INTRODUCTION TO SUPERSYMMETRY

Edward Witten[*]

Joseph Henry Laboratories
Princeton University
Princeton, New Jersey 08544

I. INTRODUCTION

Supersymmetry is a remarkable subject that has fascinated particle physicists since it was originally introduced.[1] Although supersymmetry is no longer a new idea, we still do not know in what form, if any, it plays a role in the proper description of nature.

The present status of supersymmetry might be compared very roughly to the status of non-Abelian gauge theories twenty years ago. It is a fascinating mathematical structure, and a reasonable extension of current ideas, but plagued with phenomenological difficulties.

In these lectures, I will present an introduction to supersymmetry, or at least to some aspects of this extensive subject. I also will describe some recent results. Supersymmetry has the reputation of a subject that is difficult to learn. I will try to at least partially dispel this unjustified impression.

In these lectures we will discuss only global supersymmetry, not supergravity.

A number of useful reviews of supersymmetry are available, which approach the subject from several points of view.[2] Many aspects of supersymmetry that I will not be able to discuss are treated in these review articles.

II. BOSE-FERMI SYMMETRY

If one considers a theory of two decoupled Bose fields, ϕ_1, and ϕ_2, so

$$\mathcal{L} = \int d^4 x \, \frac{1}{2}((\partial_\mu \phi_1)^2 + (\partial_\mu \phi_2)^2) \tag{1}$$

it is possible to combine ϕ_1 and ϕ_2 into a conserved current $J_\mu = \phi_1 \, \partial_\mu \phi_2 - \phi_2 \, \partial_\mu \phi_1$ even though they are not interacting. Actually, since, for example, $\partial_\alpha \phi_1$ or $\partial_\alpha \partial_\beta \phi_1$ satisfies the Klein-Gordon equation just as ϕ_1 does, one can form in the free field theory additional conserved currents such as

$$J_{\mu\alpha} = \partial_\alpha \phi_1 \, \partial_\mu \phi_2 - \phi_2 \, \partial_\mu \partial_\alpha \phi_1$$

$$J_{\mu\alpha\beta} = \partial_\alpha \partial_\beta \phi_1 \, \partial_\mu \phi_2 - \phi_2 \, \partial_\mu \partial_\alpha \partial_\beta \phi_1 \tag{2}$$

It is easy to check that these are conserved because $\nabla^2 \phi_1 = \nabla^2 \phi_2 = 0$.

The reason that J_μ is far more important than $J_{\mu\alpha}$ or $J_{\mu\alpha\beta}$ is that J_μ can still be conserved in interacting theories. One can add to $\mathcal{L}$ a term like $V(\phi_1^2 + \phi_2^2)$ that is invariant under $\delta\phi_1 = \phi_2$, $\delta\phi_2 = -\phi_1$. J_μ is still conserved.

However, when interactions are included, $J_{\mu\alpha}$ and $J_{\mu\alpha\beta}$ are no longer conserved. Moreover, it is not possible to redefine them (by adding extra terms to allow for the interactions) so that they will still be conserved. One may readily verify this in special cases. In general, it is a consequence of the Coleman-Mandula theorem.[3] Coleman and Mandula showed (basically by S matrix theory alone) that in a theory with non-zero scattering amplitudes in more than 1+1

dimensions the only possible conserved quantities that transform as
tensors under the Lorentz group are the following. The usual space-
time symmetries are certainly allowed: the energy-momentum operator
P_μ and the Lorentz transformations $M_{\alpha\beta}$ commute with the S matrix in
all of our usual theories. We may also have arbitrary Lorentz-
invariant conserved quantum numbers Q_i (electric charge, baryon
number, etcs.). Finally, if all particles are massless, the Coleman-
Mandula theorem allows conformal invariance, which is not usually
realized, however, in field theories with interactions. The Coleman-
Mandula theorem forbids "exotic" conservation laws — conservation
laws other than the usual space-time symmetries which do not commute
with Lorentz transformations.

For a proof, I refer you to reference (3); I will just give a
rough idea. The basic idea is that conservation of P_μ and $M_{\alpha\beta}$
leaves only the scattering angle unknown in (say) a two body col-
lision. Additional, exotic conservation laws would determine the
scattering angle, leaving only a discrete set of possible angles.
Since the scattering amplitude is always an analytic function of
angle, it would actually then have to vanish at all angles.

Note that the argument obviously does not apply in 1+1 dimen-
sions. In 1+1 dimensions, the only possible angles are 0 and π;
there is no such thing as analyticity as a function of scattering
angle. This is why, in 1+1 dimensions, it is possible to have in-
teracting systems, such as the sine-gordon equation, with exotic
conservation laws.

To illustrate the argument by a concrete example, suppose we
have a conserved traceless symmetric tensor $Q_{\beta\gamma}$. This is an exotic
conservation law because, transforming as a tensor, it does not com-
mute with Lorentz transformations. Its matrix element in a one
particle state of momentum p and (for simplicity) spin zero would
have to be $\langle p| \, Q_{\beta\gamma} \, |p\rangle = p_\beta \, p_\gamma - \frac{1}{4} \, g_{\beta\gamma} \, p^2$, by Lorentz invariance.

Applied to a two-body collision, with incident particles of momentum p_1, p_2, and outgoing particles of momentum q_1, q_2 (figure (1)), the conservation of $Q_{\beta\gamma}$ would tell us $p_{1\beta} \, p_{1\gamma} + p_{2\beta} \, p_{2\gamma} = q_{1\beta} \, q_{1\gamma} + q_{2\beta} \, q_{2\gamma}$.[*] This is possible only if the scattering angle is zero. With more effort, the same type of argument works even if the particles have non-zero spin.

Now going back to the exotic conserved currents that we defined in free field theory (equation (2)), the corresponding conserved charges

$$Q_\alpha = \int d^3x \, J_{o\alpha}$$

$$Q_{\alpha\beta} = \int d^3x \, J_{o\alpha\beta} \tag{3}$$

are forbidden by the Coleman-Mandula theorem for theories with a non-trivial S matrix in more than 1+1 dimension. This is why it is impossible to add interactions to the free field theory in a way that preserves the conservation of $J_{\mu\alpha}$ and $J_{\mu\alpha\beta}$.
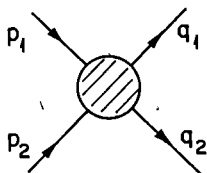

Figure 1

---

[*] It is here assumed that the matrix element of $Q_{\beta\gamma}$ in the two particle state $|p_1 \, p_2\rangle$ is the sum of the matrix elements in the states $|p_1\rangle$ and $|p_2\rangle$. This is true if $Q_{\beta\gamma}$ is the integral of a local current density, or, more generally, if $Q_{\beta\gamma}$ is defined in a way that is "not too non-local".

Instead of two non-interacting scalars, consider next a free massless charged scalar ϕ and a free two component (left handed) fermion ψ_α. The Lagrangian is

$$\mathcal{L} = \int d^4x \, (\partial_\mu \phi^* \, \partial^\mu \phi + \bar{\psi} \, i \, \not{\partial} \, \psi) \tag{4}$$

Again — although the fields are non-interacting — one can define conserved currents connecting them. One of the simplest is[*]

$$S_{\mu\alpha} = (\partial_\sigma \phi^* \, \gamma^\sigma \, \gamma_\mu \, \psi)_\alpha \tag{5}$$

It is easy to show $\partial_\mu S^\mu{}_\alpha = 0$:

$$\partial_\mu S^\mu{}_\alpha = (\partial_\mu \, \partial_\sigma \phi^* \, \gamma^\sigma \, \gamma^\mu \, \psi)_\alpha + (\partial_\sigma \phi^* \, \gamma^\sigma \, \gamma^\mu \, \partial_\mu \, \psi)_\alpha \tag{6}$$

The second term vanishes since $\gamma^\mu \, \partial_\mu \psi = 0$. For the first term, note $\partial_\mu \, \partial_\sigma \psi$ is symmetric in μ and σ, so we may replace $\gamma^\sigma \, \gamma^\mu$ by $1/2(\gamma^\mu \, \gamma^\sigma + \gamma^\sigma \, \gamma^\mu) = g^{\mu\sigma}$. We then have $g^{\mu\sigma} \, \partial_\mu \, \partial_\sigma \phi^* = 0$.

Again, additional conserved currents can be defined. The only property of ψ that we used was the Dirac equation, which is also satisfied by $\partial_\gamma \psi$, so

$$S_{\mu\gamma\alpha} = (\partial_\sigma \phi^* \, \gamma^\sigma \, \gamma_\mu \, \partial_\gamma \, \psi)_\alpha \tag{7}$$

is also conserved, that is, $\partial_\mu S^\mu{}_{\gamma\alpha} = 0$.

The exciting fact that makes supersymmetry interesting is that, although conservation of $S_{\mu\gamma\alpha}$ is always ruined when interactions are included, it is possible to add interactions in such a way that $S_{\mu\alpha}$ is conserved. A simple example is

$$\mathcal{L} = \int d^4x \, (\partial_\mu \phi^* \, \partial^\mu \phi + \bar{\psi} \, i \, \not{\partial} \, \psi - g^2 \, |\phi|^4 - g(\phi \, \psi_\alpha \, \psi^\alpha + h.c.)) \tag{8}$$

which is known as the "massless Wess-Zumino model". Although $S_{\mu\alpha}$

---

[*] $S_{\mu\alpha}$ is here a left-handed Weyl spinor, like ψ itself. We will soon go over to a Majorana basis.

as previously defined is not conserved in this model, by adding an extra term

$$S_{\mu\alpha} \rightarrow S_{\mu\alpha} + ig \, \gamma_\mu \, \phi^{*2} \, \psi^* \tag{9}$$

One preserves conservation of the current in the presence of interaction. (Here ψ^* is a <u>right</u>-handed two component spinor, the hermitian conjugate of ψ, which is <u>left</u>-handed. See, for example, the notes of Wess and Bagger[2] for more details.)

Why can $S_{\mu\alpha}$, but not $S_{\mu\gamma\alpha}$, be conserved in the presence of interactions? We can find out by studying the conserved charges

$$\hat{Q}_\alpha = \int d^3x \, S_{o\alpha}$$

$$\hat{Q}_{\gamma\alpha} = \int d^3x \, S_{o\gamma\alpha} \tag{10}$$

We cannot apply the Coleman-Mandula theorem directly to the $\hat{Q}_\alpha$ and $\hat{Q}_{\gamma\alpha}$, because the Coleman-Mandula theorem deals with conserved charges that transform as Lorentz tensors, while the $\hat{Q}$ transform as spinors (or vector-spinors). However, the Coleman-Mandula theorem can be applied to the bosonic conserved charges that can be formed from the anticommutators of the $\hat{Q}$. (It is the <u>anti</u>-commutators of the $\hat{Q}$ that we should consider, because $S_{\mu\alpha}$ and $S_{\mu\gamma\alpha}$, being linear in fermi fields, <u>anti</u>-commute at spacelike separation.)

Now $\hat{Q}_\alpha$, being a left-handed spinor, transforms as $(1/2, 0)$ under Lorentz transformations. Its hermitian adjoint, $\hat{Q}^*_\sigma$, transforms as $(0, 1/2)$. The anticommutator of $\hat{Q}$ with its adjoint $\hat{Q}^*$, which cannot vanish (since the anticommutator of any operator with its hermitian adjoint is non-zero), transforms as $(1/2, 1/2)$ under Lorentz transformations. The Coleman-Mandula theorem permits the conservation of precisely one operator that transforms as $(1/2, 1/2)$, namely the energy-momentum operator P_μ. But $\hat{Q}_{\gamma\alpha}$, a vector-spinor, has components of spin up to $3/2$. The anti-commutator $\{\hat{Q}_{\gamma\alpha}, \hat{Q}^*_{\sigma\tau}\}$, which

cannot vanish, for the reason noted above, has components of spin up to three. Since $\{\hat{Q}_{\gamma\alpha}, \hat{Q}^*_{\sigma\tau}\}$ is conserved if $\hat{Q}_{\gamma\alpha}$ is, and since the Coleman-Mandula theorem does not permit conservation of an operator of spin three in an interacting theory, $\hat{Q}_{\gamma\alpha}$ cannot be conserved in an interacting theory.

The Coleman-Mandula theorem does permit conservation of the $\hat{Q}_\alpha$. Changing basis from the $\hat{Q}_\alpha$ (not hermitian as I have defined them) to a basis of four hermitian operators Q_α that transform as a (real, Majorana) Lorentz four-spinor, the algebra of the Q_α turns out to be

$$\{Q_\alpha, \bar{Q}_\beta\} = \gamma^\mu_{\alpha\beta} \, P_\mu \tag{11}$$

One may readily check that this is so in, for example, the free field theory (4).

The right-hand side of (11) certainly contains only operators that are permitted by the Coleman-Mandula theorem, but one might ask whether there are more general possibilities. This question was answered by Haag, Sohnius, and Lopuszanski.[4] Since Q_α transforms as (1/2, 0) + (0, 1/2), its anti-commutator with itself might contain pieces transforming as (1, 0), (0, 1), or (0, 0), apart from the (1/2, 1/2) term we have already encountered. By the Coleman-Mandula theorem, the only conserved operators transforming as (1, 0) or (0, 1) are the Lorentz generators $M_{\mu\nu}$; however, Haag, Sohinius, and Lopuszanski showed that it is impossible to introduce $M_{\mu\nu}$ in (11) without violating the Jacobi identity. As for the possibility of adding to (11) operators that transform as (0, 0) (in other words, operators that commute with Lorentz transformations), this is possible, but only in the "extended supersymmetry" theories in which there are several conserved spinors $Q_{\alpha i}$. Beautiful though those theories are, I believe that they are too restrictive to describe the physics we know at energies much less than the Planck mass.[*] At such high

---

[*] The basic problem is that they require the charged fermions to form a "real representation" of the gauge group.

energies, they may, of course, be relevant. In these lectures we
will only consider the simplest theories with a single conserved
spinor Q_α and the basic algebra (11).

The algebra (11) has some dramatic consequences. Since Q_α is
hermitian, $\bar{Q}_\beta = Q_\sigma \, \gamma^o_{\sigma\beta}$; so (11) can be written

$$\{Q_\alpha, \, Q_\sigma\} \, \gamma^o_{\sigma\beta} = \gamma^\mu_{\alpha\beta} \, P_\mu \tag{12}$$

If we multiply by $\gamma^o_{\beta\alpha}$, sum over β and α, and use the facts $(\dot{\gamma}^o)^2 = 1$,
Tr $\gamma^o \, \gamma^\mu = 4\delta^{o\mu}$, we get

$$4P_o = \sum_\alpha \{Q_\alpha, \, Q_\alpha\} \tag{13}$$

Here P_o is, of course, the Hamiltonian H. For any operator A,
$\{A, \, A\} = 2A^2$. Equation (13) is equivalent to

$$H = \frac{1}{2} \sum_\alpha Q_\alpha^2 \tag{14}$$

which is one of the keys to understanding supersymmetry.

It follows immediately from equation (14) that if supersym-
metry is not spontaneously broken — if the Q_α annihilate the vacuum
$|\Omega\rangle$ — then the energy of the vacuum is zero. If $Q_\alpha |\Omega\rangle = 0$ then
obviously

$$H \, |\Omega\rangle = \frac{1}{2} \sum_\alpha Q_\alpha^2 \, |\Omega\rangle = 0 \tag{15}$$

If conversely, supersymmetry is spontaneously broken, that is, if
$Q_\alpha |\Omega\rangle \neq 0$, then

$$\langle\Omega| \, H \, |\Omega\rangle = \frac{1}{2}\sum_\alpha \langle\Omega| \, Q_\alpha^2 \, |\Omega\rangle = \frac{1}{2}\sum_\alpha |Q_\alpha \, |\Omega\rangle|^2 > 0 \tag{16}$$

so in this case the vacuum energy is positive.[*]

---

[*] Usually, one is free to add a constant to the Hamiltonian, but here
the zero of energy is fixed in a natural way by equation (14). When
asking what is the energy of the vacuum, we always have in mind the
definition (14) of H.

Combining these remarks, we see that supersymmetry is spontaneously broken 'if and only if the energy of the vacuum is greater than zero. This is often illustrated by the diagram of figure (2). In figure (2(a)) a scalar field has a vacuum expectation value, possibly breaking some internal symmetry. However, supersymmetry is unbroken because the ground state energy — the minimum of the potential — is zero. In figure (2(b)), the expectation value of the scalar field is zero, but supersymmetry is spontaneously broken because the energy of the ground state is greater than zero.

If supersymmetry plays any role in nature it must be spontaneously broken. This follows immediately from the fact that the Q_α have spin 1/2, so acting on any particle they change its spin by ±1/2:

$$Q_\alpha \,|\text{spin } s> \,\sim\, |\text{spin } s \pm 1/2> \qquad\qquad (17)$$

Since Q_α commutes with the Hamiltonian it does not change the mass of the particle on which it acts. Since we do not observe in nature the degeneracies among particles of different spin that would be predicted by (17), supersymmetry must be spontaneously broken if it is relevant to nature.

(a)   (b)

Figure 2

A puzzle immediately presents itself. Since broken supersymmetry means a positive vacuum energy, why does not a cosmological constant arise as soon as a supersymmetric theory is coupled to gravity? No good answer is known to this question, and it certainly is one of the most important questions that must be answered. Actually the positive vacuum energy of global supersymmetry does not necessarily become a positive cosmological constant after coupling to gravity, because the coupling to gravity introduces extra terms in the scalar potential; these extra terms are not positive definite. While a cancellation can occur,[5] leaving zero cosmological constant, such a cancellation apparently depends on a special choice of parameters, for which there is no known rationale.

Now, the subject of spontaneously broken supersymmetry has a special flavor, which arises from the fact that (leaving gravity aside), supersymmetry is spontaneously broken if and only if the ground state energy is positive. Suppose that in some theory an approximate calculation gives an effective potential such as the one shown in fig. (3). In this approximation, the minimum of the potential is zero, and 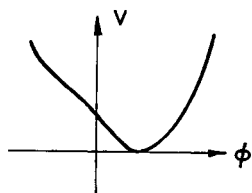supersymmetry is unbroken. Even if the approximation in question is very accurate, one cannot be certain, just from this result, that supersymmetry really is unbroken, because the errors in one's calculation may shift the potential slightly away from $V = 0$ (figure (4)) (the approximate calculation and the exact result are indicated in figure (4) by the solid and dotted line, respectively).

Figure 3

If, instead, an approximate calculation shows that supersymmetry is spontaneously broken, one can be confident in the result if
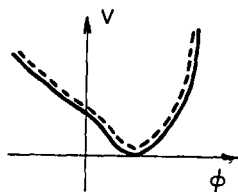
Figure 4   Figure 5
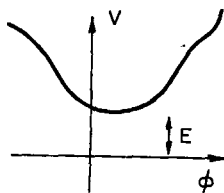
(figure (5)) the calculated vacuum energy E is much bigger than the
uncertainty ΔE in the calculation.

On the other hand, for ordinary symmetries, a reliable approxi-
mation (like perturbation theory, in a weakly coupled theory) can
reliably indicate whether the symmetry is spontaneously broken.

I have discussed these matters in much more detail in a recent
paper.[6] In that paper I also described a simple quantum mechanics
model in which tiny quantum effects shift the energy slightly away
from zero, as in figure (4).

III. SUPERSYMMETRY AND THE LARGE NUMBERS

Supersymmetry is a very beautiful idea, but I think it is fair
to say that no one knows what mysteries of nature (if any) it should
explain. To fix ideas, I will make some definite assumptions in the
next few lectures about what problems supersymmetry might solve.

Certainly, the ultimate applications may be in a very different area.

Perhaps supersymmetry is spontaneously broken at very high energies — energies that may be as high as the Planck mass 10^{19} GeV. We will assume instead in most of our discussion that supersymmetry is, more or less, within reach of the new generation of accelerators.

What would we like to explain with supersymmetry? Of the many possible answers, I will focus on Dirac's "problem of the large numbers,"[7] which nowadays is often called the "gauge hierarchy problem." As posed by Dirac, the question is why the Planck mass M_{Pl} is so much larger than the nucleon mass M_N: $M_{Pl}/M_N \sim 10^{+19}$. In a contemporary version,[8] one might ask why the mass scale M_X of presumed grand unification is so much larger than the mass scale M_W of weak interactions: $M_X/M_W \gtrsim 10^{+13}$. Starting with Dirac, most physicists have believed that such large numbers should not be postulated arbitrarily but must have a definite explanation.

We now know that the "normal" masses — the quark, lepton, and W and Z masses — are determined by SU(2) x U(1) breaking. Specifically, they are all proportional to the expectation value $<\phi>$ of the Higgs boson. So the question is really why $<\phi>$ is so tiny compared to the "large" masses — the mass scale of grand unification, or the Planck mass.

Looking at the standard Higgs potential

$$V(\phi) = \frac{\lambda}{4} \cdot \phi^4 - \frac{m_\phi^2}{2} \phi^2 \tag{18}$$

we have $<\phi> = m/\sqrt{\lambda}$, so the real problem is to explain why $m_\phi << M_{Pl}$.

Why would supersymmetry be relevant to this?

As we now understand it, bare mass terms for the quarks and leptons are forbidden by SU(2) x U(1) gauge invariance. Left-handed quarks and leptons transform as doublets of SU(2) x U(1), but right-

handed quarks and leptons are singlets. So bare masses are impossible, and the quarks and leptons get masses only from SU(2) x U(1) breaking. We do not know why $<\phi>$ is so small (relative to the large masses of physics) but if this could be explained, the lightness of the quarks and leptons would follow.

The problem with the Higgs boson is that its bare mass does not violate SU(2) x U(1), or any other gauge symmetry. Given any Higgs multiplet ϕ^i, the mass term $\sum_i \phi_i^* \phi^i$ is always gauge invariant. We would like a symmetry violated by m_ϕ, to explain why ϕ is so light. The symmetry will hopefully be spontaneously broken, but only on a very small mass scale, to explain why $<\phi>$ is tiny but non-zero.

Supersymmetry can do this if it relates the Higgs doublet $\begin{pmatrix} \phi^o \\ \phi^- \end{pmatrix}$ to fermions like $\begin{pmatrix} e^- \\ e^- \end{pmatrix}_L$ whose masses violate SU(2) x U(1). Then $m_\phi = 0$ as long as supersymmetry and SU(2) x U(1) are unbroken. And m_ϕ is small, solving the old problem of the "large numbers," if we can understand that supersymmetry and SU(2) x U(1) are weakly broken.

So that we can discuss these matters in a more tangible way, we must have available the explicit form of some supersymmetric Lagrangians. Let us consider first supersymmetric theories with fields of spin zero and spin one half only. In such theories, one may have an arbitrary number of complex scalar fields A^i, their supersymmetric partners being left-handed spinor fields ψ_L^i. We introduce a function W that depends only on the A^i, not on their complex conjugates A_j^* — in other words, W is an analytic function of the A^i. W is usually called the "superspace potential".

The Lagrangian is $\mathcal{L} = \mathcal{L}_{kinetic} + \mathcal{L}_{scalar} + \mathcal{L}_{Yukawa}$, where

$$\mathcal{L}_{kinetic} = \partial_\mu A_i^* \partial_\mu A^i + \bar{\psi}_i \, i \, \not{\partial} \, \psi^i$$

$$\mathcal{L}_{scalar} = - \sum_i \left| \frac{\partial W}{\partial A^i} \right|^2$$

$$\mathcal{L}_{Yukawa} = -\left(\frac{\partial^2 W}{\partial A^i \, \partial A^j} \, \psi_\alpha^i \, \psi^{\alpha j} + h.c. \right) \tag{19}$$

For how this construction was discovered, I refer you to the literature.[2] The easiest way to show that this describes a supersymmetric theory is to demonstrate that the supersymmetry current S^μ is conserved. This current is

$$S^\mu = \partial_\alpha A_i^* \gamma^\alpha \gamma^\mu \psi^i + i \frac{\partial W^*(A^*)}{\partial A_i^*} \gamma^\mu \psi_i^* \tag{20}$$

The first term in (20) we have already seen in free field theory; the second term is added due to the interactions.

Note that, for a renormalizable theory, W should be at most a cubic function of the A^i. If W is cubic, then (19) contains only terms of dimension four or less, and this corresponds to a renormalizable theory.

For our purposes, the most important part of (19) is the formula for the scalar potential,

$$V(A^i, A_j^*) = \sum_i \left| \frac{\partial W}{\partial A^i} \right|^2 \tag{21}$$

Let us ask, under what conditions is supersymmetry spontaneously broken at the tree level? Evidently, if for some value of the A^i the equations

$$\frac{\partial W}{\partial A^i} = 0 \tag{22}$$

are simultaneously satisfied, then for this value of the fields the potential energy vanishes, classically. Supersymmetry is unbroken. On the other hand, if the equations (22) are inconsistent — if they are not satisfied for any choice of the A^j — then the minimum of the potential is strictly positive, and supersymmetry is spontaneously broken.

The simplest example is the O'Raifeartaigh model.[9] There are three fields, A, X, and Y, and the superspace potential is.

$$W(A, X, Y) = g A Y + \lambda X (A^2 - M^2) \tag{23}$$

Here g, λ, and M are constants. This theory is technically natural, because of global symmetries.[*] The scalar potential is

$$V(A, X, Y) = \left|\frac{\partial W}{\partial A}\right|^2 + \left|\frac{\partial W}{\partial X}\right|^2 + \left|\frac{\partial W}{\partial Y}\right|^2$$

$$= g^2 |A|^2 + \lambda^2 |A^2 - M^2|^2 + |gY + 2\lambda AX|^2 \qquad (24)$$

This potential is strictly positive, and supersymmetry is spontaneously broken. In fact, $\partial W/\partial Y = 0$ only if $A = 0$, and $\partial W/\partial X = 0$ only if $A = \pm M$; these requirements are clearly inconsistent. If $g/\lambda M$ is large enough, the minimum of the potential is at $A = 0$ and the vacuum energy is $\lambda^2 M^4$, plus quantum corrections. Expanding around the minimum of the potential (and making use of our previous formulas for the Yukawa couplings as well as the scalar interactions), it is easy to see that the bosons and fermions have unequal masses, which is expected, since the positive ground state energy indicates spontaneous breaking of supersymmetry.

We will have more to say about this model later, but for the moment let us consider a model of another kind. Let us consider a theory with SU(5) symmetry, the only scalar field being a complex field $A^i{}_j$ in the adjoint representation of SU(5) (so Tr $A = 0$). The most general choice of W would be

$$W = \frac{g}{3} \text{Tr } A^3 + \frac{M}{2} \text{Tr } A^2 \qquad (25)$$

where g and M are constants. The equations $\partial W/\partial A^i{}_j = 0$ give

$$g((A^2)^i{}_j - \frac{1}{5} \delta^i{}_j \text{Tr } A^2) + M A^i{}_j = 0 \qquad (26)$$

---

[*] The theory is invariant under $A \rightarrow -A$, $Y \rightarrow -Y$ and under $Y \rightarrow e^{i\alpha}Y$, $X \rightarrow e^{i\alpha}X$. It should be noted that any transformation under which W changes only by an overall phase is a symmetry operation (the phase of W cancels out of the scalar potential and can be removed from the Yukawa couplings by a chiral transformation).

If we assume that A can be diagonalized by a unitary transformation,[*]
then it is easy to see that there are three solutions:

$$A^i{}_j = 0$$

$$A^i{}_j = \frac{M}{3g} \begin{pmatrix} 1 & & & & \\ & 1 & & & \\ & & 1 & & \\ & & & 1 & \\ & & & & -4 \end{pmatrix}$$

$$A^i{}_j = \frac{M}{g} \begin{pmatrix} 2 & & & & \\ & 2 & & & \\ & & 2 & & \\ & & & -3 & \\ & & & & -3 \end{pmatrix} \tag{27}$$

These three solutions correspond to the unbroken gauge groups SU(5),
SU(4) x U(1), and SU(3) x SU(2) x U(1), respectively. They each
correspond. to unbroken supersymmetry, and they are exactly degenerate
at zero energy at least in this approximation (figure (6)), because
they were all found by requiring $\partial W/\partial A^i{}_j = 0$.

Now, what really is the physics of this theory? It depends
entirely on the nature of the quantum corrections to the effective
potential. If really E = 0 for each of the three vacuum states, as
appears to be the case in the classical approximation, then this one
theory describes three different, inequivalent worlds. In one world,
the strong gauge group is SU(3) and a baryon is made from three
quarks; in one world, the strong gauge group is SU(4) and baryons are
bosons, made from four quarks; in one world the strong gauge group is
SU(5) and a baryon is made from five quarks.

If the quantum corrections break supersymmetry in, say, two of

---

[*]It is really when SU(5) is made into a gauge symmetry that. this
assumption is justified, because of extra terms that are then present
in the scalar potential. Since A is complex and not hermitian, it
cannot necessarily be diagonalized by a unitary transformation.

the three worlds (figure (7)), then
the true vacuum is the one in which
$E = 0$ and the supersymmetry is <u>not</u>
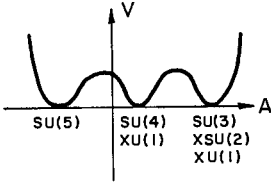spontaneously broken.

Figure 6

If supersymmetry is spontan-
eously broken, and $E \neq 0$, in each
of the three worlds, then (figure
(8)) the true vacuum is the one in
which supersymmetry is broken most
weakly and the vacuum energy is
least.

One might usually guess that such a degeneracy would be resolved,
in perturbation theory, by loop diagrams. Perhaps the vacuum energy
is of order α, coming from a one loop diagram, or of order α^2, coming
from a two loop diagram. The loop diagrams would then be inducing
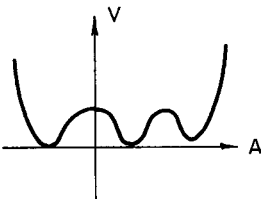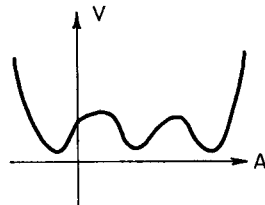supersymmetry breaking. Some particles whose masses violate super-

Figure 7

Figure 8

symmetry would get masses. Some scalars would get positive mass squared; some would get negative mass squared. If the W boson gets a mass from a Higgs doublet that gets an expectation value in this way, we would get a hierarchy of some sort, with $(M_W/M_X)^2$ perhaps of order α or α^2. This clearly would fall far short of the experimentally observed hierarchy.

However, a most remarkable theorem[10] states that this does not occur. It states that if at some point in field space the classical potential vanishes, then the effective potential vanishes at that point to all finite orders of perturbation theory. Our degeneracy is <u>not</u> lifted in perturbation theory.

This fact is an example of the "non-renormalization theorems" of supersymmetry. Here are some other examples:

(1) There is no renormalization of W in perturbation theory — neither finite nor infinite renormalization. (There is wave function renormalization. There are quantum corrections to the effective potential. There is no renormalization of W, if properly defined in terms of the coefficients of certain operators in the effective potential.)

(2) As long as supersymmetry is unbroken, any particle that is massless at the tree level is massless to all finite orders of perturbation theory — even if it was massless at the tree level only because of an arbitrary adjustment of parameters.

These theorems have been proved — on the basis of details of perturbation theory — but they are not well understood.

Returning to our model with the three degenerate vacuum states, if the relation $\langle \Omega | \, H \, | \Omega \rangle = 0$ were to break down in perturbation theory, and if this breakdown were the origin of the "light" masses, the resulting mass ratios would, as we have said, not be small enough

to solve Dirac's problem of the large numbers. However, if the mys-
terious theorem concerning non-renormalization of $\langle\Omega|\ H\ |\Omega\rangle$ were to
break down non-perturbatively — as it does in a quantum mechanics
model discussed in ref. (6) — then we might obtain a solution to
the large number problem. We might get a formula of the form

$$\frac{M_W}{M_X} \sim \exp - \frac{1}{\alpha} \tag{28}$$

and this could solve the "large number" or "gauge hierarchy" problem.

We are thus imagining a theory with two mass scales. At the
large mass scale, the scale of grand unification, a unified group
breaks down to SU(3) x SU(2) x U(1), or some other phenomenologically
acceptable group, but supersymmetry remains unbroken. At a vastly
smaller mass scale, the non-renormalization theorems break down,
supersymmetry is spontaneously broken, and the Higgs boson gets an
expectation value, breaking SU(2) x U(1) down to U(1).

What mechanism might be responsible for the tiny, non-pertur-
bative effects that make $\langle\Omega|\ H\ |\Omega\rangle$ non-zero and break supersymmetry?
With the present state of our knowledge, there are two obvious can-
didates. The obvious candidates are strong gauge forces, and grand
unified instantons.

Let us discuss first the effects of strong gauge forces. Con-
sider, as an example, the SU(5) theory that we discussed earlier.
Assume that the SU(5) coupling is gauged; this can be done in a way
compatible with supersymmetry and with our previous remarks. In each
of the three vacuum states that exist in perturbation theory, there
is an unbroken non-Abelian interaction that will become strong at low
enough energies. For instance, in the vacuum $A^i j = 0$, this is the
full SU(5) group. As we have learned in the last few years, strong
gauge forces cause a variety of non-perturbative effects, including
confinement, mass generation, and chiral symmetry breaking. Perhaps
in supersymmetric theories strong gauge forces also cause super-

symmetry breaking, presumably with the binding of a color singlet Goldstone fermion.[6,11]

Of course, in this SU(5) theory, the vacuum of most phenomenological interest is the one in which the unbroken gauge group is SU(3) x SU(2) x U(1). In this vacuum there are strong SU(3) gauge forces. But if they were responsible for supersymmetry breaking, we would get bose-fermi mass differences of order Λ_{QCD} (since that is the energy at which SU(3) becomes strong), and this is clearly unsatisfactory.

In the same theory a more promising mechanism would be "grand unified instantons," that is, instantons of the SU(5) group which do not lie in the unbroken SU(3) x SU(2) x U(1) subgroup. Such instantons are not afflicted with infrared divergences. They have a natural mass scale, the mass M_X of grand unification, and a natural, small coupling, the coupling α_G of grand unification. If such instantons were responsible for supersymmetry breaking, we would get bose-fermi mass splittings roughly of the order

$$m_B^2 - m_F^2 \sim M_X^2 \exp - \frac{2\pi}{\alpha_G} \tag{29}$$

which might be reasonable, for α_G of about 1/12.

Actually, the instanton mechanism seems to work in 2+1 dimensions, but, apparently, not in 3+1 dimensions.[6] This point is in need of further clarification.

Clearly, we must learn to analyze supersymmetry breaking in a non-perturbative way. Tomorrow, I will describe some steps in that direction. We will be able to derive some constraints on the possibility of supersymmetry breaking by strong gauge forces.

IV. $\underline{\text{Tr}}(-1)^F$

In our previous discussions we have seen that, in general, it is difficult to determine from an approximate calculation whether supersymmetry is spontaneously broken. Even if the supersymmetry seems to be unbroken in some approximation, it is always possible that a small, non-zero vacuum energy is induced by the corrections to the approximation in question.

To show that supersymmetry is unbroken in a given theory one must prove that the ground state energy is <u>exactly</u> zero. Since an approximate calculation will not accomplish this, in general, we must search for other methods. I will now describe an indirect method which in some theories can be used to prove that the ground state energy is <u>exactly</u> zero. Of course, a theory in which supersymmetry is <u>not</u> broken is not a candidate for describing nature. The purpose of proving that some theories do not break supersymmetry is to cut down on the range of options that must be considered.

As a technical convenience, we will formulate our theories in a finite volume, with periodic boundary conditions (such boundary conditions respect supersymmetry). Our goal is to find criteria under which we can prove that the ground state energy of a theory $E_0(V)$ is zero in any finite volume V. Since the large V limit of zero is zero, this implies that the ground state energy is also zero in the infinite volume limit, and therefore that supersymmetry is <u>not</u> broken in the infinite volume theory which is of real interest.

Supersymmetry implies that every state of $E = 0$ also has $\vec{P} = 0$. (Supersymmetry implies $E \geq 0$ in each frame; by Lorentz invariance this requires $E \geq |P|$; so a state of zero energy necessarily has $\vec{P} = 0$.) So, in trying to determine whether the ground state energy vanishes, we lose nothing by restricting ourselves to the sector of Hilbert space consisting of states of $\vec{P} = 0$.

Let Q be any one of the (hermitian) supersymmetry generators. There are several Q_α, of course, but we will only need one. In the $\vec{P} = 0$ sector, the supersymmetry algebra is particularly simple. It is simply

$$Q^{2'} = \frac{1}{2} H \tag{30}$$

(For nonzero $\vec{P}$, $\vec{P}$ would appear on the right-hand side of (30).)

Since $Q^2 = \frac{1}{2} H$, Q annihilates any states of zero energy. States of non-zero energy are not annihilated by Q. Rather, they are paired by the action of Q. Given any boson state $|b\rangle$ of non-zero energy E, Q acting on $|b\rangle$ gives a fermion state of $|f\rangle$. Q acting on $|f\rangle$ gives back $|b\rangle$. To be precise

$$Q |b\rangle = \sqrt{\frac{E}{2}} \ |f\rangle$$

$$Q |f\rangle = \sqrt{\frac{E}{2}} \ |b\rangle \tag{31}$$

if phases are chosen properly. (Because $Q^2 = \frac{1}{2} H$, the factors of $\sqrt{\frac{E}{2}}$ are consistent with $\langle b|b\rangle = \langle f|f\rangle = 1$, and the fact that Q is hermitian.)

I should explain that when I refer to a boson state $|b\rangle$, I do not mean a one particle state (a concept that in a finite volume is not really well defined) but any state of integral angular momentum. Likewise, $|f\rangle$ is any state of half integral angular momentum. Here, "angular momentum" refers to the 90° rotations such as $\exp - \frac{i\pi}{2} J_z$ which are well defined in the finite volume theory. If we define the operator

$$(-1)^F = \exp - 2\pi i J_z \tag{32}$$

which distinguishes bosons from germions, then a boson state $|b\rangle$ is any state that obeys $(-1)^F |b\rangle = |b\rangle$, and a fermion state $|f\rangle$ is any state that obeys $(-1)^F |f\rangle = -|f\rangle$. A boson state could be, for

example, any state that in the infinite volume limit goes over to a configuration of 92 neutrons.

Although the states of non-zero energy form bose-fermi pairs as indicated in equation (31), this is certainly not true for the states of zero energy. Any zero energy state, boson or fermion, is just annihilated by Q:

$$Q|b, \ E = 0> \ = \ Q|f, \ E = 0> \ = \ 0 \tag{33}$$

Again, this is so because $Q^2 = \frac{1}{2} H$. The zero energy states are singlets — one dimensional representations of supersymmetry.

The general form of the spectrum of a supersymmetric theory is indicated in figure (9). The states of non-zero energy are in bose-fermi pairs. The zero energy states are not paired in general since each one is separately annihilated by Q. They need not be equal in number. In the figure there are two boson states of zero energy, and one fermion.

. What happens when we change the parameters of this theory? (By "parameters", I mean the volume, the bare mass, and the coupling constant.)
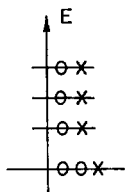
Under the change in parameters, the states of zero energy will, of course, move around in energy. However, they will move around in bose-fermi pairs. It may happen that as we change the parameters a bose state will move down to zero energy. If so, it will always be accompanied (figure (10)) by a fermi state moving down to zero

Figure 9

Circles indicate bosons; an "x" indicates a fermion.
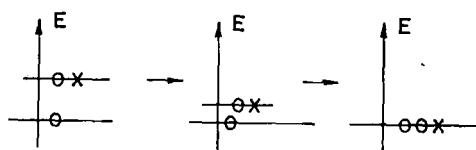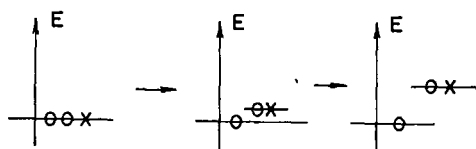
Figure 10

Figure 11

Circles indicate bosons; an "X" indicates a fermion.

energy. Conversely, as we change the parameters, a zero energy
state may get a non-zero energy. If so, as soon as it gets non-
zero energy, it must have a partner (figure (11)), because states
of non-zero energy are always in bose-fermi pairs. It is not pos-
sible however, for a zero-energy state to simply appear or disappear.
In quantum mechanics states always move around continuously in
energy.[*]

In the process of figures (10) and (11), the number of zero
energy states changes. However, the difference between the number
$n_B^{E=0}$ of zero energy states that are bosons and the number $n_F^{E=0}$
that are fermions does not change. This will be our basic tool.

Formally, the difference $n_B^{E=0} - n_F^{E=0}$ may be interpreted as
the trace of the operator, $(-1)^F$, that distinguishes bosons from
fermions. Formally, the states of non-zero energy cancel out of the
trace of $(-1)^F$ because they come in bose-fermi pairs. The trace of
$(-1)^F$ can therefore be evaluated among the zero energy states only,
and equals $n_B^{E=0} - n_F^{E=0}$. We will henceforth refer to $n_B^{E=0} - n_F^{E=0}$
as $Tr(-1)^F$. However, this is only a definition, since $Tr(-1)^F$ is
not absolutely convergent.

The fact that $Tr(-1)^F$ does not change when the parameters of a
supersymmetric theory are changed is an important fact, for the fol-
lowing reasons:

(1) If $Tr(-1)^F \neq 0$, supersymmetry is definitely not spontane-
ously broken.

(2) $Tr(-1)^F$ can be calculated reliably even in quite compli-
cated theories.

_____

[*]As discussed in detail in reference (12), there is really an impor-
tant caveat to be imposed here. One may not consider changes in
parameters that overwhelm the terms already present in the
Hamiltonian.

Let me explain these points:

If $Tr(-1)^F \neq 0$ then $n_B^{E=0} \neq 0$ or $n_F^{E=0} \neq 0$ or both. In any case, there are <u>some</u> zero energy states. Hence the ground state energy is zero and supersymmetry is unbroken. (Since $Tr(-1)^F$ is independent of the volume, a non-zero value of $Tr(-1)^F$ means that the ground state energy is zero for any V and hence also as $V \to \infty$.)

$Tr(-1)^F$ can be calculated reliably because it is independent of the parameters. We can calculate $Tr(-1)^F$ in some convenient limit, such as small volume, large bare mass, and weak coupling. Almost any theory simplifies enough in this limit (or some analogous limit) that $Tr(-1)^F$ can be calculated reliably. ·

The results can then be applied to the situation of interest — large volume, physical mass, and physical coupling — because $Tr(-1)^F$ is independent of all parameters.   .

It is important to realize that $Tr(-1)^F$ can be calculated reliably even though, in general, we may be unable to tell which states have exactly zero energy. Suppose that in some approximation we find in some theory the spectrum of figure (12(a)). In this approximation we have $n_B^{E=0} = 2$, $n_F^{E=0} = 1$, and $Tr(-1)^F = 1$. Even if the approximation is excellent we cannot be sure we have calculated $n_B^{E=0}$ or $n_F^{E=0}$ correctly. As a result of small errors in any approximation, the true answer might be that of figure (12(b)). This corresponds to $n_B^{E=0} = 1$, $n_F^{E=0} = 0$, and again $Tr(-1)^F = 1$. The hypothetical corrections to our approximation gave a small, non-zero energy to one boson and therefore (by supersymmetry) also to one fermion. The original approximation gave $n_B^{E=0}$ and $n_F^{E=0}$ incorrectly, but it gave $Tr(-1)^F$ correctly, essentially because the extra boson at $E = 0$ has no potential fermion partner and so no way to get $E \neq 0$.

Let me now explain the calculation of $Tr(-1)^F$ in a simple model. I will consider the Wess-Zumino model, which we have discussed
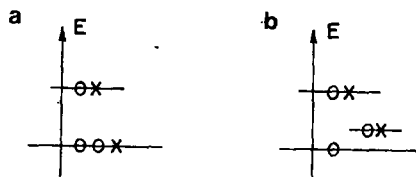
Figure 12

earlier. This model contains a single complex field ϕ (and fermi partner ψ) with

$$V(\phi) = g^2 \left| \phi^2 - \frac{a^2}{g^2} \right|^2 \tag{34}$$

(There also is a Yukawa coupling, $L_{Yuk} = g \, \phi \, \psi_\alpha \, \psi^\alpha + h.c.$). At the tree level ϕ and ψ are massive, their masses being equal to a, in lowest order. Supersymmetry is unbroken at the tree level since $V = 0$ at $\phi = \pm a/g$. Because of the non-renormalization theorems that we discussed yesterday, it is known that supersymmetry is unbroken to all finite orders of perturbation theory. Let us now prove that this is true independently of perturbation theory by showing that $Tr(-1)^F \neq 0$.

Actually, nothing could be easier. The potential has two minima, at $\phi = \pm a/g$. In each minimum of V, there is one zero energy state in perturbation theory — the "vacuum". It is a bosonic state, with

zero angular momentum. Because ϕ and ψ have non-zero mass, all other states have (at least for weak coupling) a non-zero energy, at least equal to the mass of the ϕ and ψ particles. They do not contribute to $\text{Tr}(-1)^F$. $\text{Tr}(-1)^F$ receives one contribution from the vacuum at $\phi = +a/g$ and one contribution from the vacuum at $\phi = -a/g$. Altogether $\text{Tr}(-1)^F = 2$, so supersymmetry is unbroken.

Suppose instead that we wish to calculate $\text{Tr}(-1)^F$ in the massless Wess-Zumino model — that is, with a = 0 in (34). The potential is now simply $V(\phi) = g^2 |\phi|^4$. Now $m_\phi = m_\psi = 0$, in perturbation theory. In addition to the "vacuum" (which now is a trickier concept), one can have in perturbation theory states of approximately zero energy by adding ϕ or ψ quanta to the "vacuum" in momentum eigenstates of $\vec{p} = 0$. (Such states are normalizable in finite volume.) It is difficult to count these states because any number of ϕ particles may have $\vec{p} = 0$ and it is difficult — because of the non-linearity — to know which of these states have E = 0 exactly, or to count them. This makes it difficult to calculate $\text{Tr}(-1)^F$.

The easiest way to calculate $\text{Tr}(-1)^F$ in the a = 0 model is to remember that $\text{Tr}(-1)^F$ is independent of a. So even if your interest is a = 0, you can calculate $\text{Tr}(-1)^F$ by considering a to be non-zero and large. This makes the calculation easy. So $\text{Tr}(-1)^F = 2$. This illustrates the utility of knowing $\text{Tr}(-1)^F$ to be independent of all parameters.

Going back to the massive model, with a $\neq$ 0, there actually is a simpler argument to prove that supersymmetry is unbroken for small enough g. For supersymmetry to be spontaneously broken,[2] there must be a massless Goldstone fermion. In this model, for a $\neq$ 0 and small enough g, the elementary fermion is certainly not massless. Also, for small enough g the elementary ϕ and ψ will certainly not form massless bound states.

So for small enough g, super-
symmetry must be unbroken, there
being no candidate Goldstone fer-
mion. However, one might have
believed that for large enough g
the fermion mass goes to zero and
supersymmetry is broken.   Such
behavior is very common in the
case of global symmetries, but
the fact that $\text{Tr}(-1)^F = 2$ shows
it cannot happen here — super-
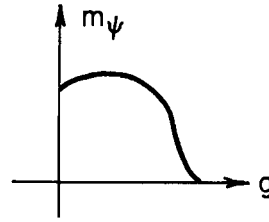symmetry is unbroken even for
strong coupling.

Figure 13

What cannot happen in supersym-
metric theories: the fermion
becomes a Goldstone fermion for
large enough g.

Also, the a = 0 model has a
massless fermion in perturbation
theory. Could not this particle become (at the non-perturbative
level, since it has been proved not to happen in perturbation theory)
a Goldstone fermion? The fact that $\text{Tr}(-1)^F = 2$ shows that this does
not occur.

Now a few simple comments:

(1) If supersymmetry is broken spontaneously at the tree level,
then at the tree level $n_B^{E=0} = n_F^{E=0} = 0$, there being no zero energy
states of any kind. So $\text{Tr}(-1)^F = 0$.

(2) If $\text{Tr}(-1)^F \neq 0$ supersymmetry is definitely unbroken. But
if $\text{Tr}(-1)^F = 0$ we do not know. We may have $n_B^{E=0} = n_F^{E=0} = 0$, and
supersymmetry broken, but we may equally well have $n_B^{E=0} = n_F^{E=0} \neq 0$,
and supersymmetry unbroken.

(3) As in the a $\neq$ 0 Wess-Zumino model, it is easy to calculate
$\text{Tr}(-1)^F$ in any model where supersymmetry is unbroken at the tree
level and all particles have mass. It is easy because, there being

no massless particles, only the "vacuum states" have $E = 0$ in per-
turbation theory. The vacuum states are all (spin zero) bosons so
$Tr(-1)^F$ is equal to the number of zeros of the classical potential
(there are some since supersymmetry is assumed unbroken at the tree
level). So in all theories with no massless particles at the tree
level, $Tr(-1)^F$ is positive, and supersymmetry is not spontaneously
broken.

Finally, let us discuss a case in which these methods really
yield interesting results — supersymmetric non-Abelian gauge theo-
ries. In the simplest such theory, the only fields are the gauge
field A_μ^a and its partner, the fermi field ψ_α^a, also in the adjoint
representation of the gauge group. The Lagrangian is

$$\mathcal{L} = \int d^4 x \ (- \frac{1}{4} \ (F_{\mu\nu}^a)^2 + \frac{1}{2} \ \bar{\psi}^a \ i \ \not{D} \ \psi^a) \tag{35}$$

The easiest way to show that this is a supersymmetric theory is to
show that the supersymmetry current

$$S_\mu = \sigma_{\alpha\beta} \ F^{\alpha\beta a} \ \gamma_\mu \ \psi^a \tag{36}$$

is conserved. This is readily demonstrated, with the aid of some
Dirac algebra and the use of fermi statistics.

This theory is tricky to deal with because of the zero momentum
modes of the massless particles.

The main problem is the gauge field. Only a finite number of
$\vec{p} = 0$ fermions can fit in the box, and we could count those states.
But the $\vec{p} = 0$ mode of the gauge field is a problem.

In infinite volume the $\vec{p} = 0$ mode of the gauge field can gauged
away. If $A_\mu = C_\mu$ (the C_μ being constants), then $A_\mu = \partial_\mu \epsilon$, with
$\epsilon = C_\mu x^\mu$. Here the constant can't be gauge away because the gauge
parameter ϵ isn't periodic.

More generally, the guage invariant

$$\text{Tr P exp i} \oint A_\mu \, d\chi^\mu \tag{37}$$

for a contour (figure (14)) that runs "around the box" is different at A = constant from its value at $A = 0$, proving that the zero momentum mode can't be gauge away.

It is possible to come to grips with this problem,[12] but in the case of the gauge group SU(N), there is a much simpler approach. One can choose boundary conditions that the zero momentum mode does not satisfy. Such boundary conditions are the "twisted boundary conditions" of 't Hooft.[13] In a special case that is general enough for our purposes, the twisted boundary conditions, in a box of length L, mean that

$$\begin{aligned}\phi(x,y,z) &= P\phi \ (x{+}L,y,z) \ P^{-1} \\ &= Q\phi \ (x,y{+}L,z) \ Q^{-1} \\ &= \ \phi \ (x,y,z{+}L) \end{aligned} \tag{38}$$

where ϕ may be A_μ or ψ. For $P = Q = 1$ this would describe conventional periodic boundary conditions. 'T Hooft instead requires P and Q to be constant matrices that obey

$$PQ = QP \exp 2\pi i/N \tag{39}$$

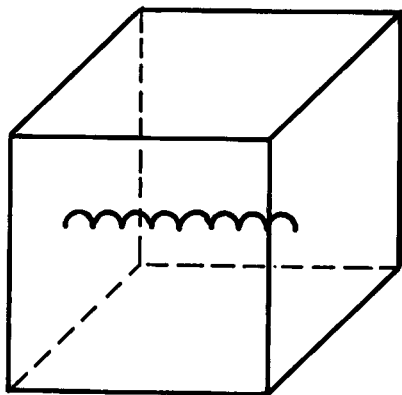Explicit matrices P and Q that obey (39) are easily found. For instance, we may take

Figure 14

$$P = \alpha \begin{pmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ & & 0 & 1 & \\ & & & \ddots & \ddots \\ 1 & & & & 1 \\ & & & & 0 \end{pmatrix} \qquad Q = \beta \begin{pmatrix} 1 & & & & \\ & e^{i\delta} & & & \\ & & e^{2i\delta} & & \\ & & & \ddots & \\ & & & & e^{(N-1)i\delta} \end{pmatrix}$$

(40)

where $\delta = 2\pi/N$, and α and β are constants chosen to ensure det P = det Q = 1.

We are entitled to adopt the twisted boundary conditions be-cause, in the large volume limit, the physics is expected to be independent of the boundary conditions. If we can show that, with twisted boundary conditions, the ground state energy vanishes for any value of the volume, then, in the large volume limit, the energy vanishes, and supersymmetry is unbroken, for any choice of the boundary conditions.

Accoarding to 't Hooft, the theory formulated as in (38) des-cribes a world with "one unit of magnetic flux·in the z direction." The motivation for that terminology is not essential for our pur-poses. For our purposes the key point is that the twisted boundary conditions eliminate the $\vec{p} = 0$ mode. It just doesn't satisfy the boundary condition.

The zero momentum mode ϕ_o of a field ϕ that satisfies (38) would have to obey

$$\phi_o = P \phi_o P^{-1} = Q \phi_o Q^{-1} \qquad (41)$$

It is easy to see that for ϕ_o in the Lie algebra of SU(N), (41) requires $\phi_o = 0$. (To commute with Q, ϕ_o must be diagonal. But a traceless, diagonal matrix that commutes with P must vanish.) There-fore, the twisted boundary conditions remove the zero momentum mode.

With this accomplished, evaluating $\text{Tr}(-1)^F$ is as easy as in the Wess–Zumino model. Only the "vacuum states" have $E = 0$. Other states have energy at least equal to the lowest allowed momentum for modes of non-zero momentum in the box of length L.

One can expand around $A_\mu^a = 0$; this "vacuum" contributes one to $\text{Tr}(-1)^F$. There are other vacuum states that can be obtained from $A_\mu^a = 0$ by topologically non-trivial gauge transformations. In a world defined by a given value of the vacuum angle θ there are still N sectors of configurations of $F_{\mu\nu}^a = 0$ (they are created from $A_\mu^a = 0$ by the gauge transformations which, according to 't Hooft, measure the electric flux in the z direction). So altogether, for SU(N), $\text{Tr}(-1)^F = N$, and supersymmetry is not spontaneously broken.

Twisted boundary conditions may be introduced for any group which has a non-trivial center. However, for groups other than SU(N), the twisted boundary conditions are not very useful in calculating $\text{Tr}(-1)^F$. The problem is that, for other groups, the twisted boundary conditions do not eliminate the zero momentum mode because the condition analogous to (41) does not imply $\phi_0 = 0$.

However, it is possible to calculate $\text{Tr}(-1)^F$ with untwisted boundary conditions, by quantizing the zero momentum modes in a Born-Oppenheimer approximation. One finds[12] that for a simple non-Abelian Lie group of rank r, $\text{Tr}(-1)^F = r+1$. Thus, spontaneous super-symmetry breaking does not occur in these theories.

What happens if additional matter fields are added? If the additional fields are in a real representation of the gauge group, bare masses m_i are possible for all of the matter fields. In this case all states containing the new quanta have non-zero energy, equal to or greater than the smallest of the m_i. The new fields do not contribute to $\text{Tr}(-1)^F$. So supersymmetry is unbroken, just as if the new fields were not present.

This argument shows that the ground state energy is zero for any non-zero values of the m_i. Taking now the limit as $m_i \to 0$, the ground state energy must remain zero. Theonly assumption needed here is that the zero mass limit should exist. We conclude that supersymmetry is unbroken for a theory with massless charged matter fields, as long as they lie in a representation of the gauge group such that they could have had bare masses.

In the very interesting case of theories with massless fields in a complex representation of the gauge group — so that gauge invariant bare masses are impossible — I do not know how to calculate $\text{Tr}(-1)^F$. For reasons explained elsewhere,[12] there are difficulties even in formulating the problem.

I should also mention that for a theory with gauge group $U(1)$, $\text{Tr}(-1)^F = 0$, as long as all charged fields have or could have had bare masses. It is nonetheless possible to prove by a variant of the concept of $\text{Tr}(-1)^F$ that dynamical supersymmetry breaking does not occur in any $U(1)$ gauge theory in which the Hamiltonian commutes with charge conjugation invariance (or in which charge conjugation invariance is broken only by Yukawa couplings, and not by the gauge couplings or the Fayet-Iliopoulos D term).

Likewise, in a theory with a simple gauge group that is spontaneously broken at the tree level to a subgroup such as $SU(3)$ x $SU(2)$ x $U(1)$ that contains a $U(1)$ factor, $\text{Tr}(-1)^F = 0$.

Clearly, these results impose significant restrictions on the possibility of supersymmetry breaking by strong gauge forces. However, loopholes remain, such as the question of matter fields in a complex representation of the gauge group. It also remains for the future to determine whether these methods can shed light on the apparent difficulties in 3+1 dimensions in supersymmetry breaking by instantons.

V. ANOTHER APPROACH TO MASS HIERARCHIES

In this last lecture I would like to describe a different approach to the gauge hierarchy problem.

It is usually assumed, in thinking about the hierarchy problem, that the "large" masses (the mass scale of grand unification and the Planck mass) are the fundamental ones. The problem then is to explain why the "small" masses (the masses of ordinary particle physics) are so small. One may seek to explain this by finding a non-perturbative mechanism that generates tiny masses. It was this possibility that motivated the discussion of dynamical supersymmetry breaking in the last few lectures.

Another possibility is to assume that the small masses are the fundamental ones. One must then explain why the large masses are so large.

This approach was actually followed by Dirac[7] in his original approach to the problem of the large numbers. Dirac assumed that the electron and proton masses were the basic ones. To account for the fact that the Planck mass is enormously larger, Dirac postulated that the Planck mass is not constant in time but increases in time as the universe grows older. The enormous present value of the Planck mass was thus attributed by Dirac to the fact that (in elementary particle units) the present universe is extremely old. Unfortunately, this beautiful idea has some serious empirical difficulties, which I will mention later.

Today I will be describing a class of supersymmetric theories[14] which spontaneously generate a mass scale vastly larger than the mass scale postulated in the Lagrangian. In these theories it is conceivable that the fundamental mass scale of nature could be comparable to the mass scale of the Weinberg-Salam model. All larger masses would be dynamically generated. As we will see, these

theories have something in common with Dirac's approach, and we can,
but need not, retrieve Dirac's idea that the "constants" of nature
are slowly varying functions of time.

In our previous discussions we emphasized dynamical (non-per-
turbative) symmetry breaking. Now, however, we will consider theo-
ries in which supersymmetry is spontaneously broken at the tree
level (classically). There will be no mystery about supersymmetry
breaking; we will simply choose a scalar potential whose minimum is
not invariant under supersymmetry. The problem will be to extract
the consequences of supersymmetry breaking.

To understand the idea, let us consider one of the simplest
models with supersymmetry spontaneously broken at the tree level —
the O'Raifeartaigh model. As we discussed before, in this model
there are three complex fields A, X, and Y of spin zero. The super-
space potential is $W(A,X,Y) = \lambda X(A^2-M^2) + g\, Y\, A$, and the ordinary
potential energy is $V(A,X,Y) = |\partial W/\partial A|^2 + |\partial W/\partial X|^2 + |\partial W/\partial Y|^2$ or

$$V(A,X,Y) = \lambda^2 \,|A^2-M^2| + g^2\,|A|^2 + |2\lambda\, AX + gY|^2 \quad . \tag{42}$$

This potential is strictly positive, so supersymmetry is spontaneously
broken.

We determine A by minimizing the first two terms. For $(g/\lambda M)^2 > 2$
one finds $A = 0$; for $(g/\lambda M)^2 < 2$ one finds $A = \sqrt{M^2 - g^2/4\lambda^2}$. Supersym-
metry is spontaneously broken in either case.

Although minimization of the potential determines A, it does not
determine X and Y uniquely. Since the only term in the potential that
depends on X and Y is $|2\lambda\, A\, X + g\, Y|^2$, we can choose any X, as long as

$$Y = -2\,\lambda\, A\, X\, /\, g \tag{43}$$

Classically, X may be arbitrarily large; the energy is minimized as
long as (43) is satisfied.

This degeneracy is not a property of this one model alone. Such degeneracies occur in many models in which supersymmetry is spontaneously broken at the tree level.

Although the energy is independent of X, the particle masses are not. A glance at (42) shows that for X >> M the mass of the A particle is approximately $2\lambda X$. If X >> M, the theory has at least two different energy scales. The small scale is the mass M in the Lagrangian. The large scale is the vacuum expectation value of X, which is undetermined by the classical Lagrangian. In an extended version of this model, we may try to interpret these as the scales of weak interactions and of grand unification, respectively.

However, we should not arbitrarily assume X >> M. We should determine X by calculating quantum corrections that remove the classical degeneracy.

Let us recall how this goes.[15] For bosons, the zero point energy per unit volume is positive and has the form

$$\frac{1}{2} \hbar \sum \omega_i = \frac{1}{2} \hbar \sum_i \int \frac{d^3k}{(2\pi)^3} \sqrt{k^2 + M_i^2} \, (X) \tag{44}$$

The sum on the right-hand side of (44) is a sum over the spin states of the various bosons; the M_i are the masses of the bosons, which depend on X. For fermions we have instead the negative energy from filling the Dirac sea. It is

$$- \frac{1}{2} \hbar \sum_i \int \frac{d^3k}{(2\pi)^3} \sqrt{k^2 + M_i^2} \, (X) \tag{45}$$

where now the sum runs over the fermion spin states.

If supersymmetry is unbroken at the tree level, the bosons and fermions have the same masses. Then (44) and (45) cancel. This is an illustration of the "non-renormalization" theorems: the ground state energy is zero to all finite orders if it is zero classically. We are, however, interested in cases in which supersymmetry is

spontaneously broken at the tree level; the integrals do not cancel.

Both integrals (44) and (45) are quartically divergent. The quartic divergence cancels because the bosons and fermions have the same number of spin states. The quadratic divergence cancels because of a sum rule of supersymmetric theories. Finally, the logarithmic divergence is removed by the renormalization counterterms that make the Green's functions finite.

After renormalization, and after doing the integrals, the final, finite expression for the $O(\hbar)$ corrections to the energy is

$$\Delta V(X) = \hbar \sum_i \frac{(-1)^F}{64\pi^2} M_i(x)^4 \ln \frac{M_i^2(x)}{\mu^2} \tag{46}$$

where $(-1)^F$ is plus one for bosons, and minus one for fermions, and where μ is a renormalization mass.

Equation (46) shows that for large X, the main contribution to ΔV comes from the particles that become heavy as $X \to \infty$, namely A and its supersymmetric partner ψ_A. It is not difficult to evaluate their contribution. Assuming for simplicity that $g/\lambda M \gg 1$, the potential to this order, including the lowest order result and the one loop correction, is

$$V(X) = \lambda^2 M^4 (1 + \frac{\lambda^2}{8\pi^2} \ln |X|^2/\mu^2) + O(1/|X|^2) \tag{47}$$

This result was first derived by Huq.[16]

The logarithmic correction in (47) is characteristic of renormalizable theories. Its positive coefficient means that X cannot become large — the energy increases with $|X|$. This model develops no big mass hierarchy from a spontaneous large vacuum expectation value.

The logarithmic equation (47) always arises, and always has a positive coefficient, in theories of particles of spins 0 and $\frac{1}{2}$ only.

This fact has a simple renormalization group interpretation. The logarithm in equation (47) can be understood as a replacement of the bare coupling λ^2 with an effective coupling $\bar{\lambda}^2(X) = \lambda^2 \left(1 + \frac{1}{8\pi^2} \ln |X|^2/M^2\right)$. The coefficient of the logarithm is positive because theories of spin 0 and spin $\frac{1}{2}$ fields only are not asymptotically free. The effective coupling $\bar{\lambda}(X)$ increases with X.

As this reasoning suggests, a different result can be obtained by enlarging the O'Raifeartaigh model to include non-Abelian gauge fields. I first wish to discuss this in a qualitative way. In a gauge invariant generalization of the O'Raifeartaigh model, the one loop corrections have the form

$$V(X) = a\lambda^2 M^4 (1 + (b\lambda^2 - ce^2) \ln |X|^2/M^2) \qquad (48)$$

where λ is a scalar coupling, e is the gauge coupling, and a, b, and c are positive constants, of order one.

We see that if $b\lambda^2 - ce^2 < 0$, a runaway behavior occurs. It is favorable for X to become large since the potential is a decreasing function of X (figure (15)). In fact, (48) seems to show that the potential V becomes negative as $X \to \infty$. This is impossible; in supersymmetric theories $V(X) \geq 0$ for all X. The fact that (48) becomes negative for very large X just means that, as one might expect, perturbation theory breaks down when $e^2 \ln |X|^2/M^2 \sim 1$.
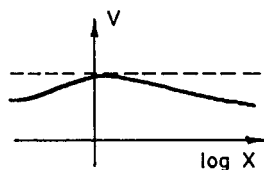

Figure 15
The classical potential (dotted line) and effective potential (solid line). The horizontal scale is logarithmic.

What happens for very large $|X|$ when perturbation theory breaks down? One possibility is that for very large $|X|$, the potential eventually ceases to decrease and begins increasing (figure (16a)). The potential would thus have a stable minimum at $e^2 \ln |X|/M \sim 1$, or equivalently at $X \sim M \exp +\frac{1}{\alpha}$. The theory would have at least two vastly different mass scales, M and X. As mentioned earlier, one might try to interpret these as the mass scales of weak inter-actions and of grand unification, respectively.

Another possibility is that the potential might decrease in-definitely as X becomes larger. Since V is bounded below by zero, it would have to approach a limit for large X (figure (16b)); the limit could be positive or zero. As we will discuss later, this cor-responds to a theory in which, as envisaged by Dirac, the constants of nature are slowly changing functions of time. In most of this talk, we will assume that the potential has a stable minimum at large X.
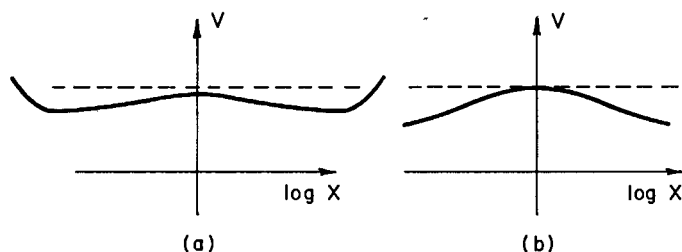

Figure 16

The potential may develop a stable minimum for very large X or may decrease indefinitely ((a) and (b) respectively). The horizontal scale is again logarithmic.

More complicated possibilities can also be imagined. Conceivably, in some theories the potential has an oscillatory dependence on X for large X.

It is possible to use the renormalization group to improve upon the one loop approximation of equation (48). In this way one can, to a large extent, decide between the various possibilities just described. One finds that the behavior of figure (16a) and also that of figure (16b) occur in a significant class of examples.[17]

I will now describe a simple model[14] which exhibits such runaway behavior. Let us consider an SU(5) theory with two complex fields A^i_j and Y^i_j in the adjoint representation of SU(5) and one singlet X. The superspace potential we choose to be

$$W(A,X,Y) = \lambda \ \text{Tr} \ A^2Y + gX \ (\text{Tr} \ A^2 - M^2) \tag{49}$$

This is the most general choice of W compatible with certain global symmetries, analogous to those of O'Raifeartaigh model.

Supersymmetry is spontaneously broken because the equations $\partial W/\partial X = 0$ and $\partial W/\partial Y^i_j = 0$ are inconsistent. As in the O'Raifeartaigh model, minimization of the potential uniquely determines A. One finds

$$A = \frac{g \ M}{\sqrt{\lambda^2 + 30g^2}} \begin{pmatrix} 2 & & & & \\ & 2 & & & \\ & & 2 & & \\ & & & -3 & \\ & & & & -3 \end{pmatrix} \tag{50}$$

However, X is once again undetermined at the tree level. One may choose any X as long as

$$Y = \frac{g}{\lambda} X \begin{pmatrix} 2 & & & & \\ & 2 & & & \\ & & 2 & & \\ & & & -3 & \\ & & & & -3 \end{pmatrix} \tag{51}$$

If X becomes large, the large expectation value of Y will strongly break SU(5) down to SU(3) x SU(2) x U(1).

To determine X, we perform a one loop calculation. This calculation is again dominated by the particles whose masses are, for large X, proportional to X. With the one loop correction included, we find

$$V(X) = V_o \left(1 + \frac{g^2}{g^2 + \lambda^2/30} \frac{29\lambda^2 - 50e^2}{80\pi^2} \ln |X|^2/M^2\right) + 0(1/|X|^2) \tag{52}$$

where V_o is the lowest order potential. We see that X will become large if $29\lambda^2 - 50e^2 < 0$.

This theory has two mass scales. The mass scale of SU(5) breaking is of order X because of the large vacuum expectation value of Y. However, supersymmetry is broken only at a mass of order M. In fact, only <A> violates supersymmetry and the ground state energy is of order M^4. Perhaps the large ratio X/M is related to the "big numbers" of physics.

Although we have obtained a hierarchy of symmetry breaking at vastly different mass scales, we have not yet addressed the original hierarchy problem. We wish SU(2) x U(1) symmetry breaking to occur as part of the low energy symmetry breaking.

To allow for this, we must include additional fields. A simple approach is to add an extra singlet Z and fields C^i and D_j in the 5 and $\bar{5}$ representation of SU(5). To the superspace potential we add a new term

$$\Delta W = \tilde{g} \, Z \, (C^i \, D_i - \tilde{M}^2) + \tilde{\lambda} A^i{}_j \, C^j \, D_i \tag{53}$$

We assume that $\tilde{M}$ is of order M. Minimization of the potential forces C and D to obtain vacuum expectation values, spontaneously breaking SU(2) x U(1) down to U(1).

In this model, SU(5) is broken to SU(3) x SU(2) x U(1) at the energy scale X, but supersymmetry is broken, and SU(2) x U(1) is broken down to U(1), at the far lower scale M. We thus obtain the desired pattern of symmetry breaking. However, there are some severe difficulties.

Perhaps the most severe difficulty is related to the fact that the color triplet components of C^i and D_j are relatively "light", with masses of order M. After introducing quarks and leptons in the model, we must couple the SU(2) x U(1) doublet components of C and D to the quarks and leptons, in order to account for their bare masses. However, SU(5) then requires that we also couple the color triplet components of C and D to the quarks and leptons. Unfortunately, these color triplets then mediate a very rapid decay of the proton, the lifetime being a small fraction of a second.

This particular problem can be avoided if we replace the coupling $A^i{}_j \, C^j \, D_i$ in equation (53) with a term $Y^i{}_j \, C^j \, D_i$. In this case, for reasons explained in reference (14), the color triplet components of C and D automatically become superheavy without any special adjustment of parameters. We can now couple C and D to quarks and leptons, so as to account for the quark and lepton masses, without inducing a rapid proton decay.

Unfortunately,[14] the new coupling also modifies the pattern of symmetry breaking. The vacuum expectation value of Y is changed. We now find that SU(5) is strongly broken down to SU(3) x U(1) x U(1) at the large mass scale X; SU(3) x U(1) x U(1) is then broken to SU(3) x U(1) at energies of order M. In this model the neutral

currents have roughly the strength that they actually have in nature, but the charged currents are greatly suppressed.

It is difficult to simultaneously obtain the right pattern of symmetry breaking and a suitably long proton lifetime. Recently, Georgi has described[18] an attempt to do this, based on a model with fields Y^{ij}_{kl} in the 75 dimensional representation of SU(5).

There is another crucial fact which must be borne in mind in thinking about the phenomenology of these models. It turns out that in the models as I have written them, the Goldstone fermion decouples in the limit as X becomes very large.* This means that although supersymmetry is broken, the bosons and fermions are degenerate to within terms of much less than M. Clearly, this is a very unacceptable state of affairs if M is interpreted as the mass scale of weak interactions.

It is possible to avoid this decoupling of the Goldstone fermion, and obtain models with bose-fermi splittings that really are of order M, by adding additional fields. However, the resulting models seem somewhat contrived.

It is equally interesting to explore the physical content of our models such as (49) more carefully, bearing in mind the decoupling of the Goldstone fermion. A more careful study of the model based on equation (49) shows that, at the tree level, many particles have masses much less than M. The SU(3) x SU(2) x U(1) gauge mesons and their fermionic partners are massless at the tree level. All components of Y except the component with a vacuum expectation value have masses of order M^2/X. (As X becomes large, A becomes heavy and decouples; as A decouples, Y becomes massless.) One would naively expect that these light particles would receive masses of order αM

---

*I wish to thank T. Banks for convincing me of this.

or perhaps $\alpha^2 M$ from one or two loop diagrams. Because of the de-
coupling of the Goldstone fermion and the non-renormalization theo-
rems of unbroken supersymmetry, this is not true. The particles
whose masses are of order M^2/X or less at the tree level receive mass
corrections at most of order $\alpha M^2/X$ from the loop diagrams.

These theories thus have <u>three</u> mass scales, X, M, and M^2/X, if
we do not attempt to tamper with the decoupling of the Goldstone
fermion that occurs for large X in the simplest models. The fascinat-
ing possibility exists that it is the <u>smallest</u> of these scales, the
scale M^2/X, which is the mass scale of particle physics as we know it.
One reason that this possibility is interesting is that at energies
of order M^2/X, supersymmetry appears to be <u>explicitly</u> (but softly)
broken. To write an effective Lagrangian describing this system in
which the supersymmetry is <u>spontaneously</u> broken, one must re-introduce
some of the degrees of freedom with masses of order M.

Of course, soft, explicit breaking of supersymmetry greatly
weakens the constraints associated with supersymmetry and thereby
alleviates the problems in finding a realistic model. The question
of finding a realistic model in which the mass scale of weak inter-
actions is the <u>smallest</u> scale, M^2/X, is therefore very much worthy
of attention. I will return to this matter elsewhere.[19] One must,
of course, decide on an appropriate identification of X and M.
Plausibly, in this approach, X might be of order the Planck mass;
M would then have to be about 10^{11} GeV.

Let us now conclude by discussing a few general questions.

In a theory in which the fundamental mass scale of nature is
relatively "small" and the mass of grand unification is a derived
quantity, we do not wish to assume that the Planck mass M_{Pl} is a
fundamental constant of nature. If it is, we are left with the large
and unexplained ratio M_{Pl}/M. Rather, we should assume that the large
value of the Planck mass is spontaneously generated along with the

large value of the mass scale of grand unification.

This possibility is not as bizarre as it might sound. The usual kinetic energy of the gravitational field is $M_o^2 R$, M_o being the bare Planck mass and R the Ricci scalar. Another possible coupling is the dimension four Brans-Dicke coupling $|X|^2 R$, X being the scalar field that will eventually obtain a large vacuum expectation value. (Such couplings actually arise almost inevitably when theories of global supersymmetry are coupled to gravity.) We thus imagine the gravitational action to be

$$\mathcal{L}_{Grav} = M_o^2 R + \lambda^2 |X|^2 R \qquad (54)$$

where λ is a constant of order one.

Clearly, the observed Planck mass would be $M_{P1}^2 = M_o^2 + \lambda^2 |X|^2$. For large X, the Planck mass is simply λX, unrelated to the original constant M_o.

Let us now return to the question of how the effective potential behaves when X becomes very large. As we discussed earlier, while it is possible that a stable minimum develops for large X, it is equally possible that the potential decreases indefinitely with increasing X, as in figure (16b).

This possibility corresponds to a world without a stable vacuum. However, it is possible to expand about a "cosmological solution" in which X is regarded as a function of time. For large times, X would change very slowly, because the potential (being bounded below) is nearly independent of X for large X.

For instance, if $V(X) \sim |X|^{-p}$ for some p, then, by solving the cosmological equations (with the gravitational action taking the form (54)), it is easy to show that for large t, $X \sim t^{2/(2+p)}$.

Since the masses and couplings of the elementary particles depend on X, this corresponds to Dirac's idea that the "constants" of nature are slowly evolving functions of time. For instance, Newton's constant $G = 1/M_{pl}2$ scales with X as $1/X^2$, because $M_{pl} \sim X$. So $G \sim t^{-4/(2+p)}$, t being the age of the universe. This may be compared to the behavior $G \sim 1/t$ which Dirac advocated in order to account for the value of G in today's universe.

Present day experiments aimed at measuring $\dot{G}/G$ are not yet sensitive enough to confirm or refute Dirac's suggestion. However, there are two good reasons to doubt that G is changing at this rate. First, if G scales like a power of t, then G was vastly larger in the early universe. This would ruin the successful calculations of nucleosynthesis in the big bang. Second, the measurements of $\dot{\alpha}/\alpha$ (α is the ordinary fine structure contant, $\alpha = e^2/\hbar c$), are far more accurate than the measurements of $\dot{G}/G$. (A recent review has been given by Dyson.[20]) A change in X would produce a change in α through renormalization effects,[14] at a rate which is ruled out by experiment if X is changing as rapidly as Dirac's approach to the hierarchy problem would require.

It is therefore most prudent to assume that the potential energy V(X) has a stable minimum at some large but finite X.

All of this, however, requires a note of caution. In particle physics discussions we usually ignore gravity. We argue that the Planck mass is larger than other masses of interest; but the real reason for ignoring gravity is, of course, that we simply do not know how to take gravitational effects into account. In the approach I have been discussing, it is not at all clear that it is valid to neglect gravity. As I have explained, in this approach the "bare" Planck mass should probably be of order M, the fundamental mass scale of the theory. The observed Planck mass is then larger in today's universe because of a sort of renormalization effect — the large,

spontaneously generated expectation value of X. But the mass scale of grand unification is also determined by X, in this approach. So when we discuss grand unification we are really working at energies above the underlying energy scale of gravita tion. It is not at all clear that it is sensible, under these conditions, to assume that gravity is a small effect. However, for the time being, it is the best that we can do.

Finally, I would like to draw attention to a possible consequence of this framework.

As I have mentioned, in the simplest models of this type, there are many particles with masses of order M^2/X — so many that one may wish to interpret M^2/X as the mass scale of particle physics as we know it. I believe that this is the most promising approach.

However, there is another approach, which in fact was the first possibility raised at the beginning of this lecture (and suggested in reference (14)). In variants of these models, arranged so that the Goldstone fermion does not decouple as X becomes large, almost all of the particles have masses of order M, perhaps multiplied by a small power of the fine structure constant α. In this case, it is plausible to assume that the fundamental mass scale M of the theory is not too much larger than the mass scale of the Weinberg-Salam theory.

But there is one particle whose mass is always in the range M^2/X. This is the X particle itself. Indeed, the effective potential depends on X only very weakly. It has the general form

$$V(X) = M^4 \, F \, (\alpha \ln |X|^2/M^2) \tag{55}$$

where F is a dimensionless function of the dimensionless variable $\alpha \ln |X|^2/M^2$ which arises in loop diagrams. To one loop order, F is given (in a particular model) in equation (52).

Taking the second derivative of (55) with respect to X, we see that the mass of the X particle is of order $\alpha M^2/X$. If M^2/X is the mass scale of weak interactions, this is not remarkably small, and the very weak couplings of the X particle (discussed below) would render it unobservable. However, if the fundamental mass scale M is identified as the scale of weak interactions, then the X particle is <u>extremely</u> light, so light that its Compton wave-length λ_X is macroscopic. For 10^{15} GeV $< X < 10^{19}$ GeV we would expect 10^2 cm $> \lambda_X$ $> 10^{-3}$ cm.

The X particle is a scalar and can have coherent couplings to matter. Actually, at the tree level, X does not couple to ordinary, light particles. X has such a large expectation value that anything that couples to X at the tree level is not light! However, coherent couplings of X to matter are generated by loop diagrams. The most important effect is a coupling of X to gluons, induced by a diagram containing heavy, color-bearing particles (figure (17)). This diagram induces a coupling of X to $\frac{\alpha_s}{<X>}$ Tr $F_{\mu\nu}$ $F^{\mu\nu}$, where $F^a_{\mu\nu}$ is the gluon field strength. As α_s Tr $F_{\mu\nu}$ $F^{\mu\nu}$ is essentially the trace of the energy momentum tensor in QCD, whose matrix element in a nucleon state is the nucleon mass m_N, we see that the coupling of X to a nucleon is of order $m_N/<X>$.
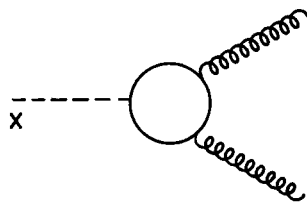

Figure 17

This may be compared to the gravitational coupling, which involves m_N/M_{Pl}. We see that if $<X> \sim M_{Pl}$, the new force is comparable to gravitation, at distances less than λ_X. If $<X> \sim 10^{15}$

GeV the force due to X exchange may be 10^8 times stronger than gravity at distances less than $\lambda_X \sim 10^{-3}$ cm!

New, coherent, short-range forces have been suggested in the recent past, on several different grounds.[21] Laboratory measurements of gravity at short distances place interesting bounds[22] but certainly do not exclude the possibility of such a new force. For instance, neither these experiments nor measurements of the Casimir effect[23] rule out the possible existence of a new force 10^8 times stronger than gravity at a distance of 10^{-3} cm.

These comments really refer to the magnitude of the X field. It may be shown that the phase of the X field behaves rather like a Peccei-Quinn-Weinberg-Wilczek axion, even if there is no underlying U(1) axial symmetry.

A previous version of these lectures was given at the International Center for Theoretical Physics July, 1981. Sections II and IV of these notes are adapted from notes circulated there. I would like to acknowledge the hospitality of the ICTP as well as that of the school on sub-nuclear physics at Erice.

REFERENCES

(1) Y.A. Gol'fand and E.P. Likhtam, JETP Lett. 13 (1971) 323;
    D.V. Volkov and V.P. Akulov, Phys. Lett. 46B (1973) 109;
    J. Wess and B. Zumino, Nucl. Phys. B70 (1974) 39.

(2) P. Fayet and S. Ferrara, Phys. Rep. 32C (1977) 249; A. Salam
    and J. Strathdee, Phys. Rev. D11 (1975) 1521, Fortschritte der
    Physik 26 (1978) 57; J. Wess and J. Bagger, preprint, Institute
    for Advanced Study (1981); P. van Nieuwenhuysen, Phys. Rept.

(3) S. Coleman and J. Mandula, Phys. Rev. 159 (1967) 1251.

(4) R. Haag, M. Sohnius, and J. Lopuszanski, Nucl. Phys. B88 (1975)
    257.

(5) B. Zumino, Coral Gables Lectures (January, 1977), Lectures at
     the 1976 Scottish Universities Summer School (Edinburgh Univer-
     sity Press, Edinburgh, Scotland, 1976); E. Cremmer, B. Julia,
     J. Scherk, P. van Nieuwenhuysen, S. Ferrara, and L. Girardello,
     Phys. Lett. 79B (1978) 231.

(6) E. Witten, Nucl. Phys. B185 (1978) 513.

(7) P.A.M. Dirac, Proc. Roy. Soc. (London) A165 (1938) 199.

(8) S. Weinberg, in Proc. of the XVII Int. Conf. on High Energy
     Physics, 1974, ed. J.R. Smith (Rutherford Laboratory, Chilton,
     Didcot, Berkshire, England, 1974).

(9) L. O'Raifeartaigh, Nucl. Phys. B96 (1975) 331.

(10) M.T. Grisaru, W. Siegel, and M. Rocek, Nucl. Phys. B159 (1979)
     429.

(11) M. Dine, W. Fischler, and M. Srednicki, Nucl. Phys. B189 (1981)
     575; S. Dimopoulos and S. Raby, Nucl. Phys. B192 (1981) 353.

(12) E. Witten, Princeton preprint, 1982.

(13) G. 't Hooft, Nucl. Phys. B153 (1979) 141.

(14) E. Witten, Phys. Lett. 105B (1981) 267.

(15) S. Coleman and E. Weinberg, Phys. Rev. D7 (1973) 788.

(16) M. Huq, Phys. Rev. D14 (1976) 3548.

(17) H. Yamagishi, Princeton preprint, 1982.

(18) H. Georgi, Harvard preprint, 1982.

(19) E. Witten, to appear.

(20) F. Dyson, in Current Trends in the Theory of Fields, ed. J.E.
     Lannutti and P.K. Williams (AIP, 1978).

(21) J.E. Kim, Phys. Rev. Lett. 43 (1979) 103; G. Feinberg and
     J. Sucher, Phys. Rev. D20 (1979) 1717; J. Scherk, Phys. Lett.
     88B (1979) 265.

(22) R. Spero, J.K. Hoskins, R. Newman, J. Pellman, and J. Schultz,
     Phys. Rev. Lett. 44 (1980) 1645; H.-T. Yu, W.-T. Ni, C.-C. Hu,
     F.-H. Liu, C.-H. Yang, and W.-N. Liu, Phys. Rev. D20 (1979) 1813.

(23) J.N. Israelachveli and K. Tabor, in Progress and Surface and
     Membrane Science, Vol. 7, ed. J.F. Danielli, M.D. Rosenberg,
     and D.A. Cadenhead (Academic Press, 1973).