Beyond the Standard Model: Supersymmetry and String Theory

Contents

1 The Standard Model and Its Discontents 1
  1.1 The Standard Model ................................. 1
  1.2 The Standard Model as an Effective Field Theory ............ 2
    1.2.1 Integrating out the $W$ and $Z$ Bosons ............... 3
    1.2.2 What Might the Standard Model Come From? ............. 4
  1.3 Experimental Status of Lepton and Baryon Conservation ....... 5
    1.3.1 Dimension five: Lepton Number Violation and Neutrino Mass 5
    1.3.2 Other Dimension Five Operators ........................ 6
    1.3.3 Dimension Six Operators: Proton Decay ................ 6
  1.4 A puzzle at the renormalizable level ........................ 7
  1.5 The Hierarchy Problem .................................. 7
  1.6 Dark Matter and Dark Energy ................................ 8
  1.7 Summary: Limitations of the Standard Model .................. 9

2 Of Anomalies, Strong CP and Axions 9
  2.1 The Chiral Anomaly ..................................... 9
    2.1.1 Return to QCD ................................... 12
    2.1.2 A Two Dimensional Detour ......................... 13
    2.1.3 The Anomaly In Two Dimensions ...................... 13
    2.1.4 The $\mathbb{CP}^N$ Model: An Asymptotically Free Theory 14
    2.1.5 The Large $N$ Limit .................................. 15
    2.1.6 The Role of Instantons ............................... 16
  2.2 Real QCD ................................................ 19
    2.2.1 The Theory and its Symmetries ....................... 19
    2.2.2 Instantons .......................................... 21
    2.2.3 Real QCD and the U(1) Problem ...................... 24
  2.3 The Strong CP Problem ................................... 25
    2.3.1 $\theta$-dependence of the Vacuum Energy ............. 25
    2.3.2 The Neutron Electric Dipole Moment .................. 26
  2.4 Possible Solutions ....................................... 27
    2.4.1 The Axion ......................................... 27

3 An Introduction to Supersymmetry 29

4 An Overview of Supersymmetry 30
  4.1 The Supersymmetry Algebra and its Representations .......... 30
  4.2 $N=1$ Lagrangians ...................................... 32
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.3</td>
<td>N=2 Theories: Exact Moduli Spaces</td>
<td>37</td>
</tr>
<tr>
<td>4.4</td>
<td>A Still Simpler Theory: N=4 Yang Mills</td>
<td>39</td>
</tr>
<tr>
<td>4.5</td>
<td>Aside: Choosing a Vacuum</td>
<td>41</td>
</tr>
<tr>
<td>4.6</td>
<td>Supersymmetry and the Hierarchy Problem</td>
<td>42</td>
</tr>
<tr>
<td>4.7</td>
<td>Supersymmetry in Superspace</td>
<td>43</td>
</tr>
<tr>
<td>4.8</td>
<td>Component Lagrangians</td>
<td>43</td>
</tr>
<tr>
<td>4.9</td>
<td>Soft Supersymmetry Breaking</td>
<td>43</td>
</tr>
<tr>
<td>4.10</td>
<td>Cancellation of Quadratic Divergences</td>
<td>43</td>
</tr>
<tr>
<td>4.11</td>
<td>More on the MSSM</td>
<td>46</td>
</tr>
<tr>
<td>4.12</td>
<td>LSP as the Dark Matter</td>
<td>47</td>
</tr>
<tr>
<td>4.13</td>
<td>Local Supersymmetry</td>
<td>48</td>
</tr>
<tr>
<td>5.1</td>
<td>Spontaneous Supersymmetry Breaking</td>
<td>51</td>
</tr>
<tr>
<td>5.1.1</td>
<td>F and D Type Breaking</td>
<td>51</td>
</tr>
<tr>
<td>6</td>
<td>Non-Renormalization Theorems</td>
<td>51</td>
</tr>
<tr>
<td>7.1</td>
<td>Supersymmetric QCD and Its Symmetries</td>
<td>51</td>
</tr>
<tr>
<td>7.2</td>
<td>$N_f &lt; N_c$: The Superpotential</td>
<td>51</td>
</tr>
<tr>
<td>7.3</td>
<td>$N_f = N_c - 1$: Instanton computation of the Superpotential</td>
<td>51</td>
</tr>
<tr>
<td>7.4</td>
<td>$N_f &lt; N_c$: Gaugino Condensation</td>
<td>51</td>
</tr>
<tr>
<td>7.5</td>
<td>Supersymmetry Breaking: The (3, 2) model</td>
<td>51</td>
</tr>
<tr>
<td>8</td>
<td>Seiberg Duality</td>
<td>51</td>
</tr>
<tr>
<td>9</td>
<td>Introduction</td>
<td>53</td>
</tr>
<tr>
<td>10</td>
<td>The Bosonic String</td>
<td>54</td>
</tr>
<tr>
<td>11</td>
<td>The Superstring</td>
<td>54</td>
</tr>
<tr>
<td>12</td>
<td>Type II Strings</td>
<td>54</td>
</tr>
<tr>
<td>13</td>
<td>Heterotic String</td>
<td>54</td>
</tr>
<tr>
<td>14</td>
<td>Compactification of String Theory</td>
<td>54</td>
</tr>
<tr>
<td>14.1</td>
<td>Tori</td>
<td>54</td>
</tr>
<tr>
<td>14.2</td>
<td>Orbifolds</td>
<td>54</td>
</tr>
<tr>
<td>14.3</td>
<td>Calabi-Yau compactifications</td>
<td>54</td>
</tr>
</tbody>
</table>
1 The Standard Model and Its Discontents

1.1 The Standard Model

The Standard Model is extremely successful. At any given time, you will find a few people excited about some slight discrepancy between theory and experiment, but this just serves to highlight the extraordinary level of agreement.

It is amazing that at a microscopic level, almost everything we know about nature is described by such a simple structure. To write down the model, one first specifies the gauge group: $SU(3) \times SU(2) \times U(1)$. Correspondingly, there are gauge bosons, $A_\mu^a$, $W_\mu^i$, and $B_\mu$. The description of the matter content is particularly compact if one describes the fermions in terms of left-handed fields (as explained in Wess and Bagger, and also in Peskin, in four dimensions, a general spin-1/2 field, whether “Majorana” or “Dirac” can be written in terms of left-moving fields). Below, we list the fields, using a notation where the first number in the parenthesis refers to the $SU(3)$ representation, the second the $SU(2)$ representation, and the subscript is the $U(1)$ charge (“hypercharge”):

Quark doublets: $Q_f = (3, 2)_{1/3}$
Antiquark singlets: $\bar{u}_f = (\bar{3}, 1)_{-4/3}$ $\bar{d}_f = (\bar{3}, 1)_{2/3}$
Lepton doublets: $L_f = (1, 2)_{-1}$
Lepton singlets: $\bar{e}_f = (1, 1)_{2}$

In addition to the fermions, (and the bosons which are required by the gauge symmetry), the simplest version of the Standard Model contains a single Higgs field, a scalar, $\phi$:

Higgs scalar: $\phi = (1, 2)_1$

Specifying the interactions is not much more complicated. First, there are gauge-invariant kinetic terms for all of the fields. For the quark doublets, for example, these are:

$$i Q \sigma^\mu D_\mu Q^*$$  \hspace{1cm} (1)

where

$$D_\mu Q = (\partial_\mu - i A_\mu - i W_\mu - i \frac{1}{3} B_\mu) Q$$  \hspace{1cm} (2)

and $A_\mu$ is a matrix-valued field, $A_\mu = A^{a}_\mu T^a$, as is $W_\mu = W^i_\mu T^i$. In terms of these fields, the field strengths can be written elegantly as:

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + i[A_\mu, A_\nu].$$  \hspace{1cm} (3)

and similarly for the $SU(2)$ and $U(1)$ fields. The kinetic terms for the gauge fields, in terms of these matrix-valued fields, take the form:

$$\mathcal{L}_A = -\frac{1}{2 g^2} \text{Tr} F_{\mu\nu}^2.$$  \hspace{1cm} (4)
Before including the Higgs fields, at the level of dimension four terms, i.e. of renormalizable terms, this is all we can write. Including the Higgs field, however, there are additional possibilities. Apart from the kinetic term for the Higgs field, we can write a potential:

\[ V(\phi) = -\mu^2|\phi|^2 + \lambda|\phi|^4. \] (5)

Again, this is completely gauge invariant.

There are also Yukawa couplings of the fermions to the Higgs field:

\[ \mathcal{L}_{yuk} = y^U_{f,f'} Q_f \bar{u}_{f'} + y^D_{f,f'} Q_f \bar{d}_{f'} \sigma_2 \phi^* + y^L_{f,f'} L_f \bar{e}_{f'} \phi. \] (6)

By redefining fields, we can simplify the matrices \( y^U, y^D \) and \( y^L \); it is customary to take \( y^U \) and \( y^L \) diagonal, and there is a conventional form for \( y^D \) (the Kobiyashi-Maskawa matrix) which we will describe in section ***.

If we require renormalizability, i.e. we require that all of the terms in the lagrangian be of dimension four or less, this is all we can write. It is remarkable that this simple structure incorporates over a century of investigation of elementary particles. Working through the consequences of this theory is very challenging, and is the subject of many excellent books and reviews. In the following sections, we will focus on a few features which might be relevant to what comes beyond.

**Exercise:** Compute the mass of the Higgs field as a function of \( \mu \) and \( \lambda \). Discuss production of Higgs particles (you do not need to do detailed calculations, but indicate the relevant Feynman graphs and make at least crude estimates of cross sections) in \( e^+ e^- \), \( \mu^+ \mu^- \) and \( p\bar{p} \) annihilation. Keep in mind that because some of the Yukawa couplings are extremely small, there may be processes generated by loop effects which are bigger than processes which arise at tree level.

### 1.2 The Standard Model as an Effective Field Theory

The standard model has some remarkable properties. The renormalizable terms respect a variety of symmetries, all of which are observed to hold to a high degree in nature:

- **Baryon number:**
  \[ Q \to e^{i\alpha Q} Q \quad \bar{u} \to e^{-i\alpha \bar{u}} \quad \bar{d} \to e^{-i\alpha \bar{d}} \] (7)

- **Three separate lepton numbers:**
  \[ L_f \to e^{-i\alpha_f} L_f \quad \bar{e}_f \to e^{i\alpha_f} \bar{e}_f. \] (8)

We did not impose these symmetries. They are simply consequences of gauge invariance, and the fact that there are only so many renormalizable terms we can write. These
symmetries are said to be “accidental,” since they don’t seem to result from any deep underlying principle.

This is already a triumph. As we will see when we consider possible extensions of the Standard Model, this did not have to be. But this success raises the question: why should we impose the requirement of renormalizability?

In the early days of quantum field theory, renormalizability was sometimes presented as a sacred principle. There was a view that field theories were fundamental, and should make sense in and of themselves. Much effort was devoted to understanding whether the theories existed in the limit that the cutoff was taken to infinity.

But there was an alternative model for understanding field theories, provided by Fermi’s original theory of the weak interactions. In this theory, the weak interactions are described by a lagrangian of the form:

$$L_{\text{weak}} = \frac{G_f}{\sqrt{2}} J^\mu J_\mu$$ (9)

Here the currents, $J^\mu$, are bilinears in the fermions; they include terms like $Q\sigma^\mu T^a Q^*$. This theory, like the Standard Model, was very successful. It took some time to actually determine the form of the currents, but for forty years, all experiments in weak interactions could be summarized in a lagrangian of this form. Only with the first indications of the $Z$ boson in $e^+e^-$ annihilation were deviations observed.

The four fermi theory is non-renormalizable. Taken seriously as a fundamental theory, it predicts violations of unitarity at TeV energy scales. But from the beginning, the theory was viewed as an effective field theory, valid only at low energies. When Fermi first proposed the theory, he assumed that the weak forces were due to exchange of particles – what we now know as the $W$ and $Z$ bosons.

1.2.1 Integrating out the $W$ and $Z$ Bosons

Within the Standard Model, we can derive the Fermi theory, and we can also understand the deviations. A traditional approach is to examine the Feynman diagram:

![Feynman Diagram](image)

This can be understood as a contribution to a scattering amplitude, but it is best understood here as a contribution to the effective action of the quarks and leptons. The currents of the Fermi theory are just the gauge currents which describe the coupling at
each vertex. The propagator, in the limit of very small momentum transfer, is just a constant. In coordinate space, this corresponds to a space-time $\delta$-function. The interaction is local. The effect is just to give the four Fermi lagrangian. One can consider effects of small finite momentum by expanding the propagator in powers of $q^2$. This will give four fermi operators with derivatives. These are suppressed by powers of $M_W$; these effects are very tiny at low energies. Still, in principle, they are there, and in fact the measurement of such terms provided the first hints of the existence of the $Z$ boson.

This can also be derived in the path integral approach. Here insert handout from 218.

The lesson is that, prior to 1987, one could view QED + QCD + the Fermi theory as a perfectly acceptable theory of the interactions. The theory would have to be viewed, however, as an effective theory, valid only up to an energy scale of order 100 GeV or so. Sufficiently precise experiments would require inclusion of operators of dimension higher than four. The natural scales for these operators would be the weak scale.

1.2.2 What Might the Standard Model Come From?

As successful as the Standard Model is, it is likely that, like the four-Fermi theory, it is the low energy limit of some underlying, more “fundamental” theory. In the second half of this book, our model for this theory will be String Theory. Consistent theories of strings, for reasons which are somewhat mysterious, are theories which describe general relativity and gauge interactions. Unlike field theory, String Theory is a finite theory. It does not require a cutoff for its definition. In principle, all physical questions have well-defined answers within the theory. If this is the correct picture for the origin of nature at extremely short distances, then the Standard Model is just its low energy limit. When we study string theory, we will understand in some detail how such a structure can emerge. For now, the main lesson we should take concerns the requirement of renormalizability: the Standard Model should be viewed as an effective theory, valid up to some energy scale, $\Lambda$. Renormalizability is not a constraint we impose upon the theory; rather, we should include operators of dimension five or higher with coefficients scaled by inverse powers of $\Lambda$. The question of the value of $\Lambda$ is an experimental one. From the success of the Standard Model, as we will see, we know that the cutoff is large. From string theory, we might imagine that $\Lambda \approx M_p = 1.2 \times 10^{18}$ GeV. But, as we will now describe, we have experimental evidence that there is new physics which we must include at scales well below $M_p$. We will also see that there are theoretical reasons to believe that there should be new physics at TeV energy scales.
1.3 Experimental Status of Lepton and Baryon Conservation

1.3.1 Dimension five: Lepton Number Violation and Neutrino Mass

To proceed systematically, we should write operators of dimension five, six, and so on. At the level of dimension five, we can write several terms which violate lepton number:

\[ \mathcal{L} = \frac{1}{\Lambda} \gamma_{f,f'} \phi \bar{f} f' \phi L_f L_{f'} + \text{c.c..} \]  

(10)

With non-zero \( \phi \), these terms give rise to neutrino masses. In the past few years, persuasive evidence has emerged that the neutrinos do have non-zero masses and mixings. This comes from the study of neutrinos coming from the sun (the “solar neutrinos”) and neutrinos produced in the upper atmosphere by cosmic rays (which produce pions which subsequently decay to muons and \( \nu_\mu \)'s, whose decays in turn produce electrons, \( \nu_\mu \)'s, and \( \nu_e \)'s. Accelerator and reactor experiments are now firming up this picture. The best current values for the neutrino masses and mixings are:

\[ m_{\nu} \ldots \]  

(11)

It is conceivable that these masses are not described by the lagrangian of eqn. 10. Instead, the masses might be “Dirac,” by which one means that there might be additional degrees of freedom, which by analogy to the \( \bar{e} \) fields we could label \( \bar{\nu} \), with very tiny Yukawa couplings to the normal neutrinos. This would truly represent a breakdown of the Standard Model: even at low energies, we would have been missing basic degrees of freedom. But this does not seem likely. If there are singlet neutrinos, \( N \), nothing would prevent them from gaining a “Majorana” mass:

\[ \mathcal{L}_{maj} = MNN. \]  

(12)

As for the leptons and quarks, there would also be a coupling of \( \nu \) to \( N \). There would now be a mass matrix for the neutrinos, involving both \( N \) and \( \nu \). For simplicity, consider the case of just one generation. Then this matrix would have the form:

\[ M_\nu = \begin{pmatrix} M & y \nu \\ y v & 0 \end{pmatrix}. \]  

(13)

Such a matrix has one large eigenvalue, of order \( M \), and one small one, of order \( \frac{y^2 v^2}{M} \). This provides a natural way to understand the smallness of the neutrino mass; it is referred to as the “seesaw mechanism.” Alternatively, we can think of integrating out the right-handed neutrino, and generating the operator of eqn. [10].

It seems more plausible that the observed neutrino mass is Majorana than Dirac, but this is a question that hopefully will be settled in time by experiment. If it is Majorana,
this suggests that there is another scale in physics, well below the Planck scale. For even if the new Yukawa couplings are of order one, the neutrino mass is of order

\[ m_\nu = 10^{-5} \text{eV}(M_p/\Lambda) \]  

(14)

If the Yukawa’s are small, as are many of the quark Yukawa couplings, the scale can be much smaller.

1.3.2 Other Dimension Five Operators

There is another class of dimension five operators which can appear in the effective lagrangian. These are magnetic and electric dipole operators. For example, for the muon magnetic dipole moment, one can have a term:

\[ \mathcal{L}_{g-2} = \frac{e}{\Lambda} F_{\mu\nu} \sigma^{\mu\nu} \mu + \text{c.c.}, \]  

(15)

where \( F_{\mu\nu} \) is the electromagnetic field (in terms of the fundamental \( SU(2) \) and \( U(1) \) fields, one can write similar gauge-invariant combinations which reduce to this at low energies). The muon magnetic moment is measured to extremely high precision, and its Standard Model contribution is calculated with comparable precision. As of this writing, there is a 2.6 \( \sigma \) discrepancy between the two. Whether this reflects new physics or not is uncertain.

Similar operators can also generate an electric dipole moment for the leptons or the quarks:

\[ \mathcal{L}_d = \frac{e}{\Lambda} \tilde{F}_{\mu\nu} \sigma^{\mu\nu} \mu + \text{c.c.} \]  

(16)

where

\[ \tilde{F}_{\mu\nu} = \frac{1}{2} \epsilon_{\mu\nu\rho\sigma} F^{\rho\sigma} \]  

(17)

These operators violate CP. There are strong experimental limits on the electric dipole moment of the neutron and of the electron which constrain the coefficient.

Operators where one of the muons is replaced by an electron can generate rare processes such as \( \mu \rightarrow e + \gamma \), on which there are quite stringent experimental bounds.

1.3.3 Dimension Six Operators: Proton Decay

Proceeding to dimension six, we can write numerous terms which violate baryon number, as well as additional lepton number violating interactions:

\[ \mathcal{L}_{6v} = \frac{1}{\Lambda^2} Q \sigma^{\mu} \bar{u}^s \sigma_\mu \bar{d}^s + \ldots \]  

(18)

\[ \sigma_\mu \]
This can lead to processes such as \( p \rightarrow \pi e \). Experiments deep underground set limits of order \( 10^{33} \) years on this process. Correspondingly, the scale \( \Lambda \) must be larger than \( 10^{15} \) GeV.

So viewing the Standard Model as an effective field theory, we see that there are many possible non-renormalizable operators which might appear, but most have scales which are tightly constrained by experiment. One might hope – or despair – that the Standard Model will provide a complete description of nature up to scales many orders of magnitude larger than we can hope to probe in experiment.

We now turn to some challenges for the Standard Model.

1.4 A puzzle at the renormalizable level

There is another puzzle, which we will touch upon now but return to and discuss in greater detail later. There are a set of operators of dimension four, which we don’t usually discuss when we introduce non-abelian gauge theories:

\[
\mathcal{L}_\theta = \theta F \tilde{F}
\]  

(19)

where

\[
\tilde{F}_{\mu \nu} = \frac{1}{2} \epsilon_{\mu \nu \rho \sigma} F^{\rho \sigma}.
\]  

(20)

We usually ignore these because classically they are inconsequential; they are total derivatives and do not modify the equations of motion. But as we will see they have real effects at the quantum level. Because they are CP violating, they turn out to be highly constrained. From the limits on the neutron electric dipole moment, one can show that \( \theta < 10^{-9} \). This is the first real puzzle we have encountered. Why such a small dimensionless number? Answering this question, as we will see, may point to new physics.

1.5 The Hierarchy Problem

The second very puzzling feature is the Higgs field. As of this writing, the Higgs field is the one piece of the Standard Model which has not been seen. Indeed, the structure we have postulated, a single Higgs doublet with a particular potential, might be viewed as somewhat artificial. We could have included several doublets, or perhaps tried to explain the breaking of the gauge symmetry through some more complicated dynamics. But there is a more serious question associated with fundamental scalar fields, raised long ago by Ken Wilson. This problem is often referred to as the “hierarchy problem.”

Consider, first, the one loop corrections to the electron mass in QED. These are logarithmically divergent. In other words,

\[
\delta m = am_0 \frac{\alpha}{4\pi} \ln(\Lambda).
\]  

(21)
We can understand this result in simple terms. In the limit \( m_0 \to 0 \), the theory has an additional symmetry, a chiral symmetry, under which \( e \) and \( \bar{e} \) transform by independent phases. This symmetry forbids a mass term, so the result must be linear in the (bare) mass. So on dimensional grounds, any divergence is at most logarithmic. This actually resolves a puzzle of classical physics, where the self-energy of the electron is linearly divergent.

But for scalars, there is no such symmetry, and corrections to masses are quadratically divergent. One can see this quickly for the Higgs self-coupling, which gives rise to a mass correction of the form:

\[
\delta m^2 = \lambda^2 \int \frac{d^4k}{(2\pi)^4} \frac{1}{k^2 - m^2}
\]  

If we view the standard model as an effective field theory, this integral should be cut off at a scale where new physics enters. We have argued that this might occur at, say, \( 10^{14} \) GeV. But in this case the correction to the Higgs mass is gigantic compared to the Higgs mass itself.

### 1.6 Dark Matter and Dark Energy

In recent years, astronomers and astrophysicists have presented persuasive evidence that the energy density of the universe is largely in some unfamiliar form; about 30% some non-baryonic pressureless matter (the dark matter) and about 60% in some form with negative pressure (the dark energy). The latter might be a cosmological constant (of which more later). The former could well be some new type of weakly interacting particle. Dark matter would indicate the existence of additional, possibly quite light degrees of freedom in nature. The dark energy is totally mysterious. If it represents a cosmological constant, it is even more puzzling than the hierarchy problem we described before. A cosmological constant represents, in essence, the energy density of the vacuum. In field theory this is quadratically divergent; it is the first divergence one encounters in quantum field theory. At one loop, it is given by an expression of the form:

\[
\Lambda = \sum_i (-1)^F \int \frac{d^3k}{2\pi^3} \frac{1}{2} \sqrt{k^2 + m_i^2}
\]  

where the sum is over all particle species (including spins). If one cuts this off, again at \( 10^{14} \) GeV, one gets a result of order

\[
\Lambda = 10^{54} \text{ GeV}^4
\]  

The measured value of the dark energy density, by contrast, is

\[
\Lambda = 10^{-47} \text{ GeV}^4.
\]  

This wide discrepancy is probably one of the most troubling problems facing fundamental physics today.
1.7 Summary: Limitations of the Standard Model

2 Of Anomalies, Strong CP and Axions

In the previous chapter, we mentioned that the Standard Model has an additional parameter, the $\theta$ parameter of QCD. This parameter is dimensionless. It is also CP violating. The $\theta$ term in the QCD lagrangian, on the other hand, is a total divergence, and it is tempting to ignore it. The reasons why this term is dynamically important are interesting and subtle. They are connected with a class of phenomena known as anomalies.

Usually, the term anomaly is used to refer to the violation at the quantum level of a symmetry which is valid classically. Anomalies will be a recurring theme of all beyond the Standard Model physics, and particularly in string theory. In this chapter, we explain how anomalies arise in four dimensional field theories in some detail, and mention how anomalies arise and the role they play and space-times of different dimensionalities.

2.1 The Chiral Anomaly

Before considering real QCD, consider a simpler theory, with only a single flavor of quark. Before making any field redefinitions, the lagrangian takes the form:

$$\mathcal{L} = -\frac{1}{4g^2} F_{\mu\nu}^2 + \bar{q} q^* + q q^* m \bar{q} q + m^* \bar{q}^* q^*. \tag{26}$$

Here, we have written the lagrangian in terms of two-component fermions, and noted that a priori, the mass need not be real,

$$m = |m| e^{i\theta}. \tag{27}$$

In this chapter, it will sometimes be convenient to work with four component fermions, and it is valuable to make contact with this language in any case. In terms of these:

$$\mathcal{L} = \text{Re } m \bar{q} q + \text{Im } m q \bar{q} \gamma_5 q. \tag{28}$$

In order to bring the mass term to the conventional form, with no $\gamma_5$’s, one could try to redefine the fermions:

$$q \rightarrow e^{-i\theta/2} q \quad \bar{q} \rightarrow e^{-i\theta/2} \bar{q}. \tag{29}$$

But in field theory, transformations of this kind are potentially fraught with difficulties because of the infinite number of degrees of freedom.

A simple calculation shows that there is a difficulty, one of the simplest manifestations of an anomaly. Suppose, first, that $M$ is very large. In that case we want to integrate out the quarks and obtain a low energy effective theory. To do this, we study the path integral:

$$Z = \int [dA_\mu] \int [dq] [d\bar{q}] e^{iS} \tag{30}$$
Again suppose \( m = e^{i \theta} M \), where \( \theta \) is small and \( M \) is real. In order to make \( m \) real, we can again make the transformations: \( q \rightarrow q e^{-i \theta/2}; \bar{q} \rightarrow \bar{q} e^{i \theta/2} \) (in four component language, this is \( q \rightarrow -i \theta/2 \gamma^5 \bar{q} \)). The result of integrating out the quark, i.e. of performing the path integral over \( q \) and \( \bar{q} \) can be written in the form:

\[
Z = \int [dA_\mu] \int e^{i S_{\text{eff}}} \tag{31}
\]

Here \( S_{\text{eff}} \) is the effective action which describes the interactions of gluons at scales well below \( M \).

Fig. 1. The triangle diagram associated with the four dimensional anomaly.

Because the field redefinition which eliminates \( \theta \) is just a change of variables in the path integral, one might expect that there can be no \( \theta \)-dependence in the effective action. But this is not the case. To see this, suppose that \( \theta \) is small, and instead of making the transformation, treat the \( \theta \) term as a small perturbation, and expand the exponential. Now consider a term in the effective action with two external gauge bosons. This is given by the Feynman diagram in fig. 1. The corresponding term in the action is given by

\[
\delta L_{\text{eff}} = -i \frac{\theta}{2} g^2 M Tr(T^a T^b) \int \frac{d^4 k}{(2\pi)^4} \gamma_5 \frac{1}{k^+} \frac{1}{\epsilon_1} \frac{1}{\epsilon_2} \frac{1}{k-\epsilon_1-q_1-M} \tag{32}
\]

Here, as in the figure, the \( q_i \)'s are the momenta of the two photons, while the \( \epsilon \)'s are their polarizations and \( a \) and \( b \) are the color indices of the gluons. To perform the integral, it is convenient to introduce Feynman parameters and shift the \( k \) integral, giving:

\[
\delta L_{\text{eff}} = -i \theta g^2 M Tr(T^a T^b) \int d\alpha_1 d\alpha_2 \int \frac{d^4 k}{(2\pi)^4} \gamma_5 (\frac{k - \alpha_1 \epsilon_1 + \alpha_2 \epsilon_2 + \epsilon_1 + M}{k^2 - M^2 + O(q_i^2)}) \frac{1}{\epsilon_1} \frac{1}{\epsilon_2} \frac{1}{k-\epsilon_1-q_1-M} \tag{33}
\]

For small \( q \), we can neglect the \( q \)-dependence of the denominator. The trace in the numerator is easy to evaluate, since we can drop terms linear in \( k \). This gives, after performing the integrals over the \( \alpha \)'s,\n
\[
\delta L_{\text{eff}} = g^2 M^2 \theta \gamma_5 (\frac{1}{k^2 - M^2}) \epsilon^{\mu\nu\rho\sigma} q_1^\mu q_2^\nu \epsilon_1^\rho \epsilon_2^\sigma \tag{34}
\]

This corresponds to a term in the effective action, after doing the integral over \( k \) and including a combinatoric factor of two from the different ways to contract the gauge bosons:

\[
\delta L_{\text{eff}} = \frac{1}{32\pi^2} \theta \gamma_5 (\frac{1}{k^2 - M^2}) \epsilon^{\mu\nu\rho\sigma} q_1^\mu q_2^\nu \epsilon_1^\rho \epsilon_2^\sigma \tag{35}
\]
Now why does this happen? At the level of the path integral, the transformation would seem to be a simple change of variables, and it is hard to see why this should have any effect. On the other hand, if one examines the diagram of fig. 1, one sees that it contains terms which are linearly divergent, and thus it should be regulated. A simple way to regulate the diagram is to introduce a Pauli-Villars regulator, which means that one subtracts off a corresponding amplitude with some very large mass $\Lambda$. However, we have just seen that the result is independent of $\Lambda$! This sort of behavior is characteristic of an anomaly.

Consider now the case that $m \ll \Lambda_{QCD}$. In this case, we shouldn’t integrate out the quarks, but we still need to take into account the regulator diagrams. For small $m$, the classical theory has an approximate symmetry under which

$$q \rightarrow e^{i\alpha}q \quad \bar{q} \rightarrow e^{i\alpha}\bar{q}$$

(in four component language, $q \rightarrow e^{i\alpha\gamma_5}q$). In particular, we can define a current:

$$j_5^\mu = \bar{q}\gamma_5\gamma_\mu q,$$

and classically,

$$\partial_\mu j_5^\mu = m\bar{q}\gamma_5 q.$$

Under a transformation by an infinitesimal angle $\alpha$ one would expect

$$\delta L = \alpha \partial_\mu j_5^\mu = m\alpha\bar{q}\gamma_5 q.$$ (39)

But what we have just discovered is that the divergence of the current contains another, $m$-independent, term:

$$\partial_\mu j_5^\mu = m\bar{q}\gamma_5 q + \frac{1}{32\pi^2}F\tilde{F}.$$ (40)

This anomaly can be derived in a number of other ways. One can define, for example, the current by “point splitting,”

$$j_5^\mu = \bar{q}(x + i\epsilon)e^{i\int_x^{x+\epsilon}dx^\mu A^\mu}q(x)$$

Because operators in quantum field theory are singular at short distances, the Wilson line makes a finite contribution. Expanding the exponential carefully, one recovers the same expression for the current. A beautiful derivation, closely related to that we have performed above, is due to Fujikawa, described in [58]. Here one considers the anomaly as arising from a lack of invariance of the path integral measure. One carefully evaluates the Jacobian associated with the change of variables $q \rightarrow q(1 + i\gamma_5\alpha)$, and shows that it yields the same result[58]. We will do a calculation along these lines in a two dimensional model shortly.

The anomaly has a number of important consequences:
• $\pi^0$ decay: the divergence of the axial isospin current,

$$ (j_5^3)^\mu = \bar{u}\gamma_5\gamma^\mu\bar{u} - \bar{d}\gamma_5\gamma^\mu d $$

(42)

has an anomaly due to electromagnetism. This gives rise to a coupling of the $\pi^0$ to two photons, and the correct computation of the lifetime was one of the early triumphs of the theory of quarks with color.

• Anomalies in gauge currents signal an inconsistency in a theory. They mean that the gauge invariance, which is crucial to the whole structure of gauge theories (e.g. to the fact that they are simultaneously unitary and lorentz invariant) is lost. The absence of gauge anomalies is one of the striking ingredients of the standard model, and it is also crucial in extensions such as string theory.

2.1.1 Return to QCD

What we have just learned is that, if in our simple model above, we require that the quark masses are real, we must allow for the possible appearance in the lagrangian of the standard model, of the $\theta$-terms of eqn. ??.

This term, however, can be removed by a $B + L$ transformation. What are the consequences of these terms? We will focus on the strong interactions, for which these terms are most important. At first sight, one might guess that these terms are in fact of no importance. Consider, first, the case of QED. Then

$$ \int d^4x F\tilde{F} $$

is a the integral of a total divergence,

$$ F\tilde{F} = \vec{E} \cdot \vec{B} = \frac{1}{2} \partial_\mu E^\mu A^\nu F^{\nu\sigma}. \quad (44) $$

As a result, this term does not contribute to the classical equations of motion. One might expect that it does not contribute quantum mechanically either. If we think of the Euclidean path integral, configurations of finite action have field strengths, $F_{\mu\nu}$ which fall off faster than $1/r^2$ (where $r$ is the Euclidean distance), and $A$ which falls off faster than $1/r$, so one can neglect surface terms in $L_\theta$.

(A parenthetical remark: This is almost correct. However, if there are magnetic monopoles there is a subtlety, first pointed out by Witten[59]. Monopoles can carry electric charge. In the presence of the $\theta$ term, there is an extra source for the electric field

\footnote{In principle we must allow a similar term, for the weak interactions. However, $B + L$ is a classical symmetry of the renormalizable interactions of the standard model. This symmetry is anomalous, and can be used to remove the weak $\theta$ term. In the presence of higher dimension $B + L$-violating terms, this is no longer true, but any effects of $\theta$ will be extremely small, suppressed by $e^{-2\pi/\alpha'$} as well as by powers of some large mass scale.}
at long distances, proportional to $\theta$ and the monopole charge. So the electric charges are given by:

$$Q = n_e e - \frac{e\theta n_m}{2\pi}$$  \hspace{1cm} (45)

where $n_m$ is the monopole charge in units of the Dirac quantum.)

In the case of non-Abelian gauge theories, the situation is more subtle. It is again true that $F^2$ can be written as a total divergence:

$$F^2 = \partial^\mu K_\mu = \epsilon_{\mu\nu\rho\sigma}(A^a_\nu F^a_\rho A^b_\sigma - \frac{2}{3} f^{abc} A^a_\nu A^b_\rho A^c_\sigma).$$  \hspace{1cm} (46)

But now the statement that $F$ falls faster than $1/r^2$ does not permit an equally strong statement about $A$. We will see shortly that there are finite action configurations – finite action classical solutions – where $F \sim 1/r^4$, but $A \to 1/r^2$, so that the surface term cannot be neglected. These terms are called instantons. It is because of this that $\theta$ can have real physical effects.

### 2.1.2 A Two Dimensional Detour

Before considering four dimensions with all of its complications, it is helpful to consider two dimensions. Two dimensions are often a poor analog for four, but for some of the issues we are facing here, the parallels are extremely close. In these two dimensional examples, the physics is more manageable, but still rich.

### 2.1.3 The Anomaly In Two Dimensions

Consider, first, electrodynamics of a massless fermion in two dimensions. Let’s investigate the anomaly. The point-splitting method is particularly convenient here. Just as in four dimensions, we write:

$$j^\mu_5 = \bar{\psi}(x + i\epsilon)e^{i\int_{x^+} A_\rho dx^\rho \gamma^\mu} \psi(x)$$  \hspace{1cm} (47)

For very small $\epsilon$, we can pick up the leading singularity in the product of $\psi(x + \epsilon)\psi$ by using the operator product expansion, and noting that (using naive dimensional analysis) the leading operator is the unit operator, with coefficient proportional to $1/\epsilon$. We can read off this term by taking the vacuum expectation value, i.e. by simply evaluating the propagator. String theorists are particularly familiar with this Green’s function:

$$\langle \bar{\psi}(x + \epsilon)\psi(x) \rangle = \frac{1}{2\pi \epsilon^2}$$  \hspace{1cm} (48)

Expanding the factor in the exponential to order $\epsilon$ gives

$$\partial_\mu j^\mu_5 = \text{naive piece} + \frac{i}{2\pi} \partial_\mu \epsilon_\rho A^\rho \text{tr} \frac{\epsilon}{\epsilon^2} \gamma^\mu \gamma^5.$$  \hspace{1cm} (49)
Taking the trace gives $\epsilon_{\mu\nu}\epsilon^\nu$; averaging $\epsilon$ over angles ($<\epsilon_{\mu}\epsilon_{\nu}> = \frac{1}{2} \eta_{\mu\nu}\epsilon^2$), yields

$$\partial_{\mu}j^\mu_5 = \frac{1}{4\pi} \epsilon_{\mu\nu}F^{\mu\nu}.$$  \hspace{1cm} (50)

**Exercise:** Fill in the details of this computation, being careful about signs and factors of 2.

This is quite parallel to the situation in four dimensions. The divergence of the current is itself a total derivative:

$$\partial_{\mu}j^\mu_5 = \frac{1}{2\pi} \epsilon_{\mu\nu} \partial_{\mu}A^{\nu}.$$  \hspace{1cm} (51)

So it appears possible to define a new current,

$$J^\mu = j^\mu_5 - \frac{1}{2\pi} \epsilon_{\mu\nu} \partial_{\mu}A^{\nu}$$  \hspace{1cm} (52)

However, just as in the four dimensional case, this current is not gauge invariant. There is a familiar field configuration for which $A$ does not fall off at infinity: the field of a point charge. Indeed, if one has charges, $\pm \theta$ at infinity, they give rise to a constant electric field, $F_{0i} = e\theta$. So $\theta$ has a very simple interpretation in this theory.

It is easy to see that physics is periodic in $\theta$. For $\theta > q$, it is energetically favorable to produce a pair of charges from the vacuum which shield the charge at $\infty$.

### 2.1.4 The CP$^N$ Model: An Asymptotically Free Theory

The model we have considered so far is not quite like QCD in at least two ways. First, there are no instantons; second, the coupling $e$ is dimensionful. We can obtain a theory closer to QCD by considering the CP$^N$ model (our treatment here will follow closely the treatment in Peskin and Schroeder’s problem 13.3[58]). This model starts with a set of fields, $z_i$, $i = 1, \ldots N + 1$. These fields live in the space $CP^N$. This space is defined by the constraint:

$$\sum_i |z_i|^2 = 1;$$  \hspace{1cm} (53)

in addition, the point $z_i$ is equivalent to $e^{i\alpha}z_i$. To implement the first of these constraints, we can add to the action a lagrange multiplier field, $\lambda(x)$. For the second, we observe that the identification of points in the “target space,” $CP^N$, must hold at every point in ordinary space-time, so this is a $U(1)$ gauge symmetry. So introducing a gauge field, $A_{\mu}$, and the corresponding covariant derivative, we want to study the lagrangian:

$$\mathcal{L} = \frac{1}{g^2} [||D_{\mu}z_i||^2 - \lambda(x)(|z_i|^2 - 1)]$$  \hspace{1cm} (54)
Note that there is no kinetic term for $A_{\mu}$, so we can simply eliminate it from the action using its equations of motion. This yields

$$\mathcal{L} = \frac{1}{g^2} [\partial_{\mu} z_j]^2 + [z_{j}^{*} \partial_{\mu} z_j]^2] \quad (55)$$

It is easier to proceed, however, keeping $A_{\mu}$ in the action. In this case, the action is quadratic in $z$, and we can integrate out the $z$ fields:

$$Z = \int [dA][d\lambda][dz_j] e^{\int [-\mathcal{L}]} = \int [dA][d\lambda] e^{\int d^2x \Gamma_{\text{eff}}[A,\lambda]} \quad (56)$$

$$= \int [dA][d\lambda] e^{\int [-N \text{tr log}(-D^2 - \lambda) - \frac{1}{g^2} \int d^2x \lambda]}$$

2.1.5 The Large $N$ Limit

By itself, the result of eqn. 56 is still rather complicated. The fields $A_{\mu}$ and $\lambda$ have complicated, non-local interactions. Things become much simpler if one takes the “large $N$ limit”, a limit where one takes $N \to \infty$ with $g^2 N$ fixed. In this case, the interactions of $\lambda$ and $A_{\mu}$ are suppressed by powers of $N$. For large $N$, the path integral is dominated by a single field configuration, which solves

$$\frac{\delta \Gamma_{\text{eff}}}{\delta \lambda} = 0 \quad (57)$$

or, setting the gauge field to zero,

$$N \int \frac{d^2k}{(2\pi)^2} \frac{1}{k^2 + \lambda} = \frac{1}{g^2}, \quad (58)$$

giving

$$\lambda = m^2 = M \exp[-\frac{2\pi}{g^2 N}]. \quad (59)$$

Here, $M$ is a cutoff required because the integral in eqn. 58 is divergent. This result is remarkable. One has exhibited dimensional transmutation: a theory which is classically scale invariant contains non-trivial masses, related in a renormalization-group invariant fashion to the cutoff. This is the phenomenon which in QCD explains the masses of the proton, neutron, and other dimensionful quantities. So the theory is quite analogous to QCD. We can read off the leading term in the $\beta$-function from the familiar formula:

$$m = Me^{-\int \frac{dg}{\pi \beta(g)}} \quad (60)$$

so, with

$$\beta(g) = -\frac{1}{2\pi} g^3 b_0 \quad (61)$$
we have $b_0 = 1$.

But most important for our purposes, it is interesting to explore the question of $\theta$-dependence. Again, in this theory, we could have introduced a $\theta$ term:

$$L_\theta = \frac{\theta}{2\pi} \int d^2 \epsilon_{\mu\nu} F^{\mu\nu},$$

(62)

where $F_{\mu\nu}$ can be expressed in terms of the fundamental fields $z_j$. As usual, this is the integral of a total divergence. But precisely as in the case of $1+1$ dimensional electrodynamics we discussed above, this term is physically important. In perturbation theory in the model, this is not entirely obvious. But using our reorganization of the theory at large $N$, it is. The lowest order action for $A_\mu$ is trivial, but at one loop (order $1/N$), one generates a kinetic term for $A$ through the usual vacuum polarization loop:

$$L_{\text{kin}} = \frac{N}{2\pi m^2} F_{\mu\nu}^2.$$  

(63)

At this order, the effective theory consists of the gauge field, then, with coupling $e^2 = \frac{2\pi m^2}{N}$, and some coupling to a dynamical, massive field $\lambda$. As we have already argued, $\theta$ corresponds to a non-zero background electric field due to charges at infinity, and the theory clearly can have non-trivial $\theta$-dependence.

There is, in addition, the possibility of including other light fields, for example massless fermions. In this case, one can again have an anomalous $U(1)$ symmetry. There is then no $\theta$-dependence, since it is possible to shield any charge at infinity. But there is non-trivial breaking of the symmetry. At low energies, one has now a theory with a fermion coupled to a dynamical $U(1)$ gauge field. The breaking of the associated $U(1)$ in such a theory is a well-studied phenomenon[60].

**Exercise:** Complete Peskin and Schroeder, Problem 13.3.

2.1.6 The Role of Instantons

There is another way to think about the breaking of the $U(1)$ symmetry and $\theta$-dependence in this theory. If one considers the Euclidean functional integral, it is natural to look for stationary points of the integration, i.e. for classical solutions of the Euclidean equations of motion. In order that they be potentially important, it is necessary that these solutions have finite action, which means that they must be localized in Euclidean space and time. For this reason, such solutions were dubbed “instantons” by ’t Hooft. Such solutions are not difficult to find in the $CP^N$ model; we will describe them briefly below. These solutions carry non-zero values of the topological charge,

$$\frac{1}{2\pi} \int d^2 x \epsilon_{\mu\nu} F_{\mu\nu} = n$$

(64)
and have an action $2\pi n$. As a result, they contribute to the $\theta$-dependence; they give a contribution to the functional integral:

$$Z_{\text{inst}} = e^{\frac{-2\pi n}{\theta}} e^{in\theta} \int [d\delta z] e^{-\delta z_i \delta \bar{z}_i} e^{\delta z_j + \ldots}$$  \hspace{1cm} (65)

It follows that:

- Instantons generate $\theta$-dependence.
- In the large $N$ limit, instanton effects are, formally, highly suppressed, much smaller that the effects we found in the large $N$ limit.
- Somewhat distressingly, the functional integral above can not be systematically evaluated. The problem is that the classical theory is scale invariant, as a result of which, instantons come in a variety of sizes. $\int [d\delta z]$ includes an integration over all instanton sizes, which diverges for large size (i.e. in the infrared). This prevents a systematic evaluation of the effects of instantons in this case. At high temperatures[61], it is possible to do the evaluation, and instanton effects are, indeed, systematically small.

It is easy to construct the instanton solution in the case of $CP^1$. Rather than write the theory in terms of a gauge field, as we have done above, it is convenient to parameterize the theory in terms of a single complex field, $\phi$. One can, for example, define $\phi = z_1/z_2$. Then, with a bit of algebra, one can show that the action for $\phi$ takes the form:

$$L = (\partial_\mu \phi \partial_\mu \phi^*) \frac{1}{1 + \phi^* \phi} - \frac{\phi^* \phi}{(1 + \phi^* \phi)^2}. \hspace{1cm} (66)$$

One can think of the field $\phi$ as living on the space with metric given by the term in parenthesis, $g_{\phi \phi^*}$. One can show that this is the metric one obtains if one stereographically maps the sphere onto the complex plane. This mapping, which you may have seen in your math methods courses, is just:

$$z = \frac{x_1 + ix_2}{1 - x_3}; \hspace{1cm} (67)$$

The inverse is

$$x_1 = \frac{z + z^*}{1 + |z|^2} \quad x_2 = \frac{z - z^*}{i(1 + |z|^2)} \quad x_3 = \frac{|z|^2 - 1}{|z|^2 + 1}. \hspace{1cm} (68)$$

It is straightforward to write down the equations of motion:

$$\partial^2 \phi g_{\phi \phi} + \partial_\mu \phi (\partial_\mu \phi \partial_\phi + \partial_\mu \bar{\phi} \partial_{\bar{\phi}}) = 0. \hspace{1cm} (69)$$
Now calling the space time coordinates $z = x_1 + ix_2, \; z^* = x_1 - ix_2$, you can see that if $\phi$ is analytic, the equations of motion are satisfied! So a simple solution, which you can check has finite action, is

$$\phi(z) = \rho z.$$  \hfill (70)

In addition to evaluating the action, you can evaluate the topological charge,

$$\frac{1}{2\pi} \int d^2x \epsilon_{\mu\nu} F^{\mu\nu} = 1$$  \hfill (71)

for this solution. More generally, the topological charge measures the number of times that $\phi$ maps the complex plane into the complex plane; for $\phi = z^n$, for example, one has charge $n$.

**Exercise:** Verify that the action of eqn. 66 is equal to

$$\mathcal{L} = g_{\phi,\phi^*} \partial_\mu \phi \partial_\mu \phi^*$$  \hfill (72)

where $g$ is the metric of the sphere in complex coordinates, i.e. it is the line element $dx_1^2 + dx_2^2 + dx_3^2$ expressed as $g_{z,z} dz \, dz + g_{z,z^*} dz \, dz^* + g_{z^*,z^*} dz^* \, dz + g_{dz, dz^*} dz^* \, dz$. A model with an action of this form is called a “Non-linear Sigma Model;” the idea is that the fields live on some “target” space, with metric $g$. Verify eqns. 66,68.

More generally, $\phi = \frac{az+b}{cz+d}$ is a solution with action $2\pi$. The parameters $a, \ldots d$ are called collective coordinates. They correspond to the symmetries of translations, dilations, and rotations, and special conformal transformations (forming the group $SL(2,\mathbb{C})$). In other words, any given finite action solution breaks the symmetries. In the path integral, the symmetry of Green’s functions is recovered when one integrates over the collective coordinates. For translations, this is particularly simple. If one studies a Green’s function,

$$<\phi(x)\phi(y)> \approx \int d^2x_o \phi_{cl}(x-x_o)\phi_{cl}(y-y_o)e^{-S_o}$$  \hfill (73)

The precise measure is obtained by the Fadeev-Popov method. Similarly, the integration over the parameter $\rho$ yields a factor

$$\int d\rho \rho^{-1} e^{-\frac{2\pi}{\nu^2(\omega)}} \ldots$$  \hfill (74)

Here the first factor follows on dimensional grounds. The second follows from renormalization-group considerations. It can be found by explicit evaluation of the functional determinant[62]. Note that, because of asymptotic freedom, this means that typical Green’s functions will be divergent in the infrared.

There are many other features of this instanton one can consider. For example, one can consider adding massless fermions to the model, by simply coupling them in a gauge-invariant way to $A_\mu$. The resulting theory has a chiral $U(1)$ symmetry, which is anomalous. In the presence of an instanton, one can easily construct normalizable
fermion zero modes (the Dirac equation just becomes the statement that $\psi$ is analytic). As a result, Green’s functions computed in the instanton background do not respect the axial $U(1)$ symmetry. But rather than get too carried away with this model (I urge you to get a little carried away and play with it a bit), let’s proceed to four dimensions, where we will see very similar phenomena.

2.2 Real QCD

The model of the previous section mimics many features of real QCD. Indeed, we will see that much of our discussion can be carried over, almost word for word, to the observed strong interactions. This analogy is helpful, given that in QCD we have no approximation which gives us control over the theory comparable to that which we found in the large $N$ limit of the $CP^N$ model. As in that theory:

- There is a $\theta$ parameter, which appears as an integral over the divergence of a non-gauge invariant current.
- There are instantons, which indicate that there should be real $\theta$-dependence. However, instanton effects cannot be considered in a controlled approximation, and there is no clear sense in which $\theta$-dependence can be understood as arising from instantons.
- There is another approach to the theory, which shows that the $\theta$-dependence is real, and allows computation of these effects. In QCD, this is related to the breaking of chiral symmetries.

2.2.1 The Theory and its Symmetries

While it is not in the spirit of much of this school, which is devoted to the physics of heavy quarks, it is sufficient, to understand the effects of $\theta$, to focus on only the light quark sector of QCD. For simplicity in writing some of the formulas, we will consider two light quarks; it is not difficult to generalize the resulting analysis to the case of three. It is believed that the masses of the $u$ and $d$ quarks are of order $5$ MeV and $10$ MeV, respectively, much lighter than the scale of QCD. So we first consider an idealization of the theory in which these masses are set to zero. In this limit, the theory has a symmetry $SU(2)_L \times SU(2)_R$. This symmetry is spontaneously broken to a vector $SU(2)$. The three resulting Goldstone bosons are the $\pi$ mesons. Calling

$$q = \begin{pmatrix} u \\ d \end{pmatrix} \quad \bar{q} = \begin{pmatrix} \bar{u} \\ \bar{d} \end{pmatrix},$$

(75)

the two $SU(2)$ symmetries act separately on $q$ and $\bar{q}$ (thought of as left handed fermions). The order parameter for the symmetry breaking is believed to be the condensate:

$$\mathcal{M}_o = \langle \bar{q} q \rangle.$$  

(76)
This indeed breaks the symmetry down to the vector sum. The associated Goldstone bosons are the $\pi$ mesons. One can think of the Goldstone bosons as being associated with a slow variation of the expectation value in space, so we can introduce a composite operator

$$\mathcal{M} = \bar{q}q = M e^{i\frac{z_a(x)\chi_a}{f_\pi}} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

The quark mass term in the lagrangian is then (for simplicity writing $m_u = m_d = m_q$; more generally one should introduce a matrix)

$$m_q \mathcal{M}.$$  \hfill (78)

Expanding $\mathcal{M}$ in powers of $\pi/f_\pi$, it is clear that the minimum of the potential occurs for $\pi_a = 0$. Expanding to second order, one has

$$m_\pi^2 f_\pi^2 = m_q M_0.$$  \hfill (79)

But we have been a bit cavalier about the symmetries. The theory also has two $U(1)$'s;

$$q \rightarrow e^{i\alpha}q \quad \bar{q} \rightarrow e^{i\alpha}\bar{q} \quad \bar{q} \rightarrow e^{-i\alpha}\bar{q}$$

The first of these is baryon number and it is not chiral (and is not broken by the condensate). The second is the axial $U(1)_A$; It is also broken by the condensate. So, in addition to the pions, there should be another approximate Goldstone boson. The best candidate is the $\eta$, but, as we will see below (and as you will see further in Thomas's lectures), the $\eta$ is too heavy to be interpreted in this way. The absence of this fourth (or in the case of three light quarks, ninth) Goldstone boson is called the $U(1)$ problem.

The $U(1)_A$ symmetry suffers from an anomaly, however, and we might hope that this has something to do with the absence of a corresponding Goldstone boson. The anomaly is given by

$$\partial_\mu j_5^\mu = \frac{2}{32\pi^2} F \bar{F}$$

Again, we can write the right hand side as a total divergence,

$$F \bar{F} = \partial_\mu K^\mu$$

where

$$K_\mu = \epsilon_{\mu\nu\rho\sigma} (A_\nu^a F_{\rho\sigma} - \frac{2}{3} f^{abc} A_\nu^a A_\rho^b A_\sigma^c).$$

So if it is true that this term accounts for the absence of the Goldstone boson, we need to show that there are important configurations in the functional integral for which the rhs does not vanish rapidly at infinity.
2.2.2 Instantons

It is easiest to study the Euclidean version of the theory. This is useful if we are interested in very low energy processes, which can be described by an effective action expanded about zero momentum. In the functional integral,

$$ Z = \int [dA][dq][d\bar{q}] e^{-S} $$

(84)

it is natural to look for stationary points of the effective action, i.e. finite action, classical solutions of the theory in imaginary time. The Yang-Mills equations are complicated, non-linear equations, but it turns out that, much as in the $CP^N$ model, the instanton solutions can be found rather easily[63]. The following tricks simplify the construction, and turn out to yield the general solution. First, note that the Yang-Mills action satisfies an inequality:

$$ \int (F \pm \tilde{F})^2 = \int (F^2 + \tilde{F}^2 \pm 2F\tilde{F}) = \int (2F^2 + 2F\tilde{F}) \geq 0. $$

(85)

So the action is bounded $\int F\tilde{F}$, with the bound being saturated when

$$ F = \pm \tilde{F} $$

(86)

i.e. if the gauge field is (anti) self dual.\(^2\) This equation is a first order equation, and it is easy to solve if one first restricts to an $SU(2)$ subgroup of the full gauge group. One makes the ansatz that the solution should be invariant under a combination of ordinary rotations and global $SU(2)$ gauge transformations:

$$ A_\mu = f(r^2) + h(r^2)\vec{x} \cdot \vec{\tau} $$

(87)

where we are using a matrix notation for the gauge fields. One can actually make a better guess: define the gauge transformation:

$$ g(x) = \frac{x_4 + i\vec{x} \cdot \vec{\tau}}{r} $$

(88)

and take

$$ A_\mu = f(r^2)g\partial_\mu g^{-1} $$

(89)

Then plugging in the Yang-Mills equations yields:

$$ f = \frac{r^2}{r^2 + \rho^2} $$

(90)

\(^2\)This is not an accident, nor was the analyticity condition in the $CP^N$ case. In both cases, we can add fermions so that the model is supersymmetric. Then one can show that if some of the supersymmetry generators, $Q_\alpha$, annihilate a field configuration, then the configuration is a solution. This is a first order condition; in the Yang-Mills case, it implies self-duality, and in the $CP^N$ case it requires analyticity.
where $\rho$ is an arbitrary quantity with dimensions of length. The choice of origin here is arbitrary; this can be remedied by simply replacing $x \to x - x_o$ everywhere in these expressions, where $x_o$ represents the location of the instanton.

**Exercise** Check that eqns. 89,90 solve 86.

> From this solution, it is clear why $\int \partial_\mu K^\mu$ does not vanish for the solution: while $A$ is a pure gauge at infinity, it falls only as $1/r$. Indeed, since $F = \tilde{F}$, for this solution,

$$\int F^2 = \int \tilde{F}^2 = 32\pi^2$$  \hfill (91)

This result can also be understood topologically. $g$ defines a mapping from the “sphere at infinity” into the gauge group. It is straightforward to show that

$$\frac{1}{32\pi^2} \int d^4xF\tilde{F}$$  \hfill (92)

counts the number of times $g$ maps the sphere at infinity into the group (one for this specific example; $n$ more generally). We do not have time to explore all of this in detail; I urge you to look at Sidney Coleman’s lecture, “The Uses of Instantons”[64]. To actually do calculations, ’t Hooft developed some notations which are often more efficient than those described above[62].

So we have exhibited potentially important contributions to the path integral which violate the $U(1)$ symmetry. How does this violation of the symmetry show up? Let’s think about the path integral in a bit more detail. Having found a classical solution, we want to integrate about small fluctuations about it:

$$Z = e^{-\frac{ax^2}{2} + i\theta} \int [d\delta A][dq][d\bar{q}] e^{i\delta^2 S}$$  \hfill (93)

Now $S$ contains an explicit factor of $1/g^2$. As a result, the fluctuations are formally suppressed by $g^2$ relative to the leading contribution. The one loop functional integral yields a product of determinants for the fermions, and of inverse square root determinants for the bosons. Consider, first, the integral over the fermions. It is straightforward, if challenging, to evaluate the determinants[62]. But if the quark masses are zero, the fermion functional integrals are zero, because there is a zero mode for each of the fermions, i.e. for both $q$ and $\bar{q}$ there is a normalizable solution of the equation:

$$\slashed{D}u = 0 \quad \slashed{D}\bar{u} = 0$$  \hfill (94)

and similarly for $d$ and $\bar{d}$. It is straightforward to construct these solutions:

$$u = \frac{\rho}{(\rho^2 + (x - x_o)^2)^{3/2}} \zeta$$  \hfill (95)

where $\zeta$ is a constant spinor, and similarly for $\bar{u}$, etc.
This means that in order for the path integral to be non-vanishing, we need to include insertions of enough $q$’s and $\bar{q}$’s to soak up all of the zero modes. In other words, non-vanishing Green’s functions have the form

$$\langle \bar{u}u \bar{d}d \rangle$$

and violate the symmetry. Note that the symmetry violation is just as predicted from the anomaly equation:

$$\Delta Q_5 = 4 \frac{1}{\pi^2} \int d^4xF \bar{F} = 4$$

However, the calculation we have described here is not self consistent. The difficulty is that among the variations of the fields we need to integrate over are changes in the location of the instanton (translations), rotations of the instanton, and scale transformations. The translations are easy to deal with; one has simply to integrate over $x_o$ (one must also include a suitable Jacobian factor[64]). Similarly, one must integrate over $\rho$. There is a power of $\rho$ arising from the Jacobian, which can be determined on dimensional grounds. For our Green’s function above, for example, which has dimension 6, we have (if all of the fields are evaluated at the same point),

$$\int d\rho \rho^{-7}.$$  

However, there is additional $\rho$-dependence because the quantum theory violates the scale symmetry. This can be understood by replacing $g^2 \rightarrow g^2(\rho)$ in the functional integral, and using

$$e^{-8\pi^2g^2(\rho)} \approx (\rho M)^{b_o}$$

for small $\rho$. For 3 flavor QCD, for example, $b_o = 9$, and the $\rho$ integral diverges for large $\rho$. This is just the statement that the integral is dominated by the infrared, where the QCD coupling becomes strong.

So we have provided some evidence that the $U(1)$ problem is solved in QCD, but no reliable calculation. What about $\theta$-dependence? Let us ask first about $\theta$-dependence of the vacuum energy. In order to get a non-zero result, we need to allow that the quarks are massive. Treating the mass as a perturbation, we obtain

$$E(\theta) = C\Lambda^9_{QCD} m_u m_d \cos(\theta) \int d\rho \rho^{-3} \rho^9.$$  

So again, we have evidence for $\theta$-dependence, but cannot do a reliable calculation. That we cannot do a calculation should not be a surprise. There is no small parameter in QCD to use as an expansion parameter. Fortunately, we can use other facts which we know about the strong interactions to get a better handle on both the $U(1)$ problem and the question of $\theta$-dependence.

Before continuing, however, let us consider the weak interactions. Here there is a small parameter, and there are no infrared difficulties, so we might expect instanton
effects to be small. The analog of the $U(1)_5$ symmetry in this case is baryon number. Baryon number has an anomaly in the standard model, since all of the quark doublets have the same sign of the baryon number. 't Hooft realized that one could actually use instantons, in this case, to compute the violation of baryon number. Technically, there are no finite action Euclidean solutions in this theory; this follows, as we will see in a moment, from a simple scaling argument. However, 't Hooft realized that one can construct important configurations of non-zero topological charge by starting with the instantons of the pure gauge theory and perturbing them. If one simply takes such an instanton, and plugs it into the action, one necessarily finds a correction to the action of the form

$$\delta S = \frac{1}{g^2} v^2 \rho^2.$$  \hfill (101)

This damps the $\rho$ integral at large $\rho$, and leads to a convergent result. Affleck showed how to develop this into a systematic computation[66]. Note that from this, one can see that baryon number violation occurs in the standard model, and that the rate is incredibly small, proportional to $e^{-2\pi \alpha W}$.

### 2.2.3 Real QCD and the U(1) Problem

In real QCD, it is difficult to do a reliable calculation which shows that there is not an extra Goldstone boson, but the instanton analysis we have described makes clear that there is no reason to expect one. Actually, while perturbative and semiclassical (instanton) techniques have no reason to give reliable results, there are two approximation methods available. The first is large $N$, where one now allows the $N$ of $SU(N)$ to be large, with $g^2 N$ fixed. In contrast to the case of $CP^N$, this does not permit enough simplification to do explicit computations, but it does allow one to make qualitative statements about the theory. available in QCD. Witten has pointed out a way in which one can at relate the mass of the $\eta$ (or $\eta'$ if one is thinking in terms of $SU(3) \times SU(3)$ current algebra) to quantities in a theory without quarks. The point is to note that the anomaly is an effect suppressed by a power of $N$, in the large $N$ limit. This is because the loop diagram contains a factor of $g^2$ but not of $N$. So, in large $N$, it can be treated as a perturbation, and the the $\eta$ is massless. $\partial_\mu j_5^\mu$ is like a creation operator for the $\eta$, so (just like $\partial_\mu j_3^\mu$ is a creation operator for the $\pi$ meson), so one can compute the mass if one knows the correlation function, at zero momentum, of

$$\langle \partial_\mu j_5^\mu(x) \partial_\mu j_5^\mu(y) \rangle \propto \frac{1}{N^2} \langle F(x) \bar{F}(x) F(y) \bar{F}(y) \rangle$$  \hfill (102)

To leading order in the $1/N$ expansion, this correlation function can be computed in the theory without quarks. Witten argued that while this vanishes order by order in perturbation theory, there is no reason that this correlation function need vanish in the full theory. Attempts have been made to compute this quantity both in lattice gauge theory and using
the AdS-CFT correspondence recently discovered in string theory. Both methods give promising results.

So the U(1) problem should be viewed as solved, in the sense that absent any argument to the contrary, there is no reason to think that there should be an extra Goldstone boson in QCD.

The second approximation scheme which gives some control of QCD is known as chiral perturbation theory. The masses of the $u$, $d$ and $s$ quarks are small compared to the QCD scale, and the mass terms for these quarks in the lagrangian can be treated as perturbations. This will figure in our discussion in the next section.

2.3 The Strong CP Problem

2.3.1 $\theta$-dependence of the Vacuum Energy

The fact that the anomaly resolves the $U(1)$ problem in QCD, however, raises another issue. Given that $\int d^4xF\tilde{F}$ has physical effects, the theta term in the action has physical effects as well. Since this term is CP odd, this means that there is the potential for strong CP violating effects. These effects should vanish in the limit of zero quark masses, since in this case, by a field redefinition, we can remove $\theta$ from the lagrangian. In the presence of quark masses, the $\theta$-dependence of many quantities can be computed. Consider, for example, the vacuum energy. In QCD, the quark mass term in the lagrangian has the form:

$$L_m = m_u\bar{u}u + m_d\bar{d}d + h.c.$$  \hspace{1cm} (103)

Were it not for the anomaly, we could, by redefining the quark fields, take $m_u$ and $m_d$ to be real. Instead, we can define these fields so that there is no $\theta F\tilde{F}$ term in the action, but there is a phase in $m_u$ and $m_d$. Clearly, we have some freedom in making this choice. In the case that $m_u$ and $m_d$ are equal, it is natural to choose these phases to be the same. We will explain shortly how one proceeds when the masses are different (as they are in nature). We can, by convention, take $\theta$ to be the phase of the overall lagrangian:

$$L_m = (m_u\bar{u}u + m_d\bar{d}d)\cos(\theta/2) + h.c.$$  \hspace{1cm} (104)

Now we want to treat this term as a perturbation. At first order, it makes a contribution to the ground state energy proportional to its expectation value. We have already argued that the quark bilinears have non-zero vacuum expectation values, so

$$E(\theta) = (m_u + m_d)e^{i\theta}\langle\bar{q}q\rangle.$$  \hspace{1cm} (105)

While, without a difficult non-perturbative calculation, we can’t calculate the separate quantities on the right hand side of this expression, we can, using current algebra, relate them to measured quantities. A simple way to do this is to use the effective lagrangian method (which will be described in more detail in Thomas’s lectures). The
basic idea is that at low energies, the only degrees of freedom which can readily be
excited in QCD are the pions. So parameterize $\bar{q}q$ as

$$\bar{q}q = \Sigma = <\bar{q}q > e^{i\pi(x)\pi^a}$$ (106)

We can then write the quark mass term as

$$L_m = e^{i\theta} Tr M_q \Sigma.$$ (107)

Ignoring the $\theta$ term at first, we can see, plugging in the explicit form for $\Sigma$, that

$$m_\pi^2 f_\pi^2 = (m_u + m_d) <\bar{q}q >.$$ (108)

So the vacuum energy, as a function of $\theta$, is:

$$E(\theta) = m_\pi^2 f_\pi^2 \cos(\theta).$$ (109)

This expression can readily be generalized to the case of three light quarks by similar
methods. In any case, we now see that there is real physics in $\theta$, even if we don’t
understand how to do an instanton calculation. In the next section, we will calculate a
more physically interesting quantity: the neutron electric dipole moment as a function
of $\theta$.

2.3.2 The Neutron Electric Dipole Moment

As Scott Thomas will explain in much greater detail in his lectures, the most interesting
physical quantities to study in connection with CP violation are electric dipole moments,
particularly that of the neutron, $d_n$. It has been possible to set strong experimental limits
on this quantity. Using current algebra, the leading contribution to the neutron electric
dipole moment due to $\theta$ can be calculated, and one obtains a limit $\theta < 10^{-9}$[65]. The
original paper on the subject is quite readable. Here we outline the main steps in the
calculation; I urge you to work out the details following the reference. We will simplify
the analysis by working in an exact $SU(2)$-symmetric limit, i.e. by taking $m_u = m_d = m$.
We again treat the lagrangian of [104] as a perturbation. We can also understand
how this term depends on the $\pi$ fields by making an axial $SU(2)$ transformation on the
quark fields. In other words, a background $\pi$ field can be thought of as a small chiral
transformation from the vacuum. Then, e.g., for the $\tau_3$ direction, $q \rightarrow (1 + i\pi_3\tau_3)q$ (the
$\pi$ field parameterizes the transformation), so the action becomes:

$$\frac{m}{f_\pi}\pi_3(\bar{q}\gamma_5 q + \theta \bar{q}q)$$ (110)

In other words, we have calculated a CP violating coupling of the mesons to the pions.
This coupling is difficult to measure directly, but it was observed in [65] that this coupling gives rise, in a calculable fashion, to a neutron electric dipole moment. Consider the graph of fig. 2. This graph generates a neutron electric dipole moment, if we take one coupling to be the standard pion-nucleon coupling, and the second the coupling we have computed above. The resulting Feynman graph is infrared divergent; we cut this off at \( m_\pi \), while cutting off the integral in the ultraviolet at the QCD scale. Because of this infrared sensitivity, the low energy calculation is reliable. The exact result is:

\[
d_n = g_{\pi NN} \frac{-\theta m_u m_d}{f_\pi (m_u + m_d)} \langle N_f | \bar{q} \tau^a q | N_i \rangle \ln(M_N/m_\pi) \frac{1}{4\pi^2} M_N. \tag{111}
\]

The matrix element can be estimated using ordinary \( SU(3) \), yielding \( d_n = 5.2 \times 10^{-16} \theta \) cm. The experimental bound gives \( \theta < 10^{-9} - 10^{-10} \). Understanding why CP violation is so small in the strong interactions is the “strong CP problem.”

### 2.4 Possible Solutions

What should our attitude towards this problem be? We might argue that, after all, some Yukawa couplings are as small as \( 10^{-5} \), so why is \( 10^{-9} \) so bad. On the other hand, we suspect that the smallness of the Yukawa couplings is related to approximate symmetries, and that these Yukawa couplings are telling us something. Perhaps there is some explanation of the smallness of \( \theta \), and perhaps this is a clue to new physics. In this section we review some of the solutions which have been proposed to understand the smallness of \( \theta \).

#### 2.4.1 The Axion

Perhaps the most popular explanation of the smallness of \( \theta \) involves a hypothetical particle called the axion. We present here a slightly updated version of the original idea of Peccei and Quinn[73].

Consider the vacuum energy as a function of \( \theta \), eqn. [105]. This energy has a minimum at \( \theta = 0 \), i.e. at the CP conserving point. As Weinberg noted long ago, this is almost automatic: points of higher symmetry are necessarily stationary points. As it stands, this observation is not particularly useful, since \( \theta \) is a parameter, not a dynamical variable. But suppose that one has a particle, \( a \), with coupling to QCD:

\[
\mathcal{L}_{\text{axion}} = (\partial_\mu a)^2 + \frac{(a/f_a + \theta)}{32\pi^2} F \tilde{F} \tag{112}
\]
$f_a$ is known as the axion decay constant. Suppose that the rest of the theory possesses a symmetry, called the Peccei-Quinn symmetry,

$$a \rightarrow a + \alpha$$

(113)

for constant $\alpha$. Then by a shift in $a$, one can eliminate $\theta$. $E(\theta)$ is now $V(a/f_a)$, the potential energy of the axion. It has a minimum at $\theta = 0$. The strong CP problem is solved.

One can estimate the axion mass by simply examining $E(\theta)$.

$$m_a^2 \approx \frac{m_a^2 f_a^2}{f_a^2}$$

(114)

If $f_a \sim TeV$, this yields a mass of order KeV. If $f_a \sim 10^{16}$ GeV, this gives a mass of order $10^{-9}$ eV. As for the $\theta$-dependence of the vacuum energy, it is not difficult to get the factors straight using current algebra methods. A collection of formulae, with great care about factors of 2, appears in [74]

Actually, there are several questions one can raise about this proposal:

- Should the axion already have been observed? In fact, as Scott Thomas will explain in greater detail in his lectures, the couplings of the axion to matter can be worked out in a straightforward way, using the methods of current algebra (in particular of non-linear lagrangians). All of the couplings of the axion are suppressed by powers of $f_a$. So if $f_a$ is large enough, the axion is difficult to see. The strongest limit turns out to come from red giant stars. The production of axions is “semiweak,” i.e. it only is suppressed by one power of $f_a$, rather than two powers of $m_W$; as a result, axion emission is competitive with neutrino emission until $f_a > 10^{10}$ GeV or so.

- As we will describe in a bit more detail below, the axion can be copiously produced in the early universe. As a result, there is an upper bound on the axion decay constant. In this case, as we will explain below, the axion could constitute the dark matter.

- Can one search for the axion experimentally[75]? Typically, the axion couples not only to the $F\tilde{F}$ of QCD, but also to the same object in QED. This means that in a strong magnetic field, an axion can convert to a photon. Precisely this effect is being searched for by a group at Livermore (the collaboration contains members from MIT, University of Florida) and Kyoto. The basic idea is to suppose that the dark matter in the halo consists principally of axions. Using a (superconducting) resonant cavity with a high Q value in a large magnetic field, one searches for the conversion of these axions into excitations of the cavity. The experiments have already reached a level where they set interesting limits; the next generation of experiments will cut a significant swath in the presently allowed parameter space.
• The coupling of the axion to $F \tilde{F}$ violates the shift symmetry; this is why the axion can develop a potential. But this seems rather paradoxical: one postulates a symmetry, preserved to some high degree of approximation, but which is not a symmetry; it is at least broken by tiny QCD effects. Is this reasonable? To understand the nature of the problem, consider one of the ways an axion can arise. In some approximation, we can suppose we have a global symmetry under which a scalar field, $\phi$, transforms as $\phi \rightarrow e^{i\alpha} \phi$. Suppose, further, that $\phi$ has an expectation value, with an associated (pseudo)-Goldstone boson, $a$. We can parameterize $\phi$ as:

$$\phi = f_a e^{i\alpha} / f_a \quad |\langle \phi \rangle| = f_a$$  \hspace{1cm} (115)

If this field couples to fermions, so that they gain mass from its expectation value, then at one loop we generate a coupling $aF \tilde{F}$ from integrating out the fermions. This calculation is identical to the corresponding calculation for pions we discussed earlier. But we usually assume that global symmetries in nature are accidents. For example, baryon number is conserved in the standard model simply because there are no gauge-invariant, renormalizable operators which violate the symmetry. We believe it is violated by higher dimension terms. The global symmetry we postulate here is presumably an accident of the same sort. But for the axion, the symmetry must be extremely good. For example, suppose one has a symmetry breaking operator

$$\frac{\phi^{n+4}}{M_p^n}$$  \hspace{1cm} (116)

Such a term gives a linear contribution to the axion potential of order $\frac{f_{a}^{n+3}}{M_p^{n}}$. If $f_a \sim 10^{11}$, this swamps the would-be QCD contribution $(\frac{m_\pi^2 f_\pi^2}{f_a})$ unless $n > 12$ [76]!

This last objection finds an answer in string theory. In this theory, there are axions, with just the right properties, i.e. there are symmetries in the theory which are exact in perturbation theory, but which are broken by exponentially small non-perturbative effects. The most natural value for $f_a$ would appear to be of order $M_{GUT} - M_p$. Whether this can be made compatible with cosmology, or whether one can obtain a lower scale, is an open question.

### 3 An Introduction to Supersymmetry

In a standard advanced field theory course, one learns about a number of symmetries: Poincare invariance, global continuous symmetries, discrete symmetries, gauge symmetries, approximate and exact symmetries. The latter symmetries all have the property that they commute with Lorentz transformations, and in particular they commute with rotations. So the multiplets of the symmetries always contain particles of the same spin; in particular, they always consist of either bosons or fermions.
For a long time, it was believed that these were the only allowed symmetries; this statement was even embodied in a theorem, known as the Coleman-Mandula theorem. However, physicists studying theories based on strings stumbled on a symmetry which related fields of different spin. Others quickly worked out simple field theories with this new symmetry: supersymmetry.

A good introduction to the formal theory of supersymmetry is provided by[?]. Here we will mention just a few important features. Supersymmetric field theories can be formulated in dimensions up to eleven. These higher dimensional theories will be important when we consider string theory. For now, we consider theories in four dimensions. The supersymmetry charges, because they change spin, must themselves carry spin – they are spin-1/2. They transform as doublets under the lorentz group, like the spinors $\chi$ and $\chi^*$. There can be 1, 2, 4 or 8 such spinors; correspondingly, the symmetry is said to be $N = 1, 2, 4$ or $8$ supersymmetry. Like generators of an ordinary group, the supersymmetry generators obey an algebra; unlike an ordinary bosonic group, however, the algebra involves anticommutators as well as commutators (it is said to be “graded”).

4 An Overview of Supersymmetry

4.1 The Supersymmetry Algebra and its Representations

In this lecture, we will collect a few facts that will be useful in the subsequent discussion. We won’t attempt a thorough introduction to the subject. This is provided, for example, by Lykken’s lectures[13,14], Wess and Bagger’s text[15], and Appendix B of Polchinski’s text[16].

Supersymmetry, even at the global level, is remarkable, in that the basic algebra involves the translation generators:

$$\{Q^A_\alpha, Q^*_{\dot{\beta}B}\} = 2\sigma^{\mu}_{\alpha\dot{\beta}}\delta^{AB}P_\mu$$

$$\{Q^A_\alpha, Q^*_{\beta B}\} = \epsilon_{\alpha\beta}X^{AB}.$$  \hspace{1cm} (117)

The $X^{AB}$’s are Lorentz scalars, antisymmetric in $A, B$, known as central charges.

If nature is supersymmetric, it is likely that the low energy symmetry is $N = 1$, corresponding to only one possible value for the index $A$ above. Only $N = 1$ supersymmetry has chiral representations. In addition, $N > 1$ supersymmetry, as we will see, is essentially impossible to break; this is not the case for $N = 1$. For $N = 1$, the basic representations of the supersymmetry algebra, on massless fields, are

- Chiral superfields fields: $(\phi, \psi_\alpha)$, a complex fermion and a chiral scalar
- Vector superfields: $(\lambda, A_\mu)$, a chiral fermion and a vector meson, both, in general, in the adjoint representation of the gauge group
The gravity supermultiplet: \((\psi_{\mu,\alpha}, g_{\mu\nu})\), a spin-3/2 particle, the gravitino, and the graviton.

N=1 supersymmetric field theories are conveniently described using superspace. The space consists of bosonic coordinates, \(x^\mu\), and Grassman coordinates, \(\theta_\alpha, \theta^*_\dot{\alpha}\). In the case of global supersymmetry, the description is particularly simple. The supersymmetry generators, classically, can be thought of as operators on functions of \(x^\mu, \theta, \theta^*\):

\[
Q_\alpha = \partial_\alpha - i\sigma^\mu_{\alpha\dot{\alpha}} \theta^* \theta \partial_\mu; \quad \bar{Q}_{\dot{\alpha}} = -\bar{\partial}_{\dot{\alpha}} + i\theta^{*\alpha} \sigma^\mu_{\alpha\dot{\alpha}} \partial_\mu.
\] (119)

A general superfield, \(\Phi(x, \theta, \bar{\theta})\) contains many terms, but can be decomposed into two irreducible representations of the algebra, corresponding to the chiral and vector superfields described above. To understand these, we need to introduce one more set of objects, the covariant derivatives, \(D_\alpha\) and \(\bar{D}_{\dot{\alpha}}\). These are objects which anti-commute with the supersymmetry generators, and thus are useful for writing down invariant expressions. They are given by

\[
D_\alpha = \partial_\alpha + i\sigma^\mu_{\alpha\dot{\alpha}} \theta^* \theta \partial_\mu; \quad \bar{D}_{\dot{\alpha}} = -\bar{\partial}_{\dot{\alpha}} - i\theta^{*\alpha} \sigma^\mu_{\alpha\dot{\alpha}} \partial_\mu.
\] (120)

With this definition, chiral fields are defined by the covariant condition:

\[
\bar{D}_{\dot{\alpha}} \Phi = 0.
\] (121)

Chiral fields are annihilated by the covariant derivative operators. In general, these covariant derivatives anticommute with the supersymmetry operators, \(Q_\alpha\), so the condition 121 is a covariant condition. This is solved by writing

\[
\Phi = \Phi(y) = \phi(y) + \sqrt{2} \theta \psi(y) + \theta^2 F(y).
\] (122)

where

\[
y = x^\mu + i\theta \sigma^\mu \bar{\theta}.
\] (123)

Vector superfields form another irreducible representation of the algebra; they satisfy the condition

\[
V = V^\dagger
\] (124)

Again, it is easy to check that this condition is preserved by supersymmetry transformations. \(V\) can be expanded in a power series in \(\theta^*\)s:

\[
V = i\chi - i\chi^\dagger - \theta \sigma^\mu \theta^* A_\mu + i\theta^2 \bar{\theta} \bar{\lambda} - i\bar{\theta} \theta \lambda + \frac{1}{2} \theta^2 \bar{\theta}^2 \lambda.
\] (125)

In the case of a \(U(1)\) theory, gauge transformations act by

\[
V \rightarrow V + i\Lambda - i\Lambda^\dagger
\] (126)
where $\Lambda$ is a chiral field. So, by a gauge transformation, one can eliminate $\chi$. This gauge choice is known as the Wess-Zumino gauge. This gauge choice breaks supersymmetry, much as choice of Coulomb gauge in electrodynamics breaks Lorentz invariance.

In the case of a $U(1)$ theory, one can define a gauge-invariant field strength,

$$ W_\alpha = -\frac{1}{4} D^2 D_\alpha V. $$  \hfill (127)

In Wess-Zumino gauge, this takes the form

$$ W_\alpha = -i\lambda_\alpha + \theta_\alpha D - \sigma^{\mu\beta}_\alpha F_{\mu\nu} \theta_\beta + \theta^2 \sigma^{\mu}_{\alpha\beta} \partial_\mu \lambda^{*\beta}. $$  \hfill (128)

This construction has a straightforward non-Abelian generalization in superspace, which is described in the references. When we write the lagrangian in terms of component fields below, the non-abelian generalization will be obvious.

### 4.2 N=1 Lagrangians

One can construct invariant lagrangians by noting that integrals over superspace are invariant up to total derivatives:

$$ \delta \int d^4 x \int d^4 \theta h(\Phi, \Phi^\dagger, V) = \int d^4 x d^4 \theta \left( \epsilon_\alpha Q^\alpha + \epsilon_\dagger \alpha Q^{\dagger \alpha} \right) h(\Phi, \Phi^\dagger, V) = 0. $$  \hfill (129)

For chiral fields, integrals over half of superspace are invariant:

$$ \delta \int d^4 x d^2 \theta f(\Phi) = \left( \epsilon_\alpha Q^\alpha + \epsilon_\dagger \alpha Q^{\dagger \alpha} \right) f(\Phi). $$  \hfill (130)

The integrals over the $Q_\alpha$ terms vanish when integrated over $x$ and $d^2 \theta$. The $Q^*$ terms also give zero. To see this, note that $f(\Phi)$ is itself chiral (check), so

$$ Q^*_\alpha f \propto \theta^\alpha \sigma^{\mu}_{\alpha\beta} \partial_\mu f. $$  \hfill (131)

We can then write down the general renormalizable, supersymmetric lagrangian:

$$ \mathcal{L} = \frac{1}{g^{(i)2}} \int d^2 \theta W^{(i)} + \int d^4 \theta \Phi_i^\dagger \epsilon^{\Phi_i} \Phi_i \int d^2 \theta W(\Phi_i) + \text{c.c.} $$  \hfill (132)

The first term on the right hand side is summed over all of the gauge groups, abelian and non-abelian. The second term is summed over all of the chiral fields; again, we have written this for a $U(1)$ theory, where the gauge group acts on the $\Phi_i$’s by

$$ \Phi_i \rightarrow e^{-q_i \Lambda} \Phi_i $$  \hfill (133)

but this has a simple non-abelian generalization. $W(\Phi)$ is a holomorphic function of the $\Phi_i$’s (it is a function of $\Phi_i$, not $\Phi_i^\dagger$).
In terms of component fields, his lagrangian takes the form, in the Wess-Zumino gauge:

$$\mathcal{L} = -\frac{1}{4}g_\alpha^2 F_{\mu\nu}^2 - i\lambda^\alpha \sigma^\mu D_\mu \lambda^{\alpha*} + |D_\mu \phi_i|^2 - i\bar{\psi}_i \sigma^\mu D_\mu \psi_i^* + \frac{1}{2g^2} (D^a)^2 + D^a \sum_i \phi_i^* T^a \phi_i$$

$$+ F_i^* F_i - F_i \frac{\partial W}{\partial \phi_i} + \text{cc} + \sum_{ij} \frac{1}{2} \frac{\partial^2 W}{\partial \phi_i \partial \phi_j} \bar{\psi}_i \psi_j + i\sqrt{2} \sum \lambda^a \bar{\psi}_i T^a \phi_i^*.$$  

(134)

The scalar potential is found by solving for the auxiliary $D$ and $F$ fields:

$$V = |F_i|^2 + \frac{1}{2g_a^2} (D^a)^2$$

(135)

with

$$F_i = \frac{\partial W}{\partial \phi_i^*} \quad D^a = \sum_i (g^a \phi_i^* T^a \phi_i).$$

(136)

In this equation, $V \geq 0$. This fact can be traced back to the supersymmetry algebra. Starting with the equation

$$\{Q_\alpha, Q_\beta\} = 2 P_\mu \sigma^\mu_{\alpha\beta},$$

(137)

Multiplying by $\sigma^\alpha$ and take the trace:

$$Q_\alpha Q_\alpha + Q_\alpha Q_\alpha = E.$$

(138)

If supersymmetry is unbroken, $Q_\alpha |0\rangle = 0$, so the ground state energy vanishes if and only if supersymmetry is unbroken. Alternatively, consider the supersymmetry transformation laws for $\lambda$ and $\psi$. One has, under a supersymmetry transformation with parameter $\epsilon$,

$$\delta \psi = \sqrt{2}\epsilon F + \ldots \quad \delta \lambda = \epsilon D + \ldots$$

(139)

So if either $F$ or $D$ has an expectation value, supersymmetry is broken.

We should stress that these statements apply to global supersymmetry. We will discuss the case of local supersymmetry later, but, as we will see, many of the lessons from the global case extend in a simple way to the case in which the symmetry is a gauge symmetry.

We can now very easily construct a supersymmetric version of the standard model. For each of the gauge fields of the usual standard model, we introduce a vector superfield. For each of the fermions (quarks and leptons) we introduce a chiral superfield with the same gauge quantum numbers. Finally, we need at least two Higgs doublet chiral fields; if we introduce only one, as in the simplest version of the standard model, the resulting theory possesses gauge anomalies and is inconsistent. In other words, the theory is specified by giving the gauge group ($SU(3) \times SU(2) \times U(1)$) and enumerating the chiral fields:

$$Q_f, \bar{u}_f, \bar{d}_f \quad L_f, \bar{e}_f \quad H_U, H_D.$$  

(140)
The gauge invariant kinetic terms, auxiliary \( D \) terms, and gaugino-matter Yukawa couplings are completely specified by the gauge symmetries. The superpotential can be taken to be:

\[
W = H_U(\Gamma_U) + C_{f,f'} Q_f \bar{U}_{f'} + H_D(\Gamma_D) + C_{f,f'} Q_f \bar{D}_{f'} H_D(\Gamma_E) + L_f \bar{e}_{f'}. \tag{141}
\]

As we will discuss shortly, this is not the most general lagrangian consistent with the gauge symmetries. It does yield the desired quark and lepton mass matrices, without other disastrous consequences.

Exercise: Consider the case of one generation. Show that if

\[
\langle H_U \rangle = \langle H_D \rangle = \left( \begin{array}{c} 0 \\ v \end{array} \right), \tag{142}
\]

(all others vanishing), then

\[
\langle D^a \rangle = 0; \langle F^a \rangle = 0. \tag{143}
\]

Study the spectrum of the model. Show that the superpartners of the \( W \) and \( Z \) are degenerate with the corresponding gauge bosons. (Note that for the massive gauge bosons, the multiplet includes an additional scalar). Show that the quarks and leptons gain mass, and are degenerate with their scalar partners.

The fact that the states fall into degenerate multiplets reflects that for this set of ground states (parameterized by \( v \)), supersymmetry is unbroken. That supersymmetry is unbroken follows from the fact that the energy is zero, by our earlier argument. It can also be understood by examining the transformation laws for the fields. For example,

\[
\delta \phi_i = \zeta \alpha [Q^\alpha, \phi_i] = \sqrt{2} i \zeta \psi \tag{144}
\]

but the right hand side has no expectation value. Similarly,

\[
\delta \psi_i = \sqrt{2} \zeta F_i + \sqrt{2} i \sigma^\mu \zeta \partial_\mu \phi. \tag{145}
\]

The last term vanishes by virtue of the homogeneity of the ground state; the first vanishes because \( F_i = 0 \). Similar statements hold for the other possible transformations.

This is our first example of a moduli space. Classically, at least, the energy is zero for any value of \( v \). So we have a one parameter family of ground states. These states are physically inequivalent, since, for example, the mass of the gauge bosons depends on \( v \). We will shortly explain why, in field theory, it is necessary to choose a particular \( v \), and why there are not transitions between states of different \( v \) (in any approximation in which degeneracy holds). As we will see later in these lectures, generically classical moduli spaces are not moduli spaces at the quantum level.

One can also read off from the lagrangian the couplings, not only of ordinary fields, but of their superpartners. For example, there is a Yukawa coupling of the gauginos to fermions and scalars, whose strength is governed by the corresponding gauge couplings.
There are also quartic couplings of the scalars, with gauge strength. These are indicated in fig. 3.

Before turning to the phenomenology of this “Minimal Supersymmetric Standard Model,” (MSSM), it is useful to get some more experience with the properties of supersymmetric theories. With the limited things we know, we can already derive some dramatic results. First, we can write down the most general globally supersymmetric lagrangian, with terms with at most two derivatives, but not restricted by renormalizability (in the rest of this section, the lower case \( \phi \) refers both to the chiral field and its scalar component):

\[
L = \int d^4\theta K(\phi_i^\dagger, \phi_i) + \int d^2\theta W(\phi_i) + cc + \int d^2\theta f(\phi_i)W^2_\alpha + cc. \tag{146}
\]

Here \( K \) is a general function known as the Kahler potential. \( W \) and \( f \) are necessarily holomorphic functions of the chiral fields. One can consider terms involving the covariant derivatives, \( D_\alpha \), but these correspond to terms with more than two derivatives, when written in terms of component fields.

We will often be interested in effective lagrangians of this sort, for example, in studying the low energy limit of string theory. From the holomorphy of \( W \) and \( f \), as well as from the symmetries of the models, one can often derive remarkable results. Consider, for example, the “Wess-Zumino” model, a model with a single chiral field with superpotential

\[
W = \frac{1}{2}m\phi^2 + \frac{1}{3}\lambda\phi^3. \tag{147}
\]

For general \( m \) and \( \lambda \), this model has no continuous global symmetries. If \( m = 0 \), is has an “\( \mathcal{R} \)” symmetry, a symmetry which does not commute with supersymmetry:

\[
\phi \rightarrow e^{\frac{2i\alpha}{3}}\phi \quad \theta \rightarrow e^{i\alpha}\theta \quad d\theta \rightarrow e^{-i\alpha}d\theta. \tag{148}
\]

Under this transformation,

\[
W \rightarrow e^{2i\alpha}W \tag{149}
\]

so \( \int d^2\theta W \) is invariant. This transformation does not commute with supersymmetry; recalling the form of \( Q_\alpha \) in terms of \( \theta \)'s, one sees that

\[
Q_\alpha \sim \frac{\partial}{\partial \theta_\alpha} + \ldots \rightarrow e^{i\alpha}Q_\alpha. \tag{150}
\]

Correspondingly, the fermions and scalars in the multiplet transform differently: the scalar has the same \( \mathcal{R} \) charge as the superfield, \( 2/3 \), while \( \psi \) has \( \mathcal{R} \) charge one unit less
than that of the scalar, i.e.

\[ \phi \rightarrow e^{2i\alpha/3} \phi, \quad \psi \rightarrow e^{-i\alpha/3} \psi. \]  

(151)

It is easy to check that this is a symmetry of the lagrangian, written in terms of the component fields. Correspondingly, in the quantum theory,

\[ Q_\alpha \approx \int d^3x (\sigma^{\mu}_{\alpha\dot{\alpha}} \partial_\mu \phi \bar{\psi}^{*\dot{\alpha}} + \bar{\psi}_\alpha F) \rightarrow e^{i\alpha} Q_\alpha. \]  

(152)

Symmetries of this type will play an important role in much of what follows. In general, in a theory with several chiral fields, one has

\[ \phi_i \rightarrow e^{i\alpha R_i} \phi_i, \quad W(\phi_i) \rightarrow e^{2i\alpha} W(\phi_i). \]  

(153)

If there are vector multiplets in the model, the gauge bosons are neutral under the symmetry, while the gauginos have charge +1. We will also be interested in discrete versions of these symmetries (in which, essentially, the parameter \( \alpha \) takes on only some discrete values). In the case of the Wess-Zumino model, for non-zero \( m \), a discrete subgroup survives for which \( \alpha = 3n\pi \), i.e. \( \phi \rightarrow \phi, \psi \rightarrow -\psi \). More elaborate discrete symmetries will play an important role in our discussions.

Even in the Wess-Zumino model with non-zero \( m \), we can exploit the power of continuous symmetries, by thinking of the couplings as if they were themselves background values for some chiral fields. If we assign \( R \) charge +1 to \( \phi \), we can make the theory \( R \)-invariant if we assign \( R \) charge \(-1\) to \( \lambda \). In practice, we might be interested in theories where \( \lambda \) is the scalar component of a dynamical field (this will often be the case in string theory) or we may simply view this as a trick[17]. In either case, we can immediately see that there are no corrections to the \( \phi^3 \) term in the superpotential in powers of \( \lambda \). The reason is that the superpotential must be a holomorphic function of \( \phi \) and \( \lambda \) (so it cannot involve, say, \( \lambda \lambda^\dagger \)), and it must respect the \( R \) symmetry. Because \( \lambda \) is the small parameter of the theory, we have proven a powerful non-renormalization theorem: the superpotential cannot be corrected to any order in the coupling constant. This non-renormalization theorem was originally derived by detailed consideration of the properties of Feynman graphs. What is crucial to this argument is that \( W \) is a holomorphic function of \( \phi \) and the parameters of the lagrangian.

This is not to say that nothing in the effective action of the theory is corrected from its lowest order value; non-holomorphic quantities are renormalized. For example:

\[ \int d^4 \theta \phi^\dagger \phi f(\lambda^\dagger \lambda) \]  

(154)

is allowed. In the Wess-Zumino model, this means that all of the renormalizations are determined by wave function renormalization. Finally, we should note that if \( m = 0 \) at tree level, no masses are generated for fermions or scalars in loops.
4.3 \(N=2\) Theories: Exact Moduli Spaces

We have already encountered an extensive vacuum degeneracy in the case of the MSSM. Actually, the degeneracy is much larger; there is a multiparameter family of such flat directions involving the squark, slepton and Higgs fields. For the particular example, we saw that classically the possible ground states of the theory are labeled by a quantity \(v\). States with different \(v\) are physically distinct; the masses of particles, for example, depend on \(v\). In non-supersymmetric theories, one doesn’t usually contemplate such degeneracies, and even if one had such a degeneracy, say, at the classical level, one would expect it to be eliminated by quantum effects. We will see that in supersymmetric theories, these flat directions almost always remain flat in perturbation theory; non-perturbatively, they are sometimes lifted, sometimes not. Moreover, such directions are ubiquitous. The space of degenerate ground states of a theory is called the “moduli space.” The fields whose expectation values label these states are called the moduli. In supersymmetric theories, such degeneracies are common, and are often not spoiled by quantum corrections.

In theories with \(N=1\) supersymmetry, detailed analysis is usually required to determine whether the moduli acquire potentials at the quantum level. For theories with more supersymmetries \((N > 1\) in four dimensions; \(N \geq 1\) in five or more dimensions), one can usually show rather easily that the moduli space is exact. Here we consider the case of \(N=2\) supersymmetry in four dimensions. These theories can also be described by a superspace, in this case built from two Grassman spinors, \(\theta\) and \(\tilde{\theta}\). There are two basic types of superfields\[13\], called vector and hyper multiplets. The vectors are chiral with respect to both \(D^\alpha\) and \(\tilde{D}^\alpha\), and have an expansion, in the case of a \(U(1)\) field:

\[
\psi = \phi + \tilde{\theta}^\alpha W_\alpha + \tilde{\theta}^2 \tilde{D}^2 \phi^\dagger, \tag{155}
\]

where \(\phi\) is an \(N=1\) chiral multiplet and \(W_\alpha\) is an \(N=1\) vector multiplet. The fact that \(\phi^\dagger\) appears as the coefficient of the \(\tilde{\theta}^2\) term is related to an additional constraint satisfied by \(\psi\)[13]. This expression can be generalized to non-abelian symmetries; the expression for the highest component of \(\psi\) is then somewhat more complicated[13]; we won’t need this here.

The theory possesses an \(SU(2)\) R symmetry under which \(\theta\) and \(\tilde{\theta}\) form a doublet. Under this symmetry, the scalar component of \(\phi\), and the gauge field, are singlets, while \(\psi\) and \(\lambda\) form a doublet.

I won’t describe the hypermultiplets in detail, except to note that from the perspective of \(N=1\), they consist of two chiral multiplets. The two chiral multiplets transform as a doublet of the \(SU(2)\). The superspace description of these multiplets is more complicated[13,14].

In the case of a non-abelian theory, the vector field, \(\psi^a\), is in the adjoint representation of the gauge group. For these fields, the lagrangian has a very simple expression in
superspace:
\[ L = \int d^2 \theta d^2 \bar{\theta} \psi^\alpha \bar{\psi}^\alpha, \]  
(156)
or, in terms of \( N = 1 \) components,
\[ L = \int d^2 \theta W_\alpha^2 + \int d^4 \theta \phi^\dagger e^V \phi. \]  
(157)
The theory with vector fields alone has a classical moduli space, given by the values of the fields for which the scalar potential vanishes. Here this just means that the \( D \) fields vanish. Written as a matrix,
\[ D = [\phi, \phi^\dagger], \]  
(158)
which vanishes for diagonal \( \phi \), i.e. for
\[ \phi = \frac{a}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \]
(159)
In quantum field theory, one must choose a value of \( a \). This is different than in the case of quantum mechanical systems with a finite number of degrees of freedom; this difference will be explained below. As in the case of the MSSM, the spectrum depends on \( a \). For a given value of \( a \), the massless states consist of a \( U(1) \) gauge boson, two fermions, and a complex scalar (essentially \( a \)), i.e. there is one light vector multiplet. The masses of the states in the massive multiplets depend on \( a \).

For many physically interesting questions, one can focus on the effective theory for the light fields. In the present case, the light field is the vector multiplet, \( \psi \). Roughly,
\[ \psi \approx \psi^\alpha \bar{\psi}^\alpha = a^2 + a \delta \psi^3 + \ldots \]
(160)
What kind of effective action can we write for \( \psi \)? At the level of terms with up to four derivatives, the most general effective lagrangian has the form:\(^3\)
\[ L = \int d^2 \theta d^2 \bar{\theta} f(\psi) + \int d^8 \theta \mathcal{H}(\psi, \psi^\dagger). \]
(161)
Terms with covariant derivatives correspond to terms with more than four derivatives, when written in terms of ordinary component fields.

The first striking result we can read off from this lagrangian, with no knowledge of \( \mathcal{H} \) and \( f \), is that there is no potential for \( \phi \), i.e. the moduli space is exact. This statement is true perturbatively and non-perturbatively!

One can next ask about the function \( f \). This function determines the effective coupling in the low energy theory, and is the object studied by Seiberg and Witten[18]. We won’t review this whole story here, but indicate how symmetries and the holomorphy of

\(^3\)This, and essentially all of the effective actions we will discuss, should be thought of as Wilsonian effective actions, obtained by integrating out heavy fields and high momentum modes.
provide significant constraints (Michael Peskin’s TASI 96 lectures provide a concise introduction to this topic[19]). It is helpful, first, to introduce a background field, \( \tau \), which we will refer to as the “dilaton,” with coupling

\[
\mathcal{L} = \int d^2\theta d^2\bar{\theta} \tau \psi^a \bar{\psi}^a
\]

(162)

where

\[
\tau = \theta + \frac{i}{g^2} + \ldots
\]

(163)

\( \tau \) is a chiral field. For our purposes, \( \tau \) need not be subject to the same constraint as the vector superfield. Classically, the theory has an R-symmetry under which \( \psi^a \) rotates by a phase, \( \psi^a \rightarrow e^{i\alpha} \psi^a \). But this symmetry is anomalous. Similarly, shifts in the real part of \( \tau (\theta) \) are symmetries of perturbation theory. This insures that there is only a one-loop correction to \( f \). This follows, first, from the fact that any perturbative corrections to \( f \) must be \( \tau \)-independent. A term of the form

\[
c \ln(a) \psi^a
\]

(164)

respects the symmetry, since the shift of the logarithm just generates a contribution proportional to \( F \bar{F} \), which vanishes in perturbation theory. Beyond perturbation theory, however, we expect corrections proportional to \( ae^{-\tau} \), since this is invariant under the non-anomalous symmetry. It is these corrections which were worked out by Seiberg and Witten.

### 4.4 A Still Simpler Theory: \( N=4 \) Yang Mills

\( N = 4 \) Yang Mills theory is an interesting theory in its own right: it is finite and conformally invariant. It also plays an important role in Matrix theory, and is central to our understanding of the AdS/CFT correspondence. \( N = 4 \) Yang Mills has sixteen supercharges, and is even more tightly constrained than the \( N = 2 \) theories. There does not exist a convenient superspace formulation for this theory, so we will find it necessary to resort to various tricks. First, we should describe the theory. In the language of \( N = 2 \) supersymmetry, it consists of one vector and one hyper multiplet. In terms of \( N = 1 \) superfields, it contains three chiral superfields, \( \phi_i \), and a vector multiplet. The lagrangian is

\[
\mathcal{L} = \int d^2\theta W^2_\alpha + \int d^4\theta \sum_{i} \phi_i^\dagger e^V \phi_i + \int d^2\theta \phi_i^a \phi_j^b \phi_k^c \epsilon_{ijk} \epsilon^{abc}.
\]

(165)

In the above description, there is a manifest \( SU(3) \times U(1) \) R symmetry. Under this symmetry, the \( \phi_i \)'s have \( U(1)_R \) charge 2/3, and form a triplet of the \( SU(3) \). But the real symmetry is larger – it is \( SU(4) \). Under this symmetry, the four Weyl fermions form a 4, while the six scalars transform in the 6. Thinking of these theories as the low energy limits of toroidal compactifications of the heterotic string will later give us a
heuristic understanding of this $SU(4)$: it reflects the $O(6)$ symmetry of the compactified six dimensions. In string theory, this symmetry is broken by the compactification lattice; this reflects itself in higher derivative, symmetry breaking operators.

In the $N = 4$ theory, there is, again, no modification of the moduli space, perturbatively or non-perturbatively. This can be understood in a variety of ways. We can use the $N = 2$ description of the theory, defining the vector multiplet to contain the $N = 1$ vector and one (arbitrarily chosen) chiral multiplet. Then an identical argument to that given above insures that there is no superpotential for the chiral multiplet alone. The $SU(3)$ symmetry then insures that there is no superpotential for any of the chiral multiplets. Indeed, we can make an argument directly in the language of $N = 1$ supersymmetry. If we try to construct a superpotential for the low energy theory in the flat directions, it must be an $SU(3)$-invariant, holomorphic function of the $\phi_i$'s. But there is no such object.

Similarly, it is easy to see that there no corrections to the gauge couplings. For example, in the $N = 2$ language, we want to ask what sort of function, $f$, is allowed in

$$\mathcal{L} = \int d^2 \theta d^2 \bar{\theta} f(\psi).$$

But the theory has a $U(1)$ R invariance under which

$$\psi \rightarrow e^{2/3i\alpha} \psi \quad \theta \rightarrow e^{i\alpha} \theta \quad \bar{\theta} \rightarrow e^{-i\alpha} \bar{\theta}$$

Already, then

$$\int d^2 \theta d^2 \bar{\theta} \psi \bar{\psi}$$

is the unique structure which respects these symmetries. Now we can introduce a background dilaton field, $\tau$. Classically, the theory is invariant under shifts in the real part of $\tau$, $\tau \rightarrow \tau + \beta$. This insures that there are no perturbative corrections to the gauge couplings. More work is required to show that there are no non-perturbative corrections either.

One can also show that the quantity $\mathcal{H}$ in eqn. [161] is unique in this theory, again using the symmetries. The expression[20,21]:

$$\mathcal{H} = c \ln(\psi) \ln(\psi^\dagger),$$

respects all of the symmetries. At first sight, it might appear to violate scale invariance; given that $\psi$ is dimensionful, one would expect a scale, $\Lambda$, sitting in the logarithm. However, it is easy to see that one integrates over the full superspace, any $\Lambda$-dependence disappears, since $\psi$ is chiral. Similarly, if one considers the $U(1)$ R-transformation, the shift in the lagrangian vanishes after the integration over superspace. To see that this expression is not renormalized, one merely needs to note that any non-trivial $\tau$-dependence spoils these two properties. As a result, in the case of $SU(2)$, the four derivative terms in the lagrangian are not renormalized. Note that this argument is non-perturbative.
4.5 Aside: Choosing a Vacuum

It is natural to ask: why in field theory, in the presence of moduli, does one have to choose a vacuum? In other words, why aren’t their transition between states with different expectation values for the moduli?

\[ \langle + | e^{i\tau \theta / 2} | + \rangle = \cos(\theta / 2). \]  

(170)

If there are \( N \) such sites, the overlap behaves as

\[ \langle \theta | 0 \rangle \sim (\cos(\theta / 2))^N \]  

(171)

i.e. it vanishes exponentially rapidly with the volume.

For a continuum field theory, states with differing values of the order parameter, \( v \), also have no overlap in the infinite volume limit. This is illustrated by the theory of a scalar field with lagrangian:

\[ \mathcal{L} = \frac{1}{2} (\partial_\mu \phi)^2. \]  

(172)

For this system, the expectation value \( \phi = v \) is not fixed. The lagrangian has a symmetry, \( \phi \to \phi + \delta \), for which the charge is just

\[ Q = \int d^3x \Pi(\vec{x}) \]  

(173)

where \( \Pi \) is the canonical momentum. So we want to study

\[ \langle v | 0 \rangle = \langle 0 | e^{iQ} | 0 \rangle. \]  

(174)

We must be careful how we take the infinite volume limit. We will insist that this be done in a smooth fashion, so we will define:

\[ Q = \int d^3x \partial_\phi \phi e^{-x^2/V^2/4} \]  

(175)

\[ = -i \int d^3k \frac{\omega_k}{(2\pi)^3} \left( \frac{V^{1/3}}{\sqrt{\pi}} \right)^3 e^{-k^2V^{2/3}/4} [a(\vec{k}) - a^\dagger(\vec{k})]. \]
Now, one can evaluate the matrix element, using

\[ e^{A+B} = e^A e^B e^{\frac{1}{2}[A,B]} \]

(provided that the commutator is a c-number), giving

\[ \langle 0 | e^{iQ} | 0 \rangle = e^{-c^2 V^{2/3}}, \tag{176} \]

where \( c \) is a numerical constant. So the overlap vanishes with the volume. You can convince yourself that the same holds for matrix elements of local operators. This result does not hold in \( 0 + 1 \) and \( 1 + 1 \) dimensions, because of the severe infrared behavior of theories in low dimensions. This is known to particle physicists as Coleman’s theorem, and to condensed matter theorists as the Mermin-Wagner theorem.

\[
\{Q^A_{\alpha}, Q^*_{\beta}B\} = 2\sigma_\mu^\alpha \gamma^{AB} P_\mu \tag{177}
\]

\[
\{Q^A_{\alpha}, Q^*_{\beta B}\} = \epsilon_{\alpha\beta} X^{AB}. \tag{178}
\]

The \( X^{AB} \)'s are Lorentz scalars, antisymmetric in \( A, B \), known as central charges.

We will focus, principally, on \( N = 1 \) supersymmetry; this means that the index \( A \) above takes only one value.

## 4.6 Supersymmetry and the Hierarchy Problem

In our discussion of the Standard Model, we argued that corrections to scalar masses are quadratically divergent because, unlike fermions, there is no symmetry in the limit in which there mass becomes small. But supersymmetry is a symmetry which relates bosons to fermions, so in a world in which the laws of nature were supersymmetric, scalar masses might be stable against large radiative corrections. This feature of supersymmetry has aroused great interest.

But there is another reason to think that supersymmetry might be important to the hierarchy problem. In fact, stability is not all we want. It is all well and good to say that if a theory has two dimensional parameters, one orders of magnitude smaller than another, that quantum effects will maintain this hierarchy. But supersymmetry is prone to the appearance of such hierarchies.
To get some practice with supersymmetric theories, let’s check that in a simple model, the quadratic divergences do indeed cancel. Take a $U(1)$ theory, with (massless) chiral fields $\phi^+$ and $\phi^-$. Before doing any computation, it is easy to see that provided we work in a way which preserves supersymmetry, there can be no quadratic divergence. In the limit that the mass term vanishes, the theory has a chiral symmetry under which $\phi^+$ and $\phi^-$ rotate by the same phase,

$$\phi^\pm \to e^{i\alpha} \phi^\pm.$$  \hfill (179)

This symmetry forbids a mass term in the superpotential, $\Lambda \phi^+ \phi^-$, the only way a supersymmetric mass term could appear. The actual diagrams we need to compute are shown in fig. ???. Since we are only interested in the mass, we can take the external momentum to be zero. It is convenient to choose Landau gauge for the gauge boson. In this gauge the gauge boson propagator is

$$D_{\mu\nu} = -i(g_{\mu\nu} - \frac{q_\mu q_\nu}{q^2}) \frac{1}{q^2}$$ \hfill (180)

so the first diagram vanishes. The second, third and fourth are straightforward to work out from the basic lagrangian. One finds:

$$I_b = g^2(i)(-i) \frac{3}{(2\pi)^4} \int \frac{d^4k}{k^2}$$ \hfill (181)

$$I_c = g^2(i)(-i) \frac{\sqrt{2}}{2} \frac{1}{(2\pi)^4} \int \frac{d^4k}{k^4} \tr(k_\mu \sigma^\mu k_\nu \sigma^\nu)$$ \hfill (182)

$$= \frac{4g^2}{(2\pi)^4} \int \frac{d^4k}{k^2}$$ \hfill (183)

$$I_d = g^2(i)(-i) \frac{1}{(2\pi)^4} \int \frac{d^4k}{k^2}.$$ \hfill (184)

It is easy to see that the sum, $I_a + I_b + I_c + I_d = 0$.

Clearly if nature is supersymmetric, supersymmetry is broken. We want, then, to ask, what is the likely scale of supersymmetry breaking? We can modify our computation
above so as to address this question. Suppose that supersymmetry breaking induces a mass for the scalars, $\tilde{m}^2$, but the fermions remain massless. Then only $I_d$ changes;

$$I_d \rightarrow \frac{g^2}{(2\pi)^4} \int \frac{d^4k}{k^2 - \tilde{m}^2}$$

$$= -i \frac{g^2}{(2\pi)^4} \int \frac{d^4k_E}{k_E^2 + \tilde{m}^2}$$

$$= \tilde{m}^2 \text{ independent} + \frac{ig^2}{16\pi^2} \tilde{m}^2 \ln(\Lambda^2/\tilde{m}^2).$$

We have worked here in Minkowski space, and I have indicated factors of $i$ to assist the reader in obtaining the correct signs for the diagrams. In the second line, we have performed a Wick rotation. In the third, we have separated off a mass-independent part, since we know that this is cancelled by the other diagrams.

**Exercise:** Verify the expressions for $I_a - I_d$.

Summarizing, the one loop mass shift is

$$\delta \tilde{m}^2 = - \frac{g^2}{16\pi^2} \tilde{m}^2 \ln(\Lambda^2/\tilde{m}^2).$$

Note that the mass shift is proportional to $\tilde{m}^2$, the supersymmetry breaking mass, as we would expect since supersymmetry is restored as $\tilde{m}^2 \rightarrow 0$. In the context of the standard model, we see that the scale of supersymmetry breaking cannot be much larger than the the Higgs mass scale itself. Roughly speaking, it can’t be much larger than this scale by factors of order $1/\sqrt{\alpha_W}$, i.e., factors of order 6.

A second reason to suspect that low energy supersymmetry might have something to do with nature comes from string theory. From the lectures at this school, you have surely gained the sense that supersymmetry is intrinsic to string theory. It is natural to suspect that any consistent fundamental theory must be supersymmetric.

By itself, this is not enough to argue that supersymmetry should survive to low energies. But it has been known for some time that there are a vast array of supersymmetric solutions of string theory. A general feature of these solutions is that if supersymmetry is unbroken in some lowest order approximation, the theory remains supersymmetric to all orders. Moreover, non-supersymmetric solutions are problematic: typically the vacuum is unstable already at one loop. While one cannot claim to have understood how string dynamics might break supersymmetry and choose some particular ground state, it is almost impossible to imagine how a sensible ground state could emerge in a non-supersymmetric vacuum. So if string theory describes nature, low energy supersymmetry is almost certainly a prediction.

Finally, there are some small experimental hints that supersymmetry might be true. The most dramatic of these is the unification of couplings [?]. To understand this, we first
introduce a supersymmetric extension of the standard model, known as the “Minimal
Supersymmetric Standard Model,” or MSSM. In this model, the gauge symmetry is still
taken to be $SU(3) \times SU(2) \times U(1)$. Each gauge generator is now associated with a
vector multiplet, so there is a gaugino for each gauge boson. Similarly, all of the known
quarks and leptons are promoted to chiral multiplets. In other words, each left-handed
fermion of the standard model now has a complex scalar partner with the same quantum
numbers. Finally, instead of one Higgs doublet, there must be two, each containing a
boson and a fermion. Otherwise the model suffers from anomalies. We will denote these
by $H_U$ and $H_D$, with hypercharge $\pm 1$, respectively.

Knowing the representation content of the theory, we can work out the $\beta$-functions
of the different groups. The general expression for the one-loop $\beta$-functions is

$$b_o = \frac{11}{3} C_A - \frac{2}{3} \sum_{i=1}^{n_f} c_i^2 - \frac{1}{3} \sum_{j=1}^{n_a} c_j^2. \quad (187)$$

Here the sum over $i$ runs over all of the left-handed fermions of the theory, while that
over $j$ runs over the scalars. For the gauginos, $c_2 = C_A$, so for a supersymmetric theory
with $n_f$ chiral fields in the fundamental representation we obtain

$$b_o = 3C_A - \frac{1}{2}n_f. \quad (188)$$

For $SU(3)$, with three generations, this gives $b_o = 3$. Similarly, for $SU(2)$, one
obtains $b_o = -1$ (remember to keep track of the two Higgs doublets). For the $U(1)$,
some care is required with the normalization of the charge. Let us assume that the three
gauge groups are unified in $SU(5)$. In that case, all of the generators must be normalized
in the same way. In a singlet generation, one has a $\bar{5}$ and 10. The $\bar{5}$ contains the $\bar{d}$ quark
and the lepton doublet. For this representation, the $SU(3)$ and $SU(2)$ generators satisfy
$\text{tr}T^2 = 1/2$. The corresponding $U(1)$ generator is then

$$\tilde{Y} = \sqrt{\frac{3}{20}} \text{diag}(2/3, 2/3, 2/3, -1, -1). \quad (189)$$

In other words, $\tilde{Y}$ is related to the conventional hypercharge generator by:

$$\tilde{Y} = \sqrt{3/20}Y \quad g' = g_5 \sqrt{3/5} \quad (190)$$

(remember that the hypercharge boson is taken to couple to $Y/2$). From this it follows
that $\sin^2(\theta_w) = 3/8$.

With this information, one can now run the gauge couplings to high energy. The
simplest thing to do is to assume that all of the new particles predicted by supersymmetry
have the same mass (say a few hundred GeV up to a TeV). At one loop, the couplings
satisfy

$$\alpha^{-1}_i(M_Z) = \alpha^{-1}_i(M_{GUT}) + \frac{b_{o(i)}}{4\pi} \ln(M_{GUT}/M_Z). \quad (191)$$
Including also two loop corrections, one obtains quite good agreement. The couplings fail to unify in nonsupersymmetric theories. This might be a coincidence, but it is quite suggestive.\textsuperscript{4}

### 4.11 More on the MSSM

Above, we have introduced the fields of the MSSM, and explained their gauge quantum numbers. Let us develop this model further. The lagrangian includes, first, the gauge invariant kinetic terms for all of the fields. In a supersymmetric theory, this includes Yukawa couplings of gauginos to matter fields and quartic scalar couplings from the $D^2$ terms. The usual Yukawa couplings of fermions arise from terms in the superpotential. These can be written in the form

$$W = H_U Q y_u \bar{u} + H_D Q y_D \bar{d} + H_D L y_L \bar{e}. \quad (192)$$

In this expression, the quark and lepton superfields are understood to carry a flavor index, and the $y$'s are $3 \times 3$ matrices. If $H_U$ and $H_D$ have non-zero masses, this gives mass for quarks and squarks, leptons and sleptons.

**Exercise:** Check that there are a set of supersymmetric ground states with equal, non-zero expectation values for $H_u$ and $H_D$ i.e., that the energy vanishes if $H_U = H_D = \begin{pmatrix} 0 & \cr \cr v \end{pmatrix}$. Check that one obtains equal masses for bosons and fermions, first for the quarks and squarks, leptons and sleptons.

There are other terms which can also be present in the superpotential. These include the “$\mu$-term,” $\mu H_U H_D$. This is a supersymmetric mass term for the Higgs fields. We will see later that we need $\mu \sim M_Z$ to have a viable phenomenology.

A set of dimension four terms which are permitted by the gauge symmetries raise much more serious issues. For example, one can have terms

$$\bar{u}_f \bar{d}_g \bar{d}_h \Gamma^{fgh} + Q_I L_g \bar{d}_H \lambda^{fgh}. \quad (193)$$

These couplings violate $B$ and $L$! This is our first serious setback. In the standard model, there is no such problem. The leading operators permitted by gauge invariance are four fermi operators of dimension 6, and it is easy to imagine that they are suppressed by some very large mass scale.

If we are not going to simply give up, we need to suppress $B$ and $L$ violation at the level of dimension four terms. This presumably requires additional symmetries. There are various possibilities one can imagine.

\textsuperscript{4}Since the weak and electromagnetic couplings are far better known than $\alpha_s$, one usually computes $\alpha_s$, making some assumptions about thresholds both at the GUT scale and at the supersymmetric scale. One actually finds that, with the simplest assumptions for these, that one predicts a value of $\alpha_s$ slightly too large. This problem, and possible solutions, have been discussed in \[?\].
1. Global continuous symmetries: It is hard to see how such symmetries could be preserved in any quantum theory of gravity, and indeed in string theory, there is a theorem which asserts that there are no global continuous symmetries [\[\text{?}\]].

2. Discrete symmetries: Discrete symmetries can be gauge symmetries, and indeed such symmetries are common in string theory. These symmetries are often “R symmetries,” symmetries which do not commute with supersymmetry.

A simple (though not unique), solution to the problem of baryon and lepton number violation by dimension four operators is known as $R$-parity or “matter parity.” Under this symmetry, all ordinary particles are even, while their superpartners are odd. Imposing this symmetry immediately eliminates all of the dangerous operators. For example,

$$\int d^2\theta \bar{u}d \sim \psi_i \psi_d \tilde{d}$$

(we have changed notation again: the tilde here indicates the superpartner of the ordinary field, i.e., the squark). This operator is clearly odd under the symmetry.

More formally, we can define this symmetry as the transformation on superfields:

$$\theta_\alpha \rightarrow -\theta_\alpha \quad \text{(195)}$$

$$\begin{align*}
(Q_f, \bar{u}_f \bar{d}_f, L_f \bar{e}_f) &\rightarrow -(Q_f, \bar{u}_f \bar{d}_f, L_f \bar{e}_f) \quad \text{(196)} \\
(H_U, H_D) &\rightarrow (H_U, H_D). \quad \text{(197)}
\end{align*}$$

While imposing this symmetry solves the immediate problem, it also has a striking consequence: the lightest of the new particles predicted by supersymmetry (the LSP) is stable. In order to avoid various cosmological disasters, this particle must be electrically neutral. It is then, inevitably, very weakly interacting. This in turn means:

- The generic signature of $R$-parity conserving supersymmetric theories is events with missing energy.
- Supersymmetry is likely to produce an interesting dark matter candidate.

In most of what follows, we will assume a conserved $R$-parity.

### 4.12 LSP as the Dark Matter

A stable particle is not necessarily a good dark matter candidate. But we can make a crude calculation which indicates that the LSP density is in a suitable range to be the dark matter. Consider particles, $X$, with mass of order 100 GeV interacting with weak interaction strength. Their annihilation cross sections go as $G_F^2 E^2$. So, in the early universe, the corresponding interaction rate is of order

$$\Gamma \approx \rho_X G_F^2 E^2 \approx \rho_X G_F^2 T^2. \quad \text{(198)}$$
These interactions drop out of equilibrium when

$$\Gamma \sim H \sim T^2/M_p,$$

i.e., when

$$\rho_X \sim \frac{G_F^2}{M_p} \sim 10^{-9}.\quad(200)$$

$T$ here is of order 1 GeV, so

$$\frac{\rho_X}{\rho_\gamma} \sim 10^{-9}.\quad(201)$$

This means that the $X$ particles have a number density today similar to that of baryons. So if their masses are of order 100 GeV, their density can be of order the closure density. This estimate is quite crude, but more careful studies indicate that the LSP can be the dark matter for a broad range of parameters.

So while it is disturbing that we need to impose additional symmetries in order to avoid proton decay, it is also exciting that this leads to a possible solution of one of the most critical problems of cosmology: the identity of the dark matter.

### 4.13 Local Supersymmetry

It would take many lectures (and a more expert lecturer) to give a proper exposition of $N = 1$ supergravity. Fortunately, there are only a few facts we will need to know. First, the terms in the effective action with at most two derivatives or four fermions are completely specified by three functions:

1. The Kahler potential, $K(\phi, \phi^\dagger)$, a function of the chiral fields
2. The superpotential, $W(\phi)$, a holomorphic function of the chiral fields.
3. The gauge coupling functions, $f^a(\phi)$, which are also holomorphic functions of the chiral fields.

The lagrangian which follows from these can be found, for example, in [?, ?]. Let us focus, first, on the scalar potential. This is given by

$$V = e^K \left[ \left( \frac{\partial W}{\partial \phi_i} + \frac{\partial K}{\partial \phi_i} \right) g^{ij} \left( \frac{\partial W}{\partial \phi_j^*} + \frac{\partial K}{\partial \phi_j^*} \right) W - 3|W|^2 \right],\quad(202)$$

where

$$g^{ij} = \frac{\partial^2 K}{\partial \phi_i \partial \phi_j^*}\quad(203)$$

is the Kahler metric associated with the Kahler potential. In this equation, we have adopted units in which

$$G_N = \frac{1}{8\pi M^2}.\quad(204)$$
where \( M \approx 2 \times 10^{18} \) GeV is the reduced Planck mass we encountered earlier.

There is a standard strategy for building supergravity models. One introduces two sets of fields, the “hidden sector fields,” which will be denoted by \( Z_i \), and the “visible sector fields,” denoted \( y_a \). The \( Z_i \)'s are assumed to be connected with supersymmetry breaking, and to have only very small couplings to the ordinary fields, \( y_a \). In other words, one assumes that the superpotential, \( W \), has the form

\[
W = W_z(Z) + W_y(y),
\]

at least up to terms suppressed by \( 1/M \).

One also usually assumes that the Kahler potential has a “minimal” form,

\[
K = \sum z_i^\dagger z_i + \sum y_a^\dagger y_a.
\]

One chooses (tunes) the parameters of \( W_Z \) so that

\[
\langle F_Z \rangle \approx M_w M
\]

and

\[
\langle V \rangle = 0.
\]

Note that this means that

\[
\langle W \rangle \approx M W M^2.
\]

To understand the structure of the low energy theory in such a model, suppose first that \( W(y) = 0 \), and that \( K \) is of the “minimal” form, eq. (206). Then

\[
V(y) = e^{K(z)}|\langle W \rangle|^2 \sum |y_a|^2.
\]

In other words, the scalars all have a common mass,

\[
m_o^2 = e^{\langle K \rangle} \frac{|\langle W \rangle|^2}{M^4} \approx m_{3/2}^2.
\]

Here \( m_{3/2} \) is the mass of the gravitino. Note that with \( F \sim M M_Z, m_{3/2} \sim M_Z \).

If we now allow for a non-trivial \( W_y \), we find also \( A \) and \( B \mu \) terms. For example, the terms

\[
\frac{\partial W}{\partial y_a} y_a W_y = 3W_y
\]

if \( W \) is homogeneous of degree three. Additional contributions arise from

\[
\left\langle \frac{\partial W}{\partial z_i} \right\rangle y_j^* W^* + c.c.
\]

**Exercise:** The simplest model of the hidden sector is known as the “Polonyi model.” In this model,

\[
W = m^2(z + \beta)
\]
\[ \beta = (2 + \sqrt{3}M). \]  
(215)

Verify that the minimum of the potential for \( Z \) lies at
\[ Z = (\sqrt{3} - 1)M \]  
(216)

and that
\[ m_{3/2} = (m^2/m)e^{(\sqrt{3}-1)^2/2} \quad m_\sigma^2 = 2\sqrt{3}m_{3/2} \quad A = (3 - \sqrt{3})m_{3/2}. \]  
(217)

So far, we have not addressed the question of gaugino mass. This can arise from a non-trivial gauge coupling function,
\[ f^a = \frac{cZ}{M} \]  
(218)

which gives a gluino mass, just as it would in the global case:
\[ m_\lambda = \frac{cF_z}{M}. \]  
(219)

So these models have just the correct structure. They have soft breakings of the correct order of magnitude, and they exhibit, with our assumption of minimal kinetic term, the properties of universality and proportionality. Indeed, this is a highly predictive framework, with only (assuming MSSM particle content) 5 parameters. A large amount of work has been done on these models, including investigations of:

1. Renormalization group evolution: one finds that there is a substantial region of the parameter space in which \( SU(2) \times U(1) \) is broken in the correct fashion.

2. Unification: The predictions for unification are far better than in the minimal standard model. However, the prediction for \( \alpha_s \) tends to be somewhat larger than observed. This can be “fixed” by adding new thresholds at the high scale [?].

3. \( b-\tau \) unification: One can also consider the possibility that the \( b \) and \( \tau \) Yukawa’s unify. This is certainly not a general requirement of unification; it depends, in grand unified models, on the Higgs content, and in string theory on the details of compactification. Still, this idea seems viable (and interesting).

4. Proton decay: with detailed assumptions about the susy spectrum and the structure of grand unification, it is possible to compute the proton lifetime and compare with experimental limits.

\[ ^5 \text{More detail on all of these points is provided in many excellent review articles. See, for example, Jon Bagger’s 1995 TASI lectures, and references therein [?].} \]
5. Dark matter: Again, in a detailed model, one can identify the LSP and compute its couplings to matter. This permits a precise calculation of the abundance.

6. $b \rightarrow s\gamma$: As we have already remarked, one can compute the rate for this process in particular models. One often finds that other constraints are stronger.

5  Spontaneous Supersymmetry Breaking

5.1  Supersymmetry Breaking

5.1.1  F and D Type Breaking

6  Non-Renormalization Theorems

Statement of the theorems. Seiberg’s proof. Various applications.

7  Beyond Perturbation Theory

7.1  Supersymmetric QCD and Its Symmetries

7.2  $N_f < N_c$: The Superpotential

7.3  $N_f = N_c - 1$: Instanton computation of the Superpotential

7.4  $N_f < N_c$: Gaugino Condensation

7.5  Supersymmetry Breaking: The $(3, 2)$ model

8  Seiberg Duality
Part II: String/M Theory
9 Introduction

String theory and quantum gravity. A bit of history of string theory.
10  The Bosonic String
11  The Superstring
12  Type II Strings
13  Heterotic String
14  Compactification of String Theory
  14.1  Tori
  14.2  Orbifolds
  14.3  Calabi-Yau compactifications
15  Dynamics of String Theory at Weak Coupling
  15.1  Non-Renormalization Theorems
  15.2  Gaugino Condensation
  15.3  Obstacles to A String Phenomenology
16  Strings at Strong Coupling: Duality
  16.1  Perturbative Dualities
  16.2  D-Branes
  16.3  Strong-Weak Coupling Dualities: Some Evidence
    16.3.1  IIB Self Duality
    16.3.2  Catalog of other string Dualities
  16.4  IIA → 11 Dimensional Supergravity (M Theory)
  16.5  Strongly Coupled Heterotic String
  16.6  AdS-CFT

Very brief. Mainly AdS$_5$. 

54
17  Coda: Where are We headed

17.1  Some alternatives to Supersymmetry

17.1.1  Technicolor

17.1.2  Large Extra Dimensions

17.1.3  Warped Spaces

17.1.4  String Phenomenology: The Landscape?

17.1.5  The Tevatron, The LHC and the ILC

18  Appendices

18.1  Left Handed Fermions

18.2  Gauge Symmetry

18.3  Path Integration

References

[1] See the lectures by S. Carroll and S. Perlmutter at this school.


[55] A model of this type has been discussed by Izawa and Yanagida. this model relies on continuous R symmetries, but is easily modified to work with discrete R symmetries. Yanagida et al on R symmetries and condensation, K. Izawa and T. Yanagida, “R Invariant Dilaton Fixing,” hep-ph/9809366.


