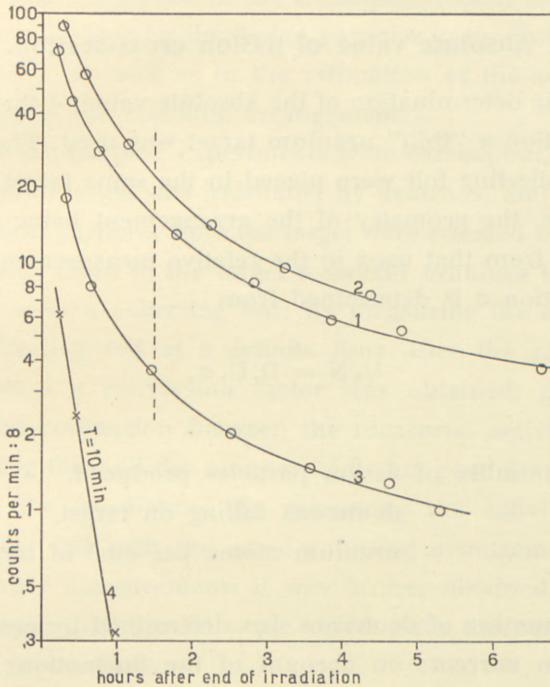


Fig. 4.

75 %. In estimating the ratio for the fission cross-section in the two elements it should be taken into account that the stopping power in the uranium layer for the fission fragments escaping from it will be somewhat larger in proportion to the amount of heavy element present. In fact, due to the greater relative number of oxygen atoms in U_3O_8 than in ThO_2 , the stopping power per atom of heavy element for the fission fragments must be expected

to be about 10 % greater in the uranium than in the thorium target. We may conclude, therefore, the fission cross-section of thorium for 9 MeV deuterons to be about 70 % of that of uranium.

5. Absolute value of fission cross-section.

For the determination of the absolute value of the fission cross-section a "thin" uranium target was used. The target and a collecting foil were placed in the same target holder as before, the geometry of the arrangement being slightly different from that used in the relative measurements. The cross-section σ is determined from

$$\frac{1}{2}N = D.U.\sigma, \quad (1)$$

where

- N = number of fission particles produced,
- D = — - deuterons falling on target,
- U = — - uranium atoms per cm^2 of target.

The number of deuterons was determined by measuring the beam current; on account of the fluctuations in the beam it was somewhat difficult to fix a mean value of the current. The uncertainty in the determination of the number of deuterons was estimated to be about 25 %.

A preliminary determination of N was obtained by assuming that the number of β -particles emitted from the fission products within some 10 hours after the end of the irradiation was 3 per atom. From a determination of the number of β -particles emitted from the collecting foil in this time interval and from the geometry of the arrangement, N could thus be determined.

In this way we got for the deuteron energy 9 MeV the value $\sigma = 5.10^{-27}$ cm² mentioned in our first note (4). In continuing the work we have, however, attempted to obtain a more reliable value for the cross-section by avoiding the uncertainties involved in the assumption about the number of β -rays emitted by the fission particles in the time interval concerned, as well as in the estimation of the solid angle entering in the counting arrangement.

For this purpose experiments were carried out, in which a uranium target was irradiated by neutrons, and in which the fission particles from the target were counted by a linear amplifier. Close to the target a second uranium target was placed with a collecting foil; by measuring the activity of the collecting foil at a definite time after the end of the irradiation a conversion factor was obtained, giving the required connection between the measured activity of the collecting foil and the number of fission particles produced during the irradiation. By measuring the activity of the collecting foil with the same counting arrangement as in the earlier measurements it was further obtained, that the knowledge of the solid angle underneath the counter was made unnecessary.

The apparatus used in the experiments with neutrons is shown in Fig. 5. In a hemispherical ionization chamber with radius 2 cm a target T_1 consisting of a brass disc with a thin layer of uranium hydroxide $((\text{NH}_4)_2\text{U}_2\text{O}_7)$ or of metallic uranium was placed at the centre; the brass disc was connected to the grid of the first valve of a linear amplifier. The hemisphere forming the other electrode of the ionization chamber was kept at a potential of 500 volts. The ionization chamber was surrounded by a cylindrical screen to shield it against high frequency disturbances.

Inside the screen the second uranium target T_2 was placed with a collecting foil, which in this case was a thin lead foil (48 mg/cm²) placed in direct contact with the target. As a consequence of the geometry of the apparatus the number of fission particles received by the collecting foil

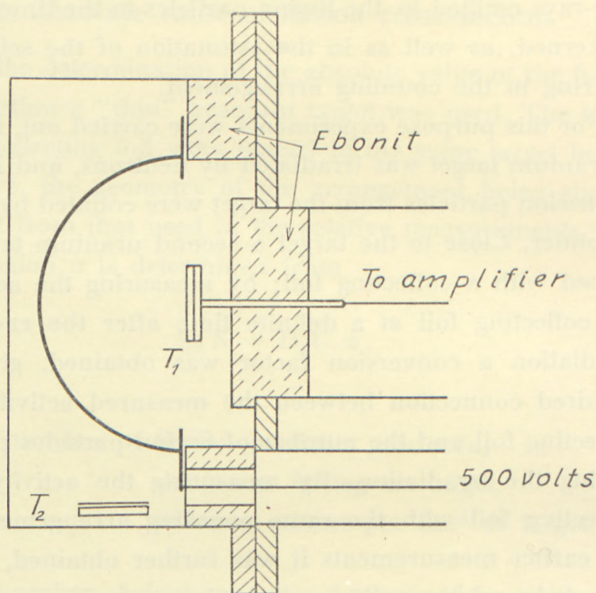


Fig. 5.

will be the same as the number of particles counted by the amplifier, when referred to the same amount of uranium in the two targets.

The ionization chamber was placed at a distance of about 30 cm from the neutron source, a beryllium target bombarded by deuterons in the cyclotron. The time of irradiation was 40 minutes as in the earlier experiments. In the present experiments, where the activity of the collecting foil was much smaller than in the experiments with deuterons, the activity of the collecting foil was measured

20 minutes after the end of the irradiation. In a particular experiment the following data were obtained: the amounts of uranium were 0,92 and 0,20 mg on T_2 and T_1 respectively. During the irradiation period of 40 minutes 2270 fission particles were counted by the amplifier; the activity of the collecting foil, when measured 20 minutes after the end of the irradiation, was 1,3 in the same unit as used previously. This gives for the ratio (number of fission particles):(activity of collecting foil) $f = \frac{2270}{1,3} \cdot \frac{0,92}{0,20} = 8000$. In four experiments with different targets the values obtained for f were $f = 8000, 5400, 8600,$ and 5300 with a mean value of $f = 7000$.

We now return to the determination of the fission cross-section for deuteron impact. In one of our experiments the beam current was 0,045 microamp. during 40 minutes and for D we get thus the value $D = 6,8 \cdot 10^{14}$. Further, U was $1,8 \cdot 10^{18}$ atoms per cm^2 , and the activity of the collecting foil (corrected for effect of neutrons and scattered deuterons), when measured 20 minutes after the end of the irradiation, was found to be 170. Thus, $170 \cdot 7000$ fission particles were collected on the foil, and the solid angle of the collecting foil as seen from the target being $\frac{1}{50} \cdot 4 \pi$, the total number of fission particles produced was $N = 170 \cdot 7000 \cdot 50 = 6,0 \cdot 10^7$. From these figures and the equation (1) we get therefore $\sigma = 2,4 \cdot 10^{-26} \text{ cm}^2$ as a mean value of the cross-section over the deuteron energy interval concerned.

In the experiments the deuteron energy spectrum was that shown in Fig. 2 b. From the energy distribution curve and from the variation of σ with deuteron energy determined above, the cross section $\sigma_{9,0}$ for the deuteron energy 9,0 MeV

can be determined by a simple numerical calculation. The result is

$$\sigma_{9,0} = 2,2 \cdot 10^{-26} \text{ cm}^2 \quad (2)$$

Besides this experiment we have performed two experiments with lower deuteron energies corresponding to the energy spectra in Figs. 3 a and 3 b. These measurements, treated in the same way as above, gave values for $\sigma_{9,0}$ of about $3 \cdot 10^{-26} \text{ cm}^2$. On account of the much larger and more uncertain corrections due to the influence of the lower part of the energy spectrum these values are, however, less reliable than the value (2). From a consideration of the probable errors in the various steps in the calculations we estimate the uncertainty of this value to be about 50 %.

6. Comparison of the results with the theory.

The theory of nuclear fission is based on the assumption that fission in competition with other nuclear transformations takes place in a highly excited intermediate compound nucleus. In neutron induced fission, this compound nucleus is just formed by the temporary capture of the neutron by the original nucleus. In deuteron induced fission, however, two types of processes of forming a compound nucleus must be taken into account. In the one process, the intermediate state is formed by the fusion of the whole deuteron with the original nucleus while, in the other process, the deuteron breaks up during the impact with the result that the proton escapes and the neutron is captured in the nucleus. In the former process, the excitation of the compound system will for heavy nuclei be much higher than the neutron binding energy as well as the

critical fission energy. In the latter process, however, the excitation will, on the average, be lower than the neutron binding energy and, therefore, for almost all nuclei be insufficient to produce fission. As pointed out by BOHR and WHEELER [2], the intermediate capture of the whole deuteron by the original nucleus will, thus, be necessary in order to obtain a considerable fission probability by deuteron impact.

For the high excitation concerned, the probability of fission of the compound nucleus will be almost independent of its energy and the fission output will, therefore, vary with varying deuteron energy in the same way as the cross-section σ_0 for the penetration of the deuteron into the original nucleus. This cross-section will be approximately given by the formula (see [2], p. 449)

$$\sigma_0 = \pi R^2 e^{-P}. \quad (3)$$

R is the nuclear radius, and P is given by

$$P = \frac{4 Z \epsilon^2}{\hbar \cdot v} (\text{arc cos } x^{1/2} - x^{1/2} (1-x)^{1/2}), \quad x = \frac{ER}{Z\epsilon^2} \quad (4)$$

where v and E denote the velocity and energy of the deuteron, respectively, and Z is the atomic number of the nucleus concerned.

The value of R to be introduced into (3) and (4) is rather uncertain; but, choosing, in agreement with the value given in [2] for uranium

$$R = 1,45 \sqrt[3]{A} \cdot 10^{-13} \text{ cm} \quad (5)$$

where A is the atomic weight, we obtain for thorium ($Z = 90$, $A = 232$) and for the abundant uranium isotope ($Z = 92$,

$A = 238$) the two curves in Fig. 6 which, at any rate as regards the order of magnitude, agree with the fission cross-section obtained for both elements by deuterons of 9 MeV. For a closer comparison of the theory with the

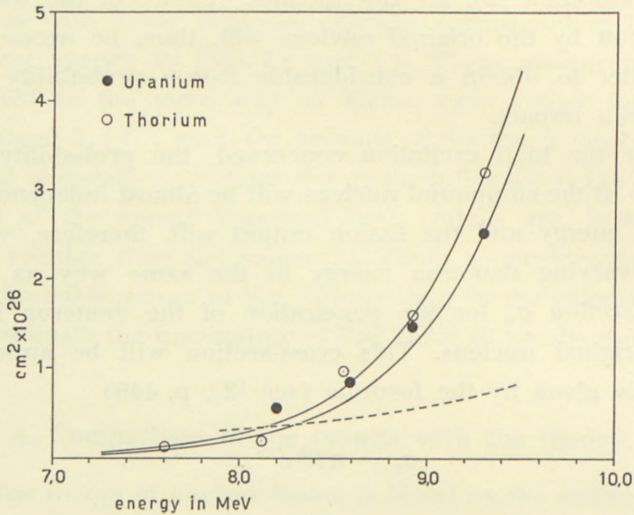


Fig. 6. The points plotted are the experimental values for the fission cross-section multiplied by a suitable factor (see text).

experiments, we have by the full and open circles indicated the experimental values from Tables II and III, respectively, the scales being in each case chosen in such a way that the point for the highest energy falls on the corresponding curve. We see that the variation of the fission output with energy for both elements approximately corresponds to formula (3). For thorium, the agreement is almost perfect; for uranium, however, the deviation from the curve is for smaller energies rather outside the limits of experimental error.

A more sensitive test of the theory can, however, be obtained from the consideration of the ratio between the

fission outputs in thorium and uranium which, for 9 MeV deuteron energy, was found to be 0,7 with an estimated uncertainty of few per cent. As regards the theoretical expectation for this ratio, it was pointed out by BOHR and WHEELER [2] that the excitation of the compound nucleus, formed by the fusion of the deuteron with the heavy nucleus, is not only sufficient for fission to occur in competition with neutron escape, but there will even be a great probability that fission may take place after the escape of a neutron from the compound nucleus. Such successive transformations have been discussed more closely by BOHR [7] who estimated the probability of fission of the compound nucleus in both steps for uranium to be nearly 1 and for thorium to be about 0,8.

In our preliminary note [4], the experimental ratio between the outputs in thorium and uranium was regarded as a confirmation of the value estimated from the theory. From the closer discussion of the experiments it appeared, however, that the difference of the cross-section for the formation of the compound system cannot be neglected in such a comparison. In fact, for 9 MeV deuteron energy, this cross-section must, as seen from the curves in Fig. 6, be expected to be about 25 % greater for thorium than for uranium. The theoretically estimated ratio between the fission outputs in thorium and uranium should, therefore, instead of 0,8 be $1,25 \cdot 0,8 = 1,0$, while the value found experimentally was, as mentioned above, only 0,7.

As pointed out to us by Professor BOHR, this rather great discrepancy between theory and experiment seems to indicate that a part of the fission effect in uranium, in contrast to thorium, has to be ascribed to a process in which the deuteron breaks up during the impact and only the

neutron is captured by the nucleus. In fact, if such processes, in spite of the low average excitation of the compound nucleus, contribute materially to the fission, we should expect a larger effect in uranium than in thorium. Thus, as known from experiments on neutron induced fission, the critical fission energy is, in thorium, almost 2 MeV higher than the neutron binding energy in the compound nucleus while, for the abundant uranium isotope (238), this difference is smaller than 1 MeV. Moreover, in the energy region concerned, where the cross-section of the process is less than 1 % of the geometrical nuclear cross-section, it is also not excluded that the presence of the lighter rare uranium isotope (235), for which the critical fission energy is still lower, may contribute sensibly to the fission effects observed.

Indications for the correctness of such views are also obtained from a consideration of the fission effects in thorium and uranium for lower energies. Thus, while the variation of the output in thorium agrees closely with the theoretical curve in Fig. 6, the values for uranium show, as already mentioned, for energies about 8 MeV rather large deviations from the curve, just as would be expected if a part of the effect in uranium was due to processes more probable for lower energies. In Fig. 6 is, by the dotted line, roughly indicated the contribution of such processes sufficient to ensure an agreement with the theoretical expectations regarding the variation with energy of the fission effect in uranium for processes initiated by deuteron capture as well as the ratio between these effects in uranium and thorium.

The experimental material so far obtained is, however, not sufficient to settle the last questions. For this purpose,

experiments over a wider energy region will be necessary. It is planned to take up such experiments as soon as the reconstruction of the cyclotron with the object of obtaining higher energies and currents is completed.

The present investigation has been carried out at the Institute of Theoretical Physics, Copenhagen, and we wish to thank Professor NIELS BOHR for his continual interest in this work and for many helpful discussions on the subject. Our thanks are also due Mr. S. HØFFER JENSEN for valuable assistance in the experiments.

REFERENCES

- [1] D. H. T. GANT, *Nature* **144**, 707 (1939).
- [2] N. BOHR and J. A. WHEELER, *Phys. Rev.* **56**, 426 (1939).
- [3] J. C. JACOBSEN, *D. Kgl. Danske Vidensk. Selskab, Math.-fys. Medd.* **XIX**, 2 (1941).
- [4] J. C. JACOBSEN and N. O. LASSEN, *Phys. Rev.* **58**, 867 (1940).
- [5] R. S. KRISHNAN and T. E. BANKS, *Nature* **145**, 860 (1940).
- [6] M. S. LIVINGSTON and H. A. BETHE, *Rev. Mod. Phys.* **9**, 268 (1937).
- [7] N. BOHR, *Phys. Rev.* **58**, 864 (1940).

Det Kgl. Danske Videnskabernes Selskab.

Mathematisk-fysiske Meddelelser. **XIX**, 7.

DIE LIBRATIONSELLIPSEN UM
DIE DREIECKSLIBRATIONSPUNKTE
IM ALLGEMEINEN DREIKÖRPER-
PROBLEM

VON

PEDER PEDERSEN



KØBENHAVN

I KOMMISSION HOS EJNAR MUNKSGAARD

1941

Printed in Denmark.
Bianco Lunos Bogtrykkeri A/S.

1. Einleitung.

Um die LAGRANGESchen Dreieckslibrationspunkte im allgemeinen Dreikörperproblem gibt es infinitesimale periodische Librationen, wenn die Bewegung auf ein rotierendes Koordinatensystem bezogen wird, das den Anfangspunkt im gemeinsamen Schwerpunkt der drei Massen hat, und das mit einer Winkelgeschwindigkeit rotiert, die dieselbe ist wie die Winkelgeschwindigkeit der Massen in der absoluten Bewegung. Von solchen Librationen gibt es zwei Sorten. Die eine Sorte kann als LAGRANGESche Librationen bezeichnet werden, weil sie beim Übergang von dem rotierenden Koordinatensystem auf das feste mit den von LAGRANGE gefundenen Lösungen identisch werden, in Übereinstimmung damit, dass die Umlaufszeit der Librationen dieselbe ist wie die Umlaufszeit des Systems in der absoluten Bewegung. Bei der zweiten Sorte Librationen stimmt die Umlaufszeit nicht mit der Umlaufszeit in der absoluten Bewegung überein; diese Librationen geben deshalb in der absoluten Bewegung keine periodischen Bahnen. Die Librationen dieser zweiten Sorte wollen wir als nicht-LAGRANGESche Librationen bezeichnen.

In einer früheren Abhandlung¹ habe ich die erwähnten Librationen behandelt. Es wurde vorausgesetzt, dass die Abstände der Massen von den Librationspunkten unend-

¹ Die Erweiterung der LAGRANGESchen Dreieckslösungen im allgemeinen Dreikörperproblem [Astr. Nachrichten 5043 und Publikation der Kopenhagener Sternwarte Nr. 35 (1920)].

lich klein 1. Ordnung waren. In den Reihenentwicklungen wurden keine Glieder mitgenommen, die höher als 1. Ordnung waren. Es ging aus der Untersuchung hervor, dass — während die LAGRANGESchen Librationen für alle Werte der Massenverhältnisse existieren — die nicht-LAGRANGESchen Librationen nur dann existieren, wenn die Massenverhältnisse die Gleichung

$$\mu_1 \mu_2 + \mu_2 \mu_3 + \mu_1 \mu_3 \leq \frac{1}{27} \quad (1)$$

befriedigen, wo μ_1 , μ_2 und μ_3 die Grösse der drei Massen angeben, in der Einheit der Gesamtmasse ausgedrückt. Diese Bedingung sagt aus, dass der gemeinsame Schwerpunkt der drei Massen ausserhalb eines Kreises oder auf demselben liegen muss, dessen Zentrum im Zentrum des gleichseitigen Dreiecks (des Librationsdreiecks) liegt und einen Radius besitzt, der nur wenig kleiner ist als der Radius des dem Dreieck umschriebenen Kreises. Die Gebiete ausserhalb des Kreises, wo der Schwerpunkt liegen kann, wenn die Gleichung (1) befriedigt werden soll, ist deshalb auf drei ganz kleine Eck-Gebiete beschränkt (Abb. 1). Die nicht-LAGRANGESchen Librationen existieren deshalb nur, wenn eine der drei Massen dominierend ist. Wenn eine der drei Massen gleich Null ist, geht die Bedingung (1) auf die im problème restreint bekannte Bedingung

$$\mu_1 \mu_2 \leq \frac{1}{27} \quad (2)$$

über.

Die in der oben zitierten Abhandlung mitgeteilte Untersuchung umfasste nur die Berechnung der Umlaufszeit in den infinitesimalen periodischen Bahnen, wogegen keine Untersuchung der Form der Librationsellipsen und ihrer Orientierung ausgeführt wurde. Veranlasst durch eine An-

frage seitens Professor E. STRÖMGREN in Bezug auf die Orientierung der Achsen habe ich jetzt die Untersuchung wieder aufgenommen. Im folgenden werde ich die erhaltenen Resultate über die Verhältnisse der Achsen und ihre Orientierung mitteilen. Der Vollständigkeit wegen sind die

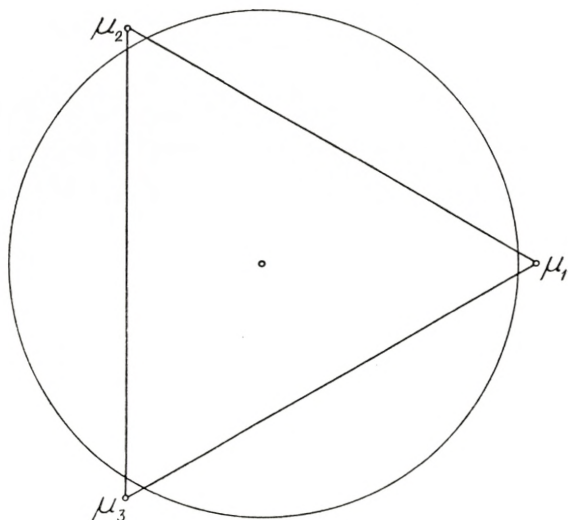


Abb. 1.

LAGRANGESCHEN Librationsellipsen auch behandelt worden. Als Grundlage der Untersuchung dienen die in der oben zitierten Abhandlung gefundenen Differentialgleichungen.

2. Die Differentialgleichungen der Bewegung.

Wir lassen die drei Massen m_1 , m_2 und m_3 zur Zeit $t = 0$ sich in den Eckpunkten eines gleichseitigen Dreiecks mit der Seite ρ befinden und geben den Anfangsgeschwindigkeiten solche Werte, dass die drei Massen für immer in den Eckpunkten eines gleichseitigen Dreiecks mit der Seite ρ verbleiben. Wir wählen ein Koordinatensystem, das den Anfangspunkt im gemeinsamen Schwerpunkt der drei

Massen hat, und das mit einer Winkelgeschwindigkeit n rotiert, die in Richtung und Grösse mit der Winkelgeschwindigkeit übereinstimmt, womit das gleichseitige Dreieck rotiert. In einem solchen Koordinatensystem werden die drei Massen feste Stellungen einnehmen. Wenn die drei Massen aber aus den Dreieckspunkten in andere Punkte in der unmittelbaren Umgebung der Dreieckspunkte verschoben werden, werden sie kleine Akzelerationen erfahren. Wenn die Koordinaten der Dreieckspunkte in dem mit der Winkelgeschwindigkeit n rotierenden Koordinatensystem x, y mit (p_1, q_1) , (p_2, q_2) , (p_3, q_3) bezeichnet werden und die Koordinaten der verschobenen Massen mit (x_1, y_1) , (x_2, y_2) , (x_3, y_3) , so haben wir die Gleichungen

$$\left. \begin{aligned} m_1 x_1 + m_2 x_2 + m_3 x_3 &= 0, \\ m_1 y_1 + m_2 y_2 + m_3 y_3 &= 0, \\ M - n^2 \rho^3 &= 0, \end{aligned} \right\} \quad (3)$$

wo

$$M = m_1 + m_2 + m_3. \quad (4)$$

Wir haben speziell:

$$\left. \begin{aligned} m_1 p_1 + m_2 p_2 + m_3 p_3 &= 0, \\ m_1 q_1 + m_2 q_2 + m_3 q_3 &= 0. \end{aligned} \right\} \quad (5)$$

Wenn die Koordinaten der drei Massen im Verhältnis zu den Dreieckspunkten mit (ξ_1, η_1) , (ξ_2, η_2) , (ξ_3, η_3) bezeichnet werden, haben wir:

$$\left. \begin{aligned} x_1 &= p_1 + \xi_1, & x_2 &= p_2 + \xi_2, & x_3 &= p_3 + \xi_3, \\ y_1 &= q_1 + \eta_1, & y_2 &= q_2 + \eta_2, & y_3 &= q_3 + \eta_3. \end{aligned} \right\} \quad (6)$$

Wegen (3) und (5) erhalten wir dann:

$$\left. \begin{aligned} m_1 \xi_1 + m_2 \xi_2 + m_3 \xi_3 &= 0, \\ m_1 \eta_1 + m_2 \eta_2 + m_3 \eta_3 &= 0. \end{aligned} \right\} \quad (7)$$

Wenn die Entfernungen der Massen von den Dreiecks-
punkten als unendlich kleine Grössen 1. Ordnung betrach-
tet werden, und wenn wir in den Reihenentwicklungen
nur Glieder bis zur 1. Ordnung inkl. mitnehmen und die

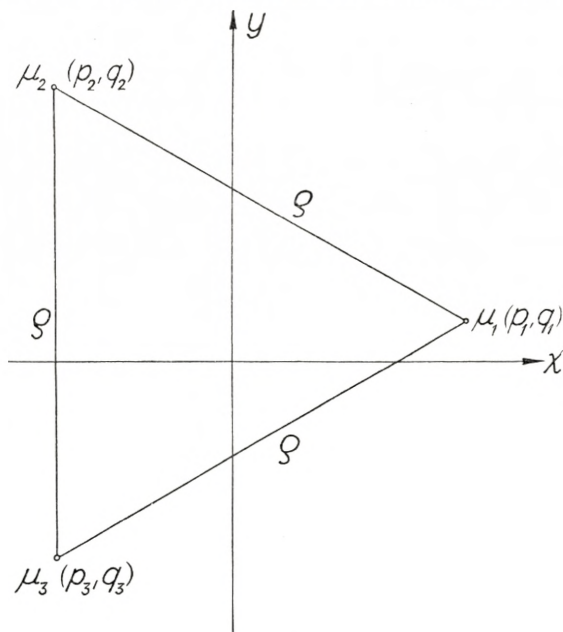


Abb. 2.

Ordinatenachse des Koordinatensystems einer Dreiecksseite
parallel legen (Abb. 2), erhalten wir die folgenden Differen-
tialgleichungen für die Bewegung der Massen m_1 und m_2 :

$$\left. \begin{aligned} \xi_1'' - 2n\eta_1' - \frac{9}{4}n^2\xi_1 - \frac{3}{4}\sqrt{3}(1-2\mu_2)n^2\eta_1 - \frac{3}{2}\sqrt{3}\mu_2n^2\eta_2 &= 0, \\ \eta_1'' + 2n\xi_1' - \frac{3}{4}n^2\eta_1 - \frac{3}{4}\sqrt{3}(1-2\mu_2)n^2\xi_1 - \frac{3}{2}\sqrt{3}\mu_2n^2\xi_2 &= 0, \\ \xi_2'' - 2n\eta_2' - \frac{9}{4}\mu_1n^2\xi_2 + \frac{9}{4}\mu_1n^2\xi_1 + \frac{3}{4}\sqrt{3}\mu_1n^2\eta_2 - \frac{3}{4}\sqrt{3}\mu_1n^2\eta_1 &= 0, \\ \eta_2'' + 2n\xi_2' - \left(3 - \frac{9}{4}\mu_1\right)n^2\eta_2 - \frac{9}{4}\mu_1n^2\eta_1 + \frac{3}{4}\sqrt{3}\mu_1n^2\xi_2 - \frac{3}{4}\sqrt{3}\mu_1n^2\xi_1 &= 0. \end{aligned} \right\} (8)$$

3. Lösung der Differentialgleichungen.

Wir suchen jetzt Lösungen der Differentialgleichungen (8) von der Form:

$$\left. \begin{aligned} \xi_1 &= A_1 e^{\lambda t}, & \xi_2 &= A_2 e^{\lambda t}, \\ \eta_1 &= B_1 e^{\lambda t}, & \eta_2 &= B_2 e^{\lambda t}. \end{aligned} \right\} \quad (9)$$

Indem $\lambda = nz$ gesetzt wird, erhalten wir aus (8) und (9) die folgenden vier Bedingungsgleichungen:

$$\left. \begin{aligned} \left[z^2 - \frac{9}{4} \right] A_1 & - \left[2z + \frac{3}{4} \sqrt{3} (1 - 2\mu_2) \right] B_1 & - \frac{3}{2} \sqrt{3} \mu_2 B_2 & = 0, \\ \frac{9}{4} \mu_1 A_1 & + \left[z^2 - \frac{9}{4} \mu_1 \right] A_2 & - \frac{3}{4} \sqrt{3} \mu_1 B_1 - \left[2z - \frac{3}{4} \sqrt{3} \mu_1 \right] B_2 & = 0, \\ \left[2z - \frac{3}{4} \sqrt{3} (1 - 2\mu_2) \right] A_1 - \frac{3}{2} \sqrt{3} \mu_2 A_2 + \left[z^2 - \frac{3}{4} \right] B_1 & & & = 0, \\ -\frac{3}{4} \sqrt{3} \mu_1 A_1 & + \left[2z + \frac{3}{4} \sqrt{3} \mu_1 \right] A_2 & - \frac{9}{4} \mu_1 B_1 + \left[z^2 - 3 + \frac{9}{4} \mu_1 \right] B_2 & = 0. \end{aligned} \right\} \quad (10)$$

Wenn es in diesem System linearer, homogener Gleichungen Wurzeln geben soll, die von Null verschieden sind, muss die Determinante des Systems gleich Null sein.

Wenn die Determinante berechnet und gleich Null gesetzt wird, erhalten wir:

$$\left. \begin{aligned} z^8 + 2z^6 + \left[1 + \frac{27}{4} (\mu_1 + \mu_2 - \mu_1^2 - \mu_1 \mu_2 - \mu_2^2) \right] z^4 & \\ + \frac{27}{4} (\mu_1 + \mu_2 - \mu_1^2 - \mu_1 \mu_2 - \mu_2^2) z^2 = 0, & \end{aligned} \right\} \quad (11)$$

eine Gleichung, die zur Bestimmung von z dienen kann.

Der Kürze wegen schreiben wir:

$$s = \frac{27}{4} (\mu_1 + \mu_2 - \mu_1^2 - \mu_1 \mu_2 - \mu_2^2) = \frac{27}{4} (\mu_1 \mu_2 + \mu_2 \mu_3 + \mu_1 \mu_3), \quad (12)$$

wonach Gleichung (11) folgender Weise geschrieben werden kann:

$$z^2 (z^2 + 1) (z^4 + z^2 + s) = 0, \quad (13)$$

eine Gleichung, die in die folgenden drei Gleichungen gespalten werden kann:

$$z^2 = 0, \quad (14)$$

$$z^2 + 1 = 0, \quad (15)$$

$$z^4 + z^2 + s = 0. \quad (16)$$

Die erste dieser Gleichungen gibt die Doppelwurzel $z = 0$, was einer periodischen Lösung mit unendlich langer Periode entspricht. Diese Lösung bedeutet, dass das Massendreieck sich um den Schwerpunkt mit unendlich kleiner Winkelgeschwindigkeit dreht.

Die zweite Gleichung gibt das rein imaginäre Wurzelpaar $z = \pm i$, d. h. $\lambda = \pm in$, also eine periodische Lösung mit derselben Umlaufszeit wie der Umlaufszeit des Librationsdreiecks in der absoluten Bewegung. Diese Lösung stellt die in der Einleitung besprochenen LAGRANGESCHEN Librationen dar.

Aus der dritten Gleichung erhalten wir:

$$z^2 = -\frac{1}{2} \pm \sqrt{\frac{1}{4} - s}. \quad (17)$$

Die Bedingung dafür, dass diese Gleichung rein imaginäre Wurzeln geben soll, ist deshalb

$$s \leq \frac{1}{4} \quad (18)$$

oder

$$\mu_1 \mu_2 + \mu_2 \mu_3 + \mu_1 \mu_3 \leq \frac{1}{27}. \quad (19)$$

Wenn diese Bedingung erfüllt ist, werden die zwei Werte von z^2 beide negativ. Jeder dieser zwei Werte von z^2 entspricht einer periodischen Lösung, für das Plus-Zeichen in der Quadratwurzel in (17) eine Klasse langperiodischer, für das Minus-Zeichen eine Klasse kurzperiodischer Lösungen. Diese zwei Klassen periodischer Lösungen stellen

die in der Einleitung besprochenen nicht-LAGRANGESCHEN Librationen dar.

Wir bezeichnen mit σ eine positive Zahl, so dass

$$z = \pm i\sigma \quad (20)$$

ein Wurzelpaar der Gleichung (13) bedeutet. Wir bezeichnen ferner mit A_1, A_2, B_1, B_2 die zu $z = i\sigma$ gehörenden Integrationskonstanten, und mit $A_{-1}, A_{-2}, B_{-1}, B_{-2}$ die zu $z = -i\sigma$ gehörenden. Periodische Lösungen der Gleichungen (8) können dann in der folgenden Form geschrieben werden:

$$\left. \begin{aligned} \xi_1 &= A_1 e^{i\sigma nt} + A_{-1} e^{-i\sigma nt}, \\ \eta_1 &= B_1 e^{i\sigma nt} + B_{-1} e^{-i\sigma nt}, \end{aligned} \right\} \quad (21)$$

$$\left. \begin{aligned} \xi_2 &= A_2 e^{i\sigma nt} + A_{-2} e^{-i\sigma nt}, \\ \eta_2 &= B_2 e^{i\sigma nt} + B_{-2} e^{-i\sigma nt}. \end{aligned} \right\} \quad (22)$$

Die Gleichungen (21) und (22) können umgeformt werden:

$$\left. \begin{aligned} \xi_1 &= a_1 \cos \sigma nt + a_{-1} \sin \sigma nt, \\ \eta_1 &= b_1 \cos \sigma nt + b_{-1} \sin \sigma nt. \end{aligned} \right\} \quad (23)$$

$$\left. \begin{aligned} \xi_2 &= a_2 \cos \sigma nt + a_{-2} \sin \sigma nt, \\ \eta_2 &= b_2 \cos \sigma nt + b_{-2} \sin \sigma nt, \end{aligned} \right\} \quad (24)$$

wo

$$\left. \begin{aligned} a_1 &= A_1 + A_{-1}, & a_{-1} &= iA_1 - iA_{-1}, \\ b_1 &= B_1 + B_{-1}, & b_{-1} &= iB_1 - iB_{-1}, \\ a_2 &= A_2 + A_{-2}, & a_{-2} &= iA_2 - iA_{-2}, \\ b_2 &= B_2 + B_{-2}, & b_{-2} &= iB_2 - iB_{-2}. \end{aligned} \right\} \quad (25)$$

Wir kehren jetzt zu den Gleichungen (10) zurück. Aus den drei ersten Gleichungen eliminieren wir A_2 und B_2 .

Wir erhalten dann eine Gleichung zwischen A_1 und B_1 :

$$\left. \begin{aligned} & \frac{3}{4} \sqrt{3} [(1 - \mu_1 - 2\mu_2) z^2 - 2\sqrt{3} (1 - \mu_1) z] A_1 \\ & - \left[z^4 + \left(\frac{13}{4} - \frac{9}{4} \mu_1 \right) z^2 + \frac{3}{2} \sqrt{3} (1 - \mu_1 - 2\mu_2) z \right] B_1 = 0, \end{aligned} \right\} (26)$$

In die Gleichung (26) wird zuerst $z = i\sigma$, dann $z = -i\sigma$ eingesetzt, indem im zweiten Fall A_1 und B_1 durch A_{-1} und B_{-1} ersetzt werden. Wir erhalten dann:

$$\left. \begin{aligned} & \frac{3}{4} \sqrt{3} [-(1 - \mu_1 - 2\mu_2) \sigma^2 - i2\sqrt{3} (1 - \mu_1) \sigma] A_1 \\ & - \left[\sigma^4 - \left(\frac{13}{4} - \frac{9}{4} \mu_1 \right) \sigma^2 + i\frac{3}{2} \sqrt{3} (1 - \mu_1 - 2\mu_2) \sigma \right] B_1 = 0, \end{aligned} \right\} (27)$$

$$\left. \begin{aligned} & \frac{3}{4} \sqrt{3} [-(1 - \mu_1 - 2\mu_2) \sigma^2 + i2\sqrt{3} (1 - \mu_1) \sigma] A_{-1} \\ & - \left[\sigma^4 - \left(\frac{13}{4} - \frac{9}{4} \mu_1 \right) \sigma^2 - i\frac{3}{2} \sqrt{3} (1 - \mu_1 - 2\mu_2) \sigma \right] B_{-1} = 0, \end{aligned} \right\} (28)$$

Aus den Gleichungen (27) und (28) und den vier ersten der Gleichungen (25) ergeben sich zwei reelle Gleichungen:

$$\left. \begin{aligned} & \frac{3}{4} \sqrt{3} (1 - \mu_1 - 2\mu_2) \sigma^2 a_1 + \frac{9}{2} (1 - \mu_1) \sigma a_{-1} \\ & + \left(\sigma^4 - \left(\frac{13}{4} - \frac{9}{4} \mu_1 \right) \sigma^2 \right) b_1 + \frac{3}{2} \sqrt{3} (1 - \mu_1 - 2\mu_2) \sigma b_{-1} = 0, \end{aligned} \right\} (29)$$

$$\left. \begin{aligned} & \frac{3}{4} \sqrt{3} (1 - \mu_1 - 2\mu_2) \sigma^2 a_{-1} - \frac{9}{2} (1 - \mu_1) \sigma a_1 \\ & + \left(\sigma^4 - \left(\frac{13}{4} - \frac{9}{4} \mu_1 \right) \sigma^2 \right) b_{-1} - \frac{3}{2} \sqrt{3} (1 - \mu_1 - 2\mu_2) \sigma b_1 = 0, \end{aligned} \right\} (30)$$

Die Gleichungen (29) und (30) dienen zur Bestimmung von b_1 und b_{-1} als Funktionen von a_1 und a_{-1} , die als arbiträre Konstanten betrachtet werden können.

Aus der ersten und der dritten Gleichung (10) erhalten wir ferner:

$$\left. \begin{aligned} \frac{3}{2} \sqrt{3} \mu_2 A_2 &= \left[2z - \frac{3}{4} \sqrt{3} (1 - 2\mu_2) \right] A_1 + \left[z^2 - \frac{3}{4} \right] B_1, \\ \frac{3}{2} \sqrt{3} \mu_2 B_2 &= \left[z^2 - \frac{9}{4} \right] A_1 - \left[2z + \frac{3}{4} \sqrt{3} (1 - 2\mu_2) \right] B_1. \end{aligned} \right\} (31)$$

In diese Gleichungen setzen wir zuerst $z = i\sigma$ und dann $z = -i\sigma$ ein. Im zweiten Falle werden A_1, B_1, A_2, B_2 durch $A_{-1}, B_{-1}, A_{-2}, B_{-2}$ ersetzt. Es ergeben sich dann die vier Gleichungen:

$$\left. \begin{aligned} \frac{3}{2} \sqrt{3} \mu_2 A_2 &= \left[i2\sigma - \frac{3}{4} \sqrt{3} (1 - 2\mu_2) \right] A_1 - \left[\sigma^2 + \frac{3}{4} \right] B_1, \\ \frac{3}{2} \sqrt{3} \mu_2 A_{-2} &= \left[-i2\sigma - \frac{3}{4} \sqrt{3} (1 - 2\mu_2) \right] A_{-1} - \left[\sigma^2 + \frac{3}{4} \right] B_{-1}, \\ \frac{3}{2} \sqrt{3} \mu_2 B_2 &= - \left[\sigma^2 + \frac{9}{4} \right] A_1 - \left[i2\sigma + \frac{3}{4} \sqrt{3} (1 - 2\mu_2) \right] B_1, \\ \frac{3}{2} \sqrt{3} \mu_2 B_{-2} &= - \left[\sigma^2 + \frac{9}{4} \right] A_{-1} - \left[-i2\sigma + \frac{3}{4} \sqrt{3} (1 - 2\mu_2) \right] B_{-1}. \end{aligned} \right\} (32)$$

Aus den zwei ersten Gleichungen (32) und den sechs ersten Gleichungen (25) erhalten wir:

$$\left. \begin{aligned} \frac{3}{2} \sqrt{3} \mu_2 a_2 &= 2\sigma a_{-1} - \frac{3}{4} \sqrt{3} (1 - 2\mu_2) a_1 - \left(\sigma^2 + \frac{3}{4} \right) b_1, \\ \frac{3}{2} \sqrt{3} \mu_2 a_{-2} &= -2\sigma a_1 - \frac{3}{4} \sqrt{3} (1 - 2\mu_2) a_{-1} - \left(\sigma^2 + \frac{3}{4} \right) b_{-1}, \end{aligned} \right\} (33)$$

die zur Bestimmung von a_2 und a_{-2} dienen können. In entsprechender Weise ergeben sich:

$$\left. \begin{aligned} \frac{3}{2} \sqrt{3} \mu_2 b_2 &= - \left(\sigma^2 + \frac{9}{4} \right) a_1 - 2\sigma b_{-1} - \frac{3}{4} \sqrt{3} (1 - 2\mu_2) b_1, \\ \frac{3}{2} \sqrt{3} \mu_2 b_{-2} &= - \left(\sigma^2 + \frac{9}{4} \right) a_{-1} + 2\sigma b_1 - \frac{3}{4} \sqrt{3} (1 - 2\mu_2) b_{-1}, \end{aligned} \right\} (34)$$

woraus b_2 und b_{-2} bestimmt werden können.

Nachdem die Koordinaten von m_1 und m_2 bestimmt sind, können die Koordinaten von m_3 mit Hilfe der Schwerpunktsintegrale (7) bestimmt werden.

Wir setzen:

$$\left. \begin{aligned} \xi_3 &= a_3 \cos \sigma nt + a_{-3} \sin \sigma nt, \\ \eta_3 &= b_3 \cos \sigma nt + b_{-3} \sin \sigma nt, \end{aligned} \right\} \quad (35)$$

und erhalten für die Bestimmung der Koeffizienten a_3, a_{-3}, b_3, b_{-3} die folgenden Gleichungen:

$$\left. \begin{aligned} \mu_1 a_1 + \mu_2 a_2 + \mu_3 a_3 &= 0, \\ \mu_1 a_{-1} + \mu_2 a_{-2} + \mu_3 a_{-3} &= 0, \\ \mu_1 b_1 + \mu_2 b_2 + \mu_3 b_3 &= 0, \\ \mu_1 b_{-1} + \mu_2 b_{-2} + \mu_3 b_{-3} &= 0. \end{aligned} \right\} \quad (36)$$

In der weiteren Untersuchung wird es notwendig sein, die LAGRANGESCHEN und die nicht-LAGRANGESCHEN Lösungen getrennt zu behandeln.

4. Die Lagrangeschen Librationsellipsen.

Das Wurzelpaar $z = \pm i$ befriedigt die Bedingungsgleichung (13) für alle Werte der Massen. Es werden deshalb entsprechende Librationsellipsen für alle Massenwerte existieren. Um diese Librationsellipsen zu bestimmen, müssen wir in die oben abgeleiteten Formeln $\sigma = 1$ einführen.

Die Gleichungen (29) und (30) werden dadurch in die folgenden zwei übergeführt:

$$\left. \begin{aligned} \frac{3}{4} \sqrt{3} (1 - \mu_1 - 2\mu_2) a_1 + \frac{9}{2} (1 - \mu_1) a_{-1} - \frac{9}{4} (1 - \mu_1) b_1 \\ + \frac{3}{2} \sqrt{3} (1 - \mu_1 - 2\mu_2) b_{-1} = 0, \end{aligned} \right\} \quad (37)$$

$$\left. \begin{aligned} \frac{3}{4} \sqrt{3} (1 - \mu_1 - 2\mu_2) a_{-1} - \frac{9}{2} (1 - \mu_1) a_1 - \frac{9}{4} (1 - \mu_1) b_{-1} \\ - \frac{3}{2} \sqrt{3} (1 - \mu_1 - 2\mu_2) b_1 = 0. \end{aligned} \right\} \quad (38)$$

Wir wollen diese zwei Gleichungen zur Bestimmung von b_1 und b_{-1} benutzen, indem wir für a_1 und a_{-1} bestimmte Werte annehmen. Man könnte hier $a_1 = \varepsilon$, $a_{-1} = 0$ setzen, wo ε eine unendlich kleine Grösse 1. Ordnung bedeutet. Die Berechnung der Ausdrücke für b_1 und b_{-1} zeigt indessen, dass diese am einfachsten werden, wenn man

$$\left. \begin{aligned} a_1 &= (7\mu_2^2 + 7\mu_3^2 - 2\mu_2\mu_3)\varepsilon, \\ a_{-1} &= 0, \end{aligned} \right\} \quad (39)$$

setzt.

Mit diesen Werten für a_1 und a_{-1} geben die Gleichungen (37) und (38):

$$\left. \begin{aligned} b_1 &= 3\sqrt{3}(\mu_2^2 - \mu_3^2)\varepsilon, \\ b_{-1} &= -8(\mu_2^2 + \mu_3^2 + \mu_2\mu_3)\varepsilon. \end{aligned} \right\} \quad (40)$$

Wenn wir dann in die Gleichungen (33) $\sigma = 1$ und die Ausdrücke in (39) und (40) einsetzen, erhalten wir:

$$\left. \begin{aligned} a_2 &= (7\mu_2^2 + 7\mu_3^2 - 2\mu_2\mu_3 - 7\mu_2 + \mu_3)\varepsilon, \\ a_{-2} &= 4\sqrt{3}\mu_3\varepsilon. \end{aligned} \right\} \quad (41)$$

Entsprechend geben die Gleichungen (34):

$$\left. \begin{aligned} b_2 &= \sqrt{3}(3\mu_2^2 - 3\mu_3^2 - 3\mu_2 + 5\mu_3)\varepsilon, \\ b_{-2} &= -4(2\mu_2^2 + 2\mu_3^2 + 2\mu_2\mu_3 - 2\mu_2 - \mu_3)\varepsilon. \end{aligned} \right\} \quad (42)$$

Endlich erhalten wir mit Hilfe der Gleichungen (36):

$$\left. \begin{aligned} a_3 &= (7\mu_2^2 + 7\mu_3^2 - 2\mu_2\mu_3 + \mu_2 - 7\mu_3)\varepsilon, \\ a_{-3} &= -4\sqrt{3}\mu_2\varepsilon, \end{aligned} \right\} \quad (43)$$

$$\left. \begin{aligned} b_3 &= \sqrt{3}(3\mu_2^2 - 3\mu_3^2 - 5\mu_2 + 3\mu_3)\varepsilon, \\ b_{-3} &= -4(2\mu_2^2 + 2\mu_3^2 + 2\mu_2\mu_3 - \mu_2 - 2\mu_3)\varepsilon. \end{aligned} \right\} \quad (44)$$

Hiermit sind die LAGRANGESCHEN Librationsellipsen vollständig bestimmt.

5. Die Nicht-Lagrangeschen Librationsellipsen.

Wir betrachten jetzt $z = \pm i\sigma$ als ein Wurzelpaar in den für die nicht-LAGRANGESchen Lösungen gültigen Bedingungsgleichung $z^4 + z^2 + s = 0$. Die positive Zahl σ soll also die Bedingungsgleichung

$$\sigma^4 - \sigma^2 + s = 0 \quad (45)$$

befriedigen.

Wir gehen wieder von den zwei Gleichungen (29) und (30) aus und wollen mit ihrer Hilfe b_1 und b_{-1} als Funktionen von a_1 und a_{-1} bestimmen. Nach Umstellung der Glieder können diese Gleichungen so geschrieben werden:

$$\left. \begin{aligned} \left(\sigma^4 - \sigma^2 - \frac{9}{4}(1 - \mu_1)\sigma^2 \right) b_1 - \frac{3}{2}\sqrt{3}(\mu_2 - \mu_3)\sigma b_{-1} \\ = \frac{3}{4}\sqrt{3}(\mu_2 - \mu_3)\sigma^2 a_1 - \frac{9}{2}(1 - \mu_1)\sigma a_{-1}, \\ \frac{3}{2}\sqrt{3}(\mu_2 - \mu_3)\sigma b_1 + \left(\sigma^4 - \sigma^2 - \frac{9}{4}(1 - \mu_1)\sigma^2 \right) b_{-1} \\ = \frac{9}{2}(1 - \mu_1)\sigma a_1 + \frac{3}{4}\sqrt{3}(\mu_2 - \mu_3)\sigma^2 a_{-1}. \end{aligned} \right\} (46)$$

Die Gleichungen können noch weiter vereinfacht werden, wenn wir

$$\alpha = 1 - \mu_1 = \mu_2 + \mu_3, \quad (47)$$

$$\beta = \mu_2 - \mu_3 \quad (48)$$

einführen und ausserdem die Gleichung (45) benutzen. Wir erhalten dann:

$$\left. \begin{aligned} \left(\frac{9}{4}\alpha\sigma^2 + s \right) b_1 + \frac{3}{2}\sqrt{3}\beta\sigma b_{-1} = -\frac{3}{4}\sqrt{3}\beta\sigma^2 a_1 + \frac{9}{2}\alpha\sigma a_{-1}, \\ \frac{3}{2}\sqrt{3}\beta\sigma b_1 - \left(\frac{9}{4}\alpha\sigma^2 + s \right) b_{-1} = \frac{9}{2}\alpha\sigma a_1 + \frac{3}{4}\sqrt{3}\beta\sigma^2 a_{-1}. \end{aligned} \right\} (49)$$

Die Determinante des Systems (49) kann durch σ und α allein ausgedrückt werden. Die Rechnung ergibt:

$$D = \sigma^8 + \left(2 - \frac{9}{2}\alpha\right)\sigma^6 + \left(\frac{81}{16}\alpha^2 + \frac{9}{2}\alpha - 3\right)\sigma^4 + \left(27\alpha - \frac{81}{4}\alpha^2\right)\sigma^2. \quad (50)$$

Der Ausdruck für D kann aber in der folgenden Weise in Faktoren aufgelöst werden:

$$D = \left[\sigma^4 - \left(1 + \frac{9}{4}\alpha\right)\sigma^2 + 9\alpha\right] \cdot \left[\sigma^2 + 3 - \frac{9}{4}\alpha\right]\sigma^2. \quad (51)$$

Diese Auflösung in Faktoren bewirkt, dass die Ausdrücke für die gesuchten Koeffizienten ausserordentlich einfach werden. Um die Ausdrücke für die übrigen Koeffizienten so einfach wie möglich zu gestalten, wollen wir für a_1 und a_{-1} die folgenden Werte wählen:

$$\left. \begin{aligned} a_1 &= \left(\sigma^2 + \frac{3}{4} + \frac{9}{4}\mu_1\right)\mu_2\mu_3\varepsilon, \\ a_{-1} &= 0. \end{aligned} \right\} \quad (52)$$

Aus den Gleichungen (49) finden wir dann:

$$\left. \begin{aligned} b_1 &= \frac{3}{4}\sqrt{3}(\mu_2 - \mu_3)\mu_2\mu_3\varepsilon, \\ b_{-1} &= -2\sigma\mu_2\mu_3\varepsilon, \end{aligned} \right\} \quad (53)$$

ferner aus (33) und (34):

$$\left. \begin{aligned} a_2 &= -\frac{1}{2}\left(\sigma^2 + 3 - \frac{9}{2}\mu_2\right)\mu_1\mu_3\varepsilon, \\ a_{-2} &= -\sqrt{3}\sigma\mu_1\mu_3\varepsilon, \end{aligned} \right\} \quad (54)$$

$$\left. \begin{aligned} b_2 &= -\frac{1}{2}\sqrt{3}\left(\sigma^2 + \frac{3}{2} + \frac{3}{2}(\mu_1 - \mu_3)\right)\mu_1\mu_3\varepsilon, \\ b_{-2} &= \sigma\mu_1\mu_3\varepsilon, \end{aligned} \right\} \quad (55)$$

und endlich mit Hilfe der Gleichungen (36):

$$\left. \begin{aligned} a_3 &= -\frac{1}{2}\left(\sigma^2 + 3 - \frac{9}{2}\mu_3\right)\mu_1\mu_2\varepsilon, \\ a_{-3} &= \sqrt{3}\sigma\mu_1\mu_2\varepsilon, \end{aligned} \right\} \quad (56)$$

$$\left. \begin{aligned} b_3 &= \frac{1}{2}\sqrt{3}\left(\sigma^2 + \frac{3}{2} + \frac{3}{2}(\mu_1 - \mu_2)\right)\mu_1\mu_2\varepsilon, \\ b_{-3} &= \sigma\mu_1\mu_2\varepsilon. \end{aligned} \right\} \quad (57)$$

Damit sind auch die nicht-LAGRANGESchen Librationsellipsen vollständig bestimmt.

6. Die Umlaufsrichtung in den Librationsellipsen.

Wenn eine elliptische Bewegung des Punktes (ξ, η) durch die Gleichungen

$$\left. \begin{aligned} \xi &= a_1 \cos \omega t + a_{-1} \sin \omega t, \\ \eta &= b_1 \cos \omega t + b_{-1} \sin \omega t, \end{aligned} \right\} \quad (58)$$

dargestellt wird, wo ω als positiv vorausgesetzt wird, wird die Umlaufsrichtung vom Vorzeichen der Determinante

$$D_1 = \begin{vmatrix} a_1 & a_{-1} \\ b_1 & b_{-1} \end{vmatrix} = a_1 b_{-1} - a_{-1} b_1 \quad (59)$$

abhängen.

Wenn D_1 positiv ist, wird die Bewegung in positiver Richtung vorsehnen, während ein negativer Wert von D_1 bedeutet, dass die Bewegung retrograd ist.

Wir wollen jetzt die Determinanten berechnen, die den drei LAGRANGESchen Librationsellipsen entsprechen. Aus (39), (40), (41), (42), (43) und (44) erhalten wir:

$$\left. \begin{aligned} D_1 &= -8(7\mu_2^2 + 7\mu_3^2 - 2\mu_2\mu_3)(\mu_2^2 + \mu_3^2 + \mu_2\mu_3)\varepsilon^2, \\ D_2 &= -8(7\mu_2^2 + 7\mu_3^2 - 2\mu_2\mu_3)(\mu_1^2 + \mu_3^2 + \mu_1\mu_3)\varepsilon^2, \\ D_3 &= -8(7\mu_2^2 + 7\mu_3^2 - 2\mu_2\mu_3)(\mu_1^2 + \mu_2^2 + \mu_1\mu_2)\varepsilon^2. \end{aligned} \right\} \quad (60)$$

Da die drei Determinanten negative Werte haben, wird die Bewegung in den LAGRANGESchen Librationsellipsen retrograd sein.

In ähnlicher Weise berechnen wir die Determinanten, die den drei nicht-LAGRANGESchen Librationsellipsen entsprechen. Aus (52), (53), (54), (55), (56) und (57) erhalten wir:

$$\left. \begin{aligned} D_4 &= -2\sigma \left(\sigma^2 + \frac{3}{4} + \frac{9}{4}\mu_1 \right) \mu_2^2 \mu_3^2 \varepsilon^2, \\ D_5 &= -2\sigma \left(\sigma^2 + \frac{3}{4} + \frac{9}{4}\mu_1 \right) \mu_1^2 \mu_3^2 \varepsilon^2, \\ D_6 &= -2\sigma \left(\sigma^2 + \frac{3}{4} + \frac{9}{4}\mu_1 \right) \mu_1^2 \mu_2^2 \varepsilon^2. \end{aligned} \right\} \quad (61)$$

Auch hier haben die drei Determinanten negative Werte, und die Bewegung in den nicht-LAGRANGESchen Librationsellipsen muss retrograd sein.

7. Das Verhältnis zwischen der grossen und der kleinen Achse in den Librationsellipsen.

Wenn eine elliptische Bewegung durch die Gleichung (58) dargestellt wird, wird das Verhältnis zwischen der grossen und der kleinen Achse denselben Wert haben wie die grösste Wurzel der Gleichung

$$f^2 - gf + 1 = 0, \quad (62)$$

wo

$$g = \frac{a_1^2 + a_{-1}^2 + b_1^2 + b_{-1}^2}{|D_1|}. \quad (63)$$

Wenn wir jetzt die für die LAGRANGESchen Librationsellipsen gültigen (39), (40), (41), (42), (43) und (44) in (63) einsetzen, findet man in allen drei Fällen:

$$g = \frac{5}{2}, \quad (64)$$

wonach (62) das Resultat gibt, dass das Verhältnis zwischen der grossen und der kleinen Achse

in den LAGRANGESCHEN Librationsellipsen gleich 2:1 ist.

Wenn wir in ähnlicher Weise die für die nicht-LAGRANGESCHEN Librationsellipsen geltenden Werte (52), (53), (54), (55), (56) und (57) einsetzen, erhalten wir in allen drei Fällen:

$$g = \frac{2\sigma^2 + 3}{2\sigma} \quad (65)$$

und daraus wieder das Verhältnis zwischen der grossen und der kleinen Achse in den *nicht-LAGRANGESCHEN* Librationsellipsen:

$$f = \frac{1}{4\sigma} \left[2\sigma^2 + 3 + \sqrt{4\sigma^4 - 4\sigma^2 + 9} \right]. \quad (66)$$

Da das Verhältnis für alle drei Ellipsen dasselbe ist, müssen die *nicht-LAGRANGESCHEN* Librationsellipsen gleichförmig sein.

In der oben zitierten Abhandlung wurde bewiesen, dass s nur von der Entfernung zwischen dem Zentrum des Librationsdreiecks und dem gemeinsamen Schwerpunkt der Massen abhängt. Da σ nur von s abhängt und f nur von σ , erhalten wir den folgenden Satz:

Das Achsenverhältnis der *nicht-LAGRANGESCHEN* Librationsellipsen hängt nur von der Entfernung vom Zentrum des Librationsdreiecks zum gemeinsamen Schwerpunkt der Massen ab.

8. Das Grössenverhältnis der Librationsellipsen.

Das Produkt der Halbachsen in der elliptischen Bahn, die durch die Gleichungen (58) definiert ist, wird gleich dem numerischen Wert der Determinante D_1 sein. Daraus folgt, dass die linearen Grössen von zwei gleichförmigen

Ellipsen sich wie die Quadratwurzeln aus den numerischen Werten der betreffenden Determinanten verhalten.

Für die LAGRANGESchen Librationsellipsen werden ihre lineare Grössen sich deshalb wie

$$\sqrt{-D_1} : \sqrt{-D_2} : \sqrt{-D_3} \quad (67)$$

verhalten.

Wenn wir in (67) die Werte (60) der drei Determinanten einsetzen, erhalten wir das Resultat, dass die linearen Grössen der LAGRANGESchen Librationsellipsen sich wie

$$\sqrt{\mu_2^2 + \mu_2 \mu_3 + \mu_3^2} : \sqrt{\mu_1^2 + \mu_1 \mu_3 + \mu_3^2} : \sqrt{\mu_1^2 + \mu_1 \mu_2 + \mu_2^2} \quad (68)$$

verhalten.

In ähnlicher Weise findet man, dass die linearen Grössen der nicht-LAGRANGESchen Librationsellipsen sich wie

$$\sqrt{-D_4} : \sqrt{-D_5} : \sqrt{-D_6} \quad (69)$$

verhalten.

Wenn wir in (69) die Werte (61) der Determinanten einsetzen, erhalten wir das Resultat, dass die linearen Grössen der *nicht*-LAGRANGESchen Librationsellipsen sich wie

$$\mu_2 \mu_3 : \mu_1 \mu_3 : \mu_1 \mu_2 \quad (70)$$

d. h. wie

$$\frac{1}{\mu_1} : \frac{1}{\mu_2} : \frac{1}{\mu_3} \quad (71)$$

verhalten, d. h., dass die linearen Grössen der *nicht*-LAGRANGESchen Librationsellipsen sich umgekehrt wie die entsprechenden Massen verhalten.

Im problème restreint gibt es bekanntlich keine LAGRANGESche Librationsellipse für die unendlich kleine Masse, während nicht-LAGRANGESche Librationsellipsen um die Libra-

tionspunkte L_4 und L_5 existieren. Dies eigentümliche Verhältnis findet nun in den oben gegebenen Resultaten seine vollständige Erklärung.

Wenn wir nämlich $\mu_3 = 0$ in (68) einsetzen, finden wir für die LAGRANGESchen Librationsellipsen die Grössenverhältnisse:

$$\mu_2 : \mu_1 : \sqrt{\mu_1^2 + \mu_1 \mu_2 + \mu_2^2}. \quad (72)$$

Da es nun in der Natur der Sache liegt, dass im problème restreint die Librationsellipsen der zwei endlichen Massen die Grösse Null haben, muss die LAGRANGESche Librationsellipse der unendlich kleinen Masse auch die Grösse Null haben.

Setzen wir dagegen in (70) $\mu_3 = 0$, erhalten wir die Verhältnisse:

$$0 : 0 : \mu_1 \mu_2. \quad (73)$$

Das heisst: die nicht-LAGRANGESche Librationsellipse der unendlich kleinen Masse kann sehr wohl eine endliche Grösse haben, obschon die nicht-LAGRANGESchen Librationsellipsen für die zwei endlichen Massen die Grösse Null haben.

9. Die Orientierung der Librationsellipsen.

Wenn wir den Winkel zwischen der Abszissenachse und einer der Achsen in der elliptischen Bahn (58) mit θ bezeichnen, haben wir:

$$\operatorname{tg} 2\theta = \frac{2(a_1 b_1 + a_{-1} b_{-1})}{a_1^2 + a_{-1}^2 - b_1^2 - b_{-1}^2}. \quad (74)$$

Für die LAGRANGESche Librationsellipse, die von der Masse m_1 beschrieben wird, erhalten wir mit Hilfe von (39) und (40):

$$\operatorname{tg} 2\theta = \frac{\sqrt{3}(\mu_3^2 - \mu_2^2)}{\mu_2^2 + 4\mu_2 \mu_3 + \mu_3^2}. \quad (75)$$

Wenn die Winkel der Achsen mit der Abszissenachse mit θ_1 und θ_2 bezeichnet werden, ergibt sich aus (75):

$$\operatorname{tg} \theta_1 = \frac{\sqrt{3}(\mu_3 - \mu_2)}{3(\mu_2 + \mu_3)}, \quad (76)$$

$$\operatorname{tg} \theta_2 = \frac{\sqrt{3}(\mu_2 + \mu_3)}{\mu_2 - \mu_3}, \quad (77)$$

wo θ_1 sich auf die kleine Achse der Ellipse und θ_2 sich auf die grosse Achse bezieht. Die Richtung vom Librationspunkt zum gemeinsamen Schwerpunkt der Massen macht aber mit der Abszissenachse einen Winkel, der durch eine Gleichung bestimmt wird, die der Gleichung (76) entspricht. Wir erhalten somit das folgende Resultat:

Die LAGRANGESCHEN Librationsellipsen haben ihre kleinen Achsen gegen den Massenschwerpunkt gerichtet.

Wenn wir jetzt in (74) die für die nicht-LAGRANGESCHEN Librationsellipsen geltenden Werte (52), (53) einsetzen, erhalten wir:

$$\operatorname{tg} 2\theta = \frac{\sqrt{3}(\mu_3 - \mu_2)}{3(\mu_2 + \mu_3) - 2}. \quad (78)$$

Um die geometrische Bedeutung von (78) abzuleiten, wollen wir die Gleichung aufstellen, wodurch der in Abb. 3 angegebene Winkel φ bestimmt werden kann. M ist das Zentrum des dem Librationsdreieck umschriebenen Kreises, und MN ist die Linie von M durch den Massenschwerpunkt.

φ wird durch die folgende Gleichung bestimmt:

$$\operatorname{tg} \varphi = \frac{\sqrt{3}(\mu_3 - \mu_2)}{3(\mu_2 + \mu_3) - 2}. \quad (79)$$

Aus (78) und (79) erhalten wir nun:

$$\operatorname{tg} 2\theta = \operatorname{tg} \varphi, \quad (80)$$

$$\theta = \frac{\varphi}{2} + k \cdot 90^\circ. \quad (81)$$

Wenn $\mu_2 > \mu_3$, liegt die grosse Achse der Ellipse im 1. und 3. Quadranten, für $\mu_2 < \mu_3$ liegt sie im 2. und 4. Quadranten.

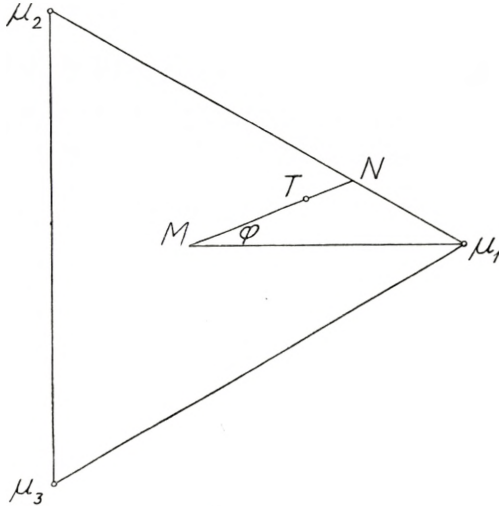


Abb. 3.

Für die *nicht-LAGRANGESCHEN* Librationsellipsen liegen die grossen Achsen parallel der Halbierungslinie des Winkels, der von den Linien vom Zentrum des Librationsdreiecks zum Massenschwerpunkt, bezw. zum betreffenden Librationspunkt, eingeschlossen ist.

In Abb. 4 sind der dem Librationsdreieck umschriebene Kreis und der Diameter durch den Schwerpunkt gezeichnet. *P* und *Q* sind die Endpunkte dieses Diameters. Man ersieht aus der Abbildung, dass die grossen Achsen der Librationsellipsen gegen *P* gerichtet sind, die kleinen Achsen gegen *Q*.

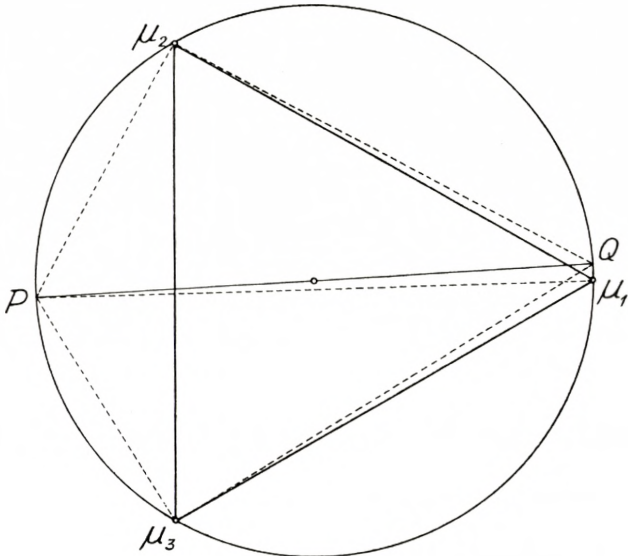


Abb. 4.

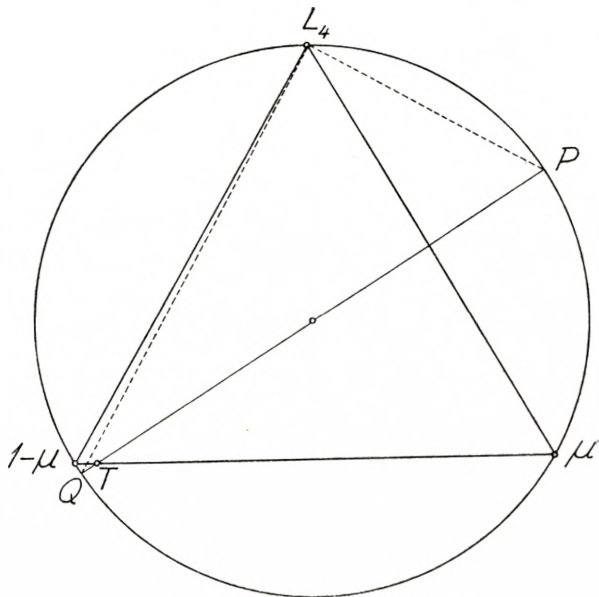


Abb. 5.

Die *nicht-LAGRANGESCHEN* Librationsellipsen haben ihre grossen Achsen gegen denselben Punkt gerichtet, die kleinen Achsen ebenso. Diese zwei Punkte sind die Endpunkte des Diameters in dem dem Librationsdreieck umschriebenen Kreis, der durch den Schwerpunkt geht. Die grossen Achsen sind gegen den Punkt gerichtet, der dem Schwerpunkt am entferntesten liegt, die kleinen Achsen gegen den Punkt, der dem Schwerpunkt am nächsten liegt.

Die oben angegebene Regel muss auch für die Librationsellipsen im problême restreint Gültigkeit haben. In Abb. 5 werden das Librationsdreieck und der ihm umschriebene Kreis gezeigt. Der Diameter PQ geht durch den gemeinsamen Schwerpunkt der zwei endlichen Massen $1 - \mu$ und μ , und die Librationsellipsen um den Librationspunkt L_4 müssen deshalb die grosse Achse gegen P und die kleine Achse gegen Q gerichtet haben.

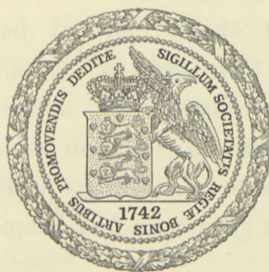
Det Kgl. Danske Videnskabernes Selskab.

Mathematisk-fysiske Meddelelser. **XIX**, 8.

ON THE
CONCEPT OF HEAT

BY

J. N. BRØNSTED



KØBENHAVN

I KOMMISSION HOS EJNAR MUNKSGAARD

1941

CONTENTS

I. Introduction	3
II. The Energetic Theory. ⁷⁷	
1. Work and Energetic Heat Evolution	5
2. The Energetics of Thermal Conduction	13
3. The Fundamental Equation	17
III. The Classical Concept of Heat.	
1. General Classical Aspects	19
2. The Spatial Process	30
3. The Chemical Process	48
4. The Thermal Process	61
IV. Conclusion	73



I. Introduction.

The old conception of heat as a substance has long ago been superseded by other ideas about the "nature" of heat. That heat is "motion", i. e. kinetic energy of the elementary particles of matter, has become the view generally accepted in all modern expositions of thermodynamics. The theory which maintains this idea is referred to as the "mechanical" or "dynamic" theory of heat.

But, it must be asked, what is it that is interpreted as the mechanical energy of the elementary particles of matter? The contention that heat is a mode of motion certainly conveys some other meaning than merely that of a definition. If this contention is to possess any physical meaning at all, there must exist already a fixed macroenergetic heat concept, to which such a microenergetic interpretation may be applied. In other words: the molecular interpretation of the "nature" of heat must be preceded by the unambiguous definition of a macroscopic concept with which, from one angle or other, one may wish to associate the term "heat". The significance of such a concept will depend upon the rigour and generality of the laws to which it is subject.

It is doubtless generally assumed that such a universal macroscopic heat definition has actually long ago been safely established. At any rate in the numerous expositions of thermodynamics and heat theory now existing apparently

no uncertainty or vagueness is encountered in the formal treatment of the fundamental chapters in which the concept of heat is introduced. A closer analysis, however, will show that the formal certitude and assurance which characterise the different methods by which heat, like the fundamental concept of work, is introduced in thermodynamics, does not imply a corresponding certainty and unambiguity in the logical comprehension of this concept. It is not perhaps a mere accident, therefore, that a true definition of heat invariably seems to be missing in the text books of thermodynamics. There is much to indicate that heat is generally regarded as a concept so elementary and familiar to the student of thermodynamics that a further penetration into its nature and an establishment of the clear and unequivocal definition which is otherwise required when introducing a new physical concept is not necessary. It is undoubtedly as a consequence of this attitude that untenable views concerning heat have established themselves, and that the development of the heat concept as well as attempts to formulate the elementary energetic basic phenomena as a whole have not so far been fully successful.

In some previous papers, to which reference will be made in the sequel, it has been attempted to give a logical description of these elementary basic phenomena. It has been possible to establish a general work principle adequately accounting for an important group of particularly simple energetic phenomena. In the cases where the work principle is no longer sufficient to describe the phenomena encountered there appears another fundamental law which can be directly formulated by means of the concept of heat, or more especially, of energetic heat evolution. In some particular respects this new concept may be described

as being identical with heat in the classical sense. In the domains in which the classical concepts are obscure or intrinsically inconsistent such a conformity will be lacking, and the problem will arise of confronting and comparing the old ideas with the new.

It is the object of the present paper to formulate and consolidate the new concept of heat on the basis of such a comparison. While in some fields, for instance in all that concerns the heat engine, the doctrines of heat are fairly consistently and uniformly presented in the various expositions, there are other cases, mostly relating to phenomena of a spatial and chemical nature, in which the formulation and interpretation of the experimental evidence regarding heat are widely divergent. In such cases it is deemed necessary to supply a more detailed documentation in order to elucidate the true character of the classical postulates.

II. The Energetic Theory.

1. Work and Energetic Heat Evolution.

In the previous papers¹⁾ the macroscopic changes in nature have been shown to consist of two groups of phenomena. The first of these groups comprises the various basic processes, i. e. processes which can be described as transport of certain "energetic quantities" from one state to another. Such processes may also be designated as partial processes of the first order. The second group comprises processes in which energetic quantities are generated or produced. These we shall describe as partial processes of the second order. As shown by experience, how-

¹⁾ J. N. BRØNSTED: D. Kgl. Danske Vidensk. Selskab, Math.-fys. Medd. XV, No. 4 (1937) and XVI, No. 10 (1939). Phil. Mag. (7) **29**, 449 (1940). J. Physical Chem. **44**, 699 (1940).

ever, only the thermal quantity entropy is subject to generation. In making this distinction we are essentially referring to stationary systems, the relation of the phenomena of motion to the mechanical basic processes, and thus to energetics as a whole, being exhaustively dealt with in the science of mechanics.

It has further been shown to be characteristic of the basic process that the processes described in traditional thermodynamics as reversible only consist of partial processes of this kind, whereas the irreversible processes comprise partial processes of both kinds. Accordingly reversible processes must be considered to be of greater simplicity than irreversible processes and should, preferably, form the introductory chapter in an exposition of thermodynamics.

Primarily, the various basic processes are characterised by the nature and the amount of the quantity transported. Furthermore, they are characterised by the states between which the transport takes place, the potential in these states being the fundamentally decisive property. The relation between quantity and potential is represented by the work expression:

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

where δA is the loss of work associated with the transport of the quantity δK , and P_1 and P_2 the potentials conjugate to this quantity in the initial and final states respectively. For example, in agreement with this general equation, we may write for the thermal basic process, which is of particular importance in this connection:

$$\delta A = (T_1 - T_2) \delta S, \quad (2)$$

δS being the amount of entropy transported and T_1 and T_2 the thermal potential, i. e. the temperature in the two states

between which the entropy transport takes place. It is important to note that the course of the basic process is determined solely on the basis of its given terminal states, independently of the path.

The physical significance of equation (1) is given by the work principle, expressed by:

$$\Sigma \delta A_1 = 0 \quad (3)$$

and established as the universal fundamental law for reversible processes. Starting from the existing definition of a quantity, i. e. the existing methods for comparing amounts of quantity of the same kind, equations (1) and (2) will permit the establishment of a scale for the conjugate potential on the basis of a conventional standard state. The conformity with (3) which on selection of suitable units for the energetic factors involved is generally observed for all kinds of work involved in reversible processes, is the experimental basis for the validity of the work principle.

The expressions set up, in accordance with their form, are valid for differential processes, i. e. processes of such slight extent that the two potentials involved are not appreciably affected in the course of the process. It is obvious that the two states 1 and 2, the potentials of which enter into the work expression, are of equal importance for the determination of the basic process and the magnitude of the work. Loss and gain of work are therefore restricted to systems in which both these states are included, i. e. systems describable as closed with regard to quantity. In the traditional treatment "work" is often spoken of in a different sense, the supply or removal of a volume δv (but hardly of any other quantity!) being identified with the

supply or expenditure of work $p\delta v$. This, however, must be considered as an unfortunate terminology, since it is apt to confuse two forms of energy which are fundamentally different in character. The energy of the form $P\delta K$ involved in the communication of a quantity and designated as neutral or equipotential energy¹⁾ is of a purely formal character, and excluded from participating in any energetic transformation. Only when K is entropy may such neutral energy appear in conjunction with the basic process as a form of energy for which the energy principle has any validity or meaning.

In the special case of $P_1 = P_2$, i. e. when the transport of a quantity takes place between two states in which the corresponding potential has the same value, the basic process is described as neutral. Such a process may take place without being accompanied by other processes. If the potentials differ from each other, the basic process is described as active, and in that case it will always be accompanied by other partial processes, which are either of the first or the second order, according to whether the total process is reversible or irreversible.

On the view advanced in the present paper, a reversible process, when not consisting in a neutral transport as described above, will comprise at least two basic processes. If each of these takes place in a separate system, the system in which the basic process is positive may be said to give off work to the system in which the basic process is negative. The two systems are said to be in work communication. It is obvious that in such a rever-

¹⁾ BRØNSTED: D. Kgl. Danske Vidensk. Selskab, Math.-fys. Medd. XV, No. 4, p. 13 (1937); XVI, No. 10, p. 57 (1939).

sible process work received or supplied will generally cause alterations of all potentials in the fractional systems in question, for which reason there has formerly been a tendency to put a different interpretation on such changes based upon the assumption of an intrinsic correlation between rise of temperature and evolution of heat. The complete reversibility, however, which characterises the phenomena described here is only to be found in coupled basic processes and will exclude a description which permits the participation of any energetic concepts other than work.

For processes of finite dimensions the loss of work is:

$$\Sigma A_i = \Sigma \int_2^1 P_i dK_i = 0. \quad (4)$$

If the reversible change of the system is made to take place in communication with work reservoirs, ΣA_i for the system will be a single-valued function of state. A system which remains unchanged with respect to its content of quantity can thus be characterised by a work function or potential energy E_p depending only on the state and given by:

$$E_{p_1} - E_{p_2} = \Sigma A_i. \quad (5)$$

It is thus possible from the point of view given here to offer a uniform and exhaustive description of the reversible processes on the basis of the concepts of quantity, potential and work, without "heat" entering at any point. More especially the reversible thermal processes may be completely described by means of the simple concepts of entropy and temperature. It is only when passing on to the irreversible processes that phenomena occur which may give rise to the formation of a new thermal concept.

As already mentioned, the irreversible processes comprise processes of both the first and the second order. This means that they are composed of basic processes in conjunction with the generation of entropy. We likewise find that the work principle is no longer valid since the total loss of work has always the same sign, which is conventionally fixed as being positive. By this convention we also obtain a rational fixation of the sign for the partial processes constituting the total reversible process as well as for the potential differences involved.

Experience having established that irreversible processes are characterised by a loss of work and a production of entropy, and this experience having been exalted into a postulate of unrestricted validity, the next step in the analysis must consist in a quantitative comparison of these two magnitudes. As the result of such an examination we shall find a very simple state of affairs, as long as the temperature of the entropy production is constant, the amount of entropy produced being proportional to the loss of work. Hence, such experiments would warrant the assumption of an equivalence of work and entropy, but the basis of such an assumption would fail if the temperature at which the entropy is generated, were changed, for in that case the same loss of work is found to generate an amount of entropy decreasing with increasing temperature. In investigating the closer correlation between the three magnitudes involved, viz. production of entropy, loss of work, and temperature, it must be remembered that according to classical thermodynamics the latter is based upon an already existing heat conception, at which we have not yet arrived by the method adopted here, and that, similarly, temperature according to the thermal scale defined in a

previous paper¹⁾ is based upon a heat definition. Only the energetic scale, which rests upon the work principle, is independent of the concept of heat and thus available for the purpose in question.

However, according to its derivation the energetic scale is a scale of temperature differences, i. e. a temperature scale, based upon an arbitrary zero. A temperature established on such a conventional basis cannot enter as a factor in the expression for a law of nature. Experiments on the irreversible evolution of entropy however show the following equation to be generally valid:

$$\delta A = T \delta S'' \quad (6)$$

in which δA is the loss of work, irrespective of its nature, $\delta S''$ the increase in entropy during the process, and T the absolute energetic temperature. This is defined as the energetic temperature read from a zero which is fixed on the basis of (6), introducing into this equation the experimental values of δA and $\delta S''$ of a single, arbitrarily chosen experiment. The T thus determined is in agreement with the Kelvin scale and the thermal scale mentioned above.

The result described here may also be expressed in the following way: Let t be the temperature in the energetic

¹⁾ BRØNSTED: D. Kgl. Danske Vidensk. Selskab, Math.-fys. Medd. XVI, No. 10, 25 (1939). The thermal temperature T is defined by the equation:

$$\frac{T_1}{T_2} = \frac{t_1 t''}{t_2 t'}$$

where T_1 and T_2 are the temperatures in the thermal scale of two heat reservoirs R_1 and R_2 and t_1 and t_2 are empirical changes in temperature (or another potential) read on two arbitrary potentiometers in R_1 and R_2 respectively, and caused by a reversible transport of entropy from one reservoir to the other. t' and t'' are corresponding changes caused by an irreversible heat transport using the same potentiometers as in the reversible case. The t 's are supposed to be small.

scale read from an arbitrary zero. Let δA be the loss of work and $\delta S''$ the production of entropy in a series of arbitrary irreversible processes. It is possible then to find a number t^* , such that the equation:

$$\delta A = (t + t^*) \delta S'' \quad (7)$$

is generally satisfied. $t + t^*$ is then the absolute temperature of the system, and t^* the absolute temperature of the conventional zero. It is obvious that δA and $\delta S''$ in equations (6) and (7) should be expressed in the same units as were used in applying the work principle to reversible thermal processes.

Equation (6) provides the basis for the creation of a new thermal concept. While entropy enters in the reversible thermal processes as a transportable quantity, the amount of entropy which leaves the initial state being equivalent to the amount of entropy which is received in the final state, a similar simple correlation between the loss of work and the production of entropy is no longer valid in the irreversible processes. What is proportional to the loss of work in these processes is not the entropy production, but the product of the entropy production and the absolute temperature. It is this macroenergetic concept, TdS , which is introduced in the energetic theory under the term heat, or, more precisely, as the energetic heat evolution. Using for this concept the symbol Q'' , we have by definition:

$$\delta Q'' = T \delta S'', \quad (8)$$

or by introduction of (6)

$$\delta A = \delta Q''. \quad (9)$$

It is true of any physical concept that only in fulfilling a certain phenomenologic law or regularity has it acquired

an actual meaning or *raison d'être*. For the heat concept, as introduced here, its simple relation to the loss of work constitutes this phenomenologic law. Equation (9) expresses this relation as the equivalence of the loss of work and the evolution of heat. Only the proportionality of these magnitudes is of course directly observable. It is the proportionality between the loss of work and the evolution of heat manifesting itself irrespective of the nature of the irreversible process which in a previous paper was introduced into energetics as the equivalence principle. There, it was also stated that the equivalence between heat and work is of a quantitative, but not of a genetic character, the sign of the basic process and of $\delta Q''$ being always positive. This postulate was introduced into energetics as the heat principle.

Combining the work and heat principles we conclude that no natural process permits a positive gain of work. Work can only be gained at the expense of an equivalent amount of work being lost. The task of technical science is restricted, therefore, to an efficacious transformation of the various types of work, utilizing the potential differences already existing in nature in conformity with the work principle. An exploitation of heat for the purpose of creating work is not rendered possible by the laws of nature.

2. The Energetics of Thermal Conduction.

The description of the irreversible processes given above is general and theoretically exhaustive. However, in the particular case where the quantity transported is identical with the quantity generated, inherently novel phenomena are bound to occur. These phenomena appear when bodies of different temperatures enter into direct thermal contact,

and are identical with the unique and important phenomena known as "thermal conduction" or "heat transport". We shall now analyse this process according to the principles given above for the irreversible processes.

We consider an entropy reservoir I of the temperature T_1 in thermal contact with an entropy reservoir II of the lower temperature T_2 , these temperatures being expressed in the absolute energetic scale. We suppose the connecting thermal conductor to be of such dimensions that the amount of entropy which may accumulate in it during the process may be disregarded. In this case there must enter into II an amount of entropy which is the sum of the entropy leaving I and that which is formed in the conductor during the thermal flow by virtue of its irreversibility.

Let us suppose that the amount of entropy S passes through a cross-section of the conductor in which the temperature is T . In the intervening space between this cross-section and a neighbouring one of the temperature $T - dT$ a loss of work will take place, which according to (1) is given by:

$$\delta A = -S dT. \quad (10)$$

On account of the irreversibility an amount of heat will be generated simultaneously in the same interval, which amount according to (8) is given by:

$$\delta Q'' = T dS, \quad (11)$$

where dS is the entropy produced in the interval.

Introducing the equivalence principle (9), according to which loss of work and evolution of heat are equal, we obtain:

$$-S dT = T dS, \quad (12)$$

or

$$TS = \text{constant}. \quad (13)$$

Accordingly, if during the process of conduction the amount of entropy given off by reservoir I is $\delta S'_1$, and that received by reservoir II is $\delta S'_2$, then by (13) these two quantities will satisfy the equation:

$$T_1 \delta S'_1 = T_2 \delta S'_2. \quad (14)$$

This result is verified by experiment.

In this equation, as in equation (6), we find a product of the form $T \delta S$ satisfying a law of great simplicity. Just as equation (6) supplied the basis for the formulation of the concept of heat evolution, and then on creation of this concept appeared as an expression of the law of the equivalence of heat and work, equation (14) may supply the basis for the formulation of the concept of heat communication, according to which this equation appears as an expression of the law of the irreversible transportability of heat. If the symbol Q' is used for heat communicated, equation (14) can be written as:

$$\delta Q'_1 = \delta Q'_2. \quad (15)$$

In case T_1 and T_2 are identical, (14) may also be taken as the expression for a neutral transport of entropy in conformity with the reversible character of this process.

In the changes which have been described as basic processes, we meet with a series of phenomena showing a complete mutual analogy. In the transference of heat described by (15) a process of quite a different character is encountered which is actually unparalleled by any other known natural phenomenon. The distinctive character of a heat transport as compared with the basic process may be illustrated by the statement that heat does not possess the character of an energetic quantity. According to the theory presented above, however, the process of thermal con-

duction is not a phenomenon differing by its intrinsic nature from irreversible processes in general. On the contrary, the analysis of the process of heat conduction given above shows how this phenomenon may be split into processes of the first and second order as typical of irreversible changes in general. Accordingly, the transportability principle represented by (14) or (15) is not a self-sufficing or independent law of heat, but is based upon the general equivalence principle, which in the form of (9) was introduced for its derivation.

In a previous paper¹⁾ the heat concept has been introduced on a somewhat different basis, the transportability being used as an actual definition of heat. This method is entirely justified, but the formulation of the heat concept given above possesses an advantage in being based on the phenomenon of irreversibility in general as contrasted with that of a specific thermal character.

One other essential feature related to the heat transport and distinguishing reversible and irreversible differential processes from each other should be mentioned. While in the reversible transport of a quantity δK the loss of work and the potentials involved in the work expression are given by the terminal states of the system in question, the potential for the entropy production is not necessarily identical with the temperature of the system in which the irreversible basic process takes place. It is possible, by means of special devices, to cause the production of entropy at an arbitrary temperature outside the system, and still from the same initial state to reach the same final state, the heat formed at a higher temperature being led irreversibly back to the system. In that case a loss of work resulting from the heat

1) D. Kgl. Danske Vidensk. Selskab, Math.-fys. Medd. XVI, No. 10 (1939).

transport in conformity with (10) will be added to the loss of work which is directly observed as a function of state; and to the heat evolution, if the entropy is formed at the temperature of the system, an equivalent evolution of heat will be added in conformity with (11). If the temperature of the system is T and the temperature at which the entropy is formed T'' , the loss of work and heat evolution will, as previously shown,¹⁾ be determined by:

$$\delta A = \delta Q'' = \delta Q \ln \frac{T''}{T}.$$

Since the final state is unaffected by possible differences in the mechanism of the process, and since the entropy transported does not belong to the initial state at all, being created during the course of the process, the additional energetic heat and work phenomena cannot be discovered from the properties of the initial and final states. Proof of their existence will only be possible through a detailed scrutiny of the special mechanism of reaction. It is essential to note that this conclusion is true of all kinds of irreversible changes.

3. The Fundamental Equation.

According to the above considerations, there are, evidently, two ways in which a system can take up heat. It can do so either by generation of heat in the system, or by communication of heat to the system. Heat taken up in the first of these ways is, as previously mentioned, always positive, i. e. heat can only be produced, not consumed. On the other hand communicated heat may obviously be either positive or negative. In either case the absorption of heat and entropy is connected by the relation:

1) D. Kgl. Danske Vidensk. Selskab, Math.-fys. Medd. XVI, No. 10, p. 40 (1939).

$$\delta Q = T dS. \quad (16)$$

It is characteristic of heat that it must be associated with processes or transformations, not with states. The contention of classical thermodynamics that heat is not a function of state, as are energy, quantity, and potential, also applies, therefore, to the extended concept of heat given by (16). Only under special conditions, primarily in "heat" or "entropy reservoirs", is such a dependence on state present. On the other hand, the thermal energy or the product TS , where S is the total entropy of the system, as well as the analogous products, corresponding to other forms of energy, are of course single-valued functions of the state of the system.

While the relation between heat and entropy communication is also accepted in classical thermodynamics, either, according to the point of departure, as a definition of heat or of entropy, the corresponding relation (8), which defines the heat evolution, and the expression (1), which defines the loss of work, are characteristic of the energetic theory. The equivalence principle (9) between the concepts thus established is therefore also foreign to the classical treatment. It should be made quite clear, however, that the actual content of the general equivalence principle, which according to (1) and (8) may be written:

$$\Sigma (P_1 - P_2) \delta K = T dS'', \quad (17)$$

is quite independent of any terminology by means of which the concepts involved are introduced into thermodynamics. Undoubtedly the terms "work" and "heat" employed to designate the magnitudes entering into equation (17) are adequate and consistent. If, however, objections should

be raised from a traditional point of view against such an application of these terms, this will not in the least affect the significance of our equivalence principle expressed by (17), since this equation is formulated solely by means of symbols of fully recognised significance. Regardless of any convention the equation represents a general and important principle, of which the classical "first law of thermodynamics" is only a special case. The genetic description of the energetic phenomena presented in the classical treatment as the "second law of thermodynamics" follows from it merely through the fixation of the sign of the magnitudes involved. For the special case of $dS'' = 0$, the equation represents the work principle introduced for the reversible processes. Taking into account the alternative possibilities with respect to the changes in entropy, the relation becomes:

$$\Sigma (P_1 - P_2) \delta K = T dS'' \geq 0 \quad (18)$$

and in this form is an expression of the three fundamental laws, the equivalence, heat, and work principles, which have previously been advanced as the fundamental laws of all macroenergetic phenomena¹⁾.

III. The Classical Phenomenon of Heat.

1. General Classical Aspects.

The basis of the heat concept, as historically evolved, is to be sought in the fact that thermal conduction and the simple mixing of substances of different temperatures are phenomena which are compatible with a material concept of heat. From investigations into such phenomena J. BLACK

¹⁾ Cf.: D. Kgl. Danske Vidensk. Selskab, Math.-fys. Medd. XVI, No. 10 (1939) and Phil. Mag. 29, 449 (1940).

succeeded in formulating the vague ideas of earlier times into a more well-founded and consistent theory in harmony with the conception of heat as a special substance, a certain elastic fluid which might be led in unchanged quantity from one place to another. Actually it is the theory of the irreversible transportability of heat that is contained in these ideas, but the general idea of the unchangeability of heat altogether was created at the same time.

Opposed to this idea of heat as an unalterable substance there early appeared the rival hypothesis that heat was an intestine commotion among the elementary particles of matter¹. RUMFORD and DAVY tried to show that heat could be formed by friction². CARNOT, however, seems to have been the first to formulate explicitly the supposition of a transition between heat and work, but his considerations on the subject³ remained unnoticed until long after the notion of the transformation, in a fixed proportion, of heat into work had been established by others. One of the pioneers in this field was the Danish scientist COLDING⁴. It is however to J. R. MAYER and J. P. JOULE that special credit should be given as the creators of the

1) See e. g. E. MACH: "Principien der Wärmelehre", 1896; M. PLANCK: "Das Princip der Erhaltung der Energie", 3. Aufl., 1913.

2) The conclusions drawn by RUMFORD and DAVY on the basis of their experiments are subjected to an interesting criticism by WILLIAM THOMSON in the article "Heat" in the Encyclopaedia Britannica, 9th Ed., 1878. Cf. Math. and Phys. Papers by Sir WILLIAM THOMSON, 3, 113 (1890).

3) S. CARNOT: "Sur la puissance motrice du feu" (Réimpression Paris 1878). Cf.: C. r. 87, p. 967 (1878).

4) L. A. COLDING: "Some Theses on the Forces". The manuscript was presented in the Royal Danish Academy of Sciences in November 1843, but was not printed until 13 years later in the "Oversigt" of the Academy No. 8, 1856. A more elaborate work: "Investigations into the General Forces in Nature and their Interdependence" appeared in D. Kgl. Danske Vidensk. Selskab, Skr., nat. og math. Afd. 5, II, p. 121 (1851). Cf. also Phil. Mag. [4] 27, 56 (1864).

transformation theory and the theory of the existence of energetic equivalence between heat and work¹⁾.

It is characteristic of these early contributions to the problem of the relation between heat and work that no reservation is made with regard to the direction of the processes in which the equivalence in question is supposed to appear. It is imagined that the inconstancy of heat appears both in the case of the production and the consumption of heat. Although the equivalence was not yet formulated as an actual "energy principle", it was the notion of reciprocity or equal genetic-energetic value of heat and work which dominated the ideas. Thus J. R. MAYER writes: "Wie die Wärme als Wirkung entsteht, bei Volumensverminderung und aufhörender Bewegung, so verschwindet die Wärme als Ursache unter dem Auftreten ihrer Wirkungen, der Bewegung, Volumensvermehrung, Lasterhebung" ²⁾ Further, in describing the function of the heat engine, he writes: "Der Aufwand von Wärme, oder die Verwandlung der Wärme in Bewegung beruht nun darauf, dass die Wärmemenge, welche von den Dämpfen aufgenommen wird, fortwährend grösser ist als die, welche von den Dämpfen bei ihrer Verdichtung an die Umgebung wieder abgesetzt wird. Die Differenz giebt die nutzbar verwendete, oder die in mechanischen Effekt verwandelte Wärme" ³⁾. "Die Leistung der Maschine ist an ein Konsumo von Wärme unzer trennlich geknüpft." ⁴⁾

1) The papers on the subject are of easiest access in: ROBERT MAYER: "Die Mechanik der Wärme", 1893, and "The Scientific Papers" of J. P. JOULE, London, 1884—87.

2) ROBERT MAYER: "Bemerkungen über die Kräfte der unbelebten Natur". Ann. Ch. Pharm. 1842; "Die Mechanik der Wärme", p. 28.

3) ROBERT MAYER: "Die organische Bewegung in ihrem Zusammenhange mit dem Stoffwechsel". 1845; "Die Mechanik der Wärme", p. 52.

4) Ibidem, p. 54.

Or, even more precisely: "Wenn wir in dieser Richtung vorgehen, und die auf mechanischem Wege entwickelte Wärmemenge, sowie die dazu verbrauchte Arbeitskraft messen, und diese Grössen mit einander vergleichen, so finden wir sofort, dass dieselben in der denkbar einfachsten Beziehung, d. h. in einem unveränderlichen geraden Verhältnisse zu einander stehen, und dass das nämliche Verhältnis auch statt hat, wenn umgekehrt mit Hilfe der Wärme wieder Arbeitskraft erzeugt wird. Diese Thatsachen in kurze klare Worte gefasst, sagen wir: "Wärme und Bewegung verwandeln sich in einander"¹⁾.

In the works of JOULE we find the same notions: His conclusions from his famous experiments on the generation of heat in chemical, electrical and mechanical processes are not confined to the directly observed phenomenon of transformation, but also include the inverse processes, in which heat is supposed to be converted into work. Thus in a paper presented at the meeting of the "British Association" in 1843 Joule wrote²⁾: "The quantity of heat capable of increasing the temperature of a pound of water by one degree of Fahrenheit's scale is equal to, and may be converted into, a mechanical force capable of raising 838 lb. to the perpendicular height of one foot". As another typical example of his ideas the following quotation from a paper of 1847 may be given³⁾: "The general rule, then, is, that wherever living force is apparently destroyed, whether by percussion, friction, or any similar means, an exact equivalent of heat is restored. The converse of

1) ROBERT MAYER: "Bemerkungen über die mechanische Äquivalent der Wärme". 1851; "Die Mechanik der Wärme", pp. 242—43.

2) Scientific Papers of JAMES PRESCOTT JOULE, I, p. 123.

3) "On Matter, Living Force, and Heat". The Scientific Papers of JAMES PRESCOTT JOULE, I, p. 265.

this proposition is also true, namely, that heat cannot be lessened or absorbed without the production of living force, or its equivalent attraction through space. Thus, for instance, in the steam engine it will be found that the power gained is at the expense of the heat of the fire, that is, that the heat occasioned by the combustion of the coal would have been greater had a part of it not been absorbed in producing and maintaining the living force of the machinery. It is right, however, to observe that this has not as yet been demonstrated by experiment. But there is no room to doubt that experiment would prove the correctness of what I have said; for I have myself proved that a conversion of heat into living force takes place in the expansion of air, which is analogous to the expansion of steam in the cylinder of the steam engine. But the most convincing proof of the conversion of heat into living force has been derived from my experiments with the electro-magnetic engine, a machine composed of magnets and bars of iron set in motion by an electric battery. I have proved by actual experiment that, in exact proportion to the force with which this machine works, heat is abstracted from the electric battery. You see, therefore, that living force may be converted into heat, and that heat may be converted into living force, or its equivalent attraction through space." "... The same quantity of heat will always be converted into the same quantity of living force".

In his well-known work on the conservation of energy¹⁾ HELMHOLTZ in the main accepts JOULE's ideas, while at the same time he particularly emphasises the mechanical

¹⁾ H. HELMHOLTZ: "Über die Erhaltung der Kraft", 1847. Ostwalds Klassiker No. 1.

conception of heat, free and latent, as "living force" and "tensional force" between the elementary particles of matter. In support of the idea of the production of "mechanical force" at the cost of heat he mentions JOULE'S experiments on the expansion of air with and without the production of work upon the surroundings, and it seems that in these experiments HELMHOLTZ sees the realisation of such a heat-work transformation. On the whole, a consumption of heat in the production of mechanical work is considered a necessary consequence of the law of the conservation of energy¹⁾. Although HELMHOLTZ is often held to have been the first to express the energy principle as a quite general law, applicable to all kinds of energy, he is in reality a stranger to the conception of the correlation between the different macroscopic forms of energy which is the main principle of the energetic theory.

The postulate of a reciprocal transformation of heat and work which is proclaimed in these early contributions to energetics as verified by experience is mainly based upon the fact that a system which undergoes a definite change gives off less or takes up more heat if work is done upon the surroundings during the process, than if this is not the case. That such a postulate can be based on this perfectly correct observation is, however, an entirely unjustified contention, since the observation only concerns the numerical, not the genetic correlation between the different kinds of energy. Nevertheless, the reciprocity postulate has exercised a decisive influence on the entire development of the concept of heat, and has been maintained, although in a modified form, right down to the

¹⁾ Cf. also WILLIAM THOMSON, *Mathematical and Physical Papers* 3, 236 (1890).

present time. To such an extent, indeed, has this idea of the conversion of heat into work become an integral part of the existing theories, that the historiography of later times has not even paused at this primitive fallacy in the establishment of the fundamental laws of thermodynamics. As late as 1849 WILLIAM THOMSON questioned the validity of JOULE'S proof of the reciprocal transformation of heat and work. Having acknowledged JOULE'S discovery concerning the production of heat he adds: "No experiment however is adduced in which the converse operation is exhibited"¹). But this, according to REYNOLDS²) is the only doubt which has ever been thrown on the sufficiency of JOULE'S demonstration. A modification of this view, however, was introduced shortly afterwards by CLAUSIUS and by WILLIAM THOMSON himself.

It was only natural that during the time immediately following JOULE'S fundamental experiments efforts should be made to establish a theory of the heat engine, based on the new principle of equivalence. The functioning of this engine, which seems to correlate heat and work in a particularly striking and significant manner, had been interpreted by CARNOT from entirely different points of view. While CARNOT had regarded a transfer of heat from a higher to a lower temperature as the basis of the mechanical work done by the engine, the newly established equivalence principle seemed to require an expenditure of heat as an indispensable condition of the work done. The difficulty of reconciling these two ideas was clearly recognised by WILLIAM THOMSON³). It was CLAUSIUS who

1) WILLIAM THOMSON: *Mathem. and Phys. Papers* I, p. 102 (1882).

2) OSBORNE REYNOLDS: *Memoirs of JAMES PRESCOTT JOULE*, Manchester Lit. a. Philos. Soc. 1892, p. 122.

3) W. THOMSON: *An Account of CARNOT'S Theory of the Motive Power of Heat*, Edinb. Trans. 16, 541 (1849).

first brought forward the alleged solution of this problem in advancing a conception of the functioning of the heat engine which aimed at reconciling the conflicting ideas.

According to the ideas of CLAUSIUS the mechanical production of work takes place at the cost of heat, the functioning of the engine being accompanied by an expenditure of heat of the same magnitude as the gain of work. In this respect there is conformity with the MAYER-JOULE principle. The new idea introduced by CLAUSIUS is that simultaneously with this heat-work transformation a compensating parallel reaction takes place in which heat is transported from a higher to a lower temperature.

In the previous papers the conception of CLAUSIUS has been subjected to a critical analysis and it will also be dealt with in more detail in one of the later chapters of the present paper. Here we shall only observe that it has exercised a considerable influence on the conception of heat as a work-producing form of energy, since not only in the case of the thermal process, with which we are here concerned, but quite generally "compensating processes" have been claimed to take place when heat is supposedly expended in the production of work. This conception of the work-producing property of heat, which is based upon the equivalence principle in its classical form, but adds a new feature of an essentially irrational character, is still prevalent at the present day. Nevertheless in numerous processes in which a transformation of heat into work is claimed to take place, the compensation process, although virtually recognised as necessary, is left entirely out of consideration.

As has been previously¹⁾ shown, the function of the heat

¹⁾ D. Kgl. Danske Vidensk. Selskab, Math.-fys. Medd. XV, No. 4 (1937).

engine is easily interpretable on the basis of the work principle in conformity with Equation (2)¹⁾. It was explicitly proved that in the thermal process, no expenditure of heat takes place any more than in other reversible processes, the idea of heat being conceived in conformity with the classical view as representing a form of energy. In a later section this question will be more closely analysed and illustrated in connection with the discussion of the CLAUSIUS theory.

As already stated, from a purely phenomenologic angle the traditional notion of heat transport and heat communication is in agreement with the ideas of the present exposition. This agreement is plainly revealed in equations (12) and (13), the validity of which is accepted also in classical thermodynamics, where they serve to define the concept of entropy. The essence of traditional ideas is the postulate that the amount of heat leaving one heat reservoir in a process of irreversible thermal communication is equal to the amount of heat taken up by the other heat reservoir. It is worth noticing, however, that such a postulate cannot be advanced as a result of an empirical observation concerning heat, unless this concept has been previously defined on a different basis. The actual facts behind the classical postulate do not permit a comparison of the two amounts of heat entering into the classical formulation, but can be expressed by the following postulate:

If two heat reservoirs, i. e. two material systems open to changes by communication of entropy only, are in a state of irreversible thermal communication, this process comprising merely the two reservoirs, not the surroundings, then the change in one reservoir accompanying a

¹⁾ D. Kgl. Danske Vidensk. Selskab, Math.-fys. Medd. XV, No. 4 (1937).

given change in the other reservoir is independent of the path by which the thermal communication takes place.

If the changes observed in the reservoirs are described as due to a transport of heat, then obviously it can only be ascertained by experience that a given amount of heat lost in one reservoir is always connected with a constant amount of heat gained in the other reservoir. Equality of these two amounts of heat is outside the possibility of experimental verification. The contention, therefore, that they are equal, although not justified by any actual demonstration, cannot lead to any conflict with experience, for on account of the irreversibility of the process of heat conduction it is impossible to subject the postulate to any experimental test.

Quite different from this, however, is the position of the classical theory as regards the phenomena of production and consumption of heat. According to the energetic theory, these phenomena should be interpreted in conformity with equation (6), which defines the evolution of heat by means of its relation to the entropy produced. As previously emphasised, on the classical view this equation will not define any amount of heat, except in certain particularly simple cases. This is revealed in a striking manner by the fact that the "heat evolution", contrary to the production of entropy, is supposed to be able to have either a positive or a negative sign. If we ask how then is formulated the classical definition of the concept of heat evolution, we encounter however the very peculiar fact that classical thermodynamics does not contain any general and universally accepted definition whatever of this concept, and that in various

quarters entirely different conceptions of heat evolution are advocated in the individual cases. Although such a lack of clarity is understandable at the moment when the very idea of the "conservation of energy" is intuitively brought into existence and therefore does not in any way reduce the importance of the achievements due to the creators of the principle of the equivalence of heat and work, the persistent absence of clarity with regard to the fundamental concepts which characterises the development in the succeeding time has had fatal consequences in the system of classical thermodynamics.

This uncertainty with regard to the concept of heat is, as previously mentioned, closely connected with the fact that the traditional treatment of the problem of heat generation does not make any distinction at all between reversible and irreversible processes. From the establishment of the equivalence theory right down to modern times, such phenomena as heat evolution by elastic compression and heat evolution by friction figure as co-ordinated and fundamentally identical phenomena. It is in this respect that the contrast between the traditional conception and the energetic theory manifests itself in the most marked and striking manner.

It is also a remarkable fact that in many domains of thermodynamics heat has not yet been entirely disentangled from the notions of the caloric theory, which attributes to heat the properties of a material quantity. In his well-known "Theory of Heat" MAXWELL¹⁾ after describing some of the properties of heat says: "We have therefore the right to speak of heat as a measurable quantity, and to treat it mathematically like other measurable quantities

¹⁾ J. CL. MAXWELL, "Theory of Heat", 5 Ed., 7 (1877).

so long as it continues to exist as heat". This statement presupposes the possibility of a heat-work-transformation and also recognises the fact that heat is not a function of state. Nevertheless it is assumed by MAXWELL — and this assumption is, as will appear from numerous instances in recent thermodynamic literature¹⁾, still commonly accepted — that heat can be treated as a property of a system in the sense that by investigating the state before and after an irreversible change of the system it would be possible to determine the amount of heat developed during that change. Such a view is actually the basis of CLAUSIUS' theory of the heat engine, of the classical interpretation of the transport of heat, and many other energetic phenomena. As already stated in chapter II. 3, heat does not, according to the energetic theory, possess the character of such a property, which is a simple consequence of the general definition of heat given by equation (II. 16).

In the following sections the principal basic processes i. e. the processes of a spatial, thermal, and chemical nature, are considered separately. The treatment comprises the necessary documentation with respect to the classical conceptions and interpretations of these processes combined with a critical analysis based on the view-points of the energetic theory.

2. The Spatial Process.

The spatial basic process in its differential form consists in the transference of a volume from one pressure to another. The process can be carried out for example by the infinitesimal displacement of a piston which separates two material states of different pressures. In the integral

¹⁾ See e. g. C.W. MILLER, "Introduction to Physical Science", p. 90 (1935).

basic process the change of state in question is given finite dimensions, whereby also the pressures in the two states separated by the piston undergo finite changes. In physics processes of this kind are designated in different ways according to the nature of the circumstances, for instance as expansion, dilatation, compression and contraction.

According to the energetic theory, no caloric or heat phenomenon is involved in the reversible course of these processes while the irreversible course is accompanied by an evolution of heat. The classical interpretation is quite different. In order to provide a clear and indisputable basis for the subsequent analysis it will be necessary to elucidate the historical status by means of a comprehensive documentation. Therefore we shall first quote a number of passages illustrating the ideas of the creators of the energy principle, or of authoritative expositions of thermodynamics on the subject of the spatial changes in gases or liquids.

J. R. MAYER¹), dealing with natural processes in which different types of energy are interconverted, says: "Verwandlung von mechanischem Effekt in Wärme bei der Kompression elastischer Flüssigkeiten, bei Stoss und Reibung" . . . "Die Verwandlung der Wärme in mechanischen Effekt erfolgt bei der Ausdehnung der Gasarten unter einem Drucke"

J. P. JOULE²): "From them" (JOULE'S principles) "we may infer that the steam while expanding in the cylinder, loses heat in quantity exactly proportional to the mechanical force which it communicates by means of the piston".

1) J. R. MAYER, "Mechanik der Wärme", p. 72.

2) J. P. JOULE, Scientific Papers I, 189. Cf. also p. 23 of the present article.

In a biography of JOULE J. G. CROWTHER¹⁾ says: "By these experiments JOULE proved to a high accuracy the assumption previously made by MAYER, that in the compression of air the whole of the work done is converted into heat".

H. v. HELMHOLTZ: When water evaporates, a weight being lifted by the expansion, more heat must be added than if the same process of expansion takes place without any production of work: "Wir haben hier also eine Verwandlung von Wärme in mechanische Energie²⁾".

Furthermore³⁾ "Giebt der Stempel nach, so wird durch den Druck des Gases mechanische Arbeit geleistet, und die im Dampf bei der Ausdehnung verschwindende Wärme wird zur Leistung dieser Arbeit verwendet". And: "Suchen wir bei der Ausdehnung eines Gases nach dem Äquivalent für den Gewinn an mechanischer Energie, so finden wir ihn in dem Wärmeverbrauch, der sich in der Abkühlung des Gases äussert".

H. C. ØRSTED. In a communication: "Experiments on the Heat developed by Water when compressed"⁴⁾ he says: "About $\frac{1^\circ}{40}$ centigrade is developed in the water for each atmosphere applied".

L. A. COLDING⁵⁾: "When heat expands a vapour or a gas, then the temperature decreases if no heat is supplied during the expansion. But the activity is not therefore lost,

1) J. G. CROWTHER, "British Scientists in the 19th Century", p. 183 (1935).

2) H. v. HELMHOLTZ, Vorles. über theor. Phys. I, 2 p. 224 (1898).

3) H. v. HELMHOLTZ, Ibid. VI, pp. 171—173 (1903).

4) H. C. ØRSTED, D. Kgl. Danske Vidensk. Selskab, Overs. for the year 1845, p. 117 (1846).

5) L. A. COLDING, D. Kgl. Danske Vidensk. Selskab, Skr., nat. og math. Afd. 5, II, p. 130 (1851).

since an amount of motion has been produced during the expansion. If a gas or liquid is compressed, an amount of motion is consumed in the process, but this amount has not disappeared either; it also manifests itself in other forms as heat etc.”.

H. V. REGNAULT¹⁾ after mentioning experiments to measure the evolution of heat in the compression of water says: “Je conclus de ces expériences que la chaleur dégagée par une pression subite de 10 atmosphères sur l’eau, est incapable d’élever sa température de $\frac{1}{50}$ de degré centigrade”.

REGNAULT thus imagines that the compression is associated with an evolution of heat, which increases the temperature of the compressed liquid.

G. A. HIRN²⁾: “Tout corps qu’on comprime s’échauffe”. “... il y a donc réellement production de chaleur dans un gaz qu’on comprime, et disparition de chaleur dans un gaz qui se détend, et non pas simplement manifestation de ce qui y existait déjà”.

JOHN TYNDALL³⁾: “Heat is known to be developed when air is compressed”. As an experimental proof of this he mentions the experiment that vapours of carbon disulfide in air ignite when the mixture is compressed. The increase of temperature is obviously interpreted as an evolution of heat in the gaseous mixture.

R. CLAUDIUS⁴⁾: “Ein permanentes Gas verschlucht, wenn es sich bei konstanter Temperatur ausdehnt, nur so viel

1) H. V. REGNAULT, Mémoires de l’Académie Royale des Sciences, Paris. 21, 462 (1847).

2) G. A. HIRN, “Théorie Mécanique de la Chaleur” (2^{me} Ed.), 1865.

3) JOHN TYNDALL, “Heat Considered as a Mode of Motion”, London 1865, p. 27.

4) R. CLAUDIUS, “Mechanische Wärmetheorie” I, 46 (1876).

Wärme, wie zu der äusseren Arbeit verbraucht wird”.

L. BOLTZMANN¹⁾. The heat amount dQ is added to a gas, increasing temperature and volume. “Wir setzen $dQ = dQ_1 + dQ_4$, wobei dQ_1 die auf Erhöhung der Molekularenergie verwendete, dQ_4 aber die auf äussere Arbeitsleistung verwendete Wärme darstellt”.

WILHELM OSTWALD²⁾. On the subject of the isothermal expansion of gas he says: “Wir haben hier den idealen Grenzfall der schon mehrfach erwähnten Maschine zur Umwandlung von Wärme in mechanische Energie ohne Temperaturdifferenz vor uns”.

M. PLANCK³⁾. Discussing CLAUSIUS' conceptions concerning the expansion of air in the case of the expanding gas being perfect, he declares: “Dann wird also alle äussere Arbeit lediglich auf Kosten der Wärme des Gases geleistet”.

PLANCK⁴⁾ strongly emphasises the fact that heat is transformed into work during the expansion of an ideal gas. After mentioning the particularly simple energetic relations of these substances he says: “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

1) L. BOLTZMANN, “Vorlesungen über Gastheorie” I, p. 54 (1923).

2) W. OSTWALD, “Lehrbuch d. allgemeinen Chemie” II. I. p. 32 (1910).

3) M. PLANCK, “Das Princip der Erhaltung der Energie” p. 63 (1913).

4) M. PLANCK, “Thermodynamik”, 1. Aufl., p. 74 (1897); 9. Aufl., p. 80 (1930).

worden ist, ohne dass sonst irgendwo ein Energieumsatz stattgefunden hat. Gegen diesen Ausspruch lässt sich nicht das mindeste Tatsächliche einwenden". Further, discussing the compression of air, he writes: "Leitet man nun, um das Gas auch auf seine alte Temperatur zurückzubringen die Kompressionswärme bei konstant gehaltenem Volumen ab". This statement seems to imply that a "heat of compression" is involved in the experiment in any case, even if no heat is led away from the system, which is in accordance with the above quotation from "Das Princip der Erhaltung der Energie".

J. BERTRAND¹⁾: Air is compressed without any rise of temperature when connected with a reservoir: "qui impose la température constante T_2 , en recueillant toute la chaleur produite par la compression".

B. WEINSTEIN²⁾: "Wie bemerkt, wandelt sich Wärme in jede andere Energie um, und wird von jeder anderen Art Energie hervorgerufen Erwärmt man ein Gas, so dehnt es sich aus, kühlt man es ab, so zieht es sich zusammen". The latter part of the quotation exemplifies the validity of the first part.

A. WASSMUTH³⁾: "Komprimiert man ein Gas, so steigt seine Temperatur, weil die aufgewendete Arbeit in Wärme umgesetzt wird".

Considering the compression of water below 4° , which takes place adiabatically and is accompanied by a decrease in temperature, he writes⁴⁾: "Wird keine Wärme von aussen

1) J. BERTRAND, "Thermodynamique", p 28 (1887).

2) B. WEINSTEIN, "Thermodynamik und Kinetik der Körper", I p. 14 (1901).

3) A. WASSMUTH, "Müller-Pouillet's Lehrb. d. Physik" III. 4 (1907), p. 645.

4) Ibidem, p. 688.

zugeführt, . . . und doch das Volum verkleinert, so muss die, wie wir sehen, hierzu nötige Wärme aus dem Innern des Körpers selbst genommen werden".

H. POINCARÉ¹⁾: The cooling effect of an expanding gas is explained as an expenditure of heat associated with a production of work: "La chaleur abandonnée par le gaz est transformée en travail".

A. EUCKEN²⁾. As an example of the equivalence of heat and work he mentions the expansion of a gas at constant temperature in thermal contact with the surroundings. Processes of this kind are accompanied by "Nebeneffekte" (CLAUSIUS' compensations), in the present case the change in the volume of the gas. For perfect gases, however, he finds "dass die bei der Expansion eines Gases gewonnene mechanische Energie tatsächlich der verbrauchten Wärmemenge äquivalent ist". Thus heat is expended to produce an equivalent amount of work.

W. C. McLEWIS³⁾: "During the process of isothermal compression, the heat which is produced and evolved is absorbed by the surroundings . . .".

Furthermore⁴⁾: "The source of the energy which has been used in the work of expansion if the temperature has been maintained constant, is the heat which has been drawn from the surroundings. If we had enclosed the gas in a "heat tight" case and allowed it to do work we should have found that its temperature would have fallen. In the latter case the work is done at the expense of the

1) H. POINCARÉ, "Thermodynamique", p. 82 (1923).

2) A. EUCKEN, Müller-Pouillet: Lehrb. d. Physik III. 1. pp. 113—14 (1926).

3) W. C. McLEWIS, "A System of Physical Chemistry" II; "Thermodynamics", p. 19 (1920).

4) Ibidem, pp. 3—4.

heat energy in the gas itself". With regard to the isothermal expansion he says: "what we have actually done is to transform heat energy into mechanical energy".

Again¹⁾ concerning the isothermal expansion: "the heat which had to be added . . . was quantitatively converted into the work done".

J. EGGERT²⁾. Discussing the expansion or compression of a gas he declares: ". . . verliert oder gewinnt das Gas—auch das ideale—jetzt ebensoviel Wärme, wie der Stempel an mechanischer Arbeit empfängt oder leistet". This applies to both isothermal and adiabatic processes.

K. JELLINEK³⁾ considers the expansion of a gas which overcomes an external pressure thus doing work at the same time as supplying kinetic energy to a heavy piston. If this takes place when the gas is in thermal contact with a large water bath, "so haben wir als Quelle der mechanischen Arbeit und kinetischen Energie die von dem Wasserbad auf das Gas übergehende äquivalente Wärmemenge anzusehen".

In a later publication⁴⁾ the same author says: "Da bekanntlich die gesamte Energie W bei der Expansion sich nicht ändert, so erkennt man, dass als Quelle für die geleistete Arbeit nur eine Wärmemenge $Q = A$ in Frage kommt, die aus dem Wärmebad in das Gasmol hineinfließt".

D. A. MACINNES⁵⁾: "The isothermal expansion of a perfect gas is therefore a means for converting heat into work quantitatively".

1) Ibidem, p. 9, footnote.

2) J. EGGERT, "Lehrb. d. physik. Chemie" I, p. 32 (1926).

3) K. JELLINEK, "Lehrb. d. physik. Chemie" I, p. 84 (1914).

4) K. JELLINEK, "Kurzes Lehrbuch der physikalischen Chemie", 31 (1938).

5) D. A. MACINNES, "The Principles of Electrochemistry", 97 (1939).

R. A. HOUSTOUN¹⁾ considering the reversible gas expansion: "... According to the first principle of thermodynamics this heat is used for two purposes: 1. It does external work. 2. It increases the intrinsic energy of the substances".

From the above quotations it will be evident that the reversible compression and expansion of gases and liquids is regarded in classical thermodynamics as exemplifying processes in which a transformation of work into heat and of heat into work takes place. In a large number of the cases quoted it is without doubt that the reciprocal transformation of heat and work is imagined to be directly associated with the adiabatic change of volume, i. e. that heat is imagined to be created or consumed within the working system itself communicating with the surroundings only as regards volume. It is positively stated that it is the change of temperature which is an indicator of the supposed production or consumption of heat in such systems. HELMHOLTZ for instance states clearly that the consumption of heat manifests itself through the fall in the temperature of the expanding gas. In his paper of 1847²⁾ he also expresses a doubt as to the correctness of CLAPEYRON's law of the relation between the heat of compression and the thermal coefficient of expansion³⁾, obviously arguing that if the "force" is to be maintained, the work of compression must result in a positive evolution of heat and, consequently, in a positive rise of temperature.

1) R. A. HOUSTOUN, "An Introduction to Mathematical Physics", 170 (1920).

2) HELMHOLTZ: "Über die Erhaltung der Kraft", Ostwalds Klassiker No. 1, p. 27.

3) E. CLAPEYRON, Journ. de l'Ecole polytechnique, 14, 23. Cahier 153 (1834).

In other cases quoted above the experiment in question is carried out isothermally the heat phenomenon being recorded in a heat reservoir communicating with the reacting gas. Whether this modification of the experiment is regarded by the authors as a necessary condition for the heat manifestation is not clear. It is natural, however, to interpret the superaddition of a heat reservoir merely as an attempt to record and, if possible, measure, by means of an ordinary calorimetric arrangement, an amount of heat, which is also in the absence of this means considered as generated or consumed. It would certainly be quite irrational to assume that a neutral passage of heat between the system and the heat reservoir should be able in any way to influence the contingency of a heat generation. That such a view, however, is actually to be found in the thermodynamic literature appears from a recent paper by F. H. MACDOUGALL¹⁾, who holds the opinion that the heat generation in a system can be generally identified with the amount of heat given off to the surroundings by the system during the process. On this view, which is considered by its author to be in harmony with traditional thermodynamics, no generation of heat by friction, electrical discharge or combustion etc. should take place, if only the system in question were adiabatically isolated from the surroundings!

With regard to the magnitude of the amount of heat produced or consumed a general and clear conception only seems to exist in the case of the system being a perfect gas. In that case the heat evolved is identified with the external mechanical loss of work. If this be designated as δA , and if δQ_a be the amount of heat produced:

1) F. H. MACDOUGALL, Journ. Phys. Chem. **44**, 713 (1940).

$$\delta A = \delta Q_a \quad (1)$$

in conformity with the principle of the equivalence of heat and work. In the case considered the evolution of heat equals the isothermal emission of heat δQ_i :

$$\delta Q_a = \delta Q_i, \quad (2)$$

whence:

$$\delta A = \delta Q_i. \quad (3)$$

It is the latter of these equations which has been verified by experiment, both the terms involved being completely energetically defined.

If the system is not a perfect gas one should expect the above-mentioned argument by HELMHOLTZ to involve as a logical consequence that the heat produced should be equivalent to the work of compression. Experience, however, shows only too plainly that for imperfect systems equation (3) is no longer satisfied. For liquids it will generally be very far from satisfied, δA and δQ_i during the compression having even in certain cases different signs, while during the expansion, at pressure <0 , such a deviation as regards the sign is almost universal. Classical thermodynamics therefore does not draw the conclusion mentioned above. It would be natural to investigate, on the basis of the classical argument, whether the invalidity of Equation (3) is due to the invalidity of (1) or (2) or possibly to that of both these equations. This problem, however, does not appear to be explicitly formulated in the classical considerations. In spite of unanimity with regard to the existence of a generation of heat in the case of volume changes in general, the fundamental question of its magnitude has not been subjected to investigation. On the other hand a conventional practice has been established in this

domain which in reality accepts equation (2) and rejects equation (1). This means that equivalence between heat and work is no longer claimed, since it is believed that the heat produced by the compression of an imperfect gas or liquid does not equal the work of compression, but equals the heat given off if the same infinitesimal change of volume were to take place isothermally. It is not improbable that this convention has been established owing to a confusion of the concepts of heat production and heat communication, as is obviously the case in the arguments of F. H. MACDOUGALL, quoted above.

This conventional determination of the generation of heat by compression, however, leads to a decided dilemma. If with HELMHOLTZ and the other classical authors we consider an expenditure of heat in the perfect expanding gas as the genetic equivalent of the external mechanical gain of work, measured for instance by the displacement of a weight perpendicular to the earth's surface, and if analogously we consider an evolution of heat as the genetic equivalent of a mechanical loss of work of the same kind by the compression, without introducing in any of the cases the spatial work as part of the transformation, we cannot, if the system under consideration deviates only infinitesimally from ideality, give an entirely different interpretation of the process. If the spatial work is excepted from the calculation in the case of a perfect system, it cannot possibly enter into it if the system is imperfect. But if this be so, an equivalent for the amount of work, positive or negative, by which the external work exceeds the expenditure of heat is obviously not available. The above-mentioned conventional estimation of the heat of compression and expansion in non-ideal systems is therefore contradictory

to the preceding classical interpretation of the spatial process and must virtually, logically analysed, be described as constituting a rupture of the principle of the equivalence between heat and work.

Classical thermodynamics makes no attempt to solve this difficulty, but evades it by introducing a new concept which represents neither work nor heat nor any other defined form of energy. Having accepted as a fact that the two magnitudes involved in (3) are no longer equivalent, the measures taken in an attempt to restore the balance consist in introducing a third term dE , writing:

$$dE = \delta A - \delta Q \quad (4)$$

and ascribing to this new term such a value that (4) is satisfied. This value is designated as the increase in internal energy of the reacting system.

It is beyond the scope of the present paper to give a detailed discussion of the content and significance of equation (4) and the concept of energy defined thereby. We may find it appropriate to describe work and heat as forms of energy but, as pointed out elsewhere,¹⁾ energy, even according to classical conceptions, may suffer alterations from other sources than through the communication of work and heat. What concerns us in the present connection is the fact that this new concept does not contribute to the solution of the problem under consideration here. This problem consists in establishing the general principles governing the transformations of the various forms of energy which are all of them already defined by strict and unambiguous definitions. The summary and unspecified character of the concept

¹⁾ BRØNSTED, D. Kgl. Danske Vidensk. Selskab, Math.-fys. Medd. XVI, No. 10, p. 50 (1939).

of internal energy makes it useless as a means of interpreting natural phenomena and establishing these general principles.

The difficulties which seem to surround the classical treatment of the simple spatial process may be ascribed to several reasons. In the first place the correlation between the change of temperature and the amount of heat in a system with the character of a calorimeter has been transferred quite automatically and without any theoretical consideration to systems of a different character. The intimate connection between the change of temperature and the supply of heat in a system of calorimetric character is due solely to the fact that entropy is the only independently variable quantity involved in the experiment. If other quantities, such as volume, are also made independently variable, the change of temperature will of course no longer depend on the supplied heat alone. Furthermore not only temperature, but also any other potential in the system will depend upon the communicated heat. The classical method of calculating heat is therefore irrational, partly because it is irrational to give precedence to temperature as compared with other quite analogous energetic potentials, partly because it is irrational and unjustifiable to transfer a correlation which is valid for simple cases to cases of a composite nature. The establishment of such a practice, not only in the case of the spatial process considered here, but quite generally, has been made possible only by the fact previously pointed out that the concept of heat generation is used in classical thermodynamics without any clear definition.

A second reason for the obscurity of the classical interpretation of the spatial basic process is due to the uncertainty with which the various types of work involved in

the process are recorded and enumerated. The compression of a gas requires some mechanical arrangement, a work reservoir, in which work is lost, for example by the lowering of a load. By the compression a volume is transferred from a higher to a lower pressure. With the latter process, however, the traditional treatment does not seem to associate any work, or rather, by virtue of a quite obscure reasoning, seems to mix it up with the work in the reservoir into a single "work of compression", the sign of which is hardly specified. Thus it is made possible to introduce heat as an equivalent for the work of compression and to clear the way for the theory of the reversible heat evolution. The obscurity of the argument actually excludes a pertinent discussion and is strongly contrasted with the simple and straightforward interpretation of the energetic theory.

Finally we shall have to consider the application of the theory of "compensations" to the case of the spatial processes. MAYER and JOULE, as already stated, believed in an unrestricted reciprocal transformation of heat and work. Later, it was enunciated by CLAUSIUS and WILLIAM THOMSON that although heat may be transformed into work, a parallel "compensating" reaction is required to make this transformation possible. The impossibility of the uncompensated transformation of heat has been formulated by OSTWALD as the impossibility of a perpetual motion machine of the second order¹⁾. With regard to the spatial processes considered here the majority of authors, however, do not make reference to any proviso concerning the heat-work transformation. It seems that such reference has only been made unambiguously by EUCKEN who, in dealing with a perfect gas, declares the volume change of the gas to be the com-

1) W. OSTWALD, Lehrbuch der allgemeinen Chemie II. I. p. 478.

compensating process: "Der Nebeneffekt, der mit diesem Vorgange verknüpft ist, besteht hier in der Volumenänderung des Gases". Strangely enough EUCKEN describes this compensation as "harmless" ("unschädlich"), an expression which is not immediately intelligible. Beyond this only very few authors reveal any kind of recognition of the necessity of a compensating reaction. For example, when PLANCK states that the heat given off by the expansion of a perfect gas is completely transformed into work, "ohne dass sonst irgendwo ein Energieumsatz stattgefunden hat", then this very positive assertion must be supposed to imply that PLANCK—like EUCKEN—though regarding the expansion of the gas as a compensating reaction, does not consider this worthy of any special mention, because the process takes place "without any energy transformation whatsoever", and possibly, for that reason, may be considered as "harmless".

This assumption, that the proof of the possibility of the transformation of heat into work—or more precisely: the transformation of heat from a single heat reservoir into work in an external work reservoir—can be based on the "harmlessness" of the compensating process, or, according to PLANCK, on the fact that no transformation of energy takes place in the compensating gas expansion, must, however, be characterised as illusory. What, in fact, is the substance of the contention that no transformation of energy takes place? Obviously it cannot be meant to express that the energy in the expanding gas does not change, for if heat is transformed into work in accordance with the equivalence principle, the change of energy in the compensating reaction must always be zero, if conflict with the energy principle is to be avoided. The meaning, then, apparently can only be that no transformation of energy of one

form into energy of another form takes place in the expansion. But, if the volume of the gas is increased by dv , this volume being taken for instance from a surrounding vacuum, then if the pressure is p , the system, not only according to the present theory, but also according to classical conceptions, will lose an amount of energy of the magnitude $p dv$ and of the particular form which is determined by the nature of the quantity, i. e. volume or spatial energy. The total energy in the compensating system having undergone no change, energy in another form must have been created simultaneously. Of this form there can be no doubt. In fact, since the compensating gas expansion is irreversible, entropy according to all stand-points has increased by a certain amount dS , which is tantamount to an increase of energy for the system of $T dS$. The compensating process is therefore clearly associated with a "transformation of energy", i. e. a loss of one form of energy and a gain of another. The expression for this transformation in the compensating process:

$$p \delta v \rightarrow T \delta S,$$

in conjunction with the expression for the main process:

$$T \delta S \rightarrow \delta A,$$

where δA is the external gain of work, shows that the term $T \delta S$ disappears in the total reaction, and that for describing this process there only remains the expression for the work principle:

$$p \delta v \rightarrow \delta A.$$

In these considerations we have been using for the products $p dv$ and $T dS$ the term "energy". This cautious mode of expression has been resorted to in order to avoid of-

fending against traditional phraseology and thus possibly diverting the reader's attention from the essential points in the argument. As $p dv$ is associated with the basic process of displacing a volume from one pressure to another and the factor dS in the product TdS is entropy generated by the irreversible expansion, the two products are of course according to the terminology of the energetic theory describable as work and heat respectively. Regardless of any terminology, however, the above reasoning shows conclusively that the spatial process considered is interpretable solely by the work principle of the energetic theory. The classical contention that heat enters into the energy transformation of the spatial process is therefore wholly unfounded. This result naturally applies no less to those formulations that entirely ignore the claim for compensations than to those which accept their presence without drawing the conclusions which they logically entail.

Since the processes here considered are reversible, "compensations" in the heat-work transformation will of course require "compensations" in the reverse process also. There would be reasons, therefore, to distinguish between two kinds of compensations, positive and negative, respectively. It is clear that the "negative" compensations would prevent heat from entering into the process of compression, just as the "positive" compensations, according to the proof given above, prevent heat from entering into the process of expansion. Thus we are obliged, even when following the classical views as closely as possible, entirely to exclude the phenomenon of heat from the reversible spatial process. If a transformation of heat into work were really to form part of these processes, then the presence of compensations would come into conflict with the generally accepted prin-

ciple that work may be transformed into heat without reservation.

Finally it should be mentioned that such phenomena as the stretching of a wire and changes of surface areas are closely related to the spatial process in imperfect systems. These phenomena have played an essential part in thermodynamic considerations but will not be discussed in the present paper, as their treatment does not offer any new aspect of special importance.

3. The Chemical Process.

The chemical basic process comprises not only the chemical reactions proper, which are characterised by certain fixed numbers of conversion, but all processes in which various forms of substance are transferred from one state to another. Even the transport of a substance from a pure state under given conditions of temperature and pressure to a pure state under other conditions of temperature and pressure is a chemical process according to this terminology. While the spatial and thermal processes involve only a single quantity, volume and entropy respectively, the chemical phenomena are more varied, as every chemical component has to be regarded as a separate quantity. It is not necessary that a chemical component should be what chemists describe as a "chemical individual". It may be an arbitrary group of substances capable of entering into the structure of the system. In principle, the amount of a component may enter also with a negative sign in a chemical component transformation. Further it should be noted that the chemical component is generally associated with the spatial and thermal quantities, i. e. with volume and entropy. The chemical basic process, which is defined by the nature and

amount of the transported component as well as by the chemical potentials in the states involved, is therefore in general associated with a simultaneous spatial and thermal basic process. The formal complications which might result herefrom in the energetic treatment are generally eliminated by a conventional definition of the composition of these quantity assemblies¹⁾. Even without such conventions they will disappear in the case of changes taking place at temperature and pressure homogeneity.

Finally may be mentioned another difference, which makes the treatment of the chemical process, as contrasted with the spatial process, more complicated. While only a single quantity, that is entropy, was admitted in the latter process as a neutral addition to the reacting system, a similar access will be generally supposed to exist both for entropy and volume in the treatment of the chemical process.

Just like the basic process in general the chemical process also can be differential or integral. The simplest differential chemical process consists in the transport of a single component between two different states, the amounts involved being of such dimensions that only infinitesimal changes in these states ensue; whereas the corresponding integral process is defined as a process of such extent that finite changes in these states take place even if these changes affect only infinitesimal fractions of the total system. When chemical processes are included it is particularly striking that a differential process may be carried out by way of an integral mechanism. It will be possible, for example, to substitute the differential basic process in which a given

¹⁾ Cf. W. GIBBS, Collected Works I. BRØNSTED, D. Kgl. Danske Vidensk. Selskab, Math.-fys. Medd. XV, No. 4, p. 18 (1937).

infinitesimal volume δv of a gas is transferred from the pressure p_1 to the pressure p_2 by an integral chemical process in which the gas contained in the volume dv is transferred to the gas of pressure p_1 , the pressure in δv falling to zero, and then gas from the pressure p_2 is transferred to dv , the pressure in this volume rising from zero to p_2 . However, this does not lead to any uncertainty regarding the concept of differential change, since a process which is differential with respect to a certain quantity, cannot also be differential with respect to another quantity.

According to the energetic theory an irreversible chemical process, like any other irreversible process, will cause an energetic evolution of heat. This applies for instance to the evaporation of a liquid in non-equilibrium with its vapour, or to the dissolution of a salt in water under decrease in temperature. On the other hand no caloric phenomenon will be associated with the reversible chemical process according to this theory. The traditional conception differs entirely from this view. For instance, with the reversible change of states it generally associates a caloric phenomenon i. e. a production or consumption of heat. CAVENDISH¹⁾, already, speaks of a generation of heat as being associated with the condensation of vapour, and recently W. JÄGER²⁾ as examples of processes in which other forms of energy are transformed into heat mentions changes in the state of aggregation and allotropic transformation and dissociation. This caloric effect is usually spoken of as "latent heat". Actually the distinction made between latent heat effects at constant temperature and reversible processes

¹⁾ Cf. J. CL. MAXWELL, "Theory of Heat", 5th Ed. 72 (1877). MAXWELL himself apparently did not share this view of a generation of heat in these processes.

²⁾ W. JÄGER, "Handbuch der Physik" (Geiger u. Scheel), IX, 474 (1926).

in which the temperature varies is unjustified, and all these reversible processes should logically be included in the general traditional idea relating heat generation to change of temperature. Precisely in the domain of chemistry, through the conventions and practice of thermochemistry, this idea of a simple correlation between heat and temperature has been demonstrated as particularly obvious. In the conventional language "evolution" and "absorption" of heat is the amount of heat which in the chemical process must be removed from or transferred to the reacting system, in order that its temperature may remain constant during the experiment. If the course of the reaction is irreversible, and also partly in the reversible case, these magnitudes are jointly designated as "Varmetoning"¹⁾, a term introduced by JULIUS THOMSEN. As distinct from energetic heat evolution the classical heat phenomenon has been described in a previous paper as thermometric²⁾.

We shall first examine the reversible chemical process, which in its simplest form consists in the transport of a substance from one state to another. If the chemical potential of the component in the initial and final states is μ_1 and μ_2 respectively and δn the amount of substance transported, then the loss of work in the process will be:

$$\delta A_{\text{ch}} = (\mu_1 - \mu_2) \delta n, \quad (5)$$

the same amount of work being gained simultaneously in an annexed work reservoir. Processes of this kind can be carried out in a particularly simple manner by means of

1) JULIUS THOMSEN, Termokemiske Resultater, 2 (1905). Thermochemischen Untersuchungen I, 5 (1882). In German: Wärmetönung. For this word there seems to be no adequate equivalent in English or French.

2) BRØNSTED, D. Kgl. Danske Vidensk. Selskab, Math.-fys. Medd., XV, No. IV, p. 39 (1937).

a galvanic concentration cell in which the transformation under consideration is electromotively active. In principle, however, the electromotive method may be used also more generally for determining chemical work. If the cell serves for differential charging of a condenser the potential of which is equal to that of the cell, the chemical work may be said to be transformed into electric work. Also, if the electric work is only intermediate, being a link in the production of mechanical or chemical forms of work or even in the irreversible heat evolution, the electric gain of work may be regarded as directly coupled with the chemical loss of work, so that it will be possible in all instances to describe adequately the reversible function of the cell as a chemical-electrical work-transformation. What is essential is of course only the fact that the reversible process can be interpreted as a transformation of chemical work into external work of some form or other depending on the experimental conditions.

In conformity with thermodynamic practice the chemical system which constitutes the cell is considered as closed with regard to chemical components. Whether it is otherwise closed as regards quantity, or whether during the differential process under consideration it is in quantity communication with the surroundings is without significance for the above interpretation. The supply of entropy or of volume in infinitesimal amounts from outside will not cause any alteration of the nature of the infinitesimal transformation of work, and the corresponding neutral energy terms $T\delta S$ og $p\delta v$ will not have any share whatsoever in the external gain of work. The latter is in all cases determined by the chemical loss of work alone in conformity with equation (II. 1). It

should be specially noted that for this reason it will not be possible to find in energy or in other thermodynamic functions comprising such energy terms a universal basis for the determinations of chemical work or the corresponding electromotive force. Before entering upon the discussion of such attempts, however, it will be appropriate to consider the application of the classical views in the particularly simple instance of the substance transported in the chemical basic process obeying the simple gas laws. It is possible, in various ways, to make this transport electromotively active in a reversible concentration cell. If the cell functions isothermally, in communication with a work reservoir and a heat reservoir of the temperature of the cell, an amount of heat numerically equivalent to the accumulated work will be removed from the heat reservoir. The whole arrangement will thus be quite analogous to an isothermal and reversible expansion of air¹⁾, and the classical ideas, if consistently maintained, should lead to identical interpretations of the two phenomena. In conformity with the above considerations concerning the spatial process, the function of the cell should therefore be interpreted as a transformation of heat from the heat reservoir into work in the work reservoir, this imaginary process being rendered possible through the simultaneous, compensating cell-reaction consisting in an irreversible transference of substance from a higher to a lower chemical potential. This consistent interpretation, however, is not to be found in the traditional treatment. An interpretation, if given at all, is generally to the effect that a "work of dilution" associated with the process of the cell appears as electric energy. This work

¹⁾ This analogy was first pointed out by W. GIBBS, *Collected Works* I, 339 (1928). Cf. also V. v. TÜRIN, *Z. physik. Ch.* 5 343 (1890).

of dilution is identical with the chemical loss of work in the energetic theory.

When dealing with such chemical processes of a more complex nature as take place in the ordinary "chemical cells", the situation is much more complicated. Although the treatment of this subject is extremely varied in classical thermodynamics, it may be said to possess the general characteristic that heat, explicitly or implicitly, is supposed to enter as a partial genetic equivalent of the electric gain of work. This view, again, has developed from the earlier attempts to interpret the function of the cell on the basis of its irreversible thermometric heat evolution.

According to the so-called THOMSON rule¹⁾, which actually was developed already by JOULE²⁾ and by HELMHOLTZ³⁾, the electric energy supplied by a galvanic cell is measured by what is now called the thermometric heat evolution. JULIUS THOMSEN⁴⁾, for instance, carried out experiments to test the validity of this rule. Its incorrectness was fully recognised in later theoretical investigations especially by HELMHOLTZ⁵⁾, and appears plainly from the energy equation for the reversible cell process:

$$\delta A_{\text{ch}} = T dS - p dv - dE, \quad (6)$$

expressing that the total or internal energy of the system is altered partly by virtue of the chemical loss of work δA_{ch}

1) WILLIAM THOMSON, Phil. Mag. [4] **2**, 429 (1851).

2) JOULE, Phil. Mag. **19**, 260 (1841), **20**, 92 (1842), "Scientific Papers" 60, 81 (1884).

3) HELMHOLTZ, "Über die Erhaltung der Kraft", 1847, Ostw. Klass. No. 1. p. 34 (1915).

4) JULIUS THOMSEN, D. Kgl. Danske Vidensk. Selskab, Skr., nat. og math. Afd. 5. V, p. 155 (1861), Ann. d. Physik u. Chemie N. F. **11**, 246 (1880).

5) HELMHOLTZ, Sitzungsber. d. Berl. Akad. 1882 p. 22, Ges. Abh. II. 598.

and partly by virtue of the amounts of entropy and volume taken up from outside sources. As the irreversible thermometric loss of heat is given by:

$$-T dS' = \delta Q' = -(dE + p dv), \quad (7)$$

(6) may also be written:

$$\delta A_{\text{ch}} = T dS + \delta Q' = T dS''. \quad (8)$$

It will be seen that the chemical loss of work and the equivalent external electric gain of work in the reversible cell function equals the thermometric loss of heat, if under the given conditions the process takes place without any change of entropy, and that this loss of work is equal to the loss of energy, if the volume also remains constant. As the JOULE-THOMSON rule involves constancy of temperature and pressure, it will therefore ordinarily be valid only in the quite special cases where these conditions coincide.

As to equation (6) it is evident that it does not give, nor can it give, any contribution to the solution of the problem as to how the function of the galvanic cell should be interpreted, since by introducing the unspecified concept of internal or total energy, the distinction between the different forms of energy, which is the requisite condition of such an interpretation, is abandoned. The equation only serves to represent a numerical correlation between amounts of energy, without considering the presence of any genetic relation. It may be said to pertain to the realm of functional energetics or energetics of states, as contrasted with energetics of processes.

The same argument applies to the relation between chemical work and another important function, the ψ -function of GIBBS, which was also independently established by HELMHOLTZ in his paper of 1882, and was called by him

the free energy and denoted by F ¹⁾. This function is defined by the equation:

$$F = E - TS. \quad (9)$$

According to HELMHOLTZ it is an expression for the internal "work store" of the system, its ability to carry out work at constant temperature. Comparison of equations (6) and (9), differentiated at constant temperature, however gives:

$$dF = -p dv - \delta A_{\text{ch}}, \quad (10)$$

from which it appears that the change in free energy comprises not only the chemical work δA_{ch} , but also a term $p dv$, which does not possess the character of work as given by (5) and is actually intransformable energy. Only if the external pressure is zero, will the term $p dv$ possess an active form, capable of representing work, but this particular condition is not involved in the characterisation of the free energy; on the contrary, the free energy is generally used as a work function at pressure and temperature homogeneity. The GIBBS ζ - or G-function also, defined by:

$$G = E - TS + pv, \quad (11)$$

will only under special conditions measure the chemical work, namely when the pressure and temperature are constant within the system. The question of the free energy and the relation of the other thermodynamic functions to the heat-work concept will, however, be discussed elsewhere, and we shall here only emphasise the fact pointed out in Section II.1 that all terms of the form PdK represent neutral energy, i. e. energy which is not transformable and to which no character of work can be ascribed. That the term $p dv$ is generally small in galvanic processes and

¹⁾ Later HELMHOLTZ in his lectures exchanged the symbol F for the symbol H . "Vorlesungen über theoretische Physik" **6**, 267 (1903).

practically without significance has of course theoretically no influence on the argument.

If we introduce this condition in (6), i. e. if we assume $p dv$ to be negligible, the isothermal loss of energy and the thermometric loss of heat are equal and accordingly (6) and (8) will coincide. It is generally under this simplifying condition that classical thermodynamics, after recognising the incorrectness of the JOULE-THOMSON rule, looks for an interpretation of the chemical process in the galvanic cell, attempting to specify the forms of energy from which the electric work of the cell originates. In this way, in conflict with the views of the energetic theory, heat is introduced as a partial equivalent for the electric work in the reversible function of the cell.

According to the generally adopted view the electric work is gained at the cost of the "chemical energy" of the cell, this energy being "converted" into work and heat. The concept of "chemical energy" is not clearly defined, but its magnitude, according to thermodynamic practice and terminology, is fixed by the thermometric heat evolution at the irreversible cell process, or eventually by the total isothermal change of energy¹⁾. In the reversible process this "chemical energy" is supposed to be transformed into electric work. The electric work, however, does not represent the total loss of energy in the cell, since the two amounts need not be equal. Their difference is supposed to appear as "heat". Here two possibilities are present, the "heat" appearing either adiabatically in a closed system or isothermally in communication with the surroundings. Thus "chemical energy" is imagined to be present in the cell and to be transformed into external electric work in the

1) Cf. e. g. W. NERNST, "Theoretische Chemie", 8.—10. Aufl., 816 (1921).

cell process, and, in case the difference between chemical energy and work is positive, into heat as well. If the difference is negative, the chemical energy available will not suffice as an equivalent for the work, and in that case the surplus of work is supposed to be formed from "heat". The interpretation is thus quite similar to that employed in the case of the spatial process, a heat-work transformation being supposed to take place in spite of the reversibility of the process, and a certain form of energy, described as "chemical" but without definition, being involved in the interpretation.

These ideas, which are still maintained even in modern expositions of chemical thermodynamics, can be traced back to the earliest attempts to arrive at a more correct understanding of the electric cell phenomena than was given by the JOULE-THOMSON rule. Among these, investigations by F. BRAUN¹⁾ occupy a prominent place. According to this author the electric work can be created at the expense of an equivalent amount of heat resulting from the chemical process in the cell. The chemical energy supplied by the process is of the same form as heat. Only part of this energy, however, is ordinarily transformed into electric work. The rest of the "chemical heat" remains in the cell. A characteristic feature of this reasoning is the introduction of the concept of "electromotive efficiency" ("Nutzeffekt"), to designate the ratio between the work gained and the heat evolved, work being quite generally identified with "die von dem betr. chemischen Process in elektrische Energie, d. h. in mechanische Arbeit verwandelbare Wärmemenge". There are cases, however, in which the electric

¹⁾ F. BRAUN, *Ann. d. Physik u. Chemie*, N. F. **5**, 182 (1878), **16**, 561 (1882), **17**, 632 (1882).

work is greater than the heat evolution, and accordingly, the efficiency is greater than unity. In such cases the view advanced by BRAUN leads to the conclusion that work is created from heat consumed in the interior of the chemical system.

The investigations of S. CZAPSKI¹⁾, which have also played an essential part in the verification of the equations of HELMHOLTZ, as a main result lead to the conclusion "dass diejenige Elemente, die nicht alle chemische Wärme in Stromarbeit umsetzen, eine mit wachsender Temperatur abnehmende elektromotorische Kraft haben, und umgekehrt diejenigen, welche zum Teil auf Kosten ihres eigenen Wärmehalts arbeiten, eine mit der Temperatur wachsende Kraft". This statement is based on the same view that the production of work in the chemical process is associated with and dependent upon an expenditure of heat. Similarly A. GOCKEL²⁾ speaks about the "fraction of heat in the chemical process which is transformed into electric work" and about the "amounts of heat appearing or disappearing at the poles of the cell".

Ideas such as those described here have remained unchanged right down to the present time and are still the basis of authoritative expositions of the thermodynamics of chemical reactions³⁾. They are sharply contrasted with the theory of the chemical basic process advanced in the introduction of this section, according to which heat is excluded from any participation in reversible processes.

Closely related to the chemical processes in reversible galvanic cells are the phenomena appearing at the passage

1) S. CZAPSKI, Ann. d. Physik u. Chemie, N. F. **21**, 209 (1884).

2) A. GOCKEL, Ann. d. Physik u. Chemie N. F. **24**, 618 (1885).

3) Cf. C. DRUCKER, "Handbuch der Experimentalphysik" (Wien-Harms) **12**. 2. Teil, p. 9 (1933).

of an electric current through the junctions of metallic conductors, as we are here dealing with the transference of electric charges from one medium to another. HELMHOLTZ¹⁾, already, supposed that a direct relation existed in these phenomena between the electric energy and heat, and extended this view to include also the thermoelectric Thomson-effect²⁾. He regards the reversible passage of current as heat-generating, so that in these combinations electric energy is transformed into heat or vice versa. This view, which is also held by WILLIAM THOMSON³⁾, is adhered to in modern expositions⁴⁾, the usual conclusion being drawn from temperature changes to caloric effects. In this, it is overlooked that a thermoelectric cell is not a system of calorimetric character. This question, however, will not be discussed any further here, since the energetics of thermoelectric phenomena will be subjected to a more detailed treatment in another connection.

The energetic phenomena connected with the chemical basic process have been studied by numerous other methods besides the electrochemical method specially dealt with above. It is generally found that the classical treatment introduces caloric effects, i. e. heat evolution and heat absorption, in processes of a reversible course, such as phase transitions, processes of dissociation and chemical reactions in general, regardless of the fact that these phenomena imply certain conditions, for instance constancy of temper-

1) HELMHOLTZ, "Über die Erhaltung der Kraft". Ostw. Klass. No. 1, p. 43.

2) HELMHOLTZ, "Vorlesungen über die theoretische Physik" VI, 263 (1903).

3) WILLIAM THOMSON, "Mathematical and Physical Papers" I, 246 (1882).

4) Cf. K. BAEDECKER, "Die elektrische Erscheinungen in metallischen Leitern", p. 77 (1911). WEINSTEIN, "Thermodynamik und Kinetik" III, 348 (1905). NERNST, "Theoretische Chemie", 10. Aufl., 814 (1921). E. LANGE, "Handbuch der Experimentalphysik" (Wien-Harms) 12, 2. Teil, 327 (1933).

ature for the course considered. In the majority of cases it must be presumed that an actual heat-work transformation is presupposed, as for instance when it is claimed that the heat of combustion of carbon may be almost completely transformed into external work¹⁾, or that heat is consumed in the process of chemical dissociation²⁾. The heat involved here is always thermometric heat effects which do not enter into any energy transformation but are merely of communicative significance. In agreement with the main aspect of the classical theory the irreversible mixing of perfect gases is assumed to take place without any evolution of heat, because it is accompanied by no temperature change. These ideas, however, are so widely disseminated and familiar that a particular documentation concerning such phenomena may presumably be omitted.

4. The Thermal Process.

The thermal basic process consists in the passage of entropy from one temperature to another. If the amount of entropy δS passes from a heat reservoir R_1 of the temperature T_1 to a heat reservoir R_2 of the temperature T_2 , a thermal work:

$$\delta A_{\text{th}} = (T_1 - T_2) \delta S \quad (12)$$

is lost, and, in case of reversibility an equivalent amount of work of another kind, for instance mechanical, is gained. This description of the thermal process is completely exhaustive, since no independent phenomena other than those mentioned participate in the process. It is a principal point in the energetic theory that no caloric phenomenon is comprehended in the reversible process. No loss

1) W. NERNST, "Theoretische Chemie", 5^{te} Aufl. p. 698 (1907).

2) Ibidem, p. 602.

of heat therefore will take place when the heat engine is operating, an equivalent for the mechanical work gained being provided solely by the entropy transport.

The concept of a thermal basic process does not form part of classical thermodynamics, its ideas in this field being embodied in the theory of the heat engine. This theory is so familiar and generally accepted that further documentation beyond that given in previous papers¹⁾ will not be needed. We may therefore directly proceed to an analysis of the classical theory as presented by CLAUDIUS.

By way of introduction it should be pointed out that the contention of CLAUDIUS of a heat-work transformation in the function of the heat engine is based on the idea that heat is motion of the elementary particles of matter. "Wenn dieses richtig ist", he says²⁾, "so muss sich auf die Wärme auch der allgemeine Satz der Mechanik anwenden lassen, dass eine vorhandene Bewegung sich in Arbeit umsetzen kann und zwar so, dass der Verlust an lebendiger Kraft der geleisteten Arbeit proportional ist". It is claimed therefore "dass zur Erzeugung von Arbeit, nicht bloss eine Änderung in der Verteilung der Wärme sondern auch ein wirklicher Verbrauch von Wärme nötig sey". This necessity of an expenditure of heat in the production of work is emphatically insisted upon in the classical theory and on the whole appears as the new idea by means of which the difficulties connected with CARNOT's view³⁾ might be duly removed.

That the "general laws of mechanics" do not take us far towards the solution of the problem of heat-work rela-

1) J. N. BRØNSTED, D. Kgl. Danske Vidensk. Selskab, Math.-fys. Medd. XV, No. 4, p. 47 (1937).

2) R. CLAUDIUS, Pogg. Ann. 79, 370 (1850).

3) Cf. WILLIAM THOMSON, Edinb. Transact. 16, 541 (1849).

tions is however a fact that appears clearly from CLAUSIUS' own treatment. Reference to these laws therefore seems rather pointless or even misleading particularly if it be kept in mind that the procedure of CLAUSIUS is otherwise based exclusively on what is described in the present paper as the macroenergetic view, entirely disregarding existing theories of the micro-structure of matter.

If we make a record of the phenomena which are directly observable as a result of the function of the heat engine operating for example by means of a CARNOT cycle in a direction which we shall describe as positive, the issue is as follows:

1. The heat or entropy reservoir R_1 of the temperature T_1 gives off an amount of heat δQ_1 and an amount of entropy δS , interrelated by the equation:

$$\delta Q_1 = T_1 \delta S. \quad (13)$$

2. The heat or entropy reservoir R_2 of the temperature T_2 receives an amount of heat δQ_2 and an amount of entropy δS , interrelated by the equation:

$$\delta Q_2 = T_2 \delta S. \quad (14)$$

3. In a mechanical work reservoir an amount of work δA_m , given by:

$$\delta A_m = \delta Q_1 - \delta Q_2 \quad (15)$$

or:

$$\delta A_m = (T_1 - T_2) \delta S \quad (16)$$

is accumulated.

This is what may be phenomenologically established. Classical thermodynamics, however, does not content itself with this description, but looks for a genetic interpretation of the phenomena from a more complicated energetic standpoint. While the present theory views the process as a transport of entropy, a contention which

merely forms a record of the observed phenomena, the function, according to the traditional view of CLAUSIUS, consists in the transformation of an amount of heat $\delta Q_1 - \delta Q_2$ into work, the heat being taken for instance from the reservoir R_1 , and a supplementary transport of heat δQ_2 from R_1 to R_2 . Thus in addition to the heat-work transformation a parallel heat transport takes place, which is regarded as a compensation for the main process: "Die eine aus dem Körper K_1 stammende Wärmemenge Q ($= \delta Q_1 - \delta Q_2$) ist in Arbeit verwandelt, und die andere Wärmemenge Q_2 ($= \delta Q_2$) ist aus dem Körper K_1 in den kälteren Körper K_2 übergegangen"¹⁾. In the thermal process the presence of such a compensation has become particularly appreciated as a condition for the realisation of the heat-work-transformation.

If the direction of the cycle is reversed the work δA , will on this view be transformed into heat at the temperature T_1 , while simultaneously heat δQ_2 is led from T_2 to the higher temperature T_1 . It is the impossibility of the latter process as an isolated phenomenon which is established by CLAUSIUS as a special principle exceeding the sphere of the energy principle, and it is the simultaneous work-heat transformation which is introduced as a "compensation" that enables heat to pass from a lower to a higher temperature.

The thermal work relation has found another classical expression in the following well-known postulate by WILLIAM THOMSON²⁾: "It is impossible, by means of inanimate

¹⁾ R. CLAUSIUS: "Die mechanische Wärmetheorie" I, p. 75 (1876). Cf. also H. v. HELMHOLTZ, *Wiss. Abhandl.* **3**, 120 (1895), M. PLANCK, "Thermodynamik", 1st Ed., p. 60 (1897), and 9th Ed., p. 66 (1930).

²⁾ WILLIAM THOMSON, *Trans. Roy. Soc. Edinb.* 1851; *Phil. Mag.* 4 1852. "Mathematical and Physical Papers" I, 174, 511 (1882).

material agency, to derive mechanical effect from any portion of matter by cooling it below the temperature of the coldest of the surrounding objects". The formulation must be interpreted as expressing the impossibility of the heat-work transformation as an isolated phenomenon, while at the same time conceding the possibility of such a process when accompanied by compensating changes.

Of the two parts into which the reversible function of the heat engine is traditionally divided the main process, i. e. the transformation of heat into work, is imaginary, while the compensating heat transport is irreversible. The imaginary processes are not known to have been subjected to special consideration in classical thermodynamics, but the irreversible processes are well characterised and familiar phenomena. It is natural therefore to commence our analysis by considering the irreversible heat transport.

As stated above, an irreversible process can always be described as a basic process associated with the formation of entropy. If the entropy ΔS is generated in a reservoir of calorimetric character at temperature T , the amount of heat $T\Delta S$ formed simultaneously can be determined by measurements on the reservoir before and after the process. If, however, the irreversible process is given merely by the initial and final states of the bodies involved, in which case we shall say that the process is terminally defined, the heat evolution will not be fixed. For the heat, equivalent to the work of the basic process considered, may have been formed in a reservoir at the higher temperature T' and from there have been conducted to the reservoir of temperature T , thus causing additional heat to be produced by the irreversible heat transport. The reason for this indetermin-

ation of the heat evolution is to be found in the fact that the entropy generated does not form part of the initial state and that the basic processes acting in the case of irreversibility are not therefore terminally established.

The heat transport is an irreversible process describable by the amount of heat Q passing from the reservoir R_1 to the reservoir R_2 and the temperatures T_1 and T_2 of these reservoirs. Owing to the general relation:

$$Q = T_1 S_1 = T_2 S_2 \quad (17)$$

the amount of entropy S_1 leaving R_1 and the amount of entropy S_2 arriving in R_2 during the heat transport are terminally determined. Furthermore the generation of entropy is fixed by:

$$\Delta S = S_2 - S_1 = Q \frac{T_1 - T_2}{T_1 T_2} \quad (18)$$

or

$$\Delta S = S_1 \frac{T_1 - T_2}{T_2} = S_2 \frac{T_1 - T_2}{T_1}. \quad (19)$$

This terminal definition of the increase of entropy exemplifies the general character of entropy as a function of state.

In accordance with the above statements concerning irreversible processes in general there are present, however, in the heat transport certain important magnitudes which are not given as functions of state. The same heat transport, describable by Q , T_1 and T_2 , may therefore be carried out by means of various basic processes. In all cases these consist in entropy transports. As contrasted with the entropy produced, however, the amount of entropy transported is not terminally fixed. Nor is this the case with the temperature of the entropy generation. But, as is easily seen, these two indeterminate magnitudes are not independent of each other.

If the entropy transported in the basic process is denoted by S and the temperature of formation for ΔS by T , then, by equation (II, 17), which expresses the equivalence principle, we may write:

$$(T_1 - T_2) S = T \Delta S, \quad (20)$$

$T \Delta S$ being the energetic heat evolution. (20) can be rewritten as:

$$S = \frac{T}{T_1 - T_2} \Delta S = \frac{T}{T_1 T_2} Q \quad (21)$$

or

$$S = S_1 \frac{T}{T_2} = S_2 \frac{T}{T_1}. \quad (22)$$

These equations express the correlation existing between the variables S and T . If T has been chosen $= T_1$, the entropy ΔS being produced in R_1 , the entropy of the basic process is given by:

$$S = \frac{T_1}{T_1 - T_2} \Delta S = S_2. \quad (23)$$

If on the other hand $T = T_2$, ΔS is produced in R_2 and:

$$S = \frac{T_2}{T_1 - T_2} \Delta S = S_1. \quad (24)$$

$T = T_1$ and $T = T_2$ represent two special instances of the realisation of the heat transport. But actually the heat transport can be accomplished in an infinite number of ways, since the generation of entropy is not confined to one of the temperatures T_1 or T_2 , but can be distributed over the two reservoirs R_1 and R_2 in arbitrary proportions. This method can be made use of for instance by introducing a third heat reservoir R of the temperature T , in which ΔS is produced. The process is then completed by transferring ΔS , in fractions $\Delta S'$ and $\Delta S''$ from R to the reservoirs R_1 and R_2 respectively. This transference should not involve any irreversible phenomenon, i. e. it must

consist of two basic processes in which loss and gain of work are equal. This condition can be expressed by:

$$\Delta S = \Delta S' + \Delta S''$$

$$(T_1 - T) \Delta S' = (T - T_2) \Delta S''$$

whence:

$$\Delta S' = \Delta S \frac{T - T_2}{T_1 - T_2} = S \frac{T - T_2}{T} \quad (25)$$

and

$$\Delta S'' = \Delta S \frac{T_1 - T}{T_1 - T_2} = S \frac{T_1 - T}{T} \quad (26)$$

the distribution ratio $\frac{x}{1-x}$, given by:

$$\frac{x}{1-x} = \frac{T - T_2}{T_1 - T} \quad (27)$$

being completely arbitrary. If the evolution and transference of entropy are continuous, the reservoir R may be given infinitesimal dimensions, so that its addition does not constitute any real change of the system. Obviously $\Delta S'$ and $\Delta S''$ can also be formed directly in R_1 and R_2 in the ratio given by x or T , without having to pass through an R-reservoir. In both cases on the analogy of (20) we obtain:

$$(T_1 - T_2) S = T_1 \Delta S' + T_2 \Delta S''$$

or for the entropy transport in dependence of x :

$$S = \frac{xT_1 + (1-x)T_2}{T_1 - T_2} \Delta S = \frac{xT_1 + (1-x)T_2}{T_1 T_2} Q. \quad (28)$$

The distribution of the entropy over the two temperatures T_1 and T_2 is thus genetically equivalent to its formation at the temperature T , if this temperature is correlated with the distribution ratio by equation (27).

If not the heat Q , but the entropy S is given in the process of heat transport, ΔQ , the energetic heat evolution, and not ΔS , will be directly defined. Since

$$\Delta Q = T \Delta S, \quad (29)$$

it will not be possible in this case, either, to calculate the temperature of the entropy formation. This is in conformity with the universal law that entropy evolution is determined by the total process and heat evolution by the basic process.

Combining (18) and (20) and (29) we may also write for the heat evolution:

$$\Delta Q = Q (T_1 - T_2) \frac{T}{T_1 T_2} = (T_1 - T_2) S = A_{th}, \quad (30)$$

where T is defined by (27) and A_{th} is the thermal loss of work.

Having thus analysed the auxiliary process of heat transport we shall now proceed to consider the imaginary main process. As previously mentioned, this process, according to CLAUSIUS, is a heat-work transformation in which work is gained at the expense of heat taken from a heat reservoir. The temperature T of this heat reservoir is not given. In the usual treatment one puts $T = T_1$ or $T = T_2$. We shall first, however, consider the general case of the temperature T being arbitrarily chosen. According to CLAUSIUS¹⁾ the heat expended is then calculated in the following way.

If in a reversible thermal process the amounts of heat Q_0 , Q_1 and Q_2 are given off by 3 heat reservoirs R_0 , R_1 and R_2 of the temperatures T_0 , T_1 and T_2 respectively, the relation:

$$\frac{Q_0}{T_0} + \frac{Q_1}{T_1} + \frac{Q_2}{T_2} = 0 \quad (31)$$

is generally valid. Putting $Q_1 = -Q_2 = Q$, the process, as regards R_1 and R_2 , may be considered a transport of the heat Q from T_1 to T_2 , while Q_0 , taken at the temperature T_0 , is being transformed into work. Equation (31) may then be written:

¹⁾ CLAUSIUS, "Die mechanische Wärmetheorie" I, 89, 97 (1876).

$$Q_0 = Q (T_1 - T_2) \frac{T_0}{T_1 T_2} = A_m. \quad (32)$$

The transference of the amount of heat Q from T_1 to T_2 will thus involve a transformation into mechanical work A_m of an amount of heat Q_0 given by (32) and consumed in a reservoir of temperature T_0 . If a special term is desired for this imaginary magnitude, Q_0 might appropriately be designated as the energetic heat consumption.

On the basis of the general expression (32) we obtain for the special cases of $T = T_1$ or T_2 , the special forms for the expenditure of heat:

$$Q_0 = Q \frac{T_1 - T_2}{T_2} \quad (33)$$

and

$$Q_0 = Q \frac{T_1 - T_2}{T_1} \quad (34)$$

respectively. These expressions indicate the amounts of heat expended in the production of work at T_1 and T_2 .

As regards the correlation between entropy and heat supplied to or removed from a heat reservoir, no distinction has been made by CLAUSIUS or other classical authors between the case in which entropy and heat are communicated and that in which these magnitudes are developed or consumed in the reservoir. Even if it is not explicitly stated, classical thermodynamics must therefore be supposed to maintain that there exists the same relation between heat and entropy in the case of the imaginary process under consideration as that universally accepted for the process of communication. The consumption of entropy S_0 corresponding to Q_0 in equations (32), (33) and (34) is then obtained on division of Q_0 by T_0 , T_1 and T_2 respectively. We obtain in all 3 cases:

$$S_0 = Q \frac{T_1 - T_2}{T_1 T_2}, \quad (35)$$

i. e. a value which is independent of the temperature at which the entropy is consumed. This expression is identical with the expression (18) for the evolution of entropy in the process of heat transport.

The result obtained may also be expressed in the following way. Since the total function of the heat engine involves no change of entropy, it is known beforehand that the expenditure of entropy in the imaginary process and the evolution of entropy in the compensation process must be identical. The entropy given by (18) will therefore be consumed concurrently with the amount of heat expressed by (32). This means that the relation between heat consumed and entropy which was supposed above to be accepted in the classical treatment, actually follows as a necessary consequence of the classical assumptions.

On the basis of these considerations the partial processes in the classical theory of the heat engine are thus seen to resolve themselves as follows:

The imaginary process is split into two partial processes, the first being a basic mechanical process representing the gain of work A_m , the second being a consumption of entropy S_0 taking place at an indeterminate temperature T_0 . The energetic consumption of heat Q_0 given by $Q_0 = T_0 S_0$, is synonymous with this expenditure of entropy.

Similarly the irreversible compensating process is split into two partial processes, the first being a basic thermal process consisting in a transport of the entropy S from T_1 to T_2 , and representing the loss of work A_{th} , the second being an evolution of entropy ΔS taking place at an indeterminate temperature T . The energetic heat evolu-

tion $\Delta Q = T\Delta S$ is synonymous with this evolution of entropy.

According to equation (32) the mechanical gain of work A_m is identical with the thermal loss of work A_{th} defined by equation (30), in perfect accordance with the demands of the energetic theory. In addition to these equivalent work phenomena which, jointly, give a complete account of the empirical function of the heat engine, the classical theory, according to the analysis offered, postulates the consumption of a certain amount of entropy in the main process and an evolution of the same amount of entropy in the compensating process. These changes, if occurring at unequal temperatures, would constitute a transport of entropy from one temperature to another and thus represent an amount of thermal work for which no energetic equivalent is available in the operation of the heat engine. Such a phenomenon would be at variance with the energy principle also in its traditional formulation. Accordingly the alleged formation and consumption of entropy must occur at the same temperature.

The recognition of this circumstance, however, entirely removes the basis of the classical interpretation, according to which the function of the heat engine consists in a heat-work transformation. If an expenditure of entropy of the same magnitude as a simultaneous evolution of entropy were to take place in different sections of the system but at the same temperature, then these processes would completely cancel and the attribution of any kind of reality to the separate phenomena would have no physical sense. Similarly one cannot attribute any reality to the concurrent phenomena of creation and consumption of the equal thermal energies $T\Delta S$

and $T_0\Delta S$, no matter what terminology is used for designating these energy terms. The classical doctrine that $T\Delta S$ is converted into work while at the same time $T\Delta S$ is generated as a consequence of the irreversible heat transport would be quite parallel to supposing that the turning of a loaded balance in a state of equilibrium might be interpreted as a process in which one of the loads were raised at the cost of an amount of heat supplied irreversibly by the lowering of the other. Assertions of this kind are physical absurdities which must be excluded from rational science. In accordance with the energetic theory we are therefore justified in contending that no heat is consumed in the operation of the heat engine and that the classical theory of the thermal production of mechanical work is based on a misinterpretation of experimental realities.

IV. Conclusion.

This concluding section contains a recapitulation of the more essential ideas to which we have been led through the present investigation. Some results arrived at or implied in the previous articles regarding the concepts of heat and work are also included.

1. The fundamental thermal concepts are defined on the basis of calorimetric phenomena. It is a requisite of the application of the calorimetric method that the various basic processes are interdistinguishable. A calorimeter, describable also as a heat or entropy reservoir, is a system the changes of which are due solely to its taking up or giving off entropy. A standard calorimeter serves to define amounts of entropy taken up or given off by it in communication with the surroundings.

2. Any irreversible process is conceivable as a transport of energetic quantities in connection with generation of entropy. If δK is the quantity transported and if it passes from a potential P_1 to a potential P_2 , these potentials being conjugate to the quantity considered, then the relation:

$$(P_1 - P_2) \delta K = T dS \quad (1)$$

where dS is the increase in entropy and T the absolute energetic temperature at which the entropy is generated, is found to be universally valid. This fundamental law of nature forms the basis for the creation of a new thermal concept TdS which is called heat. Since the generation of entropy is bound to irreversible changes and is always positive, the concept of heat is essentially that of a heat generation.

The increase in entropy dS is defined as the amount of entropy to be added to the reacting system in order to reproduce the irreversible process reversibly. It is measured by means of the standard calorimeter from which it may be reversibly transferred.

3. The product $(P_1 - P_2) \delta K$ represents work in a generalised sense. Equation (1) identifies the loss of work and the heat evolution. It is an expression, therefore, of the generalised law of energetic equivalence. In a world of reversible processes $dS = 0$, and all energetic phenomena would be describable in terms of work. In such a world there would be no rational basis for the creation of a concept of heat.

4. It is the obvious phenomenologic similarity between terms on the left hand side of equation (1) and the corresponding dissimilarity between these terms and the right hand side product TdS that has led to the new terminology

by which the terms heat and work have acquired a clear and unequivocal meaning. Classical thermodynamics does not accept the terms heat and work in the same sense. This of course does not affect the validity or significance of equation (1). The question of terminology is one of convenience and logical adaptation, not one of truth or untruth.

5. In addition to the law of the equivalence of heat and work which is recognised by classical thermodynamics in some particular instances, heat is generally characterised by the law of heat conduction. The actual contents of this law are given by the equation:

$$T_1 \delta S_1 = T_2 \delta S_2. \quad (2)$$

It has been shown, however, that equations (1) and (2) are not independent, the process of heat conduction being a special case of the process of heat generation, and that no new thermal concept, therefore, can be established on equation (2). This demonstration is of some significance, since it is the object of science to reduce as far as possible the number of axiomatic postulates upon which it rests. What is called heat transport is either simply an isothermal transport of entropy or in case of temperature inhomogeneity it is such a transport associated with a heat generation in conformity with equation (1).

6. If the law expressed by (2) is being described as the law of heat conduction or heat transport, this involves that the addition of the entropy δS to a system of temperature T is synonymous to the addition of an amount of heat $T \delta S$. If the system is a calorimeter it is customary and permissible to say that the amount of heat in the calorimeter has increased by $T \delta S$. Heat in this sense is not, however,

comprised in the definition of heat by equation (1) from which it differs fundamentally, but must be ranged with other energies of the type PdK for which no law of transformation — and also no special designations — exist. It pertains to what might be called functional energetics or energetics of state. This kind of energetics is concerned with changes in partial or fractional systems with no reference to simultaneously effected alterations in other fractional systems and is of quite formal character. Work and heat as energetic realities are dealt with in the energetics of processes. This branch of science is concerned with the interconversion and equivalence of the various forms of energy and is characterised by disregarding no part of the phenomena involved in the total change. It deals with closed or isolated systems. Indiscriminate application of the functional heat concept within the energetics of processes is a regrettable feature of classical thermodynamics.

7. If a system of pressure p receives a volume δv from an external source, traditional thermodynamics claims an amount of work $p\delta v$ to be “done” by the system, although this “work” is intransformable and the change, therefore, comes within the domain of “functional energetics”. As in the case of heat this traditional terminology involves the danger of confusing intrinsically incommensurable concepts and is actually to a considerable extent responsible for the inconsistency of classical thermodynamic ideas.

8. Reversible processes are considered by classical thermodynamics as processes with which there may be associated a transformation of heat into work or vice versa. This view is not compatible with a consistent and general definition of the concepts of heat and work. Nor is this the

case with the traditional postulates that irreversible processes involving substances which obey the ideal gas laws occur without heat evolution.

9. The heat which according to the classical view is consumed in a reversible spatial process may be taken from a heat reservoir or from the expanding system itself. In the first case the conclusion concerning a heat-work transformation is based upon the sort of fallacy of argument which was indicated under 6. In either case it is made possible by the obscurity of the classical definition of the phenomena of generation and consumption of heat and the lack of precision in what is called the "work of expansion".

10. Chemical work according to traditional ideas originates partly in "chemical energy" concealed in the system and measurable as an irreversible thermometric heat evolution, partly in heat taken from an external source. The concept of "chemical energy" is, however, completely undefined and that of chemical heat-work transformation meets with the same obstacles as the heat-work transformation in the spatial process.

The traditional interpretations of chemical and spatial processes involving substances which obey the perfect gas laws are entirely incongruous in spite of the obvious similarity of the processes.

11. The proposition of the impracticability of an uncompensated heat-work transformation, though frequently violated in the treatment of an isolated problem, seems to be generally adopted within classical thermodynamics. On the other hand, classical thermodynamics has set up and strongly advocates the doctrine that a heat-work transformation is actually realisable if supported by some sort of concurrent compensation. The nature of this

compensation is, however, not specified, in fact it has received but a very cursory attention in the historic evolution of the theory of heat.

Concerning this crucial contention in classical thermodynamics it has been shown above that the moment a heat consumption is postulated as an equivalent for the production of work, such as is the case in any imaginary classical process, the compensation, simultaneously necessitated, will emerge automatically in the shape of an irreversible process involving as such a generation of heat. Moreover the heat expended and the heat generated prove to be equal and to appear at the same temperature. Consequently the two caloric phenomena will cancel completely and there will be left no source of caloric nature for the production of work. This argument does not depend upon any terminologic circumvention, the term heat in both cases standing for energy members of the form $T\Delta S$ of quite unambiguous significance.

12. The classical notion of a generation or absorption of heat is closely bound up with the appearance of temperature changes in the reacting system. For instance, when the reaction involves a fall in temperature, heat is supposed to disappear. If heat is to be taken as a form of energy, such a view is incompatible with the principle of energetic equivalence and is therefore an obstacle to the formation of a rational and consistent concept of heat.

13. The function of the heat engine has been of decisive significance for the comprehension of heat. According to CARNOT a passage of heat from a higher to a lower temperature constituted an equivalent for the work gained by the engine. In the theory of CLAUSIUS the function of the engine was claimed to be a compensated heat-work transformation.

The energetic theory, finally, views the function as a work-work transformation exclusively controlled by the work principle.

In conformity with the general postulate that the concept of heat is foreign to reversible processes and that the theory of energetic compensation is illusory, special evidence has been brought forward to show that also in the case of a thermal process no heat is consumed in the mechanical performance of work.

14. Heat is a macroscopic entity. The maintenance of a concept of heat is vindicable only through the existence of macroenergetic phenomena and the fulfilment of macroenergetic laws. The question as to the "nature" of heat is without object. The contention that heat is "motion" or energy of molecular motion, no less than the contention that heat is a fluid, is incompatible with the classical thesis that heat does not possess the character of a single-valued function of state.

Det Kgl. Danske Videnskabernes Selskab.

Mathematisk-fysiske Meddelelser. **XIX**, 9.

UNDOR REPRESENTATION OF
THE FIVE-DIMENSIONAL MESON
THEORY

BY

IB NØRLUND



KØBENHAVN

I KOMMISSION HOS EJNAR MUNKSGAARD

1942

CONTENTS

	Page
Introduction	3
1. Fundamental equations of the five-dimensional symmetrical meson theory	7
2. The KRAMERS representation	8
3. Rotations in the five-dimensional space	9
4. Undors	11
5. The undor field equation of the symmetrical field theory	16
6. The undor field equation of the neutral meson theory	17
7. The adjoint undor	19
8. The Lagrangeian and the current vector in the symmetrical theory	20
9. Difficulties of the five-dimensional neutral meson theory	25
References	29

In 1935, YUKAWA¹⁾ pointed out that short range nuclear forces may be accounted for by a field of a new kind corresponding to the existence of particles—now generally denoted as mesons—with a rest mass intermediary between those of the electron and the proton, and with an integral spin. Since then, the meson field theory has been treated by a great number of authors and has undergone an important development in various respects.

Originally, YUKAWA used a scalar field function implying the mesons to have spin 0. Such a formalism gives, however, a repulsion between the nucleons in a 3S state of the deuteron in contradiction to the experiments which show that the ground state of the deuteron is just a 3S state. The difficulty was overcome in the further development of the theory when, according to a formalism due to PROCA²⁾, the field was described by means of a vector function. This form of the theory which leads to the value 1 of the spin of the mesons was developed by YUKAWA and SAKATA³⁾, FRÖHLICH, HEITLER and KEMMER⁴⁾, BHABHA⁵⁾, and STÜCKELBERG⁶⁾. Assuming that the spin of the meson does not exceed 1, KEMMER⁷⁾ was able to show that the meson field can be described by four—and only by four—different wave-functions characterized by their covariance properties. Besides the mentioned “scalar” and “vector” functions, “pseudoscalar” and “pseudovector” wave-functions or arbitrary combinations of all of them are possible.

In his first paper, YUKAWA considered charged mesons, only, but in the subsequent development it has been necessary also to assume the existence of neutral mesons in the theory, since it was shown by collision experiments that the interaction—at nuclear distances—between two protons is approximately the same as that between a proton and a neutron⁸⁾. As found by KEMMER⁹⁾, this experimentally observed charge independence of the nuclear forces can be accounted for in combining the wave-functions of the charged and the neutral mesons in a certain symmetrical scheme in which the nucleons interact equally strongly with neutral, positive and negative mesons.

In a paper by MØLLER and ROSENFELD¹⁰⁾ (in the following quoted as MR.) it was shown, however, that, when introducing the charged and neutral mesons in the mentioned symmetrical way, it is necessary to combine the “vector” theory with the “pseudoscalar” theory in order to avoid the appearance of singular terms in the static interaction between nucleons and, thus, to obtain a basis for an unambiguous description of the stationary states of atomic nuclei.

The charge independence of the nuclear forces may, of course, also be explained by using neutral meson fields, only. A theory on these lines has been put forward by BETHE¹¹⁾ who used a vector field function for the description of the neutral mesons. In order to get finite results in the calculations following from such a theory it is, however, necessary to “cut off” at small distances in a more or less arbitrary way. To obtain an unambiguous theory for the energy levels of the deuteron without having recourse to such a procedure it is, again, necessary to combine two types of meson fields, viz. a “scalar” and a “pseudovector”

field (cf. MR. p. 34, footnote). This theory which, in the following, will be called the "neutral meson theory" can in certain respects be regarded as dual to the "symmetrical meson theory" treated in MR.

The symmetrical theory has been brought into an equivalent and particularly compact form by BELINFANTE¹²⁾ who has written the 15 field equations as one single "undor" equation showing a close analogy to the DIRAC equation of the electron. The "undors" introduced by BELINFANTE are quantities which under LORENTZ transformations behave like products of DIRAC wave-functions (four-spinors).

Irrespective of the form in which the symmetrical theory is written, the field equations contain four universal constants, f'_1 , f'_2 , g'_1 and g'_2 , mutually independent apart from the condition

$$(f'_2)^2 = (g'_2)^2$$

which is unavoidable if the singular terms shall cancel each other in the nuclear interaction. The circumstance that the two kinds of fields appearing in the theory are in no direct connection with each other and, furthermore, the occurrence of a large number of independent constants constitute an unsatisfactory feature of the theory. Recently, MØLLER has shown in a paper¹³⁾ (quoted in the following as M.) that the field equations of the vector and the pseudoscalar theory can be comprised in one set of five-dimensional field equations which become invariant with respect to five-dimensional rotations on condition that the independence of the constants is restricted by the relations

$$f'_1 = g'_1 \text{ and } f'_2 = -g'_2.$$

The advantages of the representation used by BELINFANTE as well as the simplification obtained by a five-

dimensional treatment of the theory make it desirable to generalize the undor description of the meson theory to the case of five dimensions, a generalization which will be performed in the present paper. In this way, we obtain a rather concise and clear form of the theory and, at the same time, the relation between the neutral and the symmetrical theory is elucidated from a new point of view.

The first sections of this paper are devoted to a more detailed account of some of the earlier investigations mentioned above. Thus, section 1 contains the fundamental equations for the symmetrical theory in five dimensions, in section 2, a certain representation of the DIRAC matrices used in the following, the so-called KRAMERS representation, is introduced and, in section 3, a short account of the general transformation formulae for infinitesimal and finite rotations in five dimensions is given.

Next, section 4 contains the definition and the simplest properties of the four-dimensional undors introduced by BELINFANTE as well as of their five-dimensional generalization with the alterations which are made necessary by the condition of invariance with respect to five-dimensional rotations. Finally, the undors are expressed by the tensor field quantities and an example is given in order to illustrate the way in which calculations with undors are running.

After these preparations, it is possible to set up, in section 5, the undor field equation of the symmetrical theory and, in section 6, the undor field equation of the neutral meson theory which only by a change of sign differs essentially from the former equation. In section 7, the "adjoint" undor is introduced and expressed in terms of the original undor elements. Furthermore, the corresponding differential

equation for this adjoint undor is derived. Now, in section 8, it is shown that the field equation for the symmetrical theory may be derived from a variational principle, the corresponding Lagrangeian assuming a rather elegant form in the undor formulation. The wave-equations of the nucleons and the expression for the current vector are derived.

This formalism is especially well suited to be used for the description of the symmetrical theory but, as shown in section 9, not for the description of the neutral theory. Thus, the impossibility of obtaining in such a way a Lagrangeian principle for the neutral theory in five dimensions would seem to offer an argument in preference of the symmetrical theory.

1. Fundamental equations of the five-dimensional symmetrical meson theory.

In the five-dimensional meson theory the 10 field equations of the vector theory together with the 5 equations of the pseudoscalar theory are comprised in the following equations (cf. M. (12)):

$$\left. \begin{aligned} \mathbf{G}_{\mu\nu} &= D_\mu \mathbf{U}_\nu - D_\nu \mathbf{U}_\mu + \mathbf{S}_{\mu\nu} & (a) \\ D_\nu \mathbf{G}_{\mu\nu} + \kappa^2 \mathbf{U}_\mu &= \mathbf{M}_\mu & (b) \\ \mu, \nu &= 0, 1, 2, 3, 4; \quad \kappa = \frac{M_m c}{\hbar} \\ D_\mu &= \frac{\partial}{\partial x_\mu} \end{aligned} \right\} \quad (1)$$

where \mathbf{U}_μ and \mathbf{M}_μ are five-vectors, $\mathbf{G}_{\mu\nu}$ and $\mathbf{S}_{\mu\nu}$ are antisymmetrical five-tensors in the $(x_0, x_1, x_2, x_3, x_4)$ -space, and M_m is the mass of the meson. A heavy printed letter indicates that the quantities in question have three

independent components in isotopic space so that, for instance, $\mathbf{G}_{\mu\nu}$ signifies the symbolic vector:

$$\mathbf{G}_{\mu\nu} = (G_{\mu\nu\mathbf{1}}, G_{\mu\nu\mathbf{2}}, G_{\mu\nu\mathbf{3}}).$$

We have, thus, to do with three non-interfering meson fields, the index $\mathbf{3}$ referring to the neutral field, while the indices $\mathbf{1}$ and $\mathbf{2}$ refer to the real and imaginary part of the complex field representing the charged mesons. The symbols \mathbf{G} and \mathbf{U} describe the field quantities, while \mathbf{S} and \mathbf{M} denote the densities of the source distribution defined by

$$\left. \begin{aligned} \mathbf{M}_{\mu} &= g_1 \psi^{\dagger} \mathbf{T} \gamma_{\mu} \psi \\ \mathbf{S}_{\mu\nu} &= \frac{g_2}{2\kappa} \psi^{\dagger} \mathbf{T} [\gamma_{\mu}, \gamma_{\nu}] \psi \end{aligned} \right\} \quad (2)$$

where \mathbf{T} is the isotopic spin vector, ψ the wave-function of the nucleons, and its adjoint ψ^{\dagger} is given by

$$\psi^{\dagger} = i\psi^* \beta. \quad (3)$$

Further, $[\gamma_{\mu}, \gamma_{\nu}]$ is the commutator of the quantities γ_{μ} used in \mathbf{M} . and explicitly written down in equations (5) of the next section.

The equations (1) are evidently invariant with respect to the group of five-dimensional rotations. This group includes, of course, the complete LORENTZ group of transformations in the (x_1, x_2, x_3, x_4) -space under which the field quantities of the vector theory and of the pseudoscalar theory are transformed separately. For a general five-dimensional rotation, however, the field quantities of these two theories are mixed up.

2. The KRAMERS representation.

Let $\rho_1, \rho_2, \rho_3, \sigma_1, \sigma_2, \sigma_3$ be the ordinary DIRAC matrices, where ρ_3 and σ_3 are diagonal ("DIRAC representation").

Introducing now the variables $\rho_x, \rho_y, \rho_z, \sigma_x, \sigma_y, \sigma_z$ by the following relations

$$\left. \begin{aligned} \rho_x &= \rho_3 & \sigma_x &= \sigma_1 \\ \rho_y &= -\rho_2 & \sigma_y &= \sigma_2 \\ \rho_z &= \rho_1 & \sigma_z &= \sigma_3 \end{aligned} \right\} \quad (4)$$

it is seen that the new variables satisfy the same algebraic relations as the original variables, e. g.

$$\rho_x \rho_y = i \rho_z.$$

In the so-called KRAMERS representation which we, following BELINFANTE, shall use in the present paper, ρ_z and σ_z are chosen to be diagonal so that the matrices of $\rho_x, \rho_y, \rho_z, \sigma_x, \sigma_y, \sigma_z$ in the KRAMERS representation are the same as the matrices of $\rho_1, \rho_2, \rho_3, \sigma_1, \sigma_2, \sigma_3$ in the DIRAC representation.

For reference, we write down in the KRAMERS representation some of the quantities which will be used in the following:

$$\left. \begin{aligned} \vec{\alpha} &= \rho_z \vec{\sigma} \\ \beta &= \rho_x \\ \vec{\gamma} &= -i\beta \vec{\alpha} = -\rho_y \vec{\sigma} \\ \gamma_4 &= \beta = \rho_x \\ \gamma_0 &= \gamma_1 \gamma_2 \gamma_3 \gamma_4 = -\rho_z \end{aligned} \right\} \quad (5)$$

3. Rotations in the five-dimensional space.

For an infinitesimal rotation in the five-dimensional space,

$$\left. \begin{aligned} x'_\mu &= x_\mu + \sum_{\nu=0}^4 \varepsilon_{\mu\nu} x_\nu; & \varepsilon_{\mu\nu} &= -\varepsilon_{\nu\mu} \end{aligned} \right\} \quad (6)$$

a four-spinor transforms according to the formula (cf. M. (9), (9'))

$$\psi' = S \psi \quad (7)$$

where

$$S = 1 + \frac{1}{2} \sum_{\mu, \nu=0}^4 q_{\mu\nu} \varepsilon_{\mu\nu} \quad (8)$$

$$q_{\mu\nu} = \frac{1}{4} [\gamma_{\mu}, \gamma_{\nu}] \quad (9)$$

or, in the KRAMERS representation,

$$\{q_{\mu\nu}\} = \begin{pmatrix} 0 & -\frac{i}{2} \rho_x \sigma_x & -\frac{i}{2} \rho_x \sigma_y & -\frac{i}{2} \rho_x \sigma_z & -\frac{i}{2} \rho_y \\ \frac{i}{2} \rho_x \sigma_x & 0 & \frac{i}{2} \sigma_z & -\frac{i}{2} \sigma_y & \frac{i}{2} \rho_z \sigma_x \\ \frac{i}{2} \rho_x \sigma_y & -\frac{i}{2} \sigma_z & 0 & \frac{i}{2} \sigma_x & \frac{i}{2} \rho_z \sigma_y \\ \frac{i}{2} \rho_x \sigma_z & \frac{i}{2} \sigma_y & -\frac{i}{2} \sigma_x & 0 & \frac{i}{2} \rho_z \sigma_z \\ \frac{i}{2} \rho_y & -\frac{i}{2} \rho_z \sigma_x & -\frac{i}{2} \rho_z \sigma_y & -\frac{i}{2} \rho_z \sigma_z & 0 \end{pmatrix} \quad (10)$$

The transformation equations for a finite rotation in the (x_{α}, x_{β}) -plane are obtained by substituting (8) into

$$S(\varphi + d\varphi) = S(\varphi) \cdot S(d\varphi).$$

We get thus the equation

$$S(\varphi + d\varphi) = (1 + q_{\alpha\beta} d\varphi) S(\varphi)$$

with the only continuous solution

$$S = e^{q_{\alpha\beta} \varphi} \quad (11)$$

to be used in (7).

4. Undors.

Undors are, according to the definition of BELINFANTE, quantities transforming as products of four-component DIRAC wave-functions (four-spinors). The number of four-spinor factors in the corresponding transformation scheme is called the rank of the undor. An undor of the first rank is thus simply a four-spinor.

In the following, we shall deal with undors of the second rank, i. e. quantities with 16 components $\Psi_{k_1 k_2}$ transforming as products $\psi_{k_1} \psi_{k_2}$ ($k_1, k_2 = 1, 2, 3, 4$). It will be convenient to arrange the 16 components into a square scheme

$$\Psi = \begin{vmatrix} \Psi_{11} & \Psi_{12} & \Psi_{13} & \Psi_{14} \\ \Psi_{21} & \Psi_{22} & \Psi_{23} & \Psi_{24} \\ \Psi_{31} & \Psi_{32} & \Psi_{33} & \Psi_{34} \\ \Psi_{41} & \Psi_{42} & \Psi_{43} & \Psi_{44} \end{vmatrix} \quad (12)$$

which, however, is not to be confounded with a matrix representing an operator.

In our considerations we shall only need operators which act either on the first or on the second index of the undor. We characterize such operators by providing their matrices with the suffix ⁽¹⁾ or ⁽²⁾, respectively. Operators with the matrices $A^{(1)}$ and $A^{(2)}$, for instance, will transform the undor Ψ according to the formulae

$$\left. \begin{aligned} (A^{(1)}\Psi)_{k_1 k_2} &= \sum_{k'=1}^4 (k_1 | A | k') \Psi_{k' k_2} \\ (A^{(2)}\Psi)_{k_1 k_2} &= \sum_{k'=1}^4 (k_2 | A | k') \Psi_{k_1 k'} \end{aligned} \right\} \quad (13)$$

In accordance with (7), the transformation of an undor of the second rank by a rotation in the five-dimensional space is given by

$$\Psi' = S^{(1)} S^{(2)} \Psi. \quad (14)$$

It can be shown that, if we only consider the subgroup of LORENTZ transformations in the (x_1, x_2, x_3, x_4) -space, the transformation (14) will have the special property that the elements inside one of the four subsquares $(\Psi_{11}, \Psi_{12}, \Psi_{21}, \Psi_{22})$, $(\Psi_{31}, \Psi_{32}, \Psi_{41}, \Psi_{42})$, $(\Psi_{13}, \Psi_{14}, \Psi_{23}, \Psi_{24})$ and $(\Psi_{33}, \Psi_{34}, \Psi_{43}, \Psi_{44})$ are not mixed up with elements of the other squares. This is just the case treated by BELINFANTE in his four-dimensional theory. He was able to show that an undor of the second rank comprises—with respect to the complete LORENTZ group—a scalar, a four-vector, an antisymmetrical tensor of the second rank, a pseudo-four-vector, and a pseudo-scalar. Furthermore, he deduced a possible correspondance between the 16 components of these quantities and the 16 undor elements. Now, the meson fields and the source densities are described just by the tensors mentioned above, a fact which makes the undors so useful in the treatment of the meson theory. In the correspondance chosen by BELINFANTE, a symmetrical undor contains only the quantities appearing in the vector meson theory (the PROCA field), while an antisymmetrical undor contains the quantities of the pseudoscalar theory, so that a general undor will just represent the combination of fields dealt with in MR. This correspondance is, however, not unique since another possible correspondance would be obtained by changing the sign of all elements in one or more of the four subsquares in (12).

Passing to the five-dimensional theory, the field variables of the vector and the pseudoscalar theories are, as mentioned

in section 1, comprised to a five-vector and an antisymmetrical five-tensor of the second rank (cf. M. (10), (10'), (11), (11')). Now, if we do not confine ourselves to LORENTZ transformations but consider the full group of five-dimensional rotations, we find that the transformation (14) mixes up elements from different subsquares of the undor scheme (12). This means that some of the arbitrariness in the mentioned correspondance between undor elements and field quantities is removed.

It is thus impossible to make a straightforward generalization of the correspondance set up by BELINFANTE in replacing the field quantities of the four-dimensional theory by the corresponding quantities in five dimensions. A simple calculation shows that this would lead to obvious contradictions. Consequently, we have to find a new correspondance between the 16 undor elements, on one hand, and the antisymmetrical tensor $G_{\mu\nu}$, and the five-vector U_μ , on the other hand. Since the last mentioned quantities together have 15 components only, we add a scalar K in order to obtain the complete representation of the 16 independent undor elements. The criterion of the consistence of such a correspondance is that the transformation (14) transforms the quantities $G_{\mu\nu}$, U_μ and K as the components of a tensor, a vector and a scalar, respectively.

A straightforward calculation shows that such a correspondance can be obtained from the correspondance scheme set up by BELINFANTE by a direct generalization to five dimensions followed by a change of sign in the two subsquares forming the last two columns of the undor scheme. This correspondance is given below, in equation (15), for the meson fields and, in equation (16), for the sources giving rise to these fields.

$$\begin{aligned}
\mathbf{y}_{k_1 k_2} = & \begin{array}{cccccccc}
G_{23} + G_{14} + iG_{13} - iG_{24} & -G_{12} - G_{34} + iKU_0 - iK & -G_{01} + iG_{02} - KU_1 + iKU_2 & G_{03} - iG_{04} + KU_3 - iKU_4 \\
-G_{12} - G_{34} - iKU_0 + iK & -G_{23} - G_{14} + iG_{13} - iG_{24} & G_{03} + iG_{04} + KU_3 + iKU_4 & G_{01} + iG_{02} + KU_1 + iKU_2 \\
-G_{01} + iG_{02} + KU_1 - iKU_2 & G_{03} + iG_{04} - KU_3 - iKU_4 & G_{23} - G_{14} + iG_{13} + iG_{24} & -G_{12} + G_{34} - iKU_0 - iK \\
G_{03} - iG_{04} - KU_3 + iKU_4 & G_{01} + iG_{02} - KU_1 - iKU_2 & -G_{12} + G_{34} + iKU_0 + iK & -G_{23} + G_{14} + iG_{13} + iG_{24}
\end{array} \\
& (15)
\end{aligned}$$

$$\begin{aligned}
\mathbf{Y}_{k_1 k_2} = & \begin{array}{cccccccc}
-KS_{23} - KS_{14} - iKS_{13} + iKS_{24} & KS_{12} + KS_{34} - iM_0 + iKC & KS_{01} - iKS_{02} + M_1 - iM_2 & -KS_{03} + iKS_{04} - M_3 + iM_4 \\
KS_{12} + KS_{34} + iM_0 - iKC & KS_{23} + KS_{14} - iKS_{13} + iKS_{24} & -KS_{03} - iKS_{04} - M_3 - iM_4 & -KS_{01} - iKS_{02} - M_1 - iM_2 \\
KS_{01} - iKS_{02} - M_1 + iM_2 & -KS_{03} - iKS_{04} + M_3 + iM_4 & -KS_{23} + KS_{14} - iKS_{13} - iKS_{24} & KS_{12} - KS_{34} + iM_0 + iKC \\
-KS_{03} + iKS_{04} + M_3 - iM_4 & -KS_{01} - iKS_{02} + M_1 + iM_2 & KS_{12} - KS_{34} - iM_0 - iKC & KS_{23} - KS_{14} - iKS_{13} - iKS_{24}
\end{array} \\
& (16)
\end{aligned}$$

In the last formula, \mathbf{C} denotes a scalar which must be included in the scheme similarly to the scalar \mathbf{K} in (15).

From the equations (2), (3), (5) and the equation

$$\mathbf{C} = \frac{g_0}{\kappa} \psi^\dagger \mathbf{T} \psi, \quad (17)$$

which is the only possible expression for a scalar source, it is seen that

$$f_{op}^{-1} \mathbf{X}_{k_1 k_2} = 2 (\psi^\dagger \sigma_y)_{k_1} \mathbf{T} \psi_{k_2} \quad (18)$$

where f_{op}^{-1} is an operator dividing \mathbf{C} , $\mathbf{M}_{\mu\nu}$, and $\mathbf{S}_{\mu\nu}$ by g_0 , g_1 and g_2 , respectively.

A simple calculation shows that the correspondance (15) satisfies the criterion mentioned above. As an example, we shall bring the calculation for a rotation in the (x_0, x_1) -plane. The corresponding operator S in equation (14) is, according to (10) and (11),

$$S = \cos \frac{\Phi}{2} - i \rho_x \sigma_x \sin \frac{\Phi}{2}.$$

From (14) we then get

$$\begin{aligned} \Psi'_{11} &= \Psi_{11} \cos^2 \frac{\Phi}{2} - \Psi_{44} \sin^2 \frac{\Phi}{2} - i(\Psi_{14} + \Psi_{41}) \sin \frac{\Phi}{2} \cos \frac{\Phi}{2} \\ \Psi'_{12} &= \Psi_{12} \cos^2 \frac{\Phi}{2} - \Psi_{43} \sin^2 \frac{\Phi}{2} - i(\Psi_{13} + \Psi_{42}) \sin \frac{\Phi}{2} \cos \frac{\Phi}{2} \\ &\dots \text{ etc. (16 eqq.)} \end{aligned}$$

or, from (15),

$$\begin{aligned} &\mathbf{G}'_{23} + \mathbf{G}'_{14} + i \mathbf{G}'_{13} - i \mathbf{G}'_{24} = \\ = &\mathbf{G}_{23} + \mathbf{G}_{14} \cos \Phi + i \mathbf{G}_{13} \cos \Phi - i \mathbf{G}_{24} - i \mathbf{G}_{03} \sin \Phi - \mathbf{G}_{04} \sin \Phi \\ &- \mathbf{G}'_{12} - \mathbf{G}'_{34} + i \kappa \mathbf{U}_0 - i \mathbf{K}' = \\ = &-\mathbf{G}_{12} \cos \Phi - \mathbf{G}_{34} + i \kappa \mathbf{U}_0 \cos \Phi - i \mathbf{K} + \mathbf{G}_{02} \sin \Phi + i \kappa \mathbf{U}_1 \sin \Phi \\ &\dots \text{ etc. (16 eqq.).} \end{aligned}$$

Solving these equations we get

$$\begin{aligned}
 \mathbf{G}'_{23} &= \mathbf{G}_{23} & \mathbf{U}'_0 &= \mathbf{U}_0 \cos \varphi + \mathbf{U}_1 \sin \varphi \\
 \mathbf{G}'_{14} &= \mathbf{G}_{14} \cos \varphi - \mathbf{G}_{04} \sin \varphi & \dots & \text{etc.} \\
 \mathbf{G}'_{13} &= \mathbf{G}_{13} \cos \varphi - \mathbf{G}_{03} \sin \varphi & & \\
 \mathbf{G}'_{24} &= \mathbf{G}_{24} & \mathbf{K}' &= \mathbf{K} \\
 \mathbf{G}'_{12} &= \mathbf{G}_{12} \cos \varphi - \mathbf{G}_{02} \sin \varphi & & \\
 \mathbf{G}'_{34} &= \mathbf{G}_{34} & & \\
 \dots & \text{etc.} & &
 \end{aligned}$$

i. e. the well-known transformation formulae for a tensor, a vector and a scalar under the rotation in question.

5. The undor field equation of the symmetrical field theory.

Having obtained a correspondance between the tensor quantities of the meson theory and the elements of the undors Ψ and \mathbf{X} , we may now comprise the 15 equations (1) and a trivial scalar equation in one undor equation closely analogous to the DIRAC equation of the electron:

$$2\kappa\Psi + 2\mathbf{X} + (\gamma_{\mu}^{(1)} - \gamma_{\mu}^{(2)})D_{\mu}\Psi = 0 \quad (19)$$

which, of course, corresponds to 16 equations between the undor elements.

In fact, since the operator $\gamma_{\mu}D_{\mu}$ has the representation

$$\gamma_{\mu}D_{\mu} = \begin{pmatrix} -D_0 & 0 & iD_3 + D_4 & iD_1 + D_2 \\ 0 & -D_0 & iD_1 - D_2 & -iD_3 + D_4 \\ -iD_3 + D_4 & -iD_1 - D_2 & D_0 & 0 \\ -iD_1 + D_2 & iD_3 + D_4 & 0 & D_0 \end{pmatrix} \quad (20)$$

the introduction of (15), (16) and (20) into (19) gives the equations

$$\begin{aligned}
 & \kappa(\mathbf{G}_{23} - \mathbf{S}_{23}) + \kappa(\mathbf{G}_{14} - \mathbf{S}_{14}) + i\kappa(\mathbf{G}_{13} - \mathbf{S}_{13}) - \\
 & \quad - i\kappa(\mathbf{G}_{24} - \mathbf{S}_{24}) + \kappa(iD_3 + D_4)(\mathbf{U}_1 - i\mathbf{U}_2) + \\
 & \quad + \kappa(iD_1 + D_2)(-\mathbf{U}_3 + i\mathbf{U}_4) = 0 \\
 & \hspace{20em} (k_1 = 1, k_2 = 1) \\
 & -\kappa(\mathbf{G}_{12} - \mathbf{S}_{12}) - \kappa(\mathbf{G}_{34} - \mathbf{S}_{34}) + i(\kappa^2 \mathbf{U}_0 - \mathbf{M}_0) - i\kappa(\mathbf{K} - \mathbf{C}) + \\
 & \quad + iD_3(\mathbf{G}_{03} - i\kappa \mathbf{U}_4) + D_4(i\mathbf{G}_{04} - \kappa \mathbf{U}_3) + \\
 & \quad + iD_1(\mathbf{G}_{01} - i\kappa \mathbf{U}_2) + D_2(i\mathbf{G}_{02} - \kappa \mathbf{U}_1) = 0 \\
 & \dots \text{ etc. (16 eqq.)} \hspace{10em} (k_1 = 1, k_2 = 2)
 \end{aligned}$$

with the solution

$$\begin{aligned}
 \mathbf{G}_{23} - \mathbf{S}_{23} &= D_2 \mathbf{U}_3 - D_3 \mathbf{U}_2 & \kappa^2 \mathbf{U}_0 + D_1 \mathbf{G}_{01} + D_2 \mathbf{G}_{02} + \\
 \mathbf{G}_{14} - \mathbf{S}_{14} &= D_1 \mathbf{U}_4 - D_4 \mathbf{U}_1 & \quad + D_3 \mathbf{G}_{03} + D_4 \mathbf{G}_{04} = \mathbf{M}_0 \\
 \mathbf{G}_{13} - \mathbf{S}_{13} &= D_1 \mathbf{U}_3 - D_3 \mathbf{U}_1 & \quad \dots \text{ etc.} \\
 \mathbf{G}_{24} - \mathbf{S}_{24} &= D_2 \mathbf{U}_4 - D_4 \mathbf{U}_2 & \quad \mathbf{K} = \mathbf{C} \\
 \mathbf{G}_{12} - \mathbf{S}_{12} &= D_1 \mathbf{U}_2 - D_2 \mathbf{U}_1 \\
 \mathbf{G}_{34} - \mathbf{S}_{34} &= D_3 \mathbf{U}_4 - D_4 \mathbf{U}_3 \\
 & \dots \text{ etc.}
 \end{aligned}$$

i. e. the equations (1) and the trivial scalar equation

$$\mathbf{K} = \mathbf{C}.$$

The equation (19) differs from the undor equation of BELINFANTE (formula (20), p. 26) by the sign of $\gamma_\mu^{(2)}$ in accordance with the change of sign in the correspondance equations (15) and (16). On account of the well-known properties of the DIRAC matrices γ_μ , the equation (19) is invariant with respect to five-dimensional rotations.

6. The undor field equation of the neutral meson theory.

Like the equation (19), the undor equation

$$2\kappa\Psi + 2X + (\gamma_\mu^{(1)} + \gamma_\mu^{(2)})D_\mu\Psi = 0 \tag{21}$$

is invariant with respect to rotations in five-dimensional space. As it will be shown in the following, (21) represents the field equations of the neutral meson theory. Consequently, the undor quantities Ψ and X are not printed in heavy letters in this section, since the neutral theory only deals with one single component in the isotopic spin space (cf. Section 1).

Substituting (15) and (16) in (21) we obtain the 16 invariant field equations

$$\left. \begin{aligned} K - C + D_\mu U_\mu &= 0 & (a) \\ M_\mu &= \kappa^2 U_\mu + D_\mu K & (b) \\ \kappa(G_{\mu\nu} - S_{\mu\nu}) - (-1)^{I[\mu, \nu, 1, \kappa, \lambda]} (D_1 G_{\kappa\lambda} + D_\lambda G_{1\kappa} + D_\kappa G_{\lambda 1}) &= 0 & (c) \end{aligned} \right\} (22)$$

where $I[\mu, \nu, 1, \kappa, \lambda]$ is the number of inversions in the permutation $\mu, \nu, 1, \kappa, \lambda$ of the numbers 0, 1, 2, 3, 4.

The equations (22) may also be written as

$$\left. \begin{aligned} K - C + \text{Div } U &= 0 & (a) \\ M &= \kappa^2 U + \text{Grad } K & (b) \\ \kappa(G - S) &= \text{Rot } G & (c) \end{aligned} \right\} (23)$$

where Div and Grad are the straightforward generalizations of these differential operators to five dimensions, and Rot is defined by

$$D_1 G_{\kappa\lambda} + D_\lambda G_{1\kappa} + D_\kappa G_{\lambda 1} = \rho\sigma\tau_{1\kappa\lambda} G = (-1)^{I[1, \kappa, \lambda, \mu, \nu]} \text{Rot}_{\mu\nu} G. \quad (24)$$

If we express the antisymmetrical tensor of the second rank $G_{\mu\nu}$ by an antisymmetrical tensor $G_{1\kappa\lambda}$ of the third rank by means of a formula analogous to (24) we get, instead of (23 c), the tensor equation of the third rank

$$\kappa(G - S) = \rho\sigma\tau G. \quad (23 d)$$

Supposing now that the variables in (23) are independent of the coordinate x_0 , we find that these equations reduce to the four-dimensional equations

$$\left. \begin{aligned} K - C + D_k U_k &= 0 & (a) \\ M_k &= \kappa^2 U_k + D_k K & (b) \\ M_0 &= \kappa^2 U_0 & (c) \\ \kappa(G_{klm} - S_{klm}) &= D_k G_{lm} + D_l G_{mk} + D_m G_{kl} & (d) \\ \kappa(G_{kl} - S_{kl}) &= D_r G_{klr}, & (e) \end{aligned} \right\} (25)$$

Latin indices running only from 1 to 4. The equations (25 d) and (25 e) are obtained from (23 d) for $\iota, \kappa, \lambda \neq 0$ and $\iota = 0$, respectively. The equations (25) are just the field equations of the neutral meson theory in four dimensions.

We have, thus, for the scalar-pseudovector theory in five dimensions in (21) obtained an undor representation similar to that given by (19) for the vector-pseudoscalar theory.

7. The adjoint undor.

We define the adjoint Ψ^\dagger to an undor Ψ of the second rank in a similar way as it was done in the case of an undor of the first rank by the equation (3):

$$\Psi^\dagger = -\Psi^* \beta^{(1)} \beta^{(2)}. \quad (26)$$

Consequently, Ψ^\dagger transforms under a transformation (6) according to the formula

$$\Psi^{\dagger'} = \Psi^\dagger (S^{(1)})^{-1} (S^{(2)})^{-1} \quad (27)$$

so that $\Psi^\dagger \Psi$ is an invariant. As usual, multiplication from the right by operators $A^{(1)}$ and $A^{(2)}$ is defined by

$$\left. \begin{aligned} (\Phi A^{(1)})_{k_1 k_2} &= \sum_{k'=1}^4 \Phi_{k' k_2} (k' | A | k_1) \\ (\Phi A^{(2)})_{k_1 k_2} &= \sum_{k'=1}^4 \Phi_{k_1 k'} (k' | A | k_2) \end{aligned} \right\} \quad (28)$$

whence

$$(\Phi A) \Psi = \Phi (A \Psi). \quad (29)$$

Remembering that the field quantities \mathbf{G}_{kl} and \mathbf{U}_k ($k, l = 0, 1, 2, 3$) are real, while \mathbf{G}_{k4} , \mathbf{U}_4 and \mathbf{K} are purely imaginary, we get from (15) by a simple calculation that the elements of $\Psi_{k_1 k_2}^\dagger$ can in a simple way be expressed by the elements of $\Psi_{k_1 k_2}$, viz.:

$$\Psi^\dagger = \begin{vmatrix} \Psi_{22} & -\Psi_{12} & \Psi_{42} & -\Psi_{32} \\ -\Psi_{21} & \Psi_{11} & -\Psi_{41} & \Psi_{31} \\ \Psi_{24} & -\Psi_{14} & \Psi_{44} & -\Psi_{34} \\ -\Psi_{23} & \Psi_{13} & -\Psi_{43} & \Psi_{33} \end{vmatrix} \quad (30)$$

From the undor field equations (19) and (21) we get the following differential equations for the adjoint undors in the symmetrical and neutral theory, respectively:

$$2\kappa \Psi^\dagger + 2\mathbf{X}^\dagger - \Psi^\dagger D_\mu (\gamma_\mu^{(1)} - \gamma_\mu^{(2)}) = 0 \quad (31)$$

$$2\kappa \Psi^\dagger + 2\mathbf{X}^\dagger - \Psi^\dagger D_\mu (\gamma_\mu^{(1)} + \gamma_\mu^{(2)}) = 0. \quad (32)$$

8. The Lagrangeian and the current vector in the symmetrical theory.

The field equations (19) may be derived from the variational principle

$$\delta \int \mathcal{L}' d\Omega = 0 \quad (33)$$

where

$$\mathcal{L}' = \Psi^\dagger \left\{ \left(\kappa + \frac{\gamma_\mu^{(1)} - \gamma_\mu^{(2)}}{2} D_\mu \right) \Psi + 2 \mathbf{X} \right\} \quad (34)$$

the 16 quantities $\Psi_{k_1 k_2}$ being varied independently. In the expression (34) a scalar product like $\Psi^\dagger \Psi$ means

$$\Psi^\dagger \Psi = \sum_{k_1, k_2=1}^4 \Psi_{k_1 k_2}^\dagger \Psi_{k_1 k_2}. \quad (35)$$

The function \mathcal{L}' can, however, not be considered as a Lagrangean. In fact, if we express \mathcal{L}' in terms of the tensor field variables, we obtain

$$\mathcal{L}' = \kappa \left\{ -\frac{1}{2} \mathbf{G}_{\mu\nu} \mathbf{G}_{\mu\nu} + \kappa^2 \mathbf{U}_\mu \mathbf{U}_\mu + \mathbf{K}^2 + \mathbf{G}_{\mu\nu} \mathbf{S}_{\mu\nu} - \right. \\ \left. - 2 \mathbf{U}_\mu \mathbf{M}_\mu - 2 \mathbf{K} \mathbf{C} + \mathbf{U}_\mu D_\nu \mathbf{G}_{\mu\nu} + \frac{1}{2} \mathbf{G}_{\mu\nu} (D_\mu \mathbf{U}_\nu - D_\nu \mathbf{U}_\mu) \right\} \quad (36)$$

where now $\mathbf{G}_{\mu\nu}$, \mathbf{U}_μ and \mathbf{K} should be varied independently, but all components of $\mathbf{G}_{\mu\nu}$ cannot be regarded as canonical variables, since some of the corresponding time derivatives do not appear in \mathcal{L}' . We have, therefore, to proceed otherwise and to regard the field equations (1 a) as definitions of the $\mathbf{G}_{\mu\nu}$, so that the quantities \mathbf{U}_μ only, are independent variables in (36). At the same time, we put the scalars \mathbf{K} and \mathbf{C} equal to zero, these quantities being unimportant in this connection. Substituting the expressions (1 a) for $\mathbf{G}_{\mu\nu}$ in (36) we obtain, by partial integration, the Lagrangean

$$\mathcal{L} = \kappa \left(\frac{1}{2} \mathbf{G}_{\mu\nu} \mathbf{G}_{\mu\nu} + \kappa^2 \mathbf{U}_\mu \mathbf{U}_\mu - 2 \mathbf{U}_\mu \mathbf{M}_\mu \right). \quad (37)$$

Expressing the equations (1 a) in undor language, these relations are now to be regarded as subsidiary conditions in the variational principle (33). As the total Lagrangean we get thus

$$\bar{\mathcal{L}} = \int_{\Omega} \left\{ \Psi^\dagger \left[\left(\kappa + \frac{\gamma_\mu^{(1)} - \gamma_\mu^{(2)}}{2} D_\mu \right) \Psi + 2 \mathbf{X} \right] + \right. \\ \left. + 2 \frac{\hbar c}{i} \psi^\dagger \left(\gamma_\mu D_\mu + \frac{M_0 c}{\hbar} \right) \psi \right\} d\Omega \quad (38)$$

or, in terms of the tensor field variables:

$$\bar{\mathcal{L}} = \int_{\Omega} \left\{ \kappa \left\{ \frac{1}{2} \mathbf{G}_{\mu\nu} \mathbf{G}_{\mu\nu} + \kappa^2 \mathbf{U}_\mu \mathbf{U}_\mu - 2 \mathbf{U}_\mu \mathbf{M}_\mu + \right. \right. \\ \left. \left. + 2 \frac{\hbar c}{i} \psi^\dagger \left(\gamma_\mu D_\mu + \frac{M_0 c}{\hbar} \right) \psi \right\} d\Omega \right\} \quad (39)$$

where M_0 is an abbreviation for

$$M_0 = \frac{1 + \tau_{\mathfrak{B}}}{2} M_N + \frac{1 - \tau_{\mathfrak{B}}}{2} M_P,$$

M_N and M_P denoting the masses of the neutron and the proton, respectively.

Then we have, in accordance with M. (21) and (22), the variational principle

$$\delta \bar{\mathcal{L}} = 0 \quad (40)$$

where ψ , ψ^\dagger and \mathbf{U}_μ should be varied independently in such a way that the variations are zero at the boundary of the region Ω . The EULER equations corresponding to variations of the \mathbf{U}_μ are, as already mentioned, identical with the field equations (16). Similarly, variation of ψ and ψ^\dagger leads to the wave-equations of the nucleons. In this derivation, we can utilize the undor form (38) of $\bar{\mathcal{L}}$. Remembering that the undor \mathbf{X} , only, and not Ψ depends upon ψ and ψ^\dagger , we get

$$\left\{ \gamma_\mu D_\mu + \frac{M_0 c}{\hbar} \right\} \psi + \frac{i}{\hbar c} \Psi^\dagger \frac{\partial \mathbf{X}}{\partial \psi^\dagger} = 0 \quad (41)$$

$$\psi^\dagger \left\{ D_\mu \gamma_\mu - \frac{M_0 c}{\hbar} \right\} - \frac{i}{\hbar c} \Psi^\dagger \frac{\partial \mathbf{X}}{\partial \psi} = 0 \quad (42)$$

with

$$\Psi^\dagger \frac{\partial \mathbf{X}}{\partial \psi} = \sum_{k_1, k_2=1}^4 \Psi_{k_1 k_2}^\dagger \frac{\partial \mathbf{X}_{k_1 k_2}}{\partial \psi}.$$

In terms of the tensor field variables these equations become

$$\left\{ \gamma_\mu D_\mu + \frac{M_0 c}{\hbar} - \frac{i g_1}{\hbar c} \mathbf{U}_\mu \mathbf{T} \gamma_\mu + \frac{i g_2}{4 \kappa \hbar c} \mathbf{G}_{\mu\nu} \mathbf{T} [\gamma_\mu, \gamma_\nu] \right\} \psi = 0 \quad (43)$$

$$\psi^\dagger \left\{ D_\mu \gamma_\mu - \frac{M_0 c}{\hbar} + \frac{i g_1}{\hbar c} \mathbf{U}_\mu \mathbf{T} \gamma_\mu - \frac{i g_2}{4 \kappa \hbar c} \mathbf{G}_{\mu\nu} \mathbf{T} [\gamma_\mu, \gamma_\nu] \right\} = 0 \quad (44)$$

which are the same as (15) and (15') in M.

Multiplying (41) by ψ^\dagger from the left and (42) by ψ from the right, and adding, we get

$$D_\mu \psi^\dagger \gamma_\mu \psi = 0. \quad (45)$$

Furthermore, multiplying (41) by $\frac{1}{2} \psi^\dagger \tau_3$ from the left and (42) by $\frac{1}{2} \tau_3 \psi$ from the right, and adding, we obtain, by using the ordinary commutation relations of the isotopic spin matrices¹:

$$D_\mu \frac{1}{2} \psi^\dagger \tau_3 \gamma_\mu \psi = \frac{1}{\hbar c} (\Psi^\dagger \mathbf{A} \mathbf{X})_3. \quad (46)$$

Here we have used the formulae

$$\frac{\partial \mathbf{X}}{\partial \psi} \psi = \mathbf{X}, \quad \psi^\dagger \frac{\partial \mathbf{X}}{\partial \psi^\dagger} = \mathbf{X}$$

indicating that \mathbf{X} is a homogeneous function of the first degree in ψ and ψ^\dagger , respectively.

¹ The symbol \mathbf{A} indicates, as in MR. and M., a symbolical vector multiplication, e. g. $(\mathbf{A} \mathbf{A} \mathbf{B})_3 = A_1 B_2 - A_2 B_1$.

In a similar way, we get from (19) and (31), using (29) and (30),

$$\left. \begin{aligned} D_\mu \left(\Psi^\dagger \boldsymbol{\Lambda} \frac{\gamma_\mu^{(1)} - \gamma_\mu^{(2)}}{2} \Psi \right) &= \\ = \left(\Psi^\dagger D_\mu \frac{\gamma_\mu^{(1)} - \gamma_\mu^{(2)}}{2} \right) \boldsymbol{\Lambda} \Psi + \Psi^\dagger \boldsymbol{\Lambda} \left(\frac{\gamma_\mu^{(1)} - \gamma_\mu^{(2)}}{2} D_\mu \Psi \right) &= \\ = \mathbf{X}^\dagger \boldsymbol{\Lambda} \Psi - \Psi^\dagger \boldsymbol{\Lambda} \mathbf{X} = -2 (\Psi^\dagger \boldsymbol{\Lambda} \mathbf{X}). \end{aligned} \right\} \quad (47)$$

From (45), (46), and (47) we finally see that the five-vector

$$S_\mu = \psi^\dagger \frac{1 - \tau_3}{2} \psi - \frac{1}{4\hbar c} \left(\Psi^\dagger \boldsymbol{\Lambda} (\gamma_\mu^{(1)} - \gamma_\mu^{(2)}) \Psi \right)_3 \quad (48)$$

satisfies the divergence relation

$$D_\mu S_\mu = 0.$$

Thus, the vectors $\psi^\dagger \gamma_\mu \psi$ and S_μ may be interpreted as the five-dimensional particle-current density and charge-current density, respectively.

The expression (48) for the charge-current density is obviously the only possible. It is true that also the quantities

$$\Psi_i^\dagger \frac{\gamma_\mu^{(1)} - \gamma_\mu^{(2)}}{2} \Psi_i \quad (49)$$

have vanishing divergence but, on account of (15) and (30), it is seen that

$$\Psi^\dagger \frac{\gamma_\mu^{(1)} - \gamma_\mu^{(2)}}{2} \Psi = \mathbf{U}_\nu \mathbf{G}_{\nu\mu} + \frac{1}{2} (\mathbf{G}_{\mu\nu} \mathbf{U}_\nu - \mathbf{G}_{\nu\mu} \mathbf{U}_\nu) = 0 \quad (50)$$

and, consequently, all quantities (49) are identically equal to zero.

9. Difficulties of the five-dimensional neutral meson theory.

If we try to establish a variational principle of the neutral meson theory given by the equations (21), (22), and (23) in the same way as it was done in section 8 for the symmetrical theory, we meet with a serious difficulty. Since the equations (21) of the neutral meson theory differ from the equations (19) of the symmetrical theory by the sign of $\gamma_{\mu}^{(2)}$, only, one would expect that a variational principle of the neutral theory would be obtained from (33) by changing the sign of $\gamma_{\mu}^{(2)}$ in (34). Such a procedure, however, does not lead to the right field equation (23), but we get a set of trivial equations in which the terms containing the operators Grad, Div and Rot in (23) are missing.

The reason for this peculiar difference between the two theories is the following. In section 8, we have in the derivation of the field equations from the variational principle made use of the relation

$$D_{\mu} \left\{ \Psi^{\dagger} \frac{\partial \frac{\gamma_{\nu}^{(1)} - \gamma_{\nu}^{(2)}}{2} D_{\nu} \Psi}{\partial (D_{\mu} \Psi_{k_1 k_2})} \right\} = - \left(\frac{\gamma_{\mu}^{(1)} - \gamma_{\mu}^{(2)}}{2} D_{\mu} \Psi \right)_{k'_1 k'_2} \quad (51)$$

where k'_1, k'_2 are determined by (cf. (30))

$$\Psi^{\dagger}_{k'_1 k'_2} = \pm \Psi_{k_1 k_2}.$$

The analogous relation in the neutral theory is

$$D_{\mu} \left\{ \Psi^{\dagger} \frac{\partial \frac{\gamma_{\nu}^{(1)} + \gamma_{\nu}^{(2)}}{2} D_{\nu} \Psi}{\partial (D_{\mu} \Psi_{k_1 k_2})} \right\} = \left(\frac{\gamma_{\mu}^{(1)} + \gamma_{\mu}^{(2)}}{2} D_{\mu} \Psi \right)_{k'_1 k'_2} \quad (52)$$

and the difference in sign of the right-hand sides in (51) and (52) makes in the EULER equations all terms containing derivatives cancel each other.

The four-dimensional theory described by (25) may, however, be derived from a variational principle with the Lagrangeian

$$\mathcal{L} = \frac{\kappa}{3!} \left(\frac{1}{2} G_{klm} G_{klm} + \frac{3}{2} G_{kl} G_{kl} - G_{klm} S_{klm} - \frac{\kappa^2}{2} U_k U_k - \frac{1}{2} K^2 + KC \right) \quad (53)$$

K and G_{klm} being the independent variables, while U_k and G_{kl} are defined by (25 b) and (25 e), respectively.

Since

$$G_{\mu\nu} G_{\mu\nu} = G_{kl} G_{kl} + \frac{1}{3} G_{klm} G_{klm} \quad (54)$$

the direct generalization of the Lagrangeian to five-dimensions will be

$$\mathcal{L} = \frac{\kappa}{2} \left\{ \frac{1}{2} G_{\mu\nu} G_{\mu\nu} - G_{\mu\nu} S_{\mu\nu} - \kappa^2 U_\mu U_\mu - K^2 + 2KC \right\}. \quad (55)$$

If we vary K , the U_μ being defined by means of (23 b), we really obtain the equation (23 a); but if we vary all the $G_{\mu\nu}$ independently—which would be a possible generalization of the four-dimensional case—we get the equations

$$G_{\mu\nu} = S_{\mu\nu} \quad (56)$$

which differ from (23 c) by the essential term $\text{Rot } G$.

On the other hand, if we also in the five-dimensional theory consider the quantities G_{klm} as independent variables, the G_{kl} being defined by (25 e), we obtain the equations

$$\left. \begin{aligned} & \kappa (G_{klm} - S_{klm}) = \\ & = D_k G_{lm} + D_l G_{mk} + D_m G_{kl} - D_k S_{lm} - D_l S_{mk} - D_m S_{kl}, \end{aligned} \right\} \quad (57)$$

which deviate from the field equations (25 d) by terms containing derivatives of the source densities.

It is easily seen that it is impossible to find any kind of variational principle of the type (33) from which the right field equations could be derived by varying all the quantities $G_{\mu\nu}$ independently. Since the corresponding function \mathcal{L}' has to be an invariant, it will, in the general case, be of the shape

$$\mathcal{L}' = a G_{\mu\nu} G_{\mu\nu} + b G_{\mu\nu} S_{\mu\nu} + c \text{Rot}_{\mu\nu} G \cdot G_{\mu\nu}$$

where a , b and c are constants. If $c = 0$, we obviously have to do with the case treated above, but even if $c \neq 0$, the last term will give no contribution to the EULER equation since the variation of

$$\text{Rot}_{\mu\nu} G \cdot G_{\mu\nu}$$

on account of the definition (24) is equal to zero for any variation of the quantities $G_{\mu\nu}$ which vanish at the boundary.

On the other hand, if we only vary the G_{klm} , the G_{kl} being again defined by (25 c), the constant c has also to be put equal to zero; otherwise we would get field equations involving derivatives of the second order, since we have (cf. (54))

$$\left. \begin{aligned} G_{\mu\nu} \text{Rot}_{\mu\nu} G &= G_{kl} \text{Rot}_{kl} G + \frac{1}{3} G_{klm} \rho\sigma\tau_{klm} G = \\ &= \frac{1}{\kappa} D_r G_{klr} \text{Rot}_{kl} G + S_{kl} \text{Rot}_{kl} G + \\ &+ \frac{1}{3} G_{klm} (D_k G_{lm} + D_m G_{kl} + D_l G_{mk}) \end{aligned} \right\} (58)$$

where the quantity $\text{Rot}_{kl} G$ contains terms like $D_m G_{klm}$ (no summation!). Consequently, the first term on the right

hand of (58) will give rise to second derivatives in the EULER equation.

Thus, the equations of the five-dimensional neutral meson theory cannot be derived from any variational principle, a circumstance which may be connected with the fact that the field equations (21)-(23) consist of two separate groups: the scalar equations (a), (b) and the tensor equations (c) which even under five-dimensional rotations are not transformed into each other. Since a Lagrangeian principle is essential for the quantization of the field equations, the impossibility of formulating such principle in the neutral theory may be taken as a strong argument in favour of the symmetrical theory which, in contrast to the neutral theory, is able to satisfy the more rigorous claims resulting from a generalization to five dimensions.

I wish to express my deep gratitude to Professor NIELS BOHR for the kind interest shown in this investigation and for the possibilities given to me to work in the Institute of Theoretical Physics, Copenhagen. Furthermore, my most hearty thanks are due Dr. C. MØLLER for his constant guidance and for many valuable discussions during the performance of the present paper.

References.

- 1) H. YUKAWA, Proc. phys.-mat. Soc. Japan **17**, 48, 1935.
- 2) A. PROCA, Journ. de Physique et le Radium **7**, 347, 1936.
- 3) H. YUKAWA and S. SAKATA, Proc. phys.-mat. Soc. Japan **19**, 1084, 1937.
See also:
H. YUKAWA, S. SAKATA and M. TAKETANI, *ibid.* **20**, 319, 1938.
H. YUKAWA, S. SAKATA, M. KOBAYASHI and M. TAKETANI, *ibid.* **20**, 720, 1938.
- 4) H. FRÖHLICH, W. HEITLER and N. KEMMER, Proc. Roy. Soc. **A 166**, 154, 1938.
- 5) H. J. BHABHA, Proc. Roy. Soc. **A 166**, 501, 1938.
- 6) E. STÜCKELBERG, Helv. Phys. Acta **11**, 225 and 299, 1938.
- 7) N. KEMMER, Proc. Roy. Soc. **A 166**, 127, 1938.
- 8) M. A. TUVE, N. P. HEYDENBURG and L. R. HAFSTAD, Phys. Rev. **50**, 806, 1936.
- 9) N. KEMMER, Proc. Cambridge Phil. Soc. **34**, 354, 1938.
- 10) C. MØLLER and L. ROSENFELD, D. Kgl. Danske Vidensk. Selskab, Math.-fys. Medd. **XVII**, 8, 1940.
- 11) H. A. BETHE, Phys. Rev. **57**, 260 and 390, 1940.
- 12) F. BELINFANTE, On the Theory of Heavy Quanta. Proefschrift, 's Gravenhage 1939.
- 13) C. MØLLER, D. Kgl. Danske Vidensk. Selskab, Math.-fys. Medd. **XVIII**, 6, 1941.

Det Kgl. Danske Videnskabernes Selskab.

Mathematisk-fysiske Meddelelser. **XIX**, 10.

PROPERTIES OF MESONS
DESCRIBED BY A PSEUDOSCALAR
WAVE-FUNCTION

BY

T. S. CHANG



KØBENHAVN

I KOMMISSION HOS EJNAR MUNKSGAARD

1942

CONTENTS

	Page
Introduction.....	3
1. Elastic scattering by a static electric field	4
2. The absorption of mesons by nucleons with emission of photons	8
3. Scattering of mesons by nucleons.....	11
4. Multiple processes by which a meson in a collision with a nucleon is transformed into a pair of mesons	12
5. The lifetime of a free PS -meson	14
6. Numerical estimates	16
References.....	17

Introduction.⁽¹⁾

The development of the meson theory of nuclear forces has shown the necessity to introduce in the description of the meson field two kinds of wave-functions with the invariance properties of a vector and a pseudoscalar, respectively [1] [2]. In a recent paper [3], the writer has calculated the probability of some elementary processes involving mesons which are described by a vector wave-function and which, in the following, for brevity will be referred to as *V*-mesons. Since, however, the mesons represented by a pseudoscalar function (*PS*-mesons) appear on the same footing in the description of the nuclear forces and since they seem to play a preponderant part in the cosmic radiation available for experiment [4], it is of interest to investigate different processes involving such mesons. In particular, the determination of the lifetime of *PS*-mesons, performed in Section 5, is of importance in the discussion of the connection between the decay con-

(1) The investigation reported in this paper was carried out by Dr. Tsung-Sui Chang during his stay at the Institute of Theoretical Physics at Copenhagen in 1938–39. Shortly after his departure, an account of the work was sent by Dr. Chang to the Institute, but since then connection has been interrupted due to the war. In view of the special interest which the results of Dr. Chang's work have obtained as a consequence of recent development of the meson theory, a further delay in the publication of his paper seemed, however, most undesirable and, therefore, Dr. C. Møller and Dr. S. Rozental have kindly undertaken to prepare Dr. Chang's account for print and to furnish it with an introduction containing references to the present state of the theory. NIELS BOHR.

stants of radioactive elements and of free mesons in cosmic radiation [5].

The difference between the properties of mesons described by a pseudoscalar function and by a simple scalar function appears only as regards the interaction with other particles, while both these kinds of mesons behave exactly in the same way towards an electromagnetic field. The probability for a process in which *PS*-mesons are scattered by a static electric field and which is treated in Section 1 is, therefore, the same as for the scattering of mesons described by a scalar wave-function. For the processes, however, treated in Sections 2—5, which involve interaction of mesons with nucleons or with electrons and neutrinos, there is a marked difference between the scalar and the pseudoscalar field. In Section 2, we consider a process in which a meson hitting a nucleon is absorbed with emission of a light quantum; in Section 3, the elastic scattering of mesons by nucleons is calculated. Section 4 contains an investigation of a multiple process in which a meson hits a nucleon creating a pair of mesons. Finally, Section 5 is devoted to the evaluation of the lifetime of the *PS*-meson.

With the exception of the processes treated in Sections 1 and 5, the formulae deduced are valid in the case of relatively slow particles, only.

1. Elastic scattering by a static electric field.

We shall start by writing down, for reference, some of the expressions and relations referring to the pseudoscalar meson field and needed in the following calculations.

The meson field equations including the interaction with the electromagnetic field and the nucleons are

$$z^2 g = \left(\frac{1}{c} \frac{\partial}{\partial t} + \frac{ie}{\hbar c} A_0 \right) \chi + \left(\vec{\nabla} - \frac{ie}{\hbar c} \vec{A} \right) \vec{\xi} + 2R \quad (1)$$

$$\vec{\xi} = \left(\vec{\nabla} - \frac{ie}{\hbar c} \vec{A} \right) g - 2\vec{P} \quad (2)$$

$$\chi = - \left(\frac{1}{c} \frac{\partial}{\partial t} + \frac{ie}{\hbar c} A_0 \right) g + 2Q. \quad (3)$$

Here, (x_1, x_2, x_3, ict) is the four-dimensional space vector, $(\xi_1, \xi_2, \xi_3, i\chi)$ is a complex pseudovector which, in the case of electrically charged mesons, together with the complex pseudoscalar g is sufficient for a complete description of the meson field, (A_1, A_2, A_3, iA_0) is the electromagnetic potential vector, while $(P_1, P_2, P_3, -iQ)$ is a pseudovector, and R a pseudoscalar responsible for the interaction between the nucleons and the meson field and depending on the variables of the nucleons, only. The constant z is the reciprocal of the range of the nuclear forces and connected with the mass M_m of the meson by the equation

$$z = \frac{M_m c}{\hbar}.$$

The quantities \vec{P} , Q , and R are given by

$$\left. \begin{aligned} \vec{P} &= \frac{f_2}{z} \psi^* \vec{\sigma} \frac{\tau_1 - i\tau_2}{2} \psi \\ Q &= \frac{f_2}{z} \psi^* \varrho_1 \frac{\tau_1 - i\tau_2}{2} \psi \\ R &= f_1 \psi^* \varrho_2 \frac{\tau_1 - i\tau_2}{2} \psi \end{aligned} \right\} \quad (4)$$

where ψ is the wave-function of the nucleons, ϱ_1 , ϱ_2 and $\vec{\sigma}$ are the ordinary DIRAC matrices, and $\frac{1}{2}(\tau_1 - i\tau_2)$ is the operator transforming the nucleons from a neutron state to a proton state. Finally, f_1 and f_2 are the two constants

of the pseudoscalar theory determining the strength of the coupling between nucleons and *PS*-mesons.

The Lagrangeian is given by

$$\frac{1}{2} \vec{\xi}^* \vec{\xi}^* - \frac{1}{2} \chi \chi^* + \frac{1}{2} z^2 \varphi \varphi^* + (\vec{P} \vec{\xi}^* + \vec{\xi} \vec{P}^*) + (Q \chi^* + \chi Q^*). \quad (5)$$

By varying χ^* and $\vec{\xi}^*$, while φ^* is considered as a function of χ^* and $\vec{\xi}^*$ given by the equation complex conjugate to (1), we get in the usual way the equations (2) and (3).

The part of the Hamiltonian which involves the meson field variables is

$$\left. \begin{aligned} & \frac{1}{2} z^2 \varphi \varphi^* + \frac{1}{2} \chi \chi^* + \frac{1}{2} (\nabla \varphi) (\nabla \varphi^*) \\ & + \frac{1}{2} \frac{ie}{\hbar c} (A_0 (\varphi \chi^* - \varphi^* \chi) + \frac{1}{2} \frac{ie}{\hbar c} [\varphi^* (\vec{A} \cdot \nabla \varphi) - \varphi (\vec{A} \cdot \nabla \varphi^*)] \\ & - (\varphi^* R + R^* \varphi) - (Q \chi^* + Q^* \chi) - (\vec{P} \cdot \nabla \varphi^* + \vec{P}^* \cdot \nabla \varphi) + \\ & \quad + \frac{ie}{\hbar c} [(\vec{A} \cdot \vec{P}^*) \varphi - (\vec{A} \vec{P}) \varphi^*]. \end{aligned} \right\} \quad (6)$$

The commutation rules for the field variables are

$$[\chi(x), \varphi^*(x')] = [\chi^*(x), \varphi(x')] = 2i \hbar c \delta(x - x') \quad (7)$$

and for the FOURIER expansion of these quantities we have

$$\left. \begin{aligned} \chi &= \sum_{\vec{k}} i \sqrt{\hbar c} (k^2 + z^2)^{\frac{1}{4}} [u(\vec{k}) e^{i(\vec{k} \cdot \vec{x})} - v^*(\vec{k}) e^{-i(\vec{k} \cdot \vec{x})}] \\ \chi^* &= \sum_{\vec{k}} -i \sqrt{\hbar c} (k^2 + z^2)^{\frac{1}{4}} [u^*(\vec{k}) e^{-i(\vec{k} \cdot \vec{x})} - v(\vec{k}) e^{i(\vec{k} \cdot \vec{x})}] \\ \varphi &= \sum_{\vec{k}} \sqrt{\hbar c} (k^2 + z^2)^{-\frac{1}{4}} [u(\vec{k}) e^{i(\vec{k} \cdot \vec{x})} + v^*(\vec{k}) e^{-i(\vec{k} \cdot \vec{x})}] \\ \varphi^* &= \sum_{\vec{k}} \sqrt{\hbar c} (k^2 + z^2)^{-\frac{1}{4}} [u^*(\vec{k}) e^{-i(\vec{k} \cdot \vec{x})} + v(\vec{k}) e^{i(\vec{k} \cdot \vec{x})}] \end{aligned} \right\} \quad (8)$$

where $u(\vec{k})$, $u^*(\vec{k})$, $v(\vec{k})$ and $v^*(\vec{k})$ are the operators corresponding to annihilation and creation of positive and negat-

ive mesons with momentum $\vec{k}\hbar$. The summation is extended over all values of \vec{k} for which the function $e^{i(\vec{k}\vec{x})}$ has a whole number of periods between two opposite faces of a big cube with the volume L^3 . In (8) and in the following formulae we have, for simplicity, put $L^3 = 1$.

For elastic scattering of *PS*-mesons by a static electric field, the only important terms of the Hamiltonian are

$$\frac{1}{2}z^2\varphi\varphi^* + \frac{1}{2}\chi\chi^* + \frac{1}{2}(\vec{\nabla}\varphi)(\vec{\nabla}\varphi^*) + \frac{1}{2}\frac{ie}{\hbar c}A_0(\varphi\chi^* - \varphi^*\chi) \quad (9)$$

where the last term may be considered as a perturbation leading to a scattering process, in which the incident meson is annihilated and a new meson is created. Let us suppose that all the incident mesons are of negative charge and of momentum $\vec{k}_0\hbar$. Further, let $\vec{k}\hbar$ be the momentum of the scattered meson, and

$$\left. \begin{aligned} \epsilon_0 &= \hbar c\sqrt{k_0^2 + z^2} \\ \epsilon &= \hbar c\sqrt{k^2 + z^2} \end{aligned} \right\} \quad (10)$$

be the energies of the incident and the scattered meson, respectively. Making use of (8), we easily find that the matrix element corresponding to the process in question is given by the expression

$$-\sqrt{n_0} \int e A_0 e^{i(\vec{k}_0 - \vec{k})\vec{x}} dV \cdot \frac{1}{2} \left(\sqrt{\frac{\epsilon}{\epsilon_0}} + \sqrt{\frac{\epsilon_0}{\epsilon}} \right) \quad (11)$$

where n_0 is the number of incident mesons. Since we have conservation of energy, i. e. $\epsilon = \epsilon_0$, we get for this matrix element

$$-\sqrt{n_0} \int e A_0 e^{i(\vec{k}_0 - \vec{k})\vec{x}} dV. \quad (12)$$

In case that the incident mesons are positively charged we get the same expression (12) but with opposite sign.

The result is exactly the same as for the scattering of mesons described by a scalar field.

2. The absorption of mesons by nucleons with emission of photons.

Let us consider a process in which a proton absorbs a negatively charged meson and simultaneously is transformed into a neutron under emission of a photon, i. e. a process given by the scheme



The calculations which refer to a process involving a positively charged meson



are quite analogous and lead to the same results.

Since the theory breaks down for all processes in which the DE BROGLIE wave-length of the particles involved is very small, we shall assume that the momentum of the nucleon both before and after the process is small compared with Mc , M being the mass of the nucleon. We may, therefore, neglect the terms containing Q , R , Q^* , and R^* in the Hamiltonian. This restricts, of course, the validity of the calculations to such cases, only, in which the difference between the momenta of the meson and the photon is not too large. The remaining perturbation terms giving rise to the transition in question are now simply

$$\left. \begin{aligned} \frac{1}{2} \frac{ie}{\hbar c} [\varphi^* (\vec{A} \cdot \vec{\nabla} \varphi) - \varphi (\vec{A} \cdot \vec{\nabla} \varphi^*)] - (\vec{P} \cdot \vec{\nabla} \varphi^* + \vec{P}^* \cdot \vec{\nabla} \varphi) + \\ + \frac{ie}{\hbar c} [(\vec{A} \vec{P}^*) \varphi - (\vec{A} \vec{P}) \varphi^*]. \end{aligned} \right\} \quad (14)$$

Just as in the case of V -mesons [3], the last term leads to a direct process (13 a), while the other terms may give rise to two different second-order processes:

1) annihilation of the incident meson and creation of another negative meson M'^- with simultaneous emission of the photon, followed by the absorption of M'^- by the proton;

2) emission of a positive meson M''^+ by the proton, followed by annihilation of M''^+ and of the incident meson with creation of the photon.

We denote the momentum of the incident meson by $\vec{k}\hbar$, the momentum of the photon by $\vec{k}_\nu\hbar$, and the momenta and energies of the mesons M'^- and M''^+ by $\vec{k}'\hbar$, $\vec{k}''\hbar$, ε' and ε'' . We have, of course,

$$\left. \begin{aligned} \vec{k}' &= \vec{k} - \vec{k}_\nu \\ \vec{k}'' &= \vec{k}_\nu - \vec{k} \\ \varepsilon' &= \hbar c \sqrt{k'^2 + z^2} \\ \varepsilon'' &= \hbar c \sqrt{k''^2 + z^2}. \end{aligned} \right\} \quad (15)$$

Through a calculation similar to that indicated in [3], we get for the matrix element corresponding to the process (13 a) the expression

$$\left. \begin{aligned} V &= -\frac{1}{\varepsilon'} \left[-i \frac{f_2}{z} \hbar c \sqrt{\frac{1}{\varepsilon'}} (\psi_N^* \vec{\sigma} \psi_P \cdot \vec{k}') \right] \\ &\quad \left[\frac{1}{2} e (\hbar c)^2 \sqrt{\frac{1}{\hbar \nu}} \sqrt{\frac{1}{\varepsilon \varepsilon'}} (\vec{e} \vec{k} + \vec{e} \vec{k}') \right] \\ &\quad - \frac{1}{\varepsilon'} \left[-\frac{1}{2} e (\hbar c)^2 \sqrt{\frac{1}{\hbar \nu}} \sqrt{\frac{1}{\varepsilon \varepsilon''}} (\vec{e} \vec{k}'' - \vec{e} \vec{k}) \right] \\ &\quad \left[i \frac{f_2}{z} \hbar c \sqrt{\frac{1}{\varepsilon''}} (\psi_N^* \vec{\sigma} \psi_P \cdot \vec{k}'') \right] \\ &\quad - \frac{ie}{\hbar c} \frac{f_2}{z} (\psi_N^* \vec{\sigma} \psi_P \cdot \vec{e}) (\hbar c)^2 \sqrt{\frac{1}{\hbar \nu}} \sqrt{\frac{1}{\varepsilon}} \end{aligned} \right\} \quad (16)$$

where the unit vector \vec{e} denotes the direction of polarization of the photon emitted, and ψ_N and ψ_P are the amplitudes of the plane waves representing the states of the neutron and the proton, respectively. From (15) we have $\vec{k}' = -\vec{k}''$ and $\epsilon' = \epsilon''$ and, denoting the energy $k\hbar c$ of the photon by ϵ , we get for the matrix element V the expression

$$V = i \frac{f_2}{z} e \hbar c \frac{\sqrt{2\pi}}{\epsilon} \psi_N^* \vec{\sigma} \psi_P \cdot \left[\frac{2\vec{k}' (\vec{k} \vec{e})}{k'^2 + z^2} - \vec{e} \right]. \quad (17)$$

According to the usual perturbation theory, the probability per unit time of the emission of a photon into a solid angle $d\Omega_\nu$ is equal to

$$\frac{2\pi}{\hbar} \sum_{\vec{e}} |V|^2 \cdot \frac{\nu^2 d\Omega_\nu}{c^3} \left(\frac{d\nu}{d\epsilon} \right) \quad (18)$$

and, since

$$\frac{d\epsilon}{d\nu} = h$$

we get, remembering that the nucleon momentum is small,

$$\frac{e^2}{\hbar^2 c} \left(\frac{f_2}{z} \right)^2 \frac{1}{\pi} \left[1 - \frac{2k^2 z^2 \sin^2(\vec{k} \vec{\nu})}{(k'^2 + z^2)^2} \right] d\Omega_\nu. \quad (19)$$

An integration over all directions of the photon gives the final expression

$$4 \frac{e^2}{\hbar^2 c} \left(\frac{f_2}{z} \right)^2 \left[1 + \frac{z^2}{k_\nu^2} + \frac{z^2}{kk_\nu} \log \frac{k_\nu - k}{z} \right] \quad (20)$$

which, as may be seen, for large energies of the incident meson tends towards

$$4 \frac{e^2}{\hbar^2 c} \left(\frac{f_2}{z} \right)^2. \quad (20')$$

For the inverse process

$$N + h\nu \rightarrow P + M^- \quad (21)$$

we get similar results. The probability per unit time for the absorption of the photon and the emission of the meson into a solid angle $d\Omega_k$ is, with the same notations as above, equal to

$$\frac{1}{2\pi} \frac{e^2}{\hbar^2 c} \left(\frac{f_2}{z}\right)^2 \sqrt{\frac{k^2}{k^2+z^2}} \left[1 - \frac{4z^2(\vec{k}\cdot\vec{e})^2}{(k'^2+z^2)^2}\right] d\Omega_k \quad (22)$$

whence the total probability is obtained by integrating over $d\Omega_k$:

$$2 \frac{e^2}{\hbar^2 c} \left(\frac{f_2}{z}\right)^2 \sqrt{\frac{k^2}{k^2+z^2}} \left[1 + \frac{z^2}{k_\nu^2} + \frac{z^2}{k k_\nu} \log \frac{k_\nu - k}{z}\right]. \quad (23)$$

3. Scattering of mesons by nucleons.

For definiteness, we shall consider the process [6]

$$M_0^- + P \rightarrow P + M^- \quad (24 a)$$

the process

$$M_0^+ + N \rightarrow N + M^+ \quad (24 b)$$

being quite analogous.

Let $\vec{k}_0 \hbar$ be the momentum of the incident meson M_0^- , $\vec{k} \hbar$ that of the scattered meson M^- , and ϵ_0 and ϵ the corresponding energies. Let, furthermore, ψ_0 , ψ and ψ_N be the amplitudes of the plane waves representing the initial, final and intermediate (neutron) states of the nucleon.

Again we assume that the momenta of the nucleon in these states are small compared with Mc . The matrix element for the process turns out to be

$$V = \frac{1}{\epsilon_0} \left[i \frac{f_2}{z} \hbar c \sqrt{\frac{1}{\epsilon}} \psi^* \vec{\sigma} \psi_N \cdot \vec{k} \right] \left[-i \frac{f_2}{z} \hbar c \sqrt{\frac{1}{\epsilon_0}} \psi_N^* \vec{\sigma} \psi_0 \cdot \vec{k}_0 \right] \quad (25)$$

whence, after summation over the possible spin states, we get

$$|V|_{\text{sum}}^2 = \left(\frac{f_2}{z}\right)^4 k^2 k_0^2 \frac{1}{\sqrt{k^2 + z^2}} \frac{1}{\sqrt{k_0^2 + z^2}}. \quad (26)$$

Since the mass of the meson is small compared with M , the law of conservation of energy gives

$$k_0 = k \quad (27)$$

and, thus, we get for the total scattering probability per unit time the expression

$$\frac{1}{\pi} \frac{1}{\hbar^2 c} \left(\frac{f_2}{z}\right)^4 \frac{k_0^5}{\sqrt{k_0^2 + z^2}}. \quad (28)$$

4. Multiple processes, by which a meson in a collision with a nucleon is transformed into a pair of mesons.

In this section, too, we confine ourselves to the consideration of the first of the two analogous processes

$$\left. \begin{aligned} M_0^+ + N &\rightarrow P + M^+ + M^- \\ M_0^- + P &\rightarrow N + M^+ + M^- \end{aligned} \right\} \quad (29)$$

The process is a three-stage process which can take place in two different ways just as in the case of V -mesons [6]. Let the momenta and the energies of the mesons M_0^+ , M^+ and M^- be $\vec{k}_0 \hbar$, $\vec{k}^+ \hbar$, $\vec{k}^- \hbar$, ε_0^+ , ε^+ and ε^- , respectively. Under similar conditions as before, the law of conservation of energy gives

$$\varepsilon_0^+ = \varepsilon^+ + \varepsilon^-. \quad (30)$$

We denote the amplitudes of the wave-functions of the nucleons in the initial state by ψ_0 , and in the final state by ψ . Similarly, the corresponding amplitudes of the intermediate states are denoted by ψ'_I and ψ''_{II} or ψ''_I and ψ''_{II} , according to the process taking place in the one or the other of the two ways mentioned above.

For the matrix element we get

$$\left. \begin{aligned}
 & V = \frac{1}{\varepsilon_0^+ (\varepsilon_0^+ - \varepsilon^+)} \left(-\frac{f_2}{z} \right)^3 \\
 & \left\{ \psi^* \vec{\sigma} \psi'_{II} (-i\vec{k}^-) \cdot \frac{\hbar c}{\sqrt{\varepsilon^-}} \cdot \psi'^* \vec{\sigma} \psi'_I (-i\vec{k}^+) \frac{\hbar c}{\sqrt{\varepsilon^+}} \psi'_I \vec{\sigma} \psi_0 (i\vec{k}_0) \frac{\hbar c}{\sqrt{\varepsilon_0^+}} \right\} \\
 & + \frac{1}{(-\varepsilon^-) (-\varepsilon^- - \varepsilon^+)} \left(-\frac{f_2}{z} \right)^3 \\
 & \left\{ \psi^* \vec{\sigma} \psi''_{II} \cdot i\vec{k}_0 \cdot \frac{\hbar c}{\sqrt{\varepsilon_0^+}} \cdot \psi''^* \vec{\sigma} \psi'_I (-i\vec{k}^+) \frac{\hbar c}{\sqrt{\varepsilon^+}} \psi''^* \vec{\sigma} \psi_0 (-i\vec{k}^-) \cdot \frac{\hbar c}{\sqrt{\varepsilon^-}} \right\}
 \end{aligned} \right\} \quad (31)$$

The first term corresponds to the succession of events in which M_0^+ is first absorbed, and M^+ and M^- are then emitted, while the second term is connected with a process in which the absorption of M_0^+ takes place after the emission of M^+ and M^- .

After the summation over the possible spin states is carried out, we obtain

$$\left. \begin{aligned}
 & |V|_{\text{sum}}^2 = \left(\frac{f_2}{z} \right)^6 (\hbar c)^6 2 \frac{1}{(\varepsilon^- \varepsilon_0^+)^2} \frac{1}{\varepsilon^- \varepsilon^+ \varepsilon_0^+} \\
 & \cdot \left\{ (\vec{k}_0^+ \vec{k}^+)^2 (k^-)^2 + (\vec{k}^+ \vec{k}^-)^2 k_0^2 + (\vec{k}^- \vec{k}_0^+)^2 (k^+)^2 - \right. \\
 & \left. - 2 (\vec{k}_0^+ \vec{k}^+) (\vec{k}^+ \vec{k}^-) (\vec{k}^- \vec{k}_0^+) \right\}
 \end{aligned} \right\} \quad (32)$$

The probability per unit time of the process taking place is, as usual, given by the expression

$$\frac{2\pi}{\hbar} |V|_{\text{sum}}^2 \frac{k^{+2} k^{-2} d\Omega^+ d\Omega^-}{(8\pi^3)^2} dk^- \frac{dk^+}{d\varepsilon}. \quad (33)$$

Substituting the value of (32) and performing the integration over the angles $d\Omega^+$ and $d\Omega^-$ in the direction of the momenta \vec{k}^+ and \vec{k}^- , we are finally led to the expression

$$\frac{7}{9} \frac{2}{\pi^3} \frac{1}{\hbar} \left(\frac{f_2}{z}\right)^6 \frac{1}{\hbar^2 c^2} \frac{(k_0^+)^2 (k^+)^3 (k^-)^4 dk^-}{\{(k_0^+)^2 + z^2\}^{3/2} \{(k^-)^2 + z^2\}^{3/2}} \quad (34)$$

for the probability per unit time for the emission of the meson M^- with a momentum lying between k^- and $k^- + dk^-$, the momentum k^+ in this last formula being connected with k^- by the formula (30).

5. The lifetime of a free PS -meson.

According to the fundamental assumptions of the meson theory, a meson may disintegrate spontaneously into an electron and a neutrino. This assumption directly verified by experiments [7] is also of importance for the theory of β -disintegration both as regards the V -mesons [8] and the PS -mesons [9]. In order to calculate the probability of disintegration of a PS -meson [10] we have to consider the terms in the Hamiltonian which describe the interaction between PS -mesons and light particles. These terms are obtained by replacing in the Hamiltonian (5) the quantities \vec{P}, Q, R by the analogous quantities⁽¹⁾

$$\left. \begin{aligned} \vec{P} &= \frac{\tilde{f}_2}{z} \tilde{\psi}^* \tilde{\sigma} \frac{\tilde{\tau}_1 - i\tilde{\tau}_2}{2} \tilde{\psi} \\ \vec{Q} &= \frac{\tilde{f}_2}{z} \tilde{\psi}^* \tilde{q}_1 \frac{\tilde{\tau}_1 - i\tilde{\tau}_2}{2} \tilde{\psi} \\ \vec{R} &= \tilde{f}_1 \tilde{\psi}^* \tilde{q}_2 \frac{\tilde{\tau}_1 - i\tilde{\tau}_2}{2} \tilde{\psi}_1 \end{aligned} \right\} \quad (35)$$

where, in general, the symbol \tilde{A} denotes the same quantity with respect to light particles as A with respect to nucleons.

⁽¹⁾ It should be remarked that the constants \tilde{f}_1 and \tilde{f}_2 in this paper are chosen in a different way as those in ref. [9], so that the ratio between these constants is of opposite sign.

In the process in question a negative meson, say, with the momentum $\vec{k}\hbar$ will be annihilated, raising at the same time a light particle from a neutrino-state with the momentum \vec{p}_σ , the energy E_σ , and described by the wave-function $\psi_\sigma e^{\frac{i}{\hbar}(\vec{p}_\sigma \cdot \vec{x})}$ to an electron state with the momentum \vec{p}_s , the energy E_s and described by the wave-function $\psi_s e^{\frac{i}{\hbar}(\vec{p}_s \cdot \vec{x})}$. The corresponding matrix element is evidently given by

$$V = -\frac{\check{f}_2}{z} i \frac{\hbar c}{\sqrt{\epsilon}} \psi_s^* \check{\sigma} \psi_\sigma \cdot \vec{k} - \frac{\check{f}_2}{z} i \sqrt{\epsilon} \psi_s^* \check{q}_1 \psi_\sigma - \check{f}_1 \frac{\hbar c}{\sqrt{\epsilon}} \psi_s^* \check{q}_2 \psi_\sigma. \quad (36)$$

From the laws of conservation of momentum and energy it follows that

$$\vec{p}_\sigma + \vec{k}\hbar = \vec{p}_s \quad (37)$$

and

$$\hbar c \sqrt{k^2 + z^2} + E_\sigma = E_s. \quad (38)$$

For the frame of reference in which the meson is at rest ($\vec{k} = 0$, $\vec{p}_s = \vec{p}_\sigma$), the above expression for V reduces to

$$V = -\sqrt{\frac{\hbar c}{z}} (\check{f}_2 i \psi_s^* \check{q}_1 \psi_s + \check{f}_1 \psi_s^* \check{q}_2 \psi_s). \quad (39)$$

Squaring and summing over the possible states of the electron and the neutrino, we get

$$\left. \begin{aligned} |V|_{\text{sum}}^2 &= \sqrt{\frac{\hbar c}{z}} \frac{1}{E_s E_\sigma} \\ &\{E_s E_\sigma (\check{f}_1^2 + \check{f}_2^2) + 2E_\sigma mc^2 \check{f}_1 \check{f}_2 + c^2 (\vec{p}_s \cdot \vec{p}_\sigma) (\check{f}_2^2 - \check{f}_1^2)\}. \end{aligned} \right\} \quad (40)$$

Substituting this value into the expression for the probability per unit time of disintegration of the meson

$$\frac{2\pi}{\hbar} \int \frac{p_\sigma^2 dp_\sigma d\Omega}{h^3} |V|_{\text{sum}}^2 \delta(E_s - E_\sigma - \hbar c z) \quad (41)$$

we obtain for the decay constant λ_{PS} of a PS -meson the formula

$$\lambda_{PS} = \frac{z}{4\pi\hbar} \left(1 - \frac{m^2 c^2}{z^2 \hbar^2}\right)^2 \left(\tilde{f}_1 + \frac{mc}{z\hbar} \tilde{f}_2\right)^2. \quad (42)$$

The lifetime $\tau_{PS} = 1/\lambda_{PS}$ thus found is that of a meson at rest. For a meson moving with velocity v the lifetime will be given by the expression

$$\frac{\tau_{PS}}{\sqrt{1 - \frac{v^2}{c^2}}}.$$

6. Numerical estimates.

For a numerical estimate of the probabilities or the corresponding cross-sections (the probability divided by the velocity of the incident particles) it is necessary to know the value of the constant f_2 . Taking the value determined from the theory of nuclear forces (ref. [1], formulae (54) and (107)), and putting

$$\frac{1}{z} = \frac{1}{150} \frac{\hbar}{mc} \infty 2,6 \times 10^{-13} \text{ cm}$$

we get for the process 2 a cross-section of the order of $4,5 \cdot 10^{-27} \text{ cm}^2$ for $k = 0,5z$ and $2,0 \cdot 10^{-27} \text{ cm}^2$ for $k = 2z$. For the process 3, the cross-section is $1,0 \cdot 10^{-27} \text{ cm}^2$ for $k_0 = 0,5z$, $6,5 \cdot 10^{-26} \text{ cm}^2$ for $k_0 = 2z$ and $5 \cdot 10^{-25} \text{ cm}^2$ for $k_0 = 5z$. Finally, the cross-section for the process 4 is of the order of 10^{-31} cm^2 for $k_0 = 2z$, and 10^{-26} cm^2 for $k_0 = 5z$.

The lifetime of the PS -mesons calculated in Section 5 depends upon the constants \tilde{f}_1 and \tilde{f}_2 . For the discussion of this question and its connection with the lifetimes of β -radioactive elements, compare ref. [5].

In conclusion, the writer wishes to express his gratitude to Professor NIELS BOHR for the possibilities offered to him during his stay at the Institute of Theoretical Physics at Copenhagen. Furthermore, I wish to thank most heartily Dr. C. MØLLER for his continued interest in this investigation and for many helpful discussions.

References.

- [1] C. MØLLER and L. ROSENFELD, D. Kgl. Danske Vidensk. Selskab, Math.-fys. Medd. **XVII**, 8 (1940).
- [2] C. MØLLER, *ibid.* **XVIII**, 6 (1941).
- [3] T. S. CHANG, Proc. Cambridge Phil. Soc. **36**, 34 (1940).
- [4] R. F. CHRISTY and S. KUSAKA, Phys. Rev. **59**, 414 (1941).
J. R. OPPENHEIMER, *ibid.* **59**, 462 (1941).
L. NORDHEIM, *ibid.* **59**, 554 (1941).
- [5] S. ROZENTAL, Phys. Rev., in print.
- [6] For similar calculations for V -mesons, see W. HEITLER, Proc. Roy. Soc. A **166**, 529 (1938).
- [7] E. J. WILLIAMS and G. E. ROBERTS, Nature **145**, 102 (1940).
- [8] H. YUKAWA, S. SAKATA and M. TAKETANI, Proc. Phys. Math. Soc. Japan **20**, 319 (1938).
- [9] S. ROZENTAL, D. Kgl. Danske Vidensk. Selskab, Math.-fys. Medd. **XVIII**, 7 (1941).
- [10] As regards the V -mesons, see:
H. YUKAWA, S. SAKATA and M. TAKETANI, Proc. Phys. Math. Soc. Japan **20**, 319 (1938).
H. YUKAWA, S. SAKATA, M. KOBAYASHI and M. TAKETANI, *ibid.* **20**, 726 (1938).
H. A. BETHE and L. W. NORDHEIM, Phys. Rev. **57**, 998 (1940).

Det Kgl. Danske Videnskabernes Selskab.

Mathematisk-fysiske Meddelelser. **XIX**, 11.

ÜBER DEN EINFLUSS DES
AKKOMMODATIONSKOEFFIZIENTEN
AUF DIE WÄRMELEITUNG UND
RADIOMETERKRAFT IN GASEN

VON

SOPHUS WEBER



KØBENHAVN

I KOMMISSION HOS EJNAR MUNKSGAARD

1942

Printed in Denmark.
Bianco Lunos Bogtrykkeri A/S.

1. In einer früheren Arbeit¹⁾ wurde nachgewiesen, daß der Wärmetransport in Gasen im MAXWELL'schen Zustande auch aufgrund der einfachen kinetischen Theorie richtig berechnet werden kann, wenn angenommen wird, dass die Moleküle, welche von dem mittleren Abstand λ kommen, wo λ die mittlere freie Weglänge ist, nicht die Temperatur des Gases im Abstände λ , sondern die Temperatur des Gases im mittleren Abstand $\frac{1}{2}k\lambda$ besitzen; bei der praktischen Durchführung der Berechnungen kann also angenommen werden, dass die Moleküle nicht von dem Abstand λ , sondern von dem Abstand $\frac{1}{2}k\lambda$ kommen.

Für die Wärmeleitfähigkeit des Gases im MAXWELL'schen Zustande wurde hiernach abgeleitet:

$$K = k \cdot \varepsilon \cdot p \lambda, \quad (1)$$

Für einatomige Gase ist $k = \frac{15}{4}$, und für mehratomige Gase:

$$k = \frac{9\gamma - 5}{\gamma + 1}, \quad \text{wo } \gamma = \frac{c_p}{c_v},$$

p ist der Druck des Gases und ε die von MARTIN KNUDSEN eingeführte molekulare Wärmeleitfähigkeit; wenn der Zustand des Gases sich ändert und vom MAXWELL'schen

¹⁾ SOPHUS WEBER: D. Kgl. Danske Vidensk. Selskab, Math.-fys. Medd. XVI, 9, 1939, S. 14.

in den molekularen KNUDSEN-Zustand übergeht, wird k sich wahrscheinlich ändern von $k = \frac{15}{4}$ in $k = 2$, in Übereinstimmung damit, dass das Verhältnis $f = \frac{K}{c_v \cdot \eta}$ sich von $f = \frac{5}{2}$ in $f = \frac{4}{3}$ ändert.

2. Wir wollen zunächst die Wärmeleitung zwischen zwei unendlich grossen, parallelen Platten A und B (vgl. Fig. 1) im Abstand d und mit den Temperaturen T_1 und T_2 bzw. den Akkommodationskoeffizienten a_1 und a_0 betrach-

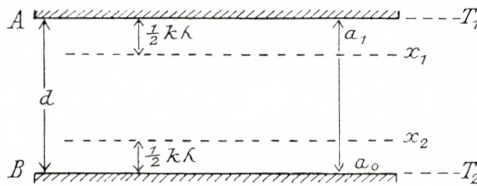


Fig. 1.

ten. Wir erhalten dann für den Wärmetransport per cm^2 , q , wenn die Temperaturen des Gases im Abstand $\frac{1}{2} k \lambda$ von den Platten mit X_1 und X_2 bezeichnet werden:

$$q = a_1 \cdot p (T_1 - X_1) \varepsilon_{x_1} = a_0 \cdot p (X_2 - T_2) \varepsilon_{x_2} = \frac{K(X_1 - X_2)}{d - 2 \cdot \frac{1}{2} k \lambda}.$$

Bei kleinen Temperaturdifferenzen können ε_{x_1} , ε_{x_2} und K bei der Temperatur $T = \frac{1}{2}(T_1 + T_2)$ befindlich angenommen werden.

Mit $K = k \varepsilon p \lambda$ und nach Elimination von X_1 und X_2 erhalten wir hieraus:

$$q = \frac{K(T_1 - T_2)}{d \left[1 + \left(\frac{1}{a_1} + \frac{1 - a_0}{a_0} \right) \frac{k \lambda}{d} \right]} = \frac{K(T_1 - T_2)}{d + \left(\frac{2 - a_1}{2 a_1} + \frac{2 - a_0}{2 a_0} \right) k \lambda}.$$

Diese Formel für die Wärmeleitung zwischen zwei unendlich grossen, parallelen Platten kann, wenn wir setzen:

$$q_p = \frac{q}{T_1 - T_2}, \quad q_0 = \frac{a_1 a_0}{a_1 + a_0 - a_1 a_0} \cdot \varepsilon p \quad \text{und} \quad q_\infty = \frac{K}{d},$$

auch folgendermassen geschrieben werden:

$$\frac{1}{q_p} = \frac{1}{q_0} + \frac{1}{q_\infty}$$

oder, weil $K = k \varepsilon p \lambda$:

$$q_p = \frac{q}{T_1 - T_2} = \frac{a_1 a_0}{a_1 + a_0 - a_1 a_0} \cdot \varepsilon p \cdot \frac{1}{1 + \frac{1}{k} \cdot \frac{a_1 a_0}{a_1 + a_0 - a_1 a_0} \cdot \frac{d}{\lambda}}$$

Im KNUDSEN-Zustand, d. h. $\frac{d}{\lambda} = 0$, wird $k = 2$ und also:

$$\lim_{\frac{d}{\lambda} \rightarrow 0} q_p = \frac{q}{T_1 - T_2} = \frac{a_1 a_0}{a_1 + a_0 - a_1 a_0} \cdot \varepsilon p \cdot \frac{1}{1 + \frac{1}{2} \frac{a_1 a_0}{a_1 + a_0 - a_1 a_0} \cdot \frac{d}{\lambda}} \quad (2)$$

Die obenstehende Formel (2) kann man auch auf direkte Weise ableiten, wenn man von dem rein molekularen Zustand, d. h. $\frac{d}{\lambda} = 0$, ausgeht und den Einfluss vereinzelter Zusammenstösse berechnet.

Da das Beobachtungsmaterial zur Nachprüfung der Formel (2) leider fehlt, werde ich hierauf nicht näher eingehen.

3. Betrachten wir die Wärmeleitung zwischen zwei coaxialen Zylindern mit den Radien r_0 und r_1 ($r_0 > r_1$) (die SCHLEIERMACHER'sche Methode)²⁾, finden wir in Analogie zu den parallelen Platten für einatomige Gase (Fig. 2):

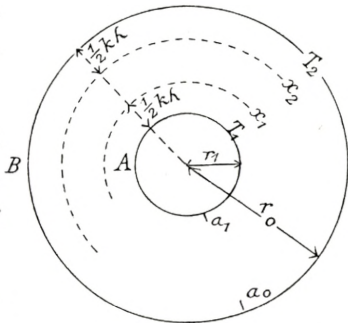


Fig. 2.

²⁾ A. SCHLEIERMACHER: W. A., 34, 1888, S. 623.

$$\begin{aligned}
 q' &= 2\pi r_1 a_1 \varepsilon p (T_1 - X_1) = 2\pi r_0 a_0 \varepsilon p (X_2 - T_2) = \\
 &= 2\pi K \frac{X_1 - X_2}{\ln \frac{r_0 - \frac{1}{2}k\lambda}{r_1 + \frac{1}{2}k\lambda}}.
 \end{aligned}$$

Nach Elimination von X_1 und X_2 erhält man:

$$q' = 2\pi K \frac{(T_1 - T_2)}{\ln \frac{r_0 - \frac{1}{2}k\lambda}{r_1 + \frac{1}{2}k\lambda} + \frac{1}{a_1} \cdot \frac{k\lambda}{r_1} + \frac{1}{a_0} \cdot \frac{k\lambda}{r_0}}. \quad (3)$$

Für $\frac{1}{2}k \frac{\lambda}{r_1} \rightarrow 0$ gewinnen wir hieraus die bekannte SCHLEIERMACHER'sche Formel mit der Korrektur für den Temperatursprung an den beiden Oberflächen:

$$q' = 2\pi K \cdot \frac{(T_1 - T_2)}{\ln \frac{r_0}{r_1} + k \frac{2 - a_1}{2 a_1} \cdot \frac{\lambda}{r_1} + k \frac{2 - a_0}{2 a_0} \cdot \frac{\lambda}{r_0}}, \quad (3a)$$

die bei den Bestimmungen der Wärmeleitfähigkeit der Gase sehr oft³⁾ angewandt wird.

Für $\frac{1}{2}k \frac{\lambda}{r_0} \rightarrow 0$ kann die Formel (3) auch wie folgt geschrieben werden:

$$q' = 2\pi K \cdot \frac{T_1 - T_2}{\ln \frac{r_0 + \frac{1}{2}k\lambda}{r_1 + \frac{1}{2}k\lambda} + k \frac{\lambda}{r_1} \left(\frac{1}{a_1} + \frac{1 - a_0}{a_0} \frac{r_1}{r_0} \right)}. \quad (4)$$

Diese Formel (4) ist für $\frac{\lambda}{r_1} \rightarrow 0$ praktisch identisch mit der Formel (3); für $\frac{r_0}{\lambda} \rightarrow 0$ gibt die Formel (4), weil $K = k \varepsilon p \lambda$, den Grenzwert:

³⁾ SOPHUS WEBER: Ann. d. Phys. 54, 1917, S. 325.

— — — — 54, 1917, S. 437.

$$q'_0 = \frac{2 \pi k \varepsilon p \lambda (T_1 - T_2)}{k \frac{\lambda}{r_1} \left\{ \frac{1}{a_1} + \frac{1 - a_0}{a_0} \cdot \frac{r_1}{r_0} \right\}} = 2 \pi r_1 \varepsilon p \frac{T_1 - T_2}{\frac{1}{a_1} + \frac{1 - a_0}{a_0} \cdot \frac{r_1}{r_0}}. \quad (5)$$

Dieser Ausdruck für q'_0 ist, wie später durch direkte Ableitung gezeigt wird, der richtige Grenzwert für die Wärmeabgabe zwischen zwei unendlich langen, coaxialen Zylindern, wenn das Gas im KNUDSEN-Zustand ist.

Es ist also zu erwarten, dass die Formel (4) im gesamten Druckgebiet:

$$0 \leq \frac{r_1}{\lambda} \leq \infty$$

mit grosser Annäherung, jedenfalls für einatomige Gase, gilt; wahrscheinlich ist sie auch eine brauchbare Annäherung für mehratomige Gase.

Aus Formel (4) erhalten wir die Formel:

$$\frac{1}{q_p} = \frac{1}{q_0} + \frac{1}{q_\alpha}, \quad (6)$$

wenn wir

$$q_0 = \frac{q'_0}{2 \pi r_1 (T_1 - T_2)} = \frac{a_1 \varepsilon p}{1 + a_1 \frac{1 - a_0}{a_0} \cdot \frac{r_1}{r_0}}$$

nennen, und

$$q_\alpha = \frac{2 \pi K}{2 \pi r_1 (T_1 - T_2)} \cdot \frac{T_1 - T_2}{\ln \frac{r_0 + \frac{1}{2} k \lambda}{r_1 + \frac{1}{2} k \lambda}} = \frac{K}{r_1} \cdot \frac{1}{\ln \frac{r_0 + \frac{1}{2} k \lambda}{r_1 + \frac{1}{2} k \lambda}}$$

q_p ist also die Wärmeabgabe per cm^2 und Sek. bei einer Temperaturdifferenz von einem Grad und bei einem Gasdruck p . Aus Formel (6) geht hervor, dass die Temperaturabhängigkeit von q_p im gesamten Druckgebiet ziemlich kompliziert wird, weil:

$$q_0 \propto \frac{1}{\sqrt{T}} \quad \text{und} \quad q_c \propto T^{\frac{1}{2}+n},$$

wo n durch die Temperaturvariation der inneren Reibung, η , des Gases bestimmt ist; wie bekannt, ist diese gegeben durch:

$$\eta = \eta_0 \left(\frac{T}{273.1} \right)^{\frac{1}{2}+n}.$$

Wir sehen hieraus, dass q_p bei einem bestimmten Druck von der mittleren Temperatur T des Gases unabhängig wird, weil die Zunahme von $\frac{1}{q_0}$ dann durch die Abnahme von $\frac{1}{q_c}$ kompensiert wird; dies erhellt auch aus den Messreihen von MARTIN KNUDSEN⁴⁾ und R. E. H. RASMUSSEN⁵⁾.

4. Die Richtigkeit von Formel (5) für $\frac{r_0}{\lambda}$ bzw. $\frac{r_1}{\lambda} = 0$ als Grenzwert für den KNUDSEN-Zustand können wir durch direkte Ableitung dieser Formel nachweisen.

Zu diesem Zweck (Fig. 3) betrachten wir die molekulare

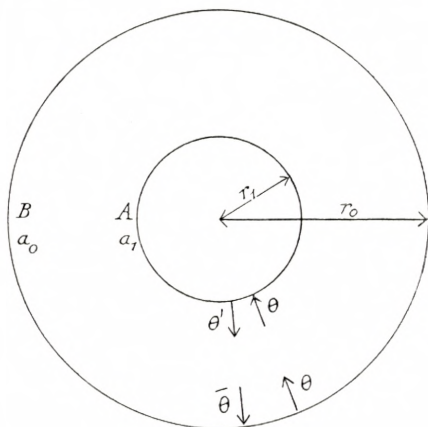


Fig. 3.

Wärmeleitung zwischen zwei coaxialen Zylindern A und B mit den Temperaturen T_1 und T_2 , bzw. den Akkommodationskoeffizienten a_1 und a_0 ; die Radien sind r_1 und r_0 , während die Länge, L , der Zylinder als unendlich angenommen wird. Wir nennen die mittlere Temperatur der

⁴⁾ MARTIN KNUDSEN: D. Kgl. Danske Vidensk. Selskab, Math.-fys. Medd. XI, 1, 1930, S. 59.

⁵⁾ R. E. H. RASMUSSEN: Persönliche Mitteilung.

Moleküle, die von der Oberfläche A kommen, θ' , und die mittlere Temperatur der Moleküle, die von der Oberfläche B kommen, θ ; alle Moleküle, die gegen A stossen, kommen von B , und wir haben demnach:

$$\theta' = \theta + a_1(T_1 - \theta). \quad (7)$$

Die Moleküle, welche gegen B stossen, kommen teilweise von A und teilweise von anderen Teilen der Oberfläche B ; wir nennen die mittlere Temperatur der auf B einfallenden Moleküle $\bar{\theta}$ und erhalten:

$$\theta = \bar{\theta} + a_0(\bar{\theta} - T_2). \quad (8)$$

Im stationären Zustand muss die Stosszahl per cm^2 , $\nu = \frac{1}{4}N\Omega$, im Raum überall dieselbe sein; die totale Anzahl von Molekülen, die gegen A stossen, ist $2\pi r_1 L\nu$, und gegen B , $2\pi r_0 L\nu$, wovon $2\pi r_1 L\nu$ von A kommen, während also $2\pi(r_0 - r_1)L\nu$ von anderen Teilen von B herrühren; die mittlere Temperatur $\bar{\theta}$ wird demnach bestimmt durch:

$$2\pi r_0 L \cdot \nu \bar{\theta} = 2\pi r_1 L \cdot \nu \theta' + 2\pi(r_0 - r_1)L \cdot \nu \theta$$

oder:

$$\bar{\theta} = \frac{r_1}{r_0} \cdot \theta' + \left(1 - \frac{r_1}{r_0}\right) \cdot \theta. \quad (9)$$

Aus den Gleichungen (7), (8), und (9) erhält man durch einfache Umstellung:

$$\theta = T_1 - \frac{1}{1 + a_1 \frac{1 - a_0}{a_0} \cdot \frac{r_1}{r_0}} (T_1 - T_2) \quad (10)$$

$$\theta' = T_1 - \frac{1 - a_1}{1 + a_1 \frac{1 - a_0}{a_0} \cdot \frac{r_1}{r_0}} (T_1 - T_2), \quad (11)$$

$$\theta' - \theta = \frac{a_1}{1 + a_1 \frac{1 - a_0}{a_0} \cdot \frac{r_1}{r_0}} (T_1 - T_2) \quad (12)$$

und

$$\theta = \theta' + \frac{a_0}{a_0 + (1 - a_0) \cdot \frac{r_1}{r_0}} (T_2 - \theta'). \quad (13)$$

Wird der innere Zylinder durch ein flaches Band der Breite B ersetzt, sehen wir, dass diese Formeln alle gelten, wenn nur $\frac{r_1}{r_0}$ durch $\frac{B}{\pi r_0}$ ersetzt wird.

Aus Formel (13) erhellt, dass durch die wiederholten Zurückwerfungen die Oberfläche B gegenüber den Molekülen, die von A kommen, so wirkt, als ob der Akkommodationskoeffizient von B :

$$a = \frac{a_0}{a_0 + (1 - a_0) \cdot \frac{r_1}{r_0}} = \frac{1}{1 + \frac{1 - a_0}{a_0} \cdot \frac{r_1}{r_0}}$$

wäre.

Für $\frac{r_1}{r_0} \rightleftharpoons 0$ wirkt also die Oberfläche B als eine »absolut schwarze« oder »absolut raue« Oberfläche, d. h. der Akkommodationskoeffizient der Oberfläche ist: $a = 1$.

Die Wärmeabgabe von A per cm Länge wird:

$$q'_0 = 2\pi r_1 \cdot \varepsilon p \cdot (\theta' - \theta)$$

oder:

$$q'_0 = 2\pi r_1 \cdot \varepsilon p \cdot \frac{(T_1 - T_2)}{\frac{1}{a_1} + \frac{1 - a_0}{a_0} \cdot \frac{r_1}{r_0}}. \quad (14)$$

Für $a_1 = a_0$ ist diese Formel in Übereinstimmung mit den von MARTIN KNUDSEN und SMOLUCHOWSKI⁶⁾ abgeleiteten Formeln; dasselbe gilt für $\frac{r_1}{r_0} \rightleftharpoons 0$.

⁶⁾ J. H. A. TER HEERDT: Dissertation, Utrecht 1923, S. 146.

Für konzentrische Kugelschalen mit den Radien r_0 und r_1 erhält man auf dieselbe Weise:

$$q'_0 = 4\pi r_1^2 \cdot \varepsilon p \cdot \frac{T_1 - T_2}{\frac{1}{a_1} + \frac{1 - a_0}{a_0} \cdot \left(\frac{r_1}{r_0}\right)^2}, \quad (15)$$

eine Formel, die für die bekannte Abkühlungsmethode zur Bestimmung der Wärmeleitfähigkeit in Gasen von Bedeutung sein kann.

Die Formeln (14) und (15) sind für alle Gase gültig; nur ist der Wert von ε , das molekulare Wärmeleitvermögen der verschiedenen Gase, verschieden.

Betrachten wir ein flaches Band der Breite B und mit verschiedenen Akkommodationskoeffizienten, a_2 und a_1 , der beiden Seiten des Bandes, erhalten wir annäherungsweise, wenn $\frac{B}{2r_0} \rightarrow 1$, für die Wärmeabgabe per cm^2 , Sek. und per Grad Temperaturdifferenz:

$$q = \frac{q'}{2B(T_1 - T_2)} = \frac{1}{2} \varepsilon p \left\{ \frac{a_2}{1 + a_2 \frac{1 - a_0}{a_0} \cdot \frac{B}{\pi r_0}} + \frac{a_1}{1 + a_1 \frac{1 - a_0}{a_0} \cdot \frac{B}{\pi r_0}} \right\}. \quad (16)$$

In dieser Formel ist der Wert von ε durch die Temperatur T an der Stelle, wo der Druck p gemessen wird, bestimmt. Da die Akkommodationskoeffizienten für Wärmeleitung und Radiometerkraft laut den Untersuchungen von **MARTIN KNUDSEN**⁷⁾ dieselben sind, findet man annäherungsweise, wenn $\frac{B}{2r_0} \rightarrow 1$, für die KNUDSEN'sche Radiometerkraft, $R_1 = \frac{R}{B(T_1 - T_2)}$, wenn $\frac{T_1 - T_2}{T}$ eine kleine Grösse ist:

⁷⁾ MARTIN KNUDSEN: loc. cit. (4), S. 8.

$$R_1 = \frac{R}{B(T_1 - T_2)} = \left. \begin{aligned} &= \frac{1}{4} \cdot \frac{p}{T} \left\{ \frac{a_2}{1 + a_2 \frac{1 - a_0}{a_0} \cdot \frac{B}{\pi r_0}} - \frac{a_1}{1 + a_1 \frac{1 - a_0}{a_0} \cdot \frac{B}{\pi r_0}} \right\} \end{aligned} \right\} (17)$$

Auch in diesem Falle sind p und T korrespondierende Grössen, während R die Radiometerkraft per cm Länge des Bandes bezeichnet.

Wir werden später im Zusammenhang mit den experimentellen Untersuchungen von R. E. H. RASMUSSEN⁸⁾ auf diese Formeln zurückkommen.

5. Wir können nun die Formeln (3 a) und (6) mit dem vorliegenden Versuchsmaterial für einatomige Gase vergleichen; die Wärmeleitfähigkeit von sehr reinem Neon wurde früher von mir in einem grossen Druckgebiet⁹⁾, bzw. Gebiet von $\frac{r_1}{\lambda}$, untersucht, u. a. mit einem Doppelapparat nach der Methode von GOLDSCHMIDT¹⁰⁾. Der Durchmesser des verwendeten Platindrahtes war $2r_1 = 0,005240$ cm; der innere Durchmesser des umgebenden Glasrohres war $2r_0 = 1,526$ cm; den gefundenen Wärmeverlust per cm Länge des Drahtes und per Grad Temperaturunterschied, korrigiert für die Ableitung durch die Enden des Drahtes und für die Strahlung, nennen wir $D_{\text{obs.}}$; diese Grösse ist in Tabelle I, Kolonne 4, angegeben. Aufgrund der Untersuchungen von KNUDSEN hat der Akkommodationskoeffizient einer Glasoberfläche und der einer glatten, blanken Platinoberfläche denselben Wert; für Neon wird in diesem Falle, da es sich um glatte Oberflächen handelt, $a = a_1 = a_2 = 0,653$

⁸⁾ R. E. H. RASMUSSEN: D. Kgl. Danske Vidensk. Selskab, Math.-fys. Medd. XI, 9, 1932; XIII, 9, 1935, S. 16.

⁹⁾ SOPHUS WEBER: Commun. Leiden. Suppl. 42 b, 1918, S. 34.

¹⁰⁾ R. GOLDSCHMIDT: Phys. Zeitschrift 12, 1911, S. 418.

gesetzt. Da Neon ein einatomiges Gas ist, wird $k = 15/4 = 3,75$. Nach den beiden Formeln (3 a) und (6) erhalten wir für $T_1 - T_2 = 1$, wenn:

$$D_\infty = \frac{2\pi K}{\ln \frac{r_0}{r_1}},$$

nach Formel (3 a):

$$D'_\infty = D_{\text{obs.}} \left[1 + \frac{k\lambda}{\ln \frac{r_0}{r_1}} \cdot \frac{2-a}{2a} \left(\frac{1}{r_1} + \frac{1}{r_0} \right) \right],$$

und nach Formel (6):

$$D''_\infty = D_{\text{obs.}} \left[\frac{\ln \frac{r_0 + \frac{1}{2}k\lambda}{r_1 + \frac{1}{2}k\lambda}}{\ln \frac{r_0}{r_1}} + \frac{k\lambda}{\ln \frac{r_0}{r_1}} \left(\frac{1}{ar_1} + \frac{1-a}{a} \cdot \frac{1}{r_0} \right) \right].$$

Nach der richtigen Formel muss also D_∞ konstant werden. In der folgenden Tabelle sehen wir die Resultate für Neon; die Temperatur des Heizdrahtes war $9^\circ,20\text{ C}$ und die Temperatur des umgebenden Bades 0° C .

Tabelle I.
Neon bei 0° C .

	p cm Hg	$\frac{r_1}{\lambda}$	$\Delta T = T_1 - T_2$	$10^6 \cdot D_{\text{obs.}}$	$10^6 \cdot D'_\infty$ Form. (3 a)	$10^6 \cdot D''_\infty$ Form. (6)
1)	33,791	92,5	$9^\circ,20\text{ C}$	816,2	822,2	822,3
2)	20,182	55,1	—	811,8	821,9	822,1
3)	10,181	27,8	—	801,8	821,5	822,0
4)	5,729	15,7	—	787,2	821,5	822,6
5)	1,894	5,2	—	720,6	816,0	822,9
6)	0,709	1,9	—	584,0	790,0	821,0
7)	0,215	0,6	—	329,2	707,8	817,0

Es erhellt aus dieser Tabelle, dass Formel (6) einen weit besseren Anschluss gibt als Formel (3 a), die nur für $\frac{r_1}{\lambda} \rightarrow \infty$ gültig ist; während die Formel (3 a) bei $\frac{r_1}{\lambda} = 0,6$ eine Abweichung von etwa 15 % gibt, ist die Abweichung für Formel (6) nur etwa 0,5 %; diese letztere Abweichung ist in Übereinstimmung mit der Messgenauigkeit des Druckes p .

Aus einer anderen Messreihe der gleichen Untersuchung mit demselben Apparat bei einer Badtemperatur von $99^{\circ},81\text{ C}$ (vgl. loc. cit. (9) S. 35) geht auch hervor, dass Formel (6) mit den Messungen innerhalb der Messgenauigkeit vollständig übereinstimmt.

6. Einen sehr einfachen Fall von Wärmeleitung zwischen coaxialen Zylindern haben wir in dem Hitzdrahtmanometer von MARTIN KNUDSEN¹¹⁾, in welchem der geheizte Draht ein Wollastondraht ist; wir können diesen Fall durch $\frac{r_1}{r_0} \cong 0$ charakterisieren, und erhalten dann:

$$q = 2\pi r_1 \cdot a_1 \varepsilon p (T_1 - T_2) \cdot \frac{1}{1 + \frac{a_1}{k} \cdot \frac{r_1}{\lambda} \ln \frac{r_0 + \frac{1}{2} k\lambda}{r_1 + \frac{1}{2} k\lambda}}, \quad (18)$$

oder für den Fall, dass $\frac{r_1}{\lambda} \rightarrow 0$ und $\frac{\lambda}{r_0} \rightarrow 0$,

$$q = 2\pi r_1 \cdot a_1 \varepsilon p (T_1 - T_2) \left\{ 1 - \frac{a_1}{k} \cdot \frac{r_1}{\lambda} \ln \frac{r_0}{r_1 + \frac{1}{2} k\lambda} \right\},$$

woraus:

$$\frac{dq}{dp} = 2\pi r_1 \cdot a_1 \varepsilon (T_1 - T_2) \left\{ 1 - 2 \frac{a_1}{k} \cdot \frac{r_1}{\lambda} \left(\ln \frac{r_0}{r_1 + \frac{1}{2} k\lambda} + \frac{1}{2} \right) \right\}.$$

¹¹⁾ MARTIN KNUDSEN: Ann. d. Phys. 35, 1911, S. 389.

MARTIN KNUDSEN¹²⁾ hat die Wärmeabgabe eines Hitzdrahtmanometers in Wasserstoff mit einem Wollastondraht $r_1 = 0,000225$ cm und $r_0 = 0,535$ cm gemessen in dem ganzen Druckgebiet $0 \leq p \leq 1$ atm.; in der folgenden Tabelle II sieht man die beobachteten und die nach Formel (18) berechneten Werte für das ganze Gebiet, in welchem $\frac{r_1}{\lambda}$ sich von 0,002 auf 20 ändert.

Tabelle II.

Wasserstoff: $(p\lambda)_0 = 11,31$, $a_1\varepsilon = 3,80 \cdot 10^{-6}$, $k = 3,34$.

p (Bar)	$\frac{q}{2\pi r_1} \cdot \frac{1}{T_1 - T_2} \cdot 10^6$	$\frac{q}{2\pi r_1} \cdot \frac{1}{T_1 - T_2} \cdot \frac{10^6}{p}$ obs.	$\frac{q}{2\pi r_1} \cdot \frac{1}{T_1 - T_2} \cdot \frac{10^6}{p}$ ber.	$\frac{r_1}{\lambda}$
107,3	407,7	3,800	3,799	$2,14 \cdot 10^{-3}$
214,8	815,6	3,797	3,797	4,30 —
322,4	1220,5	3,786	3,795	6,40 —
428,6	1621,3	3,783	3,792	$8,50 \cdot 10^{-3}$
534,8	2015,2	3,768	3,789	$1,06 \cdot 10^{-2}$
640,0	2413,5	3,771	3,786	1,27 —
744,6	2804,8	3,767	3,783	1,48 —
848,6	3192,3	3,762	3,779	1,69 —
950,9	3576,0	3,761	3,775	1,89 —
1053,0	3954,2	3,754	3,771	2,10 —
3200,0	11651	3,641	3,689	6,36 —
6355,0	22046	3,48	3,56	$1,26 \cdot 10^{-1}$
13450,0	42117	3,13	3,26	2,68 —
35000,0	83900	2,40	2,55	$6,97 \cdot 10^{-1}$
66800,0	121900	1,83	1,93	1,33
129000,0	170900	1,325	1,290	2,57
505000,0	211100	0,418	0,421	10,2
1016000,0 (1 atm.)	223900	0,220	0,220	20,2

Eine Abnahme von $k = 3,34$ wird wohl nur bei den allerniedrigsten Drucken in Frage kommen, wodurch jedoch die Übereinstimmung zwischen den beobachteten und den berechneten Werten in diesem Gebiet nur besser werden kann.

12) MARTIN KNUDSEN: ANN. D. PHYS. 34, 1911, S. 636.

Wir sehen hieraus, dass die Übereinstimmung zwischen den berechneten und den observierten Werten befriedigend ist, wenn die Unsicherheit der observierten Werte bei höherem Druck in Betracht gezogen wird; es wäre für die Verwendung des Hitzdrahtmanometers von Interesse, das gesamte Druckgebiet mit verschiedenen Gasen durchzumessen und mit Formel (18) zu vergleichen.

7. MARTIN KNUDSEN¹³⁾ hat in seiner bekannten Untersuchung über Radiometerkraft und Akkommodationskoeffizient u. a. auch die Wärmeabgabe von drei Platinbändern gemessen, die in einem Glaszylinder mit einem Durchmesser von $2r_0 = 6,40$ cm vertikal ausgespannt waren. Band I war auf beiden Seiten blank, Band II auf beiden Seiten schwarz, während das dritte Band auf einer Seite blank und auf der anderen Seite schwarz war. Band III war in der Mittelachse ausgespannt, während Band I und II etwas exzentrisch und gegenüber Band III symmetrisch montiert waren; die Abmessungen der drei Bänder waren:

	Länge	Breite	Dicke
Band III blank/schwarz . .	15,31 cm	0,2484 cm	ca. 0,00026 cm
— II schwarz/schwarz	15,20	- 0,2444	- -
— I blank/blank	15,40	- 0,2495	- -

Für diese drei Bänder hat MARTIN KNUDSEN in einem grossen Druckgebiet bei verschiedenen Temperaturdifferenzen ($T_1 - T_2$) die Wärmeabgabe in Wasserstoff und Helium gemessen; T_2 war ca. 20° C. Wir schreiben nach KNUDSEN die Wärmeabgabe per cm^2 , per Grad Temperaturdifferenz und per Sek.:

¹³⁾ MARTIN KNUDSEN: loc. cit. (4), S. 59.

$$Q' = \frac{1}{100} \cdot \frac{q'}{T_1 - T_2} \text{ Erg/cm}^2, \text{ Sek.}, \text{ Grad} \quad \text{für Band I}$$

$$Q'' = \frac{1}{100} \cdot \frac{q''}{T_1 - T_2} \text{ Erg/cm}^2, \text{ Sek.}, \text{ Grad} \quad \text{für Band II}$$

$$Q''' = \frac{1}{100} \cdot \frac{q'''}{T_1 - T_2} \text{ Erg/cm}^2, \text{ Sek.}, \text{ Grad} \quad \left\{ \begin{array}{l} \text{für beide Seiten von} \\ \text{Band III} \end{array} \right.$$

q' , q'' und q''' sind die gemessenen Wärmeabgaben per cm^2 und Sek.; $T_1 - T_2$ ist die gemessene Temperaturdifferenz.

Aus den Messungen geht hervor, dass bei den niedrigsten Drucken $Q''' = Q' + Q''$, wie auch zu erwarten war, da bei den niedrigsten Drucken die Wärmeabgabe nicht davon beeinflusst wird, dass die Bänder I und II ausserhalb der Achse des Glaszylinders stehen; bei höheren Drucken ist $Q' + Q''$ grösser als Q''' .

Wir können in diesem Falle $Q' + Q'' = (1 + \sigma) Q'''$ setzen, wo $(1 + \sigma)$ einen Formfaktor darstellt, der davon herrührt, dass Band I und II ausserhalb der Achse stehen; man findet bei den höchsten Drucken im Mittel für Wasserstoff und Helium $1 + \sigma = 1 + 0,090$.

Wie auch aus der Radiometerkraftmessung hervorgeht, muss durch die verschiedene Oberflächenbeschaffenheit der beiden Seiten des Bandes III um das Band herum eine horizontale Gasströmung entstehen; aus dem experimentellen Material geht aber hervor, dass diese Strömung die Wärmeabgabe des Bandes nur wenig beeinflusst, da die Wärmeabgabe, wie von KNUDSEN angegeben, bei hohem Druck mit der gewöhnlichen Wärmeleitfähigkeit des ruhenden Gases übereinstimmt; andererseits beeinflusst diese Strömung in hohem Masse die Radiometerkraft, die durch diese Strömung um das Band herum teilweise ausgeglichen wird.

Tabelle III.

Wasserstoff:

$$T_1 - T_2 = 75^\circ \text{C}, T_2 = 20^\circ \text{C}, T = \frac{T_1 + T_2}{2} = 57,5^\circ \text{C}.$$

p (Bar)	$\log_{10} p$	Q''	Q'	$Q'' - Q'$
30,7	1,487	75,6	33,1	42,5
55,3	1,743	126,0	56,7	69,3
99,7	1,999	199,6	93,0	106,6
179,7	2,255	299,1	152,0	147,1
324,0	2,511	407,1	230,0	177,1
584,0	2,766	498,5	321,5	177,0
1290,0	3,111	598,0	464,0	134,0
3670,0	3,565	648,0	573,0	75,0
7190,0	3,857	666,0	619,0	47,0
14390,0	4,158	673,6	644,0	29,0
3,04	0,478	9,97	4,07	5,90
10,80	1,034	32,18	14,36	17,82
15,93	1,202	46,60	19,43	27,17
20,29	1,307	58,94	25,06	33,90

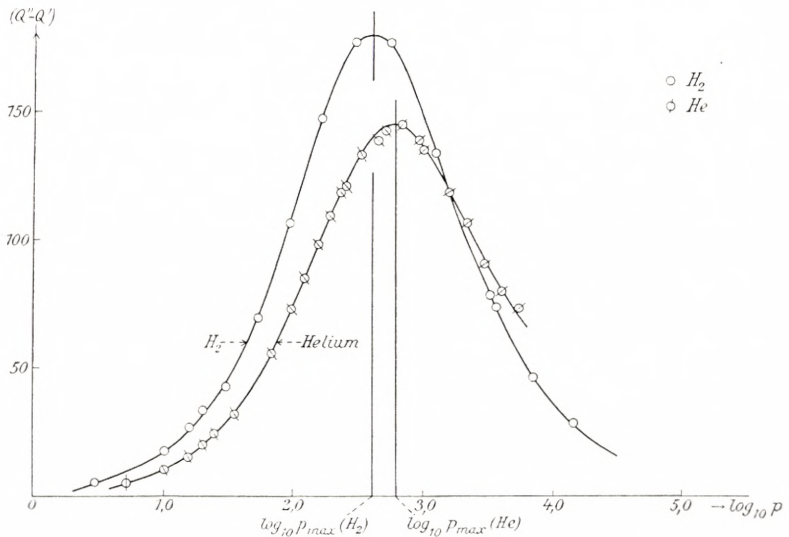


Fig. 4.

Tabelle IV.

Helium:

$$T_1 - T_2 = 82^\circ \text{ C}, \quad T_2 = 20^\circ \text{ C}, \quad T = \frac{T_1 + T_2}{2} = 61^\circ,0 \text{ C}.$$

p (Bar)	$\log_{10} p$	Q''	Q'	$Q'' - Q'$
5,205	0,717	9,72	4,36	5,36
10,34	1,014	18,74	8,20	10,54
15,40	1,188	27,37	11,94	15,43
20,30	1,308	35,55	15,49	20,06
25,10	1,400	43,10	18,70	24,40
35,7	1,553	57,3	25,1	32,2
68,9	1,838	102,7	46,5	56,2
99,7	1,999	139,0	65,5	73,5
128,2	2,108	166,6	81,4	85,2
163,5	2,214	198,2	100,0	98,2
204,1	2,310	228,4	119,4	109,0
271,0	2,433	268,8	148,0	120,8
245,0	2,389	257,0	138,7	118,3
357,0	2,553	315,5	182,0	133,5
463,0	2,666	356,3	217,0	139,3
530,0	2,724	374,0	231,4	142,6
712,0	2,853	420,0	274,7	145,3
956,0	2,981	457,4	317,8	139,6
1283	3,108	495,8	360,7	135,1
1711	3,233	526,4	406,8	119,6
2283	3,359	552,0	445,0	107,0
3064	3,486	571,0	480,0	91,0
4113	3,614	586,4	506,0	80,4
5521	3,742	602,0	528,0	74,0

In den obenstehenden Tabellen III und IV sind die mittleren beobachteten Werte von Q'' und Q' in Wasserstoff und Helium angegeben; die Mittelwerte von Wasserstoff stimmen mit den Temperaturen $T_1 = 95^\circ \text{ C}$ und $T_2 = 20^\circ \text{ C}$ überein; für Helium ist $T_1 = 102^\circ \text{ C}$ und $T_2 = 20^\circ \text{ C}$.

Fig. 4 ist eine graphische Darstellung der Abhängigkeit ($Q'' - Q'$) und $\log_{10} p$ für Wasserstoff und Helium; es zeigt

sich, dass man eine symmetrische Kurve erhält; das Maximum von $(Q'' - Q')$ liegt für Wasserstoff bei $\log_{10} p_{\max} = 2,625$ und für Helium bei $\log_{10} p_{\max} = 2,790$. Wir erhalten hiernach:

für Wasserstoff:

$$p_{\max} = 420 \text{ Bar}$$

$$\text{und } \left(\frac{q'' - q'}{T_1 - T_2} \right)_{\max} = 180 \cdot 10^2 \text{ Erg/cm}^2, \text{ Sek., Grad,}$$

für Helium:

$$p_{\max} = 615 \text{ Bar}$$

$$\text{und } \left(\frac{q'' - q'}{T_1 - T_2} \right)_{\max} = 145 \cdot 10^2 \text{ Erg/cm}^2, \text{ Sek., Grad.}$$

Wir können nunmehr den Verlauf der Kurven und die Maximumwerte für p und $\frac{q'' - q'}{T_1 - T_2}$ nach der Theorie bzw. Formel (6) berechnen. Wir haben, wenn wir $10^2 \cdot Q'' = {}_2q_p$ und $10^2 \cdot Q' = {}_1q_p$ nennen:

$$\frac{1}{{}_1q_p} = \frac{1}{{}_1q_0} + \frac{1}{{}_1q_\alpha} \quad \text{und} \quad \frac{1}{{}_2q_p} = \frac{1}{{}_2q_0} + \frac{1}{{}_2q_\alpha}.$$

Da Band II und Band I symmetrisch zur Achse stehen, ist ${}_1q_\alpha = {}_2q_\alpha$, wenn der Druck p derselbe ist. In diesen Formeln können wir, weil $\frac{r_1}{r_0}$ klein gegenüber 1 ist, setzen:

$${}_1q_0 = a_1 \varepsilon p \quad \text{und} \quad {}_2q_0 = a_2 \varepsilon p,$$

während:

$${}_1q_\alpha = {}_2q_\alpha = \frac{2\pi K}{2\pi r_1} \cdot \frac{1}{r_0 + \frac{1}{2}k\lambda} \cdot \ln \frac{r_1 + \frac{1}{2}k\lambda}{r_0 + \frac{1}{2}k\lambda}.$$

Hier ist $\pi r_1 = b$, wo b die Breite des Bandes und K die Wärmeleitfähigkeit des Gases ist; nach dem vorhergehenden ist $K = k \varepsilon p \lambda$.

Wir erhalten hieraus:

$${}_2q_p - {}_1q_p = \frac{{}_2q_0 - {}_1q_0}{1 + \frac{{}_2q_0 + {}_1q_0}{q_\alpha} + \frac{{}_2q_0 \cdot {}_1q_0}{q_\alpha^2}} \quad (19)$$

oder:

$${}_2q_p - {}_1q_p = \frac{(a_2 - a_1) \varepsilon p}{1 + (a_2 + a_1) \left(\frac{\varepsilon}{q_\alpha}\right) \cdot p + a_1 a_2 \left(\frac{\varepsilon}{q_\alpha}\right)^2 \cdot p^2}$$

oder:

$${}_2q_p - {}_1q_p = \frac{(a_2 - a_1) q_\alpha}{\frac{1}{\left(\frac{\varepsilon}{q_\alpha}\right) \cdot p} + (a_2 + a_1) + a_1 a_2 \left(\frac{\varepsilon}{q_\alpha}\right) \cdot p}. \quad (20)$$

Hieraus folgt:

$$P_{\max} = \frac{q_\alpha}{\varepsilon} \cdot \frac{1}{\sqrt{a_1 a_2}} \quad (21)$$

und:

$$({}_2q_p - {}_1q_p)_{\max} = \frac{a_2 - a_1}{(\sqrt{a_2} + \sqrt{a_1})^2} \cdot q_\alpha, \quad (22)$$

woraus:

$$f' = \frac{({}_2q_p - {}_1q_p)_{\max}}{P_{\max}} = \frac{a_2 - a_1}{(\sqrt{a_2} + \sqrt{a_1})^2} \sqrt{a_1 a_2} \cdot \varepsilon. \quad (23)$$

Verwenden wir P_{\max} und $({}_2q_p - {}_1q_p)_{\max}$ als Einheiten für p und $({}_2q_p - {}_1q_p)$, erhalten wir:

$$P = \frac{{}_2q_p - {}_1q_p}{({}_2q_p - {}_1q_p)_{\max}} = \frac{(\sqrt{a_2} + \sqrt{a_1})^2}{\sqrt{a_1 a_2}} \cdot \frac{1}{\frac{P}{P_{\max}} + \frac{P_{\max}}{P} + \frac{a_1 + a_2}{\sqrt{a_1 a_2}}}.$$

Setzen wir $x = \ln \frac{P}{P_{\max}}$ und $\delta = \frac{a_1 + a_2}{\sqrt{a_1 a_2}}$, finden wir:

$$P = \frac{{}_2q_p - {}_1q_p}{({}_2q_p - {}_1q_p)_{\max}} = \frac{2 + \delta}{e^x + e^{-x} + \delta},$$

woraus erhellt, dass P in bezug auf sein Maximum bei $x = 0$ symmetrisch ist, da der Wert für $x = m$ und

$x = -m$ derselbe wird; dies ist in Übereinstimmung mit den experimentellen Kurven.

Wir können hiernach die theoretischen Werte von p_{\max} und $({}_2q_p - {}_1q_p)_{\max}$ mit den beobachteten Werten vergleichen.

A. Wasserstoff:

$$p_{\max, \text{obs.}} = 420 \text{ Bar};$$

$$({}_2q_p - {}_1q_p)_{\max, \text{obs.}} = 18000 \text{ Erg/cm}^2, \text{ Sek., Grad}$$

und also

$$f'_{\text{obs.}} = \frac{18000}{420} = 42,85.$$

Aufgrund der KNUDSEN'schen Werte:

$$a_2 = 0,735 \quad \text{und} \quad a_1 = 0,315$$

und

$$\varepsilon_T = \frac{7555}{\sqrt{273 + T}} = 415,5$$

erhalten wir:

$$f'_{\text{theor.}} = \frac{a_2 - a_1}{(\sqrt{a_2} + \sqrt{a_1})^2} \sqrt{a_1 a_2} \cdot \varepsilon_T = 41,75,$$

also eine befriedigende Übereinstimmung.

Für

$$q'_\alpha = \frac{\pi \cdot K_0 \left(\frac{T}{T_0}\right)^{0,70}}{r_0 + \frac{1}{2} k \lambda_{\max}} \cdot \frac{b \ln}{\frac{b}{\pi} + \frac{1}{2} k \lambda_{\max}}$$

finden wir den Wert:

$$q'_\alpha = 81140, \text{ da } K_0 = 4170 \cdot 10^{-7} \cdot 4,19 \cdot 10^7,$$

$$\frac{T}{T_0} = \frac{T_1 + T_2}{2 T_0} = 1,210, \quad \frac{1}{2} k \lambda_{\max} = 0,064, \quad (p\lambda)_T = 14,22$$

und $b = 0,2470 \text{ cm}$. Wo aber Band I und II nicht in der Achse stehen, wird $q_\alpha = q'_\alpha (1 + \sigma) = 81140 \cdot 1,090 = 88440$.

Hieraus bekommt man:

$$p_{\max, \text{ber.}} = \frac{q_\alpha}{\varepsilon} \cdot \frac{1}{\sqrt{a_1 a_2}} = \frac{88440}{415,5} \cdot \frac{1}{0,481} = 442 \text{ Bar}$$

und

$$\begin{aligned} ({}_2q_p - {}_1q_p)_{\max, \text{ber.}} &= \frac{a_2 - a_1}{(\sqrt{a_2} + \sqrt{a_1})^2} \cdot q_\alpha = \frac{0,42}{2,012} \cdot 88440 \\ &= 18460 \text{ Erg/cm}^2, \text{ Sek., Grad.} \end{aligned}$$

Die Übereinstimmung ist, wenn die Genauigkeit der Bestimmungen in Betracht gezogen wird, ziemlich befriedigend.

B. Helium:

Für Helium wurde experimentell gefunden:

$$\begin{aligned} p_{\max, \text{obs.}} &= 615 \text{ Bar.} \quad \text{und} \quad ({}_2q_p - {}_1q_p)_{\max, \text{obs.}} \\ &= 14500 \text{ Erg/cm}^2, \text{ Sek., Grad,} \end{aligned}$$

woraus:

$$f'_{\text{obs.}} = \frac{14500}{615} = 23,57.$$

Aufgrund der KNUDSEN'schen Werte:

$$a_2 = 0,909, \quad a_1 = 0,411 \quad \text{und} \quad \varepsilon_T = \frac{3637,5}{\sqrt{273 + T}} = 199,1$$

erhalten wir:

$$f'_{\text{theor.}} = \frac{a_2 - a_1}{(\sqrt{a_1} + \sqrt{a_2})^2} \sqrt{a_2 a_1} \cdot \varepsilon_T = 23,83.$$

Für Helium wird:

$$q'_\alpha = \frac{\pi \cdot K_0 \left(\frac{T}{T_0}\right)^{0,65}}{b \ln \frac{r_0 + \frac{1}{2} k \lambda_{\max}}{\frac{b}{\pi} + \frac{1}{2} k \lambda_{\max}}} = 67160,$$

$$\begin{aligned} \text{da } K_0 &= 3430 \cdot 10^{-7} \cdot 4,19 \cdot 10^7, \quad \frac{T}{T_0} = \frac{T_1 + T_2}{2 T_0} = 1,223, \quad \frac{1}{2} k \lambda_{\max} \\ &= 0,0683, \quad (p \lambda)_T = 22,36, \quad b = 0,2470 \text{ cm.} \end{aligned}$$

Hieraus erhält man:

$$q_{\alpha} = (1 + \sigma) q'_{\alpha} = 1,090 \cdot 67160 = 73200,$$

woraus:

$$P_{\max, \text{ber.}} = \frac{q_{\alpha}}{\varepsilon} \cdot \frac{1}{\sqrt{a_1 a_2}} = \frac{73200}{199,1} \cdot \frac{1}{0,611} = 601 \text{ Bar}$$

und

$$\begin{aligned} ({}_2q_p - {}_1q_p)_{\max} &= \frac{a_2 - a_1}{(\sqrt{a_1} + \sqrt{a_2})^2} \cdot q_{\alpha} = \frac{0,498}{2,543} \cdot 73200 \\ &= 14300 \text{ Erg/cm}^2, \text{ Sek., Grad.} \end{aligned}$$

Auch für Helium ist die Übereinstimmung zwischen den observierten und den theoretischen Werten nicht unbefriedigend. Es kann hiernach angenommen werden, dass die theoretische Formel (20) als erste Annäherung für das ganze Gebiet $0 \leq \frac{b}{\lambda} \leq \infty$ gelten darf; für einen vollständigen Vergleich über das ganze Gebiet ist das experimentelle Material nicht voll ausreichend.

8. Die Richtigkeit der Formel (14):

$$q_0 = 2\pi r_1 \cdot \varepsilon p \cdot \frac{T_1 - T_2}{\frac{1}{a_1} + \frac{1 - a_0}{a_0} \cdot \frac{r_1}{r_0}},$$

ist für verschiedene Werte von $\frac{r_1}{r_0}$, a_0 und a_1 noch nicht hinreichend experimentell untersucht; nach den KNUDSEN'schen Untersuchungen darf aber die Formel, jedenfalls für $a_0 = a_1$ und für $\frac{r_1}{r_0} \rightarrow 0$ als richtig angenommen werden; für $\frac{r_1}{r_0} \rightarrow 1$ erhält man den Grenzwert der Formel (2) für $\frac{d}{\lambda} = 0$; in diesem Falle wird $r_0 - r_1 = d$.

R. E. H. RASMUSSEN¹⁴⁾ hat eine Reihe experimenteller Unter-

¹⁴⁾ R. E. H. RASMUSSEN: D. Kgl. Danske Vidensk. Selskab, Math.-fys. Medd. XIII, 9, 1935, S. 16.

suchungen über die KNUDSEN'sche Radiometerkraft durchgeführt, welche auf verschiedenen Werten des Akkommodationskoeffizienten der beiden Seiten des Radiometersystems beruht; gleichzeitig mit der Radiometerkraft hat RASMUSSEN auch die Wärmeabgabe des verwendeten Platinbandes gemessen¹⁵⁾; das Radiometersystem bestand aus einem vertikalen Platinband, ca. 1 cm breit, dessen eine Seite mit Platinschwarz rauh gemacht war; das Band war in der Mitte eines vertikalen Messingrohres mit rechtwinkligem Querschnitt ausgespannt (Fig. 5); die inneren Abmessungen des Messingrohres waren $ab = l = 3,2$ cm und $ac = m = 1,48$ cm; die Temperatur T_1 des geheizten Platinbandes wurde durch den elektrischen Widerstand W bestimmt, und die Temperatur T_2 des Messingrohres wurde mit einem Quecksilberthermometer gemessen, dessen Behälter mit WOOD-Metall in leitende Verbindung mit dem Messingrohr gebracht war; das ganze Messsystem war in einem weiten Glasbehälter angebracht. MARTIN KNUDSEN fand bei seinen Messungen in einem sehr grossen Behälter für die blanke Seite eines ähnlichen Bandes in Wasserstoff: $\alpha_1 = 0,315$; leider wurde bei den Versuchen von RASMUSSEN der Akkommodationskoeffizient α_2 für die raue Seite allein nicht gesondert in einem grossen Behälter gemessen; da das Messingrohr aus glattem Metall besteht, darf man annehmen, dass für Wasserstoff $\alpha_0 = 0,315$ gesetzt werden kann.

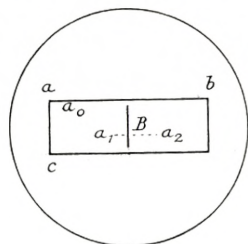


Fig. 5.

Die Abmessungen und Konstanten des Systems waren:

¹⁵⁾ Diese Messungen sind bisher nicht veröffentlicht, doch hat Dr. RASMUSSEN sie mir in lebenswürdiger Weise zur Verfügung gestellt, wofür ich ihm auch hier meinen besten Dank aussprechen möchte.

Breite des Platinbandes	$B = 0,999$ cm
Länge - —	$L = 14,83$ cm
Dicke - —	$d = \text{ca. } 0,0003$ cm
a_1 (blanke Seite) —	$a_1 = 0,315$
a_0 (Messingrohr)	$a_0 = 0,315$
a_2 (schwarze Seite des Platinb.)	nicht bekannt
Innenmasse des Messingrohres	$m = 1,48$ cm
— - —	$l = 3,20$ cm
Widerstand des Platinbandes .	W in Ohm
Heizstrom des Platinbandes . . .	i in Amp.
Wärmeverlust des Bandes per cm ² , Sek. und per Grad Tem- peraturdifferenz	$q = \frac{W \cdot i^2}{T_1 - T_2} \cdot \frac{10^7}{2BL} \text{ Erg/cm}^2, \text{ Sek., Grad}$
Radiometerkraft per cm ² und Grad	

In der folgenden Tabelle V sind die Messungen für den totalen Wärmeverlust des geheizten Bandes q' und die Radiometerkraft R_1 angegeben:

Durch ein rechnerisches oder graphisches Verfahren ist es einfach hieraus die Werte von:

$$\lim \left(\frac{dq'}{dp} \right)_{p=0} \quad \text{und} \quad \lim \left(\frac{dR_1}{dp} \right)_{p=0}$$

zu bestimmen; man erhält:

$$\lim \left(\frac{dq'}{dp} \right)_{p=0} = 167,0 \quad \text{und} \quad \lim \left(\frac{dR_1}{dp} \right)_{p=0} = 2,60 \cdot 10^{-4},$$

da $2BL = 29,66$ cm².

Mit Hilfe der Formeln (16) und (17) können wir aus den observierten Werten von $\left(\frac{dq'}{dp} \right)_{p=0}$ und $\left(\frac{dR_1}{dp} \right)_{p=0}$ den Wert von a_2 bestimmen; aus der Wärmeabgabe finden wir $a_2 = 0,622$, während die Radiometerkraft $a_2 = 0,799$ ergibt, also eine grosse Differenz.

Tabelle V.

Messungen mit Wasserstoff:

$T_2 = 20^\circ \text{C}$, $T_1 - T_2 = \text{ca. } 80^\circ \text{C}$.

p (Bar)	Δp (Bar)	$q' = \frac{W \cdot i^2}{T_1 - T_2} \cdot \text{Watt/Grad}$	$10^4 \cdot R_1$ Dyn/cm ² , Grad
0,00		0,00000	0,0
2,75	2,75	0,00132	8,4
5,48	2,73	0,00262	16,2
8,21	2,73	0,00387	24,6
10,92	2,71	0,00504	29,2
13,62	2,70	0,00620	35,3
16,31	2,69	0,00732	41,1
19,00	2,69	0,00836	46,3
21,70	2,70	0,00933	50,6
0,00		0,00000	0,0
5,25	5,25	0,00242	11,8
10,49	5,24	0,00485	27,5
15,70	5,21	0,00726	39,4
20,90	5,20	0,00939	48,3
26,1	5,20	0,01142	56,6
31,2	5,10	0,01337	62,9
36,3	5,10	0,01524	71,0

Verwendet man einfachheitshalber für a_2 den Wert: $a_2 = 0,735$, den MARTIN KNUDSEN auch durch direkte Messung für die schwarze Seite des von ihm verwendeten und auf gleiche Weise hergestellten Bandes gefunden hat, dann bekommt man mit $B = 1$ und $\pi r_0 = \frac{1}{2}(l + m) = 4,68$ aus der Formel für die Wärmeabgabe:

$$\left(\frac{dq}{dp}\right)_{p=0} = \frac{1}{2} \varepsilon_{T_2} \left\{ \frac{a_2}{1 + a_2 \frac{1 - a_0}{a_0} \cdot \frac{B}{\pi r_0}} + \frac{a_1}{1 + a_1 \frac{1 - a_0}{a_0} \cdot \frac{B}{\pi r_0}} \right\},$$

$\left(\frac{dq}{dp}\right)_{p=0} = 181,5$, also einen etwa 10 % grösseren Wert als den beobachteten.

Für die Radiometerkraft R_1 hat man:

$$\left(\frac{dR_1}{dp}\right)_{p=0} = \frac{1}{4T_2} \left\{ \frac{a_2}{1+a_2} \frac{1-a_0}{a_0} \cdot \frac{B}{\pi r_0} - \frac{a_1}{1+a_1} \frac{1-a_0}{a_0} \cdot \frac{B}{\pi r_0} \right\},$$

woraus man mit denselben Werten von a_2 , a_1 , a_0 , B und πr_0 :

$$\left(\frac{dR_1}{dp}\right)_{p=0} = 2,33 \cdot 10^{-4}$$

bekommt, d. h. einen etwa 10 % kleineren Wert als den beobachteten. Diese Abweichungen liegen innerhalb der Genauigkeit, womit diese schwierigen Messungen ausgeführt werden konnten; wir dürfen also annehmen, dass der Wert $a_2 = 0,735$ dem richtigen Wert von a_2 nahe kommt.

Es erhellt hieraus, obwohl der Wert von a_2 sehr unsicher ist und die Formeln (16) und (17) bei den RASMUSSEN'schen Versuchsbedingungen nur Annäherungsformeln sind, dass es im KNUDSEN-Zustand des Gases notwendig ist, auch den Wert von a_0 zu berücksichtigen.

9. Um den Zusammenhang zwischen der Wärmeabgabe und der KNUDSEN'schen Radiometerkraft klarzulegen, auch

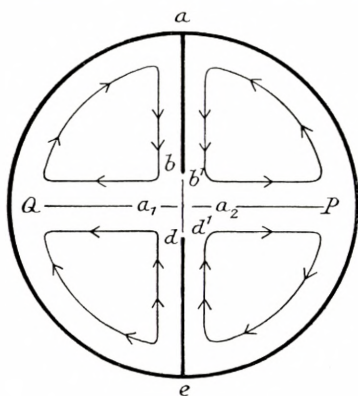


Fig. 6.

dann, wenn der molekulare Zustand nicht mehr besteht, werden wir erst die Aufstellung Fig. 6 betrachten; diese Aufstellung ist im Prinzip dieselbe wie die von RASMUSSEN¹⁶⁾ verwendete; sie ist jedoch für die Diskussion etwas übersichtlicher und würde ausserdem eine grössere Stabilität der Messungen gewährleisten.

¹⁶⁾ R. E. H. RASMUSSEN: loc. cit. (14), S. 15.

Das zylindrische Rohr mit dem Radius r_0 und der Temperatur T_2 ist in zwei Teile oder Hälften, P und Q , geteilt und zwar durch die festen Zwischenwände ab und de und durch das seitwärts bewegliche Radiometerband $b^1d^1 = B$, dessen beide Seiten verschiedene Akkommodationskoeffizienten a_1 und a_2 besitzen; die Abstände bb^1 und dd^1 werden so klein wie praktisch möglich gehalten.

Wird das Platinband b^1d^1 der Breite B durch einen elektrischen Strom, i Amp., bis zu dem Widerstand W , bzw. der Temperatur T_1 , geheizt, entsteht eine stationäre Wärmeströmung von B nach der Umgebung; gleichzeitig unterliegt das Band B einer KNUDSEN'schen Radiometerkraft, die in den Versuchen mit Hilfe einer magnetischen Kompensationsmethode¹⁷⁾ gemessen wird. Der Vorteil dieser Methode ist, dass das Band zwischen den festen Wänden ab und de auf seinem Platz bleibt und hierdurch den Strömungszustand nicht beeinflusst. Wird die Breite des Bandes gegenüber dem Umkreis des Raumes P , bzw. Q , klein gewählt, wird der Druckunterschied zwischen dem Raum P und Q nicht allzu gross, und wir dürfen dann annehmen, dass dieser Druckunterschied nicht durch eine hydrodynamische Strömung von P nach Q durch die Öffnungen bb^1 und dd^1 beeinflusst, bzw. teilweise ausgeglichen wird; aufgrund der thermischen Gleitung entlang der Wände ab und de werden bei dieser Aufstellung, unabhängig von einander, Gasströmungen in P und Q entstehen, wie dies in der Zeichnung angedeutet ist. Ist $a_2 > a_1$, wird die Strömungsgeschwindigkeit in der Hälfte P laut der Theorie der thermischen Gleitung etwas grösser sein als in dem Raum Q .

Wir betrachten erst die Wärmeabgabe per cm^2 , Sek.

¹⁷⁾ MARTIN KNUDSEN: loc. cit. (4), S. 33.

und per Grad Temperaturdifferenz unter der Voraussetzung, dass $\frac{T_1 - T_2}{T}$ eine kleine Grösse ist; wir nennen die Wärmeabgabe per cm^2 , Sek. Grad von der Seite mit dem Akkommodationskoeffizienten a_2 , ${}_2q_p$ und von der anderen Seite ${}_1q_p$; diese sind bestimmt durch:

$${}_2q_p(T_1 - T_2)B = a_2 \cdot \varepsilon_{t_2} \cdot p(T_1 - t_2)B$$

und

$${}_1q_p(T_1 - T_2)B = a_1 \cdot \varepsilon_{t_1} \cdot p(T_1 - t_1)B,$$

worin t_2 und t_1 die Temperaturen des Gases im mittleren Abstand $\frac{1}{2}k\lambda$ von dem Band bezeichnen.

Laut der Untersuchung von MARTIN KNUDSEN sind die Werte von a_2 und a_1 dieselben für die Wärmeleitung und die Radiometerkraft; hieraus erhalten wir für die KNUDSEN'sche Radiometerkraft per cm Länge:

$$R = {}_2R_p - {}_1R_p = \frac{1}{4}a_2 \cdot \frac{p}{t_2}(T_1 - t_2)B - \frac{1}{4}a_1 \cdot \frac{p}{t_1}(T_1 - t_1)B$$

oder

$$R_1 = \frac{R}{B(T_1 - T_2)} = \frac{1}{4\varepsilon} \cdot \frac{1}{T} ({}_2q_p - {}_1q_p), \quad (24)$$

weil wir in erster Annäherung, wenn $\frac{T_1 - T_2}{T}$ eine kleine Grösse gegenüber 1 ist, setzen können:

$$t_1 = t_2 = \frac{T_1 + T_2}{2} = T$$

und

$$\varepsilon_{t_1} = \varepsilon_{t_2} = \varepsilon_T = \varepsilon.$$

Für ${}_2q_p - {}_1q_p$ haben wir aber den folgenden Ausdruck abgeleitet:

$${}_2q_p - {}_1q_p = (a_2 - a_1) \frac{\varepsilon p}{1 + (a_2 + a_1) \left(\frac{\varepsilon}{q_\alpha}\right) p + a_1 a_2 \left(\frac{\varepsilon}{q_\alpha}\right)^2 p^2},$$

sodass wir erhalten:

$$\frac{R}{T_1 - T_2} = \frac{1}{4} (a_2 - a_1) \frac{B}{T} \cdot \frac{p}{1 + (a_2 + a_1) \left(\frac{\varepsilon}{q_\alpha}\right) p + a_1 a_2 \left(\frac{\varepsilon}{q_\alpha}\right)^2 p^2}. \quad (25)$$

Hieraus erhalten wir auf die bekannte Weise:

$$\left(\frac{R}{T_1 - T_2}\right)_{\max} = \frac{1}{4 T} \cdot \frac{a_2 - a_1}{(\sqrt{a_1} + \sqrt{a_2})^2} \cdot \frac{q_\alpha}{\varepsilon} \cdot B \quad (26)$$

und

$$p_{\max} = \frac{q_\alpha}{\varepsilon} \cdot \frac{1}{\sqrt{a_1 a_2}} \quad (27)$$

mit

$$q_\alpha = \frac{\pi K}{B} \cdot \frac{1}{r_0 + \frac{1}{2} k \lambda} \cdot \frac{1}{\ln \frac{B}{\pi} + \frac{1}{2} k \lambda} \quad (28)$$

Wir sehen hieraus, dass in erster Annäherung $p_{\max} \propto \frac{1}{B}$, und dass $\left(\frac{R}{T_1 - T_2}\right)_{\max}$ unabhängig von der Breite B des Bandes ist.

Ausserdem erhalten wir die folgende Formel:

$$\left(\frac{R}{T_1 - T_2}\right)_{\max} : p_{\max} = \frac{1}{4 T} \cdot \frac{a_2 - a_1}{(\sqrt{a_1} + \sqrt{a_2})^2} \sqrt{a_1 a_2} \cdot B. \quad (29)$$

RASMUSSEN hat für seine Aufstellung, die prinzipiell dieselbe, wie die in Fig. 6 skizzierte, ist, nicht nur die Wärmeabgabe und die Radiometerkraft im molekularen Zustande, sondern auch diese Grössen bei höheren Drucken gemessen; bei diesen Messungen hat er die Temperaturdifferenzen $T_1 - T_2 = 40^\circ \text{C}$, 80°C und 120°C verwendet, während $T_2 = 273,1 + 20^\circ$ war. Die Messresultate sind in Tabelle VI wiedergegeben; in Kolonne 6 ist der Mittelwert, $\frac{\bar{R}}{T_1 - T_2}$, gegeben. Wenn die Genauigkeit der Messungen in Betracht

gezogen wird, kann dieser Wert am besten für $T_1 - T_2 = 80^\circ \text{C}$ angenommen werden.

Die Wärmeabgabe $\bar{q} = \frac{1}{2}({}_2q + {}_1q)$ in Kolonne 7 ist auch der Mittelwert der gemessenen Wärmeabgaben bei $T_1 - T_2 = 40^\circ \text{C}$, 80°C und 120°C , und entspricht also der Temperaturdifferenz $T_1 - T_2 = 80^\circ \text{C} = \overline{AT}$.

Tabelle VI.

Wasserstoff:

$$T_1 - T_2 = \overline{AT} = 80^\circ, \quad T_2 = 293^\circ,1, \quad \frac{T_1 + T_2}{2} = T = 333^\circ,1$$

$$\bar{q} = \frac{1}{2}({}_2q + {}_1q)$$

p (Bar)	$\log_{10} p$	$10^4 \cdot \frac{R}{AT}$	$10^4 \cdot \frac{R}{AT}$	$10^4 \cdot \frac{R}{AT}$	$10^4 \cdot \frac{\bar{R}}{AT}$	$\bar{q} = \frac{W \cdot i^2}{AT} \cdot 10^4$
		$AT = 40^\circ$	$AT = 80^\circ$	$AT = 120^\circ$	$\overline{AT} = 80^\circ$	
50,0	1,699	99	87	86	91	192
97,4	1,989	144	119	123	129	316
142,4	2,154	151	134	135	140	418
301,0	2,479	132	122	125	126	619
452	2,655	109	96	111	105	749
596	2,775	90	89	95	91	839
3460	3,539	36	26	30	31	1261
1950	3,290	56	40	43	46	1164
964	2,984	73	58	63	65	1026
715	2,854	82	66	79	76	935
531	—	—	—	—	—	843
394	—	—	—	—	—	744
38,2	1,582	80	75	71	75	156
73,9	1,869	111	113	107	110	264
107,7	2,032	131	130	123	128	350
139,7	2,145	135	136	131	134	418
199,0	2,299	167	154	140	154	518
64,6	1,810	119	114	99	111	240
125,9	2,100	156	141	128	142	390
183,8	2,264	153	142	131	142	490
347	2,540	133	122	117	124	690
499	2,698	95	100	100	98	814
814	2,911	80	68	78	75	—

Fig. 7 ist eine graphische Darstellung von $\frac{\bar{R}}{T_1 - T_2} \cdot 10^4$ in seiner Abhängigkeit von $\log_{10} p$; hieraus geht hervor, dass man eine symmetrische Kurve erhält, was auch nach der Formel (25) zu erwarten war. Wir finden hieraus:

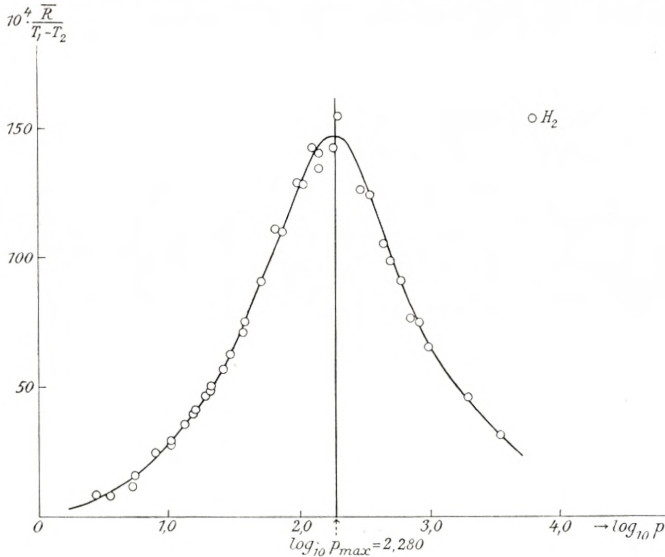


Fig. 7.

$$R'_{\max, \text{ obs.}} = \left(\frac{\bar{R}}{T_1 - T_2} \right)_{\max, \text{ obs.}} = 145 \cdot 10^{-4} \text{ Dyn/cm, Grad}$$

und $\log_{10} p_{\max} = 2,280$

oder

$$p_{\max} = 190 \text{ Bar.}$$

Wir erhalten also:

$$R'_{\max, \text{ obs.}} / p_{\max, \text{ obs.}} = 0,763 \cdot 10^{-4}.$$

Theoretisch finden wir für diese Grösse:

$$R'_{\max, \text{ ber.}} / p_{\max, \text{ ber.}} = \frac{1}{4 T} \cdot \frac{a_2 - a_1}{(\sqrt{a_1} + \sqrt{a_2})^2} \sqrt{a_1 a_2} \cdot B = 0,758 \cdot 10^{-4}$$

mit $a_2 = 0,735$, $a_1 = 0,315$ und $T = \frac{T_1 + T_2}{2} = 333^\circ,1$.

Durch die Messungen der Wärmeabgabe ist es auch möglich,

für das Messsystem den Wert von $\frac{q_\alpha}{\varepsilon}$ bei $p_{\max} = 190$ Bar zu erhalten. Eine graphische Darstellung von $\bar{q} = \frac{W \cdot t^2}{AT} \cdot 10^4$ in seiner Abhängigkeit von p ergibt eine glatte Kurve (Fig. 8),

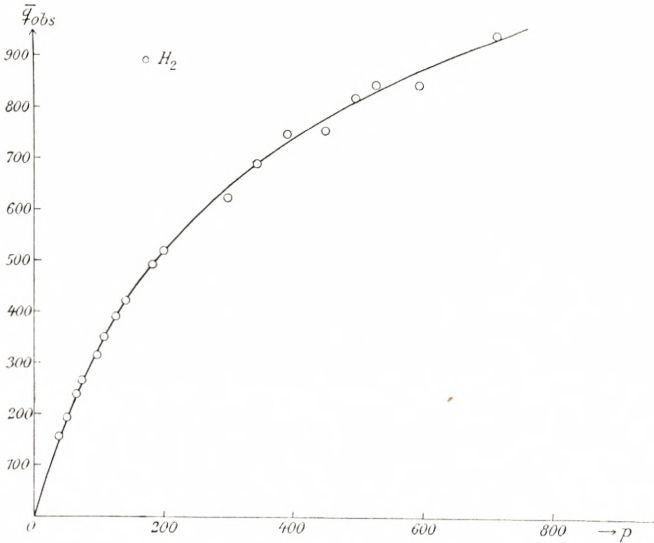


Fig. 8.

die für $p = 190$ Bar gibt: $q_p = 0,0510 \cdot \frac{10^7}{29,66} = 17200$;
hieraus findet man mit Hilfe der Formel:

$$\frac{1}{q_p} = \frac{1}{q_0} + \frac{1}{q_\alpha},$$

$$q_\alpha = 36740,$$

da

$$q_0 = \frac{1}{2} (a'_1 + a'_2) \varepsilon_T \cdot p = 32340,$$

weil

$$\varepsilon_T = 414,1 \quad \text{und} \quad \frac{1}{2} (a'_2 + a'_1)$$

$$= \frac{1}{2} \left\{ \frac{a_2}{1 + a_2 \frac{1 - a_0}{a_0} \cdot \frac{B}{\pi r_0}} + \frac{a_1}{1 + a_1 \frac{1 - a_0}{a_0} \cdot \frac{B}{\pi r_0}} \right\} = 0,411.$$

Hieraus erhält man:

$$\frac{q_{\alpha}}{\varepsilon_T} = \frac{36740}{414,1} = 89,0.$$

Verwendet man diesen Wert für $\frac{q_{\alpha}}{\varepsilon}$ zur Berechnung von R'_{\max} und p_{\max} , bekommt man:

$$R'_{\max, \text{ber.}} = \frac{1}{4 T} \cdot \frac{a_2 - a_1}{(\sqrt{a_1} + \sqrt{a_2})^2} \cdot \frac{q_{\alpha}}{\varepsilon} \cdot B = 140 \cdot 10^{-4} \begin{cases} \text{Dyn/cm,} \\ \text{Grad} \end{cases}$$

und

$$p_{\max, \text{ber.}} = \frac{q_{\alpha}}{\varepsilon} \cdot \frac{1}{\sqrt{a_1 a_2}} = 185 \text{ Bar.}$$

Wir ersehen hieraus, dass die Übereinstimmung zwischen den beobachteten und den berechneten Werten für R'_{\max} und p_{\max} befriedigend ist.

10. Man könnte sich vielleicht fragen, inwiefern es erlaubt ist, die Formel:

$$\frac{1}{q_p} = \frac{1}{q_0} + \frac{1}{q_{\alpha}} \quad \text{mit} \quad q_{\alpha} = \frac{\pi K}{B} \cdot \frac{1}{\ln \frac{2(l+m) + \pi k \lambda}{2B + \pi k \lambda}}$$

für die Aufstellung von RASMUSSEN anzuwenden; wir wollen dies an einigen der Messungen bei höherem Druck nachprüfen. RASMUSSEN fand bei dem Druck $p = 3460$ Bar die totale Wärmeabgabe per Grad Temperaturdifferenz des Bandes zu:

$$\frac{W \cdot i^2}{T_1 - T_2} = 0,1261,$$

woraus:

$$q_{p, \text{obs.}} = \frac{0,1261}{29,66} \cdot 10^7 = 42480 \text{ Erg/cm}^2, \text{ Sek., Grad.}$$

Wir finden für $p = 3460$ Bar $q_0 = \frac{1}{2}(a'_2 + a'_1) \varepsilon_T p = 0,411 \cdot 414,1 \cdot 3460 = 589300$ und

$$q_\alpha = \frac{2\pi K_{60}}{2B \ln \frac{9,36 + \pi k \lambda}{2,00 + \pi k \lambda}} = 41315 \text{ Erg/cm}^2, \text{ Sek., Grad,}$$

weil $\pi k \lambda = 0,05$ und $K_{60} = 4783 \cdot 10^{-7} \cdot 4,19 \cdot 10^7$.

Mit diesen Werten wird aus der Formel:

$$\frac{1}{q_p} = \frac{1}{q_0} + \frac{1}{q_\alpha},$$

gewonnen:

$$q_{p, \text{ber.}} = 38577 \text{ Erg/cm}^2, \text{ Sek., Grad.}$$

Wir ersehen hieraus, dass bei dem Druck $p = 3460$ Bar der beobachtete Wert von q_p etwa 10 % grösser ist als der berechnete; ob diese Abweichung ihre Ursache in dem Strömungszustand bei höherem Druck oder in einem Formfaktor hat, kann nicht mit Sicherheit gesagt werden; wahrscheinlich sind beide Ursachen von Bedeutung. Für die Messung bei dem Druck $p = 1950$ Bar finden wir in derselben Weise, dass $q_{p, \text{ber.}}$ etwa 6 % kleiner ist als der beobachtete Wert. Es erhellt jedenfalls hieraus, dass man für den RASMUSSEN'schen Apparat annäherungsweise die Formel:

$$\frac{1}{q_p} = \frac{1}{q_0} + \frac{1}{q_\alpha},$$

verwenden darf.

Wir sehen aus dem obenstehenden, dass die rein molekulartheoretische Behandlung der KNUDSEN'schen Radiometerkraft bei höheren Drucken — wenn der Apparat so eingerichtet ist, dass die entstehenden Strömungen nicht den molekularkinetischen Druckunterschied beeinflussen können — die richtige Beschreibung der Experimente und

des Entstehens der Radiometerkraft geben kann¹⁸⁾. Ganz anders werden aber die Verhältnisse, wenn die entstehenden Strömungen sich frei entwickeln und dadurch den sich bildenden Druckunterschied teilweise ausgleichen können, z. B. wenn dasselbe Band B in einem sehr grossen Behälter aufgestellt wird, so dass das Gas frei um das Band strömen kann; in diesem Falle wird der Strömungszustand so wie von MARTIN KNUDSEN angegeben¹⁹⁾, und das Radiometerproblem wird in erster Linie, jedenfalls solange die Strömungen laminar bleiben, ein rein hydrodynamisches Problem. Nachdem das hydrodynamische Problem gelöst ist, kann im allgemeinen auch das Radiometerproblem gelöst werden.

Wenn das Band in einem unbegrenzten Raum frei aufgestellt ist, wird die maximale Radiometerkraft R'_{\max} viel kleiner; dasselbe gilt auch für den Wert von p_{\max} . Mit einem praktisch gesprochen identischen Band: $a_1 = 0,315$, $a_2 = 0,715$, $B = 1,001$ cm, in einem sehr grossen Behälter fand R. E. H. RASMUSSEN²⁰⁾:

$$R'_{\max} = \left(\frac{R}{T_1 - T_2} \right)_{\max} = 30,4 \cdot 10^{-4} \text{ Dyn/cm, Grad}$$

und $p_{\max} = 33 \text{ Bar}$,

also ungefähr nur ein Fünftel der oben erwähnten experimentellen Werte in dem engen Rohr.

In einer folgenden Abhandlung soll die ziemlich komplizierte Theorie für diesen Fall näher beleuchtet und mit den experimentellen Ergebnissen verglichen werden.

¹⁸⁾ vgl. EARLE H. KENNARD: Kinetic Theory of Gases 1938, S. 334.

¹⁹⁾ MARTIN KNUDSEN: loc. cit. (4) Fig. 5, S. 58.

²⁰⁾ R. E. H. RASMUSSEN: loc. cit. (14), S. 13.

Zusammenfassung.

Ausgehend von der theoretischen Formel $K = \frac{5}{2} c_v \cdot \eta$ für die Wärmeleitfähigkeit der einatomigen Gase habe ich früher nachgewiesen, dass in einem Gase, in dem ein Temperaturgradient herrscht, die Moleküle, welche von dem mittleren Abstände λ kommen, die mittlere kinetische Energie besitzen, die der Temperatur im mittleren Abstand $\frac{1}{2} k \lambda$ entspricht; für einatomige Gase ist $k = \frac{15}{4}$; für mehratomige Gase wird $k = \frac{9\gamma - 5}{\gamma + 1}$, worin $\gamma = \frac{c_p}{c_v}$.

Für coaxiale Zylinder mit den Radien r_1 und r_0 wird hierdurch eine Formel für den Wärmetransport q_p zwischen diesen abgeleitet. Diese Formel kann folgendermassen geschrieben werden:

$$\frac{1}{q_p} = \frac{1}{q_0} + \frac{1}{q_\alpha},$$

worin:

$$q_\alpha = \frac{2\pi K(T_1 - T_2)}{\ln \frac{r_0 + \frac{1}{2}k\lambda}{r_1 + \frac{1}{2}k\lambda}} \quad \text{und} \quad q_0 = 2\pi r_1 \varepsilon p \frac{T_1 - T_2}{\frac{1}{\alpha_1} + \frac{1 - \alpha_0}{\alpha_0} \cdot \frac{r_1}{r_0}}.$$

Für $\frac{r_0}{\lambda}$ bzw. $\frac{r_1}{\lambda} \rightarrow \infty$ geht diese Formel in die bekannte SCHLEIERMACHER'sche Formel mit einer Korrektur für den Temperatursprung über; für $\frac{r_0}{\lambda}$ bzw. $\frac{r_1}{\lambda} \Rightarrow 0$ gibt diese Formel den richtigen Grenzwert q_0 für die rein molekulare Wärmeleitung.

Die Formel wird mit dem vorliegenden Beobachtungsmaterial für die Wärmeleitfähigkeit des reinen Neons und mit dem Beobachtungsmaterial für das Hitzdrahtmanometer von MARTIN KNUDSEN verglichen wie auch mit den ex-

perimentellen Resultaten von MARTIN KNUDSEN über den Einfluss des Akkommodationskoeffizienten auf die Wärmeleitung von Helium und Wasserstoff; die Übereinstimmung kann als befriedigend angesehen werden.

Zum Schluss wird der theoretische Zusammenhang zwischen dem Wärmetransport q_p und der KNUDSEN'schen Radiometerkraft R in einem besonderen, von R. E. H. RASMUSSEN experimentell untersuchten Fall geklärt. Ist $\frac{T_1 - T_2}{T} \ll 1$, stimmt die abgeleitete Formel:

$$\begin{aligned} \frac{R}{T_1 - T_2} &= \frac{1}{4 \cdot \varepsilon} \cdot \frac{B}{T} \{ {}_2q_p - {}_1q_p \} = \\ &= \frac{1}{4} (a_2 - a_1) \frac{B}{T} \frac{p}{1 + (a_2 + a_1) \left(\frac{\varepsilon}{q_\alpha} \right) p + a_1 a_2 \left(\frac{\varepsilon}{q_\alpha} \right)^2 \cdot p^2} \end{aligned}$$

mit dem vorliegenden Beobachtungsmaterial befriedigend überein.

Auch bei dieser Gelegenheit möchte ich der Direktion des Dänischen Carlsbergfonds meinen aufrichtigen Dank für die mir gewährte Unterstützung aussprechen.

Det Kgl. Danske Videnskabernes Selskab.

Mathematisk-fysiske Meddelelser. **XIX**, 12.

EINLEITUNG IN DIE
ALLGEMEINE KONGRUENZLEHRE

VON

JOHANNES HJELMSLEV

DRITTE MITTEILUNG



KØBENHAVN

I KOMMISSION HOS EJNAR MUNKSGAARD

1942

Printed in Denmark.
Bianco Lunos Bogtrykkeri A/S

I. Einleitung.

1. Die allgemeine Kongruenzlehre, wie wir sie auffassen und über welche wir früher zwei Mitteilungen¹⁾ veröffentlicht haben, wurde auf das folgende Axiomensystem gegründet.

I. Es gibt Punkte. Es gibt Punktmenge, welche gerade Linien (Geraden) heissen. Es gibt Transformationen, welche Bewegungen heissen. Jede Bewegung ist eine Zuordnung, durch welche jeder Geraden und jedem auf ihr gelegenen Punkte eine Gerade und ein auf ihr gelegener Punkt umkehrbar eindeutig entspricht. Die Umkehrung einer Bewegung ist auch eine Bewegung. Die Bewegungen bilden eine Gruppe. Zwei Figuren, die durch eine Bewegung auseinander abgeleitet werden können, sollen kongruent heissen.

II. Ausser der Identität gibt es eine und nur eine Bewegung, welche alle Punkte einer geraden Linie fest lässt. Diese Bewegung heisst Spiegelung an der geraden Linie. Die Linie wird als Achse der Spiegelung bezeichnet. Jede Gerade ist die Achse einer Spiegelung. Jeder Punkt ausserhalb der Achse geht bei der Spiegelung in einen von ihm verschiedenen Punkt über.

III. Eine Gerade b heisst senkrecht zu einer von ihr verschiedenen Geraden a (in Zeichen $b \perp a$), wenn b bei

¹⁾ Einleitung in die allgemeine Kongruenzlehre, Erste Mitteilung. D. Kgl. Danske Vidensk. Selskab, Math.-fys. Medd. VIII, 11, 1929; Zweite Mitteilung, ibd. X, 1, 1929.

der Spiegelung an a in sich selbst übergeht. Durch jeden Punkt geht eine und nur eine Gerade b , die senkrecht zu einer gegebenen Geraden a ist; die beiden Geraden haben stets einen und nur einen Punkt gemein.

IV. Wenn zwei Punkte A und B einer geraden Linie l angehören, haben sie immer eine Spiegelungsachse m , derart dass A und B bei der Spiegelung an m miteinander vertauscht werden, während l in sich selbst übergeführt wird ($l \perp m$). Diese Spiegelungsachse schneidet l in einem Punkt M , welcher als Mittelpunkt von A und B (oder von AB) auf l bezeichnet wird.

V. Zwei kongruente Punktreihen $ABC \dots$ und $AB'C' \dots$ auf einer oder auf zwei Geraden, mit dem gemeinsamen Punkt A können immer durch eine Spiegelung ineinander übergeführt werden.

In unserem Axiomensystem ist nicht ausdrücklich die Forderung aufgestellt, dass jeder Punkt durch Bewegung in jeden anderen Punkt übergehen kann. Es zeigt sich aber, dass diese Eigenschaft im Axiomensystem implizite enthalten ist. Durch zwei beliebige Punkte A , B können nämlich zwei zueinander senkrechte Geraden gelegt werden, und diese Geraden schneiden einander in einem Punkt C . Nach IV gibt es dann eine Spiegelung, welche A in C überführt, und eine Spiegelung, welche C in B überführt, somit auch eine Bewegung, welche A in B überführt.

2. Für die geraden Linien hingegen ist die entsprechende Eigenschaft keine Folge des Axiomensystems, was schon durch das einfache Beispiel, wo die Geometrie als die Cartesische Ebene mit ihren rationalen Punkten und Geraden definiert ist und die Bewegungen die rationalen orthogonalen Substitutionen sind, dargetan wird.

Bei dieser allgemeinen Fassung des Axiomsystems hat sich aber eine Schwierigkeit herausgestellt, welche mir bei den früheren Erwägungen entgangen war. In der ersten Mitteilung ist nämlich ohne Beweis behauptet worden, dass drei Geraden r, s, t durch denselben Punkt hindurchgehen müssen, wenn zwei von ihnen einen eindeutigen Schnittpunkt haben und ausserdem die Bewegung rst involutorisch ist. Es scheint aber zweifelhaft, ob dieser Satz aus den aufgestellten Axiomen I—V allein ableitbar ist. Setzt man voraus, was wir tatsächlich im folgenden tun wollen, dass alle Geraden kongruent sind, so ist der Satz sofort beweisbar. Wir bemerken aber, dass das folgende weniger aussagende Axiom VI genügen wird:

VI. Es gibt zwei zueinander senkrechte kongruente Geraden,

d. h. es gibt zwei zueinander senkrechte Geraden a, b , derart dass die eine, a , durch eine Bewegung in die andere, b , übergeführt werden kann. Geht hierbei der Schnittpunkt O der beiden Geraden als Punkt von a in den Punkt P von b über, so existiert eine Spiegelungsachse $p \perp b$ der beiden Punkte O, P , und es gibt somit auf a und b zwei kongruente Punktreihen mit gemeinsamem Punkt O , d. h. a und b haben eine Spiegelungsachse.

Es zeigt sich nun zunächst, dass unser neues Axiom VI die folgende weitergehende Eigenschaft nach sich zieht:

Je zwei zueinander senkrechte Geraden haben eine Spiegelungsachse.

Beweis. Wir wissen schon, dass die beiden gegebenen, zueinander senkrechten Geraden a, b mit dem Schnittpunkt O , eine Spiegelungsachse m haben. Durch O ziehen wir nun zwei beliebige andere zueinander senkrechte Geraden c, d . Es soll dann bewiesen werden, dass auch c, d

eine Spiegelungsachse haben. Wir bestimmen eine Gerade x durch die Gleichung

$$am = cx;$$

aus $ab = cd$ folgt nun (wegen $a \perp b$, $c \perp d$)

$$(am)(mb) = (cx)(xd),$$

oder, weil $am = mb = cx$ ist,

$$cx = xd,$$

d. h. x ist Spiegelungsachse von c , d ; und hiermit haben wir den Beweis erbracht.

Da nun ferner der Punkt O durch Bewegung in jeden beliebigen Punkt übergeführt werden kann, folgt sofort, dass je zwei zueinander senkrechte Geraden eine Spiegelungsachse haben, w. z. b. w.

Und nun schreiten wir zum Beweis des angekündigten Satzes:

Wenn 3 Geraden a , b , c eine involutorische Bewegung abc darstellen, und wenn zwei von ihnen, a , b , einen eindeutigen Schnittpunkt O haben, dann geht die dritte Gerade notwendig durch O .

Das Spiegelbild P von O in c geht bei der Bewegung $abc = cba$ in O über; hieraus folgt aber, dass P bei der Bewegung $abcc = ab$ fest bleiben muss. Die Bewegung ab lässt sich somit durch zwei Spiegelungen x , y , $x \perp a$ durch P , y durch P , ersetzen; aus $ab = xy$ folgt aber

$$by = ax,$$

und da $x \perp a$ ist, ist also auch $y \perp b$, und der Schnittpunkt von b , y muss mit dem Schnittpunkt von a , x zusammenfallen, das heisst, er fällt in O . Die beiden Geraden

x, y gehen also durch O und sind also die Lote der beiden Geraden a, b in O .

Ist nun m Spiegelungsachse der beiden zueinander senkrechten Geraden a und x , so führt die Bewegung am die Gerade a in x über; sie führt aber auch die Gerade b in y über, d. h. a und b werden durch die Bewegung am in x und y übergeführt. Da a und b nur einen Punkt O gemein haben, gilt dies auch von x und y d. h. $P = O$, oder P ist in c enthalten.

3. Das Axiomsystem enthält keine Axiome über die Bestimmung der geraden Linie durch zwei Punkte, und die Möglichkeit wurde nicht ausgeschlossen, dass zwei Punkte in gewissen Fällen keine Verbindungsgerade haben. Wir haben z. B. den Beweis erbracht, dass die beiden Punkte auch in diesen Fällen in dem Sinne einen Mittelpunkt haben, dass die beiden Punkte durch eine Umwendung um diesen Punkt ineinander übergehen. Ferner können wir gewissen Dreieckssätzen eine solche Formulierung geben, dass auch diejenigen Fälle in Betracht gezogen werden, wo die drei Ecken A, B, C nicht notwendig Verbindungsgeraden haben. Wenn z. B. A und B wenigstens eine Verbindungsgerade haben, während es offen bleibt, wie sich die Paare BC, CA , bezüglich der Existenz von Verbindungsgeraden verhalten, kann man von der Mittelsenkrechten p von AB und von den Mittelpunkten M von BC, N von CA , reden; p ist ein- oder mehrdeutig bestimmt, je nachdem die Gerade AB ein- oder mehrdeutig bestimmt ist; M und N sind immer eindeutig bestimmt. Es gilt dann der Satz, dass eine Gerade g existiert, welche durch M und N hindurchgeht und auf p senkrecht steht. Ist AB mehrdeutig, so wird $MN = g$ auch mehrdeutig, und die beiden Geraden entsprechen ein-

ander derart, dass sie stets eine (veränderliche) gemeinsame Normale p durch den festen Mittelpunkt von AB haben.

4. Im folgenden sollen aber nun unsere Untersuchungen stets auf reelle Geometrien beschränkt werden, d. h. auf solche Geometrien, wo je zwei Punkte wenigstens eine Verbindungsgerade haben; ferner sollen die elementaren Gesetze der Anordnung für die Punkte einer geraden Linie Gültigkeit haben; und endlich setzen wir auch die gewöhnlichen Regeln über die Teilung der Ebene durch jede ihrer Geraden in zwei Halbebenen voraus. Es folgt dann (vgl. Erste Mitteilung S. 17), dass zwei Geraden, welche zwei Punkte A, B gemein haben, beide die ganze Strecke AB enthalten müssen.

Schliesslich wollen wir bei den folgenden Untersuchungen auch stets voraussetzen, dass alle Geraden kongruent sind, dass also durch Bewegungen nicht nur jeder Punkt in jeden anderen, sondern auch jede Gerade in jede andere übergeführt werden kann.

II. Nachbarpunkte und Nachbargeraden.

5. Eine Strecke AB soll ordinär heissen, wenn die Gerade AB eindeutig ist, dagegen singulär, wenn dies nicht der Fall ist.

Zwei verschiedene, von einem Punkt O ausgehende Halbstrahlen a und b bilden einen Winkel mit dem Scheitel O . Wir setzen hier voraus, dass die Halbstrahlen zwei verschiedenen Geraden angehören; diese Geraden sind durch die Halbstrahlen a, b eindeutig bestimmt; durch Umwendung um O werden nämlich die Halbstrahlen a, b in die entgegengesetzten Halbstrahlen übergeführt. Wenn die beiden Halbstrahlen a, b zwei mehrdeutig schneidenden Geraden

angehören, soll der Winkel singulär heissen, spitz, wenn a und b mehrere Punkte, und dann auch Strecken gemein haben, und rund, wenn a und b nur den Punkt O gemein haben. Andere Winkel heissen ordinär. Ein runder singulärer Winkel ist Nebenwinkel eines spitzen singulären Winkels.

6. Ein singulärer spitzer Winkel kann in sich selbst verschoben werden. Ist O_1 ein von O verschiedener gemeinsamer Punkt von a und b , und ist M der Mittelpunkt von OO_1 , dann wird der Winkel durch die Schiebung $OM^1) = MO_1$ derart in sich selbst verschoben, dass der Scheitel O in O_1 übergeht. Alle singulären Winkel, deren Schenkel in den Schenkeln eines anderen singulären (spitzen) Winkels enthalten sind, sind also diesem Winkel kongruent (und unter einander kongruent). Auch ein runder Winkel kann in ähnlicher Weise in sich selbst verschoben werden.

7. Als Scheitelement eines singulären Winkels bezeichnen wir die Menge aller gemeinsamen Punkte der beiden durch die Schenkel des Winkels bestimmten Geraden. Dasselbe Element soll auch das Schnittelement der beiden Geraden genannt werden. Das Schnittelement ist ein konvexer Bereich d. h. je zwei Punkte der Menge sind die Endpunkte einer Strecke, deren Punkte alle der Menge angehören. Kongruente singuläre Winkel haben kongruente Schnittelemente.

8. Wenn zwei singuläre spitze Winkel nicht kongruent sind, so spricht man von einem grösseren und einem kleineren Winkel in dem Sinne, dass der kleinere Winkel in den grösseren so hineingelegt werden kann, dass die beiden Winkel einen Schenkel gemein haben. Man sieht dann,

1) Das ist die Schiebung, welche aus den Umwendungen O und M zusammengesetzt ist

dass der kleinere Winkel entweder dasselbe oder ein grösseres Scheitelelement als der grössere Winkel hat.

9. Drei Punkte A , B , C , welche nicht in derselben Geraden enthalten sind, bestimmen ein Dreieck ABC mit den

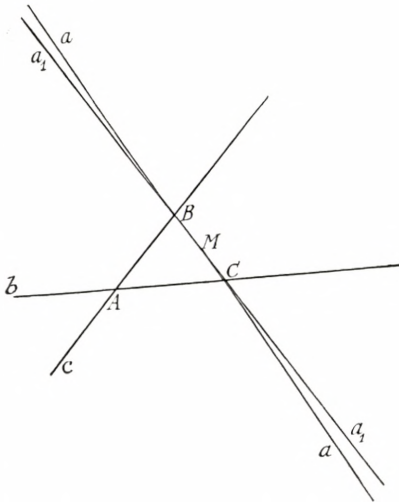


Fig. 1.

Seiten BC , CA , AB , wobei wir nur an die Strecken denken. Die Winkel des Dreiecks werden hingegen erst durch gerade Linien festgelegt, welche je zwei der Punkte verbinden: eine Gerade a durch B und C , eine Gerade b durch C und A , und eine Gerade c durch A und B . Sind diese Geraden alle eindeutig bestimmt (die Seiten sollen dann als ordinär bezeichnet werden), so sind auch die Winkel

eindeutig bestimmt: z. B. $\angle A$ durch die Halbstrahlen von A durch B und von A durch C . Wenn aber die Verbindungsgeraden der Ecken mehrdeutig sind, werden auch die Winkel mehrdeutig. Die Summe der Winkel ist aber hiervon unabhängig. Wenn wir uns nämlich vergegenwärtigen, welche Veränderung es bewirkt, wenn eine Seitengerade a durch eine andere a_1 ersetzt wird, so ergibt sich, dass die durch die Seitengeraden a , b , c festgelegten Winkel

$$\beta = CBA, \quad \gamma = ACB$$

und die durch die Seitengeraden a_1 , b , c festgelegten Winkel

$$\beta_1 = CBA, \quad \gamma_1 = ACB,$$

dieselbe Summe haben.

In der Tat ist

$$\beta_1 = \beta + x, \quad \gamma_1 = \gamma - x_1,$$

wo x den von den beiden Halbstrahlen von B durch C längs a bzw. a_1 gebildeten Winkel und x_1 den von den Halbstrahlen von C durch B längs a bzw. a_1 gebildeten Winkel bedeuten. Diese Winkel x und x_1 sind aber untereinander gleich; denn sie gehen bei der Umwendung um den Mittelpunkt M von BC ineinander über.

Es ist also

$$\beta_1 + \gamma_1 = \beta + \gamma,$$

d. h.

Die Winkelsumme im Dreieck ist bei mehrdeutigen Seitengeraden von der Wahl dieser Geraden unabhängig und hängt also nur von den Ecken des Dreiecks ab.

Entsprechendes ergibt sich für die Winkelsumme eines beliebigen Polygons.

10. In jedem Dreieck mit einem rechten Winkel sind die beiden anderen Winkel spitz.

Es sei ABC (Fig. 2) ein rechtwinkliges Dreieck mit dem rechten Winkel $C = (ab)$. In A errichten wir die Normale a_1 auf b . Die beiden Geraden a, a_1 begrenzen einen Streifen, welcher die ganze Strecke AB enthält. Hieraus folgt aber, dass der Winkel bei A spitz ist. Ebenso ergibt sich dies für den Winkel bei B .

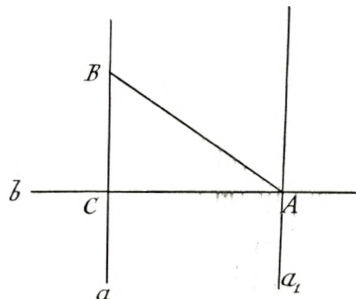


Fig. 2.

11. In einem rechtwinkligen Dreieck können nicht beide spitzen Winkel singularär sein.

Anderenfalls könnte man (Fig. 3) von den Ecken A , B aus zwei gleich grosse Strecken AA_1 und B_1B auf der

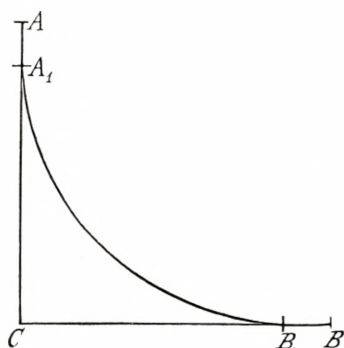


Fig. 3.

Hypotenuse AB derart abgrenzen, dass die eine Strecke AA_1 zugleich in der Kathete AC , die andere BB_1 zugleich in der Kathete BC enthalten ist. Die Bewegung $AA_1 = B_1B$ würde dann beide Kathetenlinien a und b in sich überführen, also den Punkt C fest lassen, was unmöglich ist.

12. Wie sofort ersichtlich, lässt sich der Satz zu dem folgenden erweitern:

In jedem Dreieck ABC mit einem ordinären Winkel A (wo also die Seitengeraden AB und AC nur einen Punkt A gemein haben), können die anderen Winkel niemals beide singulär sein.

13. Zwei Halbstrahlen a_1 und b_1 (Fig. 4), die von zwei Punkten A bzw. B einer Geraden g nach derselben Seite dieser Geraden ausgehen und mit der Strecke AB singuläre spitze Winkel bilden, müssen sich schneiden. Fällt

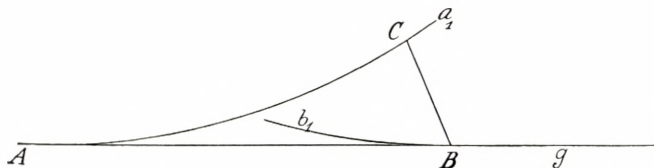


Fig. 4.

man nämlich von B die Senkrechte BC auf a_1 , so erhält man ein rechtwinkliges Dreieck ACB , dessen Winkel A singulär ist, und dessen anderer Winkel B somit ordinär

ausfallen muss. Hieraus folgt aber, dass der Halbstrahl b_1 , welcher mit BA einen singulären spitzen Winkel bildet, in den Winkel ABC eindringt; er muss daher die Seite AC schneiden oder durch A hindurchgehen.

14. Zwei Punkte A, B sollen als Nachbarpunkte (oder als benachbart) bezeichnet werden, wenn sie mehr als eine, also unendlich viele Verbindungsgeraden haben. Die Strecke AB ist dann singulär. Zieht man durch A, B zwei Geraden a, b , welche einander eindeutig schneiden (z. B. zwei zueinander senkrechte Geraden), so ist ihr Schnittpunkt C zu A und B benachbart. Sind nämlich p und q zwei sowohl durch A als auch durch B hindurchgehende Geraden, so lässt sich die Bewegung pq durch ax ersetzen, wo x durch A geht, oder durch by , wo y durch B geht. Es ist dann $ax = by$, also $bax = y$ involutorisch, und da a, b einen eindeutigen Schnittpunkt C haben, muss x durch diesen Punkt gehen, d. h. A und C haben mehrere Verbindungsgeraden. Aus C und A (oder B) lassen sich dann neue Nachbarpunkte ableiten. Wir wissen auch, dass die Mittelpunkte der Strecken AB, AC , u. s. w. Nachbarpunkte von A sind, ebenso die Punkte A^B, A^C , u. s. w. (Erste Mitteilung, S. 17).

Bemerkung. Es könnte natürlich vorkommen, dass alle Punkte unserer Ebene benachbart sind, d. h. dass je zwei Punkte unendlich viele Verbindungslinien haben, also dass alle Strecken singulär sind. Dieser Fall soll bis auf weiteres ausgeschlossen werden. Im folgenden soll also stets vorausgesetzt werden, dass wenigstens ein Punktepaar (und dann unendlich viele) existiert, dessen Verbindungslinie eindeutig ist.

15. Wir wollen nun alle Nachbarpunkte von A zu einer Menge \textcircled{A} zusammenfassen. Wir bezeichnen A als Mittel-

punkt der Menge, werden aber sofort zeigen, dass jeder Punkt der Menge als ihr Mittelpunkt betrachtet werden kann, indem wir den folgenden Satz beweisen:

Zwei Nachbarpunkte B, C eines Punktes A sind auch Nachbarpunkte voneinander.

Der Beweis verläuft in folgender Weise:

Fall 1. A, B liegen auf einer Geraden g , AC auf einer anderen Geraden $h \perp g$, oder allgemeiner: g und h haben einen eindeutigen Schnittpunkt.

Durch B und C ziehen wir eine Gerade l ; sie schneidet jedenfalls eine der beiden Geraden g, h eindeutig. Und hieraus folgt (14), dass B und C Nachbarpunkte sind. Also:

Wenn zwei Katheten eines rechtwinkligen Dreiecks singularär sind, so auch die Hypotenuse.

Fall 2. A, B, C liegen alle auf einer Geraden g . Wir errichten in A ein Lot h und tragen auf ihm mittels einer Spiegelung an einer Spiegelungsachse der beiden Geraden g, h die Strecke $AD = AB$ ab. Wir ziehen die Gerade l durch B und D senkrecht zu dieser Spiegelungsachse. Die Winkel bei B und D im Dreieck BAD müssen dann beide ordinär ausfallen, weil sie nicht beide singularär sind. Nach Fall 1 sind nun D und C Nachbarpunkte, und da die beiden Geraden l und g einander eindeutig schneiden, folgt aus 14, dass B und C benachbart sind.

Wenn zwei Strecken singularär sind, so ist auch ihre Summe und ihre Differenz singularär.

Fall 3. Dieser Fall ist der Hauptfall: A, B, C sind 3 beliebige Punkte, welche nicht in einer Geraden liegen.

Wir ziehen eine Gerade g durch B und C und fällen das Lot AA_1 auf g . Da nun A und B benachbart sind, gilt dies auch von A_1 und B . Ebenso: A und C sind benachbart, daher auch A_1 und C . Da nun A_1 zu B und C benach-

bart ist, und diese 3 Punkte auf einer Geraden liegen, folgt (Fall 2), dass B und C untereinander benachbart sind.

Hiermit ist unser Beweis vollständig erledigt.

16. Ein Nachbargebiet \textcircled{A} mit Mittelpunkt A wird nun nach diesem Satz durch irgend einen seiner Punkte bestimmt. Wenn zwei Nachbargebiete einen Punkt gemein haben, müssen sie identisch sein.

Das Nachbargebiet ist ein konvexer Bereich, hat aber keine Grenzpunkte.

Es lässt sich leicht bestätigen, dass im Nachbargebiet \textcircled{A} eine Geometrie gilt, in welcher dieselben Axiome erfüllt sind wie in unserer ursprünglichen Geometrie. Unter einer Geraden im Gebiet \textcircled{A} verstehen wir nur diejenige Teilmenge einer ursprünglichen Geraden, welche dem Gebiet angehört. In naheliegender Weise lassen sich auch andere geometrische Begriffe auf das Gebiet \textcircled{A} überführen, und nach den vorstehenden Untersuchungen gelten dann eben alle Axiome unserer Geometrie.

17. Wir wollen nun auch von Nachbargeraden sprechen. Zwei Geraden a , b sollen Nachbargeraden heißen, wenn zu jedem Punkt P der Geraden a ein Nachbarpunkt Q auf der Geraden b existiert. Das Lot PN von P auf b trifft dann diese Gerade stets in einem Nachbarpunkt Q von P .

Zwei Geraden a , b mit mehrdeutigem Schnittpunkt sind notwendig Nachbargeraden.

Die beiden Geraden a , b (Fig. 5) mögen etwa die Punkte A , B gemein haben. Von einem auf a (aber nicht auf b) gelegenen Punkt P fallen wir das Lot c auf b . Der Fusspunkt sei Q . Die Lote auf a und b in A seien n und n_1 ; ausser A haben sie noch z. B. den Punkt C gemein, welcher durch Abtragen der Strecke $AC = AB$ erhalten wird. Die

Bewegung ab ist mit der Bewegung nn_1 gleichwertig und hat also ausser A (und B) den Fixpunkt C . Wir ziehen nun die Verbindungsgerade $CP = c_1$; sie ist von a verschieden, weil C nicht auf a liegt. Die Bewegung abc_1 ist nun der Bewegung nn_1c_1 gleichwertig und daher involutorisch. Mit Hilfe dieser Tatsache lässt sich dann zeigen, dass die beiden

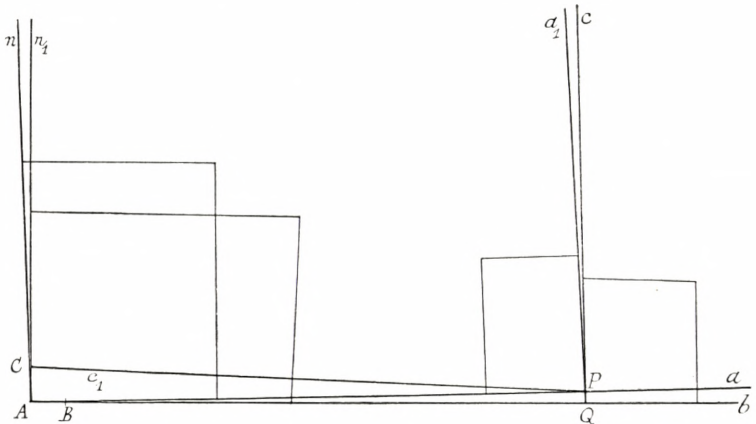


Fig. 5.

Punkte P , Q ausser c noch eine Verbindungslinie haben, nämlich die Achse c' der Spiegelung ac_1c . Hierzu verwenden wir den Satz von den drei Spiegelungen in Involution (Erste Mitteilung S. 24—25): Durch P gehen die Lote c auf b und a_1 auf a . Es gibt dann eine Verbindungslinie PQ senkrecht zu derjenigen Geraden m , welche durch die Gleichung

$$m = a_1c_1c$$

bestimmt wird. Diese Verbindungsgerade PQ ist dann als Spiegelungsachse c' derart bestimmt, dass

$$c' = Pm = Pa_1c_1c = ac_1c,$$

wie oben behauptet wurde.

Dass die Gerade c' von c verschieden ist, folgt einfach daraus, dass a von c_1 verschieden ist. Es hat sich also herausgestellt, dass die Punkte P, Q zwei verschiedene Verbindungslinien haben, d. h. die beiden Geraden a, b sind Nachbargeraden.

In der obigen Untersuchung kann der Punkt C durch irgend einen anderen gemeinsamen Punkt der beiden Geraden n, n_1 ersetzt werden. Wenn C das ganze Schnittelement (n, n_1) durchläuft, beschreibt die Gerade c_1 das projizierende Büschel $P(C)$ aus P , und gleichzeitig beschreibt die Verbindungsgerade c' von P, Q , welche durch die Gleichung

$$c' = ac_1c$$

bestimmt ist, ein entsprechendes Büschel. Da $ac_1 = c'c$, sind die beiden von c_1 und c' beschriebenen Büschel symmetrisch.

18. Zwei Nachbargeraden mit einem gemeinsamen Punkt P , schneiden sich mehrdeutig (in einem Schnittelement um P).

Beweis. PQ sei eine ordinäre Strecke auf der einen Geraden, PR ihre Projektion auf die andere. QR ist dann eine singuläre Strecke, und wenn die beiden Geraden einen eindeutigen Schnittpunkt hätten, würde das bedeuten, dass auch PQ singulär wäre, was nicht der Fall ist.

19. Haben zwei Punkte A, B eine eindeutige Verbindungslinie g , und ist B_1 Nachbarpunkt von B , so haben A und B_1 eine eindeutige Verbindungslinie g_1 , welche Nachbargerade von g ist.

Hätten nämlich g und g_1 einen eindeutigen Schnittpunkt in A , während B und B_1 Nachbarpunkte sind, so würde hieraus folgen, dass auch A und B benachbart wären, was nicht zutrifft.

20. Wenn zwei Geraden g , h Nachbargeraden derselben Geraden l sind, dann sind sie auch Nachbargeraden voneinander.

Von einem auf g gelegenen Punkt P fallen wir das Lot PQ auf l und von Q das Lot QR auf h . Da P und R Nachbarpunkte von Q sind, sind sie auch Nachbarpunkte voneinander. Jeder Punkt P von g hat somit einen Nachbarpunkt auf h , w. z. b. w.

21. Wenn zwei Punkte A und B eine eindeutige Verbindungsgerade haben und A_1 und B_1 Nachbarpunkte von A bzw. B sind, so ist die Gerade A_1B_1 Nachbargerade von AB .

Denn AB und A_1B_1 sind beide Nachbargeraden derselben Geraden AB_1 .

22. Alle Nachbarpunkte einer Geraden l (d. h. Nachbarpunkte der Punkte von l) bilden ein Nachbargebiet von l . Dieses Nachbargebiet (l) enthält alle Nachbargeraden von l , und ist durch irgend eine dieser Geraden eindeutig bestimmt.

23. Zwei Lote a , a_1 einer Geraden g , welche in zwei Nachbarpunkten auf g errichtet sind, sind Nachbargeraden voneinander.

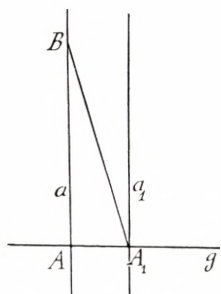


Fig. 6.

Auf a wählen wir einen Punkt B (Fig. 6) derart, dass A und B nicht Nachbarpunkte sind. Die Gerade BA_1 ist dann Nachbargerade von a (19). Sie schneidet a mehrdeutig in B (18), und hieraus folgt weiter¹⁾, dass sie

auch a_1 mehrdeutig in A_1 schneidet und somit auch Nachbargerade von a_1 ist. Hieraus folgt dann der Satz.

¹⁾ Erste Mitteilung, S. 18.

24. Wenn zwei Geraden g , h , welche einander nicht schneiden, eine singuläre Verbindungsstrecke PQ haben (also zwei Nachbarpunkte enthalten), sind sie notwendig Nachbargeraden.

Von dem auf g gelegenen Punkt P fallen wir das Lot l auf h (Fig. 7). Sein Fusspunkt sei Q_1 . Die Strecke PQ_1 ist ebenfalls singulär. Errichten wir nun in P das Lot n auf l , so wissen wir, dass n und h benachbart sind. Fällt n

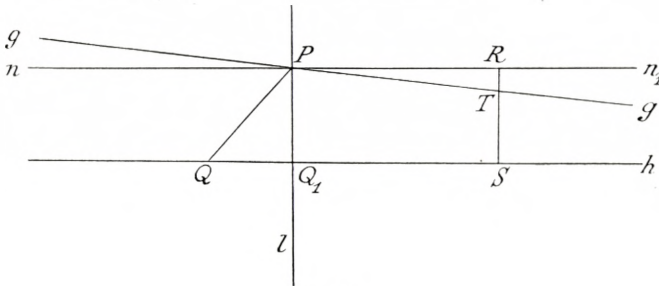


Fig. 7.

mit g zusammen, so ist die Behauptung schon bewiesen. Ist dies aber nicht der Fall, so müssen n und g einander schneiden. Es gibt dann auf g einen Halbstrahl g_1 , welcher innerhalb des Streifens nh verläuft. Derjenige Halbstrahl von n , welcher in derselben von l begrenzten Halbebene wie g_1 verläuft, sei mit n_1 bezeichnet. Auf n_1 tragen wir einen ordinären Abstand PR ab und fallen von R das Lot RS auf h . Die Strecke RS muss dann g_1 schneiden; der Schnittpunkt sei T . Die Strecke RS ist singulär, RT und TS also auch. Die Strecke PT ist hingegen ordinär. Hieraus folgt aber, dass die beiden Geraden g und h Nachbargeraden sind (21).

Es lässt sich also auch folgendes aussagen:

Zwei Geraden, von denen jede durch einen von zwei Nachbarpunkten hindurchgeht, sind ent-

weder Nachbargeraden oder sie schneiden sich eindeutig.

25. Schneidet eine Gerade c die eine, a , von zwei einander benachbarten Geraden a , b eindeutig in A , so muss sie auch die andere Gerade, b , schneiden, und zwar ebenfalls eindeutig.

Beweis. Das Lot AN von A auf b ist eine singuläre Strecke, und da b und c nicht benachbart sind, weil a und b es sind, a und c aber nicht, so müssen die beiden durch A und N hindurchgehenden nicht benachbarten Geraden b und c sich eindeutig schneiden.

Insbesondere schneidet also jede Normale auf dem einen Schenkel eines singulären Winkels den andern Schenkel eindeutig.

Eine Gerade, welche zwei benachbarte Geraden schneidet, schneidet also entweder beide eindeutig oder beide mehrdeutig.

26. Es ist nun leicht zu beweisen, dass die Lote a , b von zwei Nachbarpunkten A , B auf zwei Nachbargeraden g , h gleichfalls Nachbargeraden sind, und dass ihre Fusspunkte G , H Nachbarpunkte sind. Es ist hinreichend, den Fall zu behandeln, wo g , h einander schneiden; in den anderen Fällen braucht man nur eine dritte Nachbargerade i , welche g und h schneidet, einzuschalten. Wenn aber g und h sich schneiden, wissen wir nach dem vorigen Satz, dass die beiden Geraden a , b , weil sie durch zwei Nachbarpunkte A , B hindurchgehen, entweder Nachbargeraden sind oder einander eindeutig schneiden. Der letztere Fall kann aber hier nicht eintreffen, weil zwei Normalen, welche von einem Punkt auf zwei Geraden mit eindeutigem Schnittpunkt gefällt werden, selbst eindeutigen Schnittpunkt haben müssen.

Dass G und H Nachbarpunkte sind, schliesst man z. B. daraus, dass a und h einander in einem solchen Punkt I schneiden müssen, dass GI und IH singuläre Strecken sind. GH ist somit ebenfalls singulär.

27. Erklärung. Ein ordinärer Winkel α soll wesentlich kleiner oder grösser als ein ordinärer Winkel β genannt werden, wenn er kleiner bzw. grösser als β ist und der Unterschied zwischen den beiden Winkeln ein nicht singulärer Winkel ist.

Gelegentlich wird auch den Ausdruck gebraucht, ein Winkel sei wesentlich spitz oder wesentlich stumpf, d. h. wesentlich kleiner bzw. grösser als ein Rechter.

28. Zwei Halbstrahlen a und b , die von den Endpunkten A und B einer singulären Strecke AB auf einer Geraden c nach derselben Seite dieser Geraden ausgehen und mit der Strecke AB in A einen wesentlich spitzen und in B einen rechten Winkel bilden, müssen sich schneiden, und zwar eindeutig. Denn die Geraden a , b sind nicht benachbart, weil die Normale n in A auf c nicht zu a benachbart ist.

29. Es lässt sich dann das folgende modifizierte »5te Postulat Euklids« beweisen:

Zwei Halbgeraden a , b , welche von zwei benachbarten Punkten A , B einer Geraden c nach derselben Seite dieser Geraden ausgehen und mit der Strecke AB zwei Winkel (ac) und (bc) bilden, deren Summe wesentlich kleiner als zwei Rechte ist, müssen einander schneiden, und zwar eindeutig (Fig. 8).

Bei einer Umwendung um den Mittelpunkt M von AB geht a in a_1 über; a_2 sei der zu a_1 entgegengesetzte Halbstrahl. Der Winkel ba_2 ist dann nicht singulär, und b , a_2

sind somit nicht benachbart. Da aber a und a_2 benachbart sind, sind also a und b nicht benachbart. Und hieraus folgt die Behauptung.

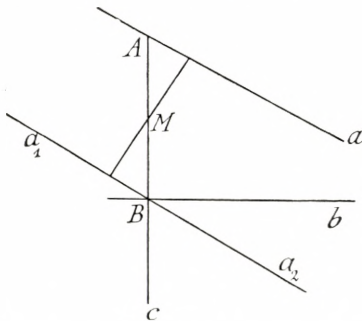


Fig. 8.

Innerhalb des Nachbarbereichs eines Punktes können wir dann jedenfalls behaupten, dass die Winkelsumme eines Dreiecks nur um einen singulären Winkel von zwei Rechten abweicht: Die Geometrie dieses Be-

reichs ist eine »fast-Euklidische«.

30. In diesem Zusammenhang möchten wir noch bemerken, dass man auch für eine beliebige Strecke AB eine spezielle Form des »5ten Postulats« aufstellen kann, nämlich die folgende:

Von zwei Halbgeraden a, b , welche von zwei beliebigen Punkten A, B einer Geraden c nach derselben Seite dieser Geraden ausgehen, bilde die eine, a , mit der Strecke AB in A einen singulären spitzen Winkel und die andere, b , in B einen Winkel, welcher wesentlich kleiner als zwei Rechte ausfällt. Dann müssen diese Halbgeraden einander schneiden, und zwar eindeutig.

Der Satz ergibt sich sofort aus dem Satz in 25, da von den beiden Nachbargeraden a, c die eine, c , von b eindeutig geschnitten wird, und daher auch die andere, a . Der Winkel (BA, b) kann auch singulär spitz sein, ein Fall der schon früher behandelt worden ist.

31. In zwei Nachbargeraden g, h lassen sich unendlich viele Rechtecke derart einschreiben,

dass zwei Gegenseiten des Rechtecks auf den beiden Geraden liegen.

In dem Falle, wo die beiden Geraden g , h einander schneiden, ist der Satz schon früher bewiesen worden (Erste Mitteilung S. 18).

In dem Fall, wo die Geraden keinen Punkt gemein haben, wählen wir auf g (Fig. 9) eine ordinäre Strecke PQ , fällen von Q das Lot QR auf h und ziehen die Gerade PR . Da die Strecke PQ ordinär, die Strecke QR hingegen

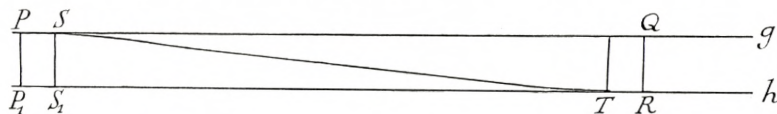


Fig. 9.

singulär ist, muss die Strecke PR ordinär sein. Die Gerade PR ist dann Nachbargerade zu g (und zu h), hat also ein Schnittlement um den Punkt P mit g und ein Schnittlement um R mit h gemein. Es lassen sich sodann zwei gleich grosse Strecken PS und TR in diesen Schnittlementen wählen, und wir wissen dann, dass die Strecke PS und ihre Projektion P_1S_1 auf h Gegenseiten eines Rechtecks sind. Hiermit ist unser Satz bewiesen.

32. In dem Falle, wo g und h sich schneiden, ist die Länge der Gegenseiten des eingeschriebenen Rechtecks durch das Schnittlement beschränkt, da nämlich gerade jeder Strecke innerhalb dieses Elements ein Seitenpaar dieser Länge entspricht. Auch wenn g und h sich nicht schneiden, kann man ein Element angeben, welches alle diejenigen Streckenlängen enthält, die bei Seitenpaaren eingeschriebener Rechtecke vorkommen können. Wir wollen dieses Element als Schiebelement der beiden Geraden g , h bezeichnen. Es gibt in der Tat Bewegungen, bei denen

alle Elemente dieser Grösse auf den beiden Geraden in sich verschoben werden. Diese Eigenschaft ist für Nachbargeraden charakteristisch.

33. Wenn die beiden Geraden g, h sich nicht schneiden, kann das Schiebelelement die ganze Gerade umfassen. Das bedeutet, dass jede Normale der einen Geraden auch Normale der anderen ist.

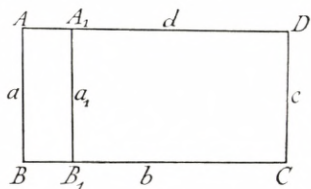


Fig. 10.

Es sei nun $ABCD$ (Fig. 10) irgend ein dreieckiges Viereck mit den Seitengeraden $a, b, c, d, d \perp a, a \perp b, b \perp c$; ferner sei a_1 eine Nachbargerade von a mit obengenannter Eigenschaft, d. h. $a_1 \perp d$ und $\perp b$. Die Bewegung aa_1c ist dann involutorisch, und wegen $aa_1 = AA_1$ ist auch AA_1c involutorisch, d. h. es gibt eine auf c senkrechte Gerade durch A, A_1 , also eine auf c senkrechte Nachbargerade zu d .

Die Abweichung des Winkels D von einem rechten Winkel ist also singulär.

Gibt es also in unserer Ebene zwei Nachbargeraden mit gemeinsamem Normalensystem, so ist die Winkelsumme eines jeden Dreiecks nur um einen singulären Winkel von zwei Rechten verschieden.

Gibt es ein Rechteck mit ordinären Seiten, so ist die Winkelsumme eines jeden Dreiecks genau zwei Rechte (Erste Mitteilung, S. 19—20).

34. Zwei Nachbargeraden g, h bestimmen eine Bewegung gh , welche jeden Punkt in einen Nachbarpunkt überführt (eine singuläre Bewegung).

Der Punkt A gehe bei der Spiegelung g in A_1 und der Punkt A_1 bei der Spiegelung h in A_2 über; dann sind A und

A_2 benachbart, weil sie Spiegelbilder von A_1 an den beiden benachbarten Geraden g, h sind.

Haben g und h keinen Punkt gemein, so hat die Bewegung gh keinen Fixpunkt. Es müssten sonst nämlich zwei Punkte P, Q mit zwei verschiedenen Spiegelungsachsen g, h existieren, und diese Spiegelungsachsen hätten einen Punkt gemein nämlich den Mittelpunkt von PQ , was unserer Annahme widerstreitet.

Haben g und h einen Punkt und somit (weil sie Nachbargeraden sind) unendlich viele Punkte gemein, so hat die Bewegung unendlich viele Fixpunkte, nämlich erstens alle gemeinsamen Punkte P von g und h und zweitens alle gemeinsamen Punkte jedes Geradenpaares g_1, h_1 mit $g_1 \perp g, h_1 \perp h$ im Punkte P . Die hierdurch bestimmte Fixpunktmenge bildet einen konvexen Bereich, welcher von jeder Geraden, die in den Bereich eindringt, in einem Element geschnitten wird, das dem Schnittelement von g und h kongruent ist¹⁾.

Zwei Nachbarpunkte A, B bestimmen eine singuläre Bewegung AB , welche jede Gerade durch A und B fest lässt.

Sind g und h zwei Nachbargeraden und l eine beliebige Gerade, so ist die bei der Bewegung ghl feste Gerade eine Nachbargerade von l . Es folgt dies daraus, dass die Gerade l bei der Bewegung gh in eine Nachbargerade l' von sich übergeht und sodann bei der Bewegung ghl in eine andere Nachbargerade l'' übergeführt wird.

35. Wenn zwei gleichartig orientierte Nachbargeraden g, h von einer dritten orientierten Geraden l geschnitten werden, so ist der Unterschied zwischen den beiden Schnittwinkeln (g, l) und (h, l)

¹⁾ Vgl. hierzu Erste Mitteilung, S. 31—32.

ein singulärer Winkel. Es folgt dies daraus, dass eine Schiebung längs l , welche den Scheitel des ersten Winkels in den Scheitel des zweiten überführt, die orientierte Gerade g in eine zur orientierten Geraden h benachbarte orientierte Gerade g' überführt.

In ähnlicher Weise erhält man den folgenden allgemeineren Satz:

Der Unterschied zwischen zwei beliebigen benachbarten Winkeln ist stets ein singulärer Winkel.

Natürlich gilt ein entsprechender Satz für benachbarte Strecken.

III. Verallgemeinerung der Saccheri-Lambertschen Sätze.

36. Wir betrachten ein Viereck $ABCD$ mit rechten Winkeln bei A , B , C . Wir setzen voraus, dass die »Katheten«, d. h. die beiden Seiten AB und BC ordinäre Strecken sind. Einander gegenüberliegende Seitengeraden des Vierecks haben dann ordinäre Lage zueinander, d. h. sie sind nicht Nachbargeraden, und die beiden Strecken CD und DA sind dann ebenfalls ordinär. Vom vierten Winkel D lässt sich nun zunächst aussagen, dass er ordinär sein muss. Wenn nämlich die Geraden AD und CD benachbart wären, müssten auch ihre Lote durch B , nämlich BA und BC , benachbart sein, was unserer Voraussetzung, dass sie aufeinander senkrecht stehen, widerstreitet.

37. Im Viereck $ABCD$ (Fig. 11) verlängern wir nun AD über D hinaus nach E und fällen die Senkrechte EF von E auf die Verlängerung von BC . Hierdurch entsteht ein neues dreieckwinkliges Viereck $ABFE$ mit ordinären

Seiten. Vergleichen wir nun die Winkel der beiden Vierecke bei D und E , so zeigt sich, dass sie entweder beide spitz, beide stumpf, oder beide rechte Winkel sind.

1) Zunächst untersuchen wir den ersten Fall, wo $\angle ADC$ spitz ist, und wollen zeigen, dass $\angle AEF$ dann auch spitz ausfallen muss (Fig. 11). Da $\angle ADC$ spitz ist, hat das Lot von A auf die Gerade CD seinen Fusspunkt Q auf der Strecke CD . Ferner hat das Lot von E auf die Gerade CD seinen Fusspunkt R auf der Verlängerung der Strecke CD über D hinaus, jedenfalls wenn die Strecke ED ordinär ist, was wir zunächst voraussetzen wollen. Die Strecke AR ist nun ordinär, und ihre Verlängerung über R hinaus bildet mit der Strecke RE einen spitzen Winkel,

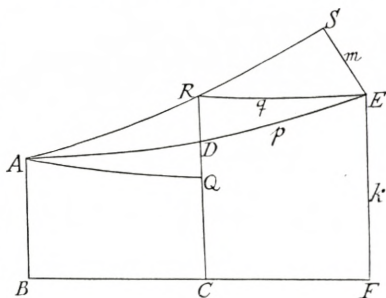


Fig. 11.

weil $\angle ARQ$ spitz und $\angle QRE$ recht ist. Das Lot m von E auf die Gerade AR muss also seinen Fusspunkt S auf dieser Verlängerung haben. Die drei Geraden AB , DC und EF sind Lote derselben Geraden BCF und somit in Involution. Vom Punkte E auf der dritten dieser Geraden haben wir Lote p , q auf die beiden ersten gefällt und die Fusspunkte A und R durch eine Gerade verbunden; ferner ist vom Punkt E das Lot m auf diese Verbindungsgerade AR gefällt. Nach dem Lotensatz (Zweite Mitteilung S. 4) besteht dann die folgende Gleichung

$$kp = qm.$$

Da nun ferner die Orientierungen

$$FEA, CDA, QDA, RDE, RAE, RES$$

deutig bestimmt. Das Lot m von E auf diese Gerade hat den Fusspunkt S . Man hat nun wie früher (nach dem Lotensatz) $kp = qm$, und da die Orientierungen

$$FEA, CDA, RDA, REA, RES$$

einander gleichwertig sind, ergibt sich durch Vergleichung der ersten und der letzten dieser Orientierungen, dass die Winkel FEA und RES Supplementwinkel sind, d. h. $\angle FEA$ ist stumpf.

Der Fall, wo R in D fällt, wird wie oben erledigt.

3) Der dritte Fall, wo $\angle ADC$ recht ist, wird unmittelbar durch den Lotensatz erledigt.

38. Aus den Sätzen von den dreieckswinkligen Vierecken folgen nun leicht die Dreieckssätze:

1°. Wenn die Winkelsumme in einem Dreieck mit ordinären Seiten und Winkeln kleiner als zwei Rechte ausfällt, dann ist sie in jedem Dreieck kleiner oder gleich zwei Rechten.

2°. Wenn die Winkelsumme in einem Dreieck mit ordinären Seiten und Winkeln grösser als zwei Rechte ausfällt, dann ist sie in jedem Dreieck grösser oder gleich zwei Rechten.

3°. Wenn die Winkelsumme in einem Dreieck mit ordinären Seiten und Winkeln zwei Rechte ist, dann ist sie in jedem Dreieck gleich zwei Rechten.

Die Sätze werden mit Hilfe der bekannten nebenstehenden Figur bewiesen: Im Dreieck ABC sind M

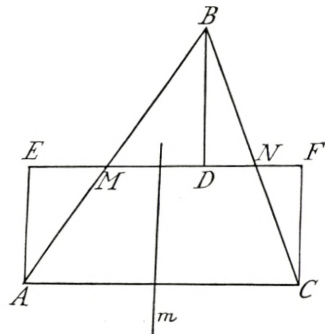


Fig. 13.

und N die Mittelpunkte der Seiten AB und BC . Da die Strecke AC ordinär ist, muss auch die Strecke MN ordinär sein,

und da die Strecken BM und BN und der Winkel MBN ordinär sind, gilt dies auch für das Lot BD und die mit diesem kongruenten Strecken AE , CF . Die Mittelsenkrechte m der Seite AC teilt das Viereck $AEFC$ in zwei kongruente ordinäre dreieckige Vierecke mit einem Winkel (bei A und bei C), welcher die Hälfte der Winkelsumme unseres Dreiecks ABC ausmacht.

Dieser Winkel ist dann spitz, stumpf oder recht, je nachdem die Winkelsumme des Dreiecks kleiner als, grösser als oder gleich zwei Rechten ist.

39. In dem Falle, wo die Winkelsumme zwei Rechte überschreitet, kann man von einem Exzess sprechen. Dieser Exzess ist jedoch immer in dem Sinne klein, dass man durch Vervielfältigung niemals einen rechten Winkel überschreiten kann. Dies folgt aus nebenstehender Figur (Fig. 14), wo aus dem Dreieck ABC , mittels Umwendungen um die Mittelpunkte der Seiten BC , CA , AB , drei andere Dreiecke BCA_1 , CAB_1 , ABC_1 abgeleitet sind.

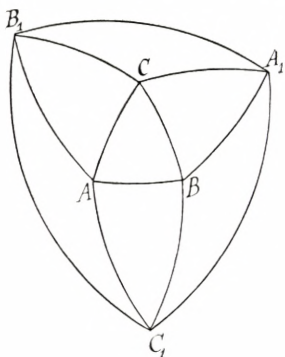


Fig. 14.

Es entsteht hierdurch ein Dreieck $A_1B_1C_1$, dessen Exzess — nach wohlbekanntem Überlegungen — wenigstens viermal so gross wie der Exzess ε des ursprünglichen Dreiecks ist. Durch Wiederholung dieses Prozesses könnte man sonach Dreiecke herstellen, deren Exzesse jedes vorgegebene Vielfache von ε überschreiten. Hieraus folgt aber gerade, dass ε im obengenannten Sinn klein

ist. Es folgt aber hieraus nicht dass der Winkel ε ein singulärer Winkel ist.

In allen Fällen, in denen das Eindeutigkeitsaxiom nicht

allgemeine Gültigkeit hat, gibt es Dreiecke, deren Winkelsumme zwei Rechte ausmacht (Erste Mitteilung S. 18), und es lässt sich um jeden Punkt der Ebene ein Bereich angeben, innerhalb dessen jedes Dreieck die Winkelsumme zwei Rechte hat.

IV. Singuläre Dreiecke.

40. Ein Dreieck mit ordinären Seiten und ordinären Winkeln soll ordinär heissen. Von ordinären Dreiecken ist leicht zu beweisen, dass jede Seite kleiner als die Summe der beiden anderen ist, und dass die Euklidischen Sätze über Ungleichheiten zwischen Seiten und ihren gegenüberliegenden Winkeln Gültigkeit haben. Für singuläre Dreiecke, d. h. Dreiecke, wo singuläre Seiten oder Winkel vorkommen, treten aber Modifikationen ein. Wir wollen hier die folgenden Sätze über singuläre Dreiecke hervorheben:

1°. In einem rechtwinkligen Dreieck ABC mit einem singulären Winkel A , dessen Scheitelele-

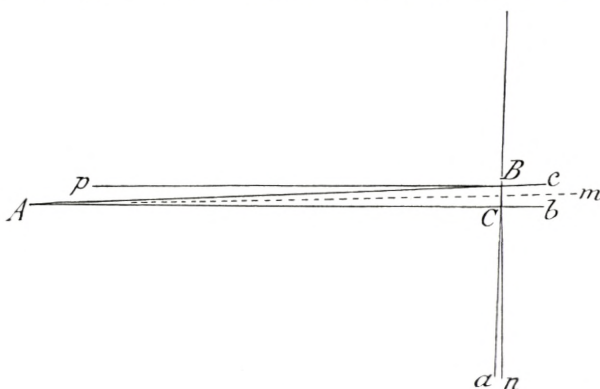


Fig. 15.

ment grösser als die gegenüberliegende Kathete BC ist, ist die Hypotenuse AB gleich der anderen Kathete AC (Fig. 15).

Die Seitengeraden sind a , b , c . In B errichten wir das Lot n auf c . Die Geraden n und a sind dann Nachbargeraden, und es lässt sich zeigen, dass sie wenigstens die Strecke BC gemein haben. Wenn wir nämlich in B das Lot p auf a errichten, wissen wir, dass p und c ein Element gemein haben, welches dem Scheitelement des Winkels A kongruent ist oder noch grösser als dieses Element ausfällt¹⁾, und da $pc = an$ ist, folgt hieraus, dass auch a und n ein ebenso grosses Element gemein haben. Da ferner BC kleiner als dieses Element ist, folgt hieraus, dass a und n jedenfalls die Strecke BC gemein haben.

Es sei nun m die Spiegelungsachse des Winkels A . Füllen wir von C das Lot n_1 auf m , so fällt n_1 zwischen a und n , und n_1 muss dann mit a jedenfalls ein ebenso grosses Element gemein haben, wie n und a , d. h. die Strecke BC ist in n_1 enthalten; n_1 muss dann c in B treffen, d. h. die Strecken AB und AC sind gleich.

2°. In einem Dreieck ABC mit zwei singulären spitzen Winkeln A und B , wo die Höhe CH von der dritten Ecke C auf die Grundlinie AB kleiner ist als die Scheitelemente der beiden Winkel A und B , ist die Grundlinie AB gleich der Summe der beiden anderen Seiten.

Der Satz ergibt sich durch Anwendung des vorigen Satzes auf die beiden rechtwinkligen Dreiecke AHC und BHC .

3°. In einem Dreieck mit einem singulären Winkel A ist die Winkelsumme fast zwei Rechte, d. h. die Abweichung von zwei Rechten ist ein singulärer Winkel.

¹⁾ Erste Mitteilung, S. 18: Jede Strecke des Elementes A ist einer Strecke kongruent, welche den beiden Geraden p, c gemeinsam ist.

Ist der Winkel A rund, so ist die Sache klar. Ist der Winkel spitz, so gilt folgendes (Fig. 16): Eine Schiebung längs der Seitengerade a , bei der B in C übergeht, ist eine singuläre Bewegung, die die Seitengerade c in eine Nachbargerade c' überführt, und da c Nachbargerade von b ist, ist auch c' Nachbargerade von b , d. h. c' und b bilden bei C einen singulären spitzen Winkel. Dieser Winkel ist aber $2R - (B + C)$, woraus der Satz folgt.

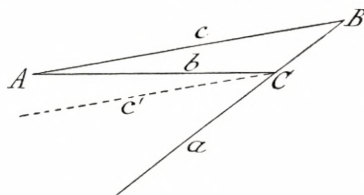


Fig. 16.

Bemerkung. Die Gerade c' muss ausserhalb des Dreiecks liegen; die Summe der beiden Dreieckswinkel B, C ist immer kleiner als $2R$. Der Satz vom Aussenwinkel ist allgemein gültig in unserer Geometrie.

V. Die Spiegelungsachse zweier Geraden.

41. Unseren Voraussetzungen zufolge kann durch Bewegung jede Gerade in jede andere Gerade übergeführt werden. Es folgt dann nach Axiom V, dass zwei Geraden mit einem gemeinsamen Punkt zwei zueinander senkrechte Spiegelungsachsen durch diesen Punkt haben.

Haben zwei Geraden a, b unendlich viele Punkte gemein, und ist P einer dieser Punkte, so bilden die beiden Geraden zwei singuläre spitze Scheitelwinkel mit dem gemeinsamen Scheitel P . Es gibt dann (Axiom V) eine Spiegelungsachse p durch P , derart dass jeder der beiden Winkel bei der Spiegelung p in sich selbst übergeht, und eine andere Spiegelungsachse q , derart dass die beiden singulären Winkel bei der Spiegelung q miteinander vertauscht werden. Die Spiegelungsachse p enthält das ganze

Scheitelelement (Schnittelement) der Geraden a , b . Die Spiegelungsachse q ist senkrecht zu p , und wenn P das ganze Scheitelelement durchläuft, durchläuft q eine ganze Reihe von Spiegelungsachsen $\perp p$.

42. Haben zwei Geraden a , b keinen Punkt gemein, so haben sie eine und nur eine Spiegelungsachse. Das

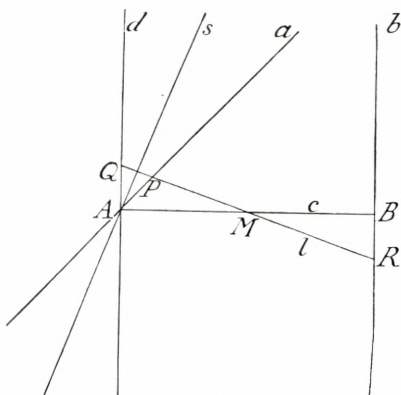


Fig. 17.

lässt sich folgendermassen einsehen.

Von einem auf a gelegenen Punkt A (Fig. 17) fallen wir das Lot c auf b ; der Fusspunkt sei B . Durch A ziehen wir eine Gerade $d \perp c$. Der Mittelpunkt der Strecke AB sei M . Diejenige Spiegelungsachse der beiden Geraden a , d , welche nicht in demjenigen Winkel (ad) gelegen ist, der M enthält, sei s . Wir legen nun durch M eine Gerade $l \perp s$. Diese Gerade l muss a schneiden, nämlich in einem Punkt P desjenigen Halbstrahls von a , welcher einen spitzen Winkel mit dem von A ausgehenden und durch B hindurchgehenden Halbstrahl von c bildet. Sie muss sodann auch die Gerade d schneiden, nämlich im Spiegelbild Q von P an s . Schliesslich schneidet dann die Gerade l auch die Gerade b , nämlich in dem aus Q durch Umwendung um M hervorgehenden Punkt R .

Die Bewegung sMR muss nun a in b überführen. Diese Bewegung ist aber eine Spiegelung s' , deren Achse senkrecht auf l steht. Hiermit haben wir eine Spiegelungsachse konstruiert. Dass zwei Spiegelungsachsen in diesem Falle nicht vorhanden sein können, ergibt sich daraus, dass ein

Punkt A von a bei den Spiegelungen an zwei Achsen in zwei Punkte B, C übergehen müsste, sodass ein Dreieck entstehen würde, bei dem ein Aussenwinkel kleiner als ein Innenwinkel ist, was nicht vorkommen kann.

VI. Die Gross-Geometrie.

43. Wir konstruieren nun eine neue Geometrie, deren »Punkte« die Punkt-Nachbargebiete und deren »Geraden« die Geraden-Nachbargebiete der ursprünglichen Geometrie sind. Die neue Geometrie soll als »Gross-Geometrie« und ihre Punkte und Geraden als »Grosspunkte« und »Grossgeraden« bezeichnet werden.

Es gilt dann der Satz:

Jeweil zwei Grosspunkte (A) , (B) bestimmen eindeutig eine Grossgerade (g) , welche die beiden Grosspunkte enthält.

Wir können in dieser Gross-Geometrie ganz wie in der alten Geometrie von Bewegungen sprechen. Jede Grossgerade (g) bestimmt eine Spiegelung, welche jedem Grosspunkt (A) einen Grosspunkt (A') zuordnet, wobei A' und A einander bei der Spiegelung g entsprechen. Jede Bewegung in der Gross-Geometrie entsteht so durch Zusammenfassung einer Menge von Bewegungen der ursprünglichen Geometrie, bei denen Nachbarpunkte in Nachbarpunkte übergehen.

44. Es zeigt sich so, dass alle Axiome der ursprünglichen Geometrie auch in der Grossgeometrie gelten. Und da zudem das Eindeutigkeitsaxiom in der Grossgeometrie erfüllt ist, gilt hier also die ganze projektive Geometrie. Ebenso können alle Begriffsbildungen aus der Kongruenzlehre auf Grund des Eindeutigkeitsaxioms: Idealpunkte, Halbdrehungen, Idealgeraden u. s. w. übernommen werden.

Der Pascalsche Satz gilt dann in der folgenden Form: Wenn drei verschiedene Grosspunkte \textcircled{A} , \textcircled{B} , \textcircled{C} auf einer Grossgeraden \textcircled{g} und drei andere von einander verschiedene Grosspunkte $\textcircled{A_1}$, $\textcircled{B_1}$, $\textcircled{C_1}$ auf einer anderen Grossgeraden $\textcircled{g_1}$ liegen, und wenn ferner die Schnittpunkte der drei Grossgeradenpaare $\textcircled{AB_1}$, $\textcircled{A_1B}$; $\textcircled{AC_1}$, $\textcircled{A_1C}$; $\textcircled{BC_1}$, $\textcircled{B_1C}$ eigentlich sind, so liegen diese Punkte auf einer Grossgeraden.

Die ganze Figur lässt sich durch eine entsprechende Figur in der ursprünglichen Ebene fixieren, in der keine zwei Punkte Nachbarpunkte und keine zwei Geraden Nachbargeraden sind.

45. Es bleibt nun noch zu untersuchen, wie sich die Beziehungen zwischen Nachbarfiguren, im besonderen Nachbarpunkten und Nachbargeraden gestalten.

Als Erweiterung einer singulären Strecke a soll der gemeinsame Punktbereich aller Geraden, welche die Strecke enthalten, bezeichnet werden.

Zwei singuläre Strecken a , b sollen ordnungsgleich genannt werden, wenn ihre Erweiterungen $E(a)$, $E(b)$ kongruent sind.

Wenn $E(a)$ grösser als $E(b)$ ausfällt, d. h. wenn $E(a)$ einen zu $E(b)$ kongruenten echten Teil enthält, sagen wir, dass a ordnungsgrösser als b (b ordnungskleiner als a) ist.

Für zwei singuläre Winkel können ähnliche Definitionen aufgestellt werden, indem man die Erweiterung eines singulären Winkels (a, b) , als das Büschel (das Büschelement) sämtlicher Geraden definiert, welche das Scheittelelement des gegebenen Winkels enthalten.

46. Den Winkel zweier Geraden a , b mit dem Schnittelelement E messen wir durch eine Strecke, welche als Ordinate des Winkels bezeichnet werde. Diese Strecke

ist das in einem Punkt P der einen Geraden a errichtete Lot PQ bis zum Schnitt Q mit der anderen; hierbei wird der Punkt P so gewählt, dass sein Abstand von einem Punkt A des Schnittelements E ordinär ist. Wir wollen diesen Abstand AP konstant wählen und ihn als Einheit (Eichmass) bezeichnen. Es ist zu beachten, dass die Wahl von A innerhalb E ohne Belang ist; wir wissen nämlich, dass eine singuläre Veränderung des Fusspunktes P innerhalb des Grössenbereichs E keinen Einfluss auf die Grösse von PQ hat, da zwei verschiedene Strecken PQ , die vorkommen können, immer Gegenseiten eines Rechtecks sind (Erste Mitteilung S. 18). Die Ordinate eines bestimmten singulären Winkels ist also nach Wahl der Einheit wirklich eindeutig bestimmt.

Singuläre Winkel werden hiernach wesentlich durch ihr Scheitelement (Schnittelement) und ihre Ordinate gekennzeichnet.

47. Der einfachste Fall unserer Geometrie ist nun derjenige, wo alle singulären Strecken ordnungsgleich sind, wo also alle Schnittelemente kongruent sind. Mit diesem Falle wollen wir anfangen.

Die Ordinate eines singulären Winkels ist dann der Grösse nach im Scheitelement des Winkels enthalten, und hieraus folgt, dass ein Rechteck mit zwei beliebig gegebenen singulären Dimensionen konstruiert werden kann. Allgemeiner:

Die Geometrie im Nachbargebiet eines Punktes A , also die Geometrie innerhalb des Grosspunktes \textcircled{A} , ist die Euklidische Geometrie.

Jede Gerade, welche in das Gebiet \textcircled{A} eindringt, hat eine Punktmenge mit dem Gebiet gemein. Diese Punktmenge soll als Gerade im Gebiet bezeichnet werden. Durch zwei

Punkte A, B des Gebiets gehen unendlich viele Geraden der ursprünglichen Ebene; da aber alle Schnittelemente kongruent sind, haben alle diese Geraden dasselbe Element gemein, d. h. innerhalb des Gebiets bestimmen zwei Punkte eine und nur eine Gerade. Da zudem Rechtecke (mit beliebigen Dimensionen) innerhalb des Gebiets existieren, ist die ganze Geometrie des Gebiets die Euklidische.

48. Ein Kreis mit ordinärem Halbmesser OA hat in A eine auf OA senkrechte Tangente t . Alle Nachbarpunkte des Punktes A auf der Tangente liegen auch auf dem Kreis. Es folgt dies unmittelbar aus dem Satz 39,1°. Also:

Ein Kreis mit ordinärem Halbmesser hat mit jeder Tangente ein Element gemein.

Zwei einander berührende Kreise mit ordinären Halbmessern haben ebenfalls ein Element gemein.

Hingegen hat jeder Kreis mit singulärem Halbmesser nur einen Punkt mit der Tangente gemein, und dies gilt auch für zwei einander berührende Kreise, von welchen

wenigstens einer singulären Halbmesser hat.

49. Es ergibt sich nun leicht folgendes:

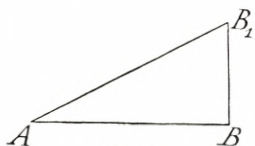


Fig. 18.

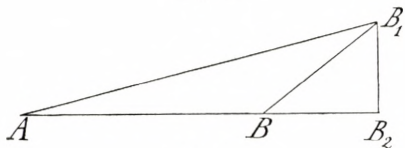


Fig. 19.

1°. Ist AB eine ordinäre Strecke (Fig. 18) und $BB_1 \perp AB$ eine singuläre Strecke, so gilt $AB = AB_1$.

2°. Ist AB eine ordinäre Strecke (Fig. 19), BB_1 eine beliebige singuläre Strecke und BB_2 deren Projektion auf AB , so gilt

$$AB_1 - AB = BB_2,$$

mit Vorzeichen in der Richtung AB gerechnet.

3°. Ist AB eine ordinäre Strecke (Fig. 18) und BB_1 eine singuläre Strecke, welche »fast senkrecht« auf AB ist, d. h. einen singulären Winkel mit der Normalen von AB in B bildet, so gilt $AB_1 = AB$ (folgt aus 1°).

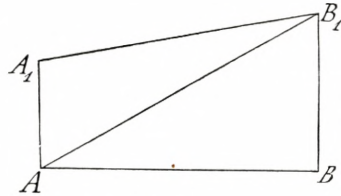


Fig. 20.

4°. Sind AB eine ordinäre Strecke (Fig. 20), AA_1 und BB_1 zwei singuläre Strecken, welche senkrecht zu AB (oder fast senkrecht zu AB) sind, so gilt $AB = A_1B_1$. Man hat nämlich nach 3°:

$$AB = AB_1 = A_1B_1.$$

Eine ordinäre Strecke ist also immer gleich ihrer Projektion auf eine Nachbargerade.

5°. Ist AB eine ordinäre Strecke (Fig. 21), A_1B_1 eine beliebige Nachbarstrecke und A_2B_2 ihre Projektion auf AB , so gilt $A_1B_1 = A_2B_2$ und

$$A_1B_1 - AB = A_2B_2 - AB = BB_2 - AA_2.$$

Fällt man die Lote AA_3 und BB_3 auf A_1B_1 , so hat man auch $AB = A_3B_3$. Die beiden Vierecke $AA_2A_1A_3$ und $BB_2B_1B_3$ sind Rechtecke.

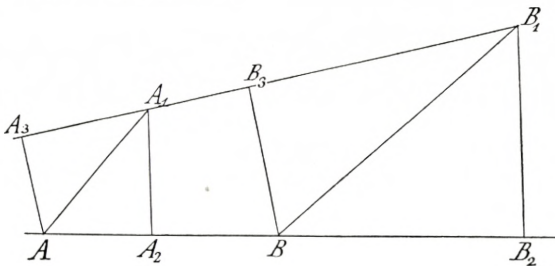


Fig. 21.

50. Wir gehen nun daran, die Abstände zwischen Nachbarpunkten zweier Nachbargeraden zu untersuchen.

Zwei Nachbargeraden g, g_1 (Fig. 22) mögen von dem gemeinsamen Punkt A ausgehen. Die Punkte B und C seien so auf g gewählt, dass die Strecken AB und AC ordinär sind. Wir errichten die Lote BB_1 und CC_1 auf g_1 , welche g_1 in B_1 und C_1 schneiden.

Wir wollen die beiden Strecken BB_1 und CC_1 vergleichen. Sie sind singularär und können deshalb durch Bewegung in

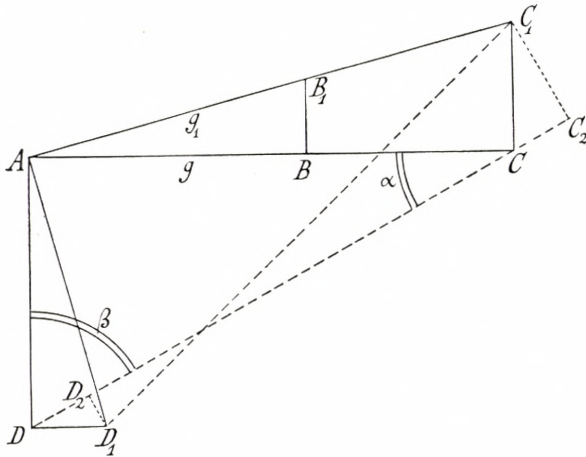


Fig. 22.

ein Nachbargebiet eines beliebigen Punktes übergeführt und sodann als Strecken einer Euklidischen Ebene angesehen werden. Infolgedessen kann man von dem Verhältnis $\frac{BB_1}{CC_1}$ reden.

Wir drehen nun das Dreieck ABB_1 um einen rechten Winkel; es gehe dabei in ADD_1 über. Die beiden Figuren ACD und AC_1D_1 sind dann kongruent, und die Nachbarstrecken CD und C_1D_1 sind einander gleich. Projiziert man C_1D_1 auf CD in C_2D_2 , so hat man also

Wir drehen nun das Dreieck ABB_1 um einen rechten Winkel; es gehe dabei in ADD_1 über. Die beiden Figuren ACD und AC_1D_1 sind dann kongruent, und die Nachbarstrecken CD und C_1D_1 sind einander gleich. Projiziert man C_1D_1 auf CD in C_2D_2 , so hat man also

$$CC_2 = DD_2,$$

und hieraus folgt

$$CC_1 \sin \alpha = DD_1 \sin \beta$$

und wegen $DD_1 = BB_1$

$$\frac{BB_1}{CC_1} = \frac{\sin \alpha}{\sin \beta}.$$

Im ordinären Dreieck ACD können wir aber die gewöhnliche Trigonometrie anwenden, also z. B. den Sinussatz

$$\frac{\sin \alpha}{\sin \beta} = \frac{s(AD)}{s(AC)} = \frac{s(AB)}{s(AC)},$$

wo $s(AB)$ und $s(AC)$ als allgemeine Bezeichnungen für sinus der Grossseiten AB und AC stehen. (In der sphärischen Geometrie ist dies der gewöhnliche sinus, in der hyperbolischen Ebene der hyperbolische sinus, und in der Euklidischen Ebene ist $s(AB) = AB$, $s(AC) = AC$ zu setzen¹⁾).

Es folgt dann

$$\frac{BB_1}{CC_1} = \frac{s(AB)}{s(AC)}.$$

51. Liegen nun zwei beliebige Nachbargeraden g, g_1 vor, wo gemeinsame Punkte nicht vorausgesetzt werden,

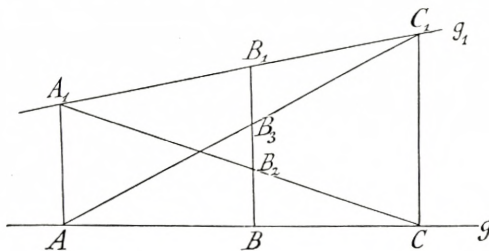


Fig. 23.

1) Die diesbezüglichen notwendigen Einzelheiten sind von F. SCHUR in seinen Grundlagen der Geometrie entwickelt. In einer weiteren Arbeit beabsichtige ich in anderer Weise auf hierher gehörige Fragen näher einzugehen.

so untersuchen wir die zwischen ihnen liegenden Ordinaten in folgender Weise. In drei beliebigen Punkten A, B, C von g (Fig. 23) mit ordinären Abständen errichten wir die Ordinaten $AA_1, BB_1, CC_1 \perp g$ und ziehen die Geraden A_1C und AC_1 , welche BB_1 in den Punkten B_2 und B_3 schneiden mögen.

Aus dem vorigen Satz folgt nun

$$BB_2 = AA_1 \cdot \frac{s(BC)}{s(AC)},$$

$$BB_3 \text{ oder } B_2B_1 = CC_1 \cdot \frac{s(AB)}{s(AC)},$$

also

$$BB_1 = BB_2 + B_2B_1 = AA_1 \cdot \frac{s(BC)}{s(AC)} + CC_1 \cdot \frac{s(AB)}{s(AC)},$$

oder (indem wir mit Vorzeichen rechnen)

$$AA_1 \cdot s(BC) + BB_1 \cdot s(CA) + CC_1 \cdot s(AB) = 0.$$

In dieser Formel können in den Abständen BC, CA, AB die Punkte A, B, C beliebig innerhalb ihrer Nachbargebiete variieren. Die Ordinaten AA_1, BB_1, CC_1 bleiben dabei ungeändert.

52. Benachbarte Dreiecke. Von zwei benachbarten Dreiecken ABC und $A_1B_1C_1$ lässt sich nun ganz allgemein aussagen, dass die Differenzen entsprechender Seiten und entsprechender Winkel dieselben Relationen erfüllen, wie die Differenziale der Seiten und Winkel des Grossdreiecks $\textcircled{A} \textcircled{B} \textcircled{C}$. Was die Seiten anbetrifft, geht dies aus den schon durchgeführten Erwägungen hervor, und für die Winkel brauchen wir nur auf die Figur (Fig. 24) hinzuweisen, wo z. B. der Winkel $B_1A_1C_1$ gleich dem Winkel (q, p) zwischen den von A auf A_1C_1 und A_1B_1 gefällten

Loten p und q ist; es folgt dies daraus, dass im Nachbargebiet \textcircled{A} von A die Euklidische Geometrie gültig ist.

Benachbarte Figuren $ABCD \dots$ und $A_1B_1C_1D_1 \dots$ lassen sich ebenso behandeln. Die zwischen ihnen bestehenden Relationen sind identisch mit den aus den Grundformeln der Grossfigur $\textcircled{A} \textcircled{B} \textcircled{C} \textcircled{D} \dots$ und deren Ableitungen (Differentialformeln) hervorgehenden.

53. Das Hauptresultat lässt sich dann folgendermassen formulieren:

In der vorgelegten Ebene denken wir uns eine Menge \mathfrak{M} von Punkten derart ausgewählt, dass keine zwei Punkte dieser Menge benachbart sind, und dass jeder Punkt der Ebene Nachbarpunkt eines Punktes von \mathfrak{M} ist.

Die Menge $\mathfrak{M} = (A, B, C, \dots)$ heisst die Grundmenge.

Irgend einer Figur der Ebene A_1, B_1, C_1, \dots entspricht dann eine bestimmte Figur A, B, C, \dots von \mathfrak{M} als Nachbarfigur.

Die Relationen der Figur $A_1B_1C_1 \dots$ bezüglich $ABC \dots$, lassen sich alle aus den Grundrelationen der Grossfigur $\textcircled{A} \textcircled{B} \textcircled{C} \dots$ und deren Ableitungen herstellen.

Die Auswahl der Grundmenge lässt sich in mannigfacher Weise vornehmen. Für jede Wahl gelten aber die obengenannten Beziehungen.

54. Dieses Hauptresultat lässt sich auch in folgender Weise ausdrücken:

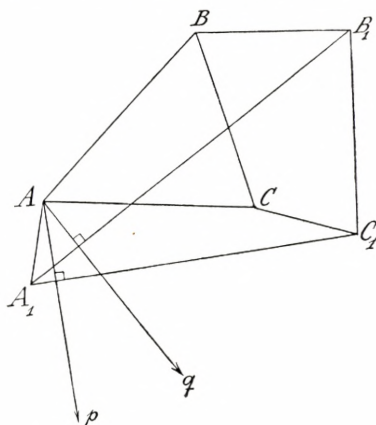


Fig. 24.

Wird in der Grossfigur $\textcircled{A}\textcircled{B}\textcircled{C}\dots$ ein Koordinatensystem eingeführt, so kann jeder Punkt der Grundfigur $ABC\dots$ durch die Koordinaten u, v des entsprechenden Punktes der Grossfigur fixiert werden; und jeder Punkt einer beliebigen Figur $A_1B_1C_1\dots$ unserer Ebene lässt sich dann durch Koordinaten der Form $u + \varepsilon u_1, v + \varepsilon v_1$ mit $\varepsilon^2 = 0$ fixieren.

Hiermit können wir den einfachsten Fall unserer Geometrie, wo alle singulären Strecken ordnungsgleich sind, als erledigt betrachten.

VII. Der allgemeine Fall.

55. Wir gehen nun daran, einige Betrachtungen über den allgemeinen Fall, wo nicht-kongruente Schnittelemente existieren, anzustellen.

Jeder singulären Strecke a entspricht ein Element $E(a)$, die Erweiterung von a mit der Eigenschaft, dass alle Geraden, welche a enthalten, auch $E(a)$ enthalten. Wir wissen, dass $E(a)$ alle Vielfachen von a enthalten muss, oder anders ausgedrückt, dass $E(a)$ wenigstens den »Eudoxischen Bereich« von a enthält; ob $E(a)$ umfassender als dieser Bereich sein kann, bleibe dahingestellt. Im folgenden setzen wir voraus, dass $E(a)$ stets mit dem Eudoxischen Bereich von a identisch ist.

In ähnlicher Weise entspricht jedem singulären Winkel α ein Büschelement $E(\alpha)$, dessen Umfang durch den Eudoxischen Bereich des Winkels gegeben sein soll.

Ebenso wie jeder Punkt eines Schnittelements als Mittelpunkt des Elements betrachtet werden kann, kann auch jede Gerade eines Büschelements als Spiegelungs-

aus, dass das Geradenbüschel (c'), welches die singuläre Strecke PQ enthält, demjenigen Strahlenbüschel $P(C)$ kongruent ist, welches das gemeinsame Element der Geraden n, n_1 vom Punkt P aus projiziert. Dieses Element ist aber dem gemeinsamen Element der Geraden a, b kongruent.

Wir ziehen nun das ganze durch die beiden Geraden a, b bestimmte Büschel (a, b) in Betracht. Sein Schnittelement sei mit $S(a, b)$ bezeichnet. Ferner betrachten wir den Punktbereich $\Sigma(P, Q)$, welcher als Durchschnitt des Büschels (a, b) mit c entsteht; wir bezeichnen ihn als Ordinatenbereich des Büschels, indem wir die Strecke AQ gleich der festen Einheit für Ordinatenmessungen gewählt denken. Ist nun $\Sigma(P, Q) < S(a, b)$, also die Voraussetzung in 17, dass die Strecke PQ kleiner als $S(a, b)$ ist, erfüllt, so hat man $AP = AQ$, und es geht dann aus der obenstehenden Untersuchung hervor, dass die beiden Elemente $S(a, b)$ und $\Sigma(P, Q)$ in der Weise reziprok verbunden sind, dass nicht nur, wie gegeben, $S(a, b)$ Schnittelement eines Büschels mit dem Ordinatenelement $\Sigma(P, Q)$ ist, sondern auch umgekehrt $\Sigma(P, Q)$ Schnittelement eines Büschels ist, dessen Ordinatenelement mit $S(a, b)$ kongruent ist.

In dieser Weise treten die Schnittelemente verschiedener Ordnung paarweise als reziproke Elemente auf. Bei dem Beweis wurde allerdings die Voraussetzung gemacht, dass die Strecke PQ kleiner als $S(a, b)$ ist. Es ist aber leicht einzusehen, dass unser Resultat auch in dem Falle, wo diese Bedingung nicht erfüllt ist, gültig bleibt. Ist nämlich $PQ > S(a, b)$, so lässt sich allerdings nicht mehr schliessen, dass $AP = AQ$ ist; ist aber $AP > AQ$, so hat man jedenfalls die Streckenungleichung

$$AP - AQ < PQ,$$

und hieraus folgt, dass Ordinatenmessungen im Büschel mit dem Schnittlelement $\Sigma(P, Q)$ unabhängig davon sind, ob man AQ oder AP als Einheit benützt.

Das grösste Schnittlelement erhält man, wenn man auf einer Geraden alle Nachbarpunkte eines Punktes zusammenfasst. Das hierzu reziproke ist dann das kleinst mögliche Schnittlelement.

Aus 39,1° geht hervor, dass jeder Kreis mit ordinärem Halbmesser mit seiner Tangente wenigstens ein kleinstes Schnittlelement gemein hat.

57. Aus den Schnittlelementen leiten wir nun »ebene Elemente« in folgender Weise ab. Auf einer Geraden sei ein Schnittlelement Σ gegeben. Ein beliebiger Punkt O von Σ kann als Mittelpunkt von Σ betrachtet werden. Wir drehen nun Σ in der Ebene um O . Der hierbei beschriebene Bereich der Ebene soll als ebenes Element von derselben Ordnung wie Σ bezeichnet werden. Es lässt sich sofort einsehen, dass dieser Bereich konvex ist, und dass jede Gerade, welche irgend einen Punkt des Bereichs enthält, aus dem ganzen Bereich ein Schnittlelement herauschneidet, welches dem erzeugenden Schnittlelement Σ kongruent ist.

Ist Σ Schnittlelement zweier Geraden a, b , so ist der von ihm erzeugte ebene Bereich mit der Fixpunktmenge der Bewegung ab identisch.

Reziproken Schnittlelementen entsprechen reziproke ebene Elemente.

Das grösste ebene Element ist der Bereich aller Nachbarpunkte eines Punktes, also genau das, was wir früher als Grosspunkt bezeichnet haben. Das kleinste ebene Element ist das hierzu reziproke Element. Das nächstgrösste und das nächstkleinste Element sind reziprok, u. s. w.

Ebene Elemente verschiedener Ordnung sollen auch als Grosspunkte verschiedener Ordnung bezeichnet werden.

58. In der inneren Geometrie eines jeden Elements sind alle ursprünglich aufgestellten Axiome gültig, wenn wir nur unter »Gerade« den Durchschnitt unseres Elements mit einer ursprünglichen Geraden verstehen.

In der Geometrie des kleinsten Elements gilt das Eindeutigkeitsaxiom; da ferner Rechtecke existieren, ist diese Geometrie die Euklidische.

In der Geometrie des grössten Elements (des Elements erster Ordnung), gilt das Eindeutigkeitsaxiom im allgemeinen nicht. Man kann aber hier wie in der ursprünglichen Geometrie »Grosspunkte« einführen, welche dann als »Grosspunkte (Elemente) zweiter Ordnung« zu bezeichnen sind, und welche natürlich mit den ebenen Elementen zweiter Ordnung der ursprünglichen Geometrie identisch sind. Man kommt so zu Untergeometrien erster, zweiter, dritter, . . . , n 'ter Ordnung. Man darf aber nicht erwarten, dass diese Reihe von Untergeometrien endlich sein muss. Es lassen sich auf vielfache Weise Beispiele aufstellen, wo die Geometrie eine unendliche Menge von solchen Elementen umfasst.

59. Als einfaches Beispiel einer Geometrie mit einer endlichen Anzahl von Elementen nennen wir eine Koordinatengeometrie, in der die Koordinaten Polynome der Form

$$a = a_0 + a_1 t + a_2 t^2 + \dots + a_n t^n$$

sind, wo die Koeffizienten $a_0, a_1, a_2, \dots, a_n$ reelle Zahlen bedeuten und t eine Unbestimmte ist, die der Relation

$t^{n+1} = 0$ genügt, und mit der man im übrigen in gewöhnlicher Weise rechnet.

Das Vorzeichen von a soll als das Vorzeichen des ersten von Null verschiedenen Koeffizienten definiert werden; dass $a > b$ ist, soll bedeuten, dass $a - b$ positiv ist; die gewöhnlichen Rechnungen mit Ungleichungen werden dann nur in der Weise modifiziert, dass aus $a > 0$, $b > 0$, nur $ab \geq 0$ folgt. Nullteiler sind alle diejenigen Polynome a , deren Grundkoeffizient a_0 gleich Null ist; ist

$$a = a_r t^r + a_{r+1} t^{r+1} + \dots + a_n t^n, \text{ mit } a_r \neq 0,$$

so soll a Nullteiler von r 'ter Ordnung heissen.

Wir beschränken uns auf die »Euklidische« Ebene, wo die Geraden in gewöhnlicher analytischer Darstellung durch lineare Gleichungen in den laufenden Koordinaten x , y und die Bewegungen durch orthogonale Substitutionen gegeben sind.

Zwei einander schneidende Nachbargeraden können dann immer durch Gleichungen

$$y = 0, y = ax,$$

wo a ein Nullteiler ist, dargestellt werden. Ihr Winkel ist durch die Ordinate $y = a$ (indem $x = 1$ gesetzt wird) bestimmt. Ihr Schnittelement ist durch alle solche x gegeben, für welche $ax = 0$ ist. Es geht hieraus hervor, dass reziproke Elemente auf der x -Achse durch die Nullteiler der Ordnungen r und $n + 1 - r$ bestimmt werden, d. h. einerseits alle Punkte, deren Abszissen Nullteiler von der Ordnung r , und andererseits alle Punkte, deren Abszissen Nullteiler von der Ordnung $n + 1 - r$ sind, bilden zwei reziproke Elemente. Die rezi-

proken Elemente in der Ebene sind hiermit auch charakterisiert.

Die hier vorgelegte Geometrie, auf deren Einzelheiten wir bei dieser Gelegenheit nicht näher eingehen, ist, wie sofort ersichtlich, mit der sogenannten Geometrie der Kurvenelemente n 'ter Ordnung gleichbedeutend.

Durch Einführung von Polynomen mit zwei oder mehreren Unbestimmten (oder von anderen algebraischen Ringen mit Nullteilern) können in ähnlicher Weise andere Beispiele von Geometrien mit endlichvielen Elementen aufgestellt werden.

60. Bei den vorstehenden Entwicklungen haben wir die Voraussetzung gemacht, dass in unserer Ebene Punktepaare mit eindeutig bestimmter Verbindungsgerade existieren. Zum Schluss wollen wir aber auch mit wenigen Worten den Fall besprechen, wo es kein Punktepaar mit eindeutiger Verbindungsgerade gibt. Man kann dann in folgender Weise vorgehen. Wir wählen eine Strecke AB als Grundstrecke und betrachten zunächst einen ebenen Bereich $E(AB)$, den Eudoxischen Bereich der Grundstrecke, welcher aus allen denjenigen Punkten besteht, die durch wiederholte Abtragung der Grundstrecke vom Punkte A aus erreicht werden können. Für diesen Bereich gilt dann, dass die Punkte A, B eine eindeutige Verbindungsgerade haben, und die Geometrie des Bereichs lässt sich dann mit den vorstehend entwickelten Hilfsmitteln behandeln.