FU JEN STUDIES

NATURAL SCIENCES

NO. 9 1975

FU JEN UNIVERSITY
TAIPEI, TAIWAN, REPUBLIC OF CHINA

FU JEN STUDIES

is published annually by the College of Natural Sciences of Fu Jen University

All correspondence regarding contributions, subscriptions and exchanges should be addressed to·

Dr. H. Hesselfeld, SVD
College of Natural Sciences
Fu Jen University
242 Hsin Chuang, Taipei,
Taiwan, R. O. C.

關于所有投稿、訂閱或交換之函件請寄：
臺北縣新莊鎮輔仁大學　理學院院長　郝思漢

Price:　US$ 2.00

FU JEN STUDIES

NATURAL SCIENCES

NO. 9 1975

CONTENTS

FU JEN UNIVERSITY

TAIPEI, TAIWAN, REPUBLIC OF CHINA

A CLASSIFICATION OF PROPER DIFFERENTIABLE MAPS

CHI-LING HSU

I. INTRODUCTION

If X and Y are separable metric spaces and $f: X \to Y$ is a map, the *branch set* B_f is the set of points at which f fails to be a local homeomorphism. The map is *proper* if for each compact set $W \subset Y$, $f^{-1}(W)$ is compact. f is *monotone* if for each $y \in Y$, $f^{-1}(y)$ is connected (it may be empty).

The author and Timourian[1] have classified proper differentiable maps locally about their branch sets in the case that they are (smooth) manifolds. We are now interested in investigating the behavior of these maps when their branch sets are not manifolds. In this note, I give examples of real analytic proper maps with branch sets $S^1 \vee S^1$, one point union of two 1-spheres, and give information about the local structure of proper differentiable maps related to the examples.

II. EXAMPLES

There are real analytic maps from torus and Klein bottle to 2-spheres which collapse $S^1 \vee S^1$ to a point.

(1) Let (x_1, x_2, x_3, x_4) be coordiantes of x in R^4, and

$$T = \{x \in R^4 \mid (x_1 - 1)^2 + x_2^2 = 1 \text{ and } (x_3 - 1)^2 + x_4^2 = 1\}.$$

Let

$$S^2 = \{(y_1, y_2, y_3) \in R^3 \mid (y_1 - 1)^2 + y_2^2 + y_3^2 = 1\}.$$

Define a map $f_1: T \to S^2$ by

$$f_1(x_1, x_2, x_3, x_4) = \frac{(x_1 x_3,\ x_1 x_4,\ x_2 x_3)}{\sqrt{(x_1 x_3 - 1)^2 + (x_1 x_4)^2 + (x_2 x_3)^2}}.$$

From reference (1, §5) we know that f_1 is a proper monotone real analytic map which maps the set

$$B = \{(x_1, x_2, x_3, x_4) \in T \mid x_1 = x_2 = 0 \text{ or } x_3 = x_4 = 0\}$$

a $S^1 \mathrm{V} S^1$, to a point $(0, 0, 0)$ and $f_1 | \mathrm{T} - \mathrm{B}$ is a homeomorphism onto $S^2 - (0, 0, 0)$.

(2) Let T and B be as above. Define a map $\alpha : \mathrm{T} \to \mathrm{R}^5$ by

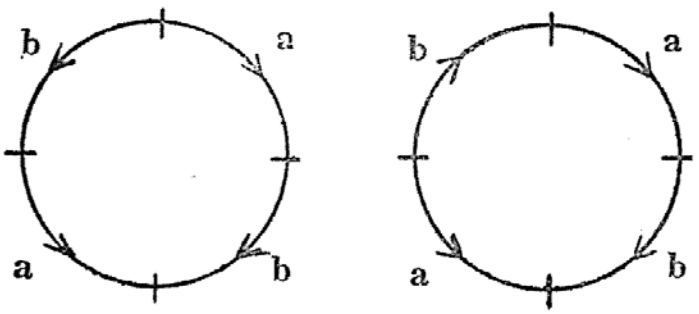
$$\alpha(x_1, x_2, x_3, x_4) = (x_1 - 1, x_2(x_3 - 1), x_2 x_4, (x_3 - 1) x_4, (x_3 - 1)^2 - x_4^2) .$$

α is a real analytic map and identifies points (x_1, x_2, x_3, x_4) and $(x_1, -x_2, 2-x_3, -x_4)$ in T. Therefore α is a $2-1$ covering map from T onto $\alpha(\mathrm{T})$. This $\alpha(\mathrm{T})$ is a Klein bottle K with a real analytic structure embedded in R^5. Let $f_2 : \mathrm{K} \to S^2$ be a map defined by

$$f_2(x_1 - 1, x_2(x_3 - 1), x_2 x_4, (x_3 - 1) x_4, (x_3 - 1)^2 - x_4^2) = (x_1 x_4^2, x_2 x_4, x_1(x_3 - 1), x_4) .$$

It is easy to see that f_2 is a well-defined map on K and f_2 is real analytic. f_2 maps the set $\alpha(\mathrm{B})$, also $S^1 \mathrm{V} S^1$, to the point $(0, 0, 0)$. For each $y \in S^2 - (0, 0, 0)$. $f_2^{-1}(y)$ is a single point. Hence f_2 is a proper monotone real analytic map which collapses the branch set $\alpha(\mathrm{B})$ to a point, and $f_2 | \mathrm{K} - \alpha(\mathrm{B})$ is a homeomorphism onto $S^2 - (0, 0, 0)$.

(3) Let M_1, M_2 be the 2-manifolds obtained from T and K, respectively, by cutting out a closed 2-disk, away from B and $\alpha(\mathrm{B})$, respectively. Then M_i $(i = 1, 2)$ is the mapping cylinder of the map $h_i : S^1 \to S^1 \mathrm{V} S^1$, $(i = 1, 2)$, indicated as



(1) (2)

where the areas marked a, b are mapped onto the circles a and b of $S^1 \mathrm{V} S^1$. Let g_i $(i = 1, 2)$ be the map $f_i | \mathrm{M}_i$ $(i = 1, 2)$ followed by the sterographic projection from $S^2 - (0, 0, 0)$ onto R^2. Then

$g_i : M_i \to R^2$ is the natural map from M_i to its quotient space obtained by identifying the images of h_i to a point, for each $i = 1, 2$.

(4) Let $r : R^2 \to R^2$ be defined by $r(z) = z^m$, where we are considering R^2 to be the space of complex numbers and m is an integer bigger than 1. Then, for each i $(i = 1, 2)$ $r \circ g_i : M_i \to R^2$ is also a proper real analytic map with branch set $S^1 \vee S^1$ and away from the branch set it is a $m - 1$ covering map.

III. THEOREM

If $f : M^2 \to N^2$ be a proper differentiable map with $f(B_f)$ discrete and $f^{-1}(y)$ a (smooth) figure eight in M^2, for each $y \in f(B_f)$. Then for each component B of B_f there exist neighborhoods U of B and V of $f(B)$ such that $f|U : U \to V$ is topologically equivalent to one of the maps in II (3) or (4).

IV. PROOF

For each component B of B_f, $f(B)$ is a single point y. Choose neighborhoods V of y and U of B such that V is a coordinate neighborhood, U is the component of $f^{-1}(V)$ containing B and $U \cap B_f = B$. Let $h \circ g$ be a factorization of $f|U$ into a monotone map followed by a light map with at most one point in its branch set. Then, either $f|U - B$ is a homeomorphism, or $g|U - B$ is a homeomorphism and h is topologically equivalent to the map r in II (4). Thus we may confine our attention to situations in which $f|U$ is a monotone map, that is we will show that the monotone map g must be topologically equivalent to one of the maps in II (3), and thus $f|U$ is topologically equivalent to one of the maps in II (4).

Let a, b denote the two oriented 1-spheres of B. Let p be the common point. Then a and b are (smooth) 1-manifolds in M^2. There are small enough closed tubular neighborhoods W_a, W_b of a and b, resp., in M^2 such that $a \cap W_b$ and $b \cap W_a$ are connected. Let ξ_a and ξ_b be the disk bundles corresponding to W_a and W_b respectively. Then ∂W_a and ∂W_b are total spaces of the corresponding S^0-sphere bundles over a and b respectively. Hence $\partial W_a - W_b$

is a S^0-sphere bundle over $a-W_b$, which is connected, hence, a 1-cell. Therefore $\partial W_a - W_b$ is a product bundle and $\partial W_a - W_b \overset{h_1}{\simeq} (a-p) \times S^0$. Similarly, $\partial W_b - W_a \overset{h_2}{\simeq} (b-p) \times S^0$. Let T_a and T_b be the neighborhoods corresponding to the disk bundles of the product sphere bundles $\partial W_a - W_b$ and $\partial W_b - W_a$ over $a-p$ and $b-p$ respectively. Then $T_a - a = (\partial W_a - W_b) \times (0, 1)$, $T_b - b = (\partial W_b - W_a) \times (0, 1)$. Let $W = W_a \cup W_b$, then $\partial W = (\partial W_a \cap \partial W_b) \cup (\partial W_a - W_b) \cup (\partial W_b - W_a)$, and $W - T_a - T_b$ is a cone about p over $\partial W_a \cap \partial W_b$. Hence $W - a \vee b = (W - T_a - T_b) \cup (T_a - a) \cup (T_a - b) = \partial W \times (0, 1)$. Since $f|W - a \vee b$ is a homeomorphism and $f(\text{int } W - a \vee b) \cup \{f(a \vee b)\}$ is a neighborhood of $f(a \vee b)$ in V, the open cone of $f(\partial W)$ is a manifold. Thus $f(\partial W)$ (and hence ∂W) is a 1-sphere. The neighborhood int W of $a \vee b$ is the mapping cylinder of a map $h : \partial W \to a \vee b$, which is defined by the two maps h_1 and h_2, and $h(\partial W_a \cap \partial W_b) = p$. Int W is a connected manifold, and $H_1(\text{int } W) = Z \times Z$. Hence int W is homeomorphic to one of the M_i in II (3). Therefore, $f|\text{int } W$ is topologically equivalent to one of the g_i in II (3).

V. REFERENCE

(1) A. C. Hsu and J. G. Timourian, Monotone real analytic maps which are not homeomorphisms, *Proc. Amer. Math. Soc.* (To appear).

SOME HARD PION RESULTS OF CURRENT ALGEBRA

JEN-I CHEN

ABSTRACT

The continuously broken chiral symmetry model suggested by Okubo and Mathur is applied to analyze K_{l_4} form factors, Goldberger-Treiman relations, and Adler-Weisberger sum rules in hard pion calculation.

§1. INRODUCTION

The latest development of the symmetry approach to high energy physics is the chiral symmetry of SU(3)XSU(3) group first proposed by Gell-Mann.[1] Although this chiral group is not an exact symmetry of hadrons in nature, Gell-Mann suggested that the current algebra underlying it is still valid in the sense that the time components of the vector and axial-vector currents satisfy the equal-time commutation relations of the quark model. Furthermore, these currents may be considered as physical quantities in the sense that their matrix elements between hadron states can be measured via electromagnetic and weak processes. Since then many papers have been published on applying current algebra to real physical processes, and scored great success in some area, but failed in others. Most of those papers have only results in the soft pion limit, that is the unphysical approximation of zero pion mass. For more crucial test of the validity of current algebra, hard pion calculations are indeed indispensable. For this purpose we employ the continuously broken chiral symmetry model suggested by Okubo and Mathur[2].

The central idea and technique of the continuously broken chiral symmetry model is that the symmetry breaking is found to be a continuous function of some parameter a. At $a=-1$, 0, 2, it is corresponding to the soft pion limit, the SU(3) limit, and the soft kaon limit ($m_K=0$) respectively. The physical point is at $a=-0.89$. And $m_\pi^2=(1+a)m_0^2$, $m_K^2=(1-a/2)m_0$. By the continuity property of sym-

metry breaking, they further suggested that the matrix elements of physical processes are also smooth functions of a in the interval $-1\leq a\leq 2$. It is the maximal smoothness assumption of the matrix elements as functions of a that allows us to extrapolate the results calculated at $a=-1$, 0, and 2, to the physical point.

§2. K_{l4} FORM FACTORS

The K_{l4} form factors are defined in the following matrix elements as usual,

$$\langle\pi^+(q_1)\pi^-(q_2)|A_\mu^{4-i5}(0)|K^+(p)\rangle=i/m_K[F_1(q_1+q_2)_\mu+F_2(q_1-q_2)_\mu +F_3(p-q_1-q_2)_\mu] \quad (1)$$

$$\langle\pi^+(q_1)|V_\mu^{6-i7}(0)|K^+(p)\rangle=-[F_+(p+q_1)_\mu+F_-(p-q_1)_\mu]$$

with form factors $F_i=F_i(t, s, u, a)$, $i=1, 2, 3$, and $F_\pm=F_\pm(t, a)$ scalar functions of momentum variables $s=(q_1+q_2)^2$, $t=(p-q_1)^2$ and $u=(q-q_1-q_2)^2$, and the symmetry breaking parameter a.

Now we proceed to calculate the results at various exact symmetry points.

1) At $a=-1$, the exact soft pion limit relations have been obtained by several authors[3],

$$F_1=F_2=-m_K F_+/f_\pi,$$
$$F_3=-m_K f_K(1+\beta)/2f_\pi, \qquad \beta=p\cdot(q_1-q_2)/p\cdot(q_1+q_2) \quad (2)$$

with $m_\pi^2=0$, $s=0$, $t=u=-m_K^2$.

2) At $a=0$, i.e. the exact SU(3) symmetry limit, we obtain the results,

$$m_\pi^2=m_K^2=m_0^2, \qquad f_\pi=f_K=f_0:$$
$$F_+=1 \text{ at } t=0; \quad (3)$$
$$F_-(t, 0)=0.$$

3) At $a=2$, $m_K=0$ and $\partial_\mu A_\mu^{4-i5}=0$. Using the standard soft meson techniques, we obtain from (1) that as $p_\mu\to 0$, $s=u$ and $t=-m_\pi^2$,

$$F_1-F_3=0, \quad F_2=m_K I_3(s)/f_K. \quad (4)$$

where the form factor $I_3(s)$ defined by

$$\langle\pi^+(q_1)\pi^-(q_2)|V_\mu^3(0)|0\rangle=I_3(s)(q_1-q_2)_\mu \quad (5)$$

and the ρ^0—dominance of V_μ^3 current leads to

$$I_3(s)=M_\rho^2/(s+M_\rho^2) \quad (6)$$

From Eqs. (2)−(6), we extrapolate the results to the physical point through the assumption of maximal smoothness. In detail, we set

$$T_1(a)=(F_2+F_+m_K/f_K)m_K^2-(F_2+I_3m_K/f_K)m_\pi^2+R_1(a)=0$$

where $R_1(a)$ is assumed to be the smoothest function of a so chosen that $T_1=0$ at $a=-1, 0, 2$. So we find that

$$R_1=-m_\pi^2m_K(1-a/2)(1-I_3(s))/f_0,$$

and get the extrapolating relation

$$(F_2+F_+m_K/f_K)m_K^2-(F_2+I_3m_K/f_K)m_\pi^2-(1-a/2)(1-I_3)m_Km_\pi^2/f_0=0 \quad (7)$$

Similarly, one can obtain

$$(F_1-F_2)m_K^2-[(F_1-F_3)-(1-a/2)(F_2-F_3)]m_\pi^2=0 \quad (8)$$

$$(F_1-F_3+F_+m_K/f_K-(1+\beta)f_Km_K/2f_\pi^2)m_K^2-(F_1-F_3)m_\pi^2 -(1-a/2)(1-\beta)m_Km_\pi^2/2f_0=0 \quad (9)$$

The arguments s, t, u, of the form factors can also be extrapolated consistently by approximating $\vec{q}_1=\vec{q}_2=\vec{p}=0$, i.e. for $-1\leq a\leq 2$,

$$s=-4m_\pi^2,\ t=-(m_K-m_\pi)^2,\ u=-(m_K-2m_\pi)^2, \quad (10)$$

which is in the physical region of the actual decay process.

Eqs. (7)−(9) can be solved for F_i

$$F_2=\frac{1}{\alpha-1}\left[\frac{F_+m_K}{f_K}-\frac{(2-a)}{2f_0}\alpha m_K+\alpha m_K I_3\left(\frac{2-a}{2f_0}-\frac{1}{f_K}\right)\right],$$

$$F_1-F_3=\frac{1}{\alpha-1}\left[\frac{m_K}{f_\pi}\left(F_+-(1+\beta)\frac{f_K}{2f_\pi}\right)-\frac{\alpha m_K}{2f_0}(2-a)(1-\beta)\right], \quad (11)$$

$$F_1-F_2=\frac{\alpha a}{2-\alpha(2-a)}(F_1-F_3), \qquad \alpha=\left(\frac{m_\pi}{m_K}\right)^2.$$

These form factors are related to experimental K_{e4} and $K_{\mu4}$ decay rates through[4]

$$\Gamma(K_{\mu4})=1.79\times10^2|F_1\sin\theta|^2+2.48\times10|F_2\sin\theta|^2+15.7|F_3\sin\theta|^2 +78.4|F_1F_3\sin^2\theta|+0.25|F_2\sin\theta|^2+8.26\times10^{-2}|2F_3\sin\theta|^2 -8.57\times10^{-1}|2F_2F_3\sin^2\theta|\ \sec^{-1}, \quad (12)$$

$$\Gamma(K_{e4})=1.60\times10^3|F_1\sin\theta|^2+3.11\times10^2|F_2\sin\theta|^2 +3.31|F_4\sin\theta|^2\sec^{-1},$$

where F_4 is the K_{e4} decay form factor of vector current.

For better understanding of our results (11) in comparison with the soft pion results, we resort to numerical estimates. Take $\sin\theta=0.25$, $f_\pi=137$ Mev, $M_\rho=765$ Mev, $t=-(m_K-m_\pi)^2$, $s=-4m_\pi^2$, $\beta=0$

and $a=-0.89$. f_0 can be calculated from the relation[2] $f_0^2=1/3(f_\pi^2+2f_K^2)$. For F_+, we take the empirical relation $F_+(t)=1-\lambda_+ t/m^2$ and the relation[5]

$$\lambda_+=\alpha\left(1.28\frac{5x^2-1}{4x^2}-1\right), \qquad x=\frac{f_K}{f_\pi}. \tag{13}$$

The final results are given in Table 1, where ΔF_i are corrections to soft pion results (2), and the F_4 contribution to K_{e4} decay is negligible.

Table 1. Values of K_{l4} Form Factors

$x=f_k/f_\pi$	1.08	1.25	Experimental Value
λ_+	0.025	0.032	
F_+	1.16	1.20	
$-F_1$	4.19	4.39	
$-F_2$	4.28	4.47	
$-F_3$	1.98	2.33	
F_1/F_2	0.98	0.98	
$-\Delta F_1$	0.01	0.06	
$-\Delta F_2$	0.10	0.14	
$-\Delta F_3$	0.03	0.08	
$\Gamma(K_{e4})$ in 10^3sec^{-1}	2.10	2.30	(2.6 ± 0.3)[6]
$\Gamma(K_{\mu4})$ 10^2sec^{-1}	2.70	3.00	(1.1 ± 0.7)
$\Gamma_0(K_{e4})$ 10^3sec^{-1}	2.08	2.24	Γ_0 is soft pion
$\Gamma_0(K_{\mu4})$ 10^2sec^{-1}	2.66	2.89	result.

From above calculations, we reach the following conclusions:

1) The corrections to soft pion results are small, at most 5% as in the case of F_2. This shows that the soft pion results of K_{e4} form factors are good approximations, consistent with experiments.
2) The values of F_i $(i=1, 2, 3)$ and so the decay rates mainly depend on the value of F_+.

§3. CORRECTIONS TO GOLDBERGER-TREIMAN RELATIONS

The matrix elements of axial-vector currents are usually defined in terms of the following form factors by invariance consideration,

$$\langle 0|A_\mu^{1+i2}(0)|\pi^-(k)\rangle = if_\pi k_\mu,$$
$$\langle 0|A_\mu^{4+i5}(0)|K^-(k)\rangle = if_K k_\mu,$$
$$\langle B_1(p_1)|A_\mu^{1+i2}(0)|B_2(p_2)\rangle = \bar{u}_1 [iG_1^{B_1B_2}(q^2)\gamma_\mu + G_2^{B_1B_2}(q^2)q_\mu + G_3^{B_1B_2}(q^3)\sigma_{\mu\nu}q_\nu]\gamma_5 u_2$$
$$\langle B_1(p_1)|A_\mu^{4+i5}(0)|B_2(p_2)\rangle = \bar{u}_1[i\bar{G}_1^{B_1B_2}\gamma_\mu + \bar{G}_2^{B_1B_2}q_\mu + \bar{G}_3^{B_1B_2}\sigma_{\mu\nu}q_\nu]\gamma_5 u_2 \qquad (14)$$
$$\langle B_1(p_1)|j_K^\pi(c)|B_2(p_2)\rangle = G_{B_1B_2\pi}^K K_{B_1B_2\pi}^K(q^2)\bar{u}_1 i\gamma_5 u_2$$
$$g_A^{B_1B_2} = G_1^{B_1B_2}(0),\ g_{\bar{A}}^{B_1B_2} = \bar{G}_1^{B_1B_2}(0),\ q_\mu = (p_1 - p_2)_\mu,$$

where B_1 and B_2 are baryon states, $G_{B_1B_2\pi/K}$ and $K_{B_1B_2\pi/K}$ are the $B_1B_2\pi$ or B_1B_2K coupling constant and vertex function respectively. There are only 12 non-zero $g's$ which are experimentally measurable via weak interaction processes.

At $a=-1$, $m_\pi^2=0$, and the well-known Goldberger-Treiman relations[7] (GTR), derived from the PCAC assumption, become exact at this point.

$$(M_1+M_2)g_A^{B_1B_2} = f_\pi G_{B_1B_2\pi} \qquad (15)$$

with M_1, M_2 masses of B_1, B_2. In the same way one can obtain the exact GTRs at $a=2$,

$$(M_1+M_2)g_{\bar{A}}^{B_1B_2} = f_K G_{B_1B_2K} \qquad (16)$$

in the soft kaon limit. At $a=0$, it is the exact SU(3) limit, $m_\pi = m_K = m_0$, $M_1=M_2=M$, $f_\pi = f_K = f_0$. The 12 $g's$ can be expressed in terms of two independent SU(3) invariant coupling constants (d, f) of the couplings of three octets BAB as

$$\begin{aligned}
g_A^{np} &= d+f & g_{\bar{A}}^{\Sigma^- n} &= d-f \\
g_A^{\Sigma^0\Sigma^+} &= -\sqrt{2}f & g_{\bar{A}}^{\Xi^-\Lambda^0} &= (-d+3f)/\sqrt{6} \\
g_A^{\Lambda^0\Sigma^+} &= \sqrt{\frac{2}{3}}d & g_{\bar{A}}^{\Xi^-\Sigma^0} &= (d+f)/\sqrt{2} \\
g_A^{\Sigma^-\Sigma^0} &= \sqrt{2}f & g_{\bar{A}}^{\Xi^0\Sigma^+} &= d+f \\
g_A^{\Sigma^-\Lambda^0} &= \sqrt{\frac{2}{3}}d & g_{\bar{A}}^{\Lambda^0 p} &= (-d-3f)/\sqrt{6}
\end{aligned} \qquad (17)$$

$$g_{A}^{\Xi^-\Xi^0}=-(d+f) \qquad g_{A}^{\Sigma^0 p}=(d-f)/\sqrt{2}$$

and G's can be expressed in terms of two invariant constants (D, F) of the BBM coupling in exactly the same relationship.

Now consider the functions defined by

$$\begin{aligned} \gamma_{\pi}^{\alpha}(a)&=(M_1+M_2)\, g_{A}^{\alpha}-f_{\pi}G_{\alpha\pi} \\ \gamma_{K}^{\alpha}(a)&=(M_1+M_2)\, g_{A}^{\alpha}-f_{K}G_{\alpha K} \end{aligned} \tag{18}$$

which are the corrections to the usual GTRs given by Eqs. (16)–(17), where the indices B_1B_2 are replaced by α. In terms of γ's, Eqs. (16)–(17) are equivalent to

$$\gamma_{\pi}^{\pi}(-1)=\gamma_{K}^{\alpha}(2)=0 \tag{19}$$

At $a=0$, the exact SU(3) limit, we obtain

$$\begin{aligned}
&\gamma_{\pi}^{np}=D'+F' && \gamma_{K}^{\Lambda^0 p}=-(D'+3F')/\sqrt{6} \\
&\gamma_{\pi}^{\Sigma^0\Sigma^+}=-\sqrt{2}\,F' && \gamma_{K}^{\Sigma^0 p}=(D'-F')/\sqrt{2} \\
&\gamma_{\pi}^{\Lambda^0\Sigma^+}=\sqrt{\frac{2}{3}}\,D' && \gamma_{K}^{\Sigma^- n}=D'-F' \\
&\gamma_{\pi}^{\Sigma^-\Sigma^0}=\sqrt{2}\,F' && \gamma_{K}^{\Xi^-\Lambda^0}=(-D'+3F')/\sqrt{6} \\
&\gamma_{\pi}^{\Sigma^-\Lambda^0}=\sqrt{\frac{2}{3}}\,D' && \gamma_{K}^{\Xi^-\Sigma^0}=(D'+F')/\sqrt{2} \\
&\gamma_{\pi}^{\Xi^-\Xi^0}=-(D'+F') && \gamma_{K}^{\Xi^0\Sigma^+}=D'+F'
\end{aligned} \tag{20}$$

where $D'=2Md+f_0D$ and $F'=2Mf+f_0F$. Obviously there should be 10 independent relations among $12\gamma(0)$'s in the form of

$$\sum_{\alpha} C_{\alpha}\gamma_{\pi}^{\alpha}(0)=\gamma_{K}^{\beta}(0) \tag{21}$$

with C_α some numerical coefficients can be read from Eq. (20).

With all the information about γ's Eqs. (19)–(21) at three symmetry limit points, the most general extrapolation by the maximal smoothness assumption would be assuming that

$$\gamma_{\pi}^{\alpha}(a)=X_{\pi}^{\alpha}a+Y_{\pi}^{\alpha},\ \gamma_{K}^{\alpha}(a)=X_{K}^{\alpha}a+Y_{K}^{\alpha}, \tag{22}$$

with X's and Y's a-independent constants. Then Eqs. (19)–(22) imply

$$\begin{aligned} &\gamma_{\pi}^{\alpha}(a)=X_{\pi}^{\alpha}(a+1) \\ &\gamma_{K}^{\alpha}(a)=X_{K}^{\alpha}(a-2) \\ &\sum_{\alpha} C_{\alpha}X_{\pi}^{\alpha}=-2X_{K}^{\beta},\ \text{or}\ \sum_{\alpha} C_{\alpha}X_{\pi}^{\alpha}=X_{\pi}^{\beta} \end{aligned} \tag{23}$$

So that we get the relations for $-1\leqslant a\leqslant 2$,

$$\frac{\sum_{\alpha} C_{\alpha}\gamma_{\pi}^{\alpha}(a)}{\gamma_{K}^{\beta}(a)}=\frac{-2(a+1)}{(a-2)}=\frac{m_{\pi}^{2}}{m_{K}^{2}}, \tag{24}$$

or

$$\sum_{\alpha} C_{\alpha}\gamma_{\pi}^{\alpha}(a)m_{K}^{2}=\gamma_{K}^{\beta}{}_{\pi}(a)m_{K}^{2}{}_{\pi} \tag{25}$$

More explicitly, we write down some of these relations,

$$\gamma_{\pi}^{np}m_{K}^{2}=-\gamma_{\pi}^{\Xi^{-}\Xi^{0}}m_{K}^{2}=\gamma_{K}^{\Xi^{0}\Sigma^{+}}m_{\pi}^{2}=\sqrt{2}\,\gamma_{K}^{\Xi^{-}\Sigma^{0}}m_{\pi}^{2}$$

$$\gamma_{K}^{\Sigma^{-}n}=\sqrt{2}\,\gamma_{K}^{\Sigma^{0}p}$$

$$\gamma_{\pi}^{\Lambda^{0}\Sigma^{+}}m_{K}^{2}=\gamma_{\pi}^{\Sigma^{-}\Lambda^{0}}m_{K}^{2}=\frac{m_{\pi}^{2}}{\sqrt{6}}(\gamma_{K}^{\Sigma^{-}n}+\gamma_{K}^{\Xi^{0}\Sigma^{+}})$$

$$\gamma_{\pi}^{\Sigma^{0}\Sigma^{+}}=-\gamma_{\pi}^{\Sigma^{-}\Sigma^{0}} \tag{26}$$

$$\gamma_{\pi}^{\Sigma^{0}\Sigma^{+}}m_{K}^{2}=\frac{m_{\pi}^{2}}{\sqrt{2}}(\gamma_{K}^{\Xi^{0}\Sigma^{+}}-\gamma_{K}^{\Sigma^{-}n})$$

$$\gamma_{\pi}^{np}m_{K}^{2}=\frac{m_{\pi}^{2}}{2}(\gamma_{K}^{\Sigma^{-}n}-\sqrt{6}\,\gamma_{K}^{\Lambda^{0}p})$$

$$\gamma_{\pi}^{\Sigma^{-}\Lambda^{0}}m_{K}^{2}=m_{\pi}^{2}\left(\frac{1}{\sqrt{6}}\gamma_{K}^{\Sigma^{-}n}-\frac{1}{4}\gamma_{K}^{\Lambda^{0}p}\right)$$

and so on for $-1\leqslant a\leqslant 2$.

Here we would like to make some comments:

1) $\gamma's$ are the first order SU(3) corrections to usual GTRs at soft pion limit. Eq.(26) are relations among these corrections. Or one obtains correct GTRs from them by solving the $g's$ in terms of $f_{\pi,K}$ and G's.
2) The last two of Eq.(26) are very similar to the ones derived by Dashen and Weinstein[8] using the perturbation method.
3) The confirmation of these results has to wait for better experimental data on coupling constants $g's$ and G's.
4) The extrapolation Eq.(22) is by no means unique. One may extrapolate the following functions

$$\Delta_{\pi}^{\alpha}(a)=1-\frac{f_{\pi}G_{\alpha\pi}}{(M_{1}+M_{2})g_{A}^{\alpha}}=\frac{\gamma_{\pi}^{\alpha}(a)}{(M_{1}+M_{2})g_{\pi}^{\alpha}}$$

$$\Delta_{K}^{\alpha}(a)=1-\frac{f_{K}G_{\alpha K}}{(M_{1}+M_{2})g_{A}^{\alpha}}=\frac{\gamma_{K}^{\alpha}(a)}{(M_{1}+M_{2})g_{A}^{\alpha}} \tag{27}$$

instead of $\gamma's$. However, the differences are only in the second order corrections.

5) The extrapolation of Eq.(22) is equivalent to assuming

$$\gamma_\pi^a(a)=\gamma_{\pi 0}^a m_\pi^2 \text{ and } \gamma_K^a(a)=\gamma_{K0}^a m_K^2 \tag{28}$$

with $\gamma_0's$ independent of a. This means the corrections to the meson-pole dominance of PCAC are still proportional to the mass of the dominating meson. This is in fact one of the basic ideas of the continuous chiral symmetry that the chiral symmetry is realized through the existence of Goldstone bosons, i.e. massless pions and kaons.

§4. ADLER-WEISBERGER'S SUM RULES

Another result of current algebra and PCAC is the Adler-Weisberger (AW) sum rule[9] which gives the renormalization ratio between the bare and the dressed coupling constants. From the commutation relation of axial charges a^{1-i2} and a^{1+i2}

$$[a^{1+i2}(t),\ a^{1-i2}(t)]=2I_3 \tag{29}$$

by taking its matrix elements between various single baryon states, one obtains the following AW sum rules in the soft pion limit, i.e. as $a=-1$,

$$\begin{aligned}
&S_p^\pi(a)=1-(g_A^{np})^2-f_\pi^2 I_p^n=0\\
&S_n^\pi(a)=1-(g_A^{np})^2+f_n^2 I_n^\pi=0\\
&S_{\Sigma^+}^\pi(a)=2-(g_A^{\Sigma^0\Sigma^+})^2-(g_A^{\Lambda^0\Sigma^+})^2-f_\pi^2 I_{\Sigma^+}^\pi=0\\
&S_{\Sigma^-}^\pi(a)=2-(g_A^{\Sigma^-\Sigma^0})^2-(g_A^{\Lambda^0\Sigma^-})^2+f_\pi^2 I_{\Sigma^-}^\pi=0\\
&S_{\Sigma^0}^\pi(a)=(g_A^{\Sigma^+\Sigma^0})^2-(g_A^{\Sigma^-\Sigma^0})^2-f_\pi I_{\Sigma^0}^\pi=0\\
&S_{\Xi^0}^\pi(a)=1-(g_A^{\Xi^-\Xi^0})^2-f_\pi^2 I_{\Xi^0}^\pi=0\\
&S_{\Xi^-}^\pi(a)=1-(g_A^{\Xi^-\Xi^0})^2+f_\pi^2 I_{\Xi^-}^\pi=0
\end{aligned} \tag{30}$$

with

$$I_B^\pi=\frac{1}{\pi}\int_{\nu_0}^\infty \frac{d\nu}{\nu}[\sigma_B^{\pi^-}(\nu,0)-\sigma_B^{\pi^+}(\nu,0)]$$

whe $\sigma_B^\pi(\nu,0)$ is the total πB cross section and ν_0 the threshold.

Similarly at $a=2$, the exact soft kaon limit, the commutation relation

$$[a^{4+i5}(t),\ a^{4-i5}(t)]=\mathrm{Q}+\mathrm{Y} \tag{31}$$

leads to the following exact sum rules,

$$\begin{aligned}
&S^{K}_{n}(a)=1-(g^{\Sigma^- n}_{A})^2-f_K^2 I^{K}_{n}=0\\
&S^{K}_{p}(a)=2-(g^{\Sigma^- p}_{A})^2-(g^{\Lambda^\circ p}_{A})^2-f_K^2 I^{K}_{p}=0\\
&S^{K}_{\Sigma^-}(a)=1-(g^{\Sigma^- n}_{A})^2+f_K^2 I^{K}_{\Sigma^-}=0\\
&S^{K}_{\Sigma^+}(a)=1-(g^{\Xi^\circ\Sigma^+}_{A})^2-f_K^2 I^{K}_{\Sigma^+}=0\\
&S^{K}_{\Sigma^\circ}(a)=(g^{\Sigma^\circ p}_{A})^2-(g^{\Xi^-\Sigma^\circ}_{A})^2-f_K^2 I^{K}_{\Sigma^\circ}=0\\
&S^{K}_{\Lambda^\circ}(a)=(g^{\Lambda^\circ p}_{A})^2-(g^{\Xi^-\Lambda^\circ}_{A})^2-f_K^2 I^{K}_{\Lambda^\circ}=0\\
&S^{K}_{\Xi^-}(a)=2-(g^{\Xi^-\Sigma^\circ}_{A})^2-(g^{\Xi^-\Lambda^\circ}_{A})^2+f_K^2 I^{K}_{\Xi^-}=0\\
&S^{K}_{\Xi^\circ}(a)=1-(g^{\Xi^\circ\Sigma^-}_{A})^2+f_K^2 I^{K}_{\Xi^\circ}
\end{aligned} \tag{32}$$

At $a=0$, $m_\pi=m_K$, $f_\pi=f_K=f_0$ and the 12 $g's$ can be expressed in terms of coupling constants (d, f) as in Eq.(17). Furthermore, we can express the 16 $\sigma's$ and so $I's$ in terms of SU(3) invariant coupling constants of the four octets $\bar{B}\bar{M}BM$,

$$\begin{aligned}
\langle BM|BM\rangle=&a_0\mathrm{Tr}\bar{B}\bar{M}\mathrm{Tr}BM+a_0'\mathrm{Tr}\bar{B}M\mathrm{Tr}B\bar{M}+a_1\mathrm{Tr}\bar{B}\bar{M}MB\\
&+a_2\mathrm{Tr}\bar{B}\bar{M}BM+a_3\mathrm{Tr}\bar{B}MB\bar{M}+a_4\mathrm{Tr}\bar{B}B\bar{M}M\\
&+a_5\mathrm{Tr}\bar{B}BM\bar{M}+a_6\mathrm{Tr}\bar{B}M\bar{M}B
\end{aligned} \tag{33}$$

through the optical theorem,

$$\sigma^{M}_{B}(\nu,\ 0)=\mathrm{Im}.\langle BM|BM\rangle. \tag{34}$$

Explicitly, we have at $a=0$,

$$\begin{aligned}
&S^{\pi}_{p}=S^{\pi}_{n}=S^{K}_{\Sigma^+}=S^{K}_{\Xi^\circ}=1-(d+f)^2+B\\
&S^{\pi}_{\Sigma^+}=S^{\pi}_{\Sigma^-}=S^{K}_{p}=S^{K}_{\Xi^-}=2-(\frac{2}{3}d^2+2f^2)+(B-A-C)\\
&S^{\pi}_{\Xi^\circ}=S^{\pi}_{\Xi^-}=1-(d+f)^2-A\\
&S^{K}_{n}=S^{K}_{\Sigma^-}=1-(d-f)^2-A\\
&S^{\pi}_{\Sigma^\circ}=S^{\pi}_{\Lambda^\circ}=0
\end{aligned} \tag{35}$$

with constants A, B, C defined by

$$A=\frac{f_0^2}{\pi}\int_{\nu_0}^{\infty}\frac{d\nu}{\nu}\mathrm{Im.}(a_1-a_6)$$

$$B=\frac{f_0^2}{\pi}\int_{\nu_0}^{\infty}\frac{d\nu}{\nu}\mathrm{Im.}(a_4-a_5) \tag{36}$$

$$C=\frac{f_0^2}{\pi}\int_{\nu_0}^{\infty}\frac{d\nu}{\nu}\mathrm{Im.}(a_0-a_0')$$

Now we proceed to obtain sum rules at the physical point by extrapolation. From Eqs.(29), (31) and (35), the most general and smooth ones are obtained by assuming AW sum rule corrections, $S_B^{\pi,K}$'s, equal to

$$S_B^{\pi}(a)=\alpha_B^{\pi}a+\beta_B^{\pi},\ S_B^{K}(a)=\alpha_B^{K}a+\beta_B^{K} \tag{37}$$

Then the constraints at $a=-1$ and 2, i.e. $S_B^{\pi}(-1)=S_B^{K}(2)=0$, imply that

$$S_B^{\pi}(a)=\alpha_B^{\pi}(a+1)$$

$$S_B^{L}(a)=\alpha_B^{K}(a-2)$$

Thus a relation at $a=0$ such as

$$\sum_B C_B S_B^{\pi}(0)=S_B^{K}(0)$$

with C_B's some known constants, will lead to the extrapolation equation,

$$\sum_B C_B S_B^{\pi}(a)m_K^2-S_B^{K}(a)m_\pi^2=0$$

In this way, from Eq.(35), we obtain the following sum rules for $-1\leqslant a\leqslant 2$,

$$S_p^{\pi}m_K^2=S_n^{\pi}m_K^2=S_{\Sigma^+}^{K}m_\pi^2=S_{\Xi^0}^{K}m_\pi^2$$

$$S_{\Sigma^+}^{\pi}m_K^2=S_{\Sigma^-}^{\pi}m_K^2=S_p^{K}m_\pi^2=S_{\Xi^-}^{K}m_n^2$$

$$S_n^{K}=S_{\Sigma^-}^{K}$$

$$S_{\Sigma^0}^{\pi}=S_{\Lambda^0}^{\pi}=0 \tag{38}$$

$$S_{\Xi^0}^{\pi}=S_{\Xi^-}^{\pi}$$

$$S_{\Sigma^0}^{K}m_\pi^2=-S_{\Lambda^0}^{K}m_\pi^2=\frac{1}{2}(S_p^{\pi}m_K^2-S_n^{K}m_\pi^2)$$

Present data on these coupling constants and cross sections are still insufficient for final confirmation of the above sum rules.

REFERENCES

(1) M. Gell-Mann, *Phys. Rev.* **125**, 1067 (1962).
(2) S. Okubo and V. S. Mathur, *Phys. Rev. Letters* **23**, 1412 (1969).
(3) S. Weinberg, *Phys. Rev. Letters* **17**, 336 (1966); C. G. Callan and S. B. Treiman, *ibid.* **16**, 153 (1966); M. Suzuki, *ibid.* **16**, 212 (1966).
(4) F. A. Berends, A. Donnachie and G. C. Oades, *Phys. Rev. Letters* **26B**, 109 (1967).
(5) V. S. Mathur and T. C. Yang, *Phys. Rev.* **D5**, 246 (1972).
(6) K. J. Sebastian, *University of Maryland Preprint* (1969).
(7) Y. Nambu, *Phys. Rev. Letters* 4, 380 (1960); M. Gell-Mann and M. Lévy, *Nuovo Cimento* **16**, 705 (1960).
(8) R. F. Dashen and M. Weinstein, *Phys. Rev.* **183**, 1261 (1969).
(9) S. L. Adler, *Phys. Rev. Letters* **14**, 1051 (1965); W. I. Weisberger, *ibid.* **14**, 1047 (1965).

"In Helgoland I had a moment of inspiration: I realized that energy was constant in terms of time. It was rather late at night. I laboriously worked it out, and it was correct. I then climbed up on to a cliff to watch the sun-rise and was happy."

WERNER HEISENBERG
Nobel Prizewinner in Physics
(1932)

AN ELEMENTARY INTRODUCTION TO THE UNIFIED GAUGE THEORIES OF WEAK AND ELECTROMAGNETIC INTERACTIONS*

I-Fu Shih

I. INTRODUCTION

There are by now a number of excellent review articles of unified gauge theories of the weak and electromagnetic interactions[1]; most of these are written for high energy physicists. This introductory paper has been prepared, on the other hand, at a relatively lower level. Only knowledge of non-relativistic quantum mechanics and a little elementary particle physics is required.

Ever since nuclear beta decays were studied by Fermi in 1934, no one specific weak interaction theory has been regarded as a complete theory. The so-called V-A theory has many good features in the lowest order of perturbation. However, it completely breaks down in higher order calculations due to divergences. These divergences occur also in quantum electrodynamics (so-called QED). Nevertheless, the QED has been proved to be a very successful theory in the sense that divergences can be systematically taken away by redefining the physical parameters like mass and charge. That is to say the QED is a renormalizable theory while weak interaction theories are not. It is thus imperative to construct a renormalizable theory for the weak interactions.

Many attempts have been made to unify weak and electromagnetic interactions. The history goes back to Fermi's (1934) original proposal of nuclear beta decay interaction, analogous to the interaction of electrons and the e.m. field. Yukawa (1935) suggested that the beta decay interaction was mediated by a vector (spin-one) boson, analogous to the photon, but with a large mass. The great obstacle standing in the way of a synthesis is the obvious discrepancy

* Work supported in part by the National Science Council of the Republic of China.

in mass between the photon and the intermediate vector boson.

In 1967, S. Weinberg, and A. Salam, proposed a model which unifies the weak and electromagnetic interactions and also is possibly renormalizable. The basic notions are local gauge invariance and spontaneous symmetry breaking. This kind of theory has become popular since 1971 and 1972 when the renormalizability of the theory was finally confirmed. From that time on, many alternative models based on the same notions have been proposed. Also extensions of these models to include strongly interacting particles are being tried. None of the models now known is sufficiently natural and realistic to win general acceptance as a complete theory; but this line of thought remains the most promising way to the final theory.

In the next section we will introduce the Klein-Gorden equation and the Dirac equation. Then in Sec. III, the basic Lagrangian formulation for fields as well as the isospin invariance of the Lagrangian density are discussed. The idea of local gauge invariance, one of the most important ideas of the theory, is presented in the next two sections; first, a simpler abelian gauge theory and then the non-abelian Yang-Mills theory. Another basic idea is spontaneous symmetry breaking which we will illustrate in Sec. VI. The Higgs mechanism is studied in Sec. VII. Then we provide in the next section a brief review of weak interaction phenomenology. Finally the Weinberg-Salam model is discussed in detail. Sec. X contains a short conclusion.

II. RELATIVISTIC WAVE EQUATIONS

In non-relativistic quantum mechanics we are familiar with the recipe to write down the wave equation, i.e. Schrödinger equation, for a system. As a simple example, let us consider the system of a single free particle. The procedure is to write down at first the Hamiltonian for the system

$$H = \frac{\vec{p}^2}{2m}, \tag{2.1}$$

where m and $\vec{p}$ are the mass and momentum respectively of the

particle. Then we multiply both sides of Eq. (2.1) by the wave function and replace H and p by the following differential operators:

$$\begin{aligned} &\mathrm{H} \to \hat{\mathrm{H}} = i\frac{\partial}{\partial t}, \\ &\vec{p} \to \hat{\vec{p}} = -i\vec{\nabla}. \end{aligned} \tag{2.2}$$

(It is convenient to set $\hbar = c = 1$. This simplifies many equations; whenever a numerical computation is needed we can easily recover $\hbar$ and c in equations by a simple dimensional analysis.) Finally, we obtain the Schrödinger equation for a free particle

$$\left(\frac{1}{2m}\nabla^2 + i\frac{\partial}{\partial t}\right)\psi(\vec{r}, t) = 0. \tag{2.3}$$

If we now follow this recipe to construct a wave equation for a relativistic free particle, we may start with the square of the Hamiltonian

$$\mathrm{H}^2 = \mathrm{E}^2 = \vec{p}^2 + m^2 \tag{2.4}$$

to avoid difficulties of interpreting the square root of operators. Therefore we will obtain a relativistic wave equation

$$\left(\frac{\partial^2}{\partial t^2} - \nabla^2 + m^2\right)\phi(\vec{r}, t) = 0. \tag{2.5}$$

This is known as the Klein-Gordon equation.

It turns out that there is another difficulty. Because, analogous to the probability density and flux of the Schrödinger equation, when we construct from Eq. (2.5) the quantities ρ and $\vec{\mathrm{S}}$ satisfying a continuity equation

$$\vec{\nabla} \cdot \vec{\mathrm{S}} + \frac{\partial \rho}{\partial t} = 0, \tag{2.6}$$

we will find that ρ is not necessarily positive.[2] Hence ρ cannot be interpreted as probability density. Owing to this difficulty, the Klein-Gordon equation was at first abandoned.

The trouble can be traced down to the fact that the Klein-Gordon equation is of second order in t, rather than of first order as is the Schrödinger equation.

In order to avoid the difficulty of negative probability density of the Klein-Gordon equation, P. A. M. Dirac tried to set up a relativistic wave equation which is of first order in t. Starting with the form

$$i\frac{\partial}{\partial t}\psi(\vec{r}, t) = \hat{\mathrm{H}}\psi(\vec{r}, t) \tag{2.7}$$

where

$$\mathrm{H} = \mathrm{E} = \sqrt{\vec{p}^2 + m^2}\ , \tag{2.8}$$

if we replace $\vec{p}$ by $-i\vec{\nabla}$, the resulting wave equation is unsymmetrical with respect to space and time derivatives, and hence not relativistic. Dirac therefore modified the Hamiltonian in such a way as to make it linear in the space derivatives,

$$\mathrm{H} = \vec{\alpha}\cdot\vec{p} + \beta m \tag{2.9}$$

where $\vec{\alpha}$ and β are independent of $\vec{r}$, t, $\vec{p}$ and E. Then the wave equation becomes

$$\left(i\frac{\partial}{\partial t} + i\vec{\alpha}\cdot\vec{\nabla} - \beta m\right)\psi(\vec{r}, t) = 0\,. \tag{2.10}$$

This is the Dirac equation.

The requirement that equation (2.9) recover the energy equation (2.4) makes it clear that $\vec{\alpha}$ and β may not be just numbers. It can be shown that $\vec{\alpha}$ and β turn out to be 4×4 anti-commutative matrices. Since $\vec{\alpha}$ and β are 4×4 matrices, the wave function ought to be a four-component column matrix

$$\psi(\vec{r}, t) = \begin{pmatrix} \psi_1(\vec{r}, t) \\ \psi_2(\vec{r}, t) \\ \psi_3(\vec{r}, t) \\ \psi_4(\vec{r}, t) \end{pmatrix}, \tag{2.11}$$

otherwise the Eq. (2.10) has no meaning.

The conserved density constructed from the Dirac equation (2.10) is $\rho = \psi^+\psi$ which is never negative as in Schrödinger theory. So ρ can be interpreted as a probability density.

The four components of ψ naturally describe a new degree of freedom, the spin, of the particle.

Until Pauli and Weisskopf reinterpreted the Klein-Gordon equation it was believed that this Dirac equation was the only valid relativistic equation. It is now recognized that both the Dirac equation and the Klein-Gordon equation are equally valid; the Dirac equation represents particles of spin $\frac{1}{2}$ and the Klein-Gordon equation those of spin zero.

We would like now to transfer these equations to a relativistic covariant form. Let us define the space-time coordinates

$$(t, x, y, z) \equiv (t, \vec{x}) . \tag{2.12}$$

This will be denoted by the contravariant four-vector

$$x^\mu \equiv (x^0, x^1, x^2, x^3) \equiv (t, \vec{x}) . \tag{2.13}$$

The covariant four-vector x_μ is obtained by changing the sign of the space components,

$$x_\mu \equiv (x_0, x_1, x_2, x_3) \equiv (t, -\vec{x}) \tag{2.14}$$

or

$$x_\mu = g_{\mu\nu} x^\nu \tag{2.15}$$

with

$$g_{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} . \tag{2.16}$$

The scalar product is

$$x^2 = x^\mu x_\mu = x_\mu x^\mu = t^2 - \vec{x}^2 . \tag{2.17}$$

Momentum vectors are similarly defined

$$p^\mu = (\mathrm{E}, \vec{p}) = (\mathrm{E}, p_x, p_y, p_z) \tag{2.18}$$

and momentum operators in coordinate representation are written as

$$p^\mu = i\partial^\mu = i\frac{\partial}{\partial x_\mu} \equiv \left(i\frac{\partial}{\partial t}, \frac{1}{i}\vec{\nabla}\right) . \tag{2.19}$$

The scalar product is

$$p^\mu p_\mu = -\partial^\mu \partial_\mu = -\frac{\partial^2}{\partial t^2} + \nabla^2 \equiv -\Box^2 . \tag{2.20}$$

If we further define

$$\gamma^0 \equiv \beta , \qquad \vec{\gamma} \equiv \beta\vec{\alpha} ; \tag{2.21}$$

and

$$\gamma^\mu = (\gamma^0, \vec{\gamma}) , \tag{2.22}$$

the Klein-Gordon equation and the Dirac equation can be written in covariant forms,

$$(\Box^2 + m^2)\, \phi(x) = 0 , \tag{2.23}$$

$$(i\gamma^\mu \partial_\mu - m)\, \psi(x) = 0 , \tag{2.24}$$

where

$$\gamma^\mu \gamma^\nu + \gamma^\nu \gamma^\mu = 2g^{\mu\nu} . \tag{2.25}$$

In quantum field theory, each particle is associated with a field (or wave function) satisfying Eqs. (2.23) or (2.24) depending on whether it is a boson or a fermion. These fields are then quantized by treating them as operators satisfying certain commutation or anticommutation rules.

III. LAGRANGIAN FORMULATION FOR FIELDS

One of the great advantages the Lagrangian formulation has is to construct theories for some phenomena which are not well understood. To be specific, the possible form of the Lagrangian density is closely limited by the requirements that it contain only the fields and their first derivatives;* that it be a Lorentz invariant; and that it contain certain symmetries to account for some particular conservation laws observed. Once the Lagrangian density is set up one can examine many of the general properties of such a theory without knowing the detailed mechanism involved.

---

* If the Lagrangian contains higher than first-order derivatives of the fields, the field equations will be higher than second order.

Fundamentally, the equation of motion must come from Hamilton's principle. It states that if J stands for the action

$$\mathrm{J} = \int_{t_1}^{t_2} \mathrm{L}\, dt \tag{3.1}$$

where L is the Lagrangian, then the dynamics of the system is determined by the condition

$$\delta \mathrm{J} = 0 \tag{3.2}$$

i.e condition that the action is stationary.

In particle mechanics the Lagrangian is a function of q_i's and $\dot{q}_i$'s; q_i's are a set of variables which is adequate to describe the system,

$$\mathrm{L} = \mathrm{L}(q_i, \dot{q}_i)\,. \tag{3.3}$$

The necessary and sufficient condition for Hamilton's principle is Euler's equation:

$$\frac{d}{dt}\left(\frac{\partial \mathrm{L}}{\partial \dot{q}_i}\right) = \frac{\partial \mathrm{L}}{\partial q_i}\,, \qquad i = 1, 2, \ldots. \tag{3.4}$$

In a theory of fields the Lagrangian density is a function of $\phi_i(x)$ and $\partial_\mu \phi_i(x)$, where $\phi_i(x)$ $(i = 1, 2, \ldots)$ are a set of fields which is adequate to describe the system,

$$\mathscr{L}(x) = \mathscr{L}(\phi_i(x), \partial_\mu \phi_i(x))\,. \tag{3.5}$$

The Lagrangian L is, of course,

$$\mathrm{L} = \int \mathscr{L}(x)\, d^3x\,. \tag{3.6}$$

Hamilton's principle leads directly to the Euler-Lagrange equation

$$\partial^\mu \frac{\partial \mathscr{L}}{\partial(\partial^\mu \phi_i(x))} = \frac{\partial \mathscr{L}}{\partial \phi_i(x)}\,, \qquad i = 1, 2, \ldots. \tag{3.7}$$

The canonical momentum corresponding to q_i in particle mechanics is defined by

$$p_i \equiv \frac{\partial \mathrm{L}}{\partial \dot{q}_i}\,, \tag{3.8}$$

and the Hamiltonian is then given by

$$\mathrm{H} = p_i \dot{q}_i - \mathrm{L}\,. \tag{3.9}$$

A canonical momentum is also defined for a field

$$\pi_i(x) \equiv \frac{\partial \mathcal{L}}{\partial(\partial^0 \phi_i(x))}\,, \tag{3.10}$$

and the Hamiltonian density is given by

$$\mathcal{H}(x) = \pi_i(x)\,\dot{\phi}_i(x) - \mathcal{L}(x) \tag{3.11}$$

where $\dot{\phi}_i(x) = \partial^0 \phi_i(x)$.

It is easy to demonstrate that the Lagrangian density, $\mathcal{L}$ for a single free boson field can be written as

$$\mathcal{L} = \frac{1}{2}\,(\partial^\mu \phi\, \partial_\mu \phi - m^2 \phi^2)\,. \tag{3.12}$$

This is Lorentz invariant and contains only ϕ and $\partial_\mu \phi$, and the field equation obtained from the Euler-Lagrange equation

$$\partial_\mu \left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)}\right) - \frac{\partial \mathcal{L}}{\partial \phi} = 0 \tag{3.13}$$

is

$$\partial_\mu(\partial^\mu \phi) + m^2 \phi = 0 \tag{3.14}$$

or

$$(\Box^2 + m^2)\,\phi = 0 \tag{3.15}$$

which is the Klein-Gordon equation as it should be.

Similarly, for a free fermion field, $\mathcal{L}$ can be written as

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu \partial_\mu - m)\,\psi \tag{3.16}$$

where $\bar{\psi} \equiv \psi^+ \gamma_0$ is called the Dirac conjugate of ψ. Considering $\bar{\psi}$ and ψ to be two independent fields, from the Euler-Lagrange equation for $\bar{\psi}$ field,

$$\partial_\mu \left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu \bar{\psi})}\right) - \frac{\partial \mathcal{L}}{\partial \bar{\psi}} = 0 \tag{3.17}$$

we obtain the Dirac equation

$$(i\gamma^\mu \partial_\mu - m)\,\psi = 0\,. \tag{3.18}$$

For the electromagnetic field, two of the Maxwell equations

$$\vec{\nabla} \times \mathrm{E} = -\frac{\partial \vec{\mathrm{B}}}{\partial t}\,, \qquad \vec{\nabla} \cdot \vec{\mathrm{B}} = 0 \tag{3.19}$$

are served to define a four-potential

$$A^{\mu}(x) = (\phi, \vec{A}) \tag{3.20}$$

i. e.

$$\vec{B} \equiv \vec{\nabla} \times \vec{A}, \qquad \vec{E} \equiv -\nabla\phi - \frac{\partial \vec{A}}{\partial t}. \tag{3.21}$$

The electromagnetic field can be expressed by the field tensor defined as following

$$F^{\mu\nu} \equiv \partial^{\nu} A^{\mu} - \partial^{\mu} A^{\nu} \tag{3.22}$$

i. e.,

$$F^{\mu\nu} = \begin{matrix} \\ \mu\downarrow \end{matrix} \overset{\nu\rightarrow}{\begin{pmatrix} 0 & E_x & E_y & E_z \\ -E_x & 0 & B_z & -B_y \\ -E_y & -B_z & 0 & B_x \\ -E_z & B_y & -B_x & 0 \end{pmatrix}}. \tag{3.23}$$

The other two Maxwell equations

$$\vec{\nabla} \cdot \vec{E} = 0, \qquad \vec{\nabla} \times \vec{B} = \frac{\partial \vec{B}}{\partial t}, \tag{3.24}$$

are the true wave equations which can be written in an elegant form,

$$\partial_{\nu} F^{\mu\nu} = 0. \tag{3.25}$$

In order to have equation (3.25) we may start with the Lagrangian density

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu}. \tag{3.26}$$

If each of the four components of $A^{\mu}(x)$ is treated as an independent dynamical degree of freedom, the Euler-Lagrange equation

$$\partial_{\nu}\left(\frac{\partial \mathcal{L}}{\partial(\partial_{\nu} A_{\mu})}\right) - \frac{\partial \mathcal{L}}{\partial A_{\mu}} = 0 \tag{3.27}$$

simply gives the field equation (3.25).

We remark here that the following Lagrangian density

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} - \frac{1}{2}\left(\frac{\partial A_{\mu}}{\partial x_{\mu}}\right)^2 \tag{3.28}$$

also leads to equation (3.25). Although both Eqs. (3.26) and (3.28) lead to the same field equation, Eq. (3.28) is discarded. The reason is that the electromagnetic theory is gauge invariant; there should be an additional symmetry in $\mathcal{L}$, i.e. symmetry under gauge transformation of the potential A^μ. Eq. (3.28) is not gauge invariant.

In elementary particle physics we encounter some more symmetries. For instant, there exist three kinds of pions, π^+, π^0 and π^- with only slightly different masses; $m_{\pi^0} = 264.37\, m_e$ and $m_{\pi^\pm} = 273.27\, m_e$. As far as only strong interactions are considered, they are deemed to be identical particles with three different charge states. The slight mass difference comes merely from electromagnetic effects. It is convenient to assign to pions a quantum number, isospin, $I = 1$. There are three states $I_z = +1, 0, -1$, just like the usual spin, which correspond to π^+, π^0 and π^-. The three kinds of pions form an isospin triplet.

When we discuss a theory involving pions we have to include all three kinds of pions. Since pions are spin zero (scalar) particles, we should introduce three scalar fields ϕ_1, ϕ_2 and ϕ_3. Also because π^+ and π^- is a particle-antiparticle pair, they should be represented by a pair of conjugate fields. We thus define

$$\begin{aligned} \phi &\equiv \frac{1}{\sqrt{2}}(\phi_1 + i\,\phi_2) \\ \phi^* &\equiv \frac{1}{\sqrt{2}}(\phi_1 - i\,\phi_2) \\ \phi_3 &\equiv \phi_3 \end{aligned} \tag{3.29}$$

standing for π^-, π^+ and π^0 respectively.

When we consider the free pion system, the Lagrangian density should be

$$\begin{aligned} \mathcal{L} &= \frac{1}{2}(\partial^\mu \phi\, \partial_\mu \phi^* - m^2_{\pi^\pm}\, \phi^\mu \phi^*_\mu) + \frac{1}{2}(\partial^\mu \phi_3 \partial_\mu \phi_3 - m_{\pi^0} \phi_3^2) \\ &= -\frac{1}{2} \sum_{j=1}^{3} (\partial^\mu \phi_j\, \partial_\mu \phi_j - m_j^2 \phi_j^2) \end{aligned} \tag{3.30}$$

where $m_1 = m_2 = m_{\pi^\pm}$ and $m_3 = m_{\pi^0}$. If only strong interactions

are considered, then $m_{\pi^\pm}=m_{\pi^0}$, or $m_1=m_2=m_3$; we can thus define an isovector

$$\phi \equiv \begin{pmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \end{pmatrix} \tag{3.31}$$

so that the Lagrangian can be written in an isospin invariant form

$$\mathcal{L} = \frac{1}{2}(\partial^\mu \phi^+ \partial_\mu \phi - m^2 \phi^+ \phi) . \tag{3.32}$$

This is invariant under arbitrary rotation in isospin space.

Similarly, the proton and the neutron have almost equal masses. As far as only strong interactions are considered they are deemed as identical particles, nucleons, having two different charge states. We can assign $I = \frac{1}{2}$ to nucleons with $I = +\frac{1}{2}$ and $-\frac{1}{2}$ representing the proton and the neutron respectively; they form an isospinor. When we consider a nucleon system the proton and neutron fields should be considered together. The Lagrangian density is

$$\mathcal{L} = \bar{\psi}_p(i\gamma^\mu \partial_\mu - m_p)\psi_p + \bar{\psi}_n(i\gamma^\mu \partial_\mu - m_n)\psi_n . \tag{3.33}$$

If we consider only strong interactions, then $m_p=m_n$; we can define an isospinor

$$\psi \equiv \begin{pmatrix} \psi_p \\ \psi_n \end{pmatrix} \tag{3.34}$$

so that Eq. (3.33) becomes

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi . \tag{3.35}$$

This is an isoscalar, i. e. invariant under arbitrary rotation in isospin space.

So far we have been discussing an unrealistic case, the free particle system. In reality there exist always interactions, so we have to incorporate the interaction part into the original free Lagrangian densities. For example, in quantum electrodynamics we are dealing with electrons interacting with e. m. fields, the Lagrangian will be

$$\mathcal{L} = \mathcal{L}_0 + \mathcal{L}_I$$
$$= \bar{\psi}_e(x)\,(i\gamma^\mu \partial_\mu - m)\,\psi_e(x) + e\,\bar{\psi}_e(x)\,\gamma_\mu\,\psi_e(x)\,A^\mu(x)\,. \quad (3.36)$$

The interaction term $\mathcal{L}_I$ is easily understood. In classic theory the Lagrangian for a relativistic electron in the e. m. field is written as

$$L = -mc^2\sqrt{1-\frac{v^2}{c^2}} - \frac{e}{c}\vec{v}\cdot\vec{A} + e\Phi$$
$$= -mc^2\sqrt{1-\frac{v^2}{c^2}} + J_\mu A^\mu \quad (3.37)$$

where J_μ is the four-current associated with the electron. The corresponding quantity in the quantum theory is simply $e\,\bar{\psi}_e\,\gamma_\mu\,\psi_e$.

For other kind of interactions which are not so well understood, e. g., the weak or the strong interactions, the interaction part of the Lagrangian density is constructed from observed conservation laws. It is well known that there is a close relationship between conservation laws and symmetries of $\mathcal{L}$. For example, energy-momentum conservation implies space-time symmetry; angular momentum conservation implies a rotational symmetry; and charge conservation implies a gauge invariance, etc. By symmetry consideration the interaction part is usually limited to only a few alternatives; in many cases it is uniquely determined. This is a great advantage of Lagrangian formulation.

IV. LOCAL GAUGE INVARIANCE

Since the class of unified theories of the weak and e. m. interactions, is based mainly on the idea of local gauge invariance of the Lagrangian density, we will start here with a simple example to illustrate the important idea in this section. A somewhat more complicated case and some related important features will be discussed in the subsequent sections.

Let us consider the Lagrangian density of a free fermion

$$\mathcal{L}_\psi = \bar{\psi}(x)\,(i\gamma^\mu \partial_\mu - m)\,\psi(x)\,. \quad (4.1)$$

This is apparently invariant under the following gauge transformation

$$\begin{aligned}\psi(x) &\to \psi'(x) = e^{-i\theta}\,\psi(x)\\ \bar{\psi}(x) &\to \bar{\psi}'(x) = \bar{\psi}(x)\,e^{i\theta}\end{aligned} \tag{4.2}$$

where θ, the transformation parameter, is a constant independent of x.

Such a gauge transformation is called the global gauge transformation or the gauge transformation of the first kind. If, however, the transformation parameter θ is a function of x, we call it a local gauge transformation or the gauge transformation of the second kind.

$\mathcal{L}_\psi$ is globally gauge invariant but it is not invariant under the local gauge transformation. In fact, when

$$\psi(x) \to \psi'(x) = e^{-i\theta(x)}\,\psi(x) \tag{4.3}$$

the $\mathcal{L}_\psi$ is transformed as follows

$$\begin{aligned}\mathcal{L}_\psi \to \mathcal{L}'_\psi &= \bar{\psi}'(x)\,(i\gamma^\mu\partial_\mu - m)\,\psi(x)\\ &= \bar{\psi}(x)\,e^{i\theta(x)}(i\gamma^\mu\partial_\mu - m)\,e^{i\theta(x)}\psi(x)\\ &= \bar{\psi}(x)\,(i\gamma^\mu\partial_\mu - m)\,\psi(x) + \bar{\psi}(x)\,\gamma^\mu(\partial_\mu\theta(x))\,\psi(x)\,.\end{aligned} \tag{4.4}$$

The last term does not vanish and hence $\mathcal{L}_\psi$ is not locally gauge invariant. The main trouble is that $\partial_\mu\psi(x)$ is not transformed covariantly, i. e.,

$$\partial_\mu\psi(x) \to \partial_\mu\psi'(x) \neq e^{-i\theta(x)}\,\partial_\mu\psi(x)\,. \tag{4.5}$$

In order to make $\mathcal{L}_\psi$ locally gauge invariant we have to introduce a new field $A_\mu(x)$ and define a "covariant derivative" D_μ to replace the ordinary derivative ∂_μ,

$$D_\mu \equiv \partial_\mu - i\,e\,A_\mu(x) \tag{4.6}$$

where e is a constant. If we perform a proper transformation for $A_\mu(x)$ simultaneouly as $\psi(x)$ undergoes a gauge transformation, it is possible to make $D_\mu\psi(x)$ transform covariantly, i. e.

$$D_\mu\psi(x) \to D'_\mu\psi'(x) = e^{-i\theta(x)}\,D_\mu\psi(x)\,. \tag{4.7}$$

Thus the Lagrangian density

$$\mathcal{L} = \bar{\psi}(x)\,(i\gamma^\mu D_\mu - m)\,\psi(x)\,, \tag{4.8}$$

with ∂_μ being replaced by D_μ, is locally gauge invariant since

$$\begin{aligned}\mathcal{L}' &= \bar{\psi}'(x)\,(i\gamma^\mu D'_\mu - m)\,\psi'(x)\\ &= \bar{\psi}(x)\,(i\gamma^\mu D_\mu - m)\,\psi(x)\,.\end{aligned} \tag{4.9}$$

It is easy to find out the correct gauge transformation for $A_\mu(x)$. Since Eq. (4.7) reads

$$\begin{aligned}D'_\mu\psi'(x) &= (\partial_\mu - ie\,A'_\mu(x))\,e^{-i\theta(x)}\psi(x)\\ &= e^{-i\theta(x)}\partial_\mu\psi(x) - e^{-i\theta(x)}(i\,\partial_\mu\theta(x))\,\psi(x)\\ &\qquad - e^{-i\theta(x)}(ie\,A'_\mu(x)\,\psi(x)\\ &= e^{-i\theta(x)}\left[\partial_\mu - ie\left(A'_\mu(x) + \frac{1}{e}\,\partial_\mu\theta(x)\right)\right]\psi(x)\\ &= e^{-i\theta(x)}(\partial_\mu - ie\,A_\mu(x))\,\psi(x)\end{aligned} \tag{4.10}$$

we see that

$$A'_\mu(x) = A_\mu(x) - \frac{1}{e}\,\partial_\mu\theta(x)\,. \tag{4.11}$$

Because $A_\mu(x)$ is a necessary field to make $\mathcal{L}$ locally gauge invariant it is called a gauge field. It is interesting to note that if we spell out D_μ in Eq. (4.8) it simply recovers the $\mathcal{L}$ we quoted for QED, Eq. (3.36). Therefore if $\psi(x)$ stands for the electron the gauge field should be identified with the photon or the e. m. field and e with the coupling constant. In fact, the transformation Eq. (4.11) is just the familiar gauge transformation of the e. m. potentials

$$\begin{aligned}&\vec{A}(x) \to \vec{A}'(x) = \vec{A}(x) + \vec{\nabla}A(x)\\ &\Phi(x) \to \Phi'(x) = \Phi(x) - \frac{1}{c}\,\frac{\partial A(x)}{\partial t}\,.\end{aligned} \tag{4.12}$$

We see that one may take the following point of view: the existence and the form of coupling of the vector field A_μ are the consequences of the requirement of local gauge invariance.

Historically, of course, the QED theory is not based on such requirement. On the contrary, we found that gauge invariance is

an important property of the QED. For this reason one can derive Ward-Takahashi identities and, in turn, prove the renormalizability of the QED theory.

To be complete we must add into $\mathscr{L}$ a gauge invariant piece concerning the free gauge field. This causes no problem since we already have the Lagrangian density of the free e. m. field, Eq. (3.26), so we may try

$$\mathscr{L}_{\text{gauge}} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} \tag{4.13}$$

with

$$F_{\mu\nu} \equiv \partial_\nu A_\mu - \partial_\mu A_\nu . \tag{4.14}$$

This happens to be gauge invariant for

$$\begin{aligned} F'_{\mu\nu} &= \partial_\nu A'_\mu - \partial_\mu A'_\nu \\ &= \partial_\nu \Big(A_\mu - \frac{1}{e} \partial_\mu \theta \Big) - \partial_\mu \Big(A_\nu - \frac{1}{e} \partial_\nu \theta \Big) \\ &= \partial_\nu A_\mu - \partial_\mu A_\nu \\ &= F_{\mu\nu} \end{aligned} \tag{4.15}$$

We conclude that with only a single fermion field, it is impossible to construct a locally gauge invariant theory. There must be a gauge field present. The complete Langrangian is therefore

$$\begin{aligned} \mathscr{L} &= \bar{\psi} (i \gamma^\mu D_\mu - m) \psi - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} \\ &= \bar{\psi} (i \gamma^\mu \partial_\mu - m) \psi + e \bar{\psi} \gamma^\mu \psi A_\mu - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} \\ &= \mathscr{L}_\psi + \mathscr{L}_{\text{int}} + \mathscr{L}_{\text{gauge}} . \end{aligned} \tag{4.16}$$

For a boson field we may similarly construct a locally gauge invariant Lagrangian density as

$$\mathscr{L} = \frac{1}{2} (D^\mu \phi D_\mu \phi - m^2 \phi^2) - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} . \tag{4.17}$$

It is important to note that in order to have a locally gauge invariant theory, the gauge field must be massless. Otherwise we have to add a mass term $\frac{1}{2} M^2 A_\mu(x) A^\mu(x)$ in $\mathscr{L}$, which spoils the gauge invariance.

V. THE YANG-MILLS THEORY

We have just discussed a most simple gauge theory where the gauge transformation

$$\psi(x) \to \psi'(x) = e^{-i\theta(x)}\psi(x) \tag{5.1}$$

forms an abelian group U(1). Since the gauge group U(1) is a one parameter group we need only one gauge field, $A_\mu(x)$.

In 1954 Yang and Mills[3] applied the idea of local gauge invariance to the theories possessing isospin symmetry. To illustrate the Yang-Mill theory, let us consider the free Lagrangian density of the pion system or the nucleon system

$$\mathcal{L}_\phi = \frac{1}{2}(\partial^\mu \phi^+ \partial_\mu \phi - m^2 \phi^+ \phi) \tag{5.2}$$

or

$$\mathcal{L}_\psi = \frac{1}{2}\bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi \tag{5.3}$$

where ϕ and ψ are isovector and isospinor, respectively,

$$\phi = \begin{pmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \end{pmatrix}, \qquad \psi = \begin{pmatrix} \psi_p \\ \psi_n \end{pmatrix}. \tag{5.4}$$

As mentioned earlier, $\mathcal{L}_\phi$ and $\mathcal{L}_\psi$ are invariant under arbitrary rotation in isospin space. Before going on it is convenient to see first what is meant by this.

Let us study the transformation of ϕ and ψ under a rotation of frame in isospin space. Mathematically, rotations in isospin space, or equally true in spin space, form a SU(2) group which is homomorphic to the ordinary rotational group O(3). Hence, instead of studying rotations in isospin space we investigate the more familiar rotations in the ordinary three dimensional Euclid space.

A N-parameter group is generated by N generators T_i ($i = 1, 2, \ldots, N$) and the group properties are specified by the algebra of these generators

$$[T_i, T_j] = i c_{ijk} T_k, \qquad i, j, k = 1, 2, 3, \ldots, N, \tag{5.5}$$

where the repeated index is summed, and c_{ijk} is called the structure constant of the group. In particular, the generators of the O(3) group are the three components of the angular momentum, J_1, J_2 and J_3. The algebra of angular momenta is the familiar one

$$[J_i, J_j] = i\,\epsilon_{ijk}\,J_k\,, \qquad i, j, k = 1, 2, 3\,, \tag{5.7}$$

i.e., the structure constant is the Levi-Civita tensor ϵ_{ijk}.

In order to see how a vector transforms under a rotation of frame, we note that a 3-dimensional matrix representation, the regular representation of the generators, can simply be obtained from the structure constant by setting the matrix elements as

$$(t_i)_{jk} = -i\,\epsilon_{ijk}\,. \tag{5.8}$$

Explicitly

$$t_1 = i\begin{pmatrix}0&0&0\\0&0&-1\\0&1&0\end{pmatrix}, \qquad t_2 = \begin{pmatrix}0&0&-1\\0&0&0\\1&0&0\end{pmatrix}, \qquad t_3 = \begin{pmatrix}0&-1&0\\1&0&0\\0&0&0\end{pmatrix}. \tag{5.9}$$

Under a rotation about some axis through an angle denoted by $\vec{\theta}$, a vector, say the position vector,

$$\vec{r} = \begin{pmatrix}x\\y\\z\end{pmatrix} \tag{5.10}$$

transforms as

$$\vec{r} \to \vec{r}' = e^{-i\vec{t}\cdot\vec{\theta}}\,\vec{r} = e^{-it_i\theta_i}\,\vec{r}\,. \tag{5.11}$$

For a rotation about the third axis through an infinitesimal angle $\delta\theta_3$ it is easy to check the validity of Eq. (5.11) since

$$\begin{pmatrix}x'\\y'\\z'\end{pmatrix} = e^{-it_3\delta\theta_3}\begin{pmatrix}x\\y\\z\end{pmatrix} = (1 - i\,\delta\theta_3 t_3)\begin{pmatrix}x\\y\\z\end{pmatrix} = \begin{pmatrix}x - \delta\theta_3 y\\ y + \delta\theta_3 x\\ z\end{pmatrix}. \tag{5.12}$$

In general, if $\phi_i(x)$ $(i = 1, 2, \ldots, n)$ form an isospin multiplet and if we write them in a n-component column vector

$$\phi = \begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_n \end{pmatrix}, \tag{5.13}$$

then ϕ transforms according to

$$\phi(x) \to \phi'(x) = e^{-iL_i\theta_i}\phi(x)\,, \tag{5.14}$$

where L_i $(i = 1, 2, 3)$ are the n-dimensional matrix representations of the generators of the SU(2) group, i.e.

$$[L_i, L_j] = i\,\epsilon_{ijk}\,L_k\,, \qquad i, j, k = 1, 2, 3\,. \tag{5.15}$$

So, the isovector $\phi(x)$ transforms according to three-dimensional representation as

$$\phi(x) \to \phi'(x) = e^{-i\vec{t}\cdot\vec{\theta}}\phi(x) \tag{5.16}$$

where t_1, t_2 and t_3 are given in Eq. (5.9). The isospinor $\psi(x)$ transforms, on the other hand, according to two-dimensional representation as

$$\psi(x) \to \psi'(x) = e^{-i\vec{\tau}\cdot\vec{\theta}}\psi(x) \tag{5.17}$$

where $\vec{\tau} = \frac{1}{2}\vec{\sigma}$, σ_i $(i = 1, 2, 3)$ are the familiar Pauli matrices,

$$\tau_1 = \frac{1}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \tau_2 = \frac{1}{2}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \tau_3 = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{5.18}$$

The Lagrangian densities Eqs. (5.2) and (5.3) are clearly invariant under rotations in isospin space, since they are invariant under transformations Eqs. (5.16) and (5.17).

These transformations are kinds of gauge transformations. It is apparent that $\mathcal{L}_\phi$ and $\mathcal{L}_\psi$ are globally invariant, although they are not locally invariant. We will discuss here the details of making $\mathcal{L}_\phi$ locally invariant; the case of $\mathcal{L}_\psi$ can be treated analogously.

In order to make the theory locally gauge invariant we have to introduce three gauge fields, $A^1_\mu(x)$, $A^2_\mu(x)$, $A^3_\mu(x)$, for the gauge group is now SU(2) which is a three-parameter group. Let us denote these gauge fields as an isovector $\vec{A}_\mu$. Then we can replace the ordinary derivative ∂_μ by a covariant one defined as

$$D_\mu \equiv \partial_\mu - i g \vec{t} \cdot \vec{A}_\mu , \tag{5.19}$$

where g is a coupling constant. Simultaneously doing a proper transformation for $\vec{A}_\mu(x)$ as $\phi(x)$ undergoes a gauge transformation will make $D_\mu \phi(x)$ covariant, i.e.

$$D_\mu \phi(x) \to D'_\mu \phi'(x) = e^{-i\vec{t}\cdot\vec{\theta}(x)} D_\mu \phi(x) . \tag{5.20}$$

Then the Lagrangian density

$$\mathcal{L} = \frac{1}{2} [(D^\mu \phi)^+ (D_\mu \phi) - m^2 \phi^+ \phi] \tag{5.21}$$

is locally gauge invariant.

The transformation of the gauge fields can be determined as follows. For simplicity, let

$$U(\theta) \equiv e^{-i\vec{t}\cdot\vec{\theta}} \tag{5.22}$$

then

$$\phi'(x) = U(\theta) \phi(x) \tag{5.23}$$

Eq. (5.20) becomes

$$\begin{aligned} D'_\mu \phi'(x) &= (\partial_\mu - i g \vec{t} \cdot \vec{A}'_\mu(x)) U(\theta) \phi(x) \\ &= (\partial_\mu U(\theta)) \phi(x) + U(\theta) \partial_\mu \phi(x) \\ &\qquad - i g \vec{t} \cdot \vec{A}'_\mu(x) U(\theta) \phi(x) \\ &= U(\theta) (\partial_\mu - i g \vec{t} \cdot \vec{A}_\mu(x)) \phi(x) , \end{aligned} \tag{5.24}$$

so we have

$$\vec{t} \cdot \vec{A}'_\mu(x) U(\theta) = U(\theta) \vec{t} \cdot \vec{A}_\mu(x) - \frac{i}{g} (\partial_\mu U(\theta)) \tag{5.25}$$

or

$$\vec{t} \cdot \vec{A}'_\mu(x) = U(\theta) \vec{t} \cdot \vec{A}_\mu(x) U^{-1}(\theta) - \frac{i}{g} (\partial_\mu U(\theta)) U^{-1}(\theta) . \tag{5.26}$$

If we consider an infinitesimal rotation and neglect the terms of the second and higher order in θ,

$$\begin{aligned} t^i A'^i_\mu &\approx (1 - i t^j \theta^j) t^k \cdot A^k_\mu (1 + i t^j \theta^j) - \frac{1}{g} t^i \partial_\mu \theta^i \\ &\approx t^k A^k_\mu + i \theta^j A^k_\mu [t^k, t^j] - \frac{1}{g} t^i \partial_\mu \theta^i \\ &\approx t^i A^i_\mu - \varepsilon_{kji} \theta^j A^k_\mu t^i - \frac{1}{g} t^i \partial_\mu \theta^i \end{aligned} \tag{5.27}$$

or

$$A'^i_\mu = A^i_\mu - \frac{1}{g}\partial_\mu \theta^i + \varepsilon_{ijk}\theta^j A^k_\mu . \tag{5.28}$$

That is, the variation of the gauge fields under the gauge transformation is

$$\delta A^i_\mu = -\frac{1}{g}\partial_\mu \theta^i + \varepsilon_{ijk}\theta^j A^k_\mu . \tag{5.29}$$

These gauge fields $A^i_\mu(x)$ are called the Yang-Mills fields. The task now is to find out an invariant piece in $\mathcal{L}$ for the free Yang-Mills fields. We may imitate the QED theory to take

$$\mathcal{L}_{Y.M.} = -\frac{1}{4}\vec{F}_{\mu\nu}\cdot\vec{F}^{\mu\nu} = -\frac{1}{4}F^i_{\mu\nu}F^{\mu\nu i} . \tag{5.30}$$

In order to have $\mathcal{L}_{Y.M.}$ invariant under SU(2) gauge transformations, or rotations in the isospin space, $F_{\mu\nu}$ should transform as an isovector.

There are two quantities we may construct out of A^i_μ's, which are anti-symmetric Lorentz tensors and also likely to transform as isovectors; i. e. $\partial_\nu A^i_\mu - \partial_\mu A^i_\nu$ and $\varepsilon_{ijk}A^j_\mu A^k_\nu$. According to Eq. (5.29), the variation of $\partial_\nu A^i_\mu$ under the local gauge transformation is

$$\delta(\partial_\nu A^i_\mu) = -\frac{1}{g}\partial_\nu\partial_\mu\theta^i + \varepsilon_{ijk}(\partial_\nu\theta^j)A^k_\mu + \varepsilon_{ijk}\theta^j\partial_\nu A^k_\mu ; \tag{5.31}$$

then

$$\begin{aligned}\delta(\partial_\nu A^i_\mu - \partial_\mu A^i_\nu) &= \varepsilon_{ijk}\theta^j(\partial_\nu A^k_\mu - \partial_\mu A^k_\nu) \\ &\quad + \varepsilon_{ijk}[(\partial_\nu\theta^j)A^k_\mu - (\partial_\mu\theta^j)A^k_\nu] .\end{aligned} \tag{5.32}$$

With the second term nonvanishing, $\partial_\nu A^i_\mu - \partial_\mu A^i_\nu$ does not transform as an isovector alone. So let us also examine the variation of $\varepsilon_{ijk}A^j_\mu A^k_\nu$,

$$\begin{aligned}\delta(\varepsilon_{ijk}\,A^j_\mu\,A^k_\nu) &= \varepsilon_{ijk}(\delta A^i_\mu)\,A^k_\nu + \varepsilon_{ijk}(\delta A^k_\nu)\,A^j_\mu \\ &= \varepsilon_{ijk}\Big[-\frac{1}{g}\,(\partial_\mu\theta^j) + \varepsilon_{jlm}\,\theta^l\,A^m_\mu\Big]A^k_\nu \\ &\quad + \varepsilon_{ijk}\Big[-\frac{1}{g}\,(\partial_\nu\theta^k) + \varepsilon_{klm}\,\theta^l\,A^m_\nu\Big]A^j_\mu \\ &= -\frac{\varepsilon_{ijk}}{g}\,[(\partial_\mu\theta^j)\,A^k_\nu - (\partial_\nu\theta^j)\,A^k_\mu] \\ &\quad + [\varepsilon_{imk}\,\varepsilon_{kjl} + \varepsilon_{ijk}\,\varepsilon_{klm}]\,\theta^l\,A^j_\mu\,A^m_\nu \qquad (5.33)\end{aligned}$$

where

$$\begin{aligned}\varepsilon_{imk}\,\varepsilon_{kjl} + \varepsilon_{ijk}\,\varepsilon_{klm} &= (\delta_{ij}\,\delta_{ml} - \delta_{il}\,\delta_{jm}) + (\delta_{il}\,\delta_{jm} - \delta_{im}\,\delta_{jl}) \\ &= \delta_{ij}\,\delta_{ml} - \delta_{im}\,\delta_{jl} \\ &= \varepsilon_{ilk}\,\varepsilon_{jmk}\,. \qquad (5.34)\end{aligned}$$

The result is

$$\begin{aligned}\delta(\varepsilon_{ijk}\,A^j_\mu\,A^k_\nu) &= -\frac{\varepsilon_{ijk}}{g}\,[(\partial_\mu\theta^j)\,A^k_\nu - (\partial_\nu\theta^j)\,A^k_\mu] \\ &\quad + \varepsilon_{ilk}\,\varepsilon_{jmk}\,\theta^l\,A^j_\mu\,A^m_\nu\,. \qquad (5.35)\end{aligned}$$

It is clear that if we combine these two quantities and define

$$F^i_{\mu\nu} \equiv \partial_\nu A^i_\mu - \partial_\mu A^i_\nu - g\,\varepsilon_{ijk}\,A^j_\mu\,A^k_\nu \qquad (5.36)$$

then

$$\begin{aligned}\delta F^i_{\mu\nu} &= \varepsilon_{ijk}\,\theta^j(\partial_\nu A^k_\mu - \partial_\mu A^k_\nu) - \varepsilon_{ilk}\,\varepsilon_{jmk}\,\theta^l\,A^j_\mu\,A^m_\nu \\ &= \varepsilon_{ijk}\,\theta^j(\partial_\nu A^k_\mu - \partial_\mu A^k_\nu) - \varepsilon_{ijk}\,\varepsilon_{kml}\,\theta^j\,A^l_\mu\,A^m_\nu \\ &= \varepsilon_{ijk}\,\theta^j\,F^k_{\mu\nu}\,, \qquad (5.37)\end{aligned}$$

i.e. $\vec{F}_{\mu\nu}$ transforms as an isovector. The Lagrangian of free Yang-Mills fields given in Eq. (5.30) is therefore invariant. In fact

$$\begin{aligned}\delta\mathcal{L}_{\text{Y.M.}} &= -\frac{1}{4}\,\delta(F^i_{\mu\nu}\,F^{\mu\nu i}) = -\frac{1}{2}\,(\delta F^i_{\mu\nu})\,F^{\mu\nu i} \\ &= \varepsilon_{ijk}\,\theta^j\,F^k_{\mu\nu}\,F^{\mu\nu i} = 0\,. \qquad (5.38)\end{aligned}$$

The last step is due to the anti-symmetric property of ε_{ijk}.

In conclusion the total Lagrangian which is locally gauge invariant is

$$\mathcal{L} = \frac{1}{2}\,[(D^\mu\phi)^+\,(D_\mu\phi) - m^2\,\phi^+\,\phi] - \frac{1}{4}\,\vec{F}_{\mu\nu}\cdot\vec{F}^{\mu\nu}\,. \qquad (5.39)$$

Similar consideration of Eq. (5.3) will lead to a locally gauge invariant Lagrangian density as

$$\mathcal{L} = \frac{1}{2}\bar{\psi}(i\gamma^\mu D_\mu - m)\psi - \frac{1}{4}\vec{F}_{\mu\nu}\cdot\vec{F}^{\mu\nu} \tag{5.40}$$

where

$$D_\mu = \partial_\mu - ig\vec{\tau}\cdot\vec{A}_\mu \tag{5.41}$$

and $F^i_{\mu\nu}$ the same as given in Eq. (5.36).

VI. SPONTANEOUS SYMMETRY BREAKING

Nature apparently possesses two different kinds of symmetries. The symmetries associated with charge and energy-momentum conservation laws are known to be exact, while those associated with isospin, strangeness, and SU(3) are approximate. To deal with approximate symmetries one usually adds to the exactly symmetric Lagrangian a small symmetry breaking term and treats it perturbatively.

It is surprising that there exists a very distinct manner in which a symmetry is broken. In the so-called spontaneous symmetry breaking the Lagrangian is invariant under some symmetry group, while the physical vacuum is not.

As a simple example let us consider a model of a single scalar field with the Lagrangian density

$$\mathcal{L} = \frac{1}{2}\partial^\mu\phi\,\partial_\mu\phi - \frac{1}{2}\mu^2\phi^2 - \frac{1}{4}\lambda\phi^4. \tag{6.1}$$

It may be helpful to begin by demonstrating the correspondence between such a field theory Eq. (6.1) and a collection of anharmonic oscillators. (It is well-known that a free field theory is like a collection of harmonic oscillators). For simplicity, let there be only one space dimension. The Lagrangian is

$$\begin{aligned} L &= \int_{-\infty}^{\infty}\mathcal{L}(x,t)\,dx \\ &= \int_{-\infty}^{\infty}dx\left[\frac{1}{2}\left(\frac{\partial\phi}{\partial t}\right)^2 - \frac{1}{2}\left(\frac{\partial\phi}{\partial x}\right)^2 - \frac{1}{2}\mu^2\phi^2 - \frac{1}{4}\lambda\phi^4\right]. \end{aligned} \tag{6.2}$$

If we divide space into unit cells of length ε labeled by the coordinate x_i; $x_i - x_{i-1} = \varepsilon$, and consider $\phi(x, t)$ as being a canonical coordinate at each x, i.e. $\phi(x_i, t) = q_i(t)$. Then

$$L = \sum_{i=-\infty}^{\infty} \left[\frac{1}{2} \left(\frac{dq_i}{dt} \right)^2 - \frac{1}{2\varepsilon^2} (q_i - q_{i-1})^2 - \frac{1}{2} \mu^2 q_i^2 - \frac{1}{4} \lambda q_i^4 \right]$$
$$\equiv T \text{ (kinetic energy)} - V \text{ (potential)} . \qquad (6.3)$$

Apparently the first term in the bracket is the kinetic term and the rest stands for the potential. The last term makes the potential anharmonic. λ is required to be positive so that field oscillations are bounded. To do any kind of perturbation calculation, we must find the minimum of the potential and start with the unperturbed harmonic oscillator solutions as the zeroth approximation (these are the "free field" solutions of field theory).

At the minimum of potential V we must have $q_i = q_{i-1}$, i.e., all q_i are equal or, in field theory, ϕ is coordinate independent. If $\mu^2 > 0$, the potential looks like Fig. (6.1) and minimum occurs at $q = 0$, or $\phi = 0$. On the other hand, if $\mu^2 < 0$, the potential looks like Fig. (6.2), in this case the minimum occurs at $q = \pm\sqrt{-\mu^2/\lambda}$ or $\phi = \pm\sqrt{-\mu^2/\lambda}$.

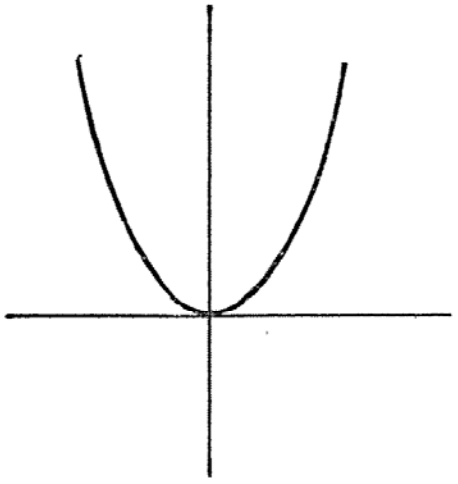


Fig. (6.1) The potential for $\mu^2 > 0$

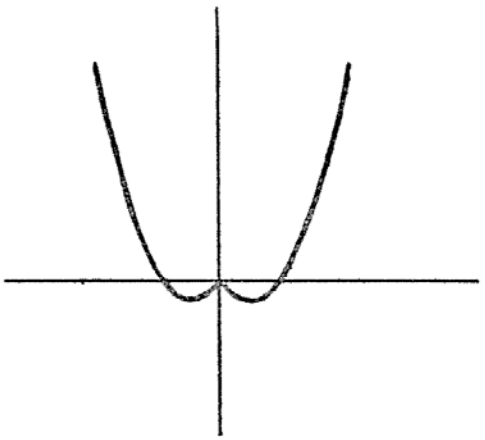


Fig. (6.2) The potential for $\mu^2 < 0$

In field theory, the ground state (lowest energy state) is the vacuum and the value of the coordinate independent field in the ground state corresponds to the vacuum expectation value of the

field. So we have now shown in a rather heuristic manner that if $\mu^2<0$ the vacuum expectation value of the field is not zero; rather, it has value $\pm\sqrt{-\mu^2/\lambda}$ to zeroth order in perturbation theory.

Let v be the vacuum expectation value of the field,

$$\langle\phi\rangle_0 \equiv v = \pm\sqrt{\frac{-\mu^2}{\lambda}}\,. \tag{6.4}$$

Either value of v may be chosen since $\mathcal{L}$ is invariant under $\phi\rightarrow-\phi$, but not both. We may by convention choose the plus sign.

The only symmetry the Lagrangian posesses is reflection invariance: $\phi\rightarrow-\phi$. Clearly the vacuum does not possess this symmetry unless $v=0$. In this sense the symmetry is spontaneously broken.

For a slightly more complicated case, let us consider the Goldstone model. The Lagrangian is given by

$$\mathcal{L} = \frac{1}{2}\partial^\mu\phi^*\partial_\mu\phi - \frac{1}{2}\mu^2\phi^*\phi - \frac{1}{4}\lambda(\phi^*\phi)^2\,, \tag{6.5}$$

where ϕ is a complex scalar field. The equation of motion is

$$(\Box^2+\mu^2)\,\phi + \lambda(\phi^*\phi)\,\phi = 0\,. \tag{6.6}$$

If we seek coordinate-independent solutions Eq. (6.6) becomes

$$(\mu^2+\lambda\,\phi^*\phi)\,\phi = 0\,. \tag{6.7}$$

As before, λ is required to be positive. If $\mu^2>0$ the solution is $\phi=0$. On the other hand, if $\mu^2<0$, we have an extra solution

$$\phi = \sqrt{\frac{-\mu^2}{\lambda}}\,e^{i\alpha}\,. \tag{6.8}$$

The phase α can be arbitrarily chosen since the Lagrangian Eq. (6.5) is invariant under the phase transformation

$$\phi(x)\rightarrow e^{-i\theta}\,\phi(x)\,. \tag{6.9}$$

It is most convenient to choose $\alpha=0$ so that ϕ is a real constant

$$\phi = \sqrt{\frac{-\mu^2}{\lambda}}\,. \tag{6.10}$$

This solution is not invariant under the phase transformation Eq. (6.9).

We may easily show that, for $\mu^2 < 0$, the non-vanishing solution Eq. (6.10), rather than $\phi = 0$, corresponds to the minimum of the potential. Therefore we face again a spontaneous symmetry breaking, i. e. the vacuum does not exhibit the symmetry which the Lagrangian possesses. When the vacuum expectation value of the field ϕ does not vanish, it is convenient to introduce a new field which has a vanishing vacuum expectation value,

$$\varphi \equiv \phi - v \,, \tag{6.11}$$

so we can use the ordinary perturbation theory in φ. If we define two real fields φ_1 and φ_2 according to

$$\varphi \equiv \varphi_1 + i\,\varphi_2 \tag{6.12}$$

the original Lagrangian Eq. (6.5) can be expressed in terms of the new fields φ_1 and φ_2,

$$\begin{aligned} \mathcal{L} = & -\frac{1}{2}\partial^\mu \varphi_1 \partial_\mu \varphi_1 - \frac{1}{2}\partial^\mu \varphi_2 \partial_\mu \varphi_2 - \varphi_1 v(\mu^2 + \lambda v^2) \\ & - \frac{1}{2}\varphi_1^2(\mu^2 + 3\lambda v^2) - \frac{1}{2}\varphi_2^2(\mu^2 + \lambda v^2) \\ & - \lambda v\,\varphi_1(\varphi_1^2 + \varphi_2^2) - \frac{1}{4}\lambda(\varphi_1^2 + \varphi_2^2)^2 \,, \end{aligned} \tag{6.13}$$

where the constant terms which have no physical consequences have been dropped. From Eq. (6.12) we see that since $v = \sqrt{-\mu^2/\lambda}$, the masses of φ_1 and φ_2 are

$$\begin{aligned} m_{\varphi_1} &= \mu^2 + 3\lambda v^2 = -2\mu^2 > 0 \,, \\ m_{\varphi_2} &= \mu^2 + \lambda v^2 = 0 \,. \end{aligned} \tag{6.14}$$

The presence of a massless scalar field φ_2 is a special case of Goldstone's theorem[4]. In general, if the Lagrangian has a symmetry corresponding to a n-parameter group and the vacuum exhibits only the symmetry associated with a subgroup having $m(<n)$ parameters, the Goldstone's theorem says that there will be $n-m$ massless scalars.

VII. THE HIGGS MECHANISM

It is interesting to combine the spontaneous broken symmetries and the local gauge invariance. The Goldstone theorem quoted in

the end of the last section fails in case the Lagrangian is invariant under some local gauge transformation and therefore possesses gauge fields.

Let us consider the Higgs model[5] which in the simplest form is just the Goldstone model Eq. (6.5) subject to local gauge invariance. According to Sec. IV, the Lagrangian is

$$\begin{aligned}\mathcal{L} &= (\partial_\mu + i e A_\mu)\,\phi^* (\partial^\mu - i e A^\mu)\,\phi \\ &\quad - \mu^2 \phi^* \phi - \lambda(\phi^* \phi)^2 - \frac{1}{4} F_{\mu\nu} F^{\mu\nu}\end{aligned} \tag{7.1}$$

where $F_{\mu\nu} = \partial_\nu A_\mu - \partial_\mu A_\nu$. This Lagrangian is invariant under local gauge transformations

$$\begin{aligned}&\phi(x) \to \phi'(x) = e^{-i\theta(x)}\,\phi(x)\,, \\ &A_\mu(x) \to A'_\mu(x) = A_\mu(x) - \frac{1}{e}\,\partial_\mu \theta(x)\,.\end{aligned} \tag{7.2}$$

In case $\mu^2 < 0$, the symmetry is spontaneously broken. The vacuum expectation value of ϕ is

$$\langle \phi \rangle_0 = \sqrt{\frac{-\mu^2}{2\lambda}} = \frac{v}{\sqrt{2}}\,. \tag{7.3}$$

(Note that different coefficients are used in Eq. (6.5) and Eq. (7.1).)

Instead of defining new fields φ_1 and φ_2 as before, it is convenient to introduce new real fields ξ and η, and parametrize ϕ exponentially as

$$\begin{aligned}\phi(x) &\equiv e^{i(\xi(x)/v)}\,\frac{v + \eta(x)}{\sqrt{2}} \\ &= \frac{1}{\sqrt{2}}\,(v + \eta(x) + i\xi(x) + \text{higher order terms})\,.\end{aligned} \tag{7.4}$$

$\eta(x)$ and $\xi(x)$ correspond to the φ_1 and φ_2 in the former case respectively as far as we keep only the linear terms.

Before inserting the Eq. (7.4) into Lagrangian (7.1) we notice that $\mathcal{L}$ is invariant under local gauge transformations (7.2). Hence, to be simpler, we may make a transformation choosing the gauge function as $\xi(x)/v$. Then

$$\phi(x) \to \phi'(x) = e^{-i(\xi(x)/v)}\,\phi(x) = \frac{1}{\sqrt{2}}\,(v+\eta(x))\,,$$
$$\mathrm{A}_\mu(x) \to \mathrm{A}'_\mu(x) = \mathrm{A}_\mu(x) - \frac{1}{ev}\,\partial_\mu\,\xi(x)\,. \tag{7.5}$$

The Lagrangian becomes

$$\begin{aligned}
\mathcal{L} &= \frac{1}{2}\,(\partial_\mu + i\,e\,\mathrm{A}'_\mu)\,(v+\eta)\cdot(\partial^\mu - i\,e\,\mathrm{A}'^\mu)\,(v+\eta) \\
&\quad - \frac{1}{2}\,\mu^2(v+\eta)^2 - \frac{1}{4}\,\lambda(v+\eta)^4 - \frac{1}{4}\,\mathrm{F}'_{\mu\nu}\,\mathrm{F}'^{\mu\nu} \\
&= -\frac{1}{4}\,\mathrm{F}'_{\mu\nu}\,\mathrm{F}'^{\mu\nu} + \frac{1}{2}\,\partial_\mu\,\eta\,\partial^\mu\,\eta + \frac{1}{2}\,e^2 v^2\,\mathrm{A}'_\mu\,\mathrm{A}'^\mu \\
&\quad + \frac{1}{2}\,e^2\,\mathrm{A}'_\mu\,\mathrm{A}'^\mu\,\eta(2v+\eta) - (\mu^2 + v^2\lambda)\,v\,\eta \\
&\quad - \frac{1}{2}\,(\mu^2 + 3v^2\lambda)\,\eta^2 - \lambda\,v\,\eta^3 - \frac{1}{4}\,\lambda\,\eta^4 + \text{const. terms} \\
&= -\frac{1}{4}\,\mathrm{F}'_{\mu\nu}\,\mathrm{F}'^{\mu\nu} + \frac{1}{2}\,\partial_\mu\,\eta\,\partial^\mu\,\eta + \frac{1}{2}\,e^2 v^2\,\mathrm{A}'_\mu\,\mathrm{A}'^\mu \\
&\quad + \frac{1}{2}\,e^2\,\mathrm{A}'_\mu\,\mathrm{A}'^\mu\,\eta(2v+\eta) + \mu^2\eta^2 - \lambda\,v\,\eta^3 \\
&\quad - \frac{1}{4}\,\lambda\,\eta^4 + \text{const. terms}\,.
\end{aligned} \tag{7.6}$$

Here the Eq. (7.3) has been used. It is clear that the ξ field, which otherwise would be the Goldstone boson, disappears and the remaining scalar field η has mass $-2\mu^2(>0)$. Also the gauge field now becomes a massive vector boson with a mass equal to $e\,v$.

This is the so-called Higgs mechanism: when a locally gauge invariant Lagrangian undergoes spontaneous symmetry breaking, the would-be Goldstone boson is eaten up by the gauge field, which, as a consequence of this inclusion becomes a massive vector boson.

The Higgs mechanism can easily be extented to a more complicated model; for example, the Yang-Mills theory with the simplest non-abelian gauge group, i.e. SU(2). The scalar fields ϕ_i $(i=1, 2, 3)$ form an isovector ϕ. The part of the invariant Lagrangian containing ϕ is

$$\begin{aligned}
\mathcal{L} &= \frac{1}{2}\,(\partial_\mu\,\phi_i + g\,\varepsilon^{ijk}\,\mathrm{A}^j_\mu\,\phi_k)\,(\partial^\mu\,\phi_i + g\,\varepsilon^{ilm}\,\mathrm{A}'^\mu\,\phi_m) \\
&\quad - \mathrm{V}(\phi^*\phi)
\end{aligned} \tag{7.7}$$

where V is an SU(2) invariant quadratic polynomial in ϕ_i's.

When the symmetry is spontaneously broken, i. e. ϕ has non-vanishing vacuum expectation values, we can always perform an isospin rotation so that it is the third component which acquires a vacuum-expectation value,

$$\langle\phi\rangle_0 = \begin{pmatrix} 0 \\ 0 \\ v \end{pmatrix}. \tag{7.8}$$

The vacuum is not invariant under the whole isospin group, i. e. SU(2). Rather, it is invariant only under the rotation about the third axis in isospin space.

We now introduce three new fields ξ_1, ξ_2 and η, and parametrize ϕ similar to the previous case,

$$\begin{aligned} \phi &\equiv e^{i(\xi_1 t_1 + \xi_2 t_2)/v} \begin{pmatrix} 0 \\ 0 \\ v+\eta \end{pmatrix} \\ &= \langle\phi\rangle_0 + \begin{pmatrix} \xi_2 \\ \xi_1 \\ \eta \end{pmatrix} + \text{higher order terms}. \end{aligned} \tag{7.9}$$

The fields ξ_1 and ξ_2 are the would-be Goldstone bosons associated with the two broken degrees of freedom. Since the Lagrangian (7.7) is invariant under local SU(2) gauge transformations, we are allowed to make the following gauge transformation:

$$\phi' = e^{-i(\xi_1 t_1 + \xi_2 t_2)/v}\,\phi; \tag{7.10}$$

thus

$$\phi' = \begin{pmatrix} 0 \\ 0 \\ v+\eta \end{pmatrix}. \tag{7.11}$$

The fields ξ_1 and ξ_2 have been "gauged away". When we write the Lagrangian in this particular gauge, ξ_1 and ξ_2 completely disappear.

$$\begin{aligned}\mathcal{L} &= \frac{1}{2}\partial_\mu \phi'_i \partial^\mu \phi'_i + \frac{1}{2} g\,\varepsilon^{ijk} A'^j_\mu \phi'_k\, \partial^\mu \phi'_i + \frac{1}{2} g\,\varepsilon^{ilm} A'^{\mu l} \phi'_m \\ &\quad + \frac{1}{2} g^2 \varepsilon^{ijk} \varepsilon^{ilm} A'^j_\mu A'^{\mu l} \phi'_k \phi'_m - V(\phi'_i \phi'_i) \\ &= \frac{1}{2}\partial_\mu \eta \partial^\mu \eta + \frac{1}{2} g^2 v^2 \varepsilon^{ij3} \varepsilon^{il3} A'^j_\mu A'^{\mu l} \\ &\quad - V((v+\eta)^2) + \text{higher order terms} \\ &= \frac{1}{2}\partial_\mu \eta \partial^\mu \eta + \frac{1}{2} g^2 v^2 (A'^2_\mu A'^{2\mu} + A'^1_\mu A'^{1\mu}) \\ &\quad - V((v+\eta)^2) + \ldots . \end{aligned} \tag{7.12}$$

From the second term we see the gauge fields corresponding to the broken degrees of freedom have acquired a mass $M = gv$.

As has been mentioned in Sec. IV, the photon which mediates the e. m. interactions can be regarded as the gauge field associated with a U(1) gauge group. It is natural to conjecture that the weak intermediate vector bosons which are supposed to mediate the weak interactions can also be regarded as the gauge fields associated with an SU(2) gauge group. A difficulty of this conjecture is that the weak intermediate vector bosons are massive, and we have emphasized that the gauge fields should be massless. The Higgs mechanism provides a method to create masses for the gauge fields without adding mass terms to the Lagrangian, which breaks the local gauge invariance explicitly.

We are now in a suitable position to discuss the unified gauge theory of the weak and electromagnetic interactions proposed by S. Weinberg and A. Salam. However, in order to fully appreciate this model, it may be pertinent to make a brief review of weak interaction phenomenology.

VIII. A BRIEF REVIEW OF WEAK INTERACTION PHENOMENOLOGY

We will only discuss the purely leptonic weak processes which are the simplest because of absence of strongly interacting particles. Until very recent there was actually only one such process firmly observed, i. e. the decay of a muon into an electron, a neutrino, and an antineutrino,

$$\mu^- \to e^- + \nu_e + \bar{\nu}_e \,. \tag{8.1}$$

Now several other processes like $\bar{\nu}_e + e \to \bar{\nu}_e + e$, and $\bar{\nu}_\mu + e \to \bar{\nu}_\mu + e$ have been reported since 1972. (There are two types of neutrinos, ν_e and ν_μ.)

The muon decay process has been successfully described by the following interaction Lagrangian[6] (we will use the names of particles to denote their fields),

$$\mathcal{L}_{\text{int}} = -\frac{G}{\sqrt{2}} [\bar{e}(x)\, \gamma_\lambda (1-\gamma_5)\, \nu_e(x)] \cdot [\bar{\nu}_\mu(x)\, \gamma^\lambda (1-\gamma_5)\, \mu(x)] + \text{h.c.}, \tag{8.2}$$

which states that the four fermions interact at one point, see Fig. (8.1).

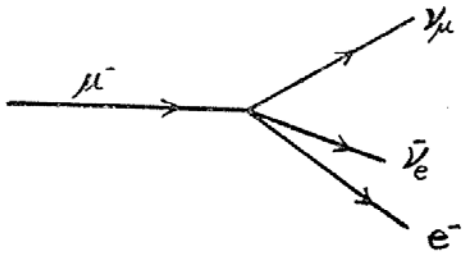


Fig. (8.1) The muon decay described by Eq. (8.2)

Although they look quite different in mass, the electron and the muon seem to share the same properties in interactions. Therefore it is natural to impose the idea of $e-\mu$ universality and extend the Lagrangian (8.2) to an elegant form,

$$\mathcal{L}^{\text{int}} = \frac{G}{\sqrt{2}}\, l_\lambda(x)\, l^{\lambda+}(x) \tag{8.3}$$

where $l_\lambda(x)$ is the leptonic current,

$$l_\lambda(x) \equiv \bar{e}(x)\, \gamma_\lambda (1-\gamma_5)\, \nu_e(x) + \bar{\mu}(x)\, \gamma_\lambda (1-\gamma_5)\, \nu_\mu(x) \tag{8.4}$$

With this interaction Lagrangian, the process like

$$\nu_\mu + e^- \to \nu_\mu + e^- \tag{8.5}$$

can occur only in the second order, see Fig. (8.2).

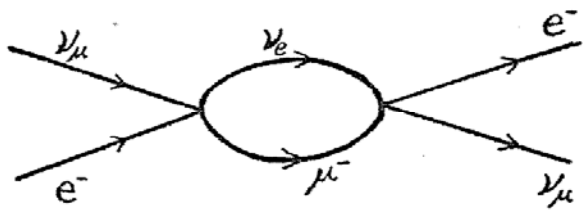


Fig. (8.2) $\nu_\mu + e^- \to \nu_\mu + e^-$ described by Eq. (8.3)

If one calculates the cross section of $\nu_\mu + e^- \to \nu_\mu + e^-$ process according to the Lagrangian (8.3) or the Feynman diagram Fig. (8.2), the result will be infinity. These kinds of divergences occur in the QED theory, too. However, in QED there is a systematic way known as renormalization to take away these divergences reasonably. The current-current weak interaction theory Eq. (8.3) is not renormalizable, i.e. it succeeds in the lowest order but fails in higher orders.

Besides the above current-current model, a different type model has been proposed by Yukawa. It is called the intermediate vector boson (IVB) theory. In this theory, the weak interactions are mediated by some intermediate vector bosons just like the strong interactions are mediated by pions or the e.m. interactions by the photon. The Lagrangian density describing the IVB theory is given by

$$\mathcal{L}_{\text{int}} = g\,[l_\lambda(x)\,\mathrm{W}^{\lambda+}(x) + \text{h.c.}] \tag{8.6}$$

where $\mathrm{W}^\lambda(x)$ stands for the IVB field or W-boson field.

The muon decay in this model occurs in the second order as shown in Fig. (8.3).

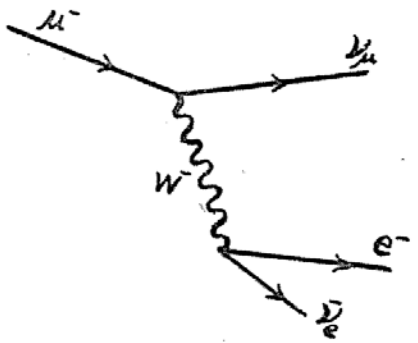


Fig. (8.3) The muon decay described by Eq. (8.6)

This second order process can be described, according to Eq. (8.6), by an effective Lagrangian

$$\mathcal{L}_{\text{int}}^{\text{eff}} = -ig^2 \int d^4x' \, \Delta^{\lambda\rho}(x-x') \, \mathrm{T}\{l_\lambda(x) \, l_\rho^+(x)\} \tag{8.7}$$

where $\Delta^{\lambda\rho}$ is the propagator of W-boson. When m_{W} is large,

$$\Delta^{\lambda\rho}(x-x') \approx \frac{-i}{m_{\text{W}}^2} g^{\lambda\rho} \, \delta^4(x-x') \tag{8.8}$$

then

$$\mathcal{L}_{\text{int}}^{\text{eff}} \approx \frac{g^2}{m_{\text{W}}^2} l_\lambda(x) \, l^{\lambda+}(x) \, . \tag{8.9}$$

Because the interaction range is inversely proportional to the mass of the intermediate particle, when m_{W} is large enough, the IVB theory Eq. (8.6) must approach the point-interaction theory or current-current theory Eq. (8.3). We may thus identify

$$\frac{g^2}{m_{\text{W}}^2} \equiv \frac{\text{G}}{\sqrt{2}} \, . \tag{8.10}$$

At low energy processes, the energies of leptons are small compared to m_{W}; these two theories are indistinguishable.

As for the process $\nu_\mu + e \to \nu_\mu + e$, it occurs at the fourth order of g; see Fig. (8.4).

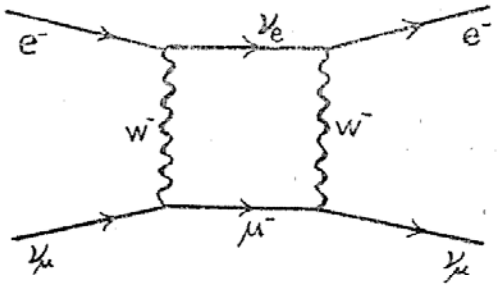


Fig. (8.4) $\nu_\mu + e \to \nu_\mu + e$ described by Eq. (8.6)

Computation of this Feynman diagram yields also infinity. The IVB theory is still unrenormalizable.

IX. THE WEINBERG-SALAM MODEL

In 1967 Weinberg, and also Salam,[7] proposed a model which unifies the weak and e.m. interactions. Such a theory is based on

the ideas of local gauge invariance and spontaneous symmetry breaking. Aside from the nice property of unification the model has been proved later to be renormalizable. Various models of this kind were subsequently suggested. We will, however, discuss here only the very original one—the Weinberg-Salam model.

Let us note that any spinor field can be divided into left- and right-handed parts, i.e.,

$$\begin{aligned}\psi &= \frac{1}{2}(1-\gamma_5)\psi + \frac{1}{2}(1+\gamma_5)\psi \\ &= \mathrm{P_L}\psi + \mathrm{P_R}\psi = \psi_\mathrm{L} + \psi_\mathrm{R}\end{aligned} \tag{9.1}$$

where $\gamma_5 \equiv i\gamma^0\gamma^1\gamma^2\gamma^3$, $\mathrm{P_L} \equiv \frac{1}{2}(1-\gamma_5)$ and $\mathrm{P_R} \equiv \frac{1}{2}(1+\gamma_5)$ are known as left- and right-handed projection operators. In terms of ψ_L and ψ_R, the leptonic current in Eq. (8.4) can be written as

$$\begin{aligned}l_\lambda(x) &= \bar{e}(x)\gamma_\lambda(1-\gamma_5)\nu_e(x) + (e \leftrightarrow \mu) \\ &= 2\left[\bar{e}_\mathrm{L}(x)\gamma_\lambda \nu_{e_\mathrm{L}}(x) + (e \leftrightarrow \mu)\right].\end{aligned} \tag{9.2}$$

Here we have used a property of the projection operator, $\mathrm{P_L^2} = \mathrm{P_L}$; also notice that

$$\bar{\psi}_\mathrm{L} = \psi_\mathrm{L}^+\gamma_0 = \psi^+\frac{1-\gamma_5}{2}\gamma_0 = \psi^+\gamma_0\frac{1+\gamma_5}{2} = \bar{\psi}\frac{1+\gamma_5}{2}. \tag{9.3}$$

Because of the $e-\mu$ universality we shall only consider the parts concerning the electron. The other parts concerning the muon can easily be spelt out whenever necessary. We may form an isospin doublet (isospinor) out of ν_{e_L} and e_L

$$\mathrm{L} \equiv \begin{pmatrix}\nu_{e_\mathrm{L}} \\ e_\mathrm{L}\end{pmatrix}, \tag{9.4}$$

then the leptonic current Eq. (9.2) can be rewritten as

$$\begin{aligned}l_\lambda &= 2\mathrm{L}^+\gamma_\lambda\tau^-\mathrm{L}, \\ l_\lambda^\dagger &= 2\mathrm{L}^+\gamma_\lambda\tau^+\mathrm{L},\end{aligned} \tag{9.5}$$

where $\tau^\pm \equiv \tau_1 \pm i\tau_2$ and τ_i are given in Eq. (5.18).

We observe that l_λ and $l_\lambda^\dagger$ are the two charged components of isospin current. If we introduce the neutral component into Eq. (8.3), $\mathscr{L}_\mathrm{int}$ becomes SU(2) invariant,

$$\mathcal{L}_{\text{int}} = \frac{G}{\sqrt{2}} [l_\lambda(x)\, l^{\lambda+}(x) + l^3_\lambda(x)\, l^{3\lambda}(x)]$$
$$= \frac{G}{\sqrt{2}} [l^1_\lambda(x)\, l^{1\lambda}(x) + l^2_\lambda(x)\, l^{2\lambda}(x) + l^3_\lambda(x)\, l^{3\lambda}(x)] \qquad (9.6)$$

where

$$l^3_\lambda = 2\mathrm{L}^+ \gamma_\lambda \tau^3 \mathrm{L}\,. \qquad (9.7)$$

In weak interactions only the left-handed fields participate. However, in e.m. interactions, the right-handed fields will also participate. Let us take e_R as a SU(2) singlet (there are no right-handed neutrinos). We define

$$\mathrm{R} \equiv e_R\,. \qquad (9.8)$$

Since the electron field has been seperated into left- and right-handed parts and the two parts transform differently under an isospin rotation, the electron cannot possess mass at the beginning, otherwise the Lagrangian is not gauge invariant at all. This can be observed by writing the Lagrangian density of a free electron in terms of left-handed and right-handed fields,

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi$$
$$= i\bar{\psi}_L \gamma^\mu \partial_\mu \psi_L + i\bar{\psi}_R \gamma^\mu \partial_\mu \psi_R - m(\bar{\psi}_L \psi_R + \bar{\psi}_R \psi_L)\,. \qquad (9.9)$$

Fortunately, we may generate electron mass in a similar way as in the Higgs mechanism.

With leptons being so grouped the Lagrangian density of free leptons is

$$\mathcal{L}_{\text{leptons}} = \bar{\mathrm{R}}\, i\gamma^\mu \partial_\mu \mathrm{R} + \bar{\mathrm{L}}\, i\gamma^\mu \partial_\mu \mathrm{L}\,. \qquad (9.10)$$

The Lagrangian is obviously invariant under SU(2) group transformation. There is another gauge group. Since we may assign a quantum number, "hypercharge", $\mathrm{Y} = -1$ to L and $\mathrm{Y} = -2$ to R according to the Gell-Mann-Nishijima relation among charge, isospin and hypercharge,

$$\mathrm{Q} = \mathrm{I}_z + \frac{1}{2}\mathrm{Y}\,, \qquad (9.11)$$

there is also a U(1) hypercharge group.

So, the gauge group of the Lagrangian (9.10) is $SU(2) \otimes U(1)$. In order to make $\mathcal{L}_{\text{leptons}}$ locally gauge invariant we have to introduce three gauge fields $A^i_\mu(x)$ for SU(2) and another $B_\mu(x)$ for U(1). As discussed in Sec. IV and V, the locally invariant form of $\mathcal{L}_{\text{leptons}}$ is

$$\begin{aligned}\mathcal{L}_{\text{leptons}} = {} & \bar{R}\, i\gamma^\mu(\partial_\mu + i g' B_\mu)\, R \\ & + \bar{L}\, i\gamma^\mu \left(\partial_\mu + \frac{i}{2} g' B_\mu - i g \frac{\vec{\tau}}{2} \cdot \vec{A}_\mu\right) L\,, \end{aligned} \tag{9.12}$$

where g and $g'/2$ are the coupling constants. The free gauge field part is

$$\mathcal{L}_{\text{gauge}} = -\frac{1}{4} F^i_{\mu\nu} F^{\mu\nu i} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu}\,, \tag{9.13}$$

with

$$\begin{aligned} F^i_{\mu\nu} &\equiv \partial_\nu A^i_\mu - \partial_\mu A^i_\nu - g\, \epsilon^{ijk} A^j_\mu A^k_\nu\,, \\ B_{\mu\nu} &\equiv \partial_\nu B_\mu - \partial_\mu B_\nu\,. \end{aligned} \tag{9.14}$$

Because the photon as a gauge field is the particle mediating e. m. interactions, it is natural to speculate that the Yang-Mills fields are the IVB mediating weak interactions. The idea seems unlikely at the first glance since IVB must be massive (the range of the weak interactions is very short), while the gauge fields are massless. Explicitly introducing mass terms in the Lagrangian will spoil the local gauge invariance. Furthermore, the massless Yang-Mills theory has been proved renormalizable while the massive one not.

Fortunately the Higgs mechanism allows us to generate masses for the gauge field without breaking the invariance explicitly. It was sensed that if the masses of the gauge fields were created merely through spontaneous breaking, the Yang-Mills theory may still be renormalizable. This conjecture was later, in 1971, proved to be correct.

We would like to create masses for three gauge fields and leave the fourth one massless which can be interpreted as the photon. To do so, Weinberg and Salam introduce into the theory a Higgs (complex) scalar doublet

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}. \tag{9.15}$$

This doublet transforms like L of Eq. (9.3) under SU(2), and has $Y = +1$ to maintain Eq. (9.11). The Lagrangian containing Higgs scalars will be

$$\mathcal{L}_{\text{scalar}} = \left(\partial_\mu \phi - i\frac{g'}{2} B_\mu \phi - i g \frac{\vec{\tau}}{2} \cdot \vec{A}_\mu \phi\right)^+ \cdot \left(\partial^\mu \phi - i\frac{g'}{2} B^\mu \phi - i g \frac{\vec{\tau}}{2} \cdot \vec{A}_\mu \phi\right) - V(\phi^+ \phi). \tag{9.16}$$

The most general form for $V(\phi^+ \phi)$ is

$$V(\phi^+ \phi) = \mu^2 \phi^+ \phi + \lambda(\phi^+ \phi)^2. \tag{9.17}$$

Remember that we started with a massless electron; in order to create the mass for the electron, there should be interactions between the Higgs scalar and the electron. By requirement of symmetry the only possible form of the interaction Lagrangian is

$$\mathcal{L}_{\text{int}} = -f[\bar{R}\phi^+ L - \bar{L}\phi R]. \tag{9.18}$$

Therefore the whole Lagrangian density of this model is

$$\mathcal{L} = \mathcal{L}_{\text{leptons}} + \mathcal{L}_{\text{gauge}} + \mathcal{L}_{\text{scalar}} + \mathcal{L}_{\text{int}}. \tag{9.19}$$

The gauge group is $SU(2) \otimes U(1)$, a four-parameter group.

Now, let μ^2 be negative so that the theory Eq. (9.19) is spontaneously broken, i. e. the vacuum expectation value of ϕ is not vanishing. We may choose

$$\langle\phi\rangle_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \tag{9.20}$$

where $v = \sqrt{-\mu^2/\lambda}$ is real. Notice that this breaks both the SU(2) and the hypercharge U(1) symmetries. The surviving symmetry is the combination charge as shown in Eq. (9.11).

We replace the four real components of ϕ by η and ξ^i $(i=1, 2, 3)$, according to

$$\phi = e^{-i(\vec{\xi}\cdot\vec{\tau})/2v} \begin{pmatrix} 0 \\ \frac{v+\eta}{\sqrt{2}} \end{pmatrix}, \tag{9.21}$$

then make a gauge transformation so that ξ^i $(i=1, 2, 3)$ are gauged away,

$$\phi \to \phi' = e^{i(\vec{\xi}\cdot\vec{\tau})/2v}\,\phi = \begin{pmatrix} 0 \\ \dfrac{v+\eta}{\sqrt{2}} \end{pmatrix} \tag{9.22}$$

$$\mathrm{L} \to \mathrm{L}' = e^{i(\vec{\xi}\cdot\vec{\tau})/2v}\,\mathrm{L} \tag{9.23}$$

$$\vec{\mathrm{A}}_\mu \to \vec{\mathrm{A}}'_\mu \tag{9.24}$$

where

$$\vec{\tau}\cdot\vec{\mathrm{A}}'_\mu = e^{i(\vec{\xi}\cdot\vec{\tau})/2v}\left[\vec{\tau}\cdot\vec{\mathrm{A}}_\mu - \frac{i}{g}\,e^{-i(\vec{\xi}\cdot\vec{\tau})/2v}\,\partial_\mu\, e^{i(\vec{\xi}\cdot\vec{\tau})/2v}\right] \cdot e^{-i(\vec{\xi}\cdot\vec{\tau})/2v} \tag{9.25}$$

and B_μ and R are unchanged. We will drop the primes on L' and A'_μ since the transformed fields are just as good as the old ones.

Now the interaction Lagrangian Eq. (9.18) becomes

$$\begin{aligned}\mathcal{L}_{\mathrm{int}} &= -f\left[\bar{e}_{\mathrm{R}}\left(0,\ \frac{v+\eta}{\sqrt{2}}\right)\begin{pmatrix}\nu_{e_{\mathrm{L}}} \\ e_{\mathrm{L}}\end{pmatrix} + (\bar{\nu}_{e_{\mathrm{L}}},\ \bar{e}_{\mathrm{L}})\begin{pmatrix} 0 \\ \dfrac{v+\eta}{\sqrt{2}}\end{pmatrix} e_{\mathrm{R}}\right] \\ &= -f\left[\frac{v}{\sqrt{2}}\,(\bar{e}_{\mathrm{R}}\,e_{\mathrm{L}} + \bar{e}_{\mathrm{L}}\,e_{\mathrm{R}}) + \frac{1}{\sqrt{2}}\,(\bar{e}_{\mathrm{R}}\,e_{\mathrm{L}} + \bar{e}_{\mathrm{L}}\,e_{\mathrm{R}})\,\eta\right] \\ &= -\frac{fv}{\sqrt{2}}\,\bar{e}\,e - \frac{f}{\sqrt{2}}\,\bar{e}\,e\,\eta\,.\end{aligned} \tag{9.26}$$

The last step comes from

$$\begin{aligned}\bar{e}_{\mathrm{R}}\,e_{\mathrm{L}} + \bar{e}_{\mathrm{L}}\,e_{\mathrm{R}} &= \bar{e}\,\frac{1-\gamma_5}{2}\,\frac{1-\gamma_5}{2}\,e + \bar{e}\,\frac{1+\gamma_5}{2}\,\frac{1+\gamma_5}{2}\,e \\ &= \bar{e}\,e\,.\end{aligned} \tag{9.27}$$

Thus the electron has acquired a mass

$$m_e = \frac{fv}{\sqrt{2}}\,, \tag{9.28}$$

and the neutrino remains massless. The part of the Lagrangian describing the Higgs scalar Eq. (9.16) has become

$$
\begin{aligned}
\mathcal{L}_{\text{scalar}} &= \left[\left(\partial_\mu - i\frac{g'}{2}B_\mu - ig\frac{\vec{\tau}}{2}\cdot\vec{A}_\mu\right)\begin{pmatrix}0\\ \frac{v+\eta}{\sqrt{2}}\end{pmatrix}\right]^+ \\
&\quad\cdot\left[\left(\partial^\mu - i\frac{g'}{2}B^\mu - ig\frac{\vec{\tau}}{2}\cdot\vec{A}^\mu\right)\begin{pmatrix}0\\ \frac{v+\eta}{\sqrt{2}}\end{pmatrix}\right] - V\left(\frac{(v+\eta)^2}{2}\right) \\
&= (0,1)\left[\frac{1}{\sqrt{2}}\partial_\mu\eta + \frac{i}{2}(g'B_\mu + g\vec{\tau}\cdot\vec{A}_\mu)\frac{v+\eta}{\sqrt{2}}\right] \\
&\quad\cdot\left[\frac{1}{\sqrt{2}}\partial^\mu\eta - \frac{i}{2}(g'B^\mu + g\vec{\tau}\cdot\vec{A}^\mu)\frac{v+\eta}{\sqrt{2}}\right]\begin{pmatrix}0\\1\end{pmatrix} \\
&\quad - \frac{1}{2}\mu^2(v+\eta)^2 - \frac{1}{4}\lambda(v+\eta)^4 \\
&= \frac{1}{2}\partial_\mu\eta\,\partial^\mu\eta \\
&\quad + \frac{(v+\eta)^2}{8}\chi_-^+\left[(g'B_\mu + g\vec{\tau}\cdot\vec{A}_\mu)(g'B^\mu + g\vec{\tau}\cdot\vec{A}^\mu)\right]\chi_- \\
&\quad - \frac{1}{2}\mu^2(v+\eta)^2 - \frac{1}{4}\lambda(v+\eta)^4 \qquad (9.29)
\end{aligned}
$$

where $\chi_- \equiv \binom{0}{1}$. The remaining scalar field η has a mass $\mu^2 + 3v^2\lambda = -2\mu^2$ (>0). The quadratic term in the gauge fields is

$$
\begin{aligned}
\mathcal{L}_{\vec{A},B} &= \frac{v^2}{8}\chi_-^+\left[(g'B_\mu + g\vec{\tau}\cdot\vec{A}_\mu)(g'B^\mu + g\vec{\tau}\cdot\vec{A}^\mu)\right]\chi_- \\
&= \frac{v^2}{8}\left[(g'B_\mu - gA_\mu^3)(g'B^\mu - gA^{3\mu})\right. \\
&\qquad \left. + g^2(A_\mu^1 A^{1\mu} + A_\mu^2 A^{2\mu})\right]. \qquad (9.30)
\end{aligned}
$$

If we define

$$
\begin{aligned}
W_\mu^\pm &= \frac{1}{\sqrt{2}}(A_\mu^1 \mp iA_\mu^2) \\
Z_\mu &= \frac{g'B_\mu - gA_\mu^3}{\sqrt{g^2+g'^2}} \qquad (9.31) \\
A_\mu &= \frac{gB_\mu + g'A_\mu}{\sqrt{g^2+g'^2}}
\end{aligned}
$$

then

$$
\mathcal{L}_{\vec{A},B} = \frac{v^2}{8}(g^2+g'^2)Z_\mu Z^\mu + \frac{v^2}{4}g^2 W_\mu^- W^{\mu+}. \qquad (9.32)
$$

It is clear that

$$m_{W^{\pm}} = \frac{1}{2} vg$$

$$m_Z = \frac{1}{2} v\sqrt{g^2 + g'^2} \tag{9.33}$$

and

$$m_A = 0$$

i.e. we will interpret A_μ as the photon, and $W_\mu^{\pm}$ and Z_μ are the charged and neutral intermediate vector bosons.

In order to find the relations between the coupling constants, g, g' in this model and those well-known electromagnetic and weak coupling constants e and G, we may express Eq. (9.12) in terms of $W_\mu^{\pm}$, Z_μ, and A_μ. From the definition Eq. (7.31), we have

$$A_\mu^1 = \frac{1}{\sqrt{2}} (W_\mu^- - W_\mu^+)$$

$$A_\mu^2 = \frac{1}{\sqrt{2}\, i} (W_\mu^- + W_\mu^+) \tag{9.34}$$

$$A_\mu^3 = \sin\theta_W A_\mu - \cos\theta_W Z_\mu$$

$$B_\mu = \cos\theta_W A_\mu + \sin\theta_W Z_\mu$$

where θ_W is called Weinberg's angle defined by

$$\tan\theta_W \equiv \frac{g'}{g}. \tag{9.35}$$

The part containing the coupling between $W_\mu^{\pm}$ and leptons in Eq. (9.12) is

$$\begin{aligned}
&\frac{g}{2} \bar{L} \gamma^\mu (\tau^1 A_\mu^1 + \tau^2 A_\mu^2) L \\
&\quad = \frac{g}{2} \Big\{ (\bar{\nu}_L \gamma^\mu e_L + \bar{e}_L \gamma^\mu \nu_L) \frac{1}{\sqrt{2}} (W_\mu^- + W_\mu^+) \\
&\qquad - i (\bar{\nu}_L \gamma^\mu e_L - \bar{e}_L \gamma^\mu \nu_L) \frac{1}{i\sqrt{2}} (W_\mu^- - W_\mu^+) \Big\} \\
&\quad = \frac{g}{\sqrt{2}} [\bar{\nu}_e \gamma^\mu e_L W_\mu^+ + \bar{e}_L \gamma^\mu \nu_L W_\mu^-].
\end{aligned} \tag{9.36}$$

Comparing with Eqs. (8.6) and (8.10) we obtain

$$\frac{\left(\frac{g}{2\sqrt{2}}\right)^2}{m_W^2} = \frac{G}{\sqrt{2}}$$

i. e.

$$\frac{G}{\sqrt{2}} = \frac{g^2}{8m_W^2} = \frac{1}{2v^2}. \tag{9.37}$$

The part containing the coupling between Z_μ, A_μ and leptons is

$$\begin{aligned}
&-g'\bar{R}\gamma^\mu B_\mu R - \frac{g'}{2}\bar{L}\gamma^\mu B_\mu L + \frac{g}{2}\bar{L}\gamma^\mu\tau^3 A_\mu^3 L \\
&\quad = -\frac{g'}{2}(2\bar{e}_R\gamma^\mu e_R + \bar{e}_L\gamma^\mu e_L + \bar{\nu}_L\gamma^\mu\nu_L)(\cos\theta_W A_\mu + \sin\theta_W Z_\mu) \\
&\qquad + \frac{g}{2}(\bar{e}_L\gamma^\mu e_L - \bar{\nu}_L\gamma^\mu\nu_L)(\sin\theta_W A_\mu - \cos\theta_W Z_\mu) \\
&\quad = \frac{Z_\mu}{2\sqrt{g^2+g'^2}}[g'^2(2\bar{e}_R\gamma^\mu e_R + \bar{e}_L\gamma^\mu e_L + \bar{\nu}_L\gamma^\mu\nu_L) \\
&\qquad\qquad - g^2(\bar{e}_L\gamma^\mu e_L - \bar{\nu}_L\gamma^\mu\nu_L)] \\
&\qquad + \frac{gg'}{\sqrt{g^2+g'^2}} A_\mu[\bar{e}_R\gamma^\mu e_R + \bar{e}_L\gamma^\mu e_L].
\end{aligned} \tag{9.38}$$

Comparing with Eq. (3.36) we can identify the electron's charge $-e$:

$$e = \frac{gg'}{\sqrt{g^2+g'^2}} = g\sin\theta_W = g'\cos\theta_W. \tag{9.39}$$

From Eq. (9.33)

$$m_{W^\pm} = \frac{1}{4}v^2g^2 = \frac{1}{4\sqrt{2}\,G}\frac{e^2}{\sin^2\theta_W} \approx \frac{38}{\sin\theta_W}\,\text{GeV} \tag{9.40}$$

$$\begin{aligned}
m_Z &= \frac{1}{2}v\sqrt{g^2+g'^2} = \frac{vg}{2\cos\theta_W} = \frac{m_{W^\pm}}{\cos\theta_W} \\
&\approx \frac{38}{\frac{1}{2}\sin 2\theta_W}\,\text{GeV}
\end{aligned} \tag{9.41}$$

That is in this model

$$\begin{aligned}
&m_Z > 76\,\text{GeV} \\
&m_{W^\pm} > 38\,\text{GeV}.
\end{aligned} \tag{9.42}$$

The most direct way of testing this model is to look for the intermediate vector bosons. However, their masses are too high to be produced by present accelerators. The possible experimental test now is to observe the neutral current, or Z-exchange effects, particularly in $\bar{\nu}_e e$ or $\bar{\nu}_\mu e$ elastic scattering. The present data neither support nor reject the Weinberg-Salam model.

X. CONCLUSION

We have studied here an original model of a wide class of unified gauge theories. A variety of other models based on the same notion have been proposed. The main differences among these models are in the choice of gauge group and classification of leptons. Present experimental data are not only unable to select one out of these various models, but also unable to justify those basic ideas like local gauge invariance and spontaneous symmetry breaking. However, from the theoretical point of view, this kind of theory is by now the most compelling one, both for aesthetic and practical reasons.

For a complete theory, it is imperative to extend these models to include the strongly interacting particles. By now, no one specific model of these extensions is regarded as natural enough.

ACKNOWLEDGEMENTS

I am grateful to Profs. H. Hesselfeld and U. E. Schnaus for carefully reading this manuscript. Also I would like to thank my wife, Jen-Mei, for many helpful suggestions and typing the manuscript.

REFERENCES

(1) B. W. Lee, in *Proceedings of the XVI International Conference on High Energy Physics*, ed. by J. D. Jackson and A. Roberts (National Accelerator Laboratory, Batavia, Ill. 1972), Vol. IV, p. 249; E. S. Abers and B. W. Lee, *Phys. Rep.* **9**, 1 (1973); S. Weinberg, *Rev. of Mod. Phys.* **46**, 255 (1974).

(2) See, e. g., H. A. Bethe and R. W. Jackiw, *Intermediate Quantum Mechanics.*

(3) C. N. Yang and R. Mills, *Phys. Rev.* **96**, 191 (1954); R. Utiyama, *Phys. Rev.* **101**, 1597 (1956); M. Gell-Mann and S. Glashow, *Ann. Phys.* **15**, 437 (1961).

(4) J. Goldstone, *Nuovo Cimento* **19**, 15 (1961).
(5) P.W. Higgs, *Phys. Rev. Letters* **12**, 132 (1964); G.S. Guralnik, C.R. Hagen and T.W.B. Kibble, *Phys. Rev. Letters* **13**, 585 (1965); P.W. Higgs, *Phys. Rev.* **145**, 1156 (1966).
(6) See, e.g., T.D. Lee and C.S. Wu, Weak Interactions, *Ann. Rev. Nucl. Sci.* **15**, 381 (1965); S. Gasiorowicz, *Elementary Particle Physics*, John Wiley and Sons, New York 1966.
(7) S. Weinberg, *Phys. Rev. Letters* **19**, 1264 (1967); A. Salam, in *Elementary Particle Theory*, ed. N. Svartholm, Almquist and Forlag, Stockholm, 1968.

NEW QUASIPARTICLE EXCITATION SPECTRUM FOR ACOUSTIC ATTENUATION IN THE INTERMEDIATE STATE

FONG-JEN LIN

ABSTRACT

Quasiparticle scattering at N—S intersurfaces in the intermediate state leads to a modified quasiparticle energy spectrum. This modification has recently been found by the author and Professor Jack R. Leibowitz in the acoustic attenuation of Indium near T_c at The Catholic University of America. A theoretical model is given which accounts for the observed attenuation structure near T_c($T_c - T \lesssim$ 5 mK) in terms of the modification of density-of-state $\rho(E)$ produced by an NSNS..."superlattice". This model predicts further observable effects.

The density-of-state is modified from BCS due to quasiparticle scattering in the SNSN..."superlattice" in the intermediate state of a superconductor.

These new features were found in ultrasonic attenuation $\alpha_t(T)$ in Indium near T_c, the superconducting transition temperature. In Fig. 1 an oscillatory $\alpha_t(T)$ is seen for $\delta T = T_c - T \lesssim 5$ mK (millidegree K). In the range of interest in Fig. 1 the acoustic phonon energy $\hbar w$ is of order 0.01 $\Delta(T)$, the BCS[1] gap, and furthermore, $\hbar w$ is comparable with the band width and separation in the new quasiparticle energy spectrum just above $\Delta(T)$, so that the modified density-of-states can have a measurable effect on the observed $\alpha_t(T)$ due to electrons. The observed $\alpha_t(T)$ structure is found to be in good agreement with that determined from a theoretical model using the new quasiparticle density-of-states.

The intermediate state of a superconductor consists of periodic NSNS...(N: normal, S: superconducting) regions. Therefore, ultrasonic attenuation is expected to be different from both pure superconducting and normal states. The theoretical model for intermediate state acoustic attenuation was given by Andreev and Bruk[2]. It

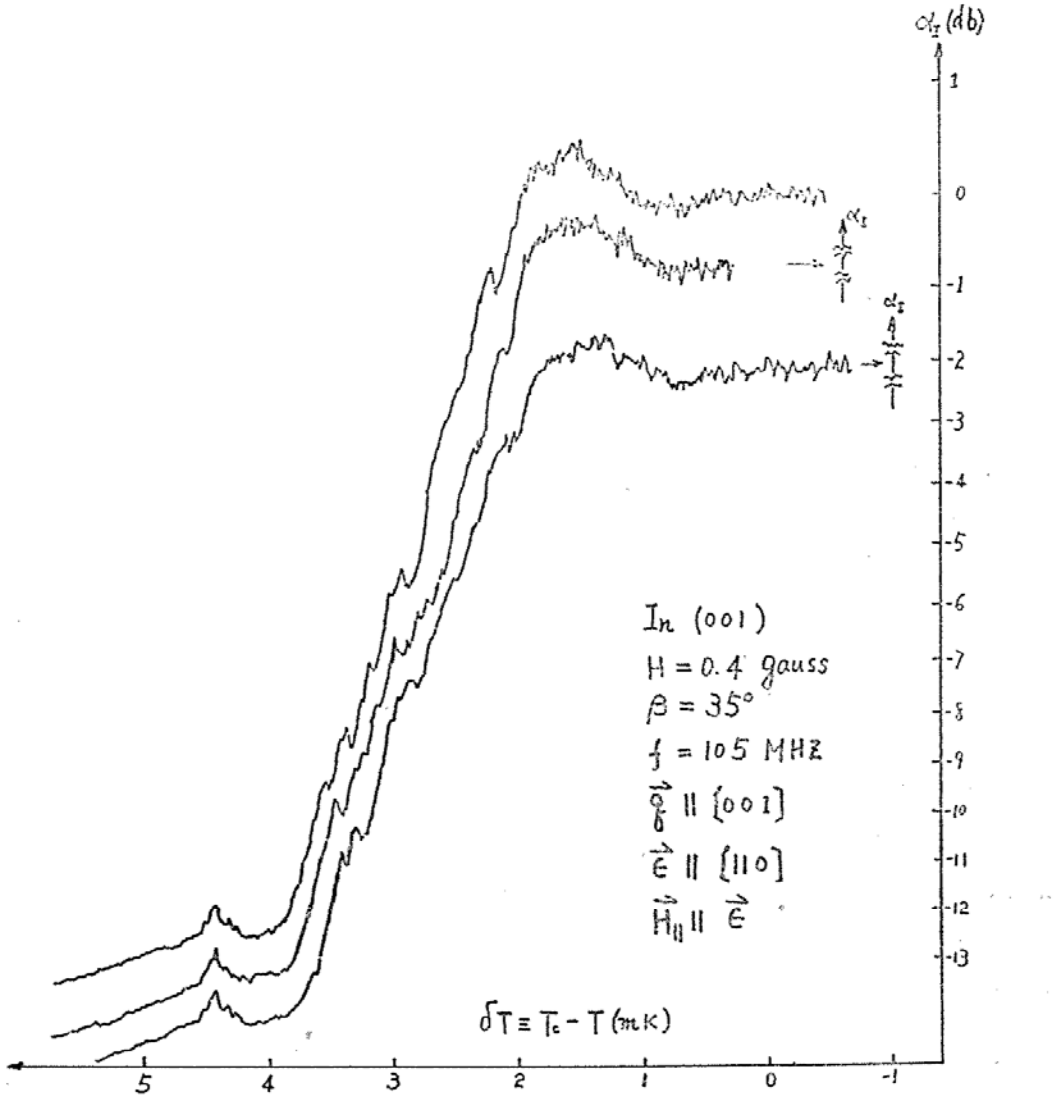


Fig. 1. Intermediate attenuation versus δT for a 105 MHZ transverse wave

takes into account the Joule heating due to N—S wall motion, and has been in resonable agreement with previous experimental studies[3,4]. However, our experimental condition is very different from previous experimental studies. Our whole intermediate state is in the temperature range very close to T_c. The value of d_n/d ranges from 1 to 0.2 for $\delta T(\equiv T_c - T)$ of 0 to 5 mK. Here d_n is the width of the normal lamina and d is that of an N—S pair. In this restricted temperature range, $\hbar w \gtrsim 0.01\Delta(T)$. To achieve this restricted condition a very small applied magnetic field $H<1$ gauss is necessary.

The ultrasonic shear wave attenuation of Indium in the intermediate state was studied as a function of δT, where the range of

δT is of the order of 10 mK. In this small temperature interval the temperature was swept at a rate of less than 3 mK per minute with the sample directly immersed in liquid helium. The intermediate state was produced by the application of a slanting magnetic field of magnitude 0.4 gauss at a direction 35° from the sample surface. Transverse acoustic waves of frequency 105 MHz and 75 MHz had a propagation vector $\vec{q}$ perpendicular to the surface of disk ((001) plane of Indium). The shear wave polarization vector $\vec{\varepsilon}$ was aligned in the [110] direction of Indium. The N—S surfaces of the intermediate state were arranged either parallel or perpendicular to the direction of polarization of the shear wave. The features of Fig. 1 were consistently repeated, even to the relative positions on the δT scale of the attenuation structure.

In addition to the Andreev's model there are other theories[5] of ultrasonic attenuation in the intermediate state, namely, (1) the geometrical absorption, treating the intermediate state as a combination of pure superconducting regions and pure normal regions. The absorption is determined entirely by the relative amount of normal and superconducting material sampled by the ultrasonic beam, and (2) the absorption due to electrons orbiting in the magnetic field. In our case the Larmor radius R is much larger than the mean free path of electrons, and the wave length of the acoustic wave is much less than the width of the normal region. Accordingly the result of (2) is the same as that of (1).

The theories existing so far can not explain our experimental result because they fail to recognize the periodic N—S laminas as a whole. Since in the intermediate state there is a periodic array consisting of alternating normal and superconducting laminas, it is not correct to investigate quasiparticle properties separately either in the normal or in superconducting region. In other words, neither the normal nor the superconducting region in the intermediate state is equivalent to the pure normal or superconducting state, because the electrons see the N—S laminar structure as a whole, unless the mean free path l is much less than the thickness of the layers d_n and d_s. In that case, the electrons can only see either the normal

region or the superconducting region.

However in our Indium sample in the liquid helium temperature range, the mean free path of electrons is of the order of 0.1 cm and our laminar spacing $d(\equiv d_n+d_s)$ is also of the order of 0.1 cm. Hence a significant fraction of the total number of electrons in the effective zone has a mean free path larger than the laminar spacing d. Therefore it is essential to treat the periodic array of N—S laminas as a whole. When this is done we expect to find the new energy spectrum to be different from that for both pure superconducting or the normal state, and an entirely new density of energy states will result. In turn this new density-of-state is expected to lead to new features of the attenuation, which are not observed when the BCS density of states is used.

The problem of energy spectrum of the intermediate state is equivalent to the energy spectrum of excitations associated with a one-dimensional step potential. We face to solve the Bogolubov equation of "electron-hole" coupled excitation. Van Gelder[6] has made a calculation of the energy spectrum of "electron-hole" coupled excitation and found the following solution for the simplified Bogolubov equation.

$$\begin{bmatrix} \left(\frac{p^2}{2m}-E_F\right) & \Delta(x,\,T) \\ \Delta(x,\,T) & -\left(\frac{p^2}{2m}-E_F\right) \end{bmatrix} \begin{bmatrix} u(r) \\ v(r) \end{bmatrix} = E \begin{bmatrix} u(r) \\ v(r) \end{bmatrix} \qquad (1)$$

where $\Delta(x,\,T)$ is the pair potential or energy gap in the superconductor, such that

$\Delta(x,\,T)=\Delta(T)$ in the superconducting region,

and

$\Delta(x,\,T)=0$ in the normal region. (2)

E_F is the Fermi energy, here $\begin{bmatrix} u \\ v \end{bmatrix}$ is the two component wave function of the "electron-hole" paired excitation with eigenvalue E, subject to the normalization condition

$$u^2+v^2=1 \qquad (3)$$

There are two special cases:

$u=1$, $v=0$ for E>0 (electron-like); and

$u=0$, $v=1$ for E<0 (hole-like).

Setting

$$\varepsilon=E/\Delta \tag{4}$$

we have the solution for $\varepsilon>1$

$$\cos[(K-k_x)d_x]=\cos(A\varepsilon)\cos[B(\varepsilon^2-1)^{1/2}] -\left[\frac{\varepsilon}{(\varepsilon^2-1)^{1/2}}\right]\sin(A\varepsilon)\sin\left[B(\varepsilon^2-1)^{1/2}\right] \tag{5}$$

and for $\varepsilon<1$

$$\cos[(K-k_x)d_x)=\cos(A\varepsilon)\cosh[B(1-\varepsilon^2)^{1/2}] -\left[\frac{\varepsilon}{(1-\varepsilon^2)^{1/2}}\right]\sin(A\varepsilon)\sinh[B(1-\varepsilon^2)^{1/2}]. \tag{6}$$

K is the Bloch momentum along the X-axis (i.e. normal to the laminas), k_x is given by

$$k_F^2=k_x^2+k_y^2+k_z^2 \tag{7}$$

d_x is the specimen dimension in x-direction, and

$$A=dn/\alpha$$

$$B=ds/\alpha$$

$$\alpha=\frac{\hbar^2 k_x}{m\Delta}.$$

Writing equations (5) and (6) symbolically as

$$F(k)=F(\varepsilon)$$

the range of value of K is

$$-1\leq F(K)\leq 1$$

Therefore, the allowed energies are such as to confine $F(\varepsilon)$ in the region

$$-1\leq F(\varepsilon)\leq 1$$

In other words, ε values such that $|F(\varepsilon)|>1$ are forbidden.

From numerical solutions, the allowed band width and band separation are of the order of 0.01 $\Delta(T)$. The allowed energies below the gap $\Delta(T)$ are narrow and rather widely separated lines, the spacing between lines being of order 0.1 $\Delta(T)$.

The transverse wave attenuation α_t by electrons in Indium near T_c when $ql>1$, as in the present case, is given by $\alpha_t=\alpha_{tE}+\alpha_{tD}$ where α_{tD} is the contribution from electromagnetic interaction, and is arising

from shear deformation of the Fermi surface. At temperatures very close to T_c, both α_t and the variation of α_t are dominated by electromagnetic interaction. Therefore, we shall consider only attenuation due to electromagnetic interaction. The ratio of α_s (attenuation in the superconducting state) to α_N (attenuation in the normal state) is[7]:

$$\frac{\alpha_s}{\alpha_N}=\frac{\frac{\sigma_{1s}}{\sigma_{1n}}(\gamma^2+1)}{\left(\gamma+\frac{\sigma_{2s}}{\sigma_{1n}}\right)^2+\left(\frac{\sigma_{1s}}{\sigma_{1n}}\right)^2} \tag{11}$$

where

$$\gamma=\frac{q^2c^2}{4\pi w\sigma_{1n}} \tag{12}$$

σ_1 and σ_2 are the real and imaginary part of conductivity σ. For temperatures very near T_c, in rough approximation

$$\frac{\alpha_s}{\alpha_N}\approx\frac{1}{\left(\frac{\sigma_{1s}}{\sigma_{1N}}\right)} \quad \text{or} \quad \frac{\delta\left(\frac{\alpha_s}{\alpha_N}\right)}{\left(\frac{\alpha_s}{\alpha_N}\right)}\approx-\frac{\delta\left(\frac{\sigma_{1s}}{\sigma_{1N}}\right)}{\left(\frac{\sigma_{1s}}{\sigma_{1N}}\right)} \tag{13}$$

The expression for $(\sigma_{1s}/\sigma_{1N})$ appropriate to the electromagnetic absorption is given by Mattis and Bardeen[8]:

$$\frac{\sigma_{1s}}{\sigma_{1N}}=\frac{2}{\frac{\hbar w}{\Delta}}\int_1^\infty \rho(\varepsilon)\,\rho\left(\varepsilon+\frac{\hbar w}{\Delta}\right)\left[1+\frac{1}{\varepsilon\left(\varepsilon+\frac{\hbar w}{\Delta}\right)}\right]$$
$$\cdot[f(\varepsilon)-f(\varepsilon+\hbar w)]d\varepsilon \tag{14}$$

where f is the Fermi function.

Calculating the density-of-state from $\rho(\varepsilon)=\frac{\hbar v_F}{\Delta}\frac{dK}{d\varepsilon}$ and numerically integrating eq.(14), $(\sigma_{1s}/\sigma_{1N})$ versus δT is no longer a smooth, monotonicly decreasing curve as in the case of the BCS calculation.

The line-like energy bands in the region $0<\varepsilon<1$ do not make direct contribution in calculating $(\sigma_{1s}/\sigma_{1N})$, because these lines are found to be much narrower than $\hbar w$ and have the spacing of about $10\hbar w$. It is also found that for $\varepsilon>1$, the influence of the SNSN... superlattice becomes important. The computations find sharp bands (band width $<\hbar w$) for $1<\varepsilon<1.1$, with band separation $\sim\hbar w$. With increasing ε the band widths increase, separation remaining of order

$\hbar w$ up to the value $\varepsilon=3$. The above mentioned modification in the spectrum is important. From eq.(14) we realize that in order to have an oscillatory variation of σ_{1s}/σ_{1N} within the range δT of a few millidegrees, two requirements must be satisfied, namely (1) the density-of-state must vary appreciably within that range of δT, and (2) $\rho(\varepsilon)$ must vary appreciably in energy of the order $\hbar w$. The new modified spectrum satisfies these two requirements. It is expected that σ_{1s}/σ_{1N} and hence α_s/α_N have oscillatory structure, and therefore the relation connecting σ_{1s}/σ_{1N} and α_s/α_N suggests that the dips in σ_{1s}/σ_{1N} versus δT will correspond to peaks in the temperature

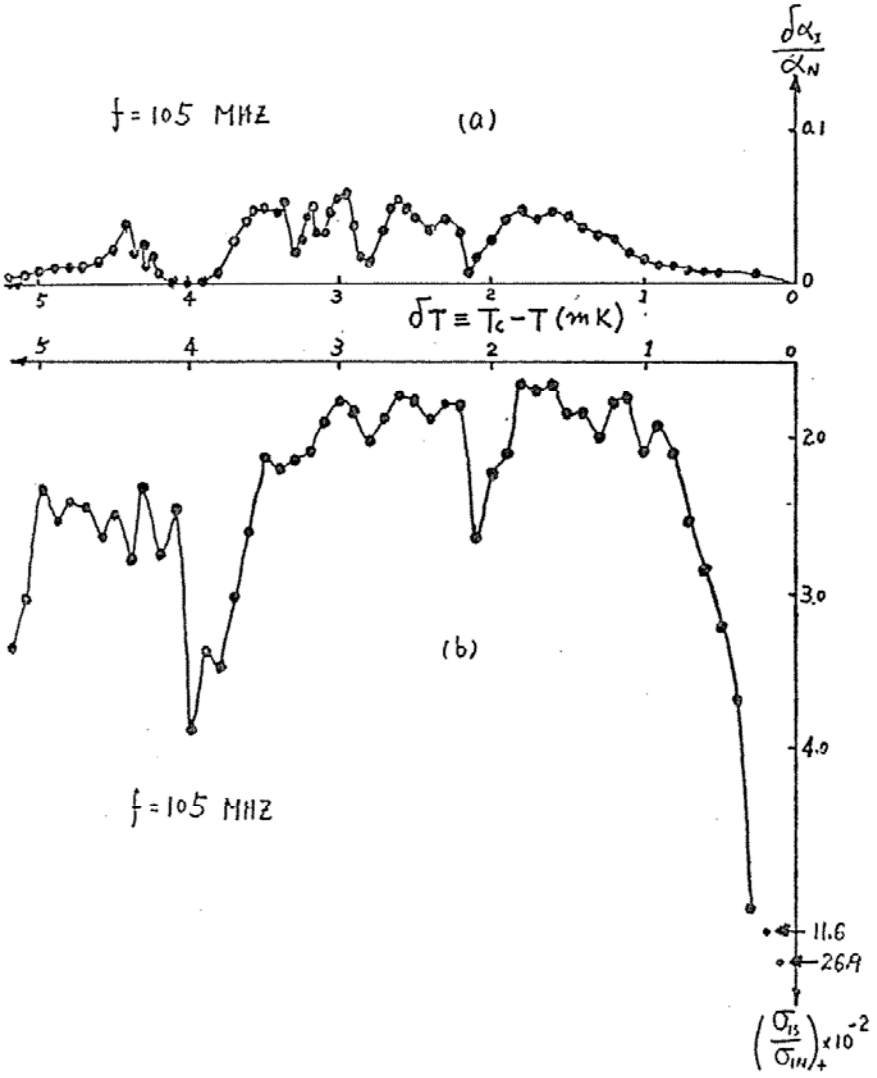


Fig. 2. Attenuation structure and conductivity ratio versus δT. (In Fig. (b) the indicated points are computed from eq. (14) up to $\varepsilon=5$).

dependence of α_s/α_N. In Fig. 2 the observed temperature dependence for the attenuation structure (arbitrary baseline) is compared with that for in the temperature range $0 \leq \delta T \leq 5$ mK. The calculation for σ_{1s}/σ_{1N} versus δT, in Fig. 2(b) was carried up to $E/\Delta=5$. It is seen that the observed attenuation structure coincides rather well with that calculated for σ_{1s}/σ_{1N} from the new theoretical model. The structure in attenuation disappears for values of $\delta T > 5$ mK, while the calculation of σ_{1s}/σ_{1N} versus δT suggests that the structure persists to the value of δT greater than 5 mK. However, the calculation assumed a demagnetization factor of 1, whereas the disk had a diameter to thickness ratio of less than 10:1. Even on assumption of an "infinite plate" geometry, for $H=0.4$ gauss, the δT range 0 to 5 mK corresponds to an H/H_c value from 1 to 0.2; and independent studies[9] of the intermediate state have demonstrated that, even for the slanting applied fields, well ordered laminas are not commonly formed for H/H_c values below about 0.2.

When $ql>1$, as it is in the present discussion, the electron-phonon coupling is restricted to the so-called effective zones, which are normal to $\vec{q}$. These electrons which are able to contribute to the attenuation are precisely those which lie in planes normal to the S—N intersurfaces and which, accordingly, suffer drastic change in the density-of-state. Since l and d are of the same order of magnitude in these present measurements, a significant fraction of the relevant electrons would be expected to respond to the BCS density-of-states or to enter bound states in a single N lamina. This gives the monotonic decreasing attenuation versus δT. Those electrons with mean-free-path greater than d would give rise to the oscillatory structure as reported here, superimposed on a monotonicly decreasing baseline.

The modification of the density-of-state should also manifest itself in α_{tD} and α_{LD}, i.e. attenuation associated with the transverse and longitudinal deformation potential. It is determined[10] that, at T_c, α_{tD} represents about 1/3 of the total attenuation in Indium at 105MHZ. It has also been shown that

$$(\alpha_s/\alpha_N)_{tD} = (\alpha_s/\alpha_N))_{LD} = (\sigma_{1s}/\sigma_{1N})_D. \tag{15}$$

Here the ratio equals $2f(\Delta)$ when the BCS density-of-states is applied[1]. Generally we have

$$\left(\frac{\sigma_{1s}}{\sigma_{1N}}\right)_D = \frac{2}{\frac{\hbar w}{\Delta}} \int_1^\infty \rho(\varepsilon)\, \rho\left(\varepsilon + \frac{\hbar w}{\Delta}\right) \left[1 - \frac{1}{\varepsilon\left(\varepsilon + \frac{\hbar w}{\Delta}\right)}\right] \cdot [f(\varepsilon) - f(\varepsilon + \hbar w)]\, d\varepsilon \tag{16}$$

This is different from eq.(14) by a minus sign in the so called coherence factor. The $(\sigma_{1s}/\sigma_{1N})_D$ amplitudes are smaller by a factor of about 10^{-1} (for computation up to $\varepsilon=3$) compared to those of eq.(14). Hence, the deformation contribution can be negleted in the treatment of transverse wave attenuation structure. However, the foregoing results suggest a striking prediction. For the case of longitudinal waves, where only deformation is important, the modified density-of-states should produce a sharply decreased attenuation immediately below T_c, where the forementioned two requirements are satisfied, and with further reduction in temperature, the attenuation should soon return to the BCS curve $\alpha_s/\alpha_N = 2f[\Delta(T)]$. Thus, a "slot" near T_c in the α_s/α_N versus T curve is predicted for the case of longitudinal waves under intermediate state conditions if $l \gg d$.

REFERENCES

(1) J. Bardeen, L. N. Cooper, and J. R. Schrieffer, *Phys. Rev.* **108**, 1175 (1957).
(2) A. F. Andreev and Yu, M. Bruk, *JETP* **23**, 942 (1966).
(3) M. Gottliev, C. K. Jones and M. Gerbuny, *Phys. Rev. Lett.* **24A**, 585 (1967).
(4) O. Sigh and M. H. Jericho, *Phys. Rev.* **5**, 3542 (1972).
(5) M., Gottliev, C.K. Jones and M. Garbuny 1970, *Physical Acoustics vol. VIII* ed. W. P. Mason (New York: Academic Press) pp. 1-49.
(6) A. P. Van Gelder, *Phys. Rev.* **181**, 787 (1969).
(7) K. Fossheim and B. Torvain, *J. Low Temperature Phys.* **1**, 341 (1969).
(8) D. C. Mattis and J. Bardeen, *Phys. Rev.* **111**, 412 (1958).
(9) T. E Faber, *Proc. Roy. Soc.* (*London*) **A248**, 460 (1958).
(10) K. Fossheim, *Phys. Rev. Lett.* **19**, 81 (1967).

"To smash the little atom
All mankind was intent;
Now every day
The Atom may
Return the compliment."

MAX BORN
Nobel Prizewinner in
Physics (1954)

VISCOSITY MEASUREMENTS OF VITREOUS MATERIALS: SILICATE GLASSES

URBAN E. SCHNAUS

ABSTRACT

Measurements of viscosity of silicate glasses by beam-bending and parallel-plate methods are reported in the range 10^5 to 10^{14} poises at temperatures between 300°C and 1000°C. Data for binary Na_2O–SiO_2 compositions of low mole-percent soda concentration indicate viscocity changes due to phase separation. Binary K_2O–SiO_2 compositions show some anomaly for concentrations less than 8% K_2O, but do not show a marked change in viscosity at temperature points recently reported for the miscibility gap for this system. Ternary and four-component silicate systems show systematic variation due to network modifying oxides.

Extensive new data are given for binary borosilicate systems with 10% to 50% silica.

INTRODUCTION

A. Factors involved in viscosity determinations

Since viscosity determines the melting conditions for vitreous materials, the temperatures for working and annealing them, the occlusion of gaseous particles in them, and the possibility of devitrification (crystallization) while handling them, it stands as one of the most important technological properties to be studied. Standard values for the working point, the softening point, and the annealing point of commercial glasses and similar amorphous materials are expressed in terms of the viscosity η, of which the c. g. s. unit, the poise (P), has dimensions gm sec^{-1} cm^{-1}, and which is defined from the equation

$$\eta = \frac{Fd}{Av} \qquad (1)$$

where F is the tangential (shearing) force acting, d the distance between two parallel planes in the material, A the area of these planes, and v the relative velocity of motion of the two planes.

For low viscosities (fairly conveniently up to 10^5 poise; with modifications to 10^8 poise) the rotating concentric-cylinder method used commonly for ordinary fluids like oil and water can also be adapted for vitreous materials at their necessarily higher temperatures and viscosities. In this method there is direct shear stress over a large area of sample.

Higher viscosities of glassy materials (10^8 to 10^{13}) poise have been most frequently measured by the fiber elongation method, which involves the rate of elongation of a rod of vitreous material under a fixed force. The equation

$$\eta = \frac{h\,m\,g\,k}{3\pi R^2\,dh/dt} \qquad (2)$$

applies to this determination: h is the length of the rod, m the load mass, g the gravitational constant, R the radius of the rod (generally less than 1 mm), dh/dt is the rate of elongation, and k a calibration constant for the instrument.

Experience has shown that within wide limits vitreous materials can be considered Newtonian liquids in which the flow rate is directly proportional to the stress. Hence the viscosity is considered to be independent of the magnitude of the load, and to be constant in time; thus having a fixed value for any given equilibrium temperature. In fiber elongation the stress is longitudinal and the viscosity measurement is less direct than for the rotating cylinder method.

Several other methods for measurements in the higher viscosity range have been developed. The two used in obtaining data given in this report were developed principally for materials that could not be drawn conveniently into fibers from a melt crucible[1][2]. The somewhat simplified versions used for the present data have been described in earlier issues of this journal[3][3a]. Another method claimed to be of comparable accuracy in this range[4] uses a loaded-rod penetrator of metal, sometimes less than 0.1 mm in diameter. For this method the equation is

$$\eta = \frac{m\,g\,(\ln \frac{h}{R} - 0.72)}{2\pi h\,(dh/dt)} \qquad (3)$$

In this case the penetration depth is h, and the other symbols are as in equation (2). For the fiber elongation and penetrometer methods, dh/dt is usually gotten from cathetometer-type measurements.

In all four of these higher-range methods of viscosity determination the basic measurements are of a linear dimension and of its time rate of change. Beam-bending and parallel-plate viscometers use an electronic device, a linear variable deflection transformer (LVDT), that supplies a signal directly proportional to change of length to a potentiometric recorder, so that dh/dt can be gotten conveniently from the recorder chart. In the parallel-plate instrument the equation involves higher powers (h^3 and h^5) of the length dimension; this affects the precision somewhat. Measurements with this instrument are most conveniently done at increasing temperature. The other methods can be used conveniently at equilibrium temperature, though the A S T M procedure for determining the annealing point by beam bending specifies decreasing temperature about 5 C°/min. With penetrometer and parallel-plate methods the stress is longitudinal; in beam-bending it is more like shear.

It is obvious that extraneous factors may influence the calculated values of viscosity obtained from all these methods. There may be friction where the samples are supported and where they are loaded. Also there may be non-uniform expansion of support rods. Temperature gradients can usually be minimized, but they are likely in all methods. In the case of fiber elongation, where the measurements of dh/dt depend entirely on tensile forces set up in a long thin rod, surface effects may cause some error. And in some cases where phase separation has been observed in a sample and two components of radically different properties seem to be produced, results of fiber elongation measurements may become quite unreliable. Viscosity measurments in the range 10^8 to 10^{11} poise are sometimes omitted for compositions where fiber elongation measurements have been found to have poor reproducibility[5]. In this case the small diameter of the fiber probably introduces a severely limiting factor.

Indication of this can be found with glass j in Fig. 1 of this report (binary soda silicate glass, with 15 mole % Na_2O) where the

η vs T relationship is much less smooth than usually shown. The light broken line added between $\log \eta = 9$ to $\log \eta = 13$, crossing the *s*, *j*, and *x* lines of the figure, is taken from the frequently-cited 1949 paper of Poole[6] for this composition.

B. Phase separation and viscosity

That phase separation can influence many different properties of glasses is becoming more widely known. The chemical durability of four-component borosilicate glasses widely used for laboratory work (e.g. Pyrex) and usually having boric oxide component between 12 and 24 weight %, with 6 to 8% alumina and soda, is now known to be due to phase separation, which is at such a fine scale that the microscopically finer borosilicate component cannot be leached out by water and most other common solvents. When subsequent heat-treatment is used, as in the case of manufacture of Vycor, the two liquid phases can be separated and the remaining silica-rich component ("thirsty" glass having open structure) can then be sintered to form, at temperature about 1000°C, a 96% silica glass without having to use temperatures near the 1700°C range required to fuse ordinary quartz.

Much research activity now centers around miscibilities for different systems of compositions, especially oxides, and various microscopic theories have been proposed to explain them[7][8]. Some of the research materials used for this report were made with a view of studying the critical viscosity anomaly found with them and the effect that this has on light scattering in them. In the earlier work a major objective was to determine more closely the fictive temperature at which density fluctuations are frozen into the glass structure. This temperature is near the annealing point, which is usually taken as $10^{13.6}$ poise. Both beam-bending and fiber-elongation procedures have been specified by the American Society for Testing and Materials for determing this point. The circled points on the 10^{13} poise line of the Arrhenius plots of this report are from beam-bending measurements using the ASTM method.

EXPERIMENTAL PROCEDURE

A. Sample preparation

Standard methods of preparation have been used, with Pt crucibles in electrically-heated furnaces melting mixtures of quartz sand and required amounts of additive reagent-grade carbonates and oxides. Melting and fining times were at least 6 hours, with special precautions for mixing and fining borate compositions to avoid water contamination. Melts were quenched by pouring onto steel plates arranged to form rectangular bars, which were then cut and ground to required size and shape for the beams and pellets needed. Many of these materials are highly hygroscopic and needed to be kept in dessicators after cooling to room temperature and being coated with oil or paraffin. In spite of these precautions there is some question about the data for 100% boric oxide (B_2O_3), and the 90% and 80% compositions of Fig. 3, glass 14, 15, and 16, as mentioned later in the discussion.

B. Method of measurement

Most of the data of this report refer to work with a parallel-plate viscometer, which uses samples in cylindrical form, about 6 mm diameter and 4 to 6 mm high. These pellets are compressed between parallel plates of Inconel metal; the viscosity is calculated from the rate at which the pellet softens and flows under vertical uniaxial compression.

For many ordinary glass compositions the viscosity values obtained by the parallel-plate method agree very well with those obtained by fiber elongation. For amorphous materials with which drawing from melts is not possible, the parallel-plate and penetrometer methods are eminently useful. Moreover, with many materials where phase separation has been found, a critical viscosity anomaly shows itself in rapid change in the rate of softening as the temperature changes. This requires use of a method that can follow more rapid change than usually measured in fiber elongation. With glasses such as the 15% binary soda silicate one (glass *j*, Fig. 1) or the

four-component borosilicate one (glass 20*, Fig. 3) the η vs T plots from several runs of the same composition show quite satisfactory reproducibility.

C. Data analysis

1. *Parallel-plate work*

At least three samples of each composition were used; in some cases as many as nine pellets of a given composition were compressed, each giving a plot of η vs T on semi-log graph paper. Heating current was regulated to increase the temperature of the sample at a rate about 5 C°/minute; the input terminals of the recorder were shorted momentarily at intervals 6 to 8 C° apart as the height of the pellet decreased under the applied vertical load. From the plot for each run values 10 C° apart were read, and a summary was prepared for each composition. On this summary average values of η were calculated for appropriate temperatures, and these averages were then used to make the Arrhenius plots of $\log_{10} \eta$ vs $10^3/T$ °K for the groups of compositions.

The Arrhenius equation for materials of this type has the form

$$\eta = \eta_0 \exp\left(\frac{E}{RT}\right) \tag{4}$$

where η_0 and E are considered to be temperature-independent coefficients called respectively the pre-exponential factor and the activation energy; R is the universal gas constant, and T the temperature in degrees Kelvin. The corresponding linear equation is

$$\log_{10} \eta = a + \frac{b(10^3)}{T}. \tag{5}$$

Least-square methods are often used to calculate the values of a and b from the linear portion of the plots.

However, it is evident from the figures given that most of the parallel-plate measurements here reported extend beyond this linear range. The NBS certificates for the standard glass compositions used for frequent checking in this work: No. 710 (soda-lime: glass *s*) and No. 717 (borosilicate: glass *s'* on Fig. 3 only), give Fulcher equation values having a third constant T_0, which is subtracted from

T in the denominator of equation (5). This enables making a fit covering the range from 10^2 to 10^{12} poise. The overall accuracy of measurements here reported is estimated about ± 5 C°.

2. *Beam-bending*

In these determinations bars of rectangular cross-section, 3 to 6 mm on a side and 6.5 cm long have been used; the method is better suited for high viscosities, 10^{10} to 10^{13} poise, and results show good consistency for annealing point determinations at decreasing temperature. Measurements at equilibrium temperatures show considerably larger scatter. But the rate of change of η can be determined quite well from summary graphical plots of η vs T for the two compositions, 15% and 20% Na_2O binary silicates, that are of particular importance in this report.

In the case of the 20% Na_2O binary composition, first plots showed nice linear Arrhenius behavior for the upper section of the pellet data and for the higher points in the beam-bending range. But there was a wide gap between the two lines. Further investigation in which the beam data was extended to higher temperatures showed that the linear behavior did not extend to these higher temperatures. Points for a curved connecting line could be determined to connect the beam data with the parallel-plate ones, joining the latter at temperature about 600°C.

The fact that the pellet data, made at increasing temperature, does not show a sharp rise at this point is probably due to a hysteresis effect, accompanied by the fact that the sensitivity of the parallel-plate apparatus is lower at lower temperatures where there is little change in height of the pellet. Additional runs with some change of pellet height and lower rate of temperature may yield new values here.

It may be noted, however, that the temperatures: 600°C for the 20% composition and 750°C for the 15% composition, where the sharp change in viscosity is observed, are also the points on the miscibility gap boundary line plotted for these compositions (Ref. 7, Fig. 1B, pg. 122).

DISCUSSION

A. Binary soda-silicate compositions

Figure 1 is a composite one incorporating data reported previously for 15, 25, 33, 38, 40, and 42 mole % Na_2 with SiO_2 along with new results: glasses *x*, *y*, and *z*, 20%, 10%, and 8% respectively. Some additional runs of the 15% composition have yielded results slightly different from what was previously shown; data here given are now more complete for this composition now well known to show phase separation. The critical viscosity anomaly is here evidenced by the separation between the linear part of the pellet data and that of the beam data for higher viscosities.

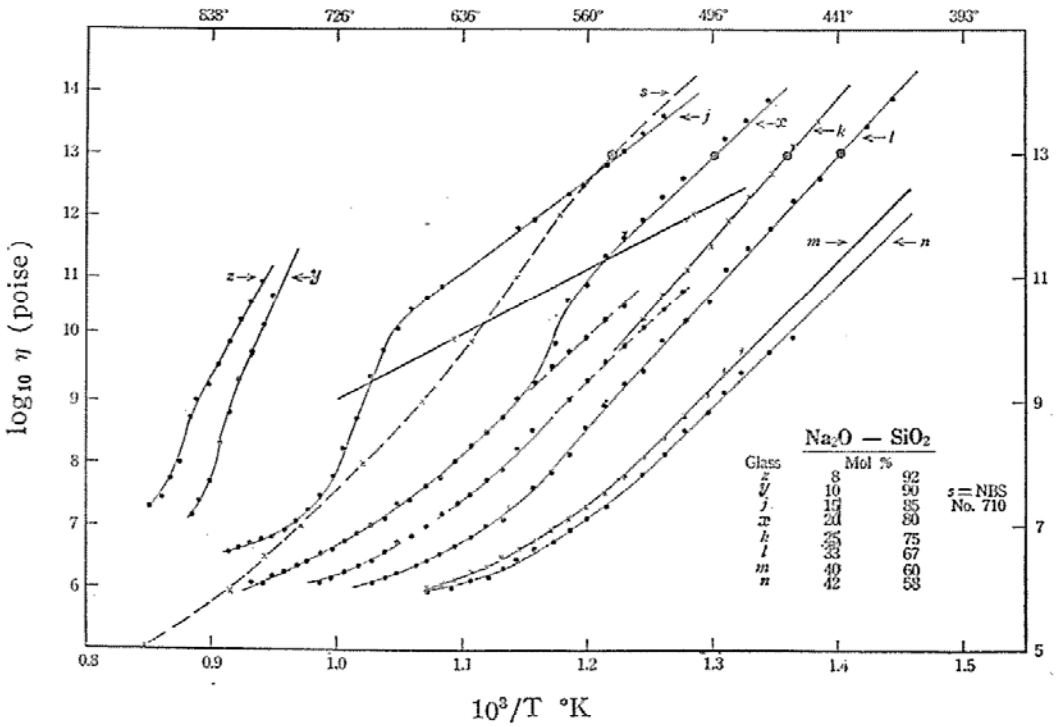


Fig. 1. Viscosity vs $10^3/T°K$ Na_2O-SiO_2 8-42 mol %

For the *y* and *z* glasses, 10% and 8% Na_2O respectively, the data are somewhat less satisfactory. The runs were made at the normal rate of heating, and at this rate of increase of temperature the softening was so rapid that there is a considerable scatter in the results. The miscibility-gap and spinodal boundaries for these binary compositions meet at 10% concentration (the top of the mis-

cibility dome: Ref. 7, Fig. 1B) and our data show a change in softening rate around 825°C for glass *y* and 845°C for glass *z*. These temperatures are in the range of those given in Ref. 7.

B. Binary potash-silicate compositions

Figure 2 combines and corrects two figures of the previous report[3a] for these compositions; the figure includes beam data for 25, 33, and 38% K_2O compositions, and pellet data for 6, 7, 8, 10, 20, 33, and 40% K_2O compositions. Schroeder *et al.*[9] indicate a critical demixing point for a composition about 7 mole % K_2O, at temperature about 540°C. This does not show up in the same way that is seen with the 15% and 20% Na_2O compositions; the behavior indicated here is strictly Arrhenius. But it should be mentioned that chemical analysis of glass *f*, 7% K_2O, disclosed serious doubt concerning the melt components; no data concerning this composition are used in Ref. 9.

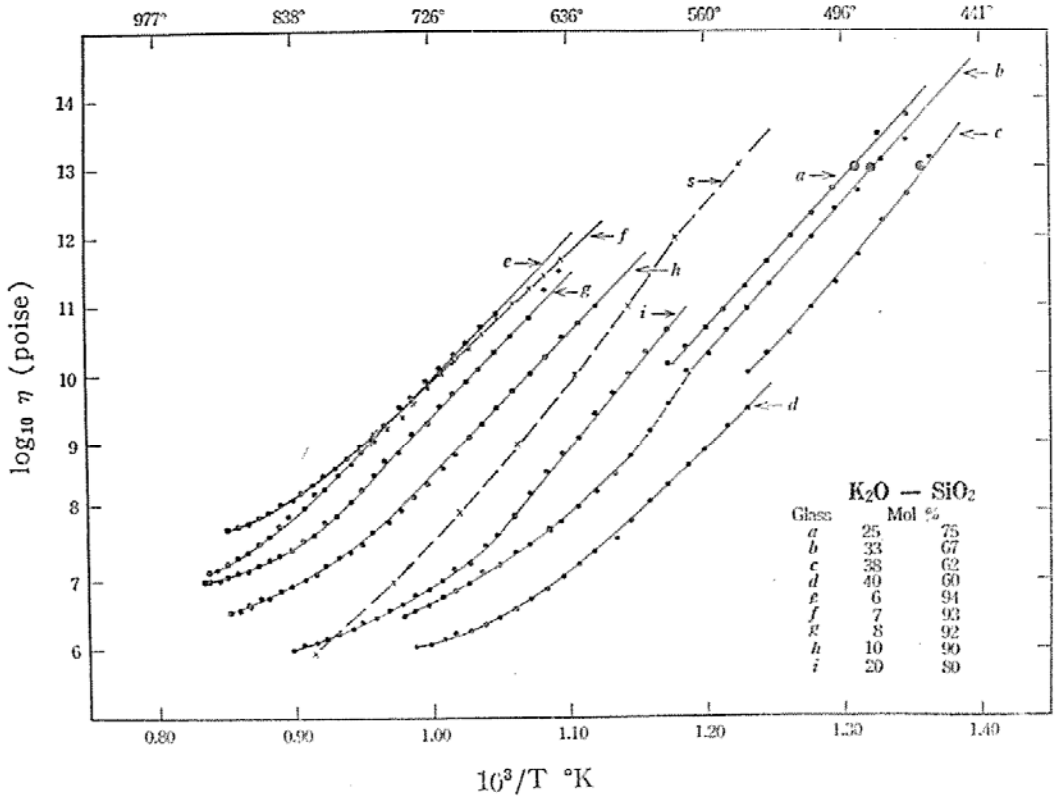


Fig. 2. Viscosity vs. $10^3/T°K$ K_2O-SiO_2 6-40 mol %

C. Binary borosilicate compositions

Vitreous B_2O_3 had been the subject of extensive study[4][10-13], but the work here reported seems to be the first so far published in which a parallel-plate viscometer was used. Glass 14 of Fig. 3 refers to data from five pellets of 100% B_2O_3 glass, the softest used in this work. Agreement with some extensive recent work is not very good. The values here given for temperatures below 380°C are significantly lower than those given in Ref. 4 and Ref. 12. Such differences are usually attributed to the presence of residual water in the melt. Considerable effort was spent to avoid this in preparing the sample materials used here, and no corrections have been made on the curves of Fig. 3.

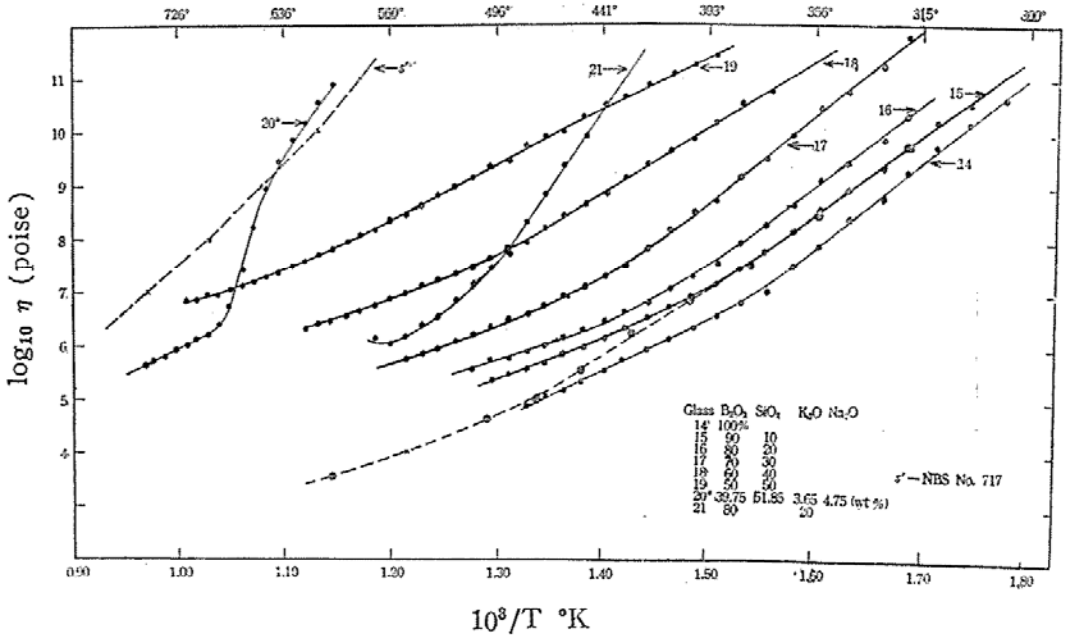


Fig. 3. Viscosity vs $10^3/T°K$ B_2O_3 with 0-50 mol % SiO_2 and with K_2O and Na_2O

It may be mentioned, however, that 1) there is agreement at higher temperatures, above 480°C and less than 10^5 viscosity; and 2) in the Arrhenius parts of these high borate concentrations our slope agrees with that of Boow (Ref. 4) and hence there is agreement in activation energy for this important vitreous material. The light broken line added at the lower part of the figure between $10^{3.5}$ and $10^{7.5}$ poise represents values given by Boow (Ref. 4, p. 48).

What is perhaps more striking with these compositions is the change in apparent activation energy as the amount of silica is increased. While increase of silica does increase the viscosity, it also decreases the slope of the Arrhenius plot. Glass 19, 50% silica, could be followed through 300 C° of softening without difficulty, though the end point, near 725°C, has viscosity near 10^7 poise. And comparision of glass 16 with glass 19, both containing 80% B_2O_3, show 20% K_2O causing significant increase in apparent activation energy for the compound while the compound having 20% SiO_2 appears to have activation energy the same as 100% B_2O_3. If all these binary borate materials are phase separated, the low temperature resistance to longitudinal stress in the potash-rich component of a sample of 20% binary borate compound is greater than for the silica rich component.

It may be noted that at these low temperatures and high viscosities the statement that "the viscosity of B_2O_3 also decreases when other oxides are added, but much less than for SiO_2"* is not verified when SiO_2 is the added oxide. For all binary borate compositions containing up to 50% SiO_2, the viscosity increases with increasing SiO_2. Whether this is true for all alkali additions remains to be seen. The work here reported does seem to verify the "boron oxide anomaly" **. The viscosity does "increase as alkali K_2O is added to B_2O_3, and the activation energy for viscous flow increases—". Possibly the size of the microstructures of the phase-separated compounds is the controlling factor here. In the alkali borates the larger structures are seen on the B_2O_3 rich sides of the miscibility gap[14], while in the borosilicates the silica-rich component has the larger microstructure, as indicated above concerning the Vycor process (pg. 72).

In a recent publication[8] Macedo and Simons make a theoretical analysis of miscibility gaps in the alkali borates, based on "the concept that molten oxides are made up of small structural units

* Doremus, R. H. *Glass Science*, John Wiley & Sons, New York, 1973, p. 109.

** Op. cit. p. 33.

which control the physical properties of their melts. ...This approach — replaces the arbitrary assumption that the glass former, say B_2O_3, mixes as a cell containing two borons and three oxygens, with the more physically acceptable concept that the cell consists of a more complex structure", in this case $[(B_2O_3)_5] = A$. According to this view, the units of phase separation in binary alkali borates are $[(Li_2O \cdot 4B_2O_3)]$ and A, $[(K_2O \cdot 3B_2O_3)]$ and A, etc.

In Ref. 7 alkali silicates are considerered to have similarly constructed cells with SiO_2 combinations. For binary soda silicate glasses, the end members are $[(SiO_2)_8]$ and $[Na_2O \cdot 3SiO_2]$, while for lithium silicates the end members are $[(SiO_2)_6]$ and $[Li_2O \cdot 2SiO_2]$. Studies of the miscibility gap in the BaO–SiO_2 system indicate that the demixing for this binary oxide system "must be considered between $[3BaO \cdot 2SiO_2]$ and $[(SiO_2)_{24}]$", the latter large multimer being required for a complex silicate structure made of SiO_4 tetrahedra that also has "the nearly spherical geometry needed for liquid structures" (Ref. 7, pg. 125). Whether such complexes exist in binary borosilicate melts and how stable they remain with changing temperature remains to be seen.

A few technical points concerning these materials containing B_2O_3 may be mentioned. While the three recorder charts for pellets of the 20% potash-borate binary composition, glass 21, Fig. 3, show excellent agreement, the material is highly unstable chemically. After the pellet was compressed in the viscometer the final disc was was found to be completely devitrified and so hygroscopic that the final thickness measurement had to be done quickly while the disc was disintegrating on exposure to the atmosphere. The early reports of viscosity measurements on alkali borate glass melts by Shartsis, Capps, and Spinner[10][15] refer to temperatures above the range of measurements reported here. At 1100°C binary alkali silicates containing K_2O, Na_2O, and LiO_2 do show monotonic decrease in viscosity with increasing alkali concentration above 15 mole % (Ref. 15, pg. 158). For temperatures between 500 and 1000°C however the picture (Ref. 10, pg. 321) is much different, and the authors report: "At 500°C only limited data are available because composi-

tions containing more than approximately 9 mole % alkali oxide were either too viscous to be measured or were devitrified." These early (1952–1953) measurements used a restrained sphere apparatus in which the viscosity was measured by "noting the motion of a platinum sphere suspended in the test liquid from one arm of a balance", the glass melt being held in a platinum crucible which also served as a conductance cell for measurements of electrical resistivity.

D. Three-component alkali-silicate glasses

1. *With added Al_2O_3*

Figure 4 summarizes work with lithium silicate compositions having 33 mole % Li_2O, and replacing SiO_2 with four different concentrations of Al_2O_3. Data from previous work with 33 mole % binary lithium silicate have been incorporated. Haller *et al.*[7] have reported metastable liquid-liquid immiscibility in 33 mole % glass of this composition, though it is practically at the boundary of the gap,

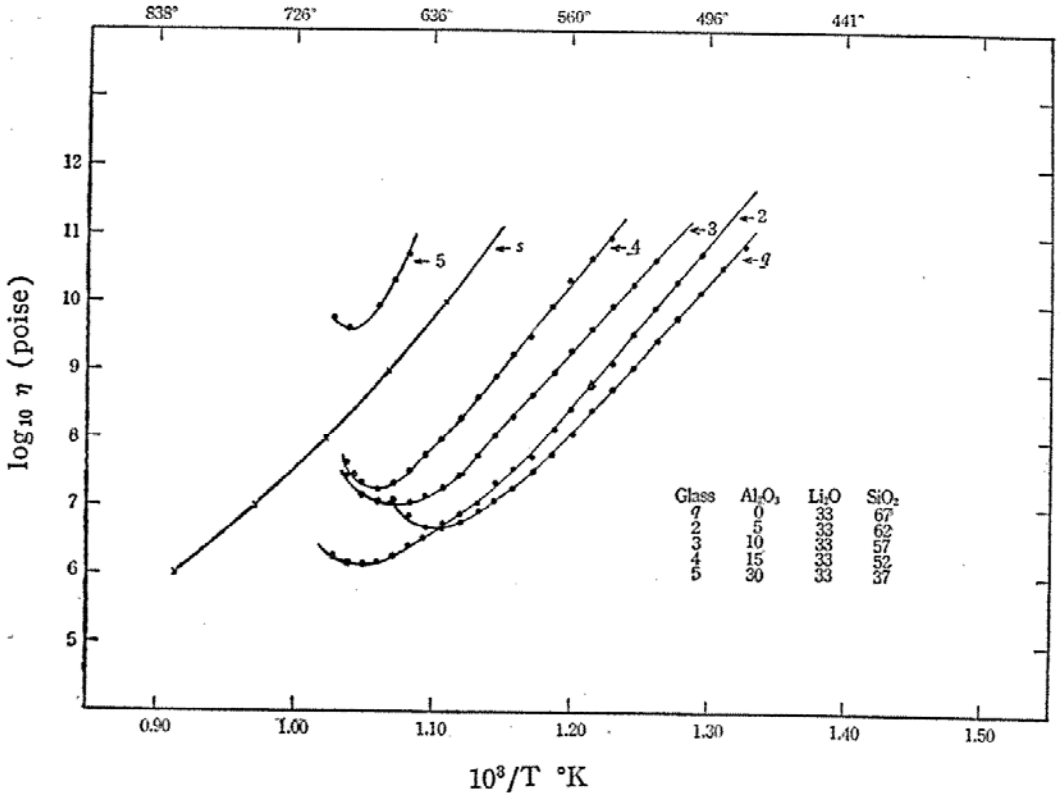


Fig. 4. Viscosity vs $10^3/T°K$ 33 mol % Li_2O with SiO_2 and 0-30% Al_2O_3

and the preparation of samples presents considerable difficulties. Binary lithium-silicate glasses are highly subject to devitrification; previous attempts to get beam bending data with them were not successful.

Data here given show the effect of alumina in producing a much more stable glass when added in small concentration. All the different concentrations show devitrification as the temperature approaches 700°C. The composition having 30 mole % alumina is an extremely "short" glass, showing no measurable softening until about 650°C.

2. *With added MgO*

Viscosity measurements of three ternary compositions having 20% Na_2O with added MgO to replace SiO_2 are summarized in Figure 5, to which data from glass x of Fig. 1, binary 20% soda silicate, has been added. The latter composition is near the limit for separation and has been discussed above, pg. 75. Whether addition of

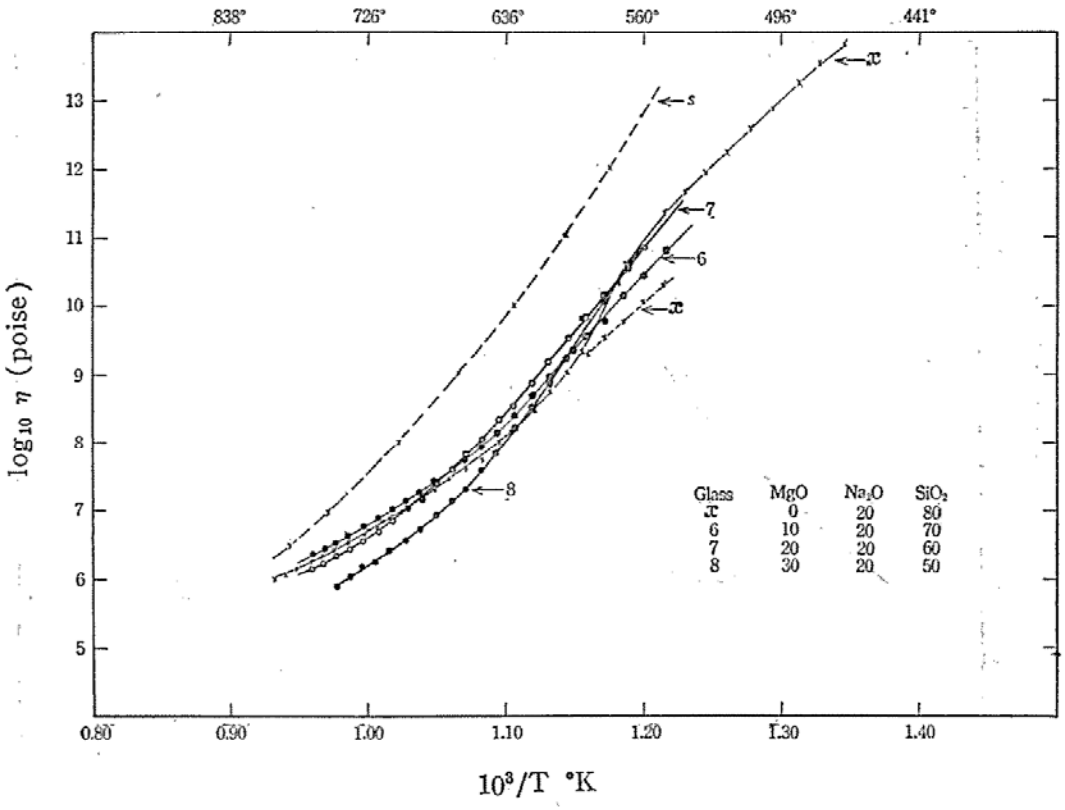


Fig. 5. Viscosity vs $10^3/T°K$ 20 mol % Na_2O with SiO_2 and 0-30% MgO

MgO plays any role in inhibiting phase separation is not evident from these data. It is clear that addition of 10% and of 20% MgO in place of SiO_2 has little effect on the viscosity of 20% Na_2O–SiO_2 glass; the lines on the Arrhenius plot for the three compositions lie very close together. Only glass 8 having 30% MgO shows significant softening at higher temperature. Beam bending measurements are needed to study the effect of added MgO at lower temperatures and higher viscosities.

3. *Pseudo-ternary alkali silicates*

Figure 6 contains an extension of work given earlier[3a] on beams of four compositions having 75% SiO_2 with varying amounts of K_2O and Na_2O. The effect of "doping" a 25% potash-silicate binary glass with $\frac{1}{2}$% of Na_2O is shown by the comparision between glass *a* and glass *t*. For the other three compositions there is no significant difference in the 10^{13} P. point; all have annealing point about 430°C. Some difference between the compositions begins to show at temperature about 460°C.

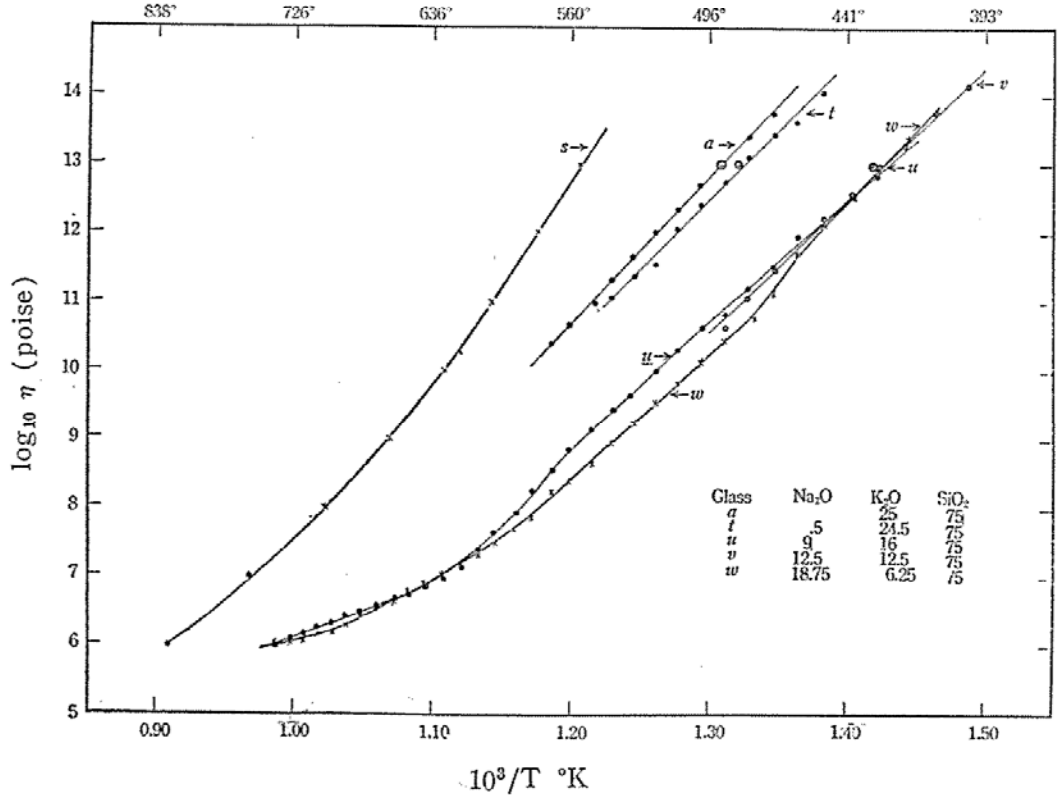


Fig. 6. Viscosity vs $10^3/T$°K 75 mol % SiO_2 with K_2O and Na_2O

Later pellet work with the two compositions *u* and *w* does show some further difference. While the slopes of the lines, and hence the activation energy for the two compositions in the Arrhenius region between 10^9 and 10^{11} poise viscosity remains the same, the glass with higher K_2O concentration shows significantly higher viscosity. But with higher temperature, about 635°C and viscosity about 10^7 poise, the difference between the two compositions again becomes insignificant.

E. Four-component silicate glasses with 14% Na_2O

Four of the glasses of Figure 7 have the same Na_2O concentration, and in two of the four the silica concentration is also the same: 74%. Two have the same concentration of CaO, 12%, but have added Al_2O_3, while the other two have smaller amounts of CaO. Glasses 10 and 11, in which Al_2O_3 is substituted for SiO_2 show the hardening effect of alumina, which can also be seen in Fig. 4. But where SiO_2

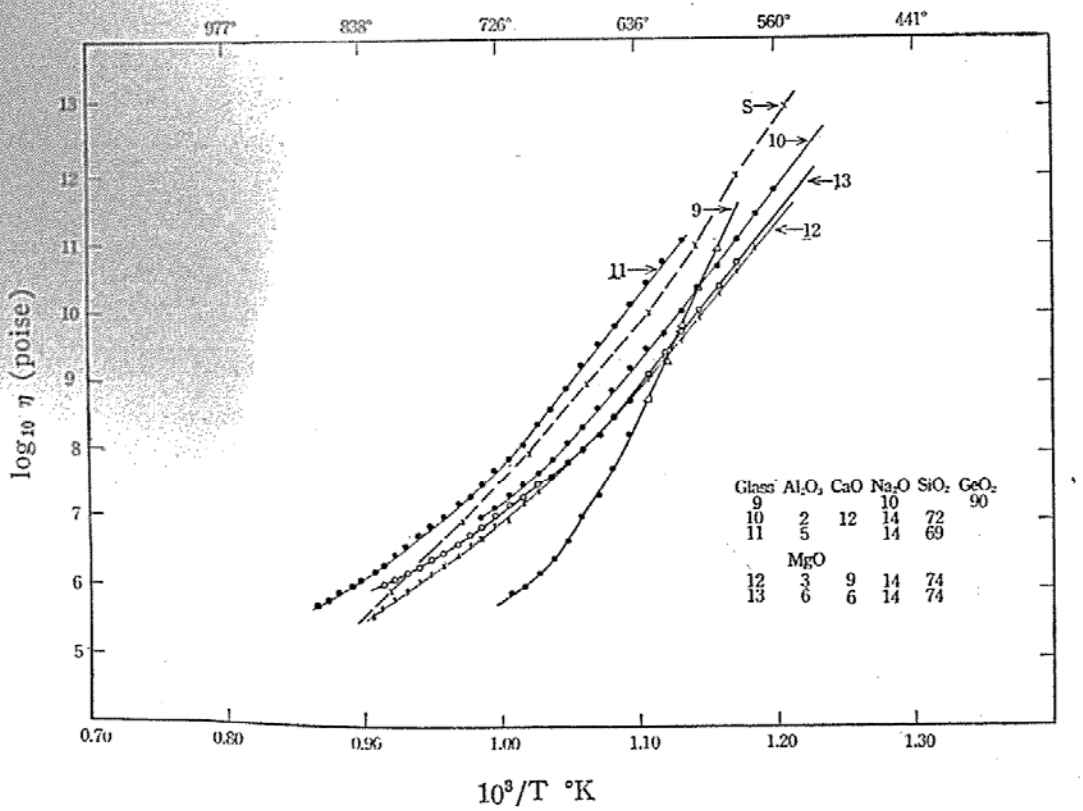


Fig. 7. Viscosity vs $10^3/T°K$ Four-component Silicate Glasses with 14% Na_2O and Al_2O_3, MgO, CaO; also 60% GeO_2 with 10% Na_2O.

and NaO_2 are kept constant, as in glass 12 and glass 13, substitution of alumina for calcium, in concentrations less than 10%, does not produce significant change in viscosity until temperature about 640°C is reached. There is no evidence here of the kind of sharp change in viscosity that can be seen with a four-component borosilicate glass like 20* of Fig. 3.

F. Binary sodium germanium glass

Three pellets of 10% Na_2O binary germanium glass were compressed; the short line of Fig. 7, glass 9, summarizes the measurements of viscosity for them. With this glass the softening could be followed much better than with the 10% sodium silicate one, glass *y*, of Fig. 1, even though only 125 C° temperature range is covered between 10^6 and 10^{11} poise viscosity. The slope of the Arrhenius plot here, however, is much different from that given for pure GeO_2 by Kurkjian and Douglas[16]. Doremus ("Glass Science", p. 105, Fig. 2) assumes the behavior to be strictly Arrhenius between 550° and 1400°C, but indicates that data were not taken between 600° and 1000°C. Why the addition of 10% Na_2O to GeO_2 should lower the softening point temperature about 300C° and almost triple the activation energy is not clear. Fontana and Plummer[17] state that addition of 10 mole % Na_2O to GeO_2 lowers the viscosity from about $10^{3.4}$ to $10^{0.4}$ poise at 1300°C; they also state that the activation energy at this temperature is lowered from about 50 to 25 Kcal/mol. No data are given by them for temperatures in the range here reported.

CONCLUSIONS

Realizing that results here reported are somewhat scattered, and regretting that additional samples were not available to give more complete data in several instances, the following points nevertheless seem worthy of mention in conclusion:

1. A "critical viscosity anomaly" indicated by sharp change in the viscosity temperature relationship and due to macroscopic effects of phase separation may be detected in viscosity measurements of amorphous materials.

2. In these measurements, the sample mass and geometry, and the method of applying stress, have considerable influence.

3. The effect of addition of SiO_2 to B_2O_3 is not entirely regular. Up to 20 mole % silica there is no change in apparent activation energy while there is relatively small monotonic increase in viscosity. For 30% silica, the additional 10% silica causes larger increase in viscosity and also an apparent increase in activation energy.

4. With silica concentration 40% and 50% even larger increase of viscosity for each additional 10% increase of silica is observed; but now the apparent activation energy decreases.

5. Addition of 20 mole % alkali oxide K_2O to B_2O_3 results in a much larger increase in viscosity and also in activation energy than addition of 20 mole % silica.

6. The difference in effect on viscosity resulting from addition of the same mole % of these two different oxides to B_2O_3 glass may well be due to difference in microstructure size and binding of phase-separated components, at least in the temperature range where viscosity is between 10^5 and 10^{11} P. for the composition.

7. Substituting increasing amounts of alumina for silica in three-component silicate glasses causes causes increasingly larger viscosity. But devitrification between 600 and 700°C is observed in all these glasses containing up to 30% alumina.

8. Substitution of magnesia for silica in three-component alkali silicate glass having 20% Na_2O has relatively insignificant effect on viscosity.

9. Binary GeO_2 glass with 10 mole % Na_2O shows widely different viscosity and activation energy than expected from limited data presently available.

Acknowledgments. The author and the members of the Physics Department at Fu Jen University, Taiwan 242, wish to thank Dr. John Schroeder, now at Roger Adams Laboratory, University of Illinois, Urbana 61801, for supplying all the research samples used in this work. These materials were made by him in his work at the Vitreous State Laboratory, The Catholic University of America, Washington, D.C. 20064. Preparation of this report has been aided

significantly through correspondence and discussions with Dr. Schroeder, under whose joint authorship some further and more theoretical reports are being prepared for publication. The first of these, dealing with the Subcriticae Viscosity Anomaly in Sodium Silicates, with J. W. Haus as a third co-author, is expected to appear soon in *Physics Letters A*.

REFERENCES

(1) Hagy, H. E. "Experimental Evaluation of Beam-Bending Method of Determining Glass Viscosities in the Range 10^8 to 10^{15} Poises"; *J. Am. Ceram. Soc.* **46** [6] 93-97 (1963).

(2) Fontana, E. H., "A Versatile Parallel-Plate Viscometer for Glass Viscosity Measurements to 1000°C"; *Am. Ceram. Soc. Bul.* **49** [6] 594-597 (1970).

(3) Schnaus, U. E., "Thermal Analysis Studies: Measurement of Viscosity of Glass"; *Fu Jen Studies* **6** 75-94 (1973).

(3a) Schnaus, U. E., "Viscosity of Alkali Silicate Glasses"; *Fu Jen Studies* 8 59-76 (1974).

(4) Boow, J., "The Viscosity of Fused Boric Oxide"; *Phys. Chem. Glasses* 8 [2] 45-55 (1967).

(5) Macedo, P. B., and Napolitano, A., "Inadequacies of Viscocity Theories for B_2O_3"; *J. Chem. Phys.* [4] 1890 (1968). Ritland's BSC Glass: Fig. 3 and Table III.

(6) Poole, J. P., "Low-Temperature Viscosity of Alkali Silicate Glasses"; *J. Am. Ceram. Soc.* **32** [7] 230-233 (1949).

(7) Haller, W., Blackburn, D. H., and Simmons, J. H., "Miscibility Gaps in Alkali-Silicate Binaries — Data and Thermodynamic Interpretation"; *J. Am. Ceram. Soc.* **57** [3] 120-126 (1974).

(8) Macedo, P. B., and Simmons, J. H., "Theoretical Analysis of Miscibility Gaps in Alkali-Borates"; *J. Res. N.B.S. (US)* **78A** [1] 53-59 (1974).

(9) Schroeder, J., Montrose, C. B., and Macedo, P. B., "Kinetics of Concentration Fluctuations in a Binary Silicate System"; *J. Chem. Phys.* (to be published).

(10) Shartsis, L., Capps, W., and Spinner, S., "Viscosity and Electrical Resistivity of Molten Alkali Borates"; *J. Am. Ceram. Soc.* **36** [10] 319-326 (1953).

(11) Mackenzie, J. D., "Structure of Liquid Boron Trioxide"; *J. Am. Ceram. Soc.* **63** [11] 1875-1878 (1959).

(12) Napolitano, A., Macedo, P. B., and Hawkins, E. G., "Viscosity and Density of Boron Trioxide"; *J. Am. Ceram.* 48 [12] 613-616 (1965).

(13) Macedo, P. B., and Litovitz, T. A., "On the Relative Roles of Free Volume and Activation Energy in the Viscosity of Liquids"; *J. Chem. Phys.* **42** [1] 245 (1965).

(14) Shaw, R. R., and Uhlmann, D. R., "Subliquidous Immiscibility in Binary Alkali Borates"; *J. Am. Ceram. Soc.* [51] 377 (1967).

(15) Shartsis, L., Spinner, S., and Capps, W., "Density, Expansivity, and Viscosity of Molten Alkali Silicates"; *J. Am. Ceram. Soc.*, **35** [6] 155-160 (1952).
(16) Kurkjian, C. R., and Douglas, R. W., "The Viscosity of Glasses in the System Na_2O-GeO_2"; *Phys. Chem. Glasses* **1** 19 (1960).
(17) Fontana, E. H., and Plummer, W. A., "A Study of Viscosity Temperature Relationships in the GeO_2 and SiO_2 Systems"; *Phys. Chem. Glasses* **7** [4] 139-146 (1966).

A STUDY ON THE POLLEN GRAINS OF HIBISCUS WITH THE SCANNING ELECTRON MICROSCOPE

FRANZ HUBER SVD, CHING-SHIA CHEN

ABSTRACT

Hibiscus is a genus of the Malvaceae. It is commonly cultivated. The flower is very showy. The same species may have various colors. From the study of the surface and morphology of the pollen grains, an analysis of the relationships in the genus has been made with the Scanning Electron Microscope. All pollen grains in five species and in one variety of Hibiscus are usually spheroidal, and have a tectum with echinate or baculate processes. The pollen grains are mostly pantoporate, a few are nonaperturate. We used these characteristics to distinguish species from species, and the varieties in the same species. We assume that color-variations in the species are varieties in the species.

INTRODUCTION

The aim of this study of the pollen grains of Hibiscus is the application of a new method of distinguishing the various species in one genus, and the varieties in the same species, by using the Scanning Electron Microscope (SEM).

In the ordinary methods of plant-classification, we rely on plant morphology. The study of pollen grains started with C. J. Fritzsche (1808–1871) and were made with the Light Microscope, by examining the sufaces and surveying basic morphological types. About 150 years had passed without progress in the study of the morphology of pollen grains. Faegri, Iversen & Waterbolk in 1964 finally established the foundation of the taxonomy of pollen grains. In 1971, N. Grew & M. Malpighi used the Scanning Electron Microscope for the first time to study pollen grains.

Using the Light Microscope it is easy to distinguish the pollen grains of different genera. Their morphology is strikingly different. It is more difficult with various species of the same genus. They are frequently very similar. Pollen grain studies in the family of

Malvaceae were made by T.C. Huang in 1968. His studies have been directed to the analysis of the relationships of genera. No such study was made of the relationships of species in one genus.

The present investigaton of one genus of Hibiscus used the SEM to distinguish the various species. Pollen grains of five species and one variety were examined and a comparison was made. The phenomenon of different colors within the same species was also considered. The study of the morphology of a plant together with the study of the morphology of the pollen grains gives undoubtedly a better understanding of the plant.

MATERIALS AND METHODS

The pollen grains used in this study were from field collections and from the herbarium material at Fu Jen University (Table 1). All pollen grains were examined in two ways, with acetolysis (method of Erdtman 1952) and without acetolysis. The pollen grains were mature. No staining was applied. The palynological terminology follows that of Erdtman (1952), Ikuse (1956), and Faegri, Iverson, Waterbolk (1964),

Table 1. Pollen grains examined and collection data. All type-specimens at the Biology Dept. Herbarium, Fu Jen University

Taxonomy	Locality	Colletor
1. Hibiscus rosa-sinensis Linn. (Chinese hibiscus)	Fu-Jen Campus	C. S. Chen 60
2. Hibiscus rosa-sinensis Linn. var. rubroplenus Sweet (Double flowered chinese hibiscus)	Fu-Jen Campus	W. C. Ju 901
3. Hibiscus schizopetalus (Mast.) Hook f. (Fringed hibiscus)	Ping-Tung	M. H. Huang 62
4. Hibiscus syriacus Linn. (Rose of sharon)	Ching-Tan	S. I. Wang 1489
5. Hibiscus taiwanensis S. Y. Hu (Taiwan Cotton-rose)	Fu-Jen Campus	C. S. Chen 65
6. Hibiscus tiliaceus Linn. (Haw tree)	Tung-Hai Campus	M. L. Yan 63

Pollen grains for scanning with the SEM were mounted with adhesives on specimen holders and coated with carbon and gold. The scanning electron micrographs were made at Fu Jen University by Mrs. Chu-Fang Lo Wang and Mr. Chung-Shung Wang with the JSM 15 Scanning Electron Microscope.

The photographs are all of the same magnification (500× and 2000×).

OBSERVATION

1. Hibiscus rosa-sinensis Linn. (Fig. 1, 2, 3, 4)

We collected two kinds of this plant in the campus. Both kinds of the plants are shrubs, up to 5 m tall. Leaves are elliptical and have a serrate margin. The apex is acute and the base is obtuse. Flowers are solitary. The petal has 5-lobes, red color. The calyx is 5-lobed. Involucral bracts are 10. There is a difference in the size of the flower; one is larger, about 13–15 cm in diameter, the petals are cherry about 7.5 cm long, 4.5 cm wide; others are small, about 8–10 cm in diameter, the petals are dark red 4.5 cm long and 3.5 cm wide.

(a) Large flower:

Grains are nonaperturate, the longest diameter is 130μ. Tectum is with baculate processes and global lophate. Bacula with prominently dilated sexine at basis are 15μ long.

(b) Small flower:

Grains are nonaperturate or pantoporate, the longest diameter is 112–120μ. Tectum is with baculate processes and nanoprocesses. Bacula with prominently dilated sexine at basis are 13.5μ long.

2. Hibiscus rosa-sinensis Linn. var. rubroplenus Sweet

We collected four kinds of this plant in the campus. All of the plants are shrubs, up to 2–3 m tall. Leaves are elliptical, have serrate margins. The apex is acute and has obtuse basis. Flowers are solitary. The petals are numerous. The colors are red, pink, yellow and orange. The calyx is 5-lobed.

The difference in color of the flowers:

(a) Red: Fig. 5, 6, 7.

Grains are pantoporate, the longest diameter is 122 μ. Tectum is with baculate processes. One part is global with pores. The others are lophate and have nanoprocesses. Bacula with prominently dilate exine at basis are 13 μ long. The pores are about 6.5 μ long and 5 μ wide.

(b) Pink: Fig. 8, 9.

Grains are pantoporate, the longest diameter is 146 μ. Tectum is with baculate processes and fine reticulum. The global has a large pore 8.5 μ long and 5.5 μ wide. Bacula with prominently dilated sexine at basis are 15 μ long.

(c) Yellow: Fig. 10, 11.

Grains are pantoporate, the longest diameter is 140 μ. Tectum is with clava processes (or gemma) and nanoprocesses. The global has a large pore, 6.5 μ long and 5 μ wide. Gemma with prominently dilated sexine at basis are 8.5 μ long.

(d) Orange: Fig. 12, 13.

Grains are pantoporate, the longest diameter is 134 μ. Tectum is with spinose processes and nanoprocesses. The global has a large pore, 6.5 μ long and 5 μ wide. Spinoids with prominently dilated sexine at basis are 10.5 μ long.

3. Hibiscus schizopetalus (Mast.) Hook. f. (Fig. 14, 15)

A shrub is up to 5 m tall. Leaves are ovate or elliptical, have a sorrate margin. The apex is acute and the base obtuse. Flowers are solitary. The petal is lobed and parted. The color is dark red. The calyx is 5-lobed.

Grains are pantoporate, the longest diameter is 142 μ. Tectum is with spinoid-baculate processes and recticulate constricted. The global has a shallow pore, 6.5 μ long and 5 μ wide. Bacula with prominently dilated sexine at basis are 11 μ long.

4. Hibiscus syriacus Linn. (Fig. 16, 17)

A shrub is up to 3 m tall. Leaves are ovate or rhomboid about 5 cm long, are irregularly dentate and often 3-lobed.

Flowers are solitary. The petal is 5–lobed. The color is lilac. Involucral bracts are 6–7, linear.

Grains are pantoporate, the longest diameter is 144 μ. Tectum is with echinate processes and nanoprocesses. Echina with prominently dilated sexine at basis are 17 μ long.

Pollen was taken from the herbarium material; therefore the surface has fungus mycelium and spores.

5. Hibiscus taiwanensis S. Y. Hu (Fig. 18, 19)

A shrub is up to 3–8 m tall. Leaves are suberbicular and 3–5 lobed. The lobes are triangular. Flowers are solitary, axillary, and 6–9 cm in diameter. The petal is pink or white. The calyx has 5–lobes and is triangular. Involucral bracts are 8, linear.

Grains are pantoporate, the longest diameter is 126 μ. Tectum is with baculate processes and nanoprocesses. Bacula with prominently dilated sexine at basis are 15 μ long.

Pollen was taken from herbarium material; therefore the surface has fungus mycelium.

6. Hibiscus tiliaceus Linn. (Fig. 20, 21, 22)

A small tree, its branchlets glabrescent. Leaves are cordate coriaceous, and 6–11 cm in diameter. They are short, stellate-pilose above, and dense short-tomentose underneath. The apex is abruptly acuminate. The margins are entirely or remotely crenulate. Flowers are in short simple cymes. The petal is 5 lobed. The color is yellow.

Grains are nonaperturate, the longest diameter is 134–168 μ. Tectum is with echinate processes and fine reticulum. Echina with prominently dialated sexine at basis are 17.5 μ long.

Pollen was taken from the herbarium material; therefore the surface has fungus mycelium and bacteria.

CONCLUSION

Hibiscus is a genus of Malvaceae. There are many species of the genus. Hibiscus is widely distributed in temperate and tropical countries. In Taiwan, both the indigenous and introduced species,

are commonly cultivated. The cultivated species in Taiwan are shrubs except for a few ones. Usually the surface of Hibiscus has stellate hairs. Leaves are palmately, 5–11 nerved. Stipule is linear. Flowera are solitary, axillary. Involucral bracts are 6–16. The calyx is campanulate, rarely tubular. The base of calyces are half connate, presistent. The corolla is showy. The petals are with various colors of cherry, yellow, red, white and pink. The apparent character of Hibiscus is the united stamen-column surrounding the style that we call monadelphous.

The results show a great similarity of all grains of the genus Hibiscus, they are usually spheroidal, with the longest diameter 112–168 μ. The tectum has echinate or baculate processes seen best from the side view. Echinae or long bacula are 8.5–17.5 μ long. Both echinae and long bacula are usually dilated at the sexine basis.

Using the SEM, we got the characteristics of five species, that are *H. rosa-sinensis*, *H. schizopetalus*, *H. syriacus*, *H. taiwanensis* and *H. tiliaceus*. According to these characteristics, we can distinguish them clearly. *H. rosa-sinensis* "large flower" (Fig. 1) and *H. tiliaceus* (Fig. 21) are nonaperturate. Because of this, the two are distinguished from the others. There is also a difference between *H. rosa-sinensis* "larger flower" and *H. tiliaceus*. The tectum of *H. tiliaceus* has a fine reticulum (Fig. 22), the tectum of *H. rosa-sinensis* "large flower" is lophtae (Fig. 2). In *H. schizopetalus*, *H. syriacus* and *H. taiwanensis*, the tectum of *H. schizopetalus* is recticulate constricted, and also the global of *H. schizopetalus* has definite pores (Fig. 14, 15), they are different from pantoporate of the two others. In remainder, the tectum of *H. syriacus* has echinate processes (Fig. 16, 17), and the tectum of *H. taiwanensis* has baculate processes (Fig. 18, 19). This is the differentiation of species in one genus.

For the differentiation of forms (varieties) in a species using the SEM, we can identify two kinds of forms. One is different in the size of the flowers. The other is different in color of the flowers. In size, *H. rosa-sinensis* has two kinds of flowers. The large is nonaperturate, and the tectum is global lophate (Fig. 1, 2). The small is nonaperturate and panto-porate in one pollen grain (Fig. 3),

and the tectum has nanoprocesses. In color, the pictures of *H. rosa-sinensis* Linn. var. *rubroplenus* Sweet show that different color have also different forms of pollen grains (Fig. 5–13). For example, the yellow flower has gemma processes in the tectum (Fig. 10, 11). The orange flower has spinose processes in the tectum (Fig. 12, 13). The red and pink flowers have baculate processes in the tectum. But the pink is definite pantoporate (Fig. 8). The red, one part is global with pores, the other is not (Fig. 5), and it has lophate and nanoprocesses in the tectum (Fig. 6, 7). They are different from the fine reticulum of the pink flower (Fig. 9). The pollen grains have distinct characteristics for each different color in the same species.

Therefore, we can differentiate species and varieties (forms). The variety in the same species must be similar and also different. In *H. rosa-sinensis,* the large flower is nonaperaturate (Fig. 1, 2). The small one has nonperturate and pantoporate pollen grains (Fig. 3, 4). In the variety of *H. rosa-sinensis* Linn. var. *rubroplenus* Sweet, the red flower has also nonaperturate and pantoporate pollen grains. Furthermore the pink, yellow and orange flower are only pantoporate (Fig. 8–13). Therefore, the pantoporate pollen grain is a distinct characteristic of the variety.

DISCUSSION

With the light microscope, the observations can not be made in such detail. Such as ora structure: ora are crassimarginate, circular costae, colpi very short, as long as or slightly longer than the diameter of the ora (Huang T. C. 1968 in Taiwania No. 14, II, p. 202). In the pictures made with the SEM (Fig. 7, 9, 11), we see the ora more in detail. They are irregular pores. They are constricted near the surface. The crassimargins make up a part of the tectum.

About baculate processes: the bacula are with prominently dilate sexine at the basis. In the light microscope, which depends on light refraction, the baculate processes show two parts (Huang T. C. 1968 in Taiwania No. 14, II, p. 207 plate 38, 1C). In the photomicrographs using the SEM, the surfaces of the bacula are smooth (Fig.

2, 4, 9, 11, 13).

This study of pollon grains with the SEM shows that classification of flowering plants can be accomplished in a very simple, very reliable, and very convenient way.

REFERENCES

(1) Liu, T. S. 1962.: Illustrations of native and introduced Ligneous plant of Taiwan I.
(2) Li, H. L. 1963.: Woody Flora of Taiwan, Livingston Publishing Co., Pennsylvania; 543-547.
(3) Huang, T. C. 1968.: Pollen grains of Formosan Plants (4), Taiwania **14**(2); 202-2027.
(4) Wang, R. L. 1972.: Palynology, Jong-Sna's Scientifical Dictionary; 297-316.

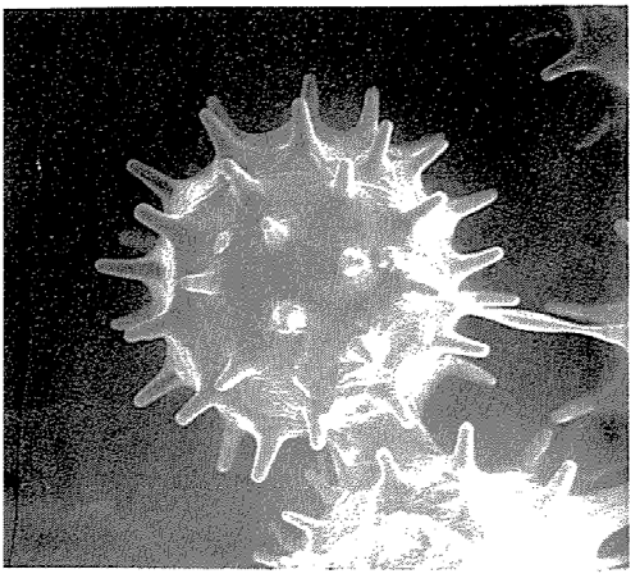


Fig. 1. Hibiscus rosa-sinensis Linn. (Large flower); which has nonaperturate and the tectum has global lophate. ×500.

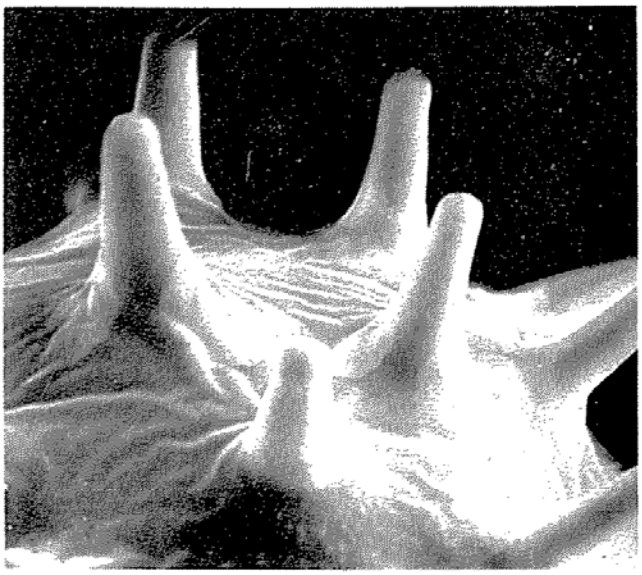


Fig. 2. Hibiscus rosa-sinensis Linn. (Large flower); which has lophate on the tectum, and with 15 μ long baculate processes. The surface of the bacula are smooth. ×2000.

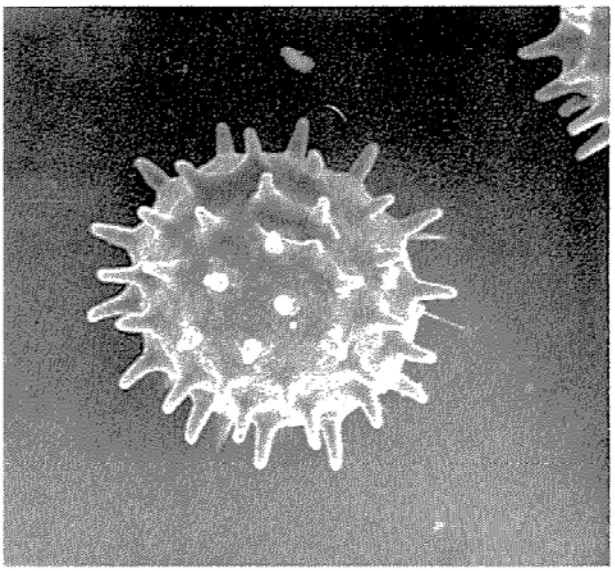


Fig. 3. Hibiscus rosa-sinensis Linn. (Small flower); which has nonaperturate in one part and other side is pantoporate. In the picture, two light pores. are shown. ×500.

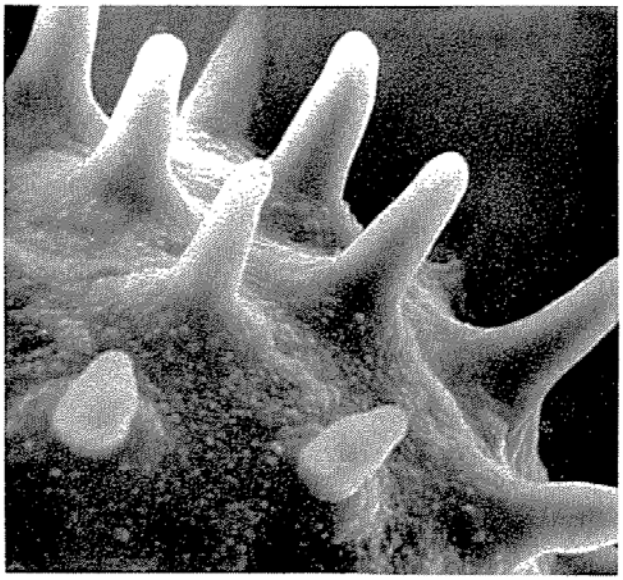


Fig. 4. Hibiscus rosa-sinensis Linn. (Small flower); which has nanoprocesses on the tectum, and with 13.5 μ long baculate processes. The bacula are smooth. ×2000.

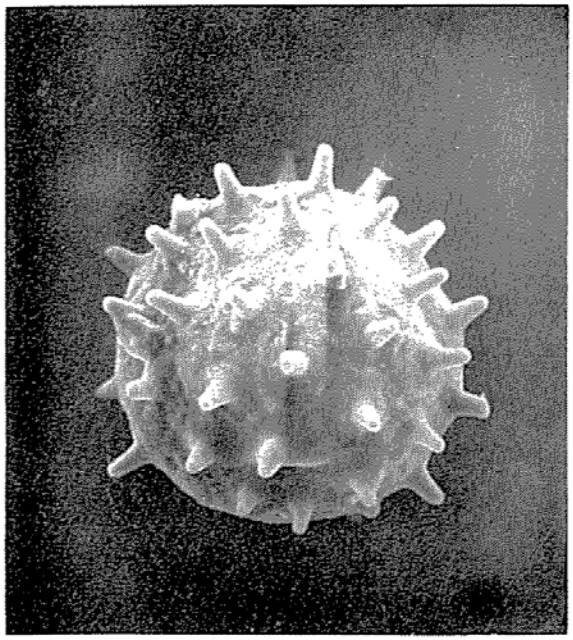


Fig. 5. Hibiscus rosa-sinensis Linn. var. rubroplenus Sweet. (Red flower); which has nonaporturate and pantoporate. One part has global with pores, the other has not, and it has lophate and nanoprocesses on the tectum. ×500.

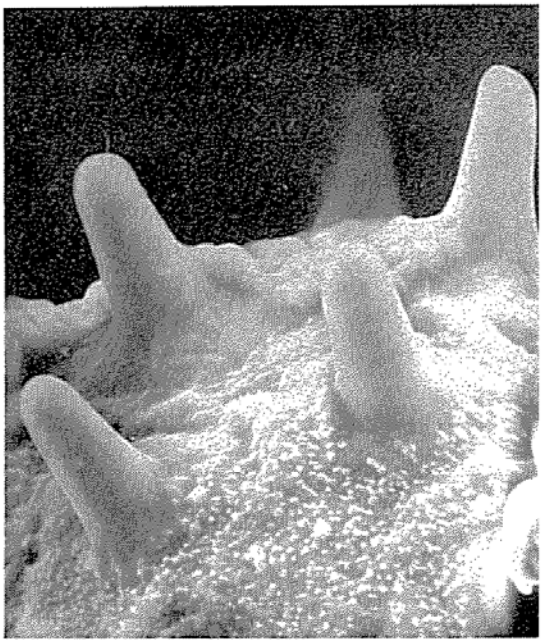


Fig. 6. Hibiscus rosa-sinensis Linn. var. rubroplenus Sweet. (Red flower); which has lophate and nanoprocesses on the tectum. Bacula are 13 μ long. ×2000.

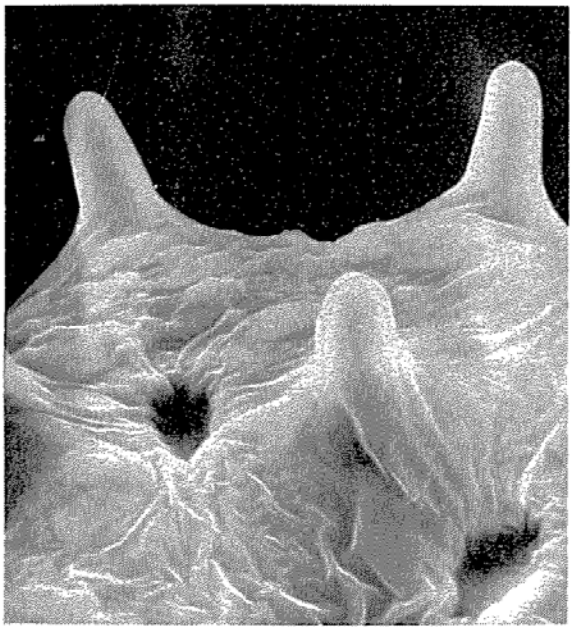


Fig. 7. Hibiscus rosa-sinensis Linn. var. rubroplenus Sweet. (Red flower); which is lophate in the tectum. The irregular pores are about 6.5 μ long and 5 μ wide. They are constricted near the surface. The crassimargins make up a part of the tectum. ×2000.

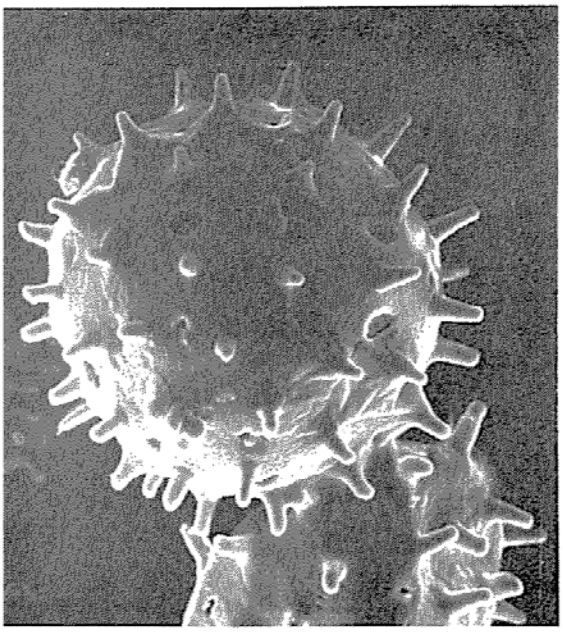


Fig. 8. Hibiscus rosa-sinensis Linn. var. rubroplenus Sweet (Pink flower); which is definite pantoporate. ×500.

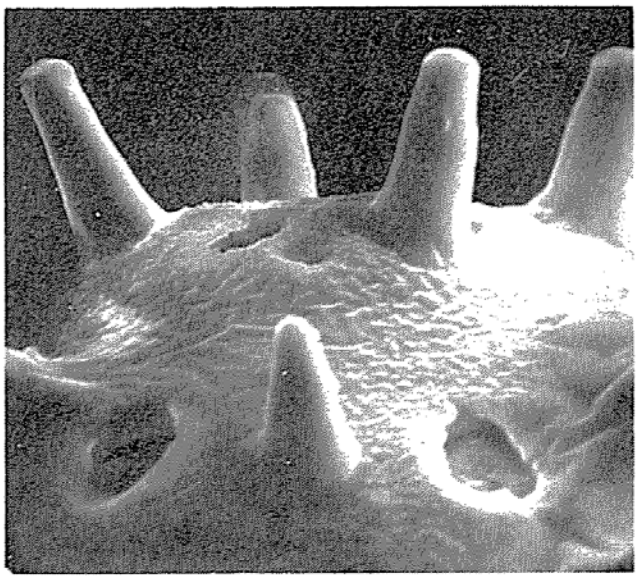


Fig. 9. Hibiscus rosa-sinensis Linn. var. rubroplenus Sweet (Pink flower); with its fine reticulum of the tectum, and with 15 μ long baculate processes. The surfaces of the bacula are smooth. The irregular pores are about 8.5 μ long and 5.5 μ wide. The crassimargins make up a part of the tectum. ×2000.

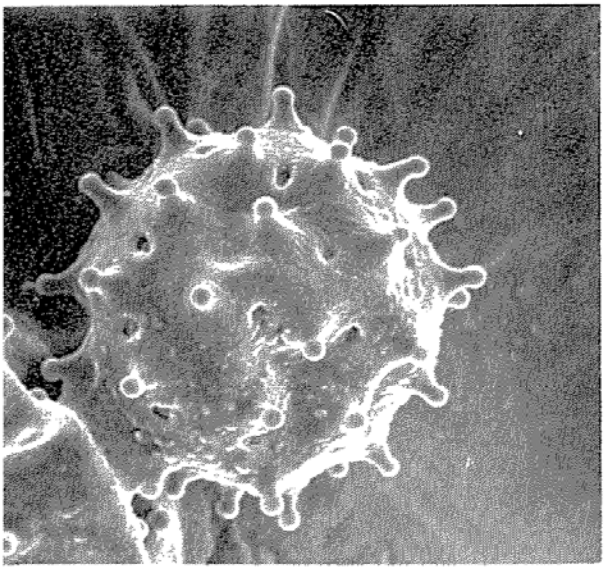


Fig. 10. Hibiscus rosa-sinensis Linn. var. rubroplenus Sweet. (Yellow flower); which is pantoporate and with gemma-processes on the tectum. ×500.

Fig. 11. Hibiscus rosa-sinensis Linn. var. rubroplenus Sweet (Yellow flower); the surface of the tectum has nanopro cesses. The gemmae are 8.5 μ long and the surface is smooth. The irregular pores are about 6.5 μ long and 5 μ wide. The crassi-margins make up a part of the tectum. ×2000.

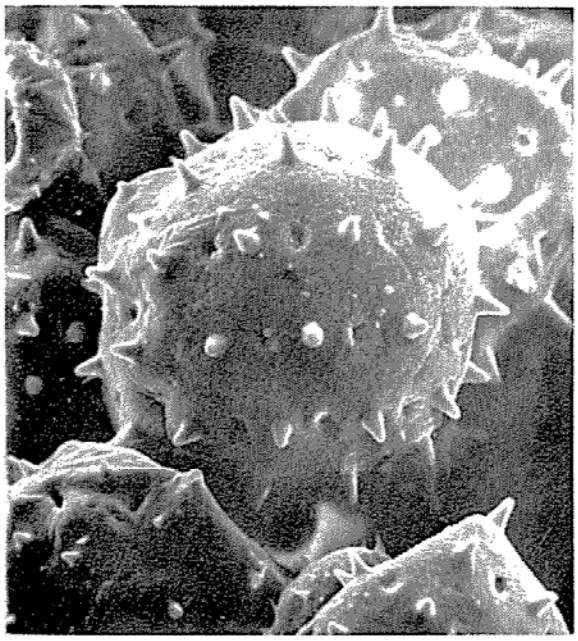


Fig. 12. Hibiscus rosa-sinensis Linn. var. rubroplenus Sweet (Orange flower); which is pantoporate. The tectum is with spinose processes and nanoprocesses. ×500.

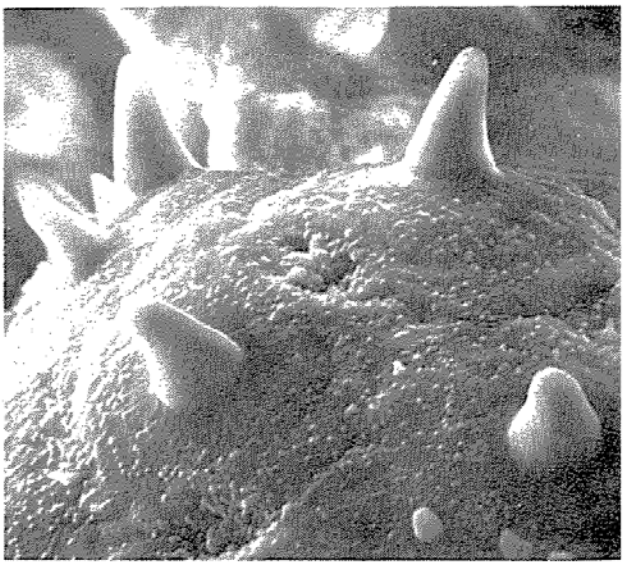


Fig. 13. Hibiscus rosa-sinensis Linn. rubroplenus Sweet (Orange flower); with its nanoprocesses of the tectum. The spinose processes are 10.5 μ long, and the surface is smooth. The pores are 6.5 μ long and 5 μ wide. ×2000.

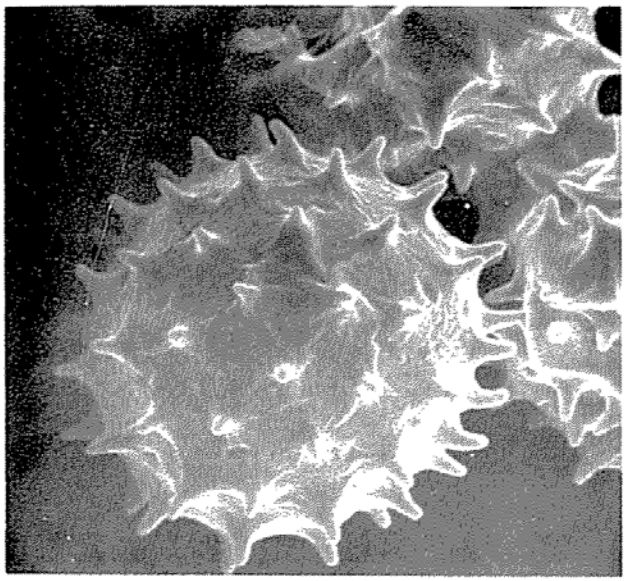


Fig. 14. Hibiscus schizopetalus (Mast.) Hook f.; which is pantoporate. The pores are definite. The tectum is with spinoid-baculate processes and is reticulate constricted. ×500.

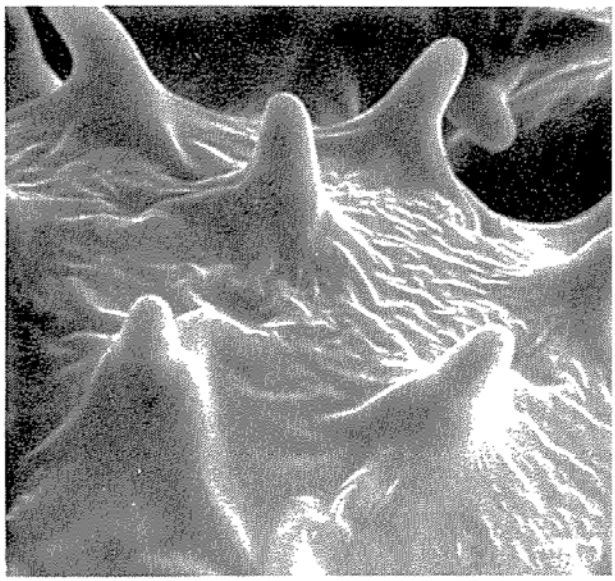


Fig. 15. Hibiscus schizopetalus (Mast.) Hook f.; which is reticulate constricted on the tectum, and with 11 μ long spinoid-baculate processes. The shallow pores are 6.5 μ long and 5 μ wide. ×2000.

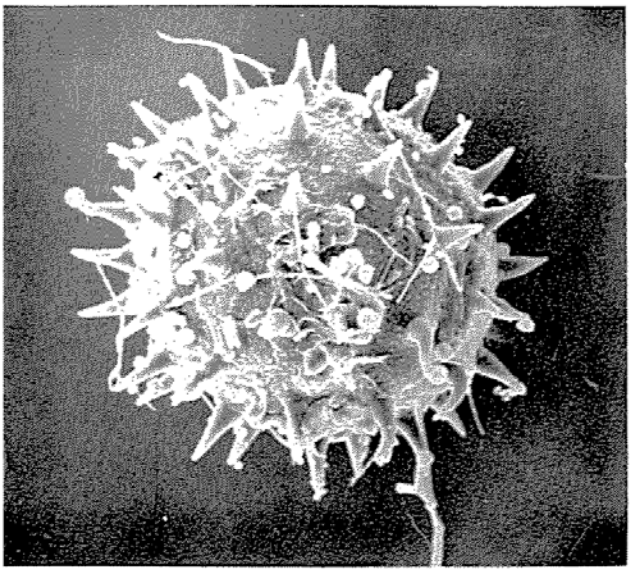


Fig. 16. Hibiscus syriacus Linn.; which has echinate processes and nanoprocesses on the tectum. The surface has fungus mycelium and spores. ×500.

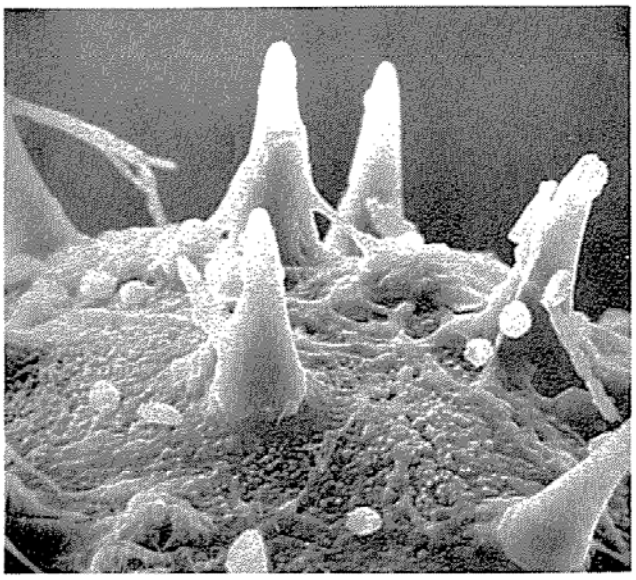


Fig. 17. Hibiscus syriacus Linn.; which has nanoprocesses on the tectum, and with 17 μ long echinate processes. The surface has fungus mycelium and spores. ×2000.

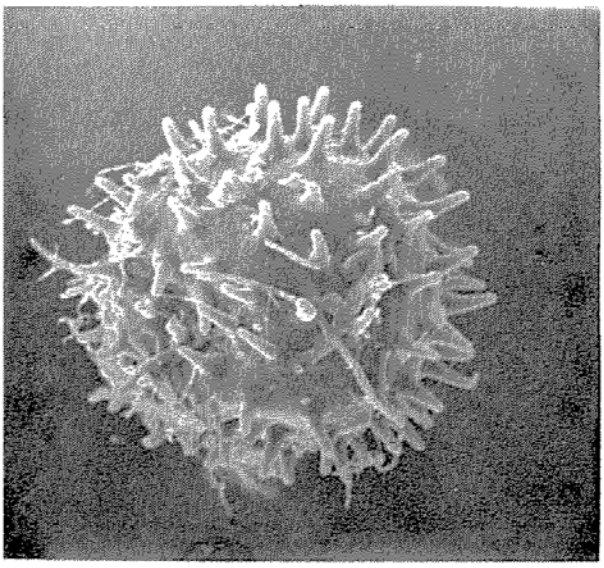


Fig. 18. Hibiscus taiwanensis S. Y. Hu; the tectum is with baculate processes and nanoprocesses. The surface has fungus mycelium. ×500.

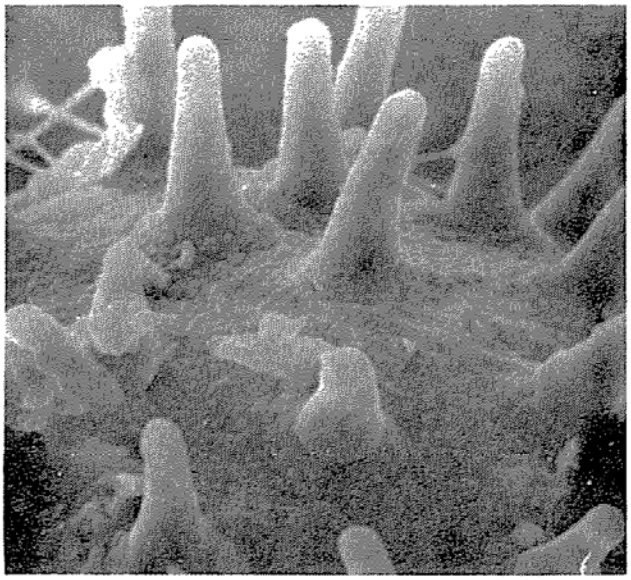


Fig. 19. Hibiscus taiwanensis S. Y. Hu; which has nanoprocesses on the tectum, and with 15 μ long baculate processes. The pores are shallow. The surface has fungus mycelium. ×2000.

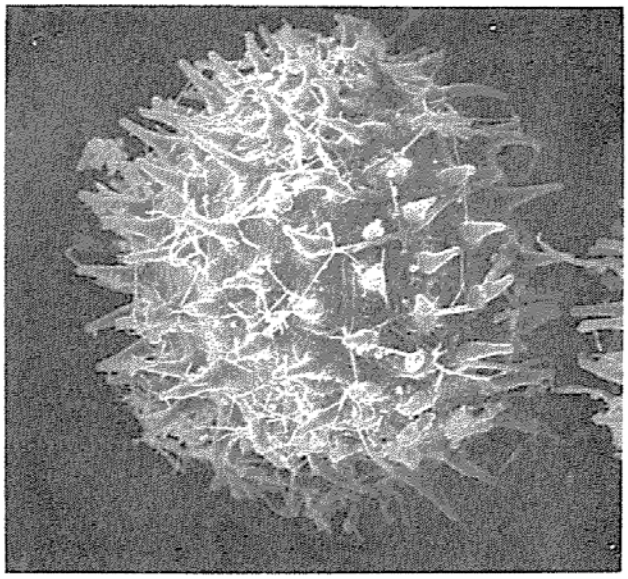


Fig. 20. Hibiscus tiliaceus Linn.; which has a diameter of 168 μ. The tectum is with echinate processes and fine reticulum. The surface has fungus mycelium. ×500.

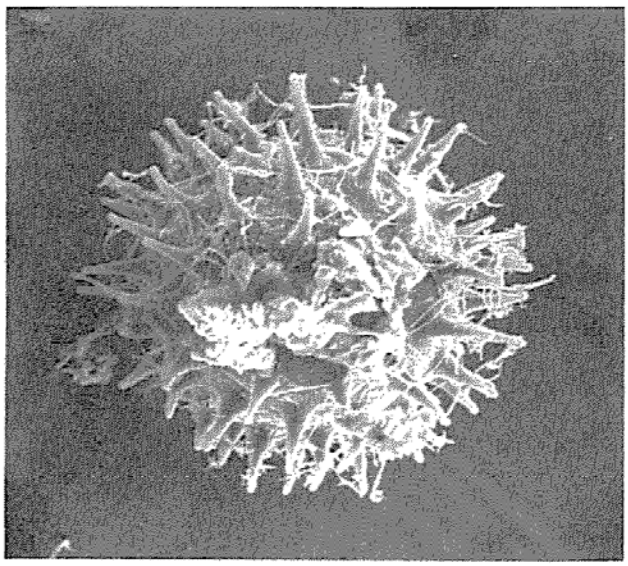


Fig. 21. Hibiscus tiliaceus Linn.; which has a diameter of 134 μ. The tectum is with echinate processes and fine reticulum. The surface has fungus mycelium. ×500.

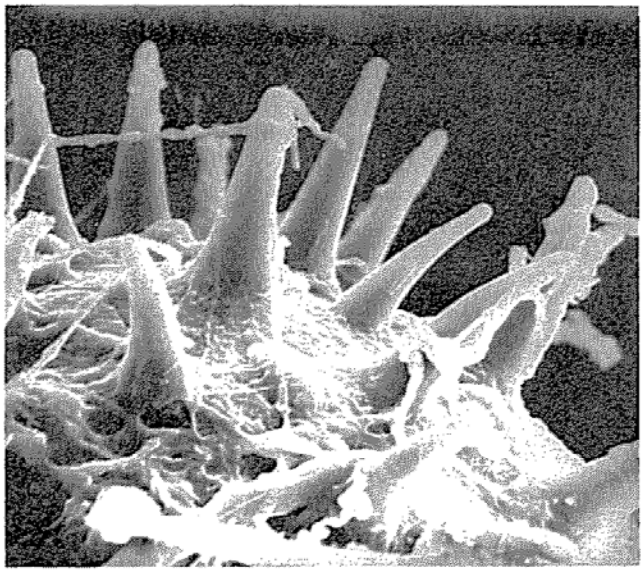


Fig. 22. Hibiscus tiliaceus Linn.; which has fine reticulum on the tectum, and has 17.5 μ long echinate processes. On the surface of echinate processes are bacteria and on the surface of the tectum is fungus mycelium. ×2000.

THE ARGININE SPARING EFFECT OF ALPHA-AMINOISOBUTYRIC ACID ON THE GROWTH OF CHICKS

TSANG-CHENG SHAO

ABSTRACT

A synthetic amnio acid, α-aminoisobutyric acid (AIB), has been known as a nonmetabolizable amino acid. It is an analogue of alanine in which the α-hydrogen is replaced by a methyl group. The amino acid has been used as an indicator for the study of amino acid absorption and transport, on the assumption that it can be transported like other neutral amino acid during intestinal absorption but can not be metabolized. In some of our experiments, we noted that AIB could modify the arginine metabolism of the chick. When feeding a casein basal diet to chicks, excess of lysine and deficiency of arginine in the casein protein caused growth retardation of the chick. The growth depression can be reversed by supplementation with arginine(1). It was suggested that high kidney arginase induced by excess of lysine accelerated the degradation of arginine in the body and caused arginine deficiency in the chick. Supplementation of arginine to the diet corrected the deficiency. We found that AIB can depress the kidney arginase of the chick and thus exhibit arginine sparing effect to some extent. The following experiments were conducted to study the effects.

MATERIALS AND METHODS

White Plymouth Rock Pullets obtained from commercial sources were fed a commercial starter ration from the day they were hatched to one week of age. At day 7, chicks within average body weight were fed the preliminary diets for an adaptation period of 4 days. At the end of the adaptation period. the chicks were further selected on the basis of body weight and distributed into groups (10 or 15 chicks) of the same average weight. The composition of the purified basal diet is given in Table 1. The experimental diet was formulated by supplementation with varying levels of L-Arginine HCl at the expense of glucose level. AIB was added into the control ration at a level of 0.5% as a treatment diet. The preliminary diet for the adaptation period was fed to duplicate groups for 7 days.

Throughout the experiments, all the chicks were housed in electrically heated batteries with feed and water supplied *ad libitum*. At the end of the experimental period, chicks were weighed and decapitated and 2 ml of artery blood were collected into vials containing a droplet (approx. 0.05 ml) of heparin solution (10 mg/ml in 0.9% NaCl solution). The blood samples were pooled in a stoppered centrifuge tube and centrifuged immediately at 2,000 rpm for 10 min. or after storage in an ice bath. After centrifuging, plasma in the supernatant was separated from crythrocytes, and 5 ml of the plasma was mixed with 2 volumes of 4.5% sulfosalicylic acid in an ice bath. The homogenate was centrifuged and the supernatant was then filtered and stored in a tightly closed bottle at –20°C.

Table 1. Basal diet

	%
Casein	20.0
Glucose monohydrate	63.66
Cellulose	3.0
Corn oil	4.0
Salts[1]	6.0
Vitamin mixture[2]	2.0
Choline chloride[3]	0.44
Glycine	0.6
L-Cystine	0.3

1. Fox-Briggs N mineral mixture (2) with additional 5 mg of $Na_2MoO_4 \cdot 2H_2O$/Kg diet.
2. Contributed in mg/Kg diet: Vitamin E conc (44,000 IU/Kg), 816.0; Vitamin A conc. (10,000 IU/g), 1,060; Vitamin D_3 conc. (20,000 ICU/g), 48; Riboflavin, 11.5; Ca pantothenate, 40.6; Niacin, 105.8; Biotin, 0.352; Folic acid, 2.2; Menadione, 20.2; Thiamine. HCl, 8.0; Pyridoxine. HCl, 14.0; Vitamin B_{12} (0.1% triturate), 40.6; p-Aminobenzoic acid, 20.0; and inositol, 1102.0.
3. A mixture of equal parts of pure choline chloride and a 25% choline cholide concentrate.

The free amino acids in the plasma and muscle samples were analyzed by a Technicon amino acid analyzer.

To study the arginase activity in kidney, the method of Smith

and Lewis[3] was used. Urea was determined colorimetrically by using 1-phenyl-1,2-propanedione-2-oxime according to the method of Van Slyke and Archibald[4].

Statistical significance between pairs of means for comparison was made by least significance difference (LSD) test[5].

RESULTS AND DISCUSSIONS

The amino acid distribution of the diet based on casein as protein source produced growth imbalance in chicks due to an excessive amount of lysine and the limiting amount of arginine in casein. From the existing concept of lysine-arginine antagonism, a high concentration of lysine in the casein protein causes arginine deficiency in the chick. Supplementation of L-arginine HCl to the casein basal diet improves feed efficiency and growth of chicks. Addition of 0,5% of AIB to the casein basal diet also improves the growth and feed efficiency of chicks (Table 2).

Table 2. Effect of dietary AIB on weight gain and feed efficiency of chicks fed a casein basal diet supplemented with L-arginine. HCl

	Exp. 1		Exp. 2	
	Wt. Gain[1]	Feed/Gain	Wt. Gain[1]	Feed/Gain
	g/chick/wk		g/chick/wk	
(1) Basal diet	35±1	3.62±0.04	39±4	3.72±0.29
(2) (1)+0.5% AIB	63±8 p	2.41±0.17	59±2 p	2.47±0.01
(3) (1)+0.5% L-arg. HCl	—	—	99±4	1.92±0.05
(4) (3)+0.5% AIB	—	—	106±5	1.76±0.01
(5) (1)+1% L-arg. HCl	110±1	1.65±0.01	109±6	1.71±0.10
(6) (5)+0.5% AIB	115±1	1.56±0.02	108±7	1.72±0.06

1. Average±range/2 of two pens of 15 chicks each. Range is the difference between the average for two pens.

p. Value is significantly different from corresponding control value ($p<0.05$).

When 0.5% AIB was added to the purified casein diet, there is a consistent elevation of plasma free arginine and depression of

plasma free ornithine (Table 3). The free arginine and ornithine in muscle was affected in a similar way and the differential effect was more pronounced.

Certain plasma amino acids showed a characteristic change when the diet was supplemented with arginine. Generally, free lysine, threonine, glycine, and valine were depressed. The addition of AIB depressed lysine and valine still further, when supplemental arginine was in the diet. (Table 3). AIB was also found to depress plasma free leucine in the presence of supplemental arginine. The depressing effect of AIB on lysine, valine and leucine was similar in muscle and plasma.

Table 3. Effect of dietary AIB on the free amino acids in the plasma and muscle of the chicks fed a casein basal diet supplemented with graded levels of arginine

Amino acid	Plasma[1]				Musclo[1]			
	L-Arg·HCl added				L-Arg·HCl added			
	0%		1.0%		0%		1.0%	
	Control	AIB[2]	Control	+AIB[2]	Control	AIB[2]	Control	+AIB[2]
	μ moles/100 ml plasma				μ moles/100 g tissues			
AIB	0	54±1	0	57±3	0	140±1	0	167± 3
Arginine	7±0	11±1	28±0	61±0[a]	21±3	26±3	134±8	472± 2[a]
Ornithine	4±0	3±1[b]	13±0	4±0[a]	14±0	11±0	46±1	20± 4[a]
Lysine	189±2	180±1	147±3	99±4[a]	882±7	1,024±23	1,059±79	793±47[b]
Threonine	257±20	272±8	111±2	87±5	427±49	435±15	311±34	179±13[b]
Glycine	81±1	93±5	70±0	65±3	447±19	437±3	380±17	243± 7[a]
Valine	57±1	57±2	48±2	39±3[b]	42±2	42±0	40±2	33±1[b]

1. Average±range/2 of two determinations, each on a pooled sample from replicate pens of 15 chicks. Range is the difference between the results for the two determinations.
2. Added at a level of 0.5%.

a. Value is significantly different from control value ($p<0.01$)
b. Value is significantly different from control value ($p<0.05$)

The improvement of growth and feed efficiency by the addition of AIB in the arginine-limiting diet or arginine-sparing effect of

Table 4. Effect of AIB injection on plasma and muscle amino acids of chicks fed a casein diet supplemented with 3% L-arginine. HCl

Amino acid	Plasma[1]		Muscle[1]	
	Saline	AIB	Saline	AIB
	μ moles/100 ml plasma		μ moles/100 g tissue	
AIB	0	34±2	0	151±4
Arginine	49±2	167±15[b]	287±21	1,461±6[a]
Ornithine	59±2	10± 1[a]	179±12	22±1[a]

1. Average±range/2 of two determinations, each on a pooled sample from replicate pens of 15 chicks. Range is the difference between the results for the two determinations.
a. Value is significantly different from control value ($p<0.01$)
b. Value is significantly different from control value ($p<0.05$)

Table 5. Effect of dietary AIB on kidney arginase and plasma urea of chicks fed a casein diet with and without supplemental arginine

		Arginase activity	Urea
		μ mole urea/hr per kidney	μg urea/ml plasma
Basal diet	Control	9,970±1,970	37±14
	0.5% AIB	780±250[a]	22±3
Basal diet	Control	20,250±1,920	129±18
+0.3%L-arg·HCl	0.5% AIB	680±240[a]	54±16[b]

1. Mean±SE is based on individual determinations on 10 chicks housed in two pens, 5 chicks in each.
2. Average±range/2 of determination on two samples of plasma, each a pooled sample for a pen of five chicks.
3. Determined by method of Archibald[9] sulfosalicylic acid in the urea standard solution.
a. Value is significantly different from control ($p<0.01$)
b. Value is significantly different from control ($p<0.05$)

AIB was apparently due to the suppression of kidney arginase in the chick. The following evidence supports this theory. First, injection of AIB in saline solution (890 mg/chick/7 days) to the chick fed a casein diet supplemented with 3% L-arginine·HCl also

showed a similar elevation of plasma arginine and depression of plasma ornithine as they were when AIB was given orally (Table 4). This will rule out the possiblity that AIB causes the elevation of plasma arginine and the depression of ornithine through the competition of AIB and other amino acids at the site of intestinal epithelial cell. Secondly, the direct evidence comes from the data in Table 5 that incorporation of 0.5% AIB into the casein diet supplemented with 3% L-arginine HCl depresses the kidney arginase and the plasma urea, despite the fact that individual variation in kidney arginase was very large. Casein is known to be a protein, containing an excess of lysine for the chick[6]. Nesheim[7] reported that there are two strains of chicks with regard to the requirement for arginine. The strain with high arginine requirement responds with high arginase activity in the kidney when an excess of lysine is added to the diet. Jones *et al.*[8] reported that excess of lysine in a casein gelatin diet or a crystalline amino acid diet caused a reduction in growth rate and an increase in kidney arginase of chicks. O'dell *et al.*[1] found that there was an inverse correlation between the growth rate and the kidney arginase activity in the chick fed a casein basal diet supplemented with a suboptimal level of arginine. If the arginine-sparing effect of AIB is brought about by the inhibition of kidney arginase, it would be of great interest, since it suggests that it could be possible to employ a synthetic nontoxic and nonmetabolizable chemical as an enzyme inhibitor to alleviate amino acid imbalance and improve growth and feed efficiency.

REFERENCES

(1) O'Dell, B. L., W. H. Amos and J. E. Savage. Relation of chick kidney arginase to growth rate and dietary arginine. *Proc. Soc. Exp. Biol. Med.* **118**, 102 (1965)

(2) Fox, M. R. S., and G. M. Briggs. Salt mixtures for purified-type diet III. An improved salt mixture for chicks. *J. Nutrition* **72**, 243 (1960)

(3) Smith, G. H., and D. Lewis. Arginine in poultry nutrition. 2, Chick arginase. *British J. Nutrition* **17**, 433 (1963)

(4) Van Slyke, D. D., and R. M. Archibald. Gasometric and photometric measurement of arginase activity. *J. Biol. Chem.* **165**, 293 (1946)

(5) Steel, R. G. D, and J. H. Torrie. *Principles and procedures of Statistics,* McGraw-Hill Book Company, New York. (1960)
(6) Jones, J. D. Lysine-arginine antagonism in the chick. *J. Nutrition* 84, 313 (1964)
(7) Nesheim, M. C. Kidney arginase activity and lysine tolerance in strains of chicks selected for a high or low requirement of arginine. *J. Nutrition* 95, 79 (1968)
(8) Jones, J. D., S. J. Petersburg and P. C. Burnett. The mechanism of the lysine-arginine antagonism in the chick: effect of lysine on digestion, kidney arginase, and liver transamidinase. *J. Nutrition* 93, 103 (1967)
(9) Archibald, R. M. Colorimetric determination of urea. *J. Biol. Chem.* 157, 507 (1945)

"I can still see the scene in the chemical laboratory at Uppsala, where the chief, himself a high-ranking chemist, pointed at a beaker containing a watery solution and inquired in a shocked tone: "And you believe the sodium atoms are just swimming around in there?" And when I answered in the affirmative, I was mustered with a brief glance that unconsciously expressed honest doubts as to my chemical good sense."

WILHELM OSTWALD
Nobel Prizewinner in
Chemistry (1909)

THE EFFECTS OF SOYBEAN HULLS ON AFLATOXIN PRODUCTION

EVAMONICA M. JAMLANG

ABSTRACT

In the presence of soybean hulls 80% of the aflatoxin produced by *A. flavus* NRRL 2999 disappeared from the medium within 24 hours after the maximum concentration was reached. Maximum aflatoxin production was attained between the third and fourth days of incubation. The hull component(s) responsible for these effects were found in the lipid soluble fraction which, when added to dehulled soybeans plus glucose ammonium nitrate medium and to rice, reproduced these effects. The soybean *per se* supports good mold growth but poor aflatoxin production. An adequate source of carbon for aflatoxin production seems to be unavailable in the soybean.

INTRODUCTION

Aflatoxins are produced on a variety of substrates by some strains of *A. flavus* and *A. parasiticus*. Studies on nutritional requirements for *in vitro* aflatoxin production in synthetic and semi-synthetic media revealed that both carbon and nitrogen sources are required, plus the presence of certain mineral elements, principally Zn, Fe, and Mg[1,5,6,13,14]. The best carbon sources are glucose and sucrose, and the best nitrogen sources are yeast extract and peptone. In *in vitro* studies using sterilized agricultural products, wheat, rice, corn and peanuts were found suitable for good aflatoxin production[10]. Significant amounts of aflatoxin have been reported to occur naturally, *in vivo*, only in oil seeds (peanuts and cottonseed) although many other agricultural products are able to support mold growth[11]. However, soybean, also an oilseed, was reported a poor substrate for both the *in vivo*[11] and *in vitro*[10] formation of aflatoxin. Composition reports on soybeans show that carbon and nitrogen sources and the required mineral elements needed for aflatoxin production are present[17,21]. These facts suggest that some factor besides lack of nutrients might be the cause of poor aflatoxin yield in soybeans.

Environmental factors that affect aflatoxin formation include temperature, length of incubation time, aeration and available moisture. Optimal temperatures are generally between 20° and 30°C. As the temperature increases, the incubation period necessary to attain maximal accumulated toxin decreases. Schroeder and Hein[20] obtained maximal yield in an autoclaved, inoculated sample of cottonseed after 8 days of incubation at 20°C but only after 3 days at 35°C. The maximal yield was greater for the 20° sample than for the 35° sample. An incubation time of 7 days was chosen for this study based on reports that the maximal yield of aflatoxin in natural substrates usually was attained after 6–8 days[19,20].

Several investigators have noticed a decline in aflatoxin concentration after different incubation times ranging from 72 hours to 8 days[2,4,19]. Ciegler *et al.*[4] found that decrease in aflatoxin concentration took place after 72 hours in fermentors by using different strains of *A. flavus*. Mycelial lysis and a rapid aeration rate were necessary conditions to produce this effect. Schroeder[19] reported aflatoxin degradation after 4–8 days by using *Aspergillus parasiticus* isolate 64–R8 in stationary culture. On the other hand, two strains of *A. flavus* (NRRL 2999 and M 001) have been called nondegraders[4], and several isolates of *A. flavus* obtained by Diener and Davis[7] maintained high aflatoxin levels after 3 weeks in stationary culture. However, Ciegler *et al.*[4] noted that nondegrading strains can be induced to degrade aflatoxin under conditions that result in mycelial lysis.

This investigation is an attempt to determine the cause of poor aflatoxin production in soybeans. If a chemical component is responsible, it may be used to prevent aflatoxin formation.

MATERIALS AND METHODS

Organism. Nine-day-old slants of *A. flavus* NRRL 2999 were prepared from a stock culture maintained on potato dextrose slants at 4°C. Ten ml of sterile distilled water was introduced into each slant and was shaken to loosen the spores. The spore suspension was used as inoculum (5 ml per sample).

Substrates. Soybean, variety "Hawkeye 63" was used as: 1) whole beans; 2) ground whole beans—beans powdered by a comminuting machine (Fitzmill, Model D, W. J. Fitzpatrick Co.); 3) ground dehulled beans—beans dehulled before grinding; 4) defatted meal—ground dehulled beans defatted by Soxhlet extraction with Skelly B for 10 hours and the residual solvent removed by air drying for 24 hours; 5) water-extracted meal—defatted meal subjected to several 4-hour extractions with water (1:5 w/w) until the extract no longer gave any precipitate with 1 N HCl and was negative to Molisch's test for carbohydrates. Soybean whey is the supernatant obtained when the water extract of ground whole soybeans was treated with 0.1 M HCl to precipitate the acid precipitable protein.

Glucose ammonium nitrate (GAN) medium containing glucose, 5%; NH_4NO_3, 1%; KH_2PO_4, 0.24%; $MgSO_4 \cdot 7H_2O$, 0.20%; $ZnSO_4 \cdot 7H_2O$, 0.0026%; $CuSO_4 \cdot 5H_2O$, 0.00026%; $Co(NO_3)_2 \cdot 6H_2O$, 0.00013%; $CaCl_2$, 0.0065%[3] was added to the soybean fractions as required.

Aflatoxin assay in different soybean fractions. Raw and sterilized (15 psi, 15 min.) duplicate samples of the soybean fractions just described (substrates 1–5) were prepared based on 80 g of whole beans. Analysis of the soybeans showed 17.7% (ca. 14 g per 80 g of beans) oil, 82.3% (ca. 66 g per 80 g beans) defatted meal and negligible weight of hulls (per 80 g of whole beans). One set of samples was supplemented with GAN medium (200 ml/80 g beans), whereas only distilled water (200 ml/80 g beans) was added to the other set.

All the soybean samples and a control sample containing only GAN medium (200 ml) in 2-liter flasks were inoculated and incubated in still culture at 25°C in semidarkness for 7 days. After 7 days the molds were killed with ca. 20 ml of $CHCl_3$, and the fermentation medium was broken up in a Waring Blendor with the addition of water to make a slurry. The slurry was filtered through Whatman #1 filter paper and was washed on the paper to give a filtrate volume of 450 ml. The filtrate was extracted with three 50-ml portions of $CHCl_3$ in a 1-liter separatory funnel, and the combined extracts were evaporated to dryness. The resulting solids were dissolved in 10 ml of fresh $CHCl_3$.

The ultraviolet absorption spectra of the aflatoxins all show maxima at 265 and 363 mμ[22]. Spectrophotometric analysis of the crude $CHCl_3$ extracts using a Beckman DK-2A recording spectrophotometer showed peaks at 260 and 355 mμ. The crude extract was applied on silica gel thin layer chromatography plates (0.5 mm thick) and was developed with 3% CH_3OH in $CHCl_3$. The resulting fluorescent bands were scraped separately and were extracted with $CHCl_3$. The concentrated extracts when analyzed in the spectrophotometer showed that certain bands that remained close to the origin on thin layer chromatography plates absorb strongly at 260 mμ, but bands that had R_f values similar to those of the aflatoxins gave peaks at 265 and 360 mμ. The peak at 355 mμ was therefore used in these experiments as a measure of the aflatoxin content of the crude extracts. Since absorbance values are directly related to concentration, the optical density values were used to compare amounts of aflatoxin rather than calculate concentration.

Determination of the effect of the hulls. Ground soybean hulls, Skelly B extracts, $CHCl_3$ extracts and ether extracts of the hulls (obtained by Soxhlet extraction of ground hulls) were added separately to ground dehulled soybeans (20 g) in 1-liter flasks. The organic solvents were evaporated with continuous stirring to insure uniform dispersion of the hull extracts. A control sample in which only Skelly B was added and afterwards evaporated was also prepared. The mixtures were sterilized and supplemented with presterilized GAN medium (50 ml). Rice samples, with and without addition of the Skelly B extract of soybean hulls, were prepared in a similar way except that GAN medium was not added. Inoculation and analysis of the crude toxin extracts from all the samples were done as described in the previous section with the corresponding reduction in the volumes of solvents used; aqueous extraction of soybean meal was continued to give a filtrate volume of 300 ml; $CHCl_3$ extractions were carried out with three 34-ml portions. The evaporated extracts were dissolved in 5 ml of fresh $CHCl_3$.

Determination of aflatoxin degradation. To determine whether a reaction can take place between aflatoxin and the hull extract *per*

se, the Skelly B extract of the hulls was freed of solvent and mixed with extracts that contained aflatoxins. Since the hull extract is not water soluble, the components were mixed in a $CHCl_3$ medium and allowed to stand for 7 days at 25°C. The ultraviolet absorption spectrum of the mixture was scanned before and after incubation.

RESULTS

Aflatoxin assay in different soybean fractions. Because of microbial competition, the raw samples did not give significant results. In most instances, the *Aspergillus* was overgrown by other fungi and bacteria, and only trace amounts of aflatoxin were detected. Exposure to ultraviolet light before inoculation and an increase of inoculum did not prevent these unfavorable factors. In general, the raw samples with added GAN medium gave better mold growth than the samples without GAN medium. Detectable absorbances per 4 g of beans are reported in Table 1. Mold growth was estimated visually and a rating of "卅" was given when the surface of the soybean culture was completely covered with thick mycelia and spores. When growth was abundant, sporulation usually started after the third day of incubation. A rating of "艹" represented growth that covered only part of the medium and "+" stands for very little growth.

The aflatoxin yields of the inoculated sterilized samples of soybeans are summarized in Table 2. All the samples with GAN medium and 80% of those without GAN medium showed good mold growth. Duplicate samples gave similar aflatoxin yields with the exception of the fivefold difference found in whole beans. Although the samples containing GAN medium consistently gave more aflatoxin than those without GAN medium, the increase in aflatoxin yield with the addition of GAN medium was appreciable only in the dehulled beans and in the high-yielding sample of whole beans. These samples gave toxin yields five- to sevenfold greater than the other fractions. The lowest aflatoxin yields were found in the water-extracted meal inspite of good mold growth. GAN medium alone produced very little aflatoxin (Table 1 and 2).

Table 1. Mold growth and aflatoxin production in raw soybean samples inoculated with spores of *A. flavus* NRRL 2999 and incubated in still culture for 7 days at 25°C

Soybean fractions	With GAN		Without GAN	
	Mold growth	A 355 mμ	Mold growth	A 355 mμ
Whole beans	++[a]	0.10[b]	+	0
	++	0.67	+	0
	++	0.12	+	0
	++	0.16	++	0.64
Ground with hulls	++	0	++	0
	+	0	+	0
	++	0.14	++	0
	++	0.08	+	0
Dehulled	++	0.14	++	0
	++	0.07	++	0
Defatted				
a) with hulls	++	0.46	0	0
	++	0.14	0	0
b) without hulls	++	0	0	0
	++	0	0	0
Water extracted meal				
a) with hulls	+++	0.22	+	0
	+++	0.11	+	0
b) without hulls	+++	0.11	+	0
	+++	0.30	+	0
GAN medium alone	+	0.03		
	+	0.05		

[a]: As estimated visually; +++, abundant growth covering entire surface of culture medium; ++, growth covering only a part of the surface; +, very little growth.

[b]: Absorbance of the $CHCl_3$ extract per 4 grams of soybeans.

Thin layer chromatography gave typical aflatoxin spots with fluorescent intensities proportional to the absorbance values observed at 355 mμ.

Determination of the effect of the hulls. Preliminary experiments in which increasing amounts of soybean hulls (2, 5, and 8 g)

Table 2. Mold growth and aflatoxin production in sterilized soybean samples inoculated with spores of *A. flavus* NRRL 2999 and incubated in still culture for 7 days at 25°C

Soybean fractions	With GAN		Without GAN	
	Mold growth	A 355 mμ	Mold growth	A 355 mμ
Whole beans	++++[a]	0.44[b]	+++	0.55
	+++	2.15	+++	0.45
	+++	0.35	+++	0.25
	+++	0.35	+++	0.05
Ground with hulls	+++	0.39	+++	0.26
	+++	0.37	+++	0.26
	+++	0.45	+++	0.12
	+++	0.26	+++	0.15
Dehulled	+++	2.95	+++	0.30
	+++	1.95	+++	0.62
	+++	2.13		
	+++	1.68		
Defatted				
a) with hulls	+++	0.29	+++	0
	+++	0.19	+++	0
b) without hulls	+++	0.24	+++	0.22
	+++	0.23	+++	0.18
Water extracted meal				
a) with hulls	+++	0.14	+	0
	+++	0.09	+	0
b) without hulls	+++	0.17	++	0
	+++	0.18	+	0
GAN medium alone	+	0.03		
	+	0.05		

[a]: As estimated visually: +++, abundant growth covering entire surface of culture medium; ++, growth covering only a part of the surface; +, very little growth.

[b]: Absorbance of the $CHCl_3$ extract per 4 grams of soybeans.

were added to dehulled soybeans showed no difference between the three amounts of hulls used and their effect on aflatoxin production. Additions were arbitrarily fixed at 5 g. However, it was found that

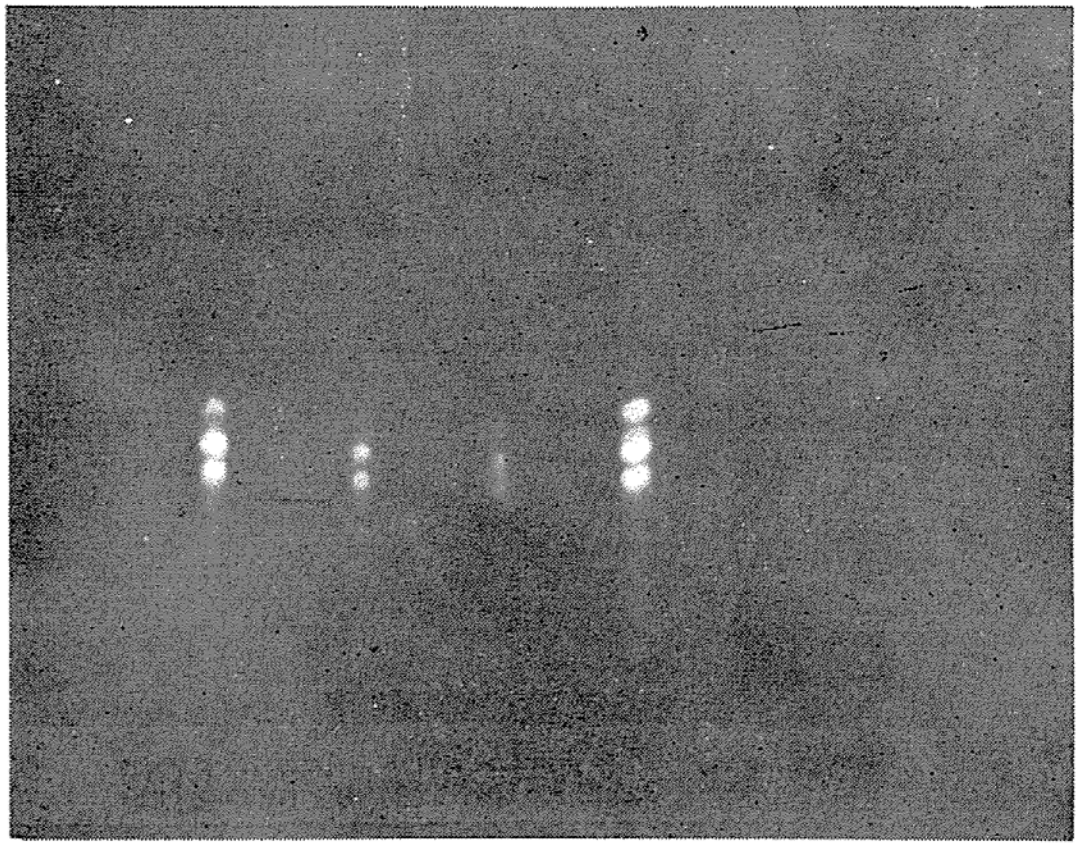


Fig. 1. Thin layer Chromatography plate developed with benzene-ethanol (9:1 v/v) showing the comparative yields of the crude aflatoxin extracts.

the hulls had to be thoroughly mixed with the dehulled beans to get a reduction in aflatoxin yield.

The Skelly B, $CHCl_3$ and ether extracts of the hulls reduced the yield of aflatoxin in dehulled beans with GAN medium (Table 3, Figure 2). The control sample did not show the same extensive decrease in aflatoxin production, indicating that the solvent is not responsible for the low toxin yield.

The Skelly B extract of the hulls reduced the aflatoxin yield in rice to one-third of the maximal yield obtained in rice alone (Table 3). Impurities that absorb at 260 mμ in the crude toxin from soybeans were absent in the crude toxin from rice thus giving a more typical ultraviolet absorption spectrum.

Figure 1 shows a decrease in fluorescent intensity of all aflatoxin spots from samples to which hull or hull extracts had been added. More aflatoxin B was produced in rice than in soybeans.

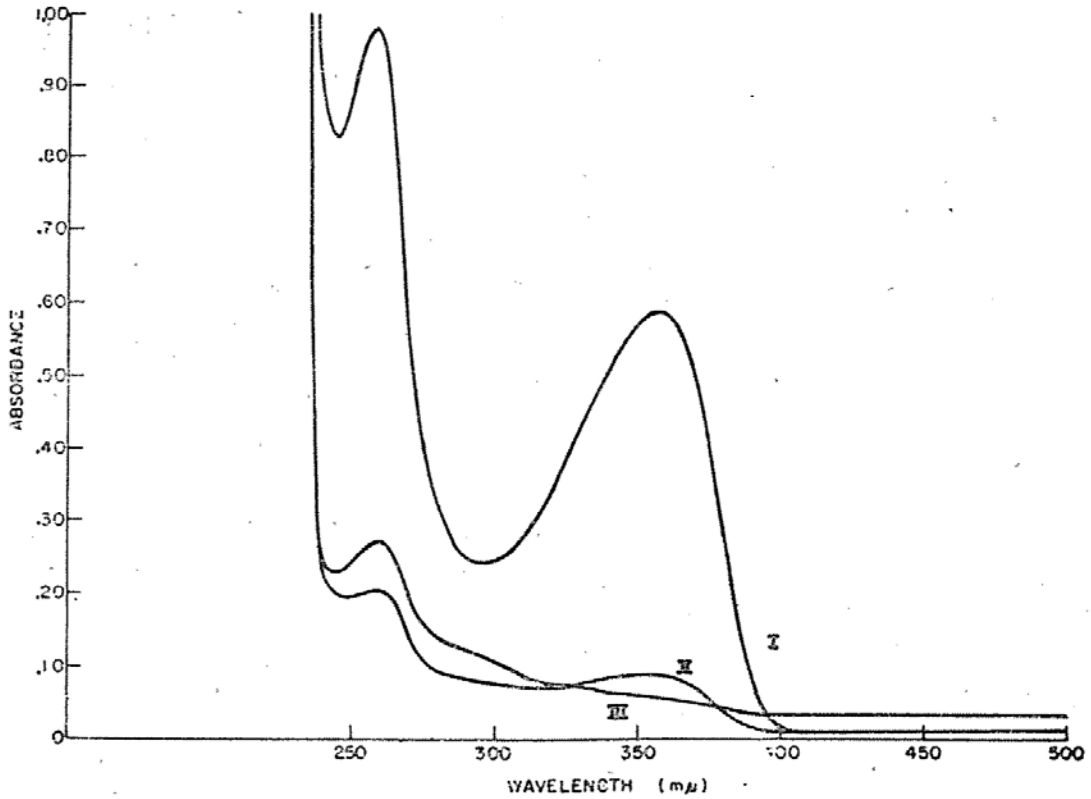


Fig. 2. UV absorption spectra of the crude aflatoxins extracted from:
I. Dehulled soybeans.
II. Dehulled soybeans with 5 g of hulls.
III. Dehulled soybeans with Skelly B extract of hulls.

Determination of aflatoxin degradation. Aflatoxin incubated for 7 days with the Skelly B extract of soybean hulls showed no decrease in ultraviolet absorption at 355 mμ although the spectrum of the incubated mixture showed some changes probably due to oxidation products.

Effect of GAN medium on aflatoxin production in soybeans. Table 2 shows poor aflatoxin yields in all soybean samples, with or without hulls, in the absence of GAN medium. The effect of the various components in GAN medium were tested by adding glucose and the remaining ingredients separately to sterilized soybean whey. The inoculated samples were incubated for 7 days at 25°C, extracted with $CHCl_3$ and the absorbance at 355 mμ was determined on the extracts. The acid precipitated protein was treated similarly with the following results:

	As at 355 mμ
1. Soybean whey with glucose only	0.32
2. Soybean whey with remaining components of GAN medium	0.04
3. Precipitated protein with glucose only	0.27
4. Precipitated protein with remaining components of GAN medium	0.04

DISCUSSION

The purpose of this investigation was to determine the cause of poor aflatoxin production in soybeans. The composition of soybean was reported to include the following: 14% sugars (3.3% sucrose, 4.8% galactose and 1.13% raffinose), 30–50% protein and trace amounts of Fe, Mn, Zn, Cu, and I[17,21]. It appears that soybeans have all the necessary components of a good substrate for aflatoxin production. Our results, however, showed low yields of aflatoxin in all the soybean fractions. GAN medium was added to all the fractions to determine if supplementary nutrients would increase the yield of aflatoxin in soybeans. Results (Table 2) showed that only certain fractions gave considerable increase in aflatoxin yield upon the addition of GAN medium; and these were the fractions in which soybean hulls were absent. Results indicate that the soybean *per se* is a poor substrate for aflatoxin production. Further results indicate that this may be due to an insufficiency or unavailability of a good carbon source since the addition of glucose alone enhanced aflatoxin production whereas addition of the rest of the components in GAN medium did not produce this effect.

However, results (Table 2) also indicate that the hull itself has an inhibitory effect on toxin production. When hull was present, aflatoxin levels were low even upon enrichment with GAN medium.

In the absence of hulls and with GAN medium a decrease in aflatoxin production was demonstrated with defatting. The role of lipids in aflatoxin production has as yet not been carefully examined by anyone although it is known that *Aspergillus flavus* is a strongly lipolytic mold[12,15,16]. Diener and Davis[8] studied the relation of

free fatty acid and aflatoxin production in peanuts and observed that aflatoxin yields and fatty acid concentrations were proportionally correlated at high relative humidity and moderate temperatures (15–40°C), but at extreme temperatures, aflatoxin formation may not correlate with fatty acid concentration. In a survey made by the Tropical Products Institute[11] on the natural occurrence of aflatoxin in agricultural products, significant amounts were found only in peanuts and cottonseed. This would suggest that aflatoxin formation *in vivo* can occur more readily in oil seeds than in other agricultural products, given optimal conditions. In our study, we are concerned mainly with the decrease in toxin production due to the hulls.

The results shown in Table 3 indicate that the active hull com-

Table 3. The effect of hulls and hull extracts on aflatoxin production in dehulled soybeans and rice

		Mold growth	A 355 mμ
A. Additions to dehulled soybeans+GAN			
No additions	A	++++[a]	2.95[b]
	B	+++	1.95
5 g hulls	A	+++	0.45
	B	+++	0.50
Skelly B extract	A	+++	0.45
	B	+++	0.30
Chloroform extract	A	+++	0.55
	B	+++	0.60
Ether extract	A	+++	0.60
	B	+++	0.60
Skelly B alone		+++	1.80
B. Rice	A	+++	3.10
	B	+++	2.50
Rice with hull extract		+++	1.10

[a]: As estimated visually: ++++, abundant growth covering entire surface of culture medium; +++, growth covering only a part of the surface; +, very little growth.

[b]: Absorbance of the $CHCl_3$ extract per 4 grams of soybeans.

ponent is lipid soluble. This could explain the fact that in an aqueous medium where the hulls were not evenly dispersed throughout the whole sample, such as in whole beans, the yield of aflatoxin was erratic and was probably dependent on whether the molds were able to penetrate the hull barrier or not. When hulls were extracted with various solvents and added to dehulled soybeans, the inhibition of aflatoxin production was of the same magnitude as that exhibited by the hull itself.

Soybean hull extract also was added to rice to determine if this effect could be duplicated in other substrates besides soybeans. Rice is reputed the best substrate for aflatoxin production for this particular strain of *A. flavus*. Under the same incubation conditions used with soybean samples, rice gave a toxin yield comparable to the highest yield obtained in dehulled soybeans supplemented with GAN medium. Upon addition of the hull extract, the yield was reduced to one-third (Table 3). These results indicated that low aflatoxin production can be induced in rice by the addition of soybean hulls.

A recent study on the production of aflatoxins in soybeans[9] showed that in nonautoclaved soybeans, the production of aflatoxin was poor and was claimed to be due to the unavailability of zinc which is bound to phytic acid. The edible portion of soybeans is said to have a phytic acid content of 690 mg per 100 g of beans as compared to 190 mg in rice. Autoclaved samples (15 min at 15 psi) gave a twenty fold increase in aflatoxin production. The authors argued that autoclaving breaks down the structure of phytic acid thus releasing zinc. In our experiments, where all the ingredients of the medium were autoclaved at 15 psi for 15 min, the presence of the hulls or hull extract caused decreased production. This decrease in production must be due to another factor besides phytic acid. Aflatoxin production in soybeans is, however, known to depend on the variety of soybeans and the toxin producing ability of the mold isolate used. Different varieties of soybeans could have different compositions and mold strains can be affected differently by these constituents of the beans.

We have investigated the possibility of aflatoxin degradation by components of the hull. Different suggestions have been given as to the nature and cause of aflatoxin degradation. These include a remetabolism of the aflatoxin by the fungus[18] and a nonenzymatic and nonspecific reaction with some sort of oxidizing system[4]. Aflatoxin degradation seems to take place only with certain strains of molds; however, Ciegler *et al.*[4] noted that nondegrading strains can be induced to degrade aflatoxin by supplying conditions that will produce mycelial lysis. Our experimental conditions were not those conducive to aflatoxin degradation and the mold strain used was a reported nondegrader. Our results show that hull components have no effect on aflatoxin that is already produced by the mold.

REFERENCES

(1) Adye, J., and R. I. Mateles. 1964. Incorporation of labelled compounds into aflatoxins. *Biochem. et Biophys. Acta* **86**, 421-422.

(2) Ashworth, L. J., Jr., H. W. Schroeder, and B. C. Langley. 1965. Aflatoxins: Environmental factors governing occurrence in Spanish peanuts. *Science* **148**, 1228-1229.

(3) Brian, P. W., A. W. Dawkins, J. F. Grove, H. G. Hemming, D. Lowe, and G. L. F. Norris. 1961. Phytotoxic compounds produced by *Fusarium equiseti*. *J. Exptl. Botany* **12**, 1-12.

(4) Ciegler, A., R. E. Peterson, A. A. Lagoda, and H. H. Hall. 1966. Aflatoxin production and degradation by *Aspergillus flavus* in 20-liter fermentors. *Appl. Microbiol.* **14**, 826-833.

(5) Davis, N. D., U. L. Diener, and P. Agnihotri. 1967. Production of aflatoxins B_1 and G_1 in chemically defined medium. *Mycopath. Mycol. Appl.* **31**, 251-256.

(6) Davis, N. D., U. L. Diener, and D. W. Eldridge. 1966. Production of aflatoxins B_1 and G_1 by *Aspergillus flavus* in a semisynthetic medium. *Appl. Microbiol.* **14**, 378-380.

(7) Diener, U. L., and N. D. Davis. 1966. Aflatoxin production by isolates of *Aspergillus flavus*. *Phytopath.* **56**, 1390-1393.

(8) Diener, U. L., and N. D. Davis. 1967. Limiting temperature and relative humidity for growth and production of aflatoxin and free fatty acids by *Aspergillus flavus* in sterile peanuts. *J. Am. Oil Chem. Soc.* **44**, 259-263.

(9) Gupta, S. K., and T. A. Venkitasubramanian. 1975. Production of aflatoxin on soybeans. *Appl. Microbiol.* **29**, 834-836.

(10) Hesseltine, C. W., O. L. Shotwell, J. J. Ellis, and R. D. Stubblefield. 1966. Aflatoxin formation by Aspergillus flavus. *Bacteriol. Rev.* **30**, 795-805.

(11) Hiscocks, E. S. 1965. *The importance of molds in the deterioration of tropical foods and feedstuffs.* P. 15-26, in G. N. Wogan (ed.), Mycotoxins

in foodstuffs. M. I. T. Press, Cambridge, Mass.
(12) Kirsh, D. 1935. Lipase production by *Penicillium oxalicum* and *Aspergillus flavus*. *Botanical Gazette* **97**, 321-333.
(13) Lee, E. G. H., P. M. Townsley, and C. C. Walden. 1966. Effect of bivalent metals on the production of aflatoxins in submerged cultures. *J. Fd. Sci.* **31**, 432-436.
(14) Mateles, R. I., and J. Adye. 1965. Production of aflatoxins in submerged culture. *Appl. Microbiol.* **13**, 208-211.
(15) Mayne, R. Y. 1956. Lipolytic microorganisms associated with prime and deteriorated cottonseed. *Appl. Microbiol.* **4**, 263-269.
(16) Mayne, R. Y. 1956. Lipolytic activity of various microorganisms isolated from cottonseed. *Appl. Microbiol.* **4**, 270-273.
(17) Morse, W. J. 1950. *Chemical composition of soybean seed*, p. 135-156, in K. S. Markley (ed.), *Soybeans and soybean products*. Interscience Publishers, Inc., New York, N. Y.
(18) Schindler, A. F., J. G. Palmer, and W. V. Eisenberg. 1967. Aflatoxin production by *Aspergillus flavus* as related to various temperatures. *Appl. Microbiol.* **15**, 1006-1009.
(19) Schroeder, H. W. 1966. Effect of corn steep liquor on mycelial growth and aflatoxin production in *Aspergillus parasiticus*. *Appl. Microbiol.* **14**, 381-385.
(20) Schroeder, H. W., and H. Hein, Jr. 1967. Aflatoxins: Production of the toxins *in vitro* in relation to temperature. *Appl. Microbiol.* **15**, 441-445.
(21) Winton, A. L., and K. B. Winton. 1932. *The structure and composition of foods*, Vol. I. John Wiley and Sons, Inc., New York, N. Y.
(22) Wogan, G. N. 1966. Chemical nature and biological effects of the aflatoxins. *Bact. Rev.* **30**, 460-470.

CONTRIBUTORS TO THIS NUMBER

Chi-Ling Hsu 徐繼玲 Ph. D. is associate professor of Mathematics at Fu Jen University

Jen-I Chen 陳振義 Ph. D. is associate professor of Physics at Fu Jen University

I-Fu Shih 施義夫 Ph. D. is associate professor of Physics at Fu Jen University

Fong-Jen Lin 林豐仁 Ph. D. is associate professor of Physics at Fu Jen University

Urban E. Schnaus, OSB 舒納思 M. S. is associate professor of Physics at Fu Jen University

Franz Huber SVD 扈伯爾 Ph. D. is professor and chairman of Biology at Fu Jen University

Ching-Shia Chen 陳擎霞 M. S. is lecturer in the Biology department at Fu Jen University

Tsang-Cheng Shao 邵燦成 Ph. D. is associate professor of Nutrition and Food Science at Fu Jen University

Evamonica M. Jamlang 何木蘭 Ph. D. is associate professor and head of Nutrition and Food Science at Fu Jen University

PRINTED BY

Ching Hua Press Co., LTD., Taipei

輔仁學報

第九號

發行者　私立輔仁大學

出版者　私立輔仁大學理學院
臺北縣新莊鎮中正路五一〇號

承印者　精華印書館股份有限公司
臺北市長沙街二段七十一號
電話:3313276, 3313429, 3313707

中華民國六十四年十二月三十日出版