

Introduction

—The Past and Future of Elementary Particle Physics—

Sachio KOMAMIYA and Yoshitaka KUNO

Department of Physics, The University of Tokyo, Tokyo 113-0033
Department of Physics, Osaka University, Osaka 560-0043

(Received May 12, 2007; accepted May 29, 2007; published November 12, 2007)

Near the end of the last century, remarkable progress was made in elementary particle physics both experimentally and theoretically, which led to the development of the Standard Model. However, we consider the model as a realization of a more fundamental theory at low energies, since there are so many unexplained physics in the model. Now, elementary particle physics is entering an exciting period in which a new paradigm of the field will be opened by the new discoveries expected in experiments at high-energy frontier colliders. The Higgs boson and supersymmetry are the main targets in these experiments. In parallel to energy frontier physics, experimental and theoretical studies attack the mystery of quark and lepton flavors. The combined efforts in cosmology and particle physics give synergy effects in understanding the history and the current state of the Universe.

KEYWORDS: the Standard Model, Higgs boson, supersymmetry, grand unification, LHC, ILC, flavor physics

DOI: 10.1143/JPSJ.76.111001

1. A Brief History and Present Status of Elementary Particle Physics

We, particle physicists, are entering an exciting period in which a new paradigm of the field will be opened on the TeV energy scale by the new discoveries expected in experiments at high-energy frontier colliders. Namely, the long-awaited discoveries of the Higgs boson and supersymmetric particles, or some new particles or phenomena alternative to these, have been strongly anticipated in the experiments at the Large Hadron Collider (LHC).1,2) The LHC, the highest-energy proton–proton colliding accelerator at CERN, will start taking data in 2008 at the full center-of-mass energy of 14 TeV. The underlying principles of the new physics are expected to be uncovered by precision measurements at the International Linear Collider (ILC), which is the highest-energy electron–positron ($e^+e^-$) colliding accelerator scheduled to be constructed via international efforts in the 2010s.3,4)

In the beginning of the last century, quantum mechanics and the relativity were established. These became the bases of elementary particle physics. In the 1920s, only the electron, proton, and photon were the elementary particles. In the 1930s and 40s, the energy frontier experiments used cosmic rays. The antiparticle was predicted by Dirac,5) and the positron was discovered by Anderson in the cloud chamber exposed to cosmic rays.6) In 1935, Yukawa predicted existence and the mass of $\pi$-mesons as intermediates for the short-range nuclear force.7) Efforts to discover Yukawa’s $\pi$-mesons was made in cosmic ray experiments, and a new lepton, the muon, was accidentally discovered in 1937. Eventually Yukawa’s pion was discovered in 1947 also in cosmic rays. Strongly interacting particles, mesons and baryons and their antiparticles, are generally called hadrons. New type of hadrons with the strangeness quantum number were also discovered in 1947 in cosmic ray.

After World War II, high-energy accelerators with beam energies over 10 GeV were constructed, and dozens of strongly interacting hadrons were discovered. To classify these hadrons, the quark model was proposed by Gell-Mann and Zweig in 1964.8) There were only three quarks (up, down and strange) in the initial quark model. The mixing of (quark) flavors was proposed by Cabbibo before the quark model was proposed.9) The experimental evidence of the quarks was discovered as constituents of nucleons by electron deep-inelastic scattering with nucleons in the early 1970s. At the time, four types of leptons, the electron, muon, electron–neutrino and muon-neutrino, were known to exist. If the quarks are considered to be elementary particles that are on the same level as leptons, the fourth quark, the charm, must have been missing.

Near the end of the twentieth century, remarkable progress was made in particle physics both experimentally and theoretically, which led to the development of the Standard Model. Rapid progress was triggered by the so-called November Revolution in 1974, which marked the discovery of the $J/\psi$ particle or the charm quark.10) This discovery clarified the direction of particle physics towards the establishment of the Standard Model. The time was ripe for the revolutionary era, since the theoretical framework of the electroweak theory had already been established by then.11) The strong interaction was formulated as Quantum Chromo Dynamics (QCD), which is a non-Abelian gauge theory based on the SU(3) gauge group.12) In these theories interactions are mediated by spin 1 gauge bosons. The mediator of the strong interaction, the gluon, was directly discovered in 1979 at the PETRA $e^+e^-$ collider. The weak bosons, $W^\pm$ and $Z^0$, were discovered in 1982 at the CERN $pp$ collider.

The existence of the third-generation quarks was predicted by Kobayashi and Maskawa in 1973,13) before the discovery of the charm quark. Their prediction is based on the argument that the CP violation observed in neutral kaon decay14) is due to the complex phase in the quark mixing matrix, and three quark generations are needed to have such a phase. The third-generation lepton $\tau$ was discovered in
1975 at the Stanford Linear Accelerator Center (SLAC) just after the discovery of the charm quark.\textsuperscript{15} For the third-generation quarks, the bottom quark was discovered in 1977 in the fixed target experiment at Fermilab.\textsuperscript{16} and finally the top quark was discovered in 1994 at the Tevatron $p\bar{p}$ collider.\textsuperscript{17} The number of the quark and lepton generations was determined to be three from the precise measurements of the cross section on and near the $Z^0$ peak at the LEP and SLC $e^+e^-$ colliders.\textsuperscript{18,19} The origin of the three generations is not yet theoretically understood.

In 1998, neutrino oscillation was discovered in the analysis of atmospheric neutrinos at SuperKamiokande.\textsuperscript{20} This is the first evidence that the neutrinos have tiny but nonzero masses. The neutrino oscillation was also observed in solar neutrinos and reactor neutrinos. The mixing pattern of neutrino flavors is measured to be very different from that of quark flavors. The smallness of neutrino masses is the hint for a new area of physics on the large-energy scale of $10^{11}$–$10^{16}$ GeV.

Recently, CP violation, which have been observed only in the neutral $K$ meson decays, was discovered in the bottom hadron sector at the $B$ factories in KEK and SLAC.\textsuperscript{21}

The Standard Model is made up of three main pillars: (1) quarks and leptons and their antiparticles are the fundamental constituents of matter, (2) their interactions (electromagnetic, weak and strong) are governed by the gauge principle, and (3) the origin of the mechanism for generating the masses of elementary particles is the Higgs boson (see Table I).

The Standard Model has been tested through many experiments with high precision, and its success has become increasingly firm. Thus far, all the quarks and leptons of 12 species in three generations have been discovered. If quarks and leptons were not point like elementary particles and they were made up of much smaller constituents, the effects would be observed at high-energy colliders. At present, the experimental upper limits of the size of quarks and leptons are on the order of $10^{-19}$ m, and they are still considered to be elementary.\textsuperscript{19}

Nevertheless, the Higgs boson has not been discovered yet. Thus, one of the pillars of the Standard Model is still missing and awaits experimental confirmation. In addition, the Standard Model does not include gravity in its framework. Since the common origin of these interactions is not known, their unification is out of the question within the Standard Model. Furthermore, it cannot explain why there are three generations and twelve kinds of quarks and leptons, why they have different masses, and why mixing between different generations occurs. Hence, there are so many unexplained free parameters, that should be determined experimentally in the model. It is obvious that the Standard Model is not the ultimate theory of particle physics; therefore we consider the model as a realization of more fundamental theory at low energies.

### 2. Particle Physics at Highest Energy Frontier

#### 2.1 Electron–positron colliders and hadron colliders

The evolution of elementary particle physics is stimulated by the development of accelerator technologies. The conventional high-energy physics experiments are fixed-target experiments, in which accelerated beams are smashed on to a stationary target. The reaction energy of the collision, which can be used for new particle production, is given by $\sqrt{2EM}$ at high energies, where $M$ is the target particle mass and $E$ is the beam particle energy. The advent of colliding beam accelerators (colliders) brought a drastic increase in reaction energy. For collider experiments the two beams are accelerated in opposite directions and the beams collide with each other. A particle detector surrounds the beam collision point with an almost $4\pi$ solid angle coverage. In this case the reaction energy is $2E$ for the head-on collision of two particles with the same mass $m$ and the same energy $E$, hence the energy efficiency is significantly higher than that for fixed target experiments. The initial colliders were constructed in the late 1960s. After the 1970s, most of the new particles, including $J/\psi$, the $\tau$-lepton, gluons, $W$- and $Z$-bosons and the top quark, were discovered in collider experiments at the highest energy frontier. There are two types of colliders: $e^+e^-$ colliders and hadron colliders (proton–proton or proton–antiproton).

Experiments at $e^+e^-$ colliders surpass those at hadron colliders in terms of simplicity of the processes and cleanness of the experimental environment. An electron and a positron annihilate each other, so that all the reaction energy can be efficiently used for new particle production. The event rate as well as background level are relatively low in $e^+e^-$ collider experiments. Since electrons and positrons do not interact strongly with matter, the experimental environment is clean.

Meanwhile, protons and antiprotons are composite particles made up of three quarks bounded by gluons; hence their interactions are complicated. The fundamental interactions take place as a collision of partons (quarks or gluons), so that only a fraction of the total energy $2E$ can be used for new particle production and the reaction energy varies for each event. Also experiments at proton–proton (proton–antiproton) colliders suffer from a high background rate and a high radiation level. In order to select a few interesting events in the high background, a sophisticated triggering system must work in a short decision time. New particles with distinctive features can only be searched for at hadron colliders.

However, in bending the electron or positron beams in a circular collider with a dipole magnetic field, the beam particles loose their energy by emitting synchrotron radiation. At high energies the energy loss by the synchrotron radiation per one turn in a circular accelerator ($\Delta E$) is given by the formula $\Delta E = (4\pi/3)e^2(E/m)^4/R$, where $E$ is

![Table 1. List of all the elementary particles in the Standard Model. Each quark has three color degree-of-freedoms, as a gluon has eight color combinations.](image)
the beam particle energy, \( m \) is the beam particle mass, and \( R \) is the bending radius of the particle in the magnetic field. The electric power used for the acceleration is proportional to \( (E/m)^{4}/R^{2} \). In any case, emission of synchrotron radiation is a serious drawback for realizing higher-energy \( e^{+}e^{-}/C_{0} \) circular colliders beyond the LEP energies.

There are two remedies for this problem. One of them is to give up \( e^{+}e^{-}/C_{0} \) collision and to use a proton as the beam particle. Protons are about 1800 times heavier than electrons or positrons, so that the energy loss per turn caused by synchrotron radiation can be reduced by a factor of \((m_{e}/m_{p})^{4}\). Since higher energy is essential to creating new heavy particles, it is worth sacrificing the cleanness of the experiment. In fact, the proton collider LHC uses the same tunnel that was previously used for the \( e^{+}e^{-}/C_{0} \) collider LEP. The center-of-mass energy is increased by a factor of 70 from that for LEP. The layout of the LHC accelerator complex is shown in Fig. 1, and the inside of the LHC accelerator tunnel is shown in Fig. 2.

Since \( e^{+}e^{-} \) collisions have so many advantages as described above, we should not simply give it up. Therefore, the other remedy is to set the bending radius \( R \) to infinity, so that no synchrotron radiation can be emitted. The concept of the \( e^{+}e^{-} \) linear collider was proposed along this line. In the linear collider, electron and positron beams are accelerated face-to-face in two separate linear accelerators, and they collide with each other in the experimental hall located at the center of the accelerator system. Circular colliders have an advantage that particles can be accelerated gradually when they circulate and pass the same accelerator cavities. For linear colliders the acceleration is one-shot, which is a serious disadvantage. Therefore, the accelerating gradient in the linear accelerator must be very high, otherwise the length of the collider becomes too long and hence the cost becomes too high. The other technical issue is to focus the beams into a very small size at the collision point in order to achieve large luminosity. Technical development of \( e^{+}e^{-} \) linear colliders began started in the 1980s. Now, the linear collider has become an international project. In 2004, ICF (International Committee for Future Accelerators) decided that the superconducting accelerator cavities be used for the linear collider, and named the project ILC (International Linear Collider). The layout of the ILC accelerator tunnels is shown in Fig. 3.

In 2008, the LHC experiments will start taking data with a center-of-mass energy of 14 TeV and will survey new physics on the TeV scale. The ILC experiments will then perform precision measurements in the clean environment of \( e^{+}e^{-} \) collisions to reveal physics principles behind the new phenomena observed at the LHC. Experiments at the LHC and ILC are expected to establish a new paradigm of particle physics, which will be as important as former breakthroughs such as the discovery of antiparticles or the establishment of the gauge principle. The most important physics expected to be discovered in the near future are given below.

### 2.2 Higgs bosons

All elementary particles in the Standard Model should be massless if gauge symmetries are not broken. The Higgs boson breaks the gauge symmetries and give mass to elementary particles.22) The weak gauge bosons (\( W^{\pm} \) and \( Z^{0} \))

---

Fig. 1. Layout of the LHC accelerator at CERN.

Fig. 2. Inside of the LHC tunnel, formerly used for the LEP \( e^{+}e^{-}/C_{0} \) collider. Superconducting dipole-magnets for the proton beams are installed in the tunnel.

Fig. 3. Layout of the ILC tunnel. In the left tunnel, the RF system such as klystrons and its power supplies are installed. This tunnel can be accessed for maintenance even during machine operation. In the right tunnel, superconducting accelerator cavities are aligned. The beam is accelerated in these cavities.
obtain masses through the Higgs mechanism, which is the direct consequence of the breakdown of the gauge symmetries.\textsuperscript{23} The quarks and leptons obtain masses from the Yukawa coupling to the Higgs boson. Past measurements constrain the mass of the Higgs boson in the Standard Model to be between 114 GeV and about 200 GeV. The upper limit of 114 GeV is from the direct Higgs boson searches at LEP, and the lower limit of about 200 GeV is from precise electroweak measurements including those from the experiments from LEP and Tevatron.\textsuperscript{19,24} Thus the Higgs boson postulated in the Standard Model will be discovered at the LHC.\textsuperscript{2} It will be the discovery of the elementary scalar particle. Elucidation of the properties of the Higgs boson will be the first step in understanding the structure of vacuum, the inflation of the Universe and dark energy, for which some kind of scalar particle may be responsible. The ILC will be a Higgs boson factory and scrutinize the production and decay of the Higgs boson, precisely determine its mass, spin, coupling constants for coupling to other elementary particles, and its self-coupling. Through these measurements, the ILC will verify that gauge symmetry breaking is the origin of elementary particle masses.\textsuperscript{4,3} Furthermore, there is a possibility of finding the direction in which to extend the Standard Model after uncovering the physics behind the electroweak symmetry breaking.

2.3 Supersymmetry

Supersymmetry\textsuperscript{25} is regarded as the most promising paradigm beyond the Standard Model. Supersymmetry is an extension of symmetry in the Poincare group, and is a symmetry between Fermion and Boson fields. If supersymmetry exists, every elementary particle has its supersymmetric partner, whose spin differs by 1/2 unit from that of the corresponding Standard Model particle. The group of supersymmetric partners can be discovered through the experiments in the TeV energy scale. There exist the following pieces of indirect evidence for TeV scale supersymmetry: (1) the Higgs boson mass, which is not protected for divergence by any symmetry, is kept sufficiently light with the advent of supersymmetry, (2) three types of interactions, strong, electromagnetic, and weak interactions, are unified at a very high energy scale of $10^{16}$ GeV with the supersymmetry, (3) the lightest supersymmetric particle is a leading candidate for dark matter. Supersymmetry, if discovered, will give an important clue to superstring theory, which is a candidate for the ultimate unified theory that includes gravity.

Experiments at the LHC are expected to discover evidence for supersymmetry at an early stage in data taking.\textsuperscript{25} If supersymmetric particles are within the reach of the ILC, their properties will be understood from precise measurements of their production and decays.\textsuperscript{4} Through these measurements, we will clearly know if dark matter can be identified as the lightest supersymmetric particle, and this influences the strategy for terrestrial dark matter search experiments. No supersymmetric partners of the Standard Model particles have far been discovered; hence the supersymmetry must be broken. By combining the information from the LHC and ILC experiments on the mass spectrum of the supersymmetric partners, the mechanism of supersymmetry breaking is expected to be studied.

2.4 Something unexpected

According to the present understanding of elementary particle physics, a light Higgs boson should exist. Many researchers also expect that supersymmetry will be discovered at an energy below 1 TeV. However, future experiments may encounter something unexpected. In particular, if we find no Higgs boson but some new particle or some new phenomenon that is responsible for the electroweak symmetry breaking or the mass generation, it will open a new frontier of physics in both theoretical and experimental studies.

One such scenario can be realized in models with extra dimensions. In superstring theories,\textsuperscript{26} there must be 6-dimensional space beyond our 4-dimensional space–time realm. Since effects of the extra dimensions are not observed anywhere, these extra dimensions can be either compactified into a very small size or the Standard Model particles can only live in our space–time realm of 4 dimensions. Some theorists are building models with extra dimensions without the ordinary Higgs boson or without TeV-scale supersymmetry. The effects of the extra dimensions might be seen in the high-energy experiments.\textsuperscript{2,4}

3. Particle Physics at Precision Frontier

Even if the light Higgs boson is discovered and supersymmetry is experimentally proven to be a new framework of physics, there still remain the following fundamental questions: Why do quarks and leptons have three generations and why do they have different masses? Why do they mix among different generations (flavor mixing)? Why do they violate the symmetry between particles and anti-particles (CP violation)? What is the origin of the neutrino masses? Did the neutrinos play an important role in creating the matter-dominant Universe? To address these fundamental questions, an experimental approach at the high-precision frontier can be explored.

3.1 Quark flavor physics

The word of “flavor” implies a quantum number to distinguish different kinds of elementary particles. The 12 different kinds of elementary particles have 12 different flavors. Flavor physics searches for a breakthrough to answer some of the questions mentioned above, in particular on flavor mixing. Here, the flavor mixing causes the transition of one type of quark to another. Parameters of the flavor mixing originate from the Yukawa couplings, which are the strength of the interactions between the Higgs particle and quarks or leptons. The Yukawa couplings in the Standard Model include many unknown parameters that need to be determined experimentally. Precision measurements of the mixing parameters are a necessary step towards understanding of flavor mixing. For quarks,\textsuperscript{27} these parameters are presented in a $3 \times 3$ unitary matrix, which is called the Cabbibo–Kobayashi–Maskawa (CKM) matrix, $V_{\text{CKM}}$.\textsuperscript{9,13} It is given by

\[
\begin{pmatrix}
\bar{d} \\
\bar{s} \\
\bar{b}
\end{pmatrix} =
\begin{pmatrix}
V_{ud} & V_{us} & V_{ub} \\
V_{cd} & V_{cs} & V_{cb} \\
V_{td} & V_{ts} & V_{tb}
\end{pmatrix}
\begin{pmatrix}
d \\
s \\
b
\end{pmatrix},
\]
where \((d', s', b')\) and \((d, s, b)\) are the eigenstates of down-type quark fields in the weak interaction (flavor eigenstate) and the mass eigenstate fields, respectively. It implies that one quark state, when the weak interaction takes effects, can be represented by a mixture of different quarks in their mass eigenstate. As a consequence, the square of the corresponding matrix element, \(|V_{q'q}|^2\), is proportional to the transition probability from one quark \(q\) to another \(q'\).

The CKM matrix, since it is a \(3 \times 3\) unitary matrix, has three angles as well as one imaginary phase, \(\delta_{\text{CKM}}\). Instead of using the three angles and one imaginary phase, the CKM matrix is usually represented in an expansion of \(\lambda = \cos \theta_C\), where \(\theta_C\) is the Cabibbo angle and \(\lambda = 0.22\). This is called the Wolfenstein parameterization, which is shown in the following.

\[
V_{\text{CKM}} = \begin{pmatrix}
1 - \frac{\lambda^2}{2} & \lambda & A\lambda^3(\rho - i\eta) \\
-\lambda & 1 - \frac{\lambda^2}{2} & A\lambda^2 \\
A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \\
\end{pmatrix} + O(\lambda^4),
\]

where \(\lambda, A, \rho,\) and \(\eta\) are parameters. The parameters of \(V_{\text{CKM}},\) in particular, \(\rho\) and \(\eta\), which were not precisely determined before, have been extensively studied and determined experimentally using \(K\) mesons \(^{26}\) and \(B\) mesons. Now, one of the long-standing questions in quark flavor physics is CP violation, which was observed in the \(K^0\) decays in 1964.\(^{14}\) One potential explanation is that this violation is due to an imaginary phase, \(\delta_{\text{CKM}}\) or equivalently \(\eta\), in the CKM matrix. To resolve this question, \(B\) factories, both at KEK in Japan and at SLAC in the US, were built for these studies. The \(B\) factories are electron–positron colliders with high beam currents at a center-of-mass energy of 10.58 GeV. In 2001, large CP violation in the \(B\) meson system was discovered at both KEK and SLAC.\(^{21}\) In addition, the parameters of \(\rho\) and \(\eta\) have been precisely determined. Figure 4 shows experimental constraints on \(\rho\) and \(\eta\) from various \(B\) decays and \(K\) decays. In Fig. 4, good consistencies among various measurements can be seen. It has been proven with good precision that the CKM theory can account for the flavor mixing and CP violation in the quark sector.

Upon the upgrading of the KEK \(B\) factory, researchers will search for deviations from the Standard Model in some key observables of flavor physics in the clean environment of a lepton collider. Deviations from the Standard Model, if discovered, will quantitatively quantify the effect of physics beyond the Standard Model. Furthermore, such discoveries may uncover new physics that cannot be detected at the energy frontier. In particular, new sources of CP violation and new right-handed currents, which may arise from supersymmetry or other physics beyond the Standard Model, can be searched for from precision measurements of the \(b \rightarrow s\) transitions that are not measured precisely at the moment.

### 3.2 Lepton flavor physics

Turning to neutrino physics,\(^{29}\) it was originally assumed in the Standard Model that the neutrinos are massless and therefore no neutrino mixing occurs. However, the discovery of neutrino oscillation phenomena at Super-Kamiokande,\(^{20}\) and later at SNO,\(^{30}\) and KamLAND,\(^{31}\) verified that neutrinos have masses and they are mixed. The neutrino mixing matrix, which is known as the Maki–Nakagawa–Sakata–Pontecorvo (MNSP) matrix,\(^ {32,33}\) is given by

\[
\begin{pmatrix}
\nu_e \\
\nu_\mu \\
\nu_\tau
\end{pmatrix}
= U
\begin{pmatrix}
\nu_1 \\
\nu_2 \\
\nu_3
\end{pmatrix},
\]

where \(\nu_l\) \((l = e, \mu, \tau)\) are the flavor eigenstates of neutrino fields and \(\nu_i\) \((i = 1, 2, 3)\) are the mass eigenstates. Since the MNSP matrix is a \(3 \times 3\) unitary matrix, it can also be presented by three angles, such as \(\theta_{12}, \theta_{23},\) and \(\theta_{13},\) and one imaginary phase \(\delta_{\text{MNSP}}.\) If neutrinos are Majorana-type particles, the neutrino mixing matrix would include two additional imaginary phases (called the Majorana phases). Here, the Majorana-type particle is a particle that is identical to its antiparticle, in contrast to the Dirac-type particle. The two large angles of \(\theta_{12}\) and \(\theta_{23}\) are determined from measurements of the oscillation of neutrinos from the sun (solar neutrinos) and those from the earth’s atmosphere (atmospheric neutrinos), respectively \((\sin^2 \theta_{23} \sim 0.5 \text{ and } \sin^2 \theta_{12} \sim 0.3),\) and the 3rd angle \(\theta_{13},\) which is the smallest, is yet to be determined. The Double Chooz experiment (in France) using reactor neutrinos, as well as the T2K experiment (in Japan) and the NOvA experiment (in the US) using accelerator-based neutrinos, are under preparation to measure the \(\theta_{12}\) angle. The studies of neutrino oscillation not only verified that neutrinos are mixed but also showed that the differences of the neutrino masses are quite small; \(\Delta m^2_{21} = m_2^2 - m_1^2 = 8.0^{+0.6}_{-0.5} \times 10^{-5} \text{ eV}^2\) and \(\Delta m^2_{31} = |m_3^2 - m_2^2| = (2.5 \pm 0.5) \times 10^{-3} \text{ eV}^2\), where \(m_i\) is the mass of neutrino \(\nu_i.\)\(^{34}\) This indicates the existence of physics at a very high energy scale if we assume the seesaw mechanism, which was proposed to explain the smallness of the neutrino masses.\(^ {35,36}\) Also, since the MNSP matrix includes an imaginary phase \(\delta_{\text{MNSP}},\) CP violation in the neutrino oscillation, which is determined from a difference in the oscillation probabilities of neutrinos and antineutrinos, can be expected to be observed, in particular, in the case when
\( \theta_{13} \) is large. It is noted that the information of the two additional CP-violating Majorana phases will not be obtained from the neutrino oscillation.

We have also understood that the flavor mixing of neutrinos shows a different pattern from that of quarks. The determination of the neutrino mixing parameters with as good precision as in the quark mixing parameters would be desirable. By combining the mixing parameters of quarks and leptons, it might be possible to gain deep insight into the flavor structures of quarks and leptons, by for instance scrutinizing the quark–lepton complementarity, \( \theta_{12} + \theta_C \sim \pi/4 \), where \( \theta_C \) is the Cabibbo angle in the quark mixing matrix.\(^{37,38}\)

Furthermore, there are the following important issues in neutrino physics ahead of us:

1. **Determination of whether neutrinos are a Dirac-type or Majorana-type particle**
   Searches for neutrinoless double beta (0\(\nu\)\(\beta\beta\)) decays, which can occur only when the neutrinos are Majorana-type, are carried out extensively for various \(\beta\beta\) decay nuclei. The rates of the 0\(\nu\)\(\beta\beta\) decays are proportional to the square of the effective electron neutrino mass, \(\langle m_\nu \rangle = \sum_i |U_{ei}|^2 m_i \exp(i \sigma_i)\), where \(\sigma_i\) is a Majorana phase of the neutrino \(\nu_i\). The present sensitivity to \(\langle m_\nu \rangle\) is on the order of 0.1 eV and future aimed sensitivities will be on the order of 0.01 eV. If the neutrinoless double beta decays are observed, the Majorana phases, which might be indirectly related to the required CP violation in leptogenesis (mentioned later), would be estimated in conjunction with the neutrino mixing parameters.

2. **Determination of the absolute masses (not the mass difference) of neutrinos**
   The direct measurement of the electron–neutrino (\(\nu_e\)) from tritium decays is planned with a sensitivity of 0.1 eV. Also, the 0\(\nu\)\(\beta\beta\) decays enable us to determine \(\langle m_\nu \rangle\) which can be used to evaluate the absolute neutrino masses.

3. **Determining the neutrino mass hierarchy**
   Two types of neutrino mass spectra can be considered; one is the case in which two neutrinos are light and one neutrino is heavy (normal hierarchy), and the other is the case in which two neutrinos are heavy and one neutrino is light (inverted hierarchy). The neutrino mass hierarchy can be determined by planned long-baseline neutrino oscillation experiments using matter effects.

Turning to charged leptons, in contrast to neutrinos, the transitions from one type to another have never been observed yet. They are lepton-flavor-violating (LFV) processes of charged leptons.\(^{39}\) Examples are \(\mu \to e\gamma\) decay, coherent \(\mu - e\) conversion in nuclei (\(\mu^- + N \to e^- + N\)), \(\tau \to \mu\gamma\) decay, and \(\tau \to e\gamma\) decay. They violate the conservation of the lepton flavor quantum numbers between the initial and final states by one unit. In the Standard Model, the LFV processes of charged leptons are heavily suppressed by small neutrino masses against the W boson mass. If they are discovered, we will be able to develop an entirely new physics of charged-lepton mixing. This will open a new way to obtain indirect evidence for supersymmetry and the seesaw mechanism and grand unification.

In order to ascertain the origin of the flavor structure, it seems likely that our understanding of physics at the energy scale of grand unification will play a key role. In the future, with all of the basic precision measurements mentioned above (for quarks, neutrinos and charged leptons) becoming available, the origin of supersymmetry breaking being understood, and proton decay modes being measured, there is a possibility of making a theoretical breakthrough on the origin of the flavor structure.

### 3.3 Precision measurements of particle properties

The highest-precision frontier in particle physics exists in the measurements of the properties of elementary particles. The objective of this kind of study is to search for any deviations from the expectations of the Standard Model, where potential deviation can be caused by new physics or new unseen heavy particles through quantum effects. Examples are the electric dipole moments (EDMs) of elementary particles,\(^{40}\) such as an electron, a neutron, a deuteron, and diamagnetic atoms, and the magnetic dipole moments (MDMs) such as those of an electron and a muon. EDM is known to violate the invariance of time reversal (T-odd and CP-odd) as well as that of space inversion (P-odd), and it has attracted much attention in terms of potentials of searching for unseen sources of CP violation to explain the baryon asymmetry in the Universe (as discussed later). In the Standard Model, EDM is estimated to be vanishingly small, and therefore, clean searches for nonzero EDMs, without the Standard Model contributions, can be realized. The most sensitive measurement of the EDM search was carried out on the \(^{199}\)Hg diamagnetic atom, whose upper limit of about \(10^{-28}\) e·cm was achieved. Although nonzero EDMs have not been observed, the next-generation experiments are planned. Regarding MDM, the precise measurement of the anomalous magnetic moment \((g - 2)/2\) of the muon at the Brookhaven National Laboratory (BNL) in the US was carried out. A sensitivity of less than 1 ppm has been achieved, and it has revealed a 3.4 \(\sigma\) discrepancy from the value predicted by the Standard Model.\(^{39}\)

The search for decays of protons (the lightest baryon), such as \(p \to e^+\pi^0\) and \(p \to K^+\nu\), which violates the conservation of the baryon numbers (B violation), is one of the most important issues in particle physics, because it would have a potential to reveal hints for the Grand Unification Theory (GUT), which is a theory that unifies the fundamental gauge forces, as well as the two different sectors of quarks and leptons, and the GUT allows the baryon-number-violating interaction. The unification of the fundamental forces is Einstein’s dream, which he was unable to achieve in his late research life. Such studies need many protons, and therefore, water has been used for most cases as a target material and at the same time, as a detector by detecting the Cherenkov lights of the decay products of protons. The Super-Kamiokande is such a water cherenkov detector of large size (50 ktons). However, the proton decays are yet to be detected, and only the upper limit of the proton half life time of \(10^{33}\) years has been obtained.\(^{41}\) Next-generation experiments are under consideration worldwide.

### 4. Particle Physics in the Intersection with Cosmology

Turning to observations in the intersection of particle
physics and cosmology, we now confront the astonishing fact that known particles, which form ordinary matter, accounts for only 4% of the total energy density of the Universe. All the rest consists of dark matter\(^{41}\) (20%) and dark energy\(^{42}\) (76%) which are not identified in terms of particle physics. It is likely that dark matter is made up of weakly interacting particles that were created in the early Universe and survived until now.\(^{43}\) These particles were gathered by the attractive gravitational force and become seed galaxies. Dark energy is carried by vacuum and is responsible for the accelerating expansion of the current Universe. Dark energy is consistent with the cosmological constant that was introduced by Einstein into the fundamental equation of general relativity. Both dark matter and dark energy should be explained in terms of particle physics in the future.

Another important issue in the intersection between particle physics and cosmology is how the matter-dominant Universe was created in the very early Universe.\(^{44}\) At the beginning of the Big Bang, it is thought that equal numbers of matter and antimatter were created. Since then, as the Universe expended and cooled, matter and antimatter annihilated each other. If the amounts of matter and antimatter were kept the same in the early Universe, there would be no matter in the present Universe and we would not exist. There must be some mechanism of producing an excess of matter over antimatter. This mechanism is called “baryogenesis” (implying baryon genesis). If baryogenesis did not occur, we humans would not exist. It is now known that baryogenesis needs Sakharov’s three conditions;\(^{45}\) namely (1) violation of the baryon numbers (B violation), (2) asymmetry between matter and antimatter (CP violation), and (3) interactions out of the thermal equilibrium. The original baryogenesis scenario based on the simple grand unification theory without neutrino masses turned out not to work and needs to be modified because of anomalous baryon-number-violating processes (sphaleron processes) in the Standard Model at high temperature. Instead, baryogenesis from massive neutrinos (leptogenesis\(^{46}\)) has now attracted attention. All these phenomena should be explained by particle physics.

5. **Summary**

In the last century, remarkable progress was made in elementary particle physics both experimentally and theoretically, which led to the establishment of the Standard Model. It is indeed a great triumph of particle physics, achieved jointly by theoretical and experimental physicists. However, we consider the Standard Model as an approximation of more fundamental theory at low energies, since there are so many unexplained issues in the Standard Model.

In this special issue on elementary particle physics, the success of the Standard Model of particle physics is described from various aspects. At the same time, future prospects of finding physics phenomena beyond the Standard Model are also emphasized. There are still many fundamental questions that are yet to be answered and should be addressed by elementary particle physics in the future. Some of the fundamental questions are listed below.

1. Are there new symmetries or new physical laws of the Universe? (supersymmetry)
2. Do all the forces become unified? (grand unification, string theory)?
3. Are there extra dimensions of space? (superstring)
4. Why are there so many kinds of particles? (flavor physics)
5. What are neutrinos telling us? (the origin of neutrino mass, Dirac or Majorana?)
6. What is dark energy? What is dark matter?
7. How did the Universe begin? (Big Bang)
8. What happened to antimatter? (baryogenesis)

By addressing these questions, we, particle physicists, are about to take the critical steps towards a revolutionary understanding of the Universe. We believe that the opportunity for new discoveries about the fundamental nature of the Universe will be enormous in the near future.

---

35. T. Yanagida: in Proc. Workshop on the Unified Theory and Baryon
Number in the Universe, ed. O. Sawada and A. Sugimoto (KEK report 79-18, 1979) p. 95.


Sachio Komamiya was born in Yokohama, Japan in 1952. He obtained his B. Sc (1976), M. Sc (1978), D. Sc (1982) degrees from the University of Tokyo. He was a research associate at Faculty of Science, the University of Tokyo (1979–1982), Wissenschaftlicher Mitarbeiter at Physikalisches Institut, Universität Heidelberg (1982–1986), a staff physicist at SLAC, Stanford University (1986–1990), an associate professor (1990–1995), a professor (1995–1999), at ICEPP, the University of Tokyo. Since 1999 he has been a professor at the University of Tokyo. He has worked on field of experimental elementary particle physics. Specially, he has worked on pioneering searches for supersymmetric particles, Higgs bosons and effects of extradimensions using accelerators at the highest-energy frontier. He also worked on experimental studies on QCD and heavy flavor physics at high energies.

Yoshitaka Kuno was born in Aichi prefecture, Japan in 1955. He obtained his B. Sc (1977), M. Sc (1979), and D. Sc (1984) degrees from the University of Tokyo. He was a research associate (1984–1985) at Faculty of Science, the University of Tokyo, a research associate (1985–1989), a research scientist (1989–1992) at TRIUMF, Canada, and an associate professor (1992–2000) at High Energy Accelerator Research Organization (KEK). Since 2000 he has been a professor at Physics Department, Graduate School of Science, Osaka University. He has worked on experimental elementary particle physics, in particular, in the field of high-precision frontier approach, such as rare kaon decays and rare muon decays.
Frontiers of Elementary Particle Physics, the Standard Model and Beyond

From Yukawa’s Pion to Spontaneous Symmetry Breaking
Yoichiro NAMBU

Enrico Fermi Institute, University of Chicago, Chicago, IL 60637, U.S.A.

(Received February 6, 2007; accepted April 9, 2007; published November 12, 2007)

An overview is given of the development of particle physics since the 1930s to the 1960s when the concept of spontaneous symmetry breaking (SSB) was established as an essential component of particle theory and eventually led to the Standard Model. A brief account of SSB as a general phenomenon in physics is also added.

KEYWORDS: particle physics, standard model, spontaneous symmetry breaking, gauge theory, Nambu–Goldstone mode, Higgs, BCS, Ginzburg–Landau

DOI: 10.1143/JPSJ.76.111002

1. Evolution of Particle Physics

What we now call particle physics started first as nuclear physics in the 1930s, and I have once proposed to call Ernest Lawrence and Hideki Yukawa the two founders of particle physics.1) The first invented the cyclotron, the second “invented” the meson theory, which together opened the door to the exploration of ever higher energy phenomena and to the discovery of particles of ever increasing masses. The problems of particles physics have consisted of three questions: a) What are the elementary constituents of matter? b) What are the interactions, i.e., the forces or fields operating among them, and their basic symmetry properties? c) What is the mathematical formalism to deal with these questions?

The present standard model has answered most of these questions successfully up to the energy scales currently available: it contains the fundamental fermions for a), the gauge fields for b), and the quantum field theory with renormalization and spontaneous symmetry breaking for c), but with the question of the auxiliary field (Higgs field) sector still awaiting verification.

Up until the 1930s, physicists had known only two elementary matter particles, the electron and the proton, plus the photon, the quantum of the only force beside gravity. The nature of atomic nuclei and that of the weak processes belonged to an unknown territory. But they thought they had already found the fundamental constituents of matter. It was a surprise and embarrassment when the neutron was discovered. On the other hand, quantum mechanics had successfully solved the mysteries of atomic phenomena which involved energies of the order of electron volts, but it could not resolve the difficulties of infinite self-energies inherited from classical theory. When dealing with nuclear phenomena involving energies a million times higher than the atomic counterparts, they inclined to believe that quantum mechanics had to be replaced with a new mechanics but did not conceive the possibility of new elementary particles. It was Yukawa2) who showed that one can keep quantum field theory but admit new particles in understanding the problem of nuclear forces. His prediction was confirmed later by the discovery of the pion.3) The difficulty of infinities was circumvented later by the idea of renormalization by Tomonaga, Schwinger, and Feynman.4)

To have a general perspective of what has happened since the 1930s, it is useful to recall the so-called three-stages theory5) by Taketani, who was a disciple and collaborator of Yukawa. He argued that the advance of physics goes through cycles of three stages: Faced with a set of new phenomena, one has first to find some regular patterns in them; then one tries to understand or explain the regularities qualitatively in terms of concrete models; in the last stage a more precise and quantitative theory is developed. But sooner or later new phenomena or discrepancies with the established theory will show up, and one starts the three stages over again. In actuality, these stages do not necessarily proceed in sequence. The three kinds of attempts may go on simultaneously. The path may take a zigzag pattern.

2. Search for Symmetries

Yukawa’s theory originally envisioned only a unique vector charged boson, but soon they found it necessary to examine various spin and neutral versions, simultaneously developing theoretical formulations as they went along. The apparent equality of forces among protons and neutrons led Kemmer6) to introduce the concept of isospin symmetry. But problems related to the divergent properties of derived nuclear forces made clearcut conclusions difficult. There was a confusion when the cosmic ray muon was discovered7) and incorrectly but understandably identified with Yukawa’s meson. As more and more new particles, now called hadrons, showed up in cosmic ray events as well as in reactions in the more powerful accelerators, there followed attempts to make sense of the rich spectrum of particles. In this period they started to search for symmetry properties of reactions even though they were only approximate symmetries. The isospin symmetry for proton and neutron was extended to include the strange hadrons,8) then to the flavor SU(3) symmetry.9) In parallel there also were attempts to understand these regularities in terms of constituent particles: the nucleon pair model of the pion,10) generalized to include Lambda baryons,11) and finally the logically simpler quark model12) which introduced a new set of fundamental constituents below the
level of hadrons. But the problem of statistics, when the hadrons are regarded as composites of quarks, led further to the concept of quarks endowed with an SU(3) of color.\textsuperscript{13} The origin of strong interaction were then to be transferred to that among the quarks, and the hadrons would become dynamical composites of them.

3. Search for Dynamical Principles

The problems c) of the mathematical formalism, may be divided into two different kinds:

1) Search for general and rigorous properties of local quantum field theory independent of the particular Lagrangian or perturbation theory, which led to results like the spin-statistics and CPT theorems. There was also a period of groping for alternatives to quantum field theory which persisted even after the successes of renormalization. These attempts include Heisenberg’s S-matrix theory, its descendants in the form of dispersion theory and the Regge trajectory theory. It is important to emphasize that all these efforts have played an important, if not direct, role in arriving at the standard model.

2) Establishment of the gauge principle for the fundamental interactions among particles. Yukawa’s meson theory was a phenomenological one from today’s viewpoint. The non-Abelian gauge theory of Yang and Mills\textsuperscript{14} was a purely theoretical construct inspired by the isospin symmetry. The crucial problem for practical applications was its masslessness. Another was that the flavor symmetries are only approximate whereas gauge symmetries are exact. As for the weak interaction, its true origins were only a subject of speculations, but when its V-A nature was established,\textsuperscript{15} there arose the possibility that it might also be attributed to a gauge field. As it has turned out, the gauge principle was then not to be applied to the flavor symmetries of the hadrons, but to the strong (color), electromagnetic and weak interactions at the level of quarks and leptons.

Regarding the problem of mass, actually there was the long known plasma mode, i.e., the massive Coulomb field, as well as its generalization to the transverse counterparts\textsuperscript{16} But these happened in ionized media, not in the relativistic “vacuum”. A crucial hint to the understanding and resolution of these problems came from the theories of superconductivity. The BCS theory\textsuperscript{17} assumed a condensate of charged pairs of electrons or holes, hence the medium was not gauge invariant. There were found intrinsically massless collective excitations of pairs (NG modes) that restored broken symmetries, and they turned into the plasmons by mixing with the Coulomb field. The Fermi sea of electrons in the BCS theory developed fermionic excitations with an energy gap, reminding one of the mass gap of the nucleons (and the quarks and leptons). Transporting these results to relativistic theories required one to abandon the concept of the relativistic “void” as Dirac once did in the interpretation of his equation. The spontaneous symmetry breaking thus emerged as a universal phenomenon in physics, and was first applied to the chiral symmetry breaking\textsuperscript{18} and (the constituent) mass generation for nucleons. Yukawa’s pion plays the role of the NG boson. The fact that the chiral symmetry is only approximate, and not gauged, makes the pion massive for reasons different from the case of the weak bosons.

An alternative and in fact older description of superconductivity was one by Ginzburg and Landau,\textsuperscript{19} which turned out to be a phenomenological (effective) representation of the BCS theory. In this transcription the complex scalar “Ginzburg–Landau” field is equivalent to the Higgs field, representing a bound pair of electrons or holes, and its phase and amplitude components correspond respectively to the massless NG and the massive Higgs type excitations. As it happens, the collective excitations of both the NG and Higgs types do exist in all phenomena of the superfluidity type. The precise transcription and utilization (or sometimes independent derivations) of these condensed matter examples to relativistic theories were carried out in due course by a number of people.\textsuperscript{20} The essential point of the problem may be explained as follows. In London’s phenomenological description, the static induced current \( j_i \) (if \( j_i = 0 \) in a superconductive medium is of the form (in arbitrary static gauge)

\[
j_i = K_\alpha A^k_i,
\]

\[
K_\alpha(q) = \left( \delta_{ik} - \frac{q_i q_k}{q^2} \right) K(q^2).
\]

The Meissner effect implies that \( K(0) \neq 0 \). There are two possibilities as to the origin of the second term: Either it is of dynamic origin so that in the nonstatic case \( q^2 \) would be replaced by \( q^2 - v^2 \alpha^2 \), implying some acoustic-type (massless) collective excitations, or else the pole remains static. But the latter could not happen unless the \( 1/q^2 \) singularity was built into the Hamiltonian from the beginning. In fact it is the Coulomb interaction (of the electron component, the compensating background charge remaining inert) that actually produces the pole In a relativistically invariant medium, the above eq. (1) will be replaced by its relativistic version:

\[
\begin{align*}
\mathbf{j}_\mu &= K_\mu j^\mu, \\
K_\mu(q) &= \left( \delta_{\mu
u} - \frac{q_\mu q_\nu}{q^2} \right) K(q^2).
\end{align*}
\]

The two alternatives still exist. It is due to the lack of manifest Lorentz invariance of gauge potentials that the second alternative escapes the NG massless boson theorem, and \( K(q) \) takes the form \( K = q^2/(q^2 - m_V^2), m_V \) being the gauge boson mass. It is now history that this mechanism of mass generation of gauge bosons was most successfully applied to the weak interaction sector, rather than the strong interaction sector, in a form of Ginzburg–Landau–Higgs (G–L–H) description of spontaneous symmetry breaking (SSB) in the electroweak unification of Glashow, Salam, and Weinberg (GSW) for the hypothetical \( W \) and \( Z \) vector bosons, which were to be confirmed experimentally later. The gauge theory of strong interactions sector also joined the GSW theory later to complete the full Standard Model. It contains two SSB phenomena: the gauged one in the weak interaction sector explicitly displayed in the G–L–H form to make the weak bosons massive, whereas the other ungauged chiral symmetry in the strong interaction sector for light quarks is only implicit and approximate.
4. General Remarks on SSB

The general properties of SSB may be characterized as follows.

1) Degeneracy of the ground state
2) The degrees of freedom $N \to \infty$ (the thermodynamic limit). Finite systems do also exhibit similar dynamical properties, but the SSB description becomes an exact one only in infinite systems.
3) Superselection rule (contraction of the Hilbert space) This means that the Hilbert space of the system is built up from one of the ground states, and other Hilbert spaces built on other ground states become inaccessible because there are no local observables that can connect them.
4) The Nambu–Goldstone (NG) modes

The lost symmetry is nevertheless visible in the form of the NG modes, which are massless since in the long wave limit it reduces to the global symmetry operation.

The conscious use of the symmetry principle in physics dates back to the 19th century. Curie\textsuperscript{21)} used symmetry considerations to derive a kind of selection rules for physical effects, an example of which is the Wiedemann effect: A conducting cylinder suffers a twist $t$ when a current $J$ is passed and simultaneously a magnetic field $B$ is applied parallel to it. The behavior of $t$ under various space reflection operations matches that of either $B$ or $J$. So he argued that an effect is possible only if its symmetries $S_{\text{eff}}$ are compatible with those of the environment $S_{\text{env}}$. The SSB, on the other hand, may be characterized by a self-diminution of $S_{\text{env}}$ from $S_L$, the symmetries of the Lagrangian $L$ of the system. It was not envisioned by Curie, but such phenomena were already known in classical systems. The spontaneous deformation of a rotating barytropic body from a sphere to a Jacobi ellipsoid is an example, which clearly shows that symmetry breaking is a dynamical problem. More relevant examples for us, however, came after Curie. The ferromagnetism is the prototype of today’s SSB, as was explained by the works of Weiss,\textsuperscript{22)} Heisenberg,\textsuperscript{23)} and others. Ferromagnetism have since served us as a standard mathematical model of SSB. It is no coincidence that Heisenberg made use of it later in his attempt at a unified theory. Examples of BCS-type SSB are superconductivity and the superfluidity of He\textsuperscript{3}. They have quasifermions, NG and Higgs bosons satisfying simple mass relations among them.\textsuperscript{24)}

Acknowledgement

This work was supported by the University of Chicago.
Experimental Electroweak Physics at Lepton Colliders

Tatsuo KAWAMOTO
ICEPP, the University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033
(Received April 29, 2007; accepted July 3, 2007; published November 12, 2007)

Progress in the electroweak physics at the e⁺e⁻ colliders are reviewed, focusing mainly on the latest and highest energy colliders LEP and SLC. The results on the Z and W boson properties are discussed, and the implications of these precision results in the standard model are shown.

KEYWORDS: electroweak, electron–positron collider, LEP, SLC
DOI: 10.1143/JPSJ.76.111003

1. Introduction

It would be fair to say that electron–positron (e⁺e⁻) colliders have made vital contributions to the progress on particle physics. The spectacular success in the 1970s includes the discovery at SPEAR of c̄c resonances (J/ψ, Ψ′, . . .), charmed mesons and τ leptons, the study of γ resonances at DORIS and CESR, and discovery of B mesons, the observation of quark and gluon jets at PEP and PETRA, followed by TRISTAN in 1980s towards higher collision energies. A great advantage of e⁺e⁻ colliders is the well defined initial state of the point-like electrons, leading to very clean final states, allowing both precise measurements and clear signatures of new phenomena. LEP¹ and SLC² are on the wave front (Fig. 1), conceived as the machines to explore physics at the electroweak scale, by producing a large number of Z bosons, and at a later stage W pairs at LEP-2. This report gives a brief summary of electroweak physics at e⁺e⁻ colliders, focusing mainly on the results from LEP and SLC.

2. Colliders and Experiments

The SLC collider at SLAC was based on the existing 3 km linear electron accelerator originally built in the 1960s. It was upgraded and modified in the 1980s to accelerate both electrons and positrons to the beam energies sufficient to produce the Z boson. The beams were brought into collision at a single interaction point after guided along two arcs. In June 1989 the first e⁺e⁻ collisions at the Z energy were recorded by the Mark II detector. By August 1989, the first results of the Z resonance parameters were produced.³

Since 1992, the new SLD detector⁴ took data at the SLC. The machine was also upgraded to provide longitudinally polarised electron beam,⁵ the feature unique to the SLC which is absent in LEP. By 1998 when the SLC was finally stopped, SLD detector collected 0.6 million Z decay events. For a large fraction of the dataset, the level of polarisation was over 70%.

The LEP collider was built at CERN in the 1980s. It was a circular accelerator of its 27 km circumference, the largest accelerator in the world. The large size was dictated by the requirement of limiting the energy loss due to synchrotron radiation, which goes as Ebeam², to a manageable level. Existing CERN accelerators such as PS and SPS were used to ramp the energy of electrons and positrons to the LEP injection energy at 20–22 GeV. The e⁺e⁻ beams were further accelerated in the LEP ring and brought into collisions at four interaction points where the four LEP detectors ALEPH,⁶ DELPHI,⁷ L3,⁸ and OPAL,⁹ were located.

LEP started its operation in 1989 at the centre-of-mass energy near to the Z mass. Until the end of 1995, four LEP experiments each collected about 4.5 million Z decay events.

In the second phase of LEP operation (LEP-2) from 1996 until 2000, centre-of-mass energy was increased progressively. Pair production of the W boson became possible for the first time in the e⁺e⁻ collision, allowing precise measurements of W boson mass and detailed studies of W pair production mechanism. Another important topic throughout the LEP programme was searches for the Higgs boson and other possible new physics signatures which could appear only at high energies. In the last year of LEP operation, the centre-of-mass energy of LEP reached to its highest energy of 209 GeV.

3. The Standard Model in Brief

At the tree level in the standard model of electroweak interaction, properties of Z and W bosons are determined practically by three independent parameters corresponding to the two gauge couplings of SU(2)L × U(1)Y and the vacuum expectation value of the Higgs field. A basic parameter set suited for the Z studies is Gf, mZ and α, since they are measured precisely. Other parameters like W mass
where $m_W$, the electroweak mixing angle $\sin^2 \theta_W$ and the $Z$ couplings to fermions are all derived using the standard model relations. For example, $m_W$ is related to Fermi constant $G_F$ by:

$$G_F = \frac{\pi \alpha}{\sqrt{2} m_w^2 \sin^2 \theta_W}, \tag{1}$$

and

$$m_w^2 = \frac{\rho m_H^2 \cos^2 \theta_W,}{(2)}$$

with $\rho = 1$ in the minimal model with one Higgs doublet. The $Z$ couplings to left- and right-handed fermions are:

$$g_L = \sqrt{2} (T_3 - Q \sin^2 \theta_W),$$

$$g_R = -\sqrt{2} \rho Q \sin^2 \theta_W. \tag{3}$$

Here $T_3$ is the third component of the weak iso-spin, and $Q$ is the fermion charge. Alternatively vector and axial-vector couplings are defined as:

$$g_V = g_L + g_R, \quad g_A = g_L - g_R. \tag{4}$$

These simple relations are modified by radiative corrections, which depend, among others, on the top mass $m_t$ and the Higgs mass $m_H$. The parameters $G_F, m_Z,$ and $m_w$ can be defined as physically measured quantities and eq. (2) holds with $\rho = 1$, introducing a corrected $\sin^2 \theta_W$. The electromagnetic coupling $\alpha$ is replaced by the running value $\alpha(m_Z)$.

Equation (1) is modified accordingly. For the $Z$ resonance in $e^+e^-$ collisions, effects on parameters $\rho$ and $\sin^2 \theta_W$ in eq. (3) can be largely absorbed by defining corresponding effective complex quantities.\(^{(1,11)}\) It is possible to split these corrected quantities into two parts; real and constant parameters (the effective parameters) and in general very small complex remnants. Furthermore, the constant effective parameters can be treated in a very good approximation\(^{(10)}\) as if they are independent parameters, allowing a (nearly) “model independent” parametrisation with the small remnants calculated using the standard model.

The basic strategy of LEP and SLC electroweak measurements is to summarise the measured observables in terms of these “model independent” pseudo-observables, and compare them to the standard model predictions to test the model and determine or constrain the standard model parameters such as $m_t$ and $m_H$.

### 4. Z Parameters

The process $e^+e^- \rightarrow \bar{f}f$ is mediated in the $s$-channel by two spin-1 neutral gauge bosons, a massless photon and a massive Z boson. The cross-section can be parametrised by a Breit–Wigner resonance for the $Z$ contribution and remaining contributions from photon exchange and $\gamma$–$Z$ interference:

$$\sigma(s) = \sigma_0^Z \frac{\lambda^2}{(s - m_Z^2)^2 + \lambda^2} + \sigma(\gamma) + \sigma(\gamma - Z), \tag{5}$$

where $s$ is the centre-of-mass energy squared, $m_Z$ and $\Gamma_Z$ are mass and total width of the Z boson, and $\sigma_0^Z$ is the peak cross-section of the resonance. Here the $s$-dependent width is used in the definition. Measurement of the resonance curve, the $Z$ lineshape, allows precise determination of the three parameters $m_Z, \Gamma_Z$, and $\sigma_0^Z$. The peak cross-section is related to the partial widths for initial and final state fermions, $\Gamma_e$ and $\Gamma_f$, with

$$\sigma_0^Z = \frac{12\pi \Gamma_e \Gamma_f}{m_Z^2 \Gamma_Z}. \tag{6}$$

By combining lineshape measurements for all visible fermion final states $f = e, \mu, \tau$, and hadrons (q), partial widths $\Gamma_f$ can be extracted.

Due to the parity violating couplings of the $Z$ to fermions, various asymmetries arise in the process $e^+e^- \rightarrow Z \rightarrow f\bar{f}$. This can be seen by writing down the cross-sections for different helicity combinations $\sigma_{ij} (i = L, R, j = L, R)$:

$$\sigma_{LL}(\cos \theta) = C(s) \cdot g_{LL}^2 G_{LL}^2 (1 + \cos \theta)^2 \tag{7}$$

$$\sigma_{LR}(\cos \theta) = C(s) \cdot g_{LL}^2 g_{RR}^2 (1 - \cos \theta)^2 \tag{8}$$

$$\sigma_{RL}(\cos \theta) = C(s) \cdot g_{RR}^2 g_{LL}^2 (1 - \cos \theta)^2 \tag{9}$$

$$\sigma_{RR}(\cos \theta) = C(s) \cdot g_{RR}^2 g_{RR}^2 (1 + \cos \theta)^2 \tag{10}$$

where $C(s)$ is the $s$-dependent common coefficient, and $i$ indicates left(L)- or right(R)-handed initial-state electron and $j$ for L or R final-state fermion $f$. Left- and right-handed couplings are each denoted by $g_{LL}$ and $g_{RL}$.

Angular distributions for unpolarised beams, ignoring the final-state helicity

Cross-section is obtained by averaging over initial-state helicities and summing over final-state helicities.

$$\sigma(\cos \theta) = \frac{1}{2} (\sigma_{LL} + \sigma_{LR} + \sigma_{RL} + \sigma_{RR})$$

$$\propto (g_{LL}^2 + g_{RR}^2) (g_{LL}^2 + g_{RR}^2)$$

$$\times (1 + \cos^2 \theta + 2A_e A_e \cos \theta), \tag{11}$$

where the asymmetry parameter $A_e$ is defined by:

$$A_e = \frac{g_{LL} g_{LL} - g_{RR} g_{RR}}{g_{LL}^2 + g_{RR}^2}. \tag{12}$$

Forward–backward asymmetry $A_{FB} = (\sigma_F - \sigma_B)/\sigma_F + \sigma_B$ arises from the term proportional to $\cos \theta$, and is given by:

$$A_{FB} = \frac{3}{4} A_e A_e. \tag{13}$$

### 4.1. Polarisation of final-state fermion, unpolarised beams

In the $t$-pair final-state, polarisation of taus can be measured based on the $V$–A charged current decays of taus. The tau polarisation $P_T(\cos \theta) = (\sigma_R - \sigma_L)/\sigma_R + \sigma_L$ is given by:

$$P_T(\cos \theta) = \frac{(\sigma_{LL} + \sigma_{RR}) - (\sigma_{LR} + \sigma_{RL})}{(\sigma_{LL} + \sigma_{RR}) + (\sigma_{LR} + \sigma_{RL})}$$

$$= -A_e (1 + \cos^2 \theta + 2A_e \cos \theta) \tag{14}$$

The forward–backward symmetric term is proportional to $A_e$, while the anti-symmetric term is proportional to $A_e$. From the measurement of $\tau$-polarisation, $A_e$ and $\Gamma_e$ are extracted simultaneously and almost independently.

### 4.2. Polarised beam, cross-section asymmetries

Measured by the SLD experiment at SLC using the polarised electron beam. Left–right cross-section asymmetry directly determines $A_e$.

$$A_{LR} = \frac{(\sigma_{LL} + \sigma_{LR}) - (\sigma_{RL} + \sigma_{RR})}{(\sigma_{LL} + \sigma_{LR}) + (\sigma_{RL} + \sigma_{RR})} = A_e. \tag{15}$$

Here $\sigma_{ij}$ is the cross-section integrated over $\cos \theta$ within forward–backward symmetric acceptance. Alternatively,
Forward–backward asymmetry of the left–right asymmetry, $A_{\text{FBLR}}$, gives the asymmetry parameter of the final-state fermion:

$$A_{\text{FBLR}} = \frac{3}{4} A_f$$

Forward–backward asymmetries are measured at LEP for charged leptons $e$, $\mu$ and $\tau$, and for $b$ and $c$ quarks. Using the measurements of $A_e$ (from electron $A_{\text{FB}}$ and $\tau$ polarisation), the parameters $A_{\mu}$, $A_{\tau}$, $A_b$, and $A_c$ can be extracted using eq. (13). At SLC $A_{\text{LR}}$ and $A_{\text{FBLR}}$ measurements determined the parameters $A_e$, $A_{\mu}$, $A_c$, and $A_{\tau}$. The LEP and SLC results form a complementary and practically complete set of the $A_f$ measurements.

If the $Z$ couplings follow the standard model structure, the ratio of vector and axial-vector couplings can be represented by the effective mixing angle $\sin^2 \theta_{\text{eff}}$

$$\frac{g_{Vf}}{g_{Af}} = 1 - \frac{2Q}{F_3} \sin^2 \theta_{\text{eff}}.$$  

Figure 2 shows the dependence of $A_f$, which is a function of $g_{Vf}/g_{Af}$, on $\sin^2 \theta_{\text{eff}}$ for leptons, $b$ and $c$ quark. For the actual value of $\sin^2 \theta_{\text{eff}} \sim 0.23$, the leptonic $A_f$ is the most sensitive to the variation of $\sin^2 \theta_{\text{eff}}$, while $A_b$ is very insensitive. Therefore $\sin^2 \theta_{\text{eff}}$ is best determined from measurements of the leptonic asymmetry parameter $A_f$.

### 4.1 Lineshape and leptonic $A_{\text{FB}}$

The $Z$ mass $m_Z$, the total width $\Gamma_Z$, and pole cross-section $\sigma_0^l$ are determined from the lineshape measurement (Fig. 3). How well these parameters be determined depends on how accurately the “lineshape” is determined. Since the hadronic final-state is the dominant decay mode of the $Z$ ($\sim 87\%$ of visible decays), $m_Z$ and $\Gamma_Z$ are mainly determined from the hadronic lineshape. The data sample from 1993–1995 is particularly important. In this running period, precision energy scans have been performed at three centre-of-mass energy points on the peak and two off-peak points approximately 1.8 GeV above and below the peak. About 36 pb$^{-1}$ of integrated luminosity were collected by each of the LEP experiments at the off-peak points. In 1994 all data were collected on peak and about 55 pb$^{-1}$ were collected. Measurements of $\sigma_0^l$ and the pole asymmetry $A_{\text{FB}}$ profit from the high statistics measurements at the peak.

Systematic uncertainties must also be under control in order to make full use of the statistical precision. Key issues are:

- Event selection efficiency and background.
- Determination of luminosity.
- Precise calibration of LEP beam energy.
- Precision calculation of radiative corrections.

All of these were studied carefully. Systematic uncertainties on the cross-section and asymmetry measurements were controlled to better than 0.1%. Luminosity determination was based on the measurement of the rate of small angle Bhabha scattering ($e^+e^- \rightarrow e^+e^-$). The experimental uncertainty was well below 0.1% while the common theory uncertainty was 0.06%. Successful calibration of LEP energy achieved a relative accuracy of $10^{-5}$ level.$^{10}$

Uncertainties on $m_Z$ and $\Gamma_Z$ arising from LEP energy uncertainty are $\pm 1.7$ and $\pm 1.2$ MeV, respectively. The cross-section formula eq. (5) is largely modified due to radiative corrections, mainly by initial-state radiation. In order to interpret the precisely measured quantities, quality of theoretical calculations must match the experimental precision. Achieved theoretical precision of the calculations as implemented in the programs ZFITTER$^{11}$ and TOPAZ0,$^{12}$ is typically to the $10^{-2}$ level.$^{13}$

Using these data, lineshape and lepton $A_{\text{FB}}$ are analysed by fitting theoretical parameterisations to the data. The standard set of the parameters are:

- $m_Z$ and $\Gamma_Z$, defined in eq. (5).
- The hadronic pole cross-section

$$\sigma_0^{\text{had}} = \frac{12\pi \Gamma_e \Gamma_{\text{had}}}{m_Z^2 \Gamma_Z^2}$$

- Partial width ratios:

$$R_\ell = \frac{\Gamma_{\text{had}}}{\Gamma_\ell}, \; \ell = e, \mu, \tau$$
Leptonic pole asymmetries.

\[ A_{FB}^{0\ell} = \frac{3}{4} A_c A_\ell \]  

(20)

Results from the four LEP experiments are combined\(^{1,4,15}\) to obtain the best results from LEP. The Z mass has been determined to a relative precision of \(10^{-5}\), \(m_Z = 91.1875 \pm 0.0021\) GeV. The ratio of invisible width, \(\Gamma_{\text{inv}} = \Gamma_Z - \Gamma_e - \Gamma_\mu - \Gamma_\tau - \Gamma_{\text{had}}\), to the leptonic width is:

\[ \frac{\Gamma_{\text{inv}}}{\Gamma_\ell} = 5.943 \pm 0.016. \]  

(21)

The uncertainty is largely due to the luminosity uncertainties. Assuming the invisible width is due to neutrinos, and using the standard model expectation for this ratio for a single neutrino species \((\Gamma_\nu/\Gamma_\ell)_{\text{SM}} = 1.99125 \pm 0.00083\), the number of light neutrino species is obtained as:

\[ N_\nu = \frac{(\Gamma_{\text{inv}}/\Gamma_\ell)}{(\Gamma_\nu/\Gamma_\ell)_{\text{SM}}} = 2.9840 \pm 0.0082 \]  

(22)

which is consistent with three. Alternatively, assuming \(N_\nu = 3\) additional contribution to the invisible width is

\[ \Delta \Gamma_{\text{inv}} = -2.7 \pm 1.6 \text{ MeV}, \]  

(23)

and the 95\% confidence level upper limit on the additional width is

\[ \Delta \Gamma_{\text{inv}} < 2.0 \text{ MeV} \]  

(24)

4.2 \( \tau \) polarisation

The fermion pairs produced via the Z resonance is polarised according to eq. (14). In the \( \tau \)-pair production, polarisation can be measured from the kinematic distributions of \( \tau \) decay products. Observed distributions are in general distorted from the original distribution due to detector effects, biases arising from event selection and radiative effects, and backgrounds. Monte Carlo simulation is used to take into account such effects. Expected distribution is parametrised as a linear combination of two distributions, one corresponding to helicity +1 and the other to −1, and the relative fraction of the two contributions, giving the best fit to the data, corresponds to the \( \tau \) polarisation.

The LEP experiments have analysed all relevant decay modes: \( \tau \rightarrow \pi \nu_\tau, \rho \nu_\tau, a_1 \nu_\tau, c_1 \nu_\tau, \mu \nu_\tau \nu_\tau \). Depending on the nature of each decay mode, an optimal kinematic observable is constructed and used for fitting the polarisation. Combined \( \tau \) polarisation measurements\(^{45}\) are shown as a function of \( \cos \theta \) in Fig. 4. The data behave as expected from eq. (14). Separately extracted \( A_c \) and \( A_\ell \) from the \( \tau \) polarisation measurements are \( A_c = 0.1498 \pm 0.0049 \) and \( A_\ell = 0.1439 \pm 0.0043 \), consistent with equality as expected from the lepton universality. The combined leptonic asymmetry parameter \( A_\ell \) from the \( \tau \) polarisation measurements is \( A_\ell = 0.1465 \pm 0.0033 \).

4.3 \( b \) and \( c \) widths and asymmetries

\( Z \) decays into \( b \) and \( c \) quarks can be identified with good efficiency and purity. Using the tagged \( b \) and \( c \) samples, the \( Z \) partial widths into \( b \) and \( c \) quarks, and asymmetries were measured.

- Ratios of \( b \) and \( c \) partial width to the hadronic width:

\[
 R_b^0 = \frac{\Gamma_b}{\Gamma_{\text{had}}}, \quad R_c^0 = \frac{\Gamma_c}{\Gamma_{\text{had}}} \]  

(25)

- Forward–backwards polarisation asymmetries:

\[
 A_{FB}^{b} = \frac{3}{4} A_c A_b, \quad A_{FB}^{c} = \frac{3}{4} A_c A_c \]  

(26)

- Asymmetry parameters for \( b \) and \( c \) from \( A_{FB}^{b} \):

\[
 A_b, \quad A_c \]  

(27)

Various tagging techniques were developed. The lepton tag is based on the weak semi-leptonic decays of \( b/c \) flavoured hadrons. Due to the large mass of \( b/c \) hadrons, these leptons tend to have high momentum \( p_\ell \) and high transverse momentum \( p_t \) with respect to the jet axis. Separation between \( b \) and \( c \) quarks is made based on different \((p_\tau, p_t)\) spectra for \( b \rightarrow \tau \), \( b \rightarrow c \rightarrow \ell \), and \( c \rightarrow \ell \). For the asymmetry measurements, charge of the lepton provide information to distinguish \( q \) from \( \bar{q} \). A powerful method of \( b \) tagging is based on the long life time (typically 1.5 ps) and large mean charged multiplicity of \( B \) hadron decays. The LEP/SLD detectors were equipped with precision vertex detectors which allow detecting signatures of displaced decay vertex, either by directly reconstructing the secondary vertex, or based on a large number of tracks with significant impact parameter. Large invariant mass of the particles from the secondary vertex is also an indication of \( B \) hadron decays. These sensitive variables are often combined using artificial neural network or likelihood technique. Figure 5 shows an example of such \( b \)-tagging variables. The SLD experiment profited from the small beam size and small beam pipe of the SLC collider which enabled tagging \( b/c \) quarks with higher efficiency than LEP experiments. This allowed SLD to perform competitive measurements of \( R_b^0 \) and \( R_c^0 \) despite much lower statistics of the \( Z \) sample (≈1/10 of LEP).

Results from the LEP experiments and SLD using different techniques are found to be consistent. A combination\(^{46}\) has been made using these measurements to produce a set of \( b \) and \( c \) quark measurements on \( R_b^0, R_c^0, A_{FB}^{b}, A_{FB}^{c}, A_b, \) and \( A_c \).
4.4 Z couplings

Results of the asymmetry parameter $A_t$ are summarised graphically in Fig. 6. Leptonic asymmetry $A_L$ is obtained from a combination of leptonic $A_{FB}$, $\tau$ polarisation and the $A_{LR}$ measurements assuming lepton universality. For $b$ quark, $A_b$ is directly from the SLD $A_{FBLR}$ measurement. These two measurements are presented in the figure by vertical and horizontal bands. Forward–backward asymmetry for $b$ quark, $A_{FB}^{bb}$, is a product of $A_t$ and $A_b$, therefore represented by the diagonal band in the $A_t$--$A_b$ plane. The three bands have a common region in the plane as indicated by the confidence level contours. However the combined value of $A_b = 0.899 \pm 0.013$ is 2.7 standard deviation away from the standard model expectation of 0.934. This is mainly because the intersect of $A_{FB}^{bb}$ and $A_t$ bands deviates largely from the standard model value. In the standard model, as seen in the figure, $A_b$ is rather stable quantity insensitive to variations of any of the standard model parameters. The large deviation of the combined contour may be simply due to fluctuation of one or both of the $A_{FB}^{bb}$ and $A_t$ measurements, or could be arising from some new physics effects which involve non-standard $b$ couplings. Direct determination of $A_b$ is not precise enough to distinguish the two possibilities.

From the asymmetries $A_t$ and partial decay widths $\Gamma_t$, vector and axial-vector couplings can be extracted; $A_t$ determines $g_{tV}$, while $\Gamma_t$ determines $g_{tA}$. Results for lepton, $b$ and $c$ couplings obtained from LEP and SLD partial widths and asymmetry measurements are shown Fig. 7. Results for the three lepton species agree well. The combined result assuming lepton universality is also shown in the plot. The vertical arrow ($\Delta \alpha$) shows the expectation where only the running of $\alpha$ is considered in the radiative correction ($g_{Al} = -1/2$). The observed value of $g_{Al}$ is significantly different from this expectation, which demonstrate the effect of electroweak correction $\Delta \rho$ on $\rho = 1$. Also shown is the expectation of the standard model. The result of lepton couplings prefers low Higgs mass. The $b$ coupling is not quite consistent with the standard model, as already seen in the $A_t$--$A_b$ plot.

Assuming the standard model structure, the ratio $g_{tV}/g_{tA}$ can be represented by the effective mixing angle $\sin^2 \theta_{eff}$ [eq. (17)]. As seen above in Fig. 2, $A_b$ (and $A_c$) are insensitive to $\sin^2 \theta_{eff}$. Therefore $A_{FB}^{bb}$ and $A_{FB}^{cc}$ can be used to determine $A_t$ via $A_t = (4/3)A_{FB}^{bb}/A_b$ with $A_b$ constrained according to the standard model relation. The effective mixing angle, $\sin^2 \theta_{eff}$, determined from $A_t$ measurements for different processes are summarised in Fig. 8. Two most precise determinations, one from $A_{LR}$ by SLD and the other from $A_{FB}^{bb}$ at LEP, differ by 3.2 standard deviation. This difference is reflected in the overall $\chi^2$ of the average; $\chi^2/d.o.f = 11.8/5$, corresponding to a probability of 3.7%. The origin of this difference is as already seen in the comparison of $A_t$ in the discussion of asymmetries. If there were some new physics contribution that affects the $b$ coupling to largely deviate from the standard model, the extraction of $\sin^2 \theta_{eff}$ from the $A_{FB}^{bb}$ does not work.

5. Measurements above the $Z$

By the end of LEP-2 operation in the year 2000, the integrated luminosity collected by each of the LEP experiments at the centre-of-mass energies above the $Z$ reached 700 pb$^{-1}$. The range of centre-of-mass energy is 130–209 GeV.

$W$ physics at LEP-2

Each LEP experiment collected about 12,000 $W$ pair events $e^+e^- \rightarrow WW$ from the full LEP-2 data sample. This allows precise measurements of $W$ properties (mass, width and decay branching ratios), and direct tests of the unique structure of three gauge boson couplings through systematic studies of production cross-section, angular distribution and $W$ helicity.

A $W$ boson decays either hadronically ($W \rightarrow q\bar{q}$) or leptonically ($W \rightarrow \ell\nu$). In the $e^+e^-$ collisions at LEP-2, all decay final states from $W$ pair production are identified with good efficiency and purity. $W$ pair cross-section and $W$ decay branching ratios are determined from data without relying much on the standard model. Figure 9 shows
The measured W pair cross-section as a function of centre-of-mass energy. The measured data are compared to the standard model expectation [solid (green) curve] obtained using the latest precision calculations. Also shown in the figure are two dashed curves which correspond to two scenarios where one or both of the gauge boson self-couplings were absent. It clearly indicates that there are in fact self-couplings of gauge bosons and the structure of the couplings are as given in the standard model. Further detailed studies have been performed which can be found elsewhere.

The W boson mass contains, through the standard model relations and radiative corrections, information on the other standard model parameters. In particular, given the precisely known values of $\sin^2\theta^\text{eff}_{\text{lept}}$, $m_Z$, and $G_F$, radiative corrections are sensitive to the values of top mass and yet unknown Higgs mass.

W mass were measured at LEP-2 by reconstructing the invariant mass of the W decay products with a constraint of the LEP beam energy. Precise calibration of LEP energy is therefore vital for the W mass measurement. Two decay modes were used. In the $WW \rightarrow q\bar{q}qq$ channel, both W bosons decay into quark pairs, leading to 4-jet final states. Another channel is the $WW \rightarrow l^+l^-q\bar{q}$ final state where one W decays into $l^+l^-$ and the other into $q\bar{q}$. Though the neutrinos are invisible, due to the beam energy constraint, complete kinematic reconstruction is possible.

Combined $m_W$ from the results of the four LEP experiments is shown in Fig. 9. The statistical uncertainty of the LEP $m_W$ is 25 MeV. Without systematic effects, it would be about 20 MeV. However, due to uncertainties on the effects of final state hadronic interactions, the weight of the $WW \rightarrow q\bar{q}qq$ channel has been reduced in the combination. Also shown is the measurement from the $pp$ collider including the new preliminary result from Tevatron RUN-II. These two measurements are based on quite different techniques, and they agree well, yielding an average value of $80.398 \pm 0.025$ GeV.

6. Higgs Boson Searches

Higgs boson was searched at LEP and LEP-2. In $e^+e^-$ collisions at the LEP/LEP-2 energies, Higgs boson is produced mainly via Higgsstrahlung process,

$$e^+e^- \rightarrow ZH.$$  \hspace{1cm} (28)

Since the cross-section decreases as the $Z$ goes off-shell, the LEP-2 Higgs searches are limited to the Higgs mass...
up to approximately $\sqrt{s} - m_Z$. The negative results of the searches yielded a LEP combined lower bound on the standard model Higgs boson of 114.4 GeV at the 95% confidence level.  

7. Global Analysis in the Standard Model

The precision electroweak measurements at LEP/SLC and other experiments can be used in the global analysis in the framework of the standard model to see whether all data are compatible with the standard model expectations for a common set of the standard model parameters, and to derive or place constraint on the parameters. Particularly interesting of these parameters is the Higgs boson mass which is the last unknown standard model parameter.

The following data are used in the analysis discussed here.  

- The Z parameters
  - lineshape and lepton asymmetry at LEP: $m_Z$, $\Gamma_Z$, $\sigma^0_{\text{had}}$, $R^0_t$, and $A^0_{\text{FB}}$.  
  - $A_t$ from $t$ polarization at LEP  
  - $A_t$ from polarized left–right asymmetry from SLD  
  - Heavy quark measurements at LEP and SLD: $R^0_b$, $R^0_t$, $A^0_{\text{FB}}$, $A^0_{\text{lep}}$, $A_b$, and $A_c$.  
  - $\sin^2 \theta^\text{eff}$ from inclusive quark forward–backward asymmetry at LEP
- W mass $m_W$ and width $\Gamma_W$ from LEP and Tevatron
- Top mass $m_t$ at Tevatron
- Light quark contribution to the running of $\alpha$: $\Delta \alpha_h^{(5)}(m_Z)^{23}$

In the determination of $\Delta \alpha_h^{(5)}(m_Z)$, data of cross-section for $e^+e^- \to$ hadrons at low energies are particularly important.

The set of the standard model parameters used are: $\alpha(m_Z)$, $m_Z$, and $G_F$ as the three basic parameters of the electroweak interaction, and $\alpha(m_Z)$ for QCD corrections. In addition $m_t$ and $m_H$ are needed to calculate radiative corrections. The latest version of ZFITTER and TOPAZ0 programs are used. Some of the measurements in the list are used to directly constrain these parameters ($m_Z$, $\Delta \alpha_h^{(5)}(m_Z)$, $m_t$), while other observables are calculated in the standard model and compared to the experimental results.

As seen in §3 effects of electroweak radiative correction appear in $\Delta \rho$ and $\sin^2 \theta^\text{eff}$. They control the size of Z partial widths $\Gamma_t$ and asymmetry parameters $A_t$. In case of the hadronic width $\Gamma_{\text{had}}$, QCD final state correction is also included, therefore it is sensitive to the strong coupling constant $\alpha_s(m_Z)$. There are additional corrections to the b quark couplings arising from the vertex correction, leading to the unique dependence of $\Gamma_t$ on $m_t$. Finally the W mass involves the correction $\Delta \rho$ to eq. (1).

The Z observables are functions of quantities such as $m_Z$, $\Gamma_t$, $\sin^2 \theta^\text{eff}$. Depending on how the observables are constructed from these quantities, the sensitivity to the standard model parameters are different. For example, the asymmetry $A_t$ is a function of $\sin^2 \theta^\text{eff}$, which is sensitive to $m_t$ and $m_H$. On the other hand partial width ratio $R^0_t = \Gamma_{\text{had}}/\Gamma_t$ is insensitive to the values of $m_t$ and $m_H$ due to a large cancellation of universal corrections included in both of $\Gamma_{\text{had}}$ and $\Gamma_t$, but it is sensitive to the strong coupling constant $\alpha_s(m_Z)$ which is involved in the QCD correction on $\Gamma_{\text{had}}$. The hadronic pole cross-section $\sigma^0_{\text{had}} = (12\pi/m_Z^2)\Gamma_t \Gamma_{\text{had}}/\Gamma_Z^2$ has only weak sensitivity to any of the standard model parameters. This is because in addition to the cancellation of universal corrections in the Z widths, the QCD correction, which is included both in $\Gamma_{\text{had}}$ and $\Gamma_t$, also cancels largely. However, $\sigma^0_{\text{had}}$ is sensitive to the invisible width, or, number of neutrino species $N_\nu$, which directly determines the total width $\Gamma_Z$.

To see how these observables place constraints on $m_t$ and $m_H$ through radiative corrections, Fig. 10 shows in the $m_{tH}$ plane the constraints set by the measurements of the four sensitive observables; $\Gamma_t$, $\sin^2 \theta^\text{eff}$, $R^0_b = \Gamma_b/\Gamma_{\text{had}}$, and $m_W$. Note that $R^0_b$ is insensitive to $m_t$ and provide constraint on $m_t$ almost independently of $m_H$. The contour shows the 68% confidence level region determined from the Z data only, without using $m_W$ and $m_t$. This can be compared to the constraint band from $m_W$. They are in agreement.

As an another way of presentation, Fig. 11 shows in the $m_r - m_{\nu W}$ plane the constraint using indirect measurements only, and comparison with the direct measurements. Direct
and indirect determinations are consistent, an important test of the standard model at the level of radiative corrections. Also shown in Fig. 11 are trajectories corresponding to given values of Higgs boson mass \( m_H \). Comparison of these lines and confidence contours indicates that the data are sensitive to the Higgs boson mass, and they prefer low \( m_H \).

Using all data including \( m_W \) and \( m_t \), a constraint on the Higgs boson mass is obtained. Figure 12 shows \( \Delta \chi^2 = \chi^2 - \chi^2_{\text{min}} \) of the fit as a function of \( m_H \). The minimum \( \chi^2/\text{d.o.f} \) is 18.2/13 which corresponds to the probability of 15.1%. Higgs mass at the minimum \( \chi^2 \) is 76 GeV. The 95% upper bound on \( m_H \), corresponding to the value which gives \( \Delta \chi^2 = 2.7 \), is 144 GeV taking into account theoretical uncertainties, while the lower bound on \( m_H \) obtained from the direct searches at LEP is 114.4 GeV at the 95% confidence level.

8. Conclusion

Twelve years of running of the two \( e^+e^- \) colliders, LEP and SLC, was very fruitful. Determination of the number of light neutrino species is one of the most important results. Direct tests of the gauge boson self-couplings should also be noted. Precision results from LEP/SLC allowed tests of the standard model to the level of radiative correction, and it seems the standard model, based on the gauge symmetry and the spontaneous symmetry breaking due to the Higgs mechanism, is well established. Agreement between indirect and direct determinations of \( m_W \) and \( m_t \) is a demonstration of such tests. Similar analysis has been repeated to predict the Higgs mass. Current knowledge on the probable Higgs mass is in the range of 114.4–144 GeV as inferred from the results of direct searches at LEP and indirect bound through the analysis of precision measurements in the framework of the standard model.

Further improvements of experimental data will come in future; new information from the current and future experiments include a new set of \( m_W \) and \( m_t \) measurements from Tevatron. LHC will further contribute to these precision measurements, and more importantly discovery and studies of the nature of the Higgs boson, and probably something new beyond the standard model. Future \( e^+e^- \) linear collider will follow to make a new round of precision and systematic survey of physics, hopefully be including new physics.

Acknowledgements

The author wishes to thank the colleagues of the LEP electroweak working group.
Tatsuo Kawamoto was born in Hiroshima, Japan in 1952. He obtained his B. Sc. (1975), M. Sc. (1978), and D. Sc. (1981) degrees from Hiroshima University. He was a research staff (1982–1994) at Faculty of Science, the University of Tokyo. Since 1994 he has been an associate professor at International Center for Elementary Particle Physics (ICEPP), the University of Tokyo. He has worked on experimental particle physics, in particular experiments at high energy e^+e^- colliders: JADE experiment at DESY and then OPAL experiment at CERN. His research interests include precision studies of electroweak interaction and searches for new phenomena beyond the standard model. Since some years, he has been active on the preparation of ATLAS experiment at Large Hadron Collider (LHC) at CERN, which will start collecting physics data in 2008.
Experimental Highlights of Electroweak Physics at Hadron Colliders

Beate HEINEMANN*

University of California, Berkeley, 366 LeConte #7300, Berkeley, CA 94720-7300, U.S.A.

(Received January 18, 2007; accepted April 24, 2007; published November 12, 2007)

Hadron colliders have played a pivotal role in establishing the electroweak sector of the standard model at high precision. Highlights were the discovery of the $W$ and $Z$ bosons in 1983, the discovery of the top quark in 1995 and the high precision measurements of the mass of the $W$ boson and the top quark. The Higgs boson is the final missing piece in the theory of electroweak interactions. In this article the discoveries and precision measurements of the $W$ boson and top quark are reviewed, and the status of searches for the Higgs boson at hadron colliders is presented.

KEYWORDS: electroweak, hadron collider, $W$ boson, top quark, Higgs boson

DOI: 10.1143/JPSJ.76.111004

1. Introduction

There have been two hadron colliders, the $SppS$ and the Tevatron, that were pivotal to establishing and to testing the electroweak sector of the standard model. In 1982 the proton–antiproton collider $SppS$ started operation at the European Center of Nuclear Physics (CERN) in Geneva/Switzerland. It was the first collider ever built where antiprotons were accelerated and collided. The proton and antiproton beams were collided at two experiments, called UA1 and UA2, with a center of mass energy of $\sqrt{s} = 540\text{GeV}$ initially and $630\text{GeV}$ later. In 1985 first collisions took place at the Tevatron proton-antiproton collider in the United States near Chicago. Here the beam energy was $900\text{GeV}$, yielding a center-of-mass energy of $1.8\text{TeV}$. However, at that time the luminosity was much smaller than at the $SppS$, and it was only in 1989 that the Tevatron physicists had sufficient data to surpass the measurements by the UA1 and UA2 collaborations. At the Tevatron there were two experiments, called CDF and D0. About $100\text{pb}^{-1}$ of integrated luminosity was collected between 1989 and 1996 in so-called Run I. Then both the accelerator and the experiments were upgraded, the energy was increased to $1.96\text{TeV}$, and about $1.5\text{fb}^{-1}$ of data have been collected in “Run II” between March 2002 and December 2006.

The electroweak force is mediated by the gauge bosons, namely the $W^\pm$ and the $Z$ boson.\(^1\) The electroweak symmetry is broken via a Higgs field that has a vacuum expectation value of $246\text{GeV}$. This mechanism of electroweak symmetry breaking causes the $W^\pm$ and the $Z$ boson to be massive, and also the fermions, e.g., the electron or the top quark. The stronger any particle interacts with the Higgs boson the more massive it is. Within the standard model there are precise predictions for the relationships between, e.g., the mass of the $W^\pm$ and $Z$ bosons, the Higgs boson, and the top quark, and the couplings of the gauge and Higgs bosons to any fermions and bosons. Thus measuring any of the masses or couplings of these particles is critical to further constrain and test the electroweak theory, and has been a major focus of hadron collider experiments.

In the following firstly hadron colliders are introduced and the experimental challenges are described. Then the discovery and the precision measurements of the $W^\pm$ and $Z$ boson quantities are highlighted. A review of the discovery and precision measurements of the top quark is presented. Finally the current knowledge of the Higgs boson is reviewed and perspectives for a discovery are given.

2. The Challenge of a Hadron Collider

Analyzing data at a hadron collider is significantly more challenging than at a lepton collider. The reason is that hadrons are not point like fundamental objects, but rather are composed of quarks and gluons. A schematic diagram of a proton–antiproton collision is shown in Fig. 1 for $W^\pm$ production.

This causes several experimental challenges:

- The composition of a hadron is known on average but unknown for any given interaction. Thus even though the $pp$ center-of-mass energy is well known the actual energy of the collision is not. The outgoing non-interacting partons are not measured in the detectors as they travel along their original direction and thus escape through the beam pipe. Therefore energy conservation which is a powerful constraint at lepton colliders can not be exploited at hadron colliders. However, since the outgoing partons carry negligible transverse momentum the constraint that is used at hadron colliders is the conservation of the momentum in the transverse plane.
- When one of the partons in the hadron undergoes a hard interaction the hadron looses its color neutralness, and

\(^1\)E-mail: BHEHeinemann@lbl.gov

---

Fig. 1. Schematic diagrams for $pp \to W^\pm \to \mu^\pm \nu$ production. The systems $X$ indicate the particles due to the proton remnant that are produced in $pp$ collision in addition to the hard scattering process.
the remaining partons interact strongly with each other or the partons from the other hadron. These rather soft interactions cause energy spray into the detectors that is unrelated to the hard scattering process and must be accounted for experimentally.

- Quarks and gluons are much more likely to interact strongly than electromagnetically or weakly. In particular at low energy the coupling constant of the strong interaction, $\alpha_s$, is much larger than the electromagnetic coupling constant, $\alpha_{em}$, or the weak constant. E.g., at $M_Z = 91$ GeV $\alpha_s \approx 0.118$ and thus more that 10 times higher than $\alpha_{em} \approx 1/128$. Therefore there is a large background from strong parton interactions to $W^\pm$ bosons in the quark decay modes. This large background makes it very difficult to observe the $W^\pm$ or $Z$ boson in its decay to quarks. Thus most measurements at hadron colliders are made using the lepton decay modes.

3. The Electroweak Gauge Bosons

The dominant processes for $W^\pm$ production at hadron colliders are $\bar{u}d \rightarrow W^- + u/d \rightarrow W^+$ and $u\bar{u} \rightarrow Z$ and $d\bar{d} \rightarrow Z$ for $Z$ production. The processes are illustrated in Fig. 2.

The best modes for observing $W^\pm$ and $Z$ bosons are the lepton decay modes. The standard model predicts about 10% of all $W^\pm$ bosons decay into an electron (or positron) and an electron-neutrino, and 10% into a muon and a muon-neutrino. Similarly, about 3% of the $Z$ boson are expected to decay into $e^+e^-$ and 3% decay into $\mu^+\mu^-$. These leptonic decay modes were used for the observation and for the subsequent precision measurements at hadron colliders.

3.1 Discovery of the $W^\pm$ and $Z$ bosons

The primary goal of the $Sp\bar{p}S$ collider and its experiments was the observation of the $W^\pm$ and $Z$ bosons. $^2$ The predictions of these bosons had been made as a consequence of the electroweak theory. $^1$ W bosons were thought to be responsible for the $\beta$ decay ($q \rightarrow q' e^+ \nu_e$) and $Z$ bosons were the prime candidate for explaining the observation of neutral currents ($\nu q \rightarrow \nu' q'$) at the Gargamelle bubble-chamber experiment in 1973. $^3$ While in these processes the intermediate vector bosons were exchanged in the $t$ channel, a resonance production in the $s$-channel could occur in quark–antiquark interactions: $gq \rightarrow W^+$ and $gq \rightarrow Z$.

For $W^\pm$ bosons the experimental strategy was to search for high energy electrons or muons and an imbalance in the total transverse energy, called “missing $E_T$” ($E_T$), measured in the event. The measurement of $E_T$ was a novel idea at that time, and the UA1 detector was designed to be as hermetic as possible to allow this quantity to be measured as this technique only works if all the transverse energy produced in the collision is detected with high efficiency. In January 1983 the UA1 collaboration published the observation of six $W^\pm \rightarrow e^\pm \nu_e$ candidate events,$^4$ and measured the $W^\pm$ boson mass to be $81 \pm 5$ GeV/c$^2$ in excellent agreement with the prediction$^5-7$ of $82.0 \pm 2.4$ GeV/c$^2$. Only a few months later, in June 1983 the UA1 collaboration published the observation of four $Z \rightarrow e^+e^-$ and candidates and one $Z \rightarrow \mu^+\mu^- + X$ event,$^8$ and measured the $Z$ boson mass as $95.2 \pm 2.5$ GeV/c$^2$. For these discoveries Carlo Rubbia and Simon van der Meer were awarded the Nobel Prize in 1984.

3.2 Measurement of the $W$ boson mass

Since the neutrino cannot be measured directly and only its transverse momentum can be inferred from the measurement of $E_T$, the $W$ boson mass can not be measured directly as the invariant mass of the lepton and the neutrino. Instead the “transverse” mass is used, which is defined as

$$m_T(l, \nu_l) = \sqrt{2p_T(l)p_T(\nu_l)(1 - \cos \phi_{l\nu})},$$

where $p_T(l)$ and $p_T(\nu_l)$ are the transverse momenta of the charged lepton and the neutrino respectively and $\phi_{l\nu}$ is the difference in azimuthal angle between the charged lepton and the neutrino.

In the publication of the observation a first mass measurement was presented at $M_W = 81 \pm 5$ GeV based on the six candidates. At the end of the $Sp\bar{p}S$ run in 1990, UA2 presented a measurement$^9$ of $M_W = 80.49 \pm 0.43$ (stat.) $\pm 0.21$ (syst.) GeV/c$^2$. This measurement used an integrated luminosity of 13 pb$^{-1}$. For this precise measurement the ratio of the $W$ boson to the $Z$ boson mass was determined to reduce systematic uncertainties, e.g., on the precision of the lepton energy determination. The $W$ boson mass was then extracted using the precise determination of the $Z$ boson mass at the LEP and SLC $e^+e^-$ colliders.$^{10-14}$

In August 1990 the CDF experiment presented the first $W^\pm$ mass measurement$^{15}$ at the Tevatron of $M_W = 79.91 \pm 0.39$ GeV/c$^2$ based on an integrated luminosity of about 4 pb$^{-1}$, for the first time improving on the precision of the UA2 experiment. At the end of the Tevatron Run I period, in 1996, a luminosity of about 100 pb$^{-1}$ was available and CDF and D0 presented measurements with substantially improved precision. The CDF and D0 measurements were combined$^{16-18}$ to yield a Tevatron average of $80.456 \pm 0.059$ GeV/c$^2$.

Recently the CDF collaboration has presented a new measurement based on a luminosity of 200 pb$^{-1}$ of Run II data. This yields the world’s single most precise measurement of $80.413 \pm 0.048$ GeV/c$^2$. It is based on a combined fit to the transverse mass and lepton $p_T$ distributions in both the electron and the muon decay modes. In the electron decay mode it is based on 63,964 candidate events, and in the muon decay mode on 51,128 candidate events. The transverse mass distributions and the corresponding mass fits are shown in Fig. 3 for electrons and muons. The measurement is most sensitive to the falling edge of the spectrum $76 < m_T < 86$ GeV/c$^2$. This is by far the highest precision measurement made at a hadron collider with the remarkable precision of only 0.059%. A summary of the $W$ mass measurements at hadron and lepton$^{19}$ colliders is presented in Fig. 4. The world average is at present $80.398 \pm 0.025$ GeV/c$^2$. 

Fig. 2. Lowest order diagrams for $W$ and $Z$ boson production.
It is interesting to review the uncertainties of the Tevatron experiments to judge whether further improvements can be expected in the future. A summary of the dominant systematic uncertainties is given in Table I for the recent CDF measurement for the fit of the transverse mass distribution.

The dominant uncertainty is the statistical precision which will be improved naturally with more data. The second largest uncertainty is the lepton energy scale. This is determined using data from the $J/\psi \rightarrow \mu^+\mu^-$, $\Upsilon \rightarrow \mu^+\mu^-$, and $Z \rightarrow \mu^+\mu^-$ decays for the track momentum scale, and by using the ratio of the calorimetric energy measurement to the track momentum measurement for electrons in $W \rightarrow ev$ decays. They will be improved as more data are collected as their precision is mostly limited by the statistical precision of these samples. Most other uncertainties are derived using the $Z \rightarrow \ell^+\ell^-$ samples, in particular the recoil and the $p_T(W)$ model will improve with increasing $Z$ sample size. The only systematic uncertainties that can not be improved with increasing statistical precision directly are the parton distribution functions and QED radiation where the analysis relies on external input. With $2\,fb^{-1}$ of luminosity it is expected that the Tevatron experiments achieves a precision of 20–30 MeV, depending on whether the uncertainties on the parton distribution functions and the QED radiation can be further reduced.

3.3 Other measurements of $W$ and $Z$ boson properties

In addition to the $W$ boson mass a large number of other interesting properties has been measured at hadron colliders. They are shortly summarized here.

The width of the $W$ boson is measured directly and indirectly. It is sensitive to possible decays of the $W$ to other new particles due to contributions from particles beyond the SM. The direct measurement uses also the transverse mass distribution but the width sensitivity arises mostly from the tail of the distribution, $100 < m_T < 200\text{ GeV}/c^2$. Combining all the results from the Tevatron$^{20}$ it is found to be $2078 \pm 87\text{ MeV}/c^2$ in good agreement with the LEP value$^{21}$ of $2128 \pm 88\text{ MeV}/c^2$ and the SM prediction$^{22}$ of $2092 \pm 3\text{ MeV}/c^2$. Indirect determinations are also in good agreement with the SM value.$^{23}$

In addition the leptonic branching ratios of the $W$,$^{23}$ the coupling of the $W$ and $Z$ bosons to photons,$^{24-26}$ the production of two $W$ or $Z$ bosons,$^{27-30}$ the forward–backward asymmetry of the $Z$ boson,$^{31}$ and several measurements related to the production of $W$ or $Z$ bosons were made.

The quark decay modes are very difficult to detect at hadron colliders since there is a huge background from QCD dijet production that obscures the signal. The quark decay

![Fig. 3](image-url) (Color online) Transverse mass distribution for $W \rightarrow ev$ (left) and $W \rightarrow \mu\nu$ (right) events. The data (points) are compared to the simulation (histogram) using the best fit $M_W$ value. Also shown are the fit values for the $W$ mass with the statistical uncertainty. The histograms near the bottom show the background from non-$W$ production.

![Fig. 4](image-url) (Color online) $W$ boson mass measurements from CDF and D0 Run I, from the LEP experiments and from CDF Run II. Also shown is the world average.

Table I. Systematic uncertainties for Run II CDF $W^\pm$ mass measurement.

<table>
<thead>
<tr>
<th>Source</th>
<th>$W \rightarrow ev$</th>
<th>$W \rightarrow \mu\nu$</th>
<th>Common</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistical</td>
<td>48</td>
<td>54</td>
<td>0</td>
</tr>
<tr>
<td>Lepton energy</td>
<td>31</td>
<td>17</td>
<td>17</td>
</tr>
<tr>
<td>Recoil and $p_T(W)$</td>
<td>11</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>Selection and backgrounds</td>
<td>12</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>Parton density functions</td>
<td>11</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>QED radiation</td>
<td>11</td>
<td>12</td>
<td>11</td>
</tr>
<tr>
<td>Total</td>
<td>62</td>
<td>60</td>
<td>26</td>
</tr>
</tbody>
</table>
modes were, however, established successfully\(^{32}\) but are not used to make precision measurements of the \(W^\pm\) and Z boson properties.

In all of these measurement the electroweak theory was confirmed and in many of them the precision is similar to that of lepton colliders.

4. The Top Quark

In the standard model top quarks are mostly produced in pairs and at the Tevatron \(q\bar{q}\) annihilation contributes about 85% and the remaining 15% arise from \(gg\) interactions. The leading order diagrams are shown in Fig. 5.

The top quark was discovered by the CDF and D0 experiments in 1995.\(^{33-35}\) It has a surprisingly large mass of about 175 GeV/c\(^2\), 40 times larger than the second heaviest quark. Its large mass makes the top quark a critical particle for the electroweak theory as the Higgs boson coupling to particles is stronger the larger the mass of that particle is.

For this mass, the top quark nearly always decays into a b quark and a W boson. The reaction chain is

\[
p p \rightarrow t \bar{t} \rightarrow W^+ b W^- b + X
\]

and in the following we call (1) the “dilepton” mode, (2) the “lepton+jets” mode, and (3) the “all-jets” mode.

4.1 Observation of the top quark

In 1995 the CDF and D0 collaborations claimed observation of the top quark. The observation was based on the dilepton and lepton+jets modes. In the lepton+jets mode CDF required an electron or muon from the dilepton and lepton+jets modes. In the lepton+jets mode the “dilepton” mode, large \(W^\pm\) boson properties.

The D0 experiment was not equipped with a silicon microstrip detector: the observation at D0 was based on the dilepton mode, on the semi-leptonic b-hadron decay tagging, and on the sum of the transverse energies of all jets in lepton+jets events. In total 17 events were observed compared to a background estimate of 3.8 ± 0.6 events, corresponding to an excess with a significance of 4.6\(\sigma\).

Both experiment found the events to be consistent with coming from a top quark, and presented measurements of the top quark mass of 176 ± 13 GeV/c\(^2\) (CDF) and 199 ± 30 GeV/c\(^2\) (D0). These direct measurements were in excellent agreement with the prediction using electroweak precision observables at lepton colliders\(^{36}\) of 150–210 GeV/c\(^2\) depending on the mass of the Higgs boson.

The analysis techniques used for the observation of the top quark are still used in similar form today by the Tevatron experiments. After an upgrade between 1996 and 2001 the D0 collaboration also inserted a silicon microstrip detector and uses this now for identifying top quarks in Run II.

4.2 Precision measurements of the top quark mass

The top quark mass has been measured in each of the signatures, i.e., the dilepton, the lepton+jets, and the all-jets mode. The main challenges of the measurement are assigning the final state particles that originate from the same top quark correctly, understanding the background and controlling the systematic uncertainties, in particular on the jet energy measurement.

The highest precision measurements are achieved in the lepton-jets mode as in this sample the background is relatively small (about 20% when requiring a b-quark tag), the samples are relatively large and nearly all final state particles can be measured. In the dilepton mode there are two neutrinos escaping which complicates the analysis, and the sample is relatively smaller due to the small branching ratio of the \(W^\pm\) to leptons. The all-jets samples is large but the large background due to QCD multi-jet production complicates measurements in this decay mode. Therefore I will focus on describing the lepton+jets measurement in this article.

In the lepton+jets mode the four-momenta of the two b-jets, the two other jets and the lepton and \(p_t\) and \(p_t\) component of the neutrino are measured. The two unknown variables thus are the longitudinal momentum of the neutrino and the mass of the top quark. In addition there are three constraints in each event:

- the lepton-neutrino invariant mass must equal the \(W^\pm\) boson mass
- the invariant mass of the non-b-quark jets must equal the \(W^\pm\) boson mass
- the invariant mass of the decay products of the top must be the same as that of the anti-top

Since there are three constraints but only two unknowns the system is constrained well enough to measure the mass of the top quark. In fact since it is over-constrained it is even possible to fit for the jet energy scale as an additional free parameter. This has an advantage since the jet energy scale is difficult to constrain with high precision \(a \text{ priori}\) to a precision better than 3%. The reason is that jets are complicated objects, and in particular the hadronization of a parton to a jet is only understood by phenomenological models based on data but there are no strict theoretical calculations one could rely on. By fitting for the jet energy scale inside the top sample itself the uncertainty becomes statistical and can improve with increasing data samples naturally. It results therefore in a higher precision of the measurement.
The current single best measurement was made by the CDF collaboration\(^58\) and used a technique called “matrix-element” method. For each event a probability distribution as function of top quark mass and jet energy scale is calculated, and the final measurement of the top mass is made by forming a likelihood as a product of the single event probabilities.\(^{39}\) The likelihood is shown in Fig. 6 as function of the top quark mass and the jet energy scale. The measurement yields \(m_{\text{top}} = 170.9 \pm 2.2 \text{(stat.)} \pm 1.4 \text{(syst.)} \text{GeV}/c^2\) where the jet energy scale uncertainty is part of the statistical error and the dominant systematic uncertainties arise from the modeling of QCD radiation, and residual uncertainties on the jet energy scale.

The top quark mass was additionally measured in the other decay channels\(^{38,40-42}\) by both CDF and D0 experiments using Run I and Run II data, and all measurements are found to agree well with each other as is seen in Fig. 7. They are thus combined\(^43\) to give the world average top quark mass of \(m_{\text{top}} = 171.4 \pm 2.1 \text{GeV}/c^2\), corresponding to a precision of 1.2%.

### 4.3 Other top quark measurements

A large number of other measurements related to the top quark have also been made and found to all be consistent with the SM predictions, e.g., measurements of the top quark production cross section,\(^{44-48}\) of the charge of the top quark,\(^{39}\) the helicity of the W boson in top quark decays\(^{50-53}\) and the branching ratio\(^{54}\) \(t \rightarrow Wb\) have been made. All of them confirmed the SM predictions and no deviation from the expectations have yet been seen.

Recently a first sign of the single top quark production process that proceeds via the electroweak interaction has been claimed.\(^{35,57}\) The results are consistent with the SM expectation\(^56\) but the precision is rather limited, and more data are required to claim conclusive observation of single top production and to test its agreement with the SM prediction with better precision.

### 5. The Higgs Boson

The Higgs Boson is often called the “holy grail” of the electroweak sector of the standard model as it plays the critical role of generating the masses of the other SM particles through the mechanism known as electroweak symmetry breaking. It is the only SM particle that has not yet been observed experimentally, and finding it is a major goal of the Tevatron and of the large hadron collider (LHC) experiments at CERN that will start operating end of 2007. The best experimental bounds so far have been set by the LEP experiments that exclude masses below 114.4 \text{GeV}/c^2 at 95% confidence level (CL).\(^{58}\)

#### 5.1 Indirect constraints on the Higgs boson mass

Within the context of the standard model the precision measurements of the \(W\) bosons and top quark mass can be used to constrain the Higgs boson mass via radiative corrections. The \(W\) boson mass is given by

\[
M_W^\text{rad} = \frac{\pi \alpha}{\sqrt{s G_F \sin^2 \theta_W}} \frac{1}{1 - \Delta r},
\]

where \(\alpha\) is the fine structure constant, \(G_F\) the Fermi constant, \(\theta_W\) the Weinberg angle, and \(\Delta r\) is a term that contains the radiative corrections. This terms \(\Delta r\) is proportional to \(m_{\text{top}}^2\) and to \(\ln(M_{\text{H}})\). Example diagrams that introduce this dependence are shown in Fig. 8.

The \(W\) boson mass is shown versus the top quark mass in Fig. 9. Shown are the direct measurements from the Tevatron and LEP2 colliders, the indirect constraints from LEP1 experiments and SLD and lines of constant Higgs boson mass. The direct and indirect measurements are in good agreement. The direct measurements, however, indicate a rather low Higgs boson mass which is already experimentally excluded by LEP2. The most probable value is determined as \(M_{\text{H}} = 80.1^{+36}_{-26} \text{GeV}/c^2\), and at 95% CL Higgs mass valuers above 153 \text{GeV}/c^2 are excluded.

---

Fig. 6. Two-dimensional likelihood of the jet energy scale versus the top quark mass. Shown is the value of minimal likelihood (cross) and contours of equal likelihood corresponding to 1σ, 2σ, 3σ, and 4σ. The jet energy scale is defined as the ratio of the jet energy scale in data and simulation.

Fig. 7. (Color online) Top mass measurements by the CDF and D0 collaborations in the individual analysis modes. Also shown is the world average value.

Fig. 8. Feynman diagrams that for radiative corrections to the \(W\) boson mass due to loops involving the top quark and the Higgs boson.
precision is required to see whether this slight tension between the direct limit on the Higgs boson mass and the indirect prediction will become more significant or whether it is just a statistical fluctuation.

The current top quark mass precision of δ(m_{top}) = 2.1 GeV/c^2 corresponds to a Higgs mass uncertainty of δ(M_H)/M_H ≈ 18% while the current world W mass precision of 25 MeV/c^2 corresponds to a Higgs mass precision of δ(M_H)/M_H ≈ 35%. Thus currently the W mass precision limits the indirect constraints on the Higgs boson mass, and the precision on the Higgs boson mass would benefit most from improvements in the W mass precision.

5.2 Direct searches for the Higgs boson

In pp collisions at the Tevatron Higgs bosons are predominantly produced via gg fusion: this cross section is about 1 pb at M_H = 120 GeV/c^2, and falls to 0.4 pb at 160 GeV/c^2. The subdominant production process is Higgs-radiation off a W or Z boson, and the cross sections are 0.15 and 0.09 pb for WH and ZH production respectively for M_H = 120 GeV. The Feynman diagrams for these processes are shown in Fig. 10. This can be compared to, e.g., the t\bar{t} production cross section of 7 pb or the W production cross section of about 2.7 nb. The main difficulty in the Higgs boson search arises from the very small signal yield, compared to the significant backgrounds that are present.

At low mass, M_H < 135 GeV/c^2, the dominant decay mode of the Higgs is to a b\bar{b} pair. This decay mode cannot be detected in the gg → H channel due to a large background from QCD b\bar{b} production. Thus at low mass the most sensitive analyses use the WH and ZH production modes in the leptonic decay channels of the W and the Z. The three most sensitive analyses are WH → ℓ⁺ℓ⁻b\bar{b}, ZH → ℓ⁺ℓ⁻b\bar{b}, and ZH → ℓ⁺ℓ⁻b\bar{b}. The invariant mass of the two b-jets is expected to peak at the H boson mass, and this is exploited to discriminate WH production from the dominant backgrounds, Wbb/c\bar{c}, and t\bar{t} production. The invariant mass distribution of the two b-jets is shown for the WH analysis in Fig. 11. A review of all low mass Higgs searches is given in ref. 59.

At larger masses, M_H > 135 GeV/c^2, the Higgs boson decays predominantly to W^+W^−. The standard model background arises mostly from the relatively low cross section SM WW production process and this analysis can make use of the large gg → H production cross section. The analysis has been performed requiring both W’s to decay to leptons: the analysis channels are WW → ℓ⁺ν_ℓℓν_ℓ, WW → μν_μν_μ, WW → μν_μν_μ. Large E_T is required to indicate the presence of neutrinos. To discriminate the signal from the SM background a large number of selection cuts are made. The final discriminant is shown in Fig. 12 for the D0 analysis of pp → H → W^+W^− → ℓ^±μ^±ν_ℓν_μ. Large E_T is the difference in azimuthal angle between the two leptons. Since the Higgs is a scalar particle and due to helicity conservation the two leptons are typically relatively close to each other, while for the backgrounds the lepton angles are less correlated.

Presently the Tevatron Higgs analyses are not sensitive to the predicted SM rate of Higgs production. Even when combining all the analyses from both CDF and D0 there is no sensitivity yet to SM Higgs production. This is seen in Fig. 13 where the ratio of the experimental cross section limit at 95% CL divided by the SM cross section is shown. At low M_H = 115 GeV/c^2 the observed limit is a factor of 10 larger than the SM prediction, and at M_H = 160 GeV/c^2 it is a factor 4 above the SM prediction.

The Tevatron will increase the integrated luminosity by about a factor 10 compared to this result. That will improve

Fig. 10. Feynman diagrams for gg → H (left), q\bar{q} → WH (middle), and q\bar{q} → ZH (right) production.

Fig. 11. (Color online) Dijet invariant mass distribution of two identified b jets for events with a W boson. The data (points) are compared to the SM prediction: the dominant background contributions are W+Heavy Flavor and t\bar{t} production. Shown as open histogram is also the SM Higgs contribution, increased by a factor 10 compared to the SM cross section.
the limits by about a factor 3 if the analyses are kept unchanged which is still not sufficient to be sensitive to SM Higgs production. Thus the experiments are working on further improving the analyses and hope to gain another factor of 3 by making experimental improvements, e.g., increasing the $b$-jet identification efficiency, improving the acceptance for leptons, improving the Higgs mass resolution and using advanced analyses techniques (e.g., neural networks) to gain additional discrimination power. If this is achieved it could be possible for the Tevatron to see a first glimpse of the Higgs boson. However, ultimately it is expected that the experiments at the LHC will for sure either detect a Higgs boson or exclude its existence at all masses. It is expected that by 2010 a Higgs boson is either found or excluded. The discovery of the Higgs boson would be yet another amazing triumph of the electroweak theory.

6. Conclusions and Outlook

The $SppS$ and Tevatron $p\bar{p}$ colliders have been critical in establishing the theory for electroweak interactions, and in testing it at high precision. Highlights are the observation of $W$ and $Z$ bosons at the $SppS$, the observation of the top quark at the Tevatron, and the subsequent precision measurements of 0.06% accuracy for the $W$ boson and 1.2% accuracy for the top quark. So far the electroweak theory has withstood all these experimental tests but just now we may start to face an interesting situation where there is a slight tension between the indirect determination of the Higgs boson mass and the direct limits.

The Tevatron collider experiments are still in progress and are expected to improve the top quark and $W$ mass precision by about a factor of two in accuracy. Further improvements of these precision measurements are expected by the experiments at the Large Hadron Collider starting in 2008. It is possible that the Tevatron will see a first hint for a Higgs boson in the coming years, but ultimately the LHC collider will be required to observe the Higgs boson conclusively and to measure its properties. These measurements will further challenge the electroweak sector of the standard model, and it remains to be seen whether the electroweak theory will continue to be so successful or whether it will break down, indicating presence of new particles and new laws of nature that yet need to be discovered.

20) Tevatron electroweak working group (for the CDF and D0 Collaborations): hep-ex/0510077.
21) The LEP electroweak working group (for the ALEPH, DELPHI, L3

Fig. 12. (Color online) Difference in azimuthal angle, $\Delta \phi(e, \mu)$, between the electron and the muon for the search for Higgs bosons in $pp \rightarrow H \rightarrow W^+W^- \rightarrow e\mu \nu \nu$ production. Shown are the data (points) and the SM background prediction, dominated by $W^+W^-$ production. Also shown as open histogram is the expected SM Higgs signal, increased by a factor 10.

Fig. 13. (Color online) Ratio of the 95% CL cross section limit divided by the SM Higgs boson cross section as function of the mass of the Higgs boson. Shown are the individual expected limits from CDF and D0 and the combined expected limit (dashed lines) and the observed limit (full line). Also shown is the limit set by LEP2. The expected limit is defined as the average outcome in a large ensemble of experiments.
and OPAL Collaborations): hep-ex/0612034.
43) Tevatron electroweak working group (CDF and D0 Collaborations): hep-ex/0608032.
44) V. M. Abazov et al. (D0 Collaboration): hep-ex/0612040; submitted to Phys. Rev. D.
59) B. Kilminster et al. (CDF and D0 Collaborations): Proc. 33rd ICHEP Conf., Moscow, August 2006; hep-ex/0611001.
60) G. Bernardi et al. (CDF and D0 Collaborations): Proc. 33rd ICHEP Conf., Moscow, August 2006; hep-ex/0612044.

Beate Heinemann received her Diploma (1996) and Ph. D. (1999) from the University of Hamburg in Germany. From 1999–2002 she had a postdoctoral fellowship from the Particle Physics and Astronomy Research Council (PPARC) at the University of Liverpool in the United Kingdom. From 2002–2004 she had a PPARC Advanced Fellowship and from 2004–2006 a fellowship from the Royal Society at the University of Liverpool. In 2006 she was appointed Associate Professor of Physics at the University of California, Berkeley.
Theoretical Overview of Flavor Physics

A. I. SANDA*  

Department of Physics, Kanagawa University, Yokohama 221-8686  

(Received February 23, 2007; accepted May 7, 2007; published November 12, 2007)

Without flavor physics, we would not have the standard model of elementary particles today. The goal of flavor physics is to get at the most fundamental theory of nature. This is done by looking for phenomena which can not be explained by the standard model. To do this, we need to study flavor physics from every angle possible. In support of this, we review how flavor physics has contributed to our understanding of nature over the past 60 years.

KEYWORDS: flavor, quarks, beauty, charm, strange  
DOI: 10.1143/JPSJ.76.111005

1. Introduction: The Discovery of the New Field

In 1940’s and 50’s, particles with very strange behavior have been found, for example, $K^-$ meson, $\Lambda^0$, $\Sigma^+$, and $\Xi^-$ baryons. When they are produced in collisions:

$$\pi + P \rightarrow \Lambda K^+, \Sigma^0 K^+, \tag{1}$$

their production rates are as large as those of typical strong interaction. But when they decay, for example, as in:

$$\Lambda \rightarrow P + \pi^-; \tag{2}$$

$$K^\pm \rightarrow \pi^\pm + \pi^0;$$

their decay rates are too slow compared to a typical strong decay like $N^0 \rightarrow P + \pi$. Thus they decay through weak interaction, and thus indeed deserved different names. Strangeness assignments are: 0 for proton, neutron, and $\pi$ meson, $-1$ for $\Lambda$, $\Sigma^0$, $\Sigma^+$, and $K^-$, $+1$ for $K^+$, $-2$ for $\Xi^-$, etc.

The strangeness is conserved in strong and electromagnetic interactions. The interactions shown in eq. (1) contain particles with opposite strangeness so that the final states as a whole are zero strangeness states — thus these production processes conserve strangeness and they get produced through strong interaction.

Particles with strangeness cannot decay to the final state with 0 strangeness through strong and electromagnetic interactions. Thus they decay through weak interaction, and they have relatively long life time. For example, particles decaying through strong interaction may have typical lifetime of order $10^{-23}$ s, while those decaying through weak interaction have typical life time of order $10^{-10–10^{-15}}$ s.

Flavor physics is a systematic study of particles with these new quantum numbers. These particles can live for long time, and they have time to expose interesting phenomena. Thus the purpose of this study is to search for yet unprobed physics hidden behind flavor changing weak decays, and to learn the fundamental laws which govern interactions of matter.

This discovery of hadrons with the internal quantum number “strangeness” marks the beginning of a most exciting epoch in particle physics that even now, sixty years later, has not yet found its conclusion.  

2. First Suggestion of Parity Violation: The $\theta–\tau$ Puzzle

Flavor physics immediately lead to something which was quite unexpected! Two very strange decays have been found for charged strange mesons, namely:

$$\theta^+ \rightarrow \pi^+ \pi^0; \tag{3}$$

$$\tau^+ \rightarrow \pi^+ \pi^0 \pi^- .$$

The problem arose when ever more precise measurements failed to find any significant difference in either the mass or the lifetime of the $\theta$ and $\tau$ mesons. This constituted the $\theta–\tau$ puzzle: how could nature assign the same mass to two distinct particles? Or even more baffling: how could nature contrive to generate the same lifetime to two distinct particles, the major decay channels of which possess totally different phase space?

To explain it further, let us review the parity symmetry. Parity transformation is almost like transforming the world to its mirror image. If $(r_1, r_2, \ldots)$ denote the coordinates of everything in this world, the parity symmetry $P$ inverts the sign of all the coordinates: $(-r_1, -r_2, \ldots)$.

The $\theta^+$ was found to be a spinless state, and therefore, the $2\pi$ final state is in a positive parity state. (2$\pi$ in a S wave state is symmetric under the interchange of these particles.) The angular distributions of the three pions from the $\tau^+$ decay revealed the final state to carry zero total angular momentum as well, but with negative parity! It was assumed that parity, like angular momentum, was conserved by the relevant forces. The parity of the initial state then coincides with that of the final state. With $\theta$ and $\tau$ exhibiting different parity, they had to be distinct objects and thus indeed deserved different names.

It is difficult it to give up what we are use to. This puzzle could be understood trivially if parity were not absolutely conserved. For then $\theta$ and $\tau$ could represent merely two decay modes of the same particle. Parity conservation had been tested extensively in strong and electromagnetic transitions. Yet the breakthrough came when Lee and Yang pointed out in 1956 that this symmetry had not been probed yet in weak transitions.

This was how new important feature of weak interaction, the parity violation, was revealed through flavor physics. But this was only a beginning!

E-mail: sanda@kanagawa-u.ac.jp
Table I. Three families of elementary particles, quarks and leptons. These quarks are called u, d, c, t; electron, muon, tau, and corresponding neutrinos νe, νμ, ντ. These particles whose constituents include a quark which carries strangeness −1, namely the strange quark s. For example, particles we mentioned in the introduction have quark assignments: \(P = (uud), N = (ddu), \pi^+ = (u\bar{d}), K^+ = (u\bar{s}), \Lambda^0 = (sdu), \Sigma^+ = (usu), \) and \(\Sigma^- = (sds).\) Here \(\hat{q}\) denotes the antiparticle of quark \(q.\) By studying \(\beta\) decays of nuclei, neutron and muon, it was found that weak interaction proceeds through \((V - A) \times (V - A)\) four-fermi interaction:

\[
\mathcal{L} = 4 \frac{G_F}{\sqrt{2}} J^\mu_\mu,
\]

where

\[
J_\mu = \bar{d} \gamma_\mu \gamma^\nu - u \mu \bar{\nu} \gamma_\nu + \bar{e} \gamma_\mu \gamma^\nu - e \nu_\nu.
\]

and \(\gamma_\nu = (1 - \gamma_3)/2.\) (In the definition of the current \(J_\mu, q\) denotes the quark field operator, and \(\hat{q}\) denotes \(q^\dagger\).) Matrix elements which cause beta decays are, for example:

\[
\mu \rightarrow e \nu_\nu \bar{\nu}_e:
\]

\[
\langle \nu_\nu \bar{\nu}_e | (\bar{\nu}_\mu \gamma_\mu \gamma^\nu - \mu) (\bar{\nu}_e \gamma^\nu - \nu_\nu) | \mu \rangle;
\]

\[
N \rightarrow Pe \bar{\nu}_e:
\]

\[
\langle (nu)e\bar{\nu}_e | (\bar{\nu}_\nu \gamma_\nu \gamma^\nu - d) (\bar{\nu}_e \gamma^\nu - \nu_\nu) | (du)u \rangle.
\]

The Lagrangian given in eq. (4) does not allow strange particles to decay. The strangeness changing interaction was introduced in eq. (4) in an ingenious way by Gell-Mann and Levy, which was verified by Cabibbo. The idea was to replace \(d\) which appears in \(\bar{d} \gamma_\mu \gamma^\nu - u \mu \bar{\nu} \gamma_\nu + \bar{e} \gamma_\mu \gamma^\nu - e \nu_\nu\) by a rotated state: \(d' = d \cos \theta_c + s \sin \theta_c,\) where \(\theta_c\) is the Cabibbo angle. In today’s language, this replacement means that there is a distinction between mass eigenstate of quarks, \(d\) and \(s\) and a state which participates in weak interaction \(d'.\) This state is called a weak eigenstate.

The current can now be written as

\[
J_\mu = J_{\mu}^{S = 0} + J_{\mu}^{S = 1} + \bar{\nu}_\mu \gamma_\mu \gamma^\nu - \nu_\nu + \bar{e} \gamma_\mu \gamma^\nu - e \nu_\nu.
\]

Here \(J_{\mu}^{S = 0} = \cos \theta_c \bar{d} \gamma_\mu \gamma^\nu - u \mu \bar{\nu} \gamma_\nu + \bar{e} \gamma_\mu \gamma^\nu - e \nu_\nu\) is a strangeness conserving current, \(J_{\mu}^{S = 1} = \sin \theta_c \bar{s} \gamma_\mu \gamma^\nu - s \mu \bar{\nu} \gamma_\nu + \bar{e} \gamma_\mu \gamma^\nu - e \nu_\nu\) is a strangeness changing current. Strangeness changing \(K\) decays are, for example:

\[
K^- \rightarrow \pi^0 \mu^+ \bar{\nu}_\mu:
\]

\[
\langle (\bar{\pi}) (\mu^+ \bar{\nu}_\mu) | (\bar{\mu} \gamma_\mu \gamma^\nu - s \bar{\nu} \gamma_\nu) | (\bar{\nu} \nu_\nu) \rangle;
\]

\[
K^0 \rightarrow \pi^+ \pi^-:
\]

\[
\langle (\bar{u} \bar{d}) | (\bar{\mu} \gamma_\mu \gamma^\nu - s \bar{\nu} \gamma_\nu) | (\bar{s} \bar{d}) \rangle.
\]

3. Flavor Physics in Terms of Quarks

Today, we know that we can make all hadrons in terms of the quark model. There are six quarks and six leptons as listed in Table I. The strange particles are those particles whose constituents include a quark which carries strangeness −1, namely the strange quark s. For example, particles we mentioned in the introduction have quark assignments: \(P = (uud), N = (ddu), \pi^+ = (u\bar{d}), K^+ = (u\bar{s}), \Lambda^0 = (sdu), \Sigma^+ = (usu), \) and \(\Sigma^- = (sds).\) Here \(\hat{q}\) denotes the antiparticle of quark \(q.\) By studying \(\beta\) decays of nuclei, neutron and muon, it was found that weak interaction proceeds through \((V - A) \times (V - A)\) four-fermi interaction:

The weak interaction Lagrangian given in eq. (4) violates \(P\) maximally, as the current consists of both vector and axial vector components \((V - A).\) It can shown that the charge conjugation symmetry, which transforms a particle state to the corresponding antiparticle state \(C\) is also violated in such a way that the combined symmetry \(CP\) is conserved.

4. The Standard Model

This theory of weak interaction, which we have been discussing, involves local four-fermi interaction, and it is unrenormalizable. We can imagine that the four-fermi interaction is generated by an exchange of a heavy charged spin 1 boson. But a field theory which contains massive spin one charged boson is known to be also unrenormalizable. The break-through came when Weinberg wrote both weak and electromagnetic interaction in terms of \(SU(2) \times U(1)\) gauge theory. The Lagrangian of this theory contains no mass terms for the gauge bosons. Their masses are generated when the neutral Higgs’ boson breaks the gauge symmetry spontaneously. In this process gauge bosons become massive. This dynamics also generates all quark masses. Since field theory with massless spin one particles is renormalizable, this theory could have a chance of being renormalizable and contain massive spin one bosons at the same time. Indeed, this has been shown to be the case.

Notice that in Table I, there is a family of leptons associated with each family of quarks. In this gauge theory, we need to consider leptons and quarks together in the family structure, otherwise infinities from anomalies arise. So, flavor theory predicts one-to-one correspondence between quarks and leptons.

Because all experimental facts are consistent with this theory, it is called the standard model (SM). No matter how difficult it is to find deviation from this theory, we know that it is not the most fundamental theory. There must be a theory to which the SM is the low energy effective theory. This is because we expect the ultimate theory to give answers to all fundamental questions. Some of the questions are as follows. Why are there three generations? Why \(m_u/m_s \sim 3 \times 10^3?\) Why neutrinos are almost massless? Why do quarks have the observed mass structure, for example, \(m_u = 5\) MeV and \(m_s = 178\) GeV?

Flavor physics tries to find answers to these questions by searching for deviations from the SM which will eventually lead us to the more fundamental theory.

5. Quantum Mechanics of Particle–Antiparticle Mixing

\(K^0\) and \(\bar{K}^0\) mesons are bound states of \(s\bar{d}\) and \(d\bar{s}\) respectively. Experimentally, they both decay to \(\pi^+\pi^-\) states. Therefore we can imagine an interaction \(K^0 \leftrightarrow \pi^+\pi^- \leftrightarrow \bar{K}^0.\) If they can turn into each other through weak interaction, they are not eigenstates of the Hamiltonian.

There are many decay channels for the \(K\) meson and the Hamiltonian is an infinite dimensional matrix. Diagonalizing it is almost impossible. However, Weisskopf–Wigner approximation comes to our rescue.

Let us assume that we have an initial state, which is made out of some linear combination of only \(K^0\) and \(\bar{K}^0.\)
\[ \Psi(0) = a(0)|K^0\rangle + b(0)|\bar{K}^0\rangle. \] (9)

This state will evolve into some admixture of \( K^0, \bar{K}^0, \) and other hadrons as time goes by. If we are interested only in \( K^0 \) and \( \bar{K}^0 \) components at some later time, define

\[ \Psi(t) = a(t)|K^0\rangle + b(t)|\bar{K}^0\rangle \equiv \begin{pmatrix} a(t) \\ b(t) \end{pmatrix}. \] (10)

The result of Weisskopf–Wigner approximation states that, at some time \( t \) which is large compared to the strong interaction time scale, \( \Psi(t) \) obeys the following Schrödinger equation:

\[ i\hbar \frac{\partial}{\partial t} \Psi = H\Psi, \] (11)

with

\[ H = \begin{pmatrix} \langle K^0 | H | K^0 \rangle & \langle K^0 | H | \bar{K}^0 \rangle \\ \langle \bar{K}^0 | H | K^0 \rangle & \langle \bar{K}^0 | H | \bar{K}^0 \rangle \end{pmatrix}. \] (12)

Let us first show that CP symmetry implies \( \langle K^0 | H | \bar{K}^0 \rangle = \langle \bar{K}^0 | H | K^0 \rangle = \Delta \). Using the unitarity of CP operator, we have \( CP^2 CP = CP^{-1} CP = I \). Then inserting these unit operators:

\[ \langle \bar{K}^0 | H | K^0 \rangle = \langle \bar{K}^0 | CP^2 CP | K^0 \rangle = \langle \bar{K}^0 | H | K^0 \rangle, \] (13)

where we have used

\[ CP|K^0\rangle = |\bar{K}^0\rangle, \] (14)

and the statement of CP conservation: \([H, CP] = 0\).

Let us now show that \( \Delta \) is complex and thus \( H \) is not hermitian. Note that \( \Delta \) contains an absorptive part. That is, for example, \( \Delta \) gets contribution from the intermediate state \( K^0 \to \pi^+\pi^- \to \bar{K}^0 \equiv \Delta_{\pi^+\pi^-} \). This contribution is complex due to the scattering of \( \pi^+\pi^- \) intermediate state. It can be shown that \( \Delta_{\pi^+\pi^-} = e^{2i\delta_0} \Delta_{\pi^+\pi^-} \), where \( \delta_0 \) is the isospin 0, S wave \( \pi\pi \) phase shift. So, \( H \) is not hermitian and it leads to the decay of the initial state \( \Psi(0) \).

As mentioned above, a typical transition which contributes to \( \Delta \) is \( K^0 \to \pi^+\pi^- \to \bar{K}^0 \) and it is second order in weak interaction. Therefore, \( \Delta \) is truly infinitesimal compared to the diagonal component which is of order \( M_K \). Nevertheless it dictates the eigenstates to be

\[ |K_1\rangle = \frac{1}{\sqrt{2}} (|K^0\rangle + |\bar{K}^0\rangle), \]
\[ |K_2\rangle = \frac{1}{\sqrt{2}} (|K^0\rangle - |\bar{K}^0\rangle). \] (15)

as long as \( |\Delta| \gg |\langle K^0 | H | K^0 \rangle - \langle \bar{K}^0 | H | \bar{K}^0 \rangle| \).

Using eq. (14), we have \( CP|K_1\rangle = \pm |K_1\rangle \). Now, consider a \( |\pi^+\pi^-\rangle \) state. \( |\pi^+\pi^-\rangle = |\pi^-\pi^+\rangle \), and \( CP|\pi^+\pi^-\rangle = |\pi^-\pi^+\rangle \). So, we have

\[ CP|\pi^+\pi^-\rangle = +|\pi^+\pi^-\rangle. \] (16)

Therefore \( K_1 \) has the same CP quantum number and it can decay to \( 2\pi \), i.e.,

\[ K_1 \to 2\pi, \]
\[ K_2 \to 2\pi. \] (17)

The leading nonleptonic channel for \( K_2 \) is then

\[ K_2 \to 3\pi. \] (18)

The phase space for eq. (18) is very restricted compared to the \( 2\pi \) decay — \( 3 \times M_2 \simeq 420 \text{ MeV} \) vs \( M(K_2) \simeq 500 \text{ MeV} \). Thus we expect the lifetime for the CP odd state \( K_2 \) to be much longer than for the CP even state \( K_1 \). Since \( K_1 \) and \( K_2 \) possess quite different lifetimes, it is customary to refer to them as \( K_S \) and \( K_L \), respectively, with \( S \) (L) referring to short-lived (long-lived). Likewise we use \( M(K_1) = M(K_S) = M_S \) and \( M(K_2) = M(K_L) = M_L \). State-of-the-art measurements gives:

\[ \tau_S = \tau(K_S) = 0.593 \pm 0.0006 \times 10^{-10} \text{ s}, \] (19)
\[ \tau_L = \tau(K_L) = (5.098 \pm 0.021) \times 10^{-8} \text{ s}. \]

It should be kept in mind, though, that this huge difference in lifetimes for neutral kaons. Fun is not over yet!

6. CP-Invariant Particle–Antiparticle Oscillations

So, we have seen the existence of two vastly different lifetimes for neutral kaons. Fun is not over yet!

For an initially pure \( K^0 \) beam the time evolution for a \( |K^0(t)\rangle \) is

\[ |K^0(t)\rangle = \frac{1}{\sqrt{2}} (|K_S(t)\rangle + |K_L(t)\rangle) = \frac{1}{\sqrt{2}} \left[ \exp \left[ -i \left( M_S - \frac{i}{2} \Gamma_S \right) t \right] |K_S(0)\rangle + \exp \left[ -i \left( M_L - \frac{i}{2} \Gamma_L \right) t \right] |K_L(0)\rangle \right]. \] (20)

where

\[ f_\pm(t) = \frac{1}{2} e^{-i M_{\pm t} e^{-i(1/2)\Gamma t}} \left[ 1 \pm e^{-i \Delta M t} e^{(1/2)\Delta \Gamma t} \right]. \] (21)

where \( \Delta M = M_2 - M_1 \), and \( \Delta \Gamma = \Gamma_2 - \Gamma_1 \), and likewise for \( |\bar{K}^0(t)\rangle \).

From this expression for the time dependent admixtures, we can compute the probability of finding a \( K^0 \) and \( \bar{K}^0 \) in an initially pure \( K^0 \) beam. We have shown these probabilities in Fig. 1.

\[ \text{Fig. 1. The probability of finding } K^0 \text{ in an initial } K^0 \text{ beam as a function of time, and the probability of finding } \bar{K}^0 \text{ in the same beam.} \]
If there is a CP violation and it is caused by the deviation from CP
symmetry requirement $\langle K^0 | H | \bar{K}^0 \rangle = \langle \bar{K}^0 | H | K^0 \rangle$, and
the second term is called the direct CP violation. If $\bar{\epsilon}$ is the
only source of CP violation, i.e., without direct CP violation, we expect $\eta_{+-} = \eta_{00}$
where
$$\eta_{00} = \frac{\Gamma(K_L \to \pi^0 \pi^0)}{\Gamma(K_S \to \pi^0 \pi^0)}.$$
Since the discovery of CP violation in $K$ decay, many experimentalists
have searched for the possible presence of the second term in
eq. (25). It will cause $\eta_{+-} \neq \eta_{00}$. The term commonly used
by physicists in this field is
$$\frac{\epsilon'}{\epsilon} = \frac{1}{6} \frac{|\eta_{+-}|^2}{|\eta_{00}|^2}.$$
(27)
The first measurement for this quantity was made by
the NA31 collaboration\cite{12} 29 years after the measurement of
$\eta_{+-}$. Recent values obtained by NA48\cite{13} and KTeV\cite{14}
collaborations averaged by PDG\cite{10} is
$$\frac{\epsilon'}{\epsilon} = (16.7 \pm 2.3) \times 10^{-4}.$$
(28)

8. Symmetries Gone Away Side
So, the parity symmetry was the first to go. Now CP
symmetry is gone. Through detailed studies of $K$ meson
decays, experiments have shown that the rate for $K^0 \to \bar{K}^0$
is different from the rate for $\bar{K}^0 \to K^0$. Thus $T$ symmetry
violation is also established.\cite{5}
$$A_T(t) = \frac{\Gamma(K^0 \to K^0(t)) - \Gamma(K^0 \to \bar{K}^0(t))}{\Gamma(K^0 \to K^0(t)) + \Gamma(K^0 \to \bar{K}^0(t))}
= (6.6 \pm 1.3 \pm 1.0) \times 10^{-3}.$$
(29)
Thus we have ended up with a curious situation. The sacred
symmetries of classical mechanics: parity and time reversal,
as well as particle antiparticle symmetry which is seen to be preserved in strong and electromagnetic interaction are violated in weak interaction.

In looking for physics beyond the SM, we should always keep in mind that the ultimate symmetry which is valid in any field theory which satisfies locality, \textit{CPT} symmetry may be violated at some level. So far, CPLEAR result is:\(^{(15)}\)

\[
A_{\text{CPT}}(t) = \frac{\Gamma(K^0 \rightarrow K^0(t)) - \Gamma(K^0 \rightarrow K^0(t))}{\Gamma(K^0 \rightarrow K^0(t)) + \Gamma(K^0 \rightarrow K^0(t))} = (0.07 \pm 0.50 \pm 0.45) \times 10^{-3}. \tag{30}
\]

### 9. Rare Decays of \(K_L\)

In the previous section, we have considered charged currents in the form of \(\bar{d} \gamma_\nu \gamma^- u\). While the charged currents were necessary to describe \(\beta\) decays, there is no reason why neutral current should not play a role. In fact in the standard model, neutral current which couples to the \(Z^0\) boson naturally exists. So, let us consider a neutral current of the form \(\bar{u} \gamma_\nu \gamma^- u + \bar{d} \gamma_\nu \gamma^- d\). The second term must be \(d'\) rather than \(d\), because \(d'\) contributes. The neutral component of the hadronic current now looks like:

\[
\frac{G_F}{\sqrt{2}} \sin \theta_c \cos \theta_c \langle \bar{d} \gamma_\nu \gamma^- s \rangle \langle \bar{u} \gamma_\nu \gamma^- u \rangle + \text{h.c.} \tag{31}
\]

Now let us look at the data. The first two decay modes shown in Table II are typical strangeness changing weak decays. But, look at the last decay, eq. (31) causes a transition

\[
\langle \bar{d} \gamma_\nu \gamma^- s \rangle \langle \bar{u} \gamma_\nu \gamma^- u \rangle \tag{32}
\]

and this decay should have a branching ratio of \(O(1\%)\) in stead of \(10^{-9}\). Why is it so suppressed? It implies that the neutral current interaction of the form eq. (31) can not exist.

This puzzle has an elegant solution proposed by Glashow, Iliopolous and Maiani.\(^{(16)}\) They assumed that there was a new quark named charm. Just as \((u, d')\) contributed in weak interactions, \((c, s')\) with \(s' = -d \sin \theta_c + s \cos \theta_c\), which corresponds to the orthogonal state to \(d'\), also should contribute. The neutral component of the hadronic current now looks like:

\[
J_{\mu}^0 = \bar{u} \gamma_\mu \gamma^- u + \bar{d} \gamma_\mu \gamma^- d + \bar{s} \gamma_\mu \gamma^- s + \bar{c} \gamma_\mu \gamma^- c
\]

\[
= \bar{u} \gamma_\mu \gamma^- u + \bar{c} \gamma_\mu \gamma^- c + \bar{d} \gamma_\mu \gamma^- d + \bar{s} \gamma_\mu \gamma^- s. \tag{33}
\]

The unwanted strangeness changing neutral current has been canceled out. A simple way to see this is to say that

\[
\begin{pmatrix}
  d' \\
  s'
\end{pmatrix}_{\text{L}} = V_L \begin{pmatrix}
  d \\
  s
\end{pmatrix}_{\text{L}} = \begin{pmatrix}
  \cos \theta_c & \sin \theta_c \\
  -\sin \theta_c & \cos \theta_c
\end{pmatrix} \begin{pmatrix}
  d \\
  s
\end{pmatrix}_{\text{L}} \tag{34}
\]

\[
\begin{pmatrix}
  d' \\
  s'
\end{pmatrix}_{\text{L}} \gamma_\mu \gamma^- \begin{pmatrix}
  d' \\
  s'
\end{pmatrix}_{\text{L}} = \begin{pmatrix}
  d \\
  s
\end{pmatrix}_{\text{L}} \gamma_\mu \gamma^- \begin{pmatrix}
  d \\
  s
\end{pmatrix}_{\text{L}} \tag{35}
\]

where we have used the fact that \(V_L\) is unitary.

So, by insisting on the absence of neutral strangeness changing current, again, flavor physics managed to predict the existence of a new quark, charm. This is the power of flavor physics. But it is not over yet!

### 10. KM Mechanism

The fact that \(\text{CP}\) violation must be due to the presence of a phase in the Hamiltonian can be easily seen as follows. Write the Hamiltonian as

\[
H = ch + c^* h', \tag{36}
\]

where \(h\) is some operator which causes a transition, and \(c\) is a complex coefficient. Note that the second term must be present due to the hermiticity of the Hamiltonian \(H\). It can be shown that if \(h\) describes interaction of particles, \(h'\) describes interaction of antiparticles, and

\[
\text{CPhCP}^{-1} = h'. \tag{37}
\]

So, \(H\) in invariant under the \(\text{CP}\) transformation \([\text{H}, \text{CP}] = 0\) if and only if \(c\) is real.

Since its experimental discovery, many theoretical ideas have been put forward to explain \(\text{CP}\) violation. Among them, multi-Higgs models, superweak theory, etc. The simplest version of the multi-Higgs model predicted \(\epsilon'/\epsilon \approx 0.05\)\(^{(17)}\) which is excluded by experiments. Superweak model predicted \(\epsilon'/\epsilon = 0\), which is again excluded by experiments. The Kobayashi–Maskawa (KM) ansatz escaped scrutiny by these measurements. In fact its prediction of \(\epsilon'/\epsilon\) is consistent with experiment, although theoretical prediction is plagued with uncertainties coming from the fact that we are unable to reliably compute strong interaction effects. In this section we describe the KM ansatz.\(^{(18)}\)

We stated that \(\text{CP}\) violation originates from the phase in the interaction Lagrangian. Could \(\text{CP}\) violating phase appear naturally in the SM? The answer is yes!

Let us examine how quarks couple to \(W\) bosons, and how quark masses arise. The Higgs’ boson interacts with quarks through Yukawa interaction:

\[
L_{\text{Yukawa}} = \sum_{i,j} (\bar{G}_U)_{ij}(\bar{U}_{\text{L}})_{ij}(\phi^0_{-})_{ij} U_{\text{R}}, \tag{38}
\]

where \(U_{\text{L}}^T = (u, c, t)^T\) and \(D_{\text{L}}^T = (d', s', b)^T\), where \(q^0\) denotes the rotated weak eigenstates. The indices \(i\) and \(j\) run over 1 to 3, the number of families. Once the neutral Higgs field acquires a vacuum expectation value, \(\phi^0 = v\), the mass matrices for the up- and down-type quarks are then proportional to the corresponding Yukawa couplings with the scale set by \(v\):

\[
M_U = vG_U \quad \text{and} \quad M_D = vG_D. \tag{39}
\]

Since the Yukawa couplings are quite arbitrary, so are the mass matrices, and in general they will contain complex elements.

To describe \(\text{CP}\) violation, there must be complex phases in the Hamiltonian. So, can the phases in these mass matrices be the origin of \(\text{CP}\) violation?
To investigate this, let us diagonalize the mass matrix:
\[
V_{\text{ud}} M_U V_{\text{ur}}^T = M_{U}^{\text{diag}},
\]
\[
V_{\text{dl}} M_D V_{\text{dr}}^T = M_{D}^{\text{diag}},
\]
where \(U_{\text{ur}}^T = (\mathbf{u}, \mathbf{c}, \mathbf{t})\gamma_5\) and \(D_{\text{dr}}^T = (\mathbf{d}, \mathbf{s}, \mathbf{b})\gamma_5\) are the mass eigenstates, and \(M_{U/D}^{\text{diag}}\) is the diagonalized down-type mass matrix, and up-type quark mass matrix, respectively. The mass eigenstates can be written in terms of weak eigenstates:
\[
U_L^m = V_L^u U_u, \quad D^m_L = V^L_d D_L.
\]

Now we write the Lagrangian in terms of particles that participate in interactions, i.e., mass eigenstates of quarks. Then the Lagrangian becomes
\[
\mathcal{L} = \bar{U}^m_L \mathcal{V}_L^u \bar{U}_d^m L^m + \bar{U}^m_L \mathcal{V}_L \bar{U}_u^m D^m R + \text{h.c.}
\]

There are plenty of phases in \(V \equiv V_L^u V_d^m\), which we call the KM matrix. Do they cause CP violating interaction? The problem is, as stated before, that experiments can only count number of particles. In quantities that can be measured, most of the phase information is lost. In particular, the phase of external states are not observable. So, we can adjust them to make the constants which appear in the Lagrangian real. Make the following transformation which does not change result of any experiment:
\[
U_L^m \rightarrow \text{diag}(e^{i\phi}, e^{i\phi_b}, e^{i\phi_c}) U_L^m,
\]
\[
D^m_L \rightarrow \text{diag}(e^{i\phi_b}, e^{i\phi_c}, e^{i\phi_d}) D^m_L,
\]
where \(\text{diag}(a, b, c)\) denotes a diagonal matrix with diagonal elements \(a, b, c\). Of course, the phases of the right handed quarks should be adjusted accordingly so that the masses remain real.

Can we tune quark phases to make all the elements of \(V\) real? Note that the number of parameters in \(V\) rises quadratically and the number of adjustable phases increases linearly with number of generations. For two generations, there are enough phases we can adjust to make \(V\) real. But, for 3 generations, there are not enough free phases to make all \(V_{ij}\) real. For a system of three generation of quarks, there will be one unremovable phase. This phase may appear in physical observables.

The existence of CP violation is a natural consequence of fact that: (1) Yukawa couplings are in general complex, and (2) there exist three generations of quarks which mix through weak interaction to all other quarks of the same charge. For three generations, there is one phase and three mixing angles in the KM matrix. All we have to do is to figure out how to measure it.

11. Why is K Meson CP Violation Small?

We have argued in the previous section that if there exist only two generations, there is no CP violation within the context of the KM ansatz. This means that for CP violating K meson decays, the third generation must play a crucial role. We know from quantum mechanics, the second order perturbative transition
\[
K^0 \rightarrow \bar{\nu} \rightarrow \bar{K}^0
\]
can occur, if there is a coupling \(K^0 \rightarrow \bar{\nu}\), and \(\bar{\nu} \rightarrow \bar{K}^0\). Remember that, in the second order perturbation theory, the intermediate state may have a different energy compared to the initial and final state energy. Indeed for K meson decay, \(\bar{\nu}\) state is the major contribution to the phase necessary for the CP violating decay \(K_L \rightarrow \pi^+\pi^-\).

We thus understand why CP violation is small in K decays. The mass of the K meson is only 500 MeV while the intermediate state of \(\bar{\nu}\) is 350 GeV. The amplitude is highly suppressed due to the fact that it is proportional to the inverse of the energy difference between initial and the intermediate state.

12. How Do We Extract Phases by Counting Number of Particles?

In physics experiments, we shoot in particles and they scatter, then we just count number of final state particles. In other words, output of experiments are just integers. We have seen that CP violation effects are controlled by phases of amplitudes. How could information about phases be obtained from integers?

Measuring intensities of light passing through two slits, Young’s double slit experiment allows us to measure phase difference of two waves passing through these slits. Intensities is proportional to the number of photons hitting the screen, i.e., integers.

Now, consider a decay in which there are two different decay paths. Experimenters can not tell which decay path has been taken, as long as they measure just the final state decay products. Having two different decay paths is analogous to having two different slits. The amplitudes for these two decay paths will interfere. Let us assume that the \(B^0\)–\(\bar{B}^0\) mixing is big enough so that their effect can be seen. Then there are two paths for a \(B^0\) meson to decay to \(f\), as shown in Fig. 3. Since we can not tell which path a particular decay took, we have to coherently add two amplitudes. These two amplitudes interfere with each other when we square the total amplitude to obtain the probabilities for the \(B^0 \rightarrow f\) and \(\bar{B}^0 \rightarrow f\) decays. For a specific decay mode \(f = \psi K_S\), we obtain:\n
\[
\Gamma(B(t) \rightarrow J/\psi K_S) = \Gamma(B(t) \rightarrow J/\psi K_S)
\]
\[
\Gamma(B(t) \rightarrow J/\psi K_S) + \Gamma(B(t) \rightarrow J/\psi K_S)
\]
\[
= -\text{Im} \left( \sum_{c} V_{ud} V_{ub} V_{ts} \sin \left( \frac{\Delta M_B \tau}{2} \right) \right)
\]

There is a nice way to summarize CP violation parameters. Note that the CKM matrix is a unitary matrix. So, we have

Fig. 3. With \(B^0\)–\(\bar{B}^0\) mixing and non vanishing amplitudes for both \(B \rightarrow f\) and \(\bar{B} \rightarrow f\) decays there are two decay chains \(B \rightarrow f\) and \(\bar{B} \rightarrow f\) just as in the two-slit interference pattern observed in Young’s experiment in optics, these two decay chains interfere according to the principle of quantum mechanics.
If we represent these three complex numbers as vectors on a complex plane, the unitarity relation is represented by a triangle shown in Fig. 4. One of the angles $\phi_1$ is given in terms of the asymmetry given in eq. (45):

$$\sin(2\phi_1) = \frac{m_s V_{ub}^* V_{cd} - m_c V_{ud}^* V_{cb}}{m_s V_{ub}^* V_{cd} + m_c V_{ud}^* V_{cb}}.$$  

(47)

Also it can be shown that $\phi_2$ is obtained from the CP asymmetry in the $B \rightarrow \pi \pi$ channel, and $\phi_3$ can also be obtained, for example, from the asymmetry in $B \rightarrow K \pi$ channel. Three quantities: like three sides of the triangle; two angles and one side; two sides and one angle; fix the triangle uniquely, and there are six experimentally measurable quantities: 3 angles and 3 sides of the triangle. The KM theory can therefore be tested by performing three consistency checks.

To arrive at our result eq. (45), we assumed that: (1) B meson exists; (2) there exists $B^0 - \bar{B}^0$ transition; (3) the B meson life time is long enough so that the time dependence of the asymmetry can be seen; (4) the phase of the KM matrix elements must be sizable for the asymmetry to be observed. With all these “ifs”, there is no wonder that nobody paid much attention to this prediction for about 7 years.

13. Discovery of Beauty that Mixes

When Kobayashi and Maskawa proposed the KM ansatz for CP violation, only two families where known to exist. But, as we have seen in §10, with only two families, all phases which appear in the mass matrix due to Yukawa interactions of Higgs’ boson can be removed by redefining the invisible quark phases. Their bold prediction was that the third family must exist.

Theorists predicted the existence of charm to complete the second family, and $J/\psi$ which is a bound state of charm quark and anti-charm quark was discovered in 1976. Theorists once again predicted the existence of the third family in order to explain CP violation. Sure enough the isodoublet of mesons ($B^+$, $B^0$), which are bound states of $b$ quark and ($u, d$), have been found in 1981.  

It was the MAC collaboration working at the PEP ring of SLAC that found the first evidence for a long beauty life time. The discovery was quickly followed by MARKII collaboration, also at PEPPII. They have observed that B decay products came not from where $B^0$ and $\bar{B}^0$ where produced, but from some distance away. It showed that B mesons traveled some observable distance before it decayed, i.e., their life times were long enough to leave a gap. To me, this was the crucial discovery which lead to many successes at the B factories 20 years later. Then came another crucial discovery, $B^0 - \bar{B}^0$ mixing. Let us take a brief trip into a memory lane. After the discovery of the B mesons, we knew that the second member of the third family had to exist. It took more than a decade of search, but sure enough, the top quark was discovered in 1995. For a long time there was a rumor that the top quark was about to be found, and its mass was at about 50 GeV. This value is huge for those who thought that B meson mass of 5 GeV was extraordinary. Theoretical calculation showed that the probability for observing mixing is proportional to $(m_t/m_w)^4$, and we thought it was too small to be observed if $m_t = 50$ GeV. If we put the correct value of $m_t \sim 174$ GeV into our old computations, it is exactly what had been observed. If we had been bold enough, we could have predicted $B^0 - \bar{B}^0$ mixing before its discovery. In spite of the fact that theorists lacked courage to bring out such a bold prediction, the ARGUS collaboration announced that they have seen it! The role has been reversed. Having seen $B^0 - \bar{B}^0$ mixing, we knew that the top quark should be found around $m_t \sim 170$ GeV. This again illustrates the power of flavor physics as long as those who work in it are courageous enough.

So, everything that was assumed in predicting the CP asymmetry in $B \rightarrow J/\psi K_S$ have been verified by experiments. The next step in understanding the CP violating phenomena is to experimentally check this prediction.

14. How Do We Actually Measure CP Violation?

First, we have to have beams of $B^+$’s and $\bar{B}^+$’s. The problem is that $B^+$’s live only for about 1.5 ps. This is so short that even light travels only 0.45 mm during this time. Before $B^+$’s decay, we have to check if we are looking at a $B^0$ decay or $\bar{B}^0$ decay, identify the decay product, and measure the time when it decayed. The way to produce $B^+$’s is to pair produce it:

$$e^+ + e^- \rightarrow B^0 + \bar{B}^0.$$  

(48)

To identify $B^0$ and $\bar{B}^0$, we take advantage of the fact that leptonic decay of $b$ ($\bar{b}$) quark always contains $\mu^-$ ($\mu^+$): $b \rightarrow c + \mu^- + \bar{\nu}_\mu$, and $\bar{b} \rightarrow \bar{c} + \mu^+ + v$. Since $B^0$ ($\bar{B}^0$) is a bound state of $bd$ ($\bar{b}\bar{d}$), leptonic decay of $B^0$ always contains $\mu^+$ and that of $\bar{B}^0$ always contains $\mu^-$.  

$$B^0 \rightarrow \mu^+ + \nu_\mu + \text{hadrons}$$

$$\bar{B}^0 \rightarrow \mu^- + \bar{\nu}_\mu + \text{hadrons}.$$  

(49)

Now, note that pair production eq. (48) always produce $B^0\bar{B}^0$ pair in the relative angular momentum 1 state, i.e., the wave function is antisymmetric in the interchange of $B^0 \leftrightarrow \bar{B}^0$. So, if we identify, for example, the leptonic B decay which contains $\mu^+$ at time $t$, then we know that, at that time $t$, the other member of the pair production was a $\bar{B}^0$. This way we managed to produce the $B^0$ beam. The $T(4S)$ is most suited for studying CP violation in B decays. At the top of the resonance, the signal to noise ratio for $B\bar{B}$ production is increased by 1 : 2.5.

We have estimated, back in 1981, that the asymmetry shown in eq. (45) is:

$$\sin(2\phi_1) > 0.15.$$

(50)
respectively. For example, experimental measurement of $\epsilon \sim 2 \times 10^{-3}$. Based on this value we estimated that we needed 100 million $\bar{B}^0 - B^0$ pairs to be produced in one year if we were to discover large CP violation during that time. This meant that we have to have a $e^+ - e^-$ collider with the luminosity of $10^{34} \text{cm}^{-2} \text{s}^{-1}$. This luminosity was 1000 times more than that of the existing state of the art machine, CESR. Not only that, the machine had to collide electrons at 9 GeV and positron at 3 GeV — it had to be an asymmetric machine. This was to boost $\tau(45)$ so that $B - \bar{B}$ pair is moving fast enough to leave a track of about 200 $\mu$m before it decayed.

### 15. Recent Results

Both KEK and SLAC have accepted this challenge in building the B factory. Sure enough they have both reached the design goal in a reasonable time period, and the CP asymmetry was discovered simultaneously at both laboratories in the year 2000. The latest number is $25)$. In addition to this result, there are many more interesting results are listed in the above mentioned reference.

![Diagram of the unitarity triangle](image)

Fig. 5. The unitarity triangle representing the unitarity relation eq. (46).

Note that this is a huge asymmetry compared to $\epsilon \sim 2 \times 10^{-3}$. Based on this value we estimated that we needed 100 million $\bar{B}^0 - B^0$ pairs to be produced in one year if we were to discover large CP violation during that time. This meant that we have to have a $e^+ - e^-$ collider with the luminosity of $10^{34} \text{cm}^{-2} \text{s}^{-1}$. This luminosity was 1000 times more than that of the existing state of the art machine, CESR. Not only that, the machine had to collide electrons at 9 GeV and positron at 3 GeV — it had to be an asymmetric machine. This was to boost $\tau(45)$ so that $B - \bar{B}$ pair is moving fast enough to leave a track of about 200 $\mu$m before it decayed.

$$\sin(2\phi_1) = 0.674 \pm 0.026.$$  \hspace{1cm} (51)

This is large compared to the lower limit we used, 15%. Since there is only one phase in the KM matrix, all measurements associated with CP violation can be represented by a point on the complex $\rho - \eta$ plane, where $\rho$, and $\eta$ are real, and imaginary part of the complex number

$$\frac{V_{ub}}{V_{cb}} = \rho + i\eta,$$  \hspace{1cm} (52)

respectively. For example, experimental measurement of $|V_{ub}/V_{cb}|$ gives a region which is a circle, with the center at the origin of the $\rho - \eta$ plane. $B^0 - \bar{B}^0$ mass mixing gives a region which is also a circle centered at $(1, 0)$ on the $\rho - \eta$ plane. The measurement of $\phi_1$ defines a unitarity triangle where the base of the triangle is 1. This is because all three sides of the triangle is divided by $|V_{cb}|$. The most recent published result which summarizes various experimental results is shown in Fig. 5. In addition to this result, there are many more interesting results are listed in the above mentioned reference.

### 16. Future

B physics is still in a very early stage. Remember K physics has been an active area of research for already 60 years. This implies that B physics, if it started in 1981 with its discovery, should continue to yield interesting results at least until 2046. The reason why B factories have been such a success is that the luminosity was set at $10^{34} \text{cm}^{-2} \text{s}^{-1}$ which is 10 times more than we needed to measure the asymmetry. Thus it is yielding much more interesting results than just the value of CP violation in $B \rightarrow \psi K_S$, $\sin(2\phi_1)$. This should be a lesson for future accelerators. The luminosity should be much as we can possibly obtain.

Physics of high enery scale should affect the way B decays through quantum corrections. Such new physics effects should be small but it may be the only way to get at the new physics if its scale is beyond the existing accelerators.

3) I. I. Bigi and A. I. Sanda: CP Violation (Cambridge University Press, 2000). For details and for other very important topics which were omitted for lack of space, we refer the reader to this text book.
Ichiro Sanda was born in Tokyo in 1944. He obtained B. Eng. (1965) from University of Illinois, and Ph. D. (1969) in physics from Princeton University. He was a research associate (1969–1974) at Columbia University and at Fermilab; an assistant professor and an associate professor (1974–1992) at Rockefeller University; a professor (1992–2006) at Nagoya University; and now a professor at Kanagawa University since 2006. He has worked in the field of elementary particle physics. In particular, he has worked on various theoretical aspects of CP violation.
1. Introduction
The history of the kaon system in the genesis of the Standard Model (SM) is well-known and for many years there was intense experimental activity in this area. But after a series of attempts to uncover a non-Superweak\(^1\) origin for the observed CP-violation failed to do so,\(^2,3\) experimental efforts in this field diminished to a much lower level for a decade. Interest revived only gradually, primarily driven by three developments:

1. The realization that the Kobayashi–Maskawa model\(^4\) of CP-violation implied a potentially observable deviation from the Superweak prediction\(^5\) for $K \rightarrow 2 \pi$\(^6\)
2. A developing appreciation for the fact that GIM suppressed\(^7\) kaon decays were potentially cleanly sensitive to SM parameters, including those characterizing CP-violation.\(^8–14\)
3. As theories were developed to address the perceived shortcomings of the SM, it turned out that many of these naturally predicted violations of lepton flavor (as in $K_L \rightarrow \mu^± \nu^±$) or other conservation laws in rare kaon decays. Prominent among these were models of dynamic symmetry breaking, particularly those under the rubric of extended technicolor.\(^15\)

2. CP-Violation in $K \rightarrow 2 \pi$
Reference 6, which predicted a ratio $\text{Re}(\epsilon'/\epsilon)$ as large as $10^{-1–10^{-2}}$, stimulated experiments at Brookhaven\(^16\) and Fermilab\(^17\) that could have observed effects of this size.\(^18\) Later predictions were more conservative and more sensitive experiments at Fermilab (E731) and CERN (NA31) were launched to pursue this measurement. Each exploited the fact that

$$\text{Re}(\epsilon'/\epsilon) \approx \frac{1}{6} \left[ 1 - \frac{\Gamma(K_L \rightarrow \pi^+\pi^0)/\Gamma(K_S \rightarrow \pi^0\pi^0)}{\Gamma(K_L \rightarrow \pi^+\pi^-)/\Gamma(K_S \rightarrow \pi^+\pi^-)} \right]$$

(1)

This implies that the precision on $\text{Re}(\epsilon'/\epsilon)$ will be 6 times better than that on the ratio of rates. The practical application of this formula was quite different in the two experiments. After a series of runs, they eventually reported marginally conflicting results: $\text{Re}(\epsilon'/\epsilon) = (23 \pm 6.5) \times 10^{-4}$ for NA31\(^19\) and $\text{Re}(\epsilon'/\epsilon) = (7.4 \pm 5.2 \pm 2.9_{\text{stat}}) \times 10^{-4}$ for E731.\(^20\) These results are less than $2\sigma$ apart, but they had qualitatively different implications for the Superweak Model. This led to another set of experiment at both FNAL and CERN, led by veterans of E731 and NA31. These experiments represented large advances on the previous round.

Figures 1 and 2 show the KTeV and NA48 experiment respectively. Both experiments featured simultaneous $K_S$ and $K_L$ beams and detectors capable of measuring all four $K \rightarrow 2 \pi$ reactions at once. In case of KTeV the average kaon momentum was about 70 GeV/c and the two beams diverged by 1.6 mrad in the horizontal so that their centers which were 14 cm apart at the beam-defining collimator became separated by 30 cm at the detector. A regenerator in one beam provided the $K_S$’s and was switched between the two beams once per minute. For NA48, the $K_S$’s were created in a special target partway down the beam line. A small fraction of the uninteracted portion of the proton beam downstream of the $K_L$ target was diverted via a silicon mono-crystal, collimated, tagged and transmitted to the $K_S$ target. The two beams converged (vertically at an angle of 0.6 mrad) at the detector. In both experiments the decay region and associated veto counters were situated in large evacuated tanks. Downstream of the KTeV tank, helium bags separated the detector elements while in NA48, the detector elements to the end of the spectrometer were situated in a helium-filled volume. Running conditions varied over the life of each experiment, but typically they saw an instantaneous fluxes of a few MHz of $K_L$. Both had four-station drift chamber spectrometers followed by high-quality electromagnetic calorimeters. KTeV used an array of pure CsI crystals and NA48 a liquid Krypton accordion-type calorimeter. NA48 had a hadron calorimeter and both experiments had muon-identification systems. The high quality of the instrumentation needed for this demanding application allowed many other interesting measurements, some of which are discussed below.

In 1996 and 1997 KTeV collected 11.1M $K_L \rightarrow \pi^+\pi^-$ and 3.3M $K_L \rightarrow \pi^0\pi^0$ decays. The corresponding numbers of $K_S$ decays were 19.2M and 5.6M. Their overall result was $\text{Re}(\epsilon'/\epsilon) = (20.7 \pm 2.8) \times 10^{-4}$. There was a subsequent run in 1999 that doubled the statistics previously collected, but the result from this data is not yet available. In 1998 and 1999 Na48 collected 14.5M $K_L \rightarrow \pi^+\pi^-$ and 3.3M $K_L \rightarrow \pi^0\pi^0$ decays. The corresponding numbers of $K_S$ decays were 22.2M and 5.2M. Their overall result, when combined with that from a lower statistics 1997 run, was $\text{Re}(\epsilon'/\epsilon) = (15.3 \pm 2.6) \times 10^{-4}$. In 2001 NA48 had an additional run under somewhat different conditions in which they collected about half as many events as in 1998–99. They obtained a result $\text{Re}(\epsilon'/\epsilon) = (13.7 \pm 3.1) \times 10^{-4}$, which combined with previous running gave $\text{Re}(\epsilon'/\epsilon) = (14.7 \pm 2.2) \times 10^{-4}$. 

**Experimental Quark Flavor Physics: Kaon Physics**
Laurence S. Littenberg
Brookhaven National Laboratory, Upton, NY 11973, U.S.A.

(Received April 17, 2007; accepted July 3, 2007; published November 12, 2007)

Selected recent results on rare kaon decays are reviewed and prospects for on-going and future experiments are discussed.

**KEYWORDS:** kaons, decays

**DOI:** 10.1143/JPSJ.76.111006
At this point the Particle Data Group\textsuperscript{25) fits for kaon data result in $\Re(e'/e) = (16.6 \pm 2.6) \times 10^{-4}$, many $\alpha$ from 0, so that it is clear that the Superweak Model\textsuperscript{5) cannot be correct, and a quest lasting more than 30 years has been successfully concluded. Unfortunately the detailed implications of this result for the origin of CP violation are not clear. At the moment one cannot rule out a Standard Model explanation for this CP-violation, but that is about all one can say. There is an enormous literature attempting to calculate $\Re(e'/e)$ in the Standard Model, but little agreement on the resulting accuracy. At the point at which ref.\textsuperscript{24} was released, the range of up-to-date predictions varied from $-10 \times 10^{-4}$ to $+30 \times 10^{-4}$. References may be found in ref.\textsuperscript{24}.

A number of experiments including KTeV and NA48 have seen $\epsilon$-level CP-violating effects in other kaon decay modes including $K_L(\epsilon3, 26, 27) \to \pi^+\pi^-\gamma$, $K_L \to \pi^+\pi^-e^+e^-$, and $K_L \to K^-\pi^+\pi^-e^+e^-$.\textsuperscript{29)}

3. CP-Violating Amplitudes in Rare Kaon Decays

A number of years before the denouement of $K \to 2\pi$, it was appreciated that CP-violation could manifest itself in a cleaner way in certain rare kaon decays. These decays had contributions from G.I.M.-suppressed\textsuperscript{7) one-loop amplitudes sensitive to fundamental SM parameters such as $m_t$ and $V_{td}$. However it was not until it became clear that $m_t$ was very large that it was realized that these processes might be accessible to measurement. Even with the large $m_t$ these decays are also suppressed enough to be potentially very sensitive to BSM physics. These processes include $K^+ \to \pi^+\nu\bar{\nu}$, $K_L \to \mu^+\mu^-$, $K_L \to \pi^0\nu\bar{\nu}$, $K_L \to \pi^0e^+e^-$, and $K_L \to \pi^0\mu^+\mu^-$. In the latter three cases the one-loop contributions violate CP. In $K_L \to \pi^0\nu\bar{\nu}$ this contribution completely dominates the decay.\textsuperscript{13)} Diagrams for such loops are shown in Fig. 3. Since the GIM-mechanism enhances the contribution of heavy quarks, in the SM these decays are sensitive
to the product of couplings \(V_{ts}^* V_{td} \equiv \lambda_1\). Although it is perhaps most natural to write the branching ratio for these decays in terms of the real and imaginary parts of \(\lambda_1\), for comparison with results from the \(B\) system it is convenient to express them in terms of the Wolfenstein parameters, \(A\), \(\rho\), and \(\eta\). Figure 4 shows the relation of these rare kaon decays to the unitarity triangle. The dashed triangle is the usual one derived from \(V_{ub}^* V_{ud} + V_{cb}^* V_{cd} + V_{ub}^* V_{td} = 0\) (\(\equiv \lambda_u + \lambda_c + \lambda_s\), the solid one indicates the information available from rare kaon decays. The apex, \((\rho, \eta)\), can be determined from either triangle, and disagreement between the \(K\) and \(B\) determinations implies physics beyond the SM. In Fig. 4 the branching ratio closest to each side of the solid triangle can be used to determine the length of that side. \(K_\ell \to \mu^+ \mu^-\), which can in principle determine the bottom of the triangle \((\rho)\), is the decay for which the experimental data is the best but for which the theory is most problematic. \(K_\ell \to \pi^0 \nu \bar{\nu}\), which determines the height of the triangle \((\eta)\) is the cleanest theoretically, but for this mode experiment falls far short of the SM-predicted level. \(K^+ \to \pi^+ \nu \bar{\nu}\), which determines the hypotenuse, is nearly as clean as \(K_\ell \to \pi^0 \nu \bar{\nu}\) and has been observed (albeit with only three events). Prospects for the latter are probably the best of all since it is both clean and already within reach experimentally.

3.1 \(K_L \to \pi^0 \nu \bar{\nu}\)

\(K_L \to \pi^0 \nu \bar{\nu}\) is considered the most attractive target in the kaon system, since

1. it is direct CP-violating to a very good approximation \(^{13,32}\) (in the SM \(B(K_L \to \pi^0 \nu \bar{\nu}) \propto \eta^2\)) and
2. the rate can be rather precisely calculated in the SM or almost any alternative. \(^{33}\)

The hadronic matrix element can be obtained from \(K_{s3}\) via an isospin transformation. \(^{34}\) Unlike the case of its charged analog, \(K^+ \to \pi^+ \nu \bar{\nu}\), discussed below, it has no significant contribution from charm. Consequently, the intrinsic theoretical uncertainty connecting \(B(K_L \to \pi^0 \nu \bar{\nu})\) to the fundamental short-distance parameters is less than 1%. \(^{35}\) In the SM \(B(K_L \to \pi^0 \nu \bar{\nu})\) is directly proportional to \((\text{Im} \lambda_1)^2\) and \(\text{Im} \lambda_1 = -J/|A(1 - (\lambda_1^2/2))|\) where \(J\) is the Jarlskog invariant. \(^{36}\) Thus a measurement of \(B(K_L \to \pi^0 \nu \bar{\nu})\) determines the area of the unitarity triangles with a precision twice as good as that on \(B(K_L \to \pi^+ \nu \bar{\nu})\) itself.

\(B(K_L \to \pi^0 \nu \bar{\nu})\) can be bounded indirectly by measurements of \(B(K^+ \to \pi^+ \nu \bar{\nu})\) through a nearly model-independent relationship pointed out by Grossman and Nir. \(^{37}\) The application of this to the E787/949 results discussed below yields \(B(K_L \to \pi^0 \nu \bar{\nu}) < 1.4 \times 10^{-8}\) at 90% CL. This is far tighter than any extant direct experimental limit. To actually observe \(K_L \to \pi^0 \nu \bar{\nu}\) at the SM level \((\sim 3 \times 10^{-11})\), one will need to improve on the current state of the art by some 4 orders of magnitude.

The first dedicated \(K_L \to \pi^0 \nu \bar{\nu}\) experiment, KEK E391, \(^{38}\) mounted at the KEK 12 GeV proton synchrotron, aimed to achieve sensitivity comparable to the indirect limit. It was meant to serve as a test for a more sensitive experiment to be performed in the future at J-PARC. \(^{39}\) E391 features a carefully designed "pencil" beam \(^{40}\) with average \(K_L\) momentum \(\sim 2\) GeV/c. Figure 5 shows a layout of the detector.

A particular challenge of the E391 approach is to achieve extremely low photon veto inefficiency in order to reject the much more copious background decay modes with additional photons, such as \(K_L \to \pi^0 \pi^0\). The photon veto system consisted of two cylinders. The inner, more upstream barrel was intended to suppress beam halo and reduce confusion from upstream \(K_L\) decays. Roughly 4% of the \(K_L\)’s decayed in the 2.4 m fiducial region between the end of the inner cylinder and the charged particle veto upstream of the photon detector. Signal photons were detected in a multielement CsI-pure crystal calorimeter. \(^{41}\) The entire apparatus operated in vacuum. Physics running began in February 2004 and two more runs occurred in 2005.
In this experiment events with two showers in the calorimeter and no additional activity were examined to determine whether the assumption of a decay vertex at any point along the fiducial section of the beamline results in a reconstructed mass consistent with a $\pi^0$. If so, the $p_T$ could then be determined. Cuts were imposed on the shower patterns and energies, the $Z_V$ and the $p_T$. In addition, events consistent with $\eta \rightarrow \gamma \gamma$ were discarded.

Figure 6 shows the distribution of residual candidates in $p_T$ vs $Z_V$, when all other cuts are applied for the first 10% of the 2004 run. The observed versus predicted events are shown in several test regions where background is expected to dominate. The agreement is statistically acceptable. No events were observed in the two divisions of the signal region, with expected backgrounds of 200 events. Signal region is the 5-sided region at the center. The observed (predicted) background events in each subregion are indicated.

The proposed future program begins with moving the E391a detector to a 16° neutral beam at J-PARC. The beam is slightly lower in energy, but much more intense with $\sim 40$ times more useful $K_L$ per hour than at KEK. The CsI calorimeter will be replaced by one with superior resolution and granularity, the vetoes thickened, and the electronics upgraded. A single event sensitivity of $8 \times 10^{-12}$ per event and a signal : background of 1.4 : 1 (assuming the SM branching ratio) are the goals. A later stage experiment with optimized beam and detector aims at a >100 event measurement with a signal : background of 5 : 1.

3.2 $K_L \rightarrow \pi^0\ell^+\ell^-$

Like $K_L \rightarrow \pi^0\nu\bar{\nu}$, $K_L \rightarrow \pi^0\mu^+\mu^-$, and $K_L \rightarrow \pi^0e^+e^-$ are GIM-suppressed neutral current reactions sensitive to short-distance SM and BSM effects, but whose experimental considerations are qualitatively different. In the SM, like $K_L \rightarrow \pi^0\nu\bar{\nu}$, they are sensitive to $\Im A_\ell$, but in general they have different sensitivity to BSM effects and the combination of the measurements of both can be quite informative in BSM scenarios.

Although they are more tractable experimentally than $K_L \rightarrow \pi^0\nu\bar{\nu}$, they are subject to a background that has no analogue in $K_L \rightarrow \pi^0\nu\bar{\nu}$: $K_L \rightarrow \gamma\gamma\ell^+\ell^-$. This process, a radiative correction to $K_L \rightarrow \gamma\ell^+\ell^-$, is $10^{-3}$ times more copious than $K_L \rightarrow \pi^0\ell^+\ell^-$. Kinematic cuts are effective, but it is quite difficult to improve the signal : background beyond about 1 : 1.45\(1\) and still maintain a practical acceptance. Both $K_L \rightarrow \gamma\gamma\ell^+\ell^-$ \[BR_{\gamma\gamma,5\text{MeV}} = (5.84 \pm 0.15\text{stat} \pm 0.32\text{sys}) \times 10^{-7}\] and $K_L \rightarrow \gamma\gamma\mu^+\mu^-$ \[BR_{\gamma\gamma,1\text{MeV}} = (10.4_{-5.9}^{+7.3}\text{stat} \pm 0.7\text{sys}) \times 10^{-9}\] observed quantity measured in $\gamma\gamma$ events and $\mu^+\mu^-$ events, respectively.

In addition to this background, there are two other contributions that complicate extracting short-distance information from $K_L \rightarrow \pi^0\ell^+\ell^-$. Recent experimental and theoretical advances have mitigated their effects but they still have substantial impact:

1. An indirect CP-violating amplitude from the $K_1$ component of $K_L$ that is proportional to $\varepsilon A(K_S \rightarrow \pi^0\ell^+\ell^-)$.
2. A contribution of similar order, but CP-conserving, is mediated by $K_L \rightarrow \pi^0\gamma\gamma$. 
The first contribution is of the same order of magnitude as the direct CP-violating amplitude and can interfere with it. In the case of $K_L \rightarrow \pi^0 e^+e^-$, it yields:\(^{48}\)
\[
B(K_L \rightarrow \pi^0 e^+e^-)_{\text{CPV}} \approx 15.7 \alpha_s^2 + 6.2 \alpha_s \frac{\text{Im} \lambda_i}{10^{-4}} + 2.4 \left( \frac{\text{Im} \lambda_i}{10^{-4}} \right)^2 \times 10^{-12} \tag{2}
\]
where
\[
B(K_S \rightarrow \pi^0 e^+e^-) \approx 5.2 \alpha_s^2 \times 10^{-9} \tag{3}
\]
For $K_L \rightarrow \pi^0 \mu^+\mu^-$, the corresponding expressions are:\(^{48,49}\)
\[
B(K_L \rightarrow \pi^0 \mu^+\mu^-)_{\text{CPV}} \approx 3.7 \alpha_s^2 + 1.6 \alpha_s \frac{\text{Im} \lambda_i}{10^{-4}} + 1.0 \left( \frac{\text{Im} \lambda_i}{10^{-4}} \right)^2 \times 10^{-12} \tag{4}
\]
and
\[
\mathcal{M}(K_L \rightarrow \pi^0 \gamma \gamma) = \frac{G_F}{4 \pi} \epsilon_\mu (k_1) \epsilon_\mu (k_2) \left[ A(k_1^2 k_2^2 - k_1 \cdot k_2 g^{\mu \nu}) + \frac{B}{m_K^2} (p_{k_1} \cdot k_1 k_2^2 p_k^\mu + p_k \cdot k_1 k_2^2 p_k^\nu - k_1 \cdot k_2 p_k^\mu p_k^\nu - g^{\mu \nu} p_k \cdot k_1 p_k \cdot k_2) \right] \tag{6}
\]

\[
\mathcal{M}(K_L \rightarrow \pi^0 e^+e^-)_{\text{CPV}} \approx [(5.3 \pm 1.8)_{\text{mix}} \pm (2.5 \pm 0.5)_{\text{int}} + (1.7 \pm 0.3)_{\text{int}}] \times 10^{-12} \tag{9}
\]

There are now measurements for both these processes from NA48, based on 7 and 6 events for the electronic and muonic cases respectively. They find $B(K_S \rightarrow \pi^0 e^+e^-) = (5.8^{+2.6}_{-1.5} \pm 0.8) \times 10^{-9}$ (ref. 50) from which eq. (3) then yields $|\alpha_s| = 1.06^{+0.26}_{-0.07}$ and $B(K_S \rightarrow \pi^0 \mu^+\mu^-) = (2.9^{+1.5}_{-1.2} \pm 0.2) \times 10^{-9}$ (ref. 51) from which eq. (5) then gives $|\alpha_s| = 1.54^{+0.40}_{-0.32} \pm 0.06$. Averaging the electron and muon results yields a best estimate of $|\alpha_s| = 1.2 \pm 0.2$.\(^{48}\)

Another contribution of potentially similar size, but CP-conserving, is mediated by $K_L \rightarrow \pi^0 \gamma \gamma$. In principle this contribution should be predictable from measurements of $K_L \rightarrow \pi^0 \gamma \gamma$, of which thousands of events have been observed. The matrix element for this decay is given by:\(^{52}\)
\[
B(K_L \rightarrow \pi^0 \mu^+\mu^-)_{\text{CPV}} \approx (3.65 \pm 0.77) \times 10^{-11} \tag{10}
\]

In the muonic mode, using the value discussed above, one gets:
\[
B(K_L \rightarrow \pi^0 \mu^+\mu^-) \approx (1.4 \pm 0.2) \times 10^{-11} \tag{11}
\]

The current experimental status of $K_L \rightarrow \pi^0 \ell^+\ell^-$ is summarized in Table I and Fig. 7. About 30% more $K_L \rightarrow \pi^0 \mu^+\mu^-$ data is expected from the KTeV 1999 run.

As can be seen from Table I and Fig. 7, background in both modes is observed at a sensitivity an order of magnitude short of what is needed to observe the direct CP-violating signal. The problems of extracting a value of Im $\lambda_i$ from these modes have been discussed in ref. 61 among other places, and summaries of various schemes to deal with these problems are given in refs. 57 and 48, but recent developments make it worthwhile to take another look:

<table>
<thead>
<tr>
<th>Mode</th>
<th>90% CL upper limit</th>
<th>Estimated background</th>
<th>Observed events</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_L \rightarrow \pi^0 e^+e^-$</td>
<td>$2.8 \times 10^{-10}$</td>
<td>$2.05 \pm 0.54$</td>
<td>3</td>
</tr>
<tr>
<td>$K_L \rightarrow \pi^0 \mu^+\mu^-$</td>
<td>$3.8 \times 10^{-10}$</td>
<td>$0.87 \pm 0.15$</td>
<td>2</td>
</tr>
</tbody>
</table>
The observation of $K_S \to \pi^0 e^+ e^-$ with higher than expected rates imply that the SM branching ratios of the $K_{L}$ modes are also higher than previously believed.

If we compare the single event sensitivity of the most recent data with that of the residual $K_{L} \to \gamma\gamma e^+ e^-$ (the least tractable background), the ratio is much less than an order of magnitude. The 1999 KTeV $K_{L} \to \pi^0 e e$ data, had a single event sensitivity of $1.04 \times 10^{-10}$ and an estimated residual $K_{L} \to \gamma\gamma e e$ background of $0.99 \pm 0.35$ events. This implies a $S : B$ (signal to background) of $1 : 2.5$. In their 1997 $K_{L} \to \pi^0 \mu \mu$ data, KTeV had a single event sensitivity of $0.75 \times 10^{-11}$ and a calculated residual $K_{L} \to \gamma\gamma \mu \mu$ background of $0.37 \pm 0.03$ events, which implies $S : B = 1 : 1.9$.

The interest in $K_{L} \to \pi^0 e^+ e^-$ has evolved from its role as a possible source of information on $\lambda_t$, to that of an arena for probing BSM effects.

With this in mind one can ask, for example, what single event sensitivity would be needed to establish a factor two effect at $3 \sigma$ in these modes, given the background levels mentioned above. For the electronic case, the answer is $10^{-12}$; for the muonic case, $0.4 \times 10^{-12}$.

Judging by the KaMi proposal at Fermilab, a next-generation experiment could reach this level in about 3 years of running. If indeed, the SM were correct, a 30% measurement of the BR would result. Thus it seems a shame that no such experiment is on the near horizon. However such an experiment has been mentioned as a possible follow-up to the CERN $K^+ \to \pi^+ \nu \bar{\nu}$ proposal discussed below.

### 3.3 $K^+ \to \pi^+ \nu \bar{\nu}$

Recent developments in theory have rendered $K^+ \to \pi^+ \nu \bar{\nu}$ nearly as clean as $K_{L} \to \pi^0 \nu \bar{\nu}$. As in the neutral case, the often problematical hadronic matrix element can be calculated to $<1\%$ via an isospin transformation from that of $K^+ \to \pi^0 e^+ e^-$. The hard GIM suppression minimizes QCD corrections and the long-distance contributions to this decay are very small. A recent discussion of the latter with references to previous work can be found in ref. 64.

$K^+ \to \pi^+ \nu \bar{\nu}$ is directly sensitive to the quantity $\lambda_t$ as can be seen in eq. (12).

$$B(K^+ \to \pi^+ \nu \bar{\nu}) = \kappa_{\nu} \left[ \left| \frac{\text{Im} \lambda_t}{\lambda} \right|^2 \frac{P_c(x_t)}{X(x_t)} + \left( \frac{\text{Re} \lambda_t}{\lambda} + \frac{\text{Re} \lambda_{\tau}}{\lambda} \right) X(x_t) \right]^2,$$

where $\lambda = \sin \theta_{\text{Cabibbo}}$, $x_t \equiv (m_t/m_W)^2$, $X(x_t)$, and $P_c$ contain the top and charm contributions respectively and will be discussed below, and $\kappa_{\nu} = (5.17 \pm 0.03) \times 10^{-11}$.

The Inami–Lim function, $X(x_t)$ characterizing the GIM suppression of the top contribution is also given in ref. 66. For the current measured value of $m_t$, $X(x_t) = 1.464 \pm 0.041$. The QCD correction to this function is $\leq 1\%$. The charm contribution is given by:

$$\lambda^4 P_c(x) = \frac{2}{3} X'_{\nu_L} + \frac{1}{3} X'_{\nu_R},$$

where the functions $X'_{\nu_L}$ are those arising from the NNLO calculation. The QCD correction leads to a $\sim 30\%$ reduction of the charm Inami–Lim function which is now known to about 10%. From eq. (12)

$$B(K^+ \to \pi^+ \nu \bar{\nu}) = (7.96 \pm 0.49_{\sigma} \pm 0.84_{\text{other}}) \times 10^{-11},$$

where the last component contains uncertainties due to the CKM matrix elements and to $m_t$, which will naturally be reduced as data is improved. This would make a high precision measurement of $B(K^+ \to \pi^+ \nu \bar{\nu})$ very interesting from a BSM point of view.

As mentioned above the branching ratio can also be written in terms of the improved Wolfenstein variables and one finds it is proportional to

$$\lambda^2 X^2(x_t) \frac{1}{\sigma} [(\sigma \bar{\sigma})^2 + (\rho_0 - \bar{\rho})^2],$$

where

$$\sigma \equiv (1 - \lambda/2)^{-2}$$

and
\[ \rho_0 \equiv 1 + \frac{P_r(X)}{A^2 X(x)} . \]

To a good approximation the amplitude is proportional to the hypotenuse of the solid triangle in Fig. 4. This is equal to the vector sum of the line proportional to \( V_{sd}/A^3 \) and that from (1,0) to the point marked \( \rho_0 \). The length \( \rho_0 - 1 \) along the real axis is proportional to the charm contribution to \( K^+ \rightarrow \pi^+ \bar{\nu}\nu \). More precisely, \( B(K^+ \rightarrow \pi^+ \bar{\nu}\nu) \) determines an ellipse of small eccentricity in the \((\rho, \eta)\) plane centered at \((\rho_0, 0)\) with axes \( r_0 \) and \( r_0/\sigma \) where

\[ r_0 \equiv \frac{1}{A^2 X(x)} \sqrt{\frac{\sigma B(K^+ \rightarrow \pi^+ \bar{\nu}\nu)}{5.3 \times 10^{-11}}} . \]  

Fig. 8. (Color online) Elevation view of the E949 detector.

\[ K^+ \rightarrow \pi^+ \bar{\nu}\nu \] has been observed in the E787/949 series of experiments at the BNL AGS. These experiments used stopped \( K^+ \) which gives direct access to the \( K^+ \) center of mass, and is conducive to hermetic vetoing. The cylindrically symmetric detector, mounted inside a 1 Tesla solenoid, is shown in Fig. 8. An 80% pure beam of \( >10^7 K^+ \) per AGS cycle traversed a Cerenkov counter for identifying \( K^+ \) and \( \pi^+ \) and was tracked by MWPC’s. It was then slowed by a BeO degrader followed by a shower counter and beam hodoscope. About a quarter of the \( K^+ \) survived to enter a scintillating fiber stopping target. A hodoscope surrounding the stopping target demanded a single charged particle leave the target after a delay of \( \sim 2 \text{ns} \). The particle was tracked by a cylindrical drift chamber giving momentum resolution \( \sim 1\% \). Additional trigger counters required it to exit the chamber radially outward and enter a cylindrical array of scintillators and straw chambers, the “Range Stack” (RS), in which it was required to stop in order for the event to be considered a \( K^+ \rightarrow \pi^+ \bar{\nu}\nu \) candidate. In this configuration, the range and kinetic energy of the particle could be measured to \( \sim 3\% \). Comparison of range, energy and momentum is a powerful discriminator of low energy particle identity. In addition, transient recorder readout of the RS photomultipliers allowed the \( \pi^+ \rightarrow \mu^+ \rightarrow e^+ \) decay chain to be used to identify \( \pi^+ \)'s. The combination of kinematic and life-cycle techniques can distinguish pions from muons with a misidentification rate of \( O(10^{-3}) \). Surrounding the RS was a cylindrical lead-scintillator veto counter array and adjacent to the ends of the drift chamber were endcap photon veto arrays of undoped CsI modules. There were also a number of auxiliary veto counters near the beamline as well as a veto in the beamline downstream of the detector.

Monte Carlo estimation of backgrounds was in general not reliable since it was necessary to estimate rejection factors as high as \( 10^4 \) for decays occurring in the stopping target. Instead, methods to measure the background from the data itself were developed, using the primary data stream as well as data from special triggers taken simultaneously. The principles adhered to included:

- The signal acceptance region was kept hidden while cuts are developed.
- Cuts were developed on 1/3 of the data (evenly distributed throughout the run) but residual background levels determined only from the remaining 2/3 after the cuts were frozen.
- Background sources were identified \textit{a priori} and two independent high-rejection cuts were developed for each background. Each cut was reversed in turn as the other was studied. After optimization, the combined effect of the cuts could then be calculated as a product.
- Background calculations were verified through comparison with data near the signal region.

In this way backgrounds could be reliably calculated at the \( 10^{-3} \) to \( 10^{-2} \) event level.

All factors in the acceptance besides those of solid angle, trigger and momentum interval were determined from data.

Three \( K^+ \rightarrow \pi^+ \bar{\nu}\nu \) events were observed, two by E787\(^{73,74}\) and one by E949.\(^{75}\) The range versus kinetic energy distribution of these events is shown in Fig. 9. The combined result was a branching ratio \( B(K^+ \rightarrow \pi^+ \bar{\nu}\nu) = (1.47^{+1.130}_{-0.889}) \times 10^{-10} \). This is about twice as high as the prediction of eq. (14), but statistically compatible with it.

The total background to the two E787 events was measured to be 0.15 of an event and that of the E949 event...
0.3 of an event. Thus E787/949 has developed methods to reduce the backgrounds to a level sufficient to make a precise measurement of $K^+ \rightarrow \pi^+\nu\bar{\nu}$.

From the first observation published in 1997, E787’s results for $B(K^+ \rightarrow \pi^+\nu\bar{\nu})$ have been rather high with respect to the SM prediction. Although there has never been a statistically significant disagreement with the latter, it has stimulated a large number of predictions in BSM theories. The $K^+ \rightarrow \pi^+\nu\bar{\nu}$ data also yield an upper limit on the process $K^+ \rightarrow \pi^+X^0$ where $X^0$ is a massless weakly interacting particle such as a familon. For E787 this was $B(K^+ \rightarrow \pi^+X^0) < 5.9 \times 10^{-11}$ at 90% CL.

E949, which ran in 2002 was based on an upgrade of the E787 detector. It was improved in a number of ways with respect to E787: thicker and more complete veto coverage, augmented beam instrumentation, higher capacity DAQ, more efficient trigger counters, upgraded chamber electronics, auxiliary gain monitoring systems, etc. Using the entire flux of the AGS for 6000 hours, E949 was designed to reach a sensitivity of $\sim 10^{-11}$/event. In 2002 the detector operated well at fluxes twice those typical of E787, but unfortunately DOE support of the experiment was terminated after that first run. Further progress in $K^+ \rightarrow \pi^+\nu\bar{\nu}$ will have to come from experiments yet to be mounted.

There are currently two initiatives for future $K^+ \rightarrow \pi^+\nu\bar{\nu}$ experiments. One is a J-PARC LOI for a higher-sensitivity stopping experiment. This is very like E787/949 in conception, but with many improvements in detail. These include a lower incident beam momentum (leading to a higher stopping efficiency and a better signal/random rate ratio), higher granularity (leading to greater rate capability and muon rejection power), brighter scintillators, and a more capable DAQ. The goal of this experiment is to observe 50 events at the SM-predicted level. The schedule for running at J-PARC would be some time after 2012.

Although the stopped-$K^+$ technique is now well-understood, and one could be reasonably sure of the outcome of any new experiment of this type, to get really large samples of $K^+ \rightarrow \pi^+\nu\bar{\nu}$ ($\geq 100$ events), it will almost certainly be necessary to go to an in-flight configuration. There have been a series of attempts to initiate such an experiment, most recently the P326 proposal to CERN. This experiment exploits newly developed tracking technology to allow the use of an extremely intense ($\sim 1$ GHz) unseparated 75 GeV/c beam. Charged beams of this intensity have been used to search for good-signature kaon decays such as $K^+ \rightarrow \pi^+\mu^+\nu\bar{\nu}$, but P326 is a departure for a poor-signature decay for which high-efficiency vetoing is required.

Figure 10 shows the layout of the proposed experiment. Protons from the 400 GeV/c SPS will impinge on a 40 cm Be target. Positive secondaries with momenta within ±1% of 75 GeV/c will be taken off in the forward direction. The ~6% of $K^+$ in the beam will be tagged by a differential Cerenkov counter (CEDAR). The 3-momenta of all tracks...
will be measured in a beam spectrometer with three sets of “GIGATRACKER” detectors (fast Si microixels and micro-mega TPCs). The expected performance is $\sigma_p = 0.5\%$, $\sigma_\phi = 16\mu m$, and $\sigma_t = 150$ ps. The beamline has been carefully designed to hold the muon halo that contributes to detector random rates to the order of 10 MHz. The beam will continue through the apparatus in vacuum. About 10% of the $K^-\bar{\nu}$ decay will be conducted in vacuum out of the detector region. Photons from $K^+$ decays will be detected in a series of ring vetoes (at wide angles), by an upgraded version of the NA48 liquid Krypton calorimeter (at intermediate angles) and by two dedicated inner veto systems. Charged decay tracks will be momentum-analyzed in a two-dipole straw-tube spectrometer (<<1% resolution on pion momentum and 50–60 $\mu m$ resolution on $\theta_{K\pi}$ are necessary). Downstream of the spectrometer a RICH filled with Ne at 1 atm will help distinguish signal pions from background muons. This is to be followed by a charged particle hodoscope of multigap glass RPC design (100 ps resolution is required). Behind the hodoscope is the “MAMUD” muon veto, a magnetized iron-scintillator sandwich device to complete the pion/muon distinction. Its 5T-m bending power serves to kick the beam out of the way of the small angle photon veto at the back of the detector.

The collaboration proposes to build this experiment in time to begin taking data in 2011. A two-year would accumulate ~100 events with a 8 : 1 signal to background.

4. Beyond the Standard Model

Although in principle LFV processes like $K_L \to \mu e$ and $K^+ \to \pi^+ \mu^+ e^-$ can proceed through neutrino mixing, the known neutrino mixing parameters limit the rate through this mechanism to a completely negligible level.\(^{110}\) Thus the observation of LFV in kaon decay would require a new mechanism. Figure 11 shows $K_L \to \mu e$ mediated by a hypothetical horizontal gauge boson $X$, compared with the kinematically very similar process $K^+ \to \mu^+ \nu$ mediated by a $W$ boson. Using measured values for $M_W$, the $K_L$ and $K^+$ decay rates and $B(K^+ \to \mu^+ \nu)$, and assuming a $V - A$ form for the new interaction, one can show:\(^{101}\)

$$M_X \approx 220\text{TeV}/c^2 \left[ \frac{g_X}{g} \right]^{1/4} \left( \frac{10^{-12}}{B(K_L \to \mu e)} \right)^{1/4}.$$  \hspace{1cm} (16)

\(^{101}\) \hspace{1cm} Table II. Current 90% CL limits on $K$ decay modes violating the SM. The violation codes are “LF” for lepton flavor, “LN” for lepton number, “G” for generation number,\(^{100}\) “H” for helicity, “N” requires new particle.

<table>
<thead>
<tr>
<th>Process</th>
<th>Violates</th>
<th>90% CL BR limit</th>
<th>$\Gamma$ limit ($s^{-1}$)</th>
<th>Experiment</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_L \to \mu e$</td>
<td>LF</td>
<td>$4.7 \times 10^{-12}$</td>
<td>$9.1 \times 10^{-5}$</td>
<td>AGS-871</td>
<td>103</td>
</tr>
<tr>
<td>$K^+ \to \pi^+ \mu^+ e^-$</td>
<td>LF</td>
<td>$1.2 \times 10^{-11}$</td>
<td>$9.7 \times 10^{-4}$</td>
<td>AGS-865</td>
<td>104</td>
</tr>
<tr>
<td>$K^- \to \pi^- \mu^- e^+$</td>
<td>LF, G</td>
<td>$5.2 \times 10^{-10}$</td>
<td>$4.2 \times 10^{-2}$</td>
<td>AGS-865</td>
<td>105</td>
</tr>
<tr>
<td>$K_L \to \pi^0 \mu e$</td>
<td>LF</td>
<td>$3.31 \times 10^{-10}$</td>
<td>$6.4 \times 10^{-3}$</td>
<td>KTeV</td>
<td>106</td>
</tr>
<tr>
<td>$K_L \to \pi^0 \mu e$</td>
<td>LF</td>
<td>$1.58 \times 10^{-10}$</td>
<td>$3.1 \times 10^{-3}$</td>
<td>KTeV</td>
<td>107</td>
</tr>
<tr>
<td>$K^+ \to \pi^+ e^+ e^-$</td>
<td>LN, G</td>
<td>$6.4 \times 10^{-10}$</td>
<td>$5.2 \times 10^{-2}$</td>
<td>AGS-865</td>
<td>105</td>
</tr>
<tr>
<td>$K^+ \to \pi^+ \mu^+ \mu^-$</td>
<td>LN, G</td>
<td>$3.0 \times 10^{-9}$</td>
<td>$2.4 \times 10^{-1}$</td>
<td>AGS-865</td>
<td>105</td>
</tr>
<tr>
<td>$K^+ \to \pi^+ \mu^+ e^-$</td>
<td>LF, LN, G</td>
<td>$5.0 \times 10^{-10}$</td>
<td>$4.0 \times 10^{-2}$</td>
<td>AGS-865</td>
<td>105</td>
</tr>
<tr>
<td>$K_L \to \mu^+ \mu^- e^0$</td>
<td>LF, LN, G</td>
<td>$4.12 \times 10^{-11}$</td>
<td>$8.0 \times 10^{-4}$</td>
<td>KTeV</td>
<td>108</td>
</tr>
<tr>
<td>$K^+ \to \pi^0 f^0$</td>
<td>N</td>
<td>$5.9 \times 10^{-11}$</td>
<td>$4.8 \times 10^{-3}$</td>
<td>AGS-787</td>
<td>74</td>
</tr>
<tr>
<td>$K^+ \to \pi^+ \nu$</td>
<td>H</td>
<td>$2.3 \times 10^{-9}$</td>
<td>$1.9 \times 10^{-1}$</td>
<td>AGS-949</td>
<td>109</td>
</tr>
</tbody>
</table>

so that truly formidable scales can be probed if $g_X \sim g$ (see also ref. 102). In addition to this generic picture, there are specific models, such as extended technicolor in which LFV at observable levels in kaon decays is quite natural.\(^{15}\)

There were a number of $K$ decay experiments primarily dedicated to lepton flavor violation at the Brookhaven AGS during the 1990’s. These advanced the sensitivity to such processes by many orders of magnitude. In addition, several “by-product” results on LFV and other BSM topics have emerged from the other kaon decay experiments of this period and from more recent ones. Table II summarizes the status of work on BSM probes in kaon decay. The relative reach of these processes is best assessed by comparing the partial rates rather than the branching ratios.

This table makes it clear that any deviation from the SM must be highly suppressed. In a real sense the kaon LFV probes have become the victims of their own success. By and large the particular theories they were designed to probe have been forced to retreat to the point where meaningful tests in the kaon system would be very difficult (although there are exceptions\(^{100}\)). The currently more popular theoretical approaches tend to predict a rather small degree of LFV in kaon decays. For example, although these decays do provide the most stringent limits on strangeness-changing $R$-violating couplings, the minimal supersymmetric extension of the Standard Model (MSSM) predicts LFV in kaon decay at levels far below the current experimental sensitivity.\(^{111}\) Decays such as $K^+ \to \pi^+ \mu^+ \mu^+$, that violate lepton number as well as flavor, are also allowed in the MSSM, but are much more strongly suppressed. However there are models involving sterile neutrinos in which such processes are conceivably observable.\(^{112}\)
There have been some recent exceptions to the waning of theoretical interest in kaon LFV, including models with extra dimensions, but even in an improved motivational climate, there would be barriers to rapid future progress. Although K fluxes significantly greater than those used in the last round of LFV experiments are currently available, commensurate rejection of background is a significant challenge.

At the moment no new kaon experiments focussed on LFV are being planned. Interest in probing LFV has largely migrated to the muon sector. Note that a new experiment to probe $K_L \rightarrow \pi^0 \mu^+ \mu^-$ would tend to be sensitive to several LFV modes as well.

One exception to the poor prospects for dedicated BSM searches in kaon decay is the search for $T$-violating (out-of-plane) $\mu^+$ polarization in $K^- \rightarrow \pi^0 \mu^+ \nu$. There’s a proposal! to continue the work of the current experiment, KEK E246, at the J-PARC facility currently under construction. The proponents seek to make an advance in precision of a factor 20 on the measurement of the polarization. Since this is an interference effect the advance is roughly equivalent to that of $\sim 400$ in BR sensitivity. This measurement is quite sensitive to BSM physics, particularly multi-Higgs models including certain varieties of supersymmetry.

5. Further Study

The subject of kaon decay is much richer than can be presented in a short review such as this. Fuller accounts can be found in refs. 120–125.

Acknowledgment

This work was supported by the U.S. Department of Energy under Contract No. DE-AC02-98CH10886.

1) The Superweak hypothesis postulates that CP violation is solely due to an extremely weak $\Delta S = 2$ interaction.


18) $\epsilon'$ is the parameter governing $\Delta S = 1$-mediated CP violation in the decay $K^0 \rightarrow \pi \pi$ whereas $\epsilon$ is the CP-violating state-mixing parameter.


Laurence Littenberg was born in New York City in 1941. He obtained his A.B. (1963) degree from Cornell University and his Ph. D. (1969) degree from the University of California at San Diego. He was a research associate at the Daresbury Laboratory in England from 1970–74 before coming to Brookhaven National Laboratory in 1974. He is presently a Senior Physicist and Associate Chair for High Energy Physics in the Brookhaven Physics Department. He has worked primarily on experiments on rare kaon decay, but has occasionally strayed into areas like photoproduction, new particle searches, and neutrino oscillations.
HERA Physics

Robin DEVENISH

Department of Physics, University of Oxford, Keble Rd, Oxford OX1 3RH, U.K.

(Received January 24, 2007; accepted February 6, 2007; published November 12, 2007)

A brief introduction is given to high energy electron–proton interactions at the HERA collider. The article focuses on deep-inelastic scattering at high momentum transfers, electroweak interactions of neutral and charged current processes with polarised $e^\pm$ beams and one example of a search for signals beyond the standard model of particle physics.

KEYWORDS: high energy physics, deep-inelastic scattering, QCD, quarks, gluons, electroweak, neutral current, charged current, isolated leptons, BSM

DOI: 10.1143/JPSJ.76.111007

1. Introduction

HERA is a unique facility providing collisions of electrons (or positrons) with beam energy of 27.5 GeV with protons of beam energy 920 GeV. It is situated at the DESY laboratory in Hamburg. During 2000–2003 the collider was upgraded to provide higher luminosity and polarised $e^\pm$ beams. This article will concentrate on the physics studied by the two general purpose detectors, ZEUS and H1. Two other projects have used HERA beams in fixed target mode — HERA-B studied charm and beauty (two of the so-called “heavy flavour quarks”) production in hadron-hadron interactions using the proton beam on a wire target and HERMES continues to study fully polarised electron-hadron scattering using the polarised $e^\pm$ beams on a polarised hadronic target.

HERA has operated in two phases, HERA-I during the period 1992–2000 for much of which the proton beam energy was slightly lower at 820 GeV and the $e^\pm$ beams were unpolarised and HERA-II as described above. The HERA collider will cease operations at the end of June 2007. As this review is intentionally brief, only the highlights of HERA physics can be covered (the topics selected inevitably reflecting the interests of the author).

2. Deep-Inelastic Scattering

Deep-inelastic scattering at HERA is the process \(ep \rightarrow eX\) in which there is a large transfer of energy and momentum from the electron to the target through the exchanged force quantum, here a virtual photon or $Z^0$, causing the target proton to break up. As the exchanged quanta have zero electric charge this type of scattering is known as “neutral current” or NC. At very large momentum transfers the process \(ep \rightarrow \gamma X\) may also occur in which the exchanged quantum is now a charged $W^\pm$ — “charged current” or CC. In this part of the HERA phase space one may study the electroweak (EW) properties of the proton’s constituents. Many of the most important results from HERA concern the strong interaction, quantum chromodynamics (QCD). By measuring the final state, both inclusively and in more detail, one can learn something directly about the constituents of the proton (quarks) that couple to EW quanta and infer quite a bit about other constituents (gluons, the quanta of the strong force field). A good analogy is to imagine the combined HERA+detector setup as an enormous electron microscope that allows one to “see” inside the proton with a resolution given roughly by the uncertainty relation \(\Delta x \sim h/|q_c|\), where \(|q_c|\) is the magnitude of the momentum transferred from electron to proton. At the largest values of momentum transfer one is probing a distance scale of \(10^{-18}\) m, or something like a thousandth of the size of a proton.

The main components of a general purpose HERA collider detector are shown in Fig. 1. This is an event display from the H1 detector of a neutral current deep-inelastic scattering event. The beam line runs horizontally through the middle of the detector, which is cylindrical. The 27.5 GeV electron beam enters from the left and the 820 GeV proton beam from the right. The beam electron has been scattered through a large angle and emerges in the proton (forward) direction as the single charged particle in the upper half. In the lower half there is a nice example of

![Fig. 1. (Color online) A deep-inelastic $ep \rightarrow eX$ event in the H1 detector. The inner open “boxes” represent the tracking detectors, the green (light grey) shows the electromagnetic calorimeter and the orange (grey) the hadronic calorimeter. All of these components are inside a superconducting solenoid of diameter 6 m providing a magnetic field of 1.15 T.](image-url)
a hadronic “jet” of charged particles that has emerged from the break up of the proton in the interaction. There is also a small amount of debris recorded around the forward beam pipe. The panels on the right of the figure show that in a NC event the momentum of the scattered electron transverse to the beam line is balanced by that of the hadronic jet—a very distinctive feature. A CC event would have a hadronic jet balanced by an undetected neutrino, thus giving events with large missing “transverse energy”, $E_T$. The main features of the detector are a charged particle tracking system within a solenoidal magnetic field, followed by a calorimeter to measure both charged and neutral particle energies. The calorimeter is usually divided into two sections, often with different technologies, the first [electromagnetic (EM)] with a depth of around 25 radiation lengths to measure the energies of electrons and photons followed by the hadronic section with a depth of up to six or seven interaction lengths to measure hadronic energy. Muons do not interact strongly and penetrate to chambers (not shown) outside the calorimeter and coil. Final state neutrinos cannot be detected. The H1 calorimeter uses liquid argon with lead absorber plates for the EM section, followed by steel plates for hadron absorption and the whole detector is optimised for electron identification and measurement. The ZEUS detector has a similar structure, but with a uranium and scintillator sampling calorimeter for both EM and hadronic sections surrounding a coil of diameter 2 m providing a field of 1.43 T and the whole detector is optimised for hadronic jet measurements.

Deep-inelastic scattering has been intimately connected with the development of QCD and has a history as old as nuclear physics—going back through the results from muon and neutrino beams at CERN and Fermilab, the seminal results from SLAC, the earlier work by Hofstadter at Stanford on nuclei to Rutherford Scattering and the discovery of the nucleus. The connection between the elastic scattering of the earliest experiments and deep-inelastic scattering is made clear in Fig. 2, the virtual photon is scattering elastically off one of the charged constituents in the proton. The inelastic $ep \rightarrow eX$ cross-section is then the incoherent sum of the elastic cross-sections over all charged partons (parton is the generic name for all constituents in the proton, charged or neutral, now known to be quarks, antiquarks and gluons). At a fixed $ep$ centre-of-mass energy two variables characterise an inelastic interaction. These could be the energy and scattering angle of the scattered beam particle, but it is better to use Lorentz invariant quantities. Using the notation in the figure to define the 4-momenta of the particles concerned, the two most commonly used variables are $Q^2 = -(k-k')^2$, the negative of the squared momentum transfer from the electron, and $x = Q^2 / (2p_p \cdot q)$, which may be interpreted as the fraction of the incoming proton’s momentum carried by the quark involved in the so-called “hard-scatter”. A third variable $y = p_p \cdot q / Q^2$, $k$, which is related to the scattering angle in the electron-quark frame, is also useful. The three variables are related by $Q^2 = x y \sqrt{s}$, where $\sqrt{s}$ is the $ep$ centre-of-mass energy and particle masses have been ignored. The double differential cross-section for $ep \rightarrow eX$ takes the form

$$\frac{d^2\sigma}{dx dQ^2} = \frac{2\pi\alpha^2}{xQ^2} \left[1 + (1-y)^2F_2(x, Q^2) - y^2F_L(x, Q^2)\right],$$

where $F_2$ and $F_L$ are the structure functions of the proton—that there are two follows from the restrictions of a parity conserving interaction (electromagnetism) between two spin-1/2 particles—at large $Q^2$ the NC cross-section requires a third structure function $xF_3$ which contains the parity violating terms from $Z^0$ exchange. Equation (1) shows that for $Q^2 \ll M^2_Z$, the cross-section factorises into a piece containing $(\alpha/\pi Q^2)^2$ coming from the virtual photon exchange with a coupling strength given by the fine-structure constant $\alpha$ and the structure functions containing the information on the dynamical structure of the proton. $F_2$ may be written as $F_2(x, Q^2) = \sum e_i^2 f_i(x, Q^2)$, where $e_i$ is the electric charge (relative to the proton’s charge) of the $i$-th quark and $f_i(x, Q^2)$ is an unknown function (the parton momentum density function or pdf) giving the probability that $i$-th quark has a momentum $xP_p$. $F_L$ has a more complicated relationship to the proton pdfs, including a dependence on the gluon momentum density, $xg(x, Q^2)$. However, as the contribution of $F_L$ to the cross-section is small, except at large values of $y$, no details are given here. More introductory information about deep-inelastic scattering in general and at HERA in particular may be found in the book by the author and Cooper-Sarkar.

The pre-QCD quark-parton model predicted the above form for $F_2$, that the $f_i$ would be independent of $Q^2$ and that $F_L = 0$. Both results were supported approximately by the first data on deep-inelastic scattering. The more general $Q^2$ form has since been proved by the “factorisation” theorems of QCD. Although the pdfs cannot be calculated completely from QCD, the rate of change with $Q^2$ is calculable. The crucial ingredients are the “splitting” functions which are derived from the basic processes: $q \rightarrow qg, g \rightarrow q\bar{q}$ and $g \rightarrow gg$. The first two are analogous to bremsstrahlung and pair production in quantum electrodynamics (QED), but the third is different and follows from the fact the gluons carry the colour charge. Very roughly $\partial f/\partial \ln Q^2 \sim a_s(Q^2) \times xg + O(a_s^2)$, where $a_s(Q^2)$ is the QCD analogue of $\alpha$. The dependence of $\alpha_s$ on $Q^2$ is a consequence of quantum fluctuations and, unlike QED, $\alpha_s$ decreases as $Q^2$ increases—the phenomenon known as “asymptotic freedom”—making perturbative calculations possible for processes with a “hard” scale (for example $Q^2$ larger than a few GeV$^2$). The importance of the pdfs is that they are intrinsic to the proton, so if determined from deep-inelastic scattering then other hard scattering processes for example $pp \rightarrow W^+X$ can be predicted. The accurate determination of parton densities is crucial for all hadron collider physics and in particular for the the large hadron collider (LHC) at CERN. The pdfs are
also determined by “global” fits to hard scattering data from high energy $ep$, $vp$, and $p\bar{p}$ experiments.\(^4\)

3. Proton Structure

The left plot of Fig. 3 is a summary of the data on $F_2$ from $ep \to eX$ deep-inelastic scattering experiments from SLAC, CERN, Fermilab and HERA. The data are plotted at fixed values of $x$ as a function of $Q^2$. HERA results populate the band from top left (small $x$ and $Q^2$) to bottom right (large $x$ and $Q^2$) with the lower energy fixed target data in the region to the left at smaller $Q^2$ values. It is clear that the prediction of the early quark-parton model is not supported by the data, for small values of $x$ $F_2$ rises with $Q^2$ and for large values of $x$ the opposite occurs. This pattern of $x$ and $Q^2$ dependence is just what is predicted by QCD. The splitting functions derived from $q \to gq$ and gluon splitting $g \to gg$ behave like $1/x$ for small $x$. The gluon density rises and, through $g \to q\bar{q}$, so does the density of $q\bar{q}$ pairs that contribute directly to $F_2$. As $Q^2$ increases at small $x$ the momentum of the proton is shared amongst an increasing number of particles and thus the momentum of any one of them decreases. The right-hand plot of Fig. 3 shows $F_2$ data at $Q^2 = 20 GeV^2$ as a function of $x$ in more detail. The quality of the agreement between the QCD calculations and the data can be seen from this plot. The rapid decrease $F_2 \to 0$ as $x \to 1$ is simply a statement that it is very unlikely for a single quark to carry almost all the proton’s momentum. In the region of the pre-HERA fixed target data, $3 \times 10^{-2} < x < 0.2$ (triangles), $F_2$ increases slowly by about 20%. In contrast in the HERA region, $3 \times 10^{-4} < x < 3 \times 10^{-2}$, $F_2$ increases by a factor of 3. This is a dramatic increase indeed and one of the most important results from HERA and it raises an interesting question — will the rising behaviour persist to yet smaller $x$-values, as the calculated curves suggest?

That the rise of $F_2$ cannot continue indefinitely follows from a very general consideration. For values of $Q^2 \ll M_T^2$ deep-inelastic scattering is equivalent to virtual-photon proton scattering at a centre-of-mass energy, $W$, given by $W^2 = Q^2(1/x - 1)$. So small $x$ corresponds to large $W$. Now unitarity (conservation of probability) limits how rapidly any total cross-section can increase with energy. Given that $\sigma(\gamma^*/p) = (4\pi^2a/Q^2)F_2$, a limit on the large $W$ behaviour of $\sigma(\gamma^*/p)$ is equivalent to a limit on the low $x$ behaviour of $F_2$. Another, more physical approach, is to realise that as the gluon density increases at low $x$, other terms will start to become important, particularly the possibility of gluon recombination $gg \to g$. The recombination cross-section is of order $\alpha_s/Q^2$, thus one might expect to see “saturation” effects when $(\alpha_s/Q^2)xg(x, Q^2) \sim \pi R^2$, where $R$ is the radius of the proton. Taking $R \sim 0.8 fm$ and $Q^2 \sim 10 GeV^2$ gives $xg \sim 2000$, quite a bit above the largest projections at the smallest values of $x$ reachable at HERA.\(^5,6\) Much effort has been expended to find other signs in the HERA data for saturation effects and, although there are some tantalising hints, there are competing explanations. What is undoubtedly true is that the rising gluon density has sparked enormous interest in trying to understand high-density gluon dynamics theoretically and to extrapolate from the HERA measurements to heavy-ion collisions (in which these effects are enhanced) at the Brookhaven relativistic heavy-ion collider (RHIC) and the LHC.

The gluon density is only measured indirectly from the $Q^2$ dependence of the inclusive structure function. The excellent tracking of the HERA detectors enables measurement of processes involving the heavy charm and beauty quarks (masses around 1.3 and 4.5 GeV/c\(^2\) respectively). This can be done by reconstruction of a $D^*$ meson\(^7\) (c$q$ combination) or by identifying events with secondary vertices or tracks with large impact parameters (both indicative of long lived heavy meson decay). For beauty measurements in particular, the identification of the electron or muon from the weak semi-leptonic decay of b-quark containing hadrons or jets with large impact parameter tracks is crucial.\(^8\) The dominant
process for the production of a heavy $q\bar{q}$ pair in deep-inelastic scattering is so-called photon–gluon fusion (shown in the left-hand plot of Fig. 4). The rate for this process depends directly on the gluon pdf. The crucial point is that a heavy $q\bar{q}$ pair is extremely unlikely to be produced in the process by which a quark “fragments” into the jet of hadrons seen in the detector. The right-hand plot of Fig. 4 shows the $Q^2$ dependence of the cross-section for $ep \rightarrow eD^*X$. The measured data are compared with a next-to-leading order (i.e., diagrams to order $\alpha^2$) QCD calculation using a gluon density derived from fitting the ZEUS structure function data. The good agreement over five orders of magnitude gives one confidence in both QCD calculations of hard processes and the consistency of the HERA measurements.

4. Electroweak Standard Model Measurements

The ability of HERA to allow the measurement of both NC ($ep \rightarrow eX$) and CC ($ep \rightarrow \nu X$) total cross-sections to values of $Q^2$ well above the mass scale of the $W$ and $Z$ weak bosons offers a nice “textbook” picture of the unity of the electromagnetic and weak interactions at large $Q^2$. This is shown in the left-hand plot of Fig. 5 which summaries the $Q^2$ dependence of the four cross-sections for NC and CC scattering with incident electron and positron beams. As discussed after eq. (1), the dominant diagram for NC scattering behaves as $\alpha/Q^2$. For the CC process the behaviour is $M_W^2 G_F/(M_W^2 + Q^2)$, where $G_F$ is the Fermi constant as measured in low energy beta decay. For $Q^2 \ll M_W^2$ ($M_W = 80.4$ GeV/$c^2$), the CC cross-section is tending to a constant value much smaller than that of the NC cross-section which is enhanced by $1/Q^2$ as $Q^2$ decreases. At large values of $Q^2$, near and above $M_W^2$, the NC and CC cross-sections are of comparable magnitude and both eventually decreasing as $1/Q^4$. This is EW unification in action — the simplest explanation is that $\alpha \sim M_W^2 G_F$ — which, with a few extra constants, is what is predicted by the Weinberg-Salam theory. The small but clear dependence of the NC cross-section on the charge of the beam $e^\pm$ follows from the parity violating $Z^0$ exchange. For the CC cross-sections parity is also violated, but the charge dependence is larger because of the quark charge (and flavour) selection by $W^\pm$ exchange as discussed in the next section.

Another textbook plot—exploiting the polarised $e^\pm$ beams at HERA-II—is shown in the right-hand plot of Fig. 5. This shows the dependence of the CC cross-section on the beam polarisation. In the standard model, neutrinos have only left-handed helicity (spin projected along the momentum direction — analogous to circular polarisation of...
light) and anti-neutrinos only right-handed helicity. The expectation is then that the $e^-p \rightarrow \nu X$ cross-section will show a $(1 - P)$ dependence (where $P$ is the beam polarisation) and the $e^+p \rightarrow \bar{\nu}X$ a $(1 + P)$ dependence. This is what the measurements in the figure show.\(^5\)\(^,\)\(^10\) \(10\) The difference in magnitude of the two cross-sections follows from the charge of the $W$: for $e^-p$ the dominant contribution comes from $\nu W^- \rightarrow d$; for $e^+p$ it is $d W^+ \rightarrow u$ but in this case the contribution is suppressed by an angular factor $(1 - y)^2$. The closeness of the data to the expected polarisation dependence can be used to extract limits on possible “right-handed” weak exchanges, beyond the standard model. Accurate measurements of both NC and CC cross-sections at large $x$ and $Q^2$ are also an important input to the fits that are used to extract proton pdfs as they depend on the charge of the valence quarks, allowing these to be measured as well.\(^11\) All these interesting measurements will be completed when the full HERA-II data set is available.

5. High $p_T$ Leptons

Beyond the standard model (BSM) searches at HERA cover a vast range of topics. No definitive signals have yet been seen, but lower limits on the masses of states in many models have been set. For example, leptoquarks (particles formed by fusing a quark and lepton) which occur in many BSM models could be seen at HERA as a resonant excitation from $QW^\pm \rightarrow q\ell$; for $e^-p$ it is $d W^+ \rightarrow u$ but in this case the contribution is suppressed by an angular factor $(1 - y)^2$. The closeness of the data to the expected polarisation dependence can be used to extract limits on possible “right-handed” weak exchanges, beyond the standard model. Accurate measurements of both NC and CC cross-sections at large $x$ and $Q^2$ are also an important input to the fits that are used to extract proton pdfs as they depend on the charge of the valence quarks, allowing these to be measured as well.\(^11\) All these interesting measurements will be completed when the full HERA-II data set is available.

<table>
<thead>
<tr>
<th>Events with $p_T^e &gt; 25\text{GeV}$</th>
<th>Electrons data/SM</th>
<th>Muons data/SM</th>
</tr>
</thead>
<tbody>
<tr>
<td>$e^-p$ H1 158 pb$^{-1}$</td>
<td>9.3 ± 0.3</td>
<td>6.2 ± 0.4</td>
</tr>
<tr>
<td>ZEUS 106 pb$^{-1}$</td>
<td>1.5 ± 0.1</td>
<td>1.5 ± 0.2</td>
</tr>
<tr>
<td>$e^-p$ H1 184 pb$^{-1}$</td>
<td>3.8 ± 0.6</td>
<td>0.3 ± 0.5</td>
</tr>
<tr>
<td>ZEUS 143 pb$^{-1}$</td>
<td>2.9 ± 0.5</td>
<td>2.1 ± 0.2</td>
</tr>
</tbody>
</table>

Table I. Summary of isolated lepton results from H1 and ZEUS.\(^13\)

$e^-p \rightarrow eWX$ followed by the leptonic decay of the $W$, see the left-hand diagram of Fig. 6 in which the $ff$ from the $W$ decay is here $\ell \bar{\nu}_\ell$. Isolated lepton events have generated a lot of interest in recent years and are an unsolved problem for the HERA community. H1 see an excess of events above the expectation from $W$ production, particularly in $e^-p$ scattering. ZEUS, on the other hand, find agreement with the standard model expectation in all channels. The discrepancy is most clear for large values of $P_T^e$, the total transverse momentum of the hadronic system in the event, as the standard model $W$ production mechanism tends to produce isolated lepton events with low values of this quantity. The right-hand plot of Fig. 6 shows the $P_T^e$ spectrum from H1 for isolated $e$ and $\mu$ events in $e^-p$ scattering. A clear excess of events above the standard model expectation is observed. Table I gives a summary of the observed and expected event numbers from both experiments, for the highest $P_T^e$ bins.\(^13\) Taking the $e$ and $\mu$ results together the H1 $e^-p$ data give an excess of around 3.4. For both ZEUS channels and for H1 $e^-p$ data there is no excess. Both collaborations have tried hard to identify any differences in detector performance or analysis technique that could explain the different results, without success. An example of a possible BSM mechanism that could match the pattern of the H1 data comes from $R$-parity violating supersymmetry using 3rd-generation fields, $e^- + d \rightarrow \bar{\nu}_e$ via $\tilde{b}$ (SUSY partner of the b-quark) exchange. The top quark then decays by $t \rightarrow \ell \nu_\ell$ ($\ell = e$ or $\mu$) with large missing energy. The equivalent process in $e^-p$ scattering could not produce the positive charge-2/3 top quark. For a more general discussion of these events in the context of $\tilde{R}_2$ SUSY, see ref. 14.
6. Conclusions

This brief survey has given a very partial overview of some of the many physics topics that have been studied in high-energy $ep$ scattering at the HERA collider. There are many other areas that have not been covered. These include “diffractive scattering” in which the proton remains intact or emerges in a low lying excited state with the same quantum numbers ($ep \to epX$) and the many interesting results on hadronic final states, QCD jet physics and the hadronic properties of the photon from photoproduction reactions. Photoproduction events, $\gamma p \to X$, are those in which an energetic and nearly real photon, radiated from the beam $e^+$ interacts with the proton. The scattered $e^+$ remains undetected in the beam pipe. Many more details about the two experiments, the published results and conference reports may be found at the H1 and ZEUS websites.\textsuperscript{15}

Acknowledgements

I would like to thank my colleagues on the ZEUS and H1 experiments for many fruitful years of physics at HERA and Mark Bell for a careful reading of the manuscript.

4) Full details on global pdf fits may be found at the Durham University HEPDATA website: http://durpdg.dur.ac.uk/HEPDATA/
13) C. Diaconu: talk in the BSM session ICHEP06, Moscow, July 2006; hep-ex/0610041; to be published in the proceedings.
15) H1: www-h1.desy.de; ZEUS: www-zeus.desy.de

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{RobinDevenish.jpg}
\caption{Robin Devenish was born in Trinidad, West Indies in 1942, but has lived in the UK since 1944. He obtained BA (1964), MA, and Ph. D. (1969) from Cambridge University. He held research associate positions at Lancaster University and UCL and was on the staff of the DESY Laboratory in Hamburg for six years. In 1979 he moved to Oxford where he has been ever since. He has worked on the photo- and electro-production of nucleon resonances, electron–positron annihilation physics and many aspects of electron–proton deep inelastic scattering.}
\end{figure}
1. Introduction

In 1998, long sought neutrino oscillation was discovered by the Super-Kamiokande group in atmospheric neutrino observations. Since then, neutrino oscillation has also been established in solar neutrino observations and long-baseline experiments utilizing reactor- and accelerator-produced neutrinos.

Neutrino oscillations imply finite neutrino mass as well as flavor mixing in the neutrino sector. In the Standard Model, neutrinos are massless because no right-handed neutrinos exist in the theory. Therefore, neutrino oscillation is the first evidence of physics beyond the Standard Model.

An unexpectedly large mixing observed in atmospheric neutrino oscillations indicated that the neutrino mixing matrix is quite different from the well-known quark mixing matrix. Comparison of the mass and mixing in the quark sector and those in the lepton sector would be an important clue to investigate physics beyond the Standard Model. Therefore, obtaining complete knowledge of neutrino mass and mixing is the goals of future neutrino physics.

From the cosmic microwave background (CMB) data alone, neutrino mass is constrained as 2–4)

\[ \sum_i \nu_i < 2 \text{ eV}. \]  

A stronger constraint

\[ \sum_i \nu_i < 0.7 \text{ eV} \]  

is obtained using the combination of CMB and other cosmological data. Therefore, the mass of the heaviest neutrino is less than 0.7–2 eV. Compared with the masses of quarks and charged leptons, the neutrino masses are extremely small. If neutrinos are Dirac particles, a minimal extension of the Standard Model may allow a Dirac neutrino mass term \( \mathcal{L}_D = m_D (\bar{\nu}_R \nu + \bar{\nu}_L \nu_L) \) by introducing the right-handed neutrino field, but the tiny neutrino mass means a tiny Yukawa coupling compared to Yukawa couplings of quarks and charged leptons. There is no explanation for this mysterious relation between the neutrino mass and other fermion masses. However, because the neutrino is a neutral particle, a Majorana mass term \( \mathcal{L}_M = M (\bar{\nu}_R \nu_R + \bar{\nu}_L \nu_L) \) may be introduced, where \( \nu_R \) is the charge conjugate of \( \nu_R \). The Majorana mass term violates the lepton number by 2, and the neutrino and antineutrino states cannot be distinguished.

Then, neutrinos are Majorana particles. The existence of both Dirac and Majorana mass terms leads to a natural explanation of the tiny left-handed neutrino mass, \( m \sim m_D^2/M \) by introducing heavy right-handed neutrinos with mass \( M > 10^{12} \text{ GeV} \), where \( m_D \) is a typical quark or charged-lepton mass. Conversely, the tiny neutrino mass indicates the existence of a very high mass scale. This is called the seesaw mechanism.

Whether neutrinos are Dirac or Majorana particles is of fundamental importance. If neutrinos are Majorana particles, “leptogenesis,” proposed by Fukugita and Yanagida, is an attractive mechanism for the origin of baryon asymmetry in the universe.

In the following, we will first discuss in §2 the present knowledge of neutrino mass and mixing. In §3, we will briefly survey the history of neutrino oscillation experiments. Then, we will review experiments which established neutrino oscillations. Atmospheric neutrino experiments and accelerator long baseline neutrino oscillation experiments which measured \( \Delta m^2_{12} \) and \( \theta_{13} \) are reviewed in §4, and solar neutrino experiments and the KamLAND reactor neutrino oscillation experiment which measured \( \Delta m^2_{12} \) and \( \theta_{12} \) in §5. In §6, future prospects of neutrino experiments are discussed. Conclusions are given in §7.

2. Neutrino Mass and Mixing

2.1 Neutrino oscillation parameters

Within the framework of three active neutrino species, the flavor eigen states \( \nu_{\alpha} (\alpha = 1, 2, 3) \) and the mass eigen states \( \nu_i (i = 1, 2, 3) \) with eigen mass \( m_i \) are related by the following equation.

\[
\begin{pmatrix}
\nu_e \\
\nu_\mu \\
\nu_\tau
\end{pmatrix} =
[U_{\alpha i}]
\begin{pmatrix}
\nu_1 \\
\nu_2 \\
\nu_3
\end{pmatrix},
\]

where \( U \) is a unitary \( 3 \times 3 \) mixing matrix. It is defined as a product of three rotation matrices: the parameters involved are three mixing angles \( \theta_{12}, \theta_{23}, \) and \( \theta_{13} \), and a CP-violating phase \( \delta \):
where $S_{ij}$ and $C_{ij}$ stands for $\sin \theta_{ij}$ and $\cos \theta_{ij}$, respectively. Historically, Maki, Nakagawa, and Sakata first formulated two-flavor neutrino oscillations in 1962. Therefore, the mixing matrix $U$ is called Maki–Nakagawa–Sakata or MNS matrix after them. Prior to them, Pontecorvo predicted neutrino-antineutrino oscillations in 1957. Therefore, $U$ is also called Maki–Nakagawa–Sakata–Pontecorvo or MNSP matrix.

If neutrinos are Majorana particles, there are additional two CP-violating phases called Majorana phases, but they have no effects on neutrino oscillations. The general expression for the probability of neutrino oscillation $\nu_\alpha \rightarrow \nu_\beta$ in vacuum at distance $L$ is given by

$$P(\nu_\alpha \rightarrow \nu_\beta) = \left| \sum_j U_{\alpha j}^* U_{\beta j} e^{-i m_j^2 / 2 E L} \right|^2,$$

where $E$ is the neutrino energy. It is apparent that the neutrino mass enters into the neutrino oscillation probability as mass squared differences, $\Delta m^2_{ij} = m_i^2 - m_j^2$. For three active neutrinos there are only two independent values of $\Delta m^2$, because

$$\Delta m^2_{12} + \Delta m^2_{23} + \Delta m^2_{31} = 0.$$  

In the analyses of neutrino oscillation experiments, the two-neutrino oscillation framework is often used. In vacuum, the neutrino oscillation probability $P_{(2\nu)}(\nu_\alpha \rightarrow \nu_\beta)$ for $\alpha \neq \beta$ is given by

$$P_{(2\nu)}(\nu_\alpha \rightarrow \nu_\beta) = \sin^2 2\theta \sin^2(1.27 \Delta m^2 L/E).$$

Here, the units of $\Delta m^2$, $E$, and $L$ are $eV^2$, $GeV$, and $km$, respectively. In fact, the atmospheric neutrino oscillation indicated that it is almost pure two-neutrino oscillation in vacuum, $\nu_\mu \leftrightarrow \nu_\tau$.

$$P_{(2\nu)}(\nu_\mu \rightarrow \nu_\tau) = \sin^2 2\theta_{\text{atm}} \sin^2(1.27 \Delta m^2_{\text{atm}} L/E).$$

The solar neutrino measurements by SNO (Sudbury Neutrino Observatory) and Super-Kamiokande showed evidence for neutrino flavor conversion. A plausible explanation is neutrino oscillation dictated by matter effect with $\Delta m^2_{\text{sol}} \sim (5-10) \times 10^{-5} eV^2$ and a nearly maximal mixing angle, $\sin^2 \theta_{\text{sol}} \sim 1$.

The solar neutrino measurements by SNO (Sudbury Neutrino Observatory) and Super-Kamiokande showed evidence for neutrino flavor conversion. A plausible explanation is neutrino oscillation dictated by matter effect with $\Delta m^2_{\text{sol}} \sim (5-10) \times 10^{-5} eV^2$ and a nearly maximal mixing angle, $\sin^2 \theta_{\text{sol}} \sim 1$.

Now, the values of $\Delta m^2_{\text{atm}}$ and $\Delta m^2_{\text{sol}}$ are clearly different, they are identified with the two independent $\Delta m^2$ in eq. (6). At this point, it is worth noting that the LSND group claimed a positive signal observed in a short baseline neutrino oscillation experiment, $\nu_\mu \rightarrow \nu_e$. If confirmed, the corresponding $\Delta m^2$ should lie in the range of $0.1-10 eV^2$. This means that there are at least three independent values of $\Delta m^2$, indicating the existence of at least four neutrino species. Since the number of light active neutrino species is constrained to three, an additional one should be a sterile neutrino species. However, this LSND “anomaly,” has not been confirmed by the recent short baseline $\nu_\mu \rightarrow \nu_e$ oscillation experiment MiniBooNE. If the oscillations of neutrinos and antineutrinos are the same, the MiniBooNE result excludes neutrino oscillations claimed by LSND at 98% CL (confidence level). Therefore, the three-neutrino framework now stands on a solid basis.

By convention, we define

$$\Delta m^2_{\text{atm}} = m^2_1 - m^2_2 > 0.$$  

With this definition, the largest component of $\nu_1$ is $\nu_e$. For larger mass splitting, we define

$$\Delta m^2_{\text{sol}} = |\Delta m^2_{23}| = |m^2_2 - m^2_3|.$$  

Note that two-flavor $\nu_\mu \leftrightarrow \nu_e$ oscillation is vacuum oscillation which is symmetric under the change of sign of $\Delta m^2_{23}$. In order to determine $\text{sign}(\Delta m^2_{23})$, matter effects are needed.

If neutrino mass spectrum is hierarchical, there are two possibilities; one is “normal mass hierarchy” $m_3 > m_2 > m_1$ and the other is “inverted mass hierarchy,” $m_3 > m_1 > m_2$. It is one of important tasks of future neutrino oscillation experiment to resolve this ambiguity. There is another possibility of “quasi-degenerate” mass spectrum, $m_1 \simeq m_2 \simeq m_3 > 0.1 eV$. The lower mass limit is obtained from the observed masses of $|\Delta m^2_{23}|$ and $\Delta m^2_{12}$. In this case, neutrinoless double beta decay experiments will have sensitivities to explore the relevant mass region in near future.

In the framework of three-neutrino oscillation, to the extent that $\Delta m^2_{12}$ is neglected, the oscillation probabilities in vacuum can be approximately expressed with the two mixing angles, $\theta_{23}$ and $\theta_{13}$, and $\Delta m^2_{23}$. The $\nu_\mu \leftrightarrow \nu_e$ oscillation probability is given as

$$P(\nu_\mu \rightarrow \nu_e) = \sin^2 2\theta_{23} \cos^2 \theta_{13} \sin^2(1.27 \Delta m^2_{23} L/E).$$

Also, $\nu_e$ disappearance probability is given as

$$P(\bar{\nu}_e \rightarrow \bar{\nu}_e) = 1 - \sin^2 2\theta_{13} \sin^2(1.27 \Delta m^2_{23} L/E).$$

Note that eq. (12) has an identical form with the two-neutrino disappearance probability in vacuum, $P_{(2\nu)}(\bar{\nu}_e \rightarrow \bar{\nu}_e)$ with a mixing angle $\theta_{13}$. From a short baseline reactor neutrino oscillation experiment, CHOOZ, $\theta_{13}$ is known to be small, $\sin^2 2\theta_{13} < 0.19$. Then, eq. (11) reduces to the two-neutrino oscillation in vacuum. From eqs. (8) and (11),

$$\theta_{23} \approx \theta_{\text{atm}}.$$  

Writing the survival probability of solar $\nu_e$ in terms of the MSW effects with two-neutrino mixing as

$$P_{(2\nu)}(\nu_{e} \rightarrow \nu_{e}) \equiv P_{(2\nu)}(\nu_{e} \rightarrow \nu_{e}; \Delta m^2_{\text{sol}}, \sin^2 2\theta_{\text{sol}}),$$

if $\theta_{13}$ is small the survival probability of solar $\nu_e$ with three-neutrino mixing is approximately written as

$$P_{(3\nu)}(\nu_{e} \rightarrow \nu_{e}) = \cos^4 \theta_{13} P_{(2\nu)}(\nu_{e} \rightarrow \nu_{e}; \Delta m^2_{\text{sol}}, \sin^2 2\theta_{13}) + \sin^4 \theta_{13}.$$
For $\theta_{13} \sim 0$, therefore,

$$\theta_{12} \approx \theta_{\text{atm}}. \quad (16)$$

Measurement of the small mixing angle $\theta_{13}$ and the CP violating phase $\delta$ in the MNS matrix and measurement of the sign of $\Delta m^2_{21}$ are important problems left for future neutrino oscillation experiments.

2.2 Absolute neutrino mass scale

The best upper limit (95% CL) so far was obtained by the Mainz tritium beta decay experiment, $^{3}\text{H} \rightarrow ^{3}\text{He} + e^- + \nu_e$, has been the most sensitive method of direct neutrino mass measurement. Because of the neutrino mixing, what is measured is the quantity

$$m_{\nu_{e}}^{2(\text{eff})} = \sum |U_{ei}|^2 m_i^2, \quad (17)$$

The best upper limit (95% CL) so far was obtained by the

If neutrinos are Majorana particles, neutrinoless double beta decay $(0\nu\beta\beta)$, $(Z, A) \rightarrow (Z + 2, A) + 2e^-$, occurs. The decay rate is proportional to the square of the effective Majorana mass $\langle m \rangle$,

$$\langle m \rangle = |m_1|U_{e1}|^2 + m_2|U_{e2}|^2 e^{i\alpha_2} + m_3|U_{e3}|^2 e^{i\alpha_3}, \quad (18)$$

where $\alpha_2$ and $\alpha_3$ are the Majorana CP-violating phases. In general, cancellations may occur for $\langle m \rangle$.

So far, the best limit of $\langle m \rangle < 0.2$–1.1 eV has been obtained with an isotope of $^{130}\text{Tm}$ using a cryogenic detector. The wide range of upper limit reflects uncertainties in the theoretical nuclear matrix elements. Although there is a claim of positive observation of neutrinoless double beta decay using $^{76}\text{Ge}$, corresponding to $\langle m \rangle = 0.1$–0.9 eV, it has not been confirmed.

From the present knowledge of the neutrino oscillation parameters, the relation between the effective Majorana mass and the lightest neutrino mass can be evaluated for each possibility of the neutrino mass spectrum, as shown in Fig. 1. If the neutrino mass spectrum is quasi-degenerate, $(0\nu\beta\beta)$ will be observed with $\langle m \rangle \geq 50$ meV. If $(0\nu\beta\beta)$ is observed with $\langle m \rangle \sim 20$–50 meV, the inverted mass hierarchy is the likely neutrino mass spectrum.

3. History of Neutrino Oscillation Experiments

Searches for neutrino oscillations using the data from accelerator neutrino experiments around 1976. Thereafter, extensive efforts have been made, but no evidence was found because the expected $\Delta m^2$ range was not relevant. Though in a few cases oscillations were reported from reactor experiments, they were not confirmed.

It was fortunate that the value of $\Delta m^2_{21}$ was well-matched with the sensitivity of the atmospheric neutrino oscillation experiment and the mixing angle was almost maximal. Otherwise, it would not have been so obvious whether the atmospheric neutrino experiment could discover the neutrino oscillation first. It should be noted, however, that there were already hints for neutrino oscillations before Super-Kamiokande’s discovery. Since 1988, Kamiokande Collaboration reported that the observed $\mu/e$ ratio (the ratio of the muon events to electron events, both induced by atmospheric neutrinos) was significantly smaller than the theoretically expected $\mu/e$ ratio. More interesting results presented by the Kamiokande Collaboration was zenith-angle dependence of $R = \text{(observed } \mu/e)/\text{(expected } \mu/e)$ for events having a visible energy $> 1.33$ GeV. Kamiokande Collaboration suggested neutrino oscillations to explain these observations. Some of the other atmospheric neutrino experiments supported Kamiokande’s observation of $R < 1$, but other experiments reported $R$ to be consistent with unity. The high statistics data from an order-of-magnitude bigger detector, Super-Kamiokande, was needed to disentangle this controversy.
4. Measurements of $|\Delta m^2_{32}|$ and $\theta_{23}$

4.1 Atmospheric neutrino results

The first evidence for the neutrino oscillation was presented by the Super-Kamiokande Collaboration in 1998. The zenith-angle distributions of the $\mu$-like events (mostly muon-neutrino initiated charged-current interactions) showed clear deficit compared to the no-oscillation expectation.

Super-Kamiokande is a 50-kton water Cherenkov detector located in the Mozumi mine in Gifu Prefecture, about 250 km west of Tokyo, at a depth of 2,700 m of water equivalent. The inside of the water tank is optically separated into the cylindrical inner detector and a 2.75 m thick outer shield region (anti-counter) by opaque sheets. This experiment started observation in April 1996. In November 2001, Super-Kamiokande suffered from an accident in which substantial number of photomultiplier tubes were lost. The detector was rebuilt within a year with about half of the original number of photomultiplier tubes. The experiment with the detector before the accident is called Super-Kamiokande-I (SK-I), and that after the accident is called Super-Kamiokande-II (SK-II). Later in 2005–2006, the detector was rebuilt again to recover the original number of photomultiplier tubes, and the experiment entered a new phase of Super-Kamiokande-III (SK-III). The complete SK-I atmospheric-neutrino results with a 92 kton-yr (1489 live-day) exposure to atmospheric neutrinos are reported in ref. 35.

Neutrino events in Super-Kamiokande are classified into fully contained (FC) events and partially contained (PC) events. The FC events are required to have their vertex position inside the 22.5 kton fiducial volume, defined to be >2 m from the PMT wall of the inner detector, and to have no visible energy deposited in the anti-counter. The PC events are those events which have at least one prong that penetrates the inner detector. Obviously, the total visible energy can be measured for the FC events, but not for the PC events.

Neutrino events are also classified according to the number of Cherenkov rings. A Monte Carlo study shows that more than 90% of the FC single-ring events have leptons which remember the flavor of the parent neutrinos; the contribution from neutral-current (NC) reactions is less than 10%. For the FC multi-ring events, the contamination of NC interactions is significant (~30%).

FC events are subjected to particle identification of the final-state particles. On the other hand, all the FC events were assumed to be $\mu$-like since the FC events comprise a 98% pure charged-current $\nu_\mu$ sample. The method adopted for the FC events identifies the particle types as $e$-like or $\mu$-like based on the pattern of each Cherenkov ring. A ring produced by an $e$-like ($e^\pm, \gamma$) particle exhibits more diffuse pattern than that produced by a $\mu$-like ($\mu^\pm, \pi^\pm$) particle, since an $e$-like particle produces an electromagnetic shower and low-energy electrons suffer considerable multiple Coulomb scattering in water.

For FC events, another classification of neutrino events is defined in terms of the visible energy, $E_{\text{vis}}$. The sub-GeV events are defined as those events with $E_{\text{vis}} < 1.33$ GeV and a lower momentum cut of 100 (200) MeV/c for $e$-like ($\mu$-like) events. The multi-GeV events are defined as those events with $E_{\text{vis}} > 1.33$ GeV.

The zenith-angle dependence of each category of events observed in SK-I is shown in Fig. 2. Here, the FC sub-GeV events are separately shown in two parts ($P \leq 400$ MeV/c and $P > 400$ MeV/c, where $P$ denotes the lepton momentum). The box histograms show the Monte Carlo expectation for the hypothesis of no neutrino oscillations. One notices that the zenith-angle distributions of the FC $\mu$-like events and PC events show a strong deviation from the Monte Carlo expectations. The zenith-angle distributions of the upward-going muons also show deviations from the Monte Carlo expectations. On the other hand, the zenith-angle distribution of the FC $e$-like events is consistent with the Monte Carlo expectation.

The observed zenith-angle dependence of atmospheric neutrino events suggests $\nu_\mu \leftrightarrow \nu_\tau$ oscillations. A two-neutrino oscillation analysis with the hypothesis of $\nu_\mu \leftrightarrow \nu_e$ was made with use of all categories of events shown in Fig. 2 as well as upward-going muon events. The best fit was obtained at a slightly unphysical value of $\sin^2 2\theta = 1.02$ and $\Delta m^2 = 2.1 \times 10^{-3}$ eV$^2$ with $\chi^2_{\text{min}}$/DOF (degrees of freedom) = 174.5/177. By limiting the parameter space to the physical region, the minimum $\chi^2$ value, $\chi^2_{\text{min}}$/DOF = 174.8/177, was obtained at $\sin^2 2\theta = 1.00$ and $\Delta m^2 = 2.1 \times 10^{-3}$ eV$^2$. On the other hand, assuming no oscillation ($\sin^2 2\theta = 0$ and $\Delta m^2 = 0$), a $\chi^2$ value was found to be 478.7 for 179 DOF. It is very striking that the observed $\nu_\mu \leftrightarrow \nu_\tau$ mixing is consistent with maximal mixing.

Figure 3 shows the 68%, 90%, and 99% CL allowed regions in the ($\sin^2 2\theta, \Delta m^2$) plane. In Fig. 2, the expectation for $\nu_\mu \leftrightarrow \nu_\tau$ oscillations with the oscillation parameters corresponding to the $\chi^2$ minimum in the physical region is shown by the solid lines.

Though the SK-I atmospheric neutrino observations gave compelling evidence for neutrino flavor conversion $\nu_\mu \leftrightarrow \nu_\tau$ which is consistent with vacuum neutrino oscillations, 35
other exotic explanations such as neutrino decay\textsuperscript{36,37} and quantum decoherence\textsuperscript{38} cannot be completely ruled out from the zenith-angle distributions alone. The firm evidence for neutrino oscillation is to confirm characteristic sinusoidal behavior of the conversion probability as a function of neutrino energy \(E\) for a fixed distance \(L\) in the case of long baseline neutrino oscillation experiments, or as a function of \(L/E\) in the case of atmospheric neutrino experiments. By selecting events with high \(L/E\) resolution, evidence for the dip in the \(L/E\) distribution was observed at the right place expected from the interpretation of the SK-I data in terms of \(\nu_\mu \leftrightarrow \nu_\tau\) oscillations\textsuperscript{39} (see Fig. 4). This dip cannot be explained by alternative hypotheses of neutrino decay and neutrino decoherence, and they are excluded at more than 3\(\sigma\) in comparison with the neutrino oscillation interpretation. At 90\% CL, the constraints obtained from the \(L/E\) analysis are \(1.9 \times 10^{-3} < \Delta m^2 < 3.0 \times 10^{-3}\) eV\(^2\) and \(\sin^2 2\theta > 0.90\). These results are consistent with the SK-I final results of the combined zenith-angle analysis of fully-contained, partially-contained, and upward-going muon events (see Fig. 3).

When the Super-Kamiokande Collaboration announced the compelling evidence for atmospheric \(\nu_\mu \leftrightarrow \nu_\tau\) oscillations in 1998,\textsuperscript{11} two other underground experiments, MACRO and Soudan 2, reported supporting evidence from respective atmospheric neutrino observations. The final results from these experiments also indicate \(\nu_\mu \leftrightarrow \nu_\tau\) oscillations consistent with the SK-I atmospheric neutrino results.

The Soudan 2 detector was a 963 ton fine-segmented iron-tracking calorimeter with a 770 ton fiducial volume, located at a depth of 2070 m of water equivalent underground of the Soudan Mine in Minnesota. The Soudan 2 group has analyzed the neutrino \(L/E\) distributions with the hypotheses of neutrino oscillations and no neutrino oscillations.\textsuperscript{40} The probability of the no oscillation hypothesis is \(5.8 \times 10^{-4}\), and the 90\% CL allowed region in the \((\sin^2 2\theta, \Delta m^2)\) plane, consistent with that from the SK-I atmospheric neutrino observations, is obtained.

MACRO was a large-area multipurpose underground detector located at the Gran Sasso Laboratory at an average overburden of 3700 m water equivalent. The active detector elements were limited streamer tube planes for tracking and liquid scintillation counters. As a neutrino detector, MACRO measured the angular distribution of the upward-going muon flux.\textsuperscript{41} The oscillation analysis of the MACRO data\textsuperscript{41,42} supports \(\nu_\mu \leftrightarrow \nu_\tau\) oscillations with parameters \(\Delta m^2 = 2.5 \times 10^{-3}\) eV\(^2\) and \(\sin^2 2\theta \sim 1\). This result is also consistent with the SK-I allowed region.

### 4.2 Results from accelerator experiments

The \(\Delta m^2 \gtrsim 2 \times 10^{-3}\) eV\(^2\) region can be explored by accelerator-based long baseline experiments with typically \(E \sim 1\) GeV and \(L \sim\) several hundred km. With a fixed baseline distance and narrower neutrino energy spectrum, the value of \(\Delta m^2\), and also with higher statistics, the mixing angle, are potentially better constrained in accelerator experiments than from atmospheric neutrino observations.

#### 4.2.1 K2K

The K2K (KEK-to-Kamioka) long baseline neutrino oscillation experiment\textsuperscript{43} is the first accelerator-based experiment with a neutrino path length extending hundreds of kilometers. A horn-focused wide-band muon neutrino beam having an average \(L/E_\nu \sim 200\) \((L = 250\) km, \(E_\nu \sim 1.3\) GeV), was produced by 12-GeV protons from the KEK-PS and directed to the Super-Kamiokande detector. K2K aimed at confirmation of the neutrino oscillation in \(\nu_\mu\) disappearance in the \(\Delta m^2 \gtrsim 2 \times 10^{-3}\) eV\(^2\) region with well-understood flux and composition of the neutrino beam. For this purpose, the energy spectrum and profile of the neutrino beam were measured by a near neutrino detector system located 300 m downstream from the production target.

The K2K experiment started stable data-taking in June 1999. Super-Kamiokande events caused by accelerator-produced neutrinos were selected based on the timing information from the global positioning system. Data were intermittently taken until November 2004. The total number of protons on target for physics analysis amounted to \(0.92 \times 10^{20}\). The observed number of beam-originated
coordinates. It is 0.0015% or the likelihood ratio of the no-oscillation case to the best fit point instead of neutrino oscillation is estimated by computing the probability that the observations are due to a statistical fluctuation consistent with muon neutrino oscillations. The probability energy spectrum of neutrinos at Super-Kamiokande are compared with the expected spectrum at Super-Kamiokande for neutrino oscillation. The measured energy spectrum shows distortion which is characteristic of neutrino oscillations. Figure 5 shows the reconstructed $E_\nu$ distribution for these events. This distribution is compared with the expected spectrum at Super-Kamiokande for no neutrino oscillation. The measured energy spectrum shows distortion which is characteristic of neutrino oscillations.

A two-flavor neutrino oscillation analysis was performed using a maximum-likelihood method. The oscillation parameters ($\sin^2 2\theta, \Delta m^2$) are estimated by maximizing the product of the likelihood for the observed number of FC events and that for the shape of the reconstructed $E_\nu$ spectrum. The best fit point lies in the unphysical region, ($\sin^2 2\theta, \Delta m^2$) = (1.2, 2.6 $\times 10^{-3}$ eV$^2$). The allowed $\Delta m^2$ region at $\sin^2 2\theta = 1.0$ is between 1.9 and 3.5 $\times 10^{-3}$ eV$^2$ at the 90% CL with the best-fit value of 2.8 $\times 10^{-3}$ eV$^2$. This region is consistent with the allowed region from the SK-I atmospheric neutrino observations. The $E_\nu$ distribution calculated with the best-fit parameters in the physical region is shown in Fig. 5. The observed number of events and energy spectrum of neutrinos at Super-Kamiokande are consistent with muon neutrino oscillations. The probability that the observations are due to a statistical fluctuation instead of neutrino oscillation is estimated by computing the likelihood ratio of the no-oscillation case to the best fit point in the physical region. It is 0.0015% or 4.3 $\sigma$.

4.2.2 MINOS

MINOS is a long baseline neutrino oscillation experiment with near and far detectors. Neutrinos are produced by using 120 GeV protons from the Fermilab Main Injector. The far detector is a 5.4 kton (total mass) iron-scintillator tracking calorimeter with toroidal magnetic field, located underground in the Soudan mine. The baseline distance is 735 km. The near detector is also an iron-scintillator tracking calorimeter with toroidal magnetic field, with a total mass of 0.98 kton. The neutrino beam is a horn-focused wide-band beam. Its energy spectrum can be varied by moving the target position relative to the first horn and changing the horn current.

MINOS started the neutrino-beam run in 2005. Initial results were published$^{44}$ using the data taken between May 2005 and February 2006. During this period, a "low-energy" option was chosen for the spectrum of the neutrino beam so that the flux was maximized in the 1–3 GeV energy range. With $1.27 \times 10^{20}$ protons on the production target, 215 events with reconstructed neutrino energy $E_\nu < 30$ GeV were observed in the far detector. This number is contrasted with an expectation of $336 \pm 14$ events for no neutrino oscillations. Figure 6 compares the reconstructed $E_\nu$ spectrum at the far detector with the predicted spectrum with and without oscillations. Figure 7 shows the 68 and 90% CL allowed regions, compared with the 90% CL allowed region at $\Delta|\sin^2 2\theta| < 0.2$.

Fig. 5. Reconstructed $E_\nu$ distribution for 1-ring $\mu$-like sample. The solid histogram shows the best-fit spectrum. The dashed histogram shows the expected spectrum without oscillation. These histograms are normalized by the number of events observed (58). This figure is taken from ref. 43.

Fig. 6. Comparison of the far detector spectrum with predictions for no oscillations for two different analysis methods (shown by the gray lines) and for $\nu_\mu \rightarrow \nu_e$ oscillations with the best-fit parameters in one of the analysis methods (shown by the black lines). The last energy bin contains events between 18–30 GeV. This figure is taken from ref. 44.

Fig. 7. Allowed region for the $\nu_\mu \rightarrow \nu_e$ oscillation parameters from the initial MINOS results. The 68 and 90% CL allowed regions are shown together with the SK-I and K2K 90% CL allowed regions. This figure is taken from ref. 44.
regions obtained from the SK-I zenith-angle dependence, the SK-I $L/E$ analysis, and the K2K results. The best fit point is in the unphysical region, with parameters $\Delta m^2 = 2.72 \times 10^{-3}$ eV$^2$ and $\sin^2 2\theta = 1.01$. The initial MINOS results are consistent with the SK-I and K2K results.

5. Measurements of $\Delta m^2_{12}$ and $\Theta_{12}$

5.1 Solar neutrino observations

Observation of solar neutrinos directly addresses the theory of stellar structure and evolution, which is the basis of the standard solar model (SSM). The Sun as a well-defined neutrino source also provides extremely important opportunities to investigate nontrivial neutrino properties such as nonzero mass and mixing, because of the wide range of matter density and the great distance from the Sun to the Earth.

The solar neutrinos are produced by some of the reactions in the pp chain or CNO cycle. There have been efforts to calculate solar neutrino fluxes from these reactions on the basis of SSM. A variety of input information is needed in the evolutionary calculations. The most elaborate SSM calculations have been developed by Bahcall and his collaborators, who define their SSM as the solar model which is constructed with the best available physics and input data. The currently preferred SSM is BS05(OP) developed by Bahcall and Serenelli. Its prediction for the fluxes from neutrino-producing reactions is given in Table I. The solar-neutrino spectra calculated with this model is shown in Fig. 8.

So far, solar neutrinos have been observed by chlorine (Homestake) and gallium (SAGE, GALLEX, and GNO) radiochemical detectors and water Cherenkov detectors using light water (Kamiokande and Super-Kamiokande) and heavy water (SNO).

A pioneering solar neutrino experiment by Davis and collaborators at Homestake using $^{37}$Cl started in the late 1960’s. This experiment exploits electron neutrino absorption on chlorine nuclei followed by their decay through orbital electron capture,

$$^{37}\text{Cl} + \nu_e \rightarrow ^{37}\text{Ar} + e^- \text{ (threshold 814 keV).}$$

The $^{37}$Ar atoms produced are radioactive, with a half life ($\tau_{1/2}$) of 34.8 days. After an exposure of the detector for two to three times $\tau_{1/2}$, the reaction products are chemically extracted and introduced into a low-background proportional counter, where they are counted for a sufficiently long period to determine the exponentially decaying signal and a constant background. Solar-model calculations predict that the dominant contribution in the chlorine experiment comes from $^8$B neutrinos, but $^7$Be, $p\bar{p}$, $^{13}$N, and $^{15}$O neutrinos also contribute (for notations, refer to Table I).

From the very beginning of the solar-neutrino observation, it was recognized that the observed flux was significantly smaller than the SSM prediction, provided nothing happens to the electron neutrinos after they are created in the solar interior. This deficit has been called “the solar-neutrino problem”.

Gallium experiments (GALLEX and GNO at Gran Sasso in Italy and SAGE at Baksan in Russia) utilize the reaction

$$^{71}\text{Ga} + \nu_e \rightarrow ^{71}\text{Ge} + e^- \text{ (threshold 233 keV).}$$

They are sensitive to the most abundant pp solar neutrinos. However, the solar-model calculations predict almost half of the capture rate in gallium is due to other solar neutrinos. GALLEX presented the first evidence of pp solar-neutrino observation in 1992. The GALLEX Collaboration formally finished observations in early 1997. Since April, 1998, a newly defined collaboration, GNO (Gallium Neutrino Observatory) continued the observations until April 2003. The complete GNO results are published in ref. 48. The GNO + GALLEX joint analysis results are also presented. SAGE initially reported very low flux, but later observed similar flux to that of GALLEX. The complete SAGE results are published in ref. 50.

In 1987, the Kamiokande experiment succeeded in real-time solar neutrino observation, utilizing $\nu e$ scattering

$$\nu_e + e^- \rightarrow \nu_e + e^-$$

in a large water-Cherenkov detector. These experiment take advantage of the directional correlation between the incoming neutrino and the recoil electron. This feature greatly helps the clear separation of the solar-neutrino signal from the background. The Kamiokande result gave the first direct evidence that the Sun emits neutrinos. Later, the high-statistics Super-Kamiokande experiment took over the Kamiokande experiment. Due to the high thresholds (7 MeV), the neutrino events are

<table>
<thead>
<tr>
<th>Reaction</th>
<th>Abbr.</th>
<th>Flux (cm$^{-2}$ s$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$pp \rightarrow de^-e^+$</td>
<td>$pp$</td>
<td>$5.99(1.00 \pm 0.01) \times 10^{10}$</td>
</tr>
<tr>
<td>$pe^- \rightarrow d\nu_e$</td>
<td>$pep$</td>
<td>$1.42(1.00 \pm 0.02) \times 10^8$</td>
</tr>
<tr>
<td>$^7\text{He}^\rightarrow \nu_e + ^7\text{Li}^+ + (y)$</td>
<td>$^7\text{Be}$</td>
<td>$7.93(1.00 \pm 0.16) \times 10^3$</td>
</tr>
<tr>
<td>$^7\text{Be}^\rightarrow ^7\text{Li}e^- + e^-$</td>
<td>$^7\text{Be}$</td>
<td>$4.84(1.00 \pm 0.11) \times 10^6$</td>
</tr>
<tr>
<td>$^7\text{B}^\rightarrow ^7\text{Be}^\nu_e + e^-$</td>
<td>$^7\text{B}$</td>
<td>$5.69(1.00 \pm 0.16) \times 10^6$</td>
</tr>
<tr>
<td>$^{13}\text{N}^\rightarrow ^{13}\text{C}e^- + e^-$</td>
<td>$^{13}\text{N}$</td>
<td>$3.07(1.00(0.31)^{+0.31}_{-0.31}) \times 10^6$</td>
</tr>
<tr>
<td>$^{15}\text{O}^\rightarrow ^{15}\text{N}e^- + e^-$</td>
<td>$^{15}\text{O}$</td>
<td>$2.33(1.00(0.62)^{+0.62}_{-0.62}) \times 10^6$</td>
</tr>
<tr>
<td>$^{17}\text{F}^\rightarrow ^{17}\text{O}e^- + e^-$</td>
<td>$^{17}\text{F}$</td>
<td>$5.84(1.00 \pm 0.52) \times 10^6$</td>
</tr>
</tbody>
</table>
in Kamiokande and 5 MeV at present in Super-Kamiokande) the experiments observe pure $^8$B solar neutrinos because hep neutrinos contribute negligibly according to the SSM. It should be noted that the reaction (21) is sensitive to all active neutrinos, $\nu = e, \mu, \tau$, and $\tau$. However, the sensitivity to $\nu_\mu$ and $\nu_\tau$ is much smaller than the sensitivity to $\nu_e$, $\sigma(\nu_\mu, e) \approx 0.16 \sigma(\nu_e,e)$.

In 1999, a new real time solar-neutrino experiment, SNO, in Canada started observation. This experiment uses 1000 tons of ultra-pure heavy water ($D_2O$) contained in a spherical acrylic vessel, surrounded by an ultra-pure $H_2O$ shield. SNO measures $^8$B solar neutrinos via the reactions

$$\nu_e + d \rightarrow e^- + p + p \quad (22)$$

and

$$\nu_x + d \rightarrow \nu_x + p + n \quad (23)$$
as well as $\nu e$ scattering, (21). The charged-current (CC) reaction, (22), is sensitive to electron neutrinos, while the NC reaction, (23), is sensitive to all active neutrinos.

The $Q$-value of the CC reaction is $-1.4 \text{MeV}$ and the electron energy is strongly correlated with the neutrino energy. Thus, the CC reaction provides an accurate measure of the shape of the $^8$B solar-neutrino spectrum. The contributions from the CC reaction and $\nu e$ scattering can be distinguished by using different $\cos \theta_e$ distributions where $\theta_e$ is the angle of the electron momentum with respect to the direction from the Sun to the Earth. While the $\nu e$ scattering events have a strong forward peak, CC events have an approximate angular distribution of $1 - 1/3 \cos \theta_e$.

The threshold of the NC reaction is 2.2 MeV. In the pure $D_2O$, the signal of the NC reaction is neutron capture in deuterium, producing a 6.25-MeV $\gamma$-ray. In this case, the capture efficiency is low and the deposited energy is close to the detection threshold of 5 MeV. In order to enhance both the capture efficiency and the total $\gamma$-ray energy (8.6 MeV), 2 tons of NaCl were added to the heavy water in the second phase of the experiment. In addition, discrete $^3$He neutron counters were installed and the NC measurement with them is made as the third phase of the SNO experiment.

Figure 9 compares the predictions of the BP2005(OP) SSM with the results of solar neutrino experiments. It is clearly seen from Fig. 9 that the results from all the solar-neutrino experiments, except the SNO’s NC result, indicate significantly less flux than expected from the solar-model predictions.\[45\]

5.2 Evidence for solar neutrino flavor conversion

The solar-neutrino problem had remained unsolved for more than 30 years. However, there have been remarkable developments in the past six years and now the solar-neutrino problem has been finally solved. In 2001, the initial SNO CC result combined with the Super-Kamiokande’s high-statistics $\nu e$ elastic scattering (ES) result\[53\] provided direct evidence for flavor conversion of solar neutrinos.\[54\] Later, SNO’s NC measurements further strengthened this conclusion.\[55,56\] From the recent salt phase measurement, the fluxes measured with CC, ES, and NC events were obtained as

$$\phi^\text{CC}_{\text{SNO}}(\nu_e) = (1.68 \pm 0.06^{+0.08}_{-0.09}) \times 10^6 \text{cm}^{-2}\text{s}^{-1}, \quad (24)$$

$$\phi^\text{ES}_{\text{SNO}}(\nu_e) = (2.35 \pm 0.22 \pm 0.15) \times 10^6 \text{cm}^{-2}\text{s}^{-1}, \quad (25)$$

$$\phi^\text{NC}_{\text{SNO}}(\nu_e) = (4.94 \pm 0.21^{+0.38}_{-0.34}) \times 10^6 \text{cm}^{-2}\text{s}^{-1}, \quad (26)$$

where the first errors are statistical and the second errors are systematic. Equation (26) is a mixing-independent result and therefore tests solar models. It shows good agreement with the $^8$B solar-neutrino flux predicted by the solar model.\[45\] Figure 10 shows the salt phase result of $\phi(\nu_{\mu or \tau})$ versus the flux of electron neutrinos $\phi(\nu_e)$ with the 68, 95, and 99% joint probability contours. The flux of non-$\nu_e$ active neutrinos, $\phi(\nu_{\mu or \tau})$, can be deduced from these results. It is

$$\phi(\nu_{\mu or \tau}) = (3.26 \pm 0.25^{+0.49}_{-0.33}) \times 10^6 \text{cm}^{-2}\text{s}^{-1}. \quad (27)$$

The non-zero $\phi(\nu_{\mu or \tau})$ is strong evidence for neutrino flavor conversion. A natural and most probable explanation of neutrino flavor conversion is neutrino oscillation. At this stage, the LMA (large mixing angle) solution of solar neutrino flavor conversion is neutrino oscillation. At this stage, the LMA (large mixing angle) solution of solar neutrino oscillation is neutrino oscillation. At this stage, the LMA (large mixing angle) solution of solar neutrino oscillation is neutrino oscillation. At this stage, the LMA (large mixing angle) solution of solar neutrino oscillation is neutrino oscillation. At this stage, the LMA (large mixing angle) solution of solar neutrino oscillation is neutrino oscillation.
event deficit expected from neutrino oscillation. This result showed clear evidence of reactors contributing to the \( \tan^2 \theta \) and \( \delta \) solar + KamLAND analysis are

\[
\begin{align*}
N_{\text{obs}} - N_{\text{BG}} & = 0.611 \pm 0.085 \pm 0.041 \quad (28)
\end{align*}
\]

with obvious notation. This result showed clear evidence of event deficit expected from neutrino oscillation.

The first KamLAND results \(^\text{58}\), with 162 ton-yr exposure were reported in December 2002. The ratio of observed to expected (assuming no neutrino oscillation) number of events was

\[
\frac{N_{\text{obs}} - N_{\text{BG}}}{N_{\text{NoOsc}}} = 0.611 \pm 0.085 \pm 0.041
\]

In June 2004, KamLAND released the results from 766 ton-yr exposure.\(^\text{59}\) In addition to the deficit of events, the observed positron spectrum showed the distortion expected from neutrino oscillation as can be seen in Fig. 11. Here, the ratio of the observed \( \bar{\nu}_e \) spectrum to the expectation without oscillation is plotted as a function of \( L_0/E \) with \( L_0 = 180 \text{ km} \). It is a sort of average distance of nuclear reactors contributing to the \( \bar{\nu}_e \) flux detected in KamLAND, determined as if all anti-neutrinos detected in KamLAND were due to a single reactor at this distance. Figure 12 shows the allowed regions in the neutrino-oscillation parameter space. The best-fit point lies in the region called LMA I.

### 5.4 Global neutrino oscillation analysis

The SNO Collaboration updated \(^\text{56}\) a global two-neutrino oscillation analysis of the solar-neutrino data including the SNO’s complete salt phase data, and global solar + KamLAND 766 ton-yr data.\(^\text{39}\) The best fit parameters for the global solar analysis are \( \Delta m^2 = 6.5_{-2.3}^{+2.1} \times 10^{-5} \text{ eV}^2 \) and \( \tan^2 \theta = 0.45^{+0.09}_{-0.08} \). The inclusion of the KamLAND data significantly constrains the allowed \( \Delta m^2 \) region, but shifts the best-fit \( \Delta m^2 \) value. The best-fit parameters for the global solar + KamLAND analysis are \( \Delta m^2 = 8.0^{+0.6}_{-0.4} \times 10^{-5} \text{ eV}^2 \) and \( \tan^2 \theta = 0.45^{+0.09}_{-0.07} \) or \( \theta = (33.9^{+2.4}_{-2.2})^\circ \).

### 6. Future Prospects

Some of the important problems left for the future neutrino experiments are: (1) Absolute neutrino mass; (2) Determining whether neutrinos are Majorana particles or Dirac particles; (3) Measurement of the small mixing angle \( \theta_{13} \) and the CP violating phase \( \delta \) in the MNS matrix; (4) Measurement of the sign of \( m_3^2 \).

Measurement of absolute neutrino mass will be challenged by a new tritium beta-decay experiment KATRIN\(^\text{61}\) with a sensitivity down to \( \sim 0.2 \text{ eV} \). As noted in §1, giving an answer to the question of whether neutrinos are Majorana or Dirac particles is of fundamental importance. Detection of neutrinoless double beta decay is the only realistic means to show the Majorana nature of neutrinos. There are a number of proposed neutrinoless double beta decay experiments aiming at reaching sensitivities to the effective Majorana mass of \( 10-50 \text{ meV} \). For these experiments, readers are referred to ref. 22, and references therein.
Items (3) and (4) are the main goals of the future neutrino oscillation experiments. If $\theta_{13}$ is not much less than the CHOOZ limit, $\sin^2 2\theta_{13} < 0.19$, the ongoing experiments MINOS and OPERA will have a chance to establish non-zero $\theta_{13}$. A new experiment under construction in Japan, T2K, will improve the sensitivity. T2K will use Super-Kamiokande as a far detector. An experiment proposed in US, NOvA will have a similar sensitivity. Both T2K and NOvA plan to use an “off-axis” beam. In this scheme, the axis of the beamline components is directed a few degrees away from the direction of the far detector. With this trick, a high-intensity, low-energy, narrow-band neutrino beam can be obtained. The peak of the neutrino energy spectrum can be adjusted by varying the off-axis angle. It will be adjusted close to the oscillation maximum in order to maximize the sensitivity to the oscillation measurement.

Short baseline reactor neutrino oscillation experiments also aim at the measurement of $\theta_{13}$. Double-CHOOZ experiment in France will have a sensitivity of $\sin^2 2\theta_{13} < 0.06$ with the far detector only, and $< 0.03$ when the near detector is completed. Daya Bay experiment in China aims at a sensitivity of $\sin^2 2\theta_{13} < 0.01$.

In the accelerator neutrino oscillation experiments with conventional neutrino beams, $\theta_{13}$ is measured using $\nu_\mu \rightarrow \nu_e$ appearance. With the same approximation used for eq. (11), the dominant term in the probability of $\nu_\mu \rightarrow \nu_e$ appearance is

$$P(\nu_\mu \rightarrow \nu_e) = \sin^2 2\theta_{13} \sin^2 \theta_{23} \sin^2 (1.27 \Delta m^2_{23} / E).$$  (29)

However, by examining the exact expression for the oscillation probability, it is understood that some of the neglected terms have rather large effects and the unknown CP-violating phase $\delta$ causes uncertainties in determining the value of $\theta_{13}$. Actually, from the measurement of $\nu_\mu \rightarrow \nu_e$ appearance, $\theta_{13}$ is given as a function of $\delta$ for a given sign of $\Delta m^2_{23}$. Therefore, a single experiment with a neutrino beam cannot determine the value of $\theta_{13}$ though it is possible to establish non-zero $\theta_{13}$. On the other hand, since the disappearance probability does not depend on the CP-violating phase, reactor $\nu_e$ disappearance experiments allow to measure $\theta_{13}$ with less ambiguities in contrast to accelerator $\nu_\mu \rightarrow \nu_e$ appearance experiments.

In order to determine the value of $\theta_{13}$ in the accelerator neutrino oscillation experiments, simultaneous measurement of the CP-violating phase $\delta$ is necessary. This will be achieved by measuring the appearance probabilities of $\nu_\mu \rightarrow \nu_e$ and $\bar{\nu}_\mu \rightarrow \bar{\nu}_e$. To measure CP violation in this way as well as to determine whether the mass hierarchy is normal or inverted one, a very high intensity neutrino beam and a Mton class detector will be required. A “super-beam” means a very high intensity, conventional neutrino beam line ($\nu_\mu$ or $\bar{\nu}_\mu$ beam obtained from decays of horn-focused secondary pions). According to the T2K Phase II (T2K-II) plan, the beam power should be increased to 4 MW and a 1 Mton water Cherenkov detector called Hyper-Kamiokande should be constructed. T2K-II is an example of a super-beam experiment.

It should be noted there are famous degeneracies of parameters which give the same appearance probabilities. There are three types of degeneracies. One is a two-fold degeneracy of $(\theta_{13}, \delta)$. Another two-fold degeneracy is caused by the sign of $\Delta m^2_{23}$. If $\theta_{23} \neq \pi / 4$, yet another two-fold degeneracy arises by the interchange of $\theta_{23} \leftrightarrow \pi/2 - \theta_{23}$. For a single experiment with a given neutrino beam spectrum and a given detector at a fixed baseline distance, it is impossible to resolve these up to eight parameter degeneracies. Multiple measurements, or synergy of different experiments, are needed to resolve the degeneracies. An idea of the T2K experiment is to use the same off-axis beam with two 0.5 Mton class water Cherenkov detectors at different baseline distances. One detector will be located at Kamioka with $L \sim 300$ km. The other detector will be located in Korea with $L \sim 1000$ km. Since the T2KK experiment will provide measurements with two different baseline distances, it is expected to have a capability of resolving some of the parameter degeneracies. In fact, it has been shown that T2KK can resolve mass hierarchy if $\sin^2 2\theta_{13} > 0.01$.

For this measurement, the baseline distance as long as 1000 km is essential because matter effects are needed. (Note that the T2K experiment with a baseline distance of $\sim 300$ km is not sensitive to mass hierarchy.) Furthermore, T2KK has high sensitivity to CP-violation measurement and also to resolving $\delta_{23}$ degeneracy if $\sin^2 2\theta_{13}$ is in the reach of the Phase-I T2K experiment.

If $\sin^2 2\theta_{13} < 0.01$, its measurement would be beyond the reach of the on-going projects (Phase-I T2K experiment and reactor experiments, etc.) Even super-beam experiments have a sensitivity down to $\sin^2 2\theta_{13} \sim$ a few $\times 10^{-3}$. To enhance the discovery reach, advanced concepts of producing neutrino beams, “neutrino factory” and “beta-beam” facility have been proposed.

A neutrino factory is an intense neutrino source based on a 20–50 GeV muon storage ring. If $10^{20–21}$ muons per year decay in the straight section of the ring, a very intense and highly collimated neutrino beam with well understood properties will be produced. The main modes of neutrino oscillations investigated with beams from the neutrino factory are $\nu_\mu \rightarrow \nu_\mu$ and $\bar{\nu}_\mu \rightarrow \bar{\nu}_e$. If positive muons are accumulated, the produced neutrino beam consists of $\nu_\mu$ and $\nu_e$ components. In the detector, the CC interactions of the beam $\nu_\mu$’s produce positive muons. On the other hand, negative muons are produced by the CC interactions of $\nu_e$’s appeared in the beam due to $\nu_e \rightarrow \nu_\mu$ oscillations. Thus, the signal of neutrino oscillations is wrong-sign muons and very clean measurement is possible.

The concept of a beta-beam was proposed by Zucchelli. In this scheme, $\bar{\nu}_e$ or $\nu_e$ beams can be produced from the beta decay of boosted ions; $^7\text{He} \rightarrow \beta^+ \nu_e$ or $^{18}\text{Ne} \rightarrow \beta^- \nu_e$. High-intensity radioactive ions can be produced, for example, at CERN ISOLDE. After acceleration, they are stored in a storage ring having a long straight section. The resulting neutrino beam is strongly collimated, has well-known energy spectrum and intensity, and contains a purely single neutrino flavor. For a recent review, see ref. 76.

Experiments using the neutrino factory beam and the beta-beam will have remarkably good physics potential. The sensitivities to the $\sin^2 2\theta_{13}$ measurement reach $\sim 10^{-4}$. For the CP-violation and mass hierarchy measurements, they have sensitivities down to $\sin^2 2\theta_{13} \sim$ a few $\times 10^{-4}$.  

111008-10
7. Conclusions

In the last 10 years, our understanding of neutrino mass and mixing has advanced remarkably. Neutrino oscillations have been firmly established. Finite neutrino mass is the first evidence of physics beyond the Standard Model. The remarkably large mixing observed in the neutrino sector is strikingly different from the mixing in the quark sector.

In order to gain complete knowledge of the MNS matrix, future neutrino oscillation experiments should measure the last mixing angle \( \theta_{13} \) and the CP-violating phase \( \delta \), and determine the sign of \( \Delta m_{21}^2 \). Measurement of the absolute neutrino mass scale and determining whether neutrinos are Majorana particles or Dirac particles are also very important goals of future neutrino experiments.
Muons: A Pillar of the Standard Model

Bradley Lee Roberts*

Department of Physics, Boston University, Boston, MA 02215, U.S.A.

(Received April 8, 2007; accepted April 18, 2007; published November 12, 2007)

Since its discovery in the 1930s, the muon has played an important role in our quest to understand the subatomic theory of matter. The muon was the first second-generation standard-model particle to be discovered, and its decay has provided information on the (Vector-Axial Vector) structure of the weak interaction, the strength of the weak interaction, \( g_\mu \), and the conservation of lepton number (flavor) in muon decay. The muon’s anomalous magnetic moment has played an important role in restricting theories of physics beyond the standard model, where at present there is a 3.4\( \sigma \) difference between the experiment and standard-model theory. Its capture on the atomic nucleus has provided valuable information on the modification of the weak current by the strong interaction which is complementary to that obtained from nuclear \( \beta \) decay.

KEYWORDS: muon, weak decay, muon capture, magnetic moment, lepton flavor violation

DOI: 10.1143/JPSJ.76.111009

1. Introduction

The muon was first observed in a Wilson cloud chamber by Kunze\(^1\) in 1933, where it was reported to be “a particle of uncertain nature”. In 1936 Anderson and Neddermeyer\(^2\) reported the presence of “particles less massive than protons but more penetrating than electrons” in cosmic rays, which was confirmed in 1937 by Street and Stevenson.\(^3\) Nishina, Tekeuchi, and Ichimiya,\(^4\) and by Crussard and Leprince-Ringuet.\(^5\) The Yukawa theory of the nuclear force had predicted such a particle, but this “mesotron” as it was called, interacted too weakly with matter to be the carrier of the strong force. Today we understand that the muon is a second generation lepton, with a mass of about 207 times that of the electron’s. Like the electron, the muon obeys quantum electrodynamics (QED), and can interact with other particles through the electromagnetic and weak forces. Unlike the electron which appears to be stable, the muon decays through the weak force.

The muon lifetime of 2.2\( \mu s \) permits one to make precision measurements of its properties, and to use it as a tool to study the semileptonic weak interaction, nuclear properties, as well as magnetic properties of condensed matter systems. The high precision to which the muonium (\( \mu^+e^- \) atom) hyperfine structure can be measured and calculated makes it a significant input parameter in the determination of fundamental constants.\(^6\) In this review, I will focus on the role of the muon in particle physics.

A beam of negative muons can be brought to rest in matter, where hydrogen-like atoms are formed, with a nuclear charge of \( Z \). The Bohr radius for a hydrogen-like atom is inversely proportional to the orbiting particle’s mass \( (r_n = [n^2 \hbar^2 c]/[m_e^2 Z a_0]) \), so that for the lowest quantum numbers of high-\( Z \) muonic atoms, the muon is well inside of the atomic electron cloud, with the Bohr radius of the 1S atomic state well inside the nucleus. The \( 2P \rightarrow 1S \) x-ray energies are shifted because of the modification of the Coulomb potential inside the nucleus, and these x-rays have provided information on nuclear root-mean-square charge radii. The Lamb shift in muonic hydrogen, \( \Delta E_{2p-2s} \), which is being measured at the Paul Scherrer Institut (PSI), is given by\(^7\) \( (209.974(6) - 5.226 R_p^2 + 0.036 R_p^3) \) meV, where \( R_p \) is the proton rms charge radius. This experiment should provide a precise measurement of \( R_p \). The weak nuclear capture, called ordinary muon capture (OMC), of the muon on the atomic nucleus following the cascade to the 1S ground state, \( \mu^- + \tilde{\eta}N \rightarrow 1S + e^- + \nu_\mu \), is the analog to the weak capture of a K-shell electron by the nucleus, and provides information on the modification of the weak interaction by hadronic matter.

The muon mass of \( \sim 106 \) MeV restricts the muon to decay into the electron, neutrinos, and photons. Thus muon decay is a purely leptonic process, and the dominant decay mode is \( \mu^- \rightarrow e^- + \nu_\mu + \bar{\nu}_e \). This three-body decay tells us that the individual lepton number, electron and muon, is conserved separately, and that the two flavors (kinds) of neutrinos are distinct particles.\(^8\) Here the \( \mu^- \) and \( e^- \) are “particles” and the \( \mu^+ \) and \( e^+ \) are the antiparticles. In the 1950s, it became possible to make pions, and thus muons, in the laboratory. The energetically favorable decay \( \mu^+ \rightarrow e^+ \gamma \) was searched for and not found\(^9\) to a relative branching ratio of \( <2 \times 10^{-5} \). Also searched for was the neutrinoless capture of a \( \mu^- \) on an atomic nucleus,\(^9\) \( \mu^- + \tilde{\eta}N \rightarrow e^- + N \), which was not found at the level of \( \sim 5 \times 10^{-4} \). Such processes are said to “violate lepton flavor”, and continue to be the object of present and planned studies reaching to sensitivities of \( 10^{-14} \) and \( 10^{-16} \), respectively.

The muon, like the electron, is a spin 1/2 lepton, with a magnetic moment given by

\[
\mu_s = g_s \left( \frac{q}{2m} \right) \gamma; \quad \mu = (1 + a) \frac{q\hbar}{2m}; \quad a = \frac{g_\mu - 2}{2};
\]

where the muon charge \( q = \pm e \), and \( g_s \), the spin \( g \)-factor is slightly greater than the Dirac value of 2. The middle equation above is useful from a theoretical point of view, as it separates the magnetic moment into two pieces: the Dirac moment which is unity in units of the appropriate magneton, \( e\hbar/2m \), and is predicted by the Dirac equation; and the...

---

*E-mail: roberts@bu.edu
three generations of leptons, the "standard model" of subatomic physics, which incorporates
It is easy to forget how we reached what is now called the
surprise. Looking at this from our 21st century perspective,
discussed below.
the electron’s, arises from radiative corrections that are
While the V–A structure of the weak interaction was first inferred from nuclear β decay, the study
of muon decay has provided a useful laboratory in which to
study the purely leptonic weak interaction, to search for
physics beyond the standard model, such as additional terms
in the interaction besides the standard-model V–A structure,
as well as looking for standard model forbidden decays like
μ → eγ. For many years, the experimental value of the
muon’s anomalous magnetic moment has served to constrain
physics beyond the standard model, and continues that
role today.

2. Muon Decay and \( G_F \)

The muon decay \( \mu^- \rightarrow e^- γν_ν_\mu \) is purely leptonic. Since \( m_μ \ll M_W \), muon decay can be described by a local four-fermion (contact) interaction. While nonrenormalizable, at low energies it provides an excellent approximation to the
full electroweak theory. The weak Lagrangian is written as a current–current interaction, where the leptonic current is of the (V–A) form, \( ĵ_u μ \mu \) \( (1 − γ_5) μ \).

Michele(14) first wrote down a parameterization of muon decay, defining five parameters, \( p, n, \xi, δ, \) and \( h \), which are combinations of the different possible couplings allowed by Lorentz invariance in muon decay. The standard model has clear predictions for these parameters and they have been measured repeatedly over the intervening years to search for physics beyond the standard model. This tradition continues today, with the TWIST experiment at TRIUMF, which is
mid 2016s, none of this was clear. Quarks were viewed by
many as a mathematical device, not as constituent particles.
Even after quarks were inferred from deep inelastic electron scattering off the proton, we only knew of the existence of three of them. While the V–A structure of the weak interaction was first inferred from nuclear β decay, the study
of muon decay has provided a useful laboratory in which to
study the purely leptonic weak interaction, to search for
physics beyond the standard model, such as additional terms
in the interaction besides the standard-model V–A structure,
as well as looking for standard model forbidden decays like
\( \mu \rightarrow eγ \). For many years, the experimental value of the
muon’s anomalous magnetic moment has served to constrain
physics beyond the standard model, and continues that
role today.

3. Nuclear Muon Capture

The weak capture of a muon on a proton has much in common with nuclear β decay. As for other low-energy weak processes, the interaction can be described as a
current–current interaction with the (V–A) leptonic current
given by \( ĵ_u μ \mu \) \( (1 − γ_5) μ \). Because the strong interaction
can induce additional couplings, the hadronic current is more complicated. The most general form of the vector current allowed by Lorentz invariance is(20)

The corresponding form of the axial-vector current is

where \( m_μ \) and \( M_N \) are the muon and nucleon masses respectively; the \( g(q^2) \) are the induced form factors: vector, weak magnetism, scalar, axial-vector, pseudoscalar and
tensor. The scalar and tensor terms are called “second class currents” because of their transformation properties under
G-parity, and in the standard model are expected to be quite small. It is traditional to set these second-class currents
equal to zero.

Nuclear β decay is sensitive to the vector, axial-vector and weak-magnetism form factors, but in muon capture the
capture rate has a measurable contribution from the induced pseudoscalar interaction, the least well known of the weak nucleon form factors. Radiative muon capture (RMC), $\mu^- + p \to n + \gamma + \nu_\mu$, should in principle be more sensitive to the induced pseudoscalar coupling than OMC, since with the three-body final state, $q^2$ can get closer to the pion pole than is possible in ordinary muon capture, which was pointed out many years ago.\textsuperscript{21,22} The interested reader is referred to the review by Goriolle and Fearing for further discussion.\textsuperscript{19}

In the past, current algebra and the Goldberger–Trieber relation expressed $g_\mu$ in terms of $g_A$. With the development of quantum chromodynamics (QCD), and chiral perturbation theory, new interest has developed in the value of $g_\mu$.\textsuperscript{20,23} The presently accepted theory value is $g_\mu(q^2 = -0.88m^2_\mu) = 8.26 \pm 0.23$.\textsuperscript{19}

The experimental history is rather interesting. For many years the radiative capture reaction $\mu^- + p \to n + \gamma + \nu_\mu$ was thought to be the “golden” channel to study. However, it is this approach that the recent MuCap experiment at PSI has used. It is this approach that the recent MuCap experiment at PSI has used.

In modern notation, the magnetic dipole moment (MDM) interaction becomes

$$\bar{u}_\mu \left[ eF_1(q^2)q_\beta + \frac{ie}{2m_\mu} F_2(q^2)q_\beta q_\gamma \right] u_\mu,$$

where $F_1(0) = 1$ and $F_2(0) = a_\mu$. The electric dipole moment (EDM) interaction is

$$\bar{u}_\mu \left[ \frac{ie}{2m_\mu} F_2(q^2) - F_3(q^2) q_\gamma \right] q_\beta q^\gamma u_\mu,$$

where $F_2(0) = a_\mu$, $F_3(0) = d_\mu$, with

$$d_\mu = \left( \frac{\eta}{2} \right) \left( \frac{eh}{2mc} \right) \approx \eta \times 4.7 \times 10^{-14} \text{ e cm}.$$ (This $\eta$, which is the EDM analogy to $g$ for the MDM, should not be confused with the Michel parameter $\eta$.)

The existence of an EDM implies that both $P$ and $T$ are violated.\textsuperscript{25,27} This can be seen by considering the non-relativistic Hamiltonian for a spin one-half particle in the presence of both an electric and magnetic field: $\mathcal{H} = -\mathbf{\mu} \cdot \mathbf{B} - \mathbf{d} \cdot \mathbf{E}$. The transformation properties of $E$, $B$, $\mathbf{\mu}$, and $\mathbf{d}$ are given in Table I, and we see that while $\mathbf{\mu} \cdot \mathbf{B}$ is even under all three, $\mathbf{d} \cdot \mathbf{E}$ is odd under both $P$ and $T$. While parity violation has been observed in many weak processes, direct $T$ violation has only been observed in the neutral kaon system.\textsuperscript{30} In the context of CPT symmetry, an EDM implies CP violation, which is allowed by the standard model for decays in the neutral kaon and $B$-meson sectors.

Observation of a non-zero electron or muon EDM would be a clear signal for new physics. To date no permanent EDM has been observed for the electron, the neutron, or an atomic nucleus, with the experimental limits given in Table II. It is interesting to note that in his original paper Dirac stated “The electric moment, being a pure imaginary, we should not expect to appear in the model. It is doubtful

<table>
<thead>
<tr>
<th>$E$</th>
<th>$B$</th>
<th>$\mu$ or $d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P$</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>$C$</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>$T$</td>
<td>+</td>
<td>–</td>
</tr>
</tbody>
</table>
whether the electric moment has any physical meaning, since the Hamiltonian . . . that we started from is real, and the imaginary part only appeared when we multiplied it up in an artificial way in order to make it resemble the Hamiltonian of previous theories”. Even in the 4th edition of his quantum mechanics book from 1958, well after the suggestion of Purcell and Ramsey31) that one should search for a permanent EDM, Dirac held fast to this point of view.

While CP violation is widely invoked to explain the muon anomaly, it is sensitive to a number of potential candidates for physics beyond the standard model.41)

- (1) muon substructure, where the contribution depends on the substructure scale $\Lambda$ as
  $$\delta a_{\mu}(\Lambda_{\mu}) \simeq \frac{m_{\mu}}{\Lambda_{\mu}}. $$

- (2) $W$-boson substructure.
- (3) new particles that couple to the muon, such as the supersymmetric partners of the weak gauge bosons.
- (4) extra dimensions

The potential contribution from supersymmetry has generated a lot of attention,42,43) the relevant diagrams are shown in Fig. 3 below. A simple model with equal masses41) gives

$$a_{\mu}^{(\text{SUSY})} \approx \frac{\alpha(M_Z)}{8\pi \sin^2 \theta_W} \frac{m_{\mu}^2}{m^2} \tan \beta \left( 1 - \frac{4\alpha \ln \frac{m}{m_{\mu}}}{\pi} \right) $$

$$\approx (\text{sgn} \mu) \times 10^{-10} \tan \beta \left( \frac{100 \text{ GeV}}{m} \right)^2,$$

where $\tan \beta$ is the ratio of the two vacuum expectation values of the two Higgs fields. If the SUSY mass scale were known, then $a_{\mu}^{(\text{SUSY})}$ would provide a clean way to determine $\tan \beta$.

---

**Table II.** Measured limits on electric dipole moments, and their standard model values.

<table>
<thead>
<tr>
<th>Particle</th>
<th>Present EDM limit (e cm)</th>
<th>Standard model value (e cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$</td>
<td>$2.9 \times 10^{-26}$ (90% CL)</td>
<td>$10^{-31}$</td>
</tr>
<tr>
<td>$e^-$</td>
<td>$\sim 1.6 \times 10^{-27}$ (90% CL)</td>
<td>$10^{-38}$</td>
</tr>
<tr>
<td>$\mu$</td>
<td>$&lt;10^{-18}$ (CERN)</td>
<td>$10^{-35}$</td>
</tr>
<tr>
<td>$^{199}\text{Hg}$</td>
<td>$2.1 \times 10^{-28}$ (95% CL)</td>
<td></td>
</tr>
</tbody>
</table>

a) Estimated.
4.1 Measurement of the anomalous magnetic dipole moment

Measurement of the magnetic anomaly uses the spin motion in a magnetic field. For a muon moving in a magnetic field, the spin and momentum rotate with the frequencies:

$$\omega_S = -\frac{g q B}{2m} \frac{q B}{\gamma m} (1 - \gamma); \quad \omega_C = -\frac{g q B}{m \gamma}. \quad (15)$$

The spin precession relative to the momentum occurs at the difference frequency, $$\omega_a$$, between the spin and cyclotron frequencies, eq. (15),

$$\omega_a = \omega_S - \omega_C = \left(\frac{g - 2}{2}\right) \frac{q B}{m} = -\frac{a_\mu}{m}. \quad (16)$$

The magnetic field in eq. (16) is the average field seen by the ensemble of muons. This technique has been used in all but the first experiments by Garwin et al., which used stopping muons, to measure the anomaly. After Garwin et al., made a 12% measurement of the anomaly, a series of experiments culminated with a 7.3 ppm measure of $$a_\mu$$.

In the third CERN experiment, a new technique was developed based on the observation that electrostatic quadrupoles could be used for vertical focusing. With the velocity transverse to the magnetic field ($$\beta \cdot B = 0$$), the spin precession formula becomes

$$\omega_a = -\frac{q}{m} \left[ a_\mu B - \left( a_\mu - \frac{1}{\gamma^2 - 1} \right) \frac{\beta \times E}{c} \right]. \quad (17)$$

For $$\gamma_{\text{magic}} = 29.3$$ ($$p_{\text{magic}} = 3.09 \text{ GeV}/c$$), the second term vanishes; one is left with the simpler result of eq. (16), hence the name “magic”, and the electric field does not contribute to the spin precession relative to the momentum. There are two major advantages of using the magic $$\gamma$$ and a uniform magnetic field: (i) the knowledge needed on the muon trajectories to determine the average magnetic field is much less than when gradients are present, and the more uniform field permits NMR techniques to realize their full accuracy, thus increasing the knowledge of the B-field. The spin precession is determined almost completely by eq. (16), which is independent of muon momentum; all muons precess at the same rate. This technique was used also in experiment E821 at the Brookhaven National Laboratory Alternating Gradient Synchrotron (AGS). The reader is referred to ref. 38 for a discussion of muon decay relevant to E821, and to ref. 40 for details of E821.

Muons are stored in a storage ring, and the arrival time and energy of the decay electrons is measured. When a single energy threshold is placed on the decay electrons, the number of high-energy electrons is modulated by the spin precession frequency, eq. (17), producing the time distribution

$$N(t, E_{\text{th}}) = N_0(E_{\text{th}}) e^{-t/(\tau(t))} \times [1 + A(E_{\text{th}}) \cos(\omega_a t + \phi(E_{\text{th}}))]. \quad (18)$$

as shown in Fig. 4. The value of $$\omega_a$$ is obtained from a least-squares fit to these data. The five-parameter function [eq. (18)] is used as a starting point, but many additional small effects must be taken into account.

In E821, both $$\mu^+$$ and $$\mu^-$$ were measured, and assuming

$$CPT$$ invariance, the final result obtained by E821, $$a_\mu^\exp = 116 592 080 (63) \times 10^{-11}$$, is shown in Fig. 5, along with the individual measurements and the standard-model value. The present standard-model value is $$a_\mu^{\text{SM(06)}} = 116 591 785 (61) \times 10^{-11}$$, and one finds $$\Delta a_\mu = 295(88) \times 10^{-11}$$, a 3.4$$\sigma$$ difference.

One candidate for the cosmic dark matter is the lightest supersymmetric partner, the neutralino, $$\chi^0$$ in Fig. 3. In the context of a constrained minimal supersymmetric model (CMSSM), $$(g - 2)_\mu$$ provides an orthogonal constraint on dark matter from that provided by the WMAP survey, as can be seen in Fig. 6.

With the apparent 3.4$$\sigma$$ difference between theory and experiment, a new experiment to improve the error by a factor of 2 to 2.5 has been proposed to Brookhaven Laboratory, but at present it is not funded. The theoretical value will continue to be improved, both with the expected availability of additional data on $$e^+e^-$$ annihilation to hadrons, and with additional work on the hadronic light-light contribution.

4.2 The search for a muon electric dipole moment

With an EDM present, the spin precession frequency relative to the momentum must be modified. The total frequency becomes $$\omega = \omega_a + \omega_{\text{EDM}}$$, where
and below the storage region would detect a time-dependent motion out of the plane of the storage ring. Electron detectors above this line, which is further restricted by the limit on the EDM. While this would be a very exciting result, it is unlikely that the EDM would be as large as that expected from even the upper limits on branching ratios respectively conversion probabilities of muon-number violating processes which involve muons and kaons. Also shown is the projected goal of the MEG (\(\mu \rightarrow e^+\gamma\)) experiment which is underway at PSI, and the projected sensitivity of the recent letter of intent to J-PARC for muon-electron conversion. (Figure from ref. 13.)

(\(g - 2\)) experiments, detectors placed in the plane of the beam would be used, in this case to make sure that the radial-E-field cancels the normal spin precession exactly. Adelmann and Kirsh\(^{52}\) have proposed that one could reach a sensitivity of \(5 \times 10^{-23} \text{ cm}\) with a small storage ring at PSI. A letter of intent at J-PARC\(^{53}\) suggested that one could reach \(<10^{-24} \text{ cm}\) there. The ultimate sensitivity would need an even more intense muon source, such as a neutrino factory.

5. The Search for Lepton Flavor Violation

The standard-model gauge bosons do not permit leptons to mix with each other, unlike the quark sector where mixing has been known for many years. Quark mixing was first proposed by Cabibbo,\(^{50}\) and extended to three generations by Kobayashi and Maskawa,\(^{55}\) which is described by a \(3 \times 3\) mixing matrix now universally called the CKM matrix. With the discovery of neutrino mass, we know that lepton flavor violation (LFV) certainly exists in the neutral lepton sector, with the determination of the mixing matrix for the three neutrino flavors having become a world-wide effort.

While the mixing observed in neutrinos does predict some level of charged lepton mixing, it is many orders of magnitude below present experimental limits.\(^{13}\) New dynamics,\(^{56-66}\) e.g., supersymmetry, do permit leptons to mix, and the observation of standard-model forbidden processes such as

\[
\mu^+ \rightarrow e^+\gamma; \quad \mu^+ \rightarrow e^+e^-e^-; \quad \mu^-N \rightarrow e^-N; \quad \mu^-e^- \rightarrow \mu^-e^-; \quad \mu^- + N' \rightarrow e^- + N'.
\]

would clearly signify the presence of new physics. The present limits on lepton flavor violation are shown in Fig. 7. If lepton mixing occurs via supersymmetry, there will be a mixing between the supersymmetric leptons (sleptons) which would also be described by a \(3 \times 3\) mixing matrix.

The schematic connection between lepton flavor violations and the dipole moments is shown in Fig. 8, and there are models that try to connect these processes.\(^{57}\)

In a large class of models, if the \(\Delta E = 1\) LFV decay goes through the transition magnetic moment, one finds...
Fig. 8. The supersymmetric contributions to the anomaly, and to $\mu \rightarrow e$ conversion, showing the relevant sleptonic matrix elements. The MDM and EDM give the real and imaginary parts of the matrix element respectively.

\[
\frac{B(\mu N \rightarrow eN)}{B(\mu \rightarrow e\gamma)} = 2 \times 10^{-3} B(A, Z), \tag{23}
\]

where $B(A, Z)$ is a coefficient of order 1 for nuclei heavier than aluminum.\(^6\) For other models, these two rates can be the same,\(^13\) so in the design of new experiments, the reach in single event sensitivity for the coherent muon conversion experiments needs to be several orders of magnitude smaller than for $\mu \rightarrow e\gamma$ to probe the former class of models with equal sensitivity. Detailed calculations of $\mu - e$ conversion rates as a function of atomic number have also been carried out,\(^69\) and if observed, measurements should be carried out in several nuclei.

From the experimental side, the next generation $\mu \rightarrow e\gamma$ experiment, MEG, is now under way at PSI,\(^70\) with a sensitivity goal of $10^{-13}$–$10^{-14}$. Since the decay occurs at rest, the photon and positron are back-to-back, and share equally the energy $m_\mu c^2$. This experiment makes use of a unique “COBRA” magnet which produces a constant bending radius for the mono-energetic $e^+$, independent of its angle. The photon is detected by a large liquid Xe scintillation detector as shown in Fig. 9.

Of the various lepton-flavor violating reactions, only coherent muon conversion does not require coincidence measurements. The decay $\mu \rightarrow 3e$, while theoretically appealing, requires a triple coincidence and sensitivity to the whole phase space of the decay, and thereby is experimentally more challenging. It is the coherent muon to electron conversion, where with adequate energy resolution, the conversion electron can be resolved from background, that with adequate muon flux can be pushed to the $10^{-18}$ or $10^{-19}$ sensitivity. Such a program has been proposed for J-PARC.\(^71\)

The muonium to antimuonium conversion [left-hand process in eq. (22)] represents a change of two units of lepton number, analogous to $K^0\bar{K}^0$ oscillations. This process was originally proposed by Pontecorvo.\(^22\) An experiment at PSI\(^73\) obtained a single event sensitivity of $P_{\text{amu}} \sim 8.2 \times 10^{-11}$ which implies a coupling $G_{\text{amu}} \lesssim 3 \times 10^{-5} G_F$ at 90\% C.L., where $G_F$ is the Fermi coupling constant. A broad range of speculative theories such as left–right symmetry, $R$-parity violating supersymmetry, etc.,\(^74\) could permit such an oscillation.

6. Summary and Conclusions

Since its discovery, the muon has provided an important tool to study the standard model, and to constrain its extensions. Experiments in the planning stage for $(g - 2)$, the search for an electric dipole moment and lepton flavor violation in muon decay or conversion will continue this tradition. Research and development for new more intense muon sources, such as the muon ionization cooling experiment (MICE),\(^75\) will further propel increases in sensitivity. Muon experiments form an important part of the precision frontier in particle physics, which will continue to provide vital information complementary to that from the highest energy colliders.

Acknowledgments

I wish to acknowledge A. Czarnecki, T. Gorringle, D. Hertzog, P. Kammel, K. Jungmann, Y. Kuno, W. Marciano, J. P. Miller, Y. Okada, and E. de Rafael, for helpful conversations, and TG, DH, KJ, PK, and JM for their excellent suggestions on this manuscript. This work was supported in part by the U.S. National Science Foundation and U.S. Department of Energy.

1) See Fig. 5 in P. Kunze: Z. Phys. 83 (1933) 1.
et al.

31) E. M. Purcell and N. F. Ramsey: Phys. Rev. 42

32) C. J. Baker

33) B. C. Regan

34) J. Bailey et al.


52) A. Adelmann and K. Kirsh: hep-ex/0606034.


Bradley Lee Roberts was born in Bristol, Virginia, U.S.A. in 1946. He obtained his B.S. in Physics from the University of Virginia in 1968, the M.S. (1970), and Ph. D. (1974) degrees from the College of William and Mary. He was a postdoctoral research associate at the Rutherford High-Energy Laboratory in England from 1974–1976, and at the Massachusetts Institute of Technology 1976–1977. Since the fall of 1977 he has been on the physics faculty at Boston University, holding the rank of Professor of Physics since 1989. He has worked in exotic (τ−, K−, ¯{ρ}, Σ−) atoms, muon physics, photonuclear physics and hyperon physics. Since 1989 he has served as co-spokesman for the Muon (g − 2) experiment, E821 at the Brookhaven Laboratory.
Frontiers of Elementary Particle Physics, the Standard Model and Beyond

Electric Dipole Moments of Elementary Particles, Nuclei, Atoms, and Molecules

Eugene D. Commings

Physics Department, University of California, Berkeley, CA 94720, U.S.A.

(Received December 13, 2006; accepted May 17, 2007; published November 12, 2007)

The significance of particle and nuclear electric dipole moments is explained in the broader context of elementary particle physics and the charge–parity (CP) violation problem. The present status and future prospects of various experimental searches for electric dipole moments are surveyed.

KEYWORDS: electric dipole moment, CP violation, parity and time reversal
DOI: 10.1143/JPSJ.76.111010

1. Introduction

Every physicist knows that the electron, proton, and neutron have spin magnetic dipole moments, as do many other particles and nuclei with non-zero spin. Can an elementary particle or nucleus also have a spin electric dipole moment (EDM)? In this review we try to explain why this question has become very important for elementary particle physics, and we briefly survey the present status and future prospects of various experimental searches for EDMs.

An EDM cannot exist unless both parity (P) and time reversal (T) invariance are violated. This can be seen from the non-relativistic Hamiltonian for the interaction of an EDM d with an electric field E, which is $H_{\text{EDM}}^{\text{NR}} = -d \cdot E$. For an elementary particle or nucleus in a non-degenerate state, the spin angular momentum $J$ is the only vector available to define a direction. Thus $d$ must be collinear with $J$, and:

$$H_{\text{EDM}}^{\text{NR}} = -d \cdot E = -d \cdot J \cdot E.$$  \hfill (1)

However $E$ is a T-even polar vector while $J$ is a T-odd axial vector. Thus $H_{\text{EDM}}^{\text{NR}}$ is odd under $P$ and $T$ transformations. The same conclusion is of course true for the relativistic generalization of $H_{\text{EDM}}^{\text{NR}}$.

Until now no EDM has been observed, and it is obvious from the present experimental upper limits, summarized in Table I, that EDMs must be extremely small. For example, the upper limit on the electron EDM $d_e$ is $8.3 \times 10^{-17} \mu_B$, where $\mu_B$ is the Bohr magneton. Nevertheless, EDMs may be non-zero, because $P$ and $T$ are in fact violated in nature. Parity nonconservation [as well as the violation of charge conjugation (C) invariance] occurs in the weak interaction. Furthermore, combined charge–parity (CP) violation is observed in neutral $K$ meson and $B$ meson decays. \(^{(1)}\) If we assume CPT invariance, for which we have very strong confidence, then this CP violation is equivalent to $T$ violation. Thus the weak interaction and the mechanism or mechanisms causing CP violation could act jointly to generate EDMs by $P,T$-odd radiative corrections to the $P$, $C$, $T$ conserving electromagnetic interaction.

Unfortunately, given the present state of our knowledge, such radiative corrections cannot be calculated with confidence. Instead, they depend on uncertain theoretical models of CP violation. According to the standard model, while CP violation in quantum chromodynamics (QCD) can in principle be large, CP violation in the electroweak sector is described phenomenologically by a single phase that appears in the Cabibbo–Kobayashi–Maskawa (CKM) quark mixing matrix.\(^{(2)}\) This gives a satisfactory account of $K$- and $B$-meson CP violation data.\(^{(3)}\) It can be shown that in this description the neutron EDM $d_n$ appears only at the three-loop level of perturbation theory\(^{(4)}\) while the electron EDM $d_e$ appears only at the four-loop level\(^{(5)}\) (and there are additional suppressions). Thus the standard model electroweak predictions: $d_n \approx 10^{-26} e\text{ cm}$, $d_e \leq 10^{-38} e\text{ cm}$ (where $e = 4.8 \times 10^{-10}$ esu is the unit of electronic charge) are many orders of magnitude smaller than the present experimental limits. Indeed, if the standard model mechanism of CP violation is the only one, then given present or foreseeable experimental capabilities, future observation of any EDM is very unlikely.

However, there are good reasons to think that additional mechanisms exist for CP violation. It is generally accepted that if the universe initially was symmetric in baryon–antibaryon number, the presently observed baryon–anti-baryon asymmetry could not have developed without a much larger CP violation than is predicted by the standard model.\(^{(6)}\) Furthermore, in many theories that attempt to go beyond the standard model, predicted EDMs are relatively large. For example, in various supersymmetric theories, new hypothetical particles and couplings appear, and along with them exist new CP violating phases. Thus in many such models the electron and neutron EDMs already appear at the

---

Table I. Experimental limits on EDMs and the $\tau$ weak dipole moment (WDM).

| Particle | EDM $|d| (e\text{ cm})$ | WDM $|d| (e\text{ cm})$ | Ref. |
|----------|----------------------|----------------------|----|
| $\nu_e$  | $2 \times 10^{-20}/f$|                      | 35 |
| $e^-$    | $1.6 \times 10^{-27}$|                      | 23 |
| $\mu^\pm$| $7 \times 10^{-10}$   |                      | 29 |
| $e^+$    | $1.1 \times 10^{-17}$| $5.8 \times 10^{-18}$| 34 |
| $p$      | $5.4 \times 10^{-24}$|                      | 19 |
| $n$      | $3 \times 10^{-26}$   |                      | 13 |
| $^9\text{Be}$ | $1.5 \times 10^{-16}$ |                      | 20 |
| $^{199}\text{Hg}$ atom  | $2 \times 10^{-28}$  |                      | 17 |

$e = 4.8 \times 10^{-10}$ esu; $F$ = form factor.

\(^{(1)}\) E-mail: commins@physics.berkeley.edu
one-loop level, and as a result predictions of $d_n$ and $d_a$ are close to present experimental limits.\textsuperscript{7,8} Thus discovery of an EDM by practical experimental methods is a real possibility within the foreseeable future, and such a discovery would provide definite evidence for physics beyond the standard model.

The search for EDMs of the neutron and of nuclei is important for a related issue of fundamental significance in QCD: the “strong CP problem”. A CP-odd term exists in the effective Lagrangian density for QCD, characterized by the “QCD CP-violating parameter” $\Theta$.\textsuperscript{3} It can be shown that this contributes to $d_n$ as: $d_n(\Theta) \approx 3 \times 10^{-10} \Theta$ e cm. Thus the present experimental limit on $d_n$ implies $\Theta \leq 1 \times 10^{-10}$. Why is $\Theta$ so small? A satisfactory answer to this question is not yet known.

2. Proper Lorentz-Invariant, Gauge-Invariant Lagrangian Density

In order to describe the interaction of the EDM of a spin-$1/2$ fermion with an electromagnetic field we need a gauge-invariant, proper-Lorentz-invariant effective Lagrangian density. First let us recall the analogous Lagrangian density in the Dirac theory.\textsuperscript{10} It is given by the well-known expression:

$$L_{\text{Pauli}} = -\kappa \frac{\mu_B}{2} \bar{\Psi} \gamma ^5 \gamma ^\mu F_{\mu \nu}.$$ (2)

Here $\Psi$ is the Dirac field for the fermion, $\bar{\Psi} = \Psi ^\dagger \gamma ^0$ is the Dirac conjugate field, $\gamma ^\mu$ are the usual $4 \times 4$ Dirac matrices,

$$F_{\mu \nu} = \partial _\mu A _\nu - \partial _\nu A _\mu = \begin{pmatrix} 0 & \varepsilon _x & \varepsilon _y & \varepsilon _z \\ -\varepsilon _x & 0 & -B_z & B_y \\ -\varepsilon _y & B_z & 0 & -B_x \\ -\varepsilon _z & -B_y & B_x & 0 \end{pmatrix}$$ (3)

is the electromagnetic field tensor, and $\kappa$ is a suitable constant. Rewriting eq. (2) in terms of $\mathcal{E}$ and $\mathcal{B}$ fields, we obtain:

$$L_{\text{Pauli}} = \kappa \mu_B \bar{\Psi} [\mathcal{E} \cdot \mathcal{B} - i \alpha \cdot \mathcal{E}] \Psi$$ (4)

where

$$\Sigma = \begin{pmatrix} \alpha & 0 \\ 0 & \alpha \end{pmatrix}, \quad \alpha = \begin{pmatrix} 0 & \sigma \\ \sigma & 0 \end{pmatrix}.$$ (5)

This Lagrangian density results in the Hamiltonian density

$$H_{\text{Pauli}} = -\kappa \mu_B \bar{\Psi} [\mathcal{E} \cdot \mathcal{B} - i \alpha \cdot \mathcal{E}] \Psi$$ (6)

and in the single-particle Hamiltonian:

$$H_{\text{Pauli}} = -\kappa \mu_B (\gamma ^0 \mathcal{E} \cdot \mathcal{B} - i y \cdot \mathcal{E}).$$ (7)

Of course, $L_{\text{Pauli}}$ of eq. (2) or (4) and $H_{\text{Pauli}}$ of eq. (6) are each $P$- and $T$-invariant. We can render them $P,T$-odd by replacing $\mathcal{E}$ by $-\mathcal{B}$ and $\mathcal{B}$ by $\mathcal{E}$, which is equivalent to the replacement of $F_{\mu \nu}$ by the tensor $F^{*}_{\mu \nu}$, where:

$$F^{*}_{\mu \nu} = \frac{1}{2} \varepsilon _{\mu \nu \rho \delta} F^{\rho \delta} = \begin{pmatrix} 0 & B_x & B_y & B_z \\ -B_x & 0 & -\varepsilon _z & \varepsilon _y \\ -B_y & \varepsilon _z & 0 & \varepsilon _x \\ -B_z & -\varepsilon _y & -\varepsilon _x & 0 \end{pmatrix}$$ (8)

and $\varepsilon _{\mu \nu \rho \delta}$ is the completely antisymmetric unit 4-tensor. Alternatively we obtain the same Lagrangian density by replacing $\sigma ^{\mu \nu}$ in eq. (2) with $i \sigma ^{\mu \nu} y ^5$ (where, as usual, $y ^5 = i y^{0} y^{1} y^{2} y^{3}$), with no change in $F_{\mu \nu}$. Making this latter transformation and replacing $\kappa \mu_B$ by $d$ we obtain the EDM Lagrangian density:

$$L_{\text{EDM}} = -\frac{d}{2} \bar{\Psi} \varepsilon ^{\mu \nu \gamma \delta} \gamma ^5 \gamma ^\mu F_{\mu \nu} = \bar{\Psi} [\mathcal{E} \cdot \mathcal{E} + i \alpha \cdot \mathcal{B}] \Psi$$ (9)

which was first described by Salpeter.\textsuperscript{11} This in turn yields the single-particle Hamiltonian:

$$H_{\text{EDM}} = -d (\gamma ^0 \mathcal{E} \cdot \mathcal{E} + i y \cdot \mathcal{B}).$$ (10)

In the non-relativistic limit the first term on the right-hand side (r.h.s.) of eq. (9) reduces to the r.h.s. of eq. (1). However when the particle of interest is relativistic, the full expression on the r.h.s. of eq. (9) must be used, and this has significant consequences, as we shall see.

3. The Neutron EDM

In all neutron EDM experiments use is made of the fact that non-relativistic polarized neutrons in collinear $\mathcal{E}$ and $\mathcal{B}$ fields undergo Larmor precession with frequency $\nu = [2 \mu_B^2 B^2 + 2d_B \mathcal{E}^2]/h$, where the $\pm$ sign corresponds to parallel (antiparallel) $\mathcal{E}$ and $\mathcal{B}$ fields. Thus the result of an EDM is revealed by an electric field-dependent shift in $\nu$ proportional to the $T$-odd pseudoscalar $\mathcal{E} \cdot \mathcal{B}$. The earliest experiments employed neutron beams and the Ramsey method of magnetic resonance with spatially separated oscillating fields and an intense electric field between them.\textsuperscript{12} More recent experiments utilize ultra-cold neutrons that typically have kinetic energies of $\approx 10^{-7}$ eV or less and undergo total internal reflection at any angle of incidence on suitable materials. These neutrons can be stored without substantial loss in closed vessels permeated by collinear $\mathcal{E}$ and $\mathcal{B}$ fields, where the oscillating fields for magnetic resonance are separated in time rather than in space. The most recent and precise of such experiments was performed at the Institut Laue–Langevin (ILL) in Grenoble, where the following result was obtained:\textsuperscript{13}

$$d_n = [+0.2 \pm 1.5 \text{(stat)} \pm 0.7 \text{(syst)}] \times 10^{-26} \text{ e cm.}$$ (11)

The statistical uncertainty on the r.h.s. of eq. (10) is mainly due to the limited number of ultra-cold neutrons that could be generated and stored, while the systematic uncertainty arises for the most part from a geometric phase effect caused by unintended gradients in $\mathcal{B}$.

New methods are needed if $d_n$ is to be determined to much better precision. Several novel projects are under development, including one at ILL and another at the Los Alamos National Laboratory (LANL). In the latter experiment,\textsuperscript{14} ultra-cold polarized neutrons are produced and stored in a bath of superfluid $^4$He containing a dilute solution of nuclear–spin-polarized $^3$He. A neutron can only exchange momentum with $^3$He and the free neutron dispersion curve and the phonon–roton dispersion curve of superfluid $^4$He intersect (see Fig. 1), since energy and momentum must both be conserved in the collision. Intersections occur at $E = k = 0$ and at $k = k^*$, $E = E^*$ (corresponding to the temperature $T^* = E^*/k_B \approx 12 \text{ K}$). Polarized neutrons enter the superfluid bath, and those with wavelength $\lambda = 2\pi/k^* \approx$...
0.89 nm are downscattered to form a polarized ultra-cold neutron sample. The probability of subsequent up-scattering by absorption of a \(^{4}\text{He}\) excitation is very small at the bath operating temperature \(T \leq 500\) mK, since the Boltzmann factor for these excitations is \(\exp(-T^4/T) \ll 1\). Thus a relatively large density of ultra cold polarized neutrons can be accumulated in the bath. The polarized \(^{3}\text{He}\) acts simultaneously as a neutron spin analyzer and as a co-magnetometer. The cross section for the reaction \(^{3}\text{He}(n,p)^{4}\text{He}\) + 764 keV is very large, but only in the n-\(^{3}\text{He}\) \(J = 0\) state (where the spins of these two species are opposed). Thus observation of this spin-dependent reaction by means of the resulting scintillations in \(^{3}\text{He}\) provides a way to detect the precession of the neutron spins in the \(E\) and \(B\) fields. A number of subtle problems are associated with this ambitious experiment. If they can be overcome, an improvement in the present limit on \(d_n\) of a factor of \(\approx 100\) seems possible.

The neutron and proton EDMs are expected to be roughly comparable in magnitude. However, as we discuss in the next section, the proton presents a completely different challenge to the experimenter because it is charged.

4. Atomic and Molecular EDMs

4.1 Schiff’s theorem

It has long been considered impractical to search for an EDM by placing a charged particle (e, p, bare nucleus, . . . ) in an electrostatic field, since the particle would quickly be accelerated out of the region of observation. (Recent proposals for storage ring searches for EDMs of charged particles are discussed in §5 and §6). What can we learn by applying an external electrostatic field to a neutral atom or molecule that contains a nucleus or unpaired electron with an EDM \(d\)? At first sight this approach appears useless, because in the limits where all atomic or molecular constituents are treated as point charges, and where non-relativistic quantum mechanics applies, the atom or molecule cannot possess an EDM \(d_n\) (cannot exhibit a linear Stark effect) to first order in \(d\). This is Schiff’s theorem\(^{15}\) which can be understood intuitively as follows: A neutral atom or molecule is not accelerated in a uniform external electric field. Thus the average force on each of the atomic or molecular constituents must be zero. In the non-relativistic, point charge limits, the only forces are electrostatic; hence the average electric field at each point charge must be zero. This happens because the external field is cancelled, on average, by the internal polarizing field.

We note in passing that Schiff’s theorem is not in conflict with existence of the so called “permanent” electric dipole moments of many polar molecules, familiar in chemistry and molecular spectroscopy. These moments do not violate \(P\) or \(T\), nor do they result in a linear Stark effect for sufficiently small applied electric fields, in the absence of degeneracy. They have entirely different observational signatures than exist for the EDMs of interest to us.

4.2 Nuclear EDMs

Schiff’s theorem is evaded for a nucleus if one takes into account magnetic hyperfine structure, and more importantly, for a nucleus of finite size if the nuclear EDM distribution is not the same as the charge distribution.\(^{19}\) In the latter case, a small residual EDM effect remains, which is expressed in terms of an additional \(P,T\)-odd electronic potential \(V_S = -e\xi S \cdot \mathbf{r}_i^3(r_i)\) that must be included in the atomic or molecular Hamiltonian. Here \(r_i\) refers to the position of the \(i\)-th electron relative to the nuclear center-of-mass, and the “Schiff moment” \(S\) is a vector proportional to the nuclear EDM depending on the difference between the normalized charge and EDM distributions:

\[
S = \sum_i I_i, I_z = I \frac{e}{10} \sum_{p=1}^{8} \left( r_p^2 - \frac{5}{3} \langle r^2 \rangle_{\text{cc}} \right) p |I, I_z = I \rangle. \tag{11}
\]

where the sum is over all nuclear protons, and \(I\) is the nuclear spin. \(S\) can be generated by an intrinsic EDM of an unpaired nucleon, and/or by \(P,T\)-odd nucleon–nucleon (NN) interactions. Generally speaking, \(S\) is largest for the heaviest nuclei, and in particular it is enhanced in octupole-deformed nuclei such as \(^{199}\text{Hg}\). In addition to the Schiff moment, a nucleus with nuclear spin \(I \geq 1\) can possess a magnetic quadrupole moment \(M\) originating from nucleonic EDMs and/or \(P,T\)-odd NN interactions. In a paramagnetic atom or molecule this would couple to the magnetic field resulting from the spin and spatial distribution of the unpaired electron. Because this interaction is magnetic, it would not be constrained by Schiff’s theorem.\(^{16}\)

The most sensitive nuclear EDM search to date\(^{17}\) was an optical pumping experiment carried out on the diamagnetic atom \(^{199}\text{Hg}\) (\(I = 1/2\)). The result is:

\[
d(199\text{Hg}) = -[1.06 \pm 0.49(\text{stat}) \pm 0.40(\text{syst})] \times 10^{-28} \ e \text{ cm} \tag{12}
\]

which gives the limit: \(|d(199\text{Hg})| \leq 2 \times 10^{-28} \ e \text{ cm}\). Calculations relating \(d(199\text{Hg})\) to \(S\) yield the result \(d(199\text{Hg}) = -2.8 \times 10^{-14}[S(e \text{ fm}^3)]\) e cm.\(^{18}\) From this and eq. (12) one obtains \(S(199\text{Hg}) = (3.8 \pm 1.8 \pm 1.4) \times 10^{-12} \ e \text{ fm}^3\). The largest contribution to \(S(199\text{Hg})\) is estimated to arise from a \(P,T\)-odd nucleon–nucleon interaction of the form \(\xi G_F(\bar{p}p)(\vec{\mu} \vec{r})/\sqrt{2}\), where \(G_F\) is Fermi’s coupling constant and \(\xi\) is a dimensionless constant. However calculations
show that there is also a contribution from the intrinsic EDM of the proton; indeed the best current limit on the proton EDM: $|d_p| \leq 5.4 \times 10^{-24} \text{e cm}$, is inferred from the experimental result (12). Very significant improvements in our knowledge of $d_p$ may come from future storage-ring experiments; see §6. We note in passing that a relatively large upper limit also exists for the $\Lambda^0$ hyperon; (20) see Table I.

4.3 The electron EDM

We next consider the unpaired atomic electron(s) in a paramagnetic atom or molecule. Sandars' analysis (21,22) has shown that Schiff’s theorem is also evaded here when relativistic effects are taken into account. The result of Sandars' analysis [which is based on the first term on the r.h.s. of eq. (9) including the factor $\gamma^0$] may be expressed in terms of the ratio $d_s/d_e$ or equivalently in terms of the effective electric field $E_{\text{eff}}$ experienced by $d_s$. It is convenient to write $E_{\text{eff}} = Q\Pi$ where $Q$ is a factor that includes relativistic effects as well as details of atomic (or molecular) structure, while $\Pi$ is the degree of polarization of the atom or molecule by the external electric field $E_{\text{ext}}$. For paramagnetic atoms with valence electrons in $s_{1/2}$ or $p_{1/2}$ orbitals, such as Cs and Tl in their ground states,

$$Q \approx 4 \times 10^{10} \text{V/cm} \times (Z/80)^3,$$  

(13)

where $Z$ is the atomic number. Also, for such atoms $\Pi \approx 10^{-3}(E_{\text{ext}}/100\text{KV/cm})$, which is only $\approx 10^{-3}$ for the maximum attainable laboratory fields $E_{\text{ext}} \approx 100\text{KV/cm}$. Since for paramagnetic atoms in all practical situations, $\Pi$ is proportional to $E_{\text{ext}}$, the ratio $E_{\text{eff}}/E_{\text{ext}}$ is a constant, and is usually called the enhancement factor $R = d_s/d_e$. For the ground states of alkali atoms and for thallium, one finds $|R| \approx 10Z^3\alpha^2$ where $\alpha$ is the fine structure constant. Although in these cases $\Pi \ll 1$, $|R|$ can greatly exceed unity for sufficiently large $Z$. For example, for the thallium atom ($Z = 81$), one calculates $R = -585$.

Equation (13) also applies for a wide range of heavy polar diatomic paramagnetic molecules with valence electrons in $\sigma$ or $\pi$ orbitals, such as YbF in the ground $^2\Sigma_{1/2}$ state, or PbO in the metastable $a(1)^3\Sigma_1$ state. (In these cases $Z$ is the atomic number of the heavy nucleus.) The main difference between atoms and molecules occurs in the factor $\Pi$. In a typical polar diatomic molecule, nearly complete polarization ($\Pi \approx 1$) can be achieved with relatively modest external fields: $(E_{\text{ext}} \approx 10^2-10^4 \text{V/cm})$ because of the very close spacing between adjacent spin-rotational levels of opposite parity. When $\Pi \approx 1$, $E_{\text{eff}}$ for a paramagnetic molecule such as YbF or PbO is approximately 3 orders of magnitude larger than the maximum attainable with atoms.

$P,T$-odd electron–nucleon (eN) interactions can also contribute to $d_s$ in diamagnetic or paramagnetic atoms or molecules. These, as well as the $P,T$-odd NN interactions, can appear in one or several non-derivative coupling forms: “scalar”, “tensor”, and “pseudoscalar”. ($P,T$-odd electron-electron interactions are also possible but these only yield an extremely small contribution.) Finally, $C,T$-odd (P-even) eN and NN interactions, and possible $T$-odd beta decay couplings could cause a $P,T$-odd atomic or molecular EDM through radiative corrections involving the usual weak interactions of the standard model. For a paramagnetic atom or molecule the most important contribution to $d_s$, in addition to $d_e$ itself, is the scalar $P,T$-odd eN interaction. (16)

The most sensitive search for $d_s$ to date employed $^{205}\text{Tl}$ in an atomic beam magnetic resonance experiment with separated oscillating fields. (23) The result is:

$$d_s = (6.9 \pm 7.4) \times 10^{-28} \text{e cm},$$  

(14)

assuming $R(^{205}\text{Tl}) = -585$ and no contribution from the scalar $P,T$-odd eN interaction. Equation (14) yields the limit: $|d_s| \leq 1.6 \times 10^{-27} \text{e cm}$. At present, many new searches for $d_s$ are in progress, using cesium, francium, YbF, PbO*, and the molecular ion HfF*•. (24) These experiments with free atoms or molecules employ various standard methods of atomic, molecular, and optical physics: laser and rf spectroscopy, optical pumping, atomic and molecular beams, ion trapping, atom trapping and cooling, etc. In another search for $d_s$, one applies a large electric field to the paramagnetic solid gadolinium gallium garnet (GGG). (25) In principle, the interaction of the EDMs of the unpaired electrons with the electric field at sufficiently low temperature can yield a net magnetization of GGG which can be detected by a superconducting quantum interference device (SQUID) magnetometer. (It has also been proposed separately that application of a sufficiently large electric field to a gaseous sample of diamagnetic diatomic molecules could generate an observable $P,T$-odd magnetization. (26))

In a complementary experiment, application of an external magnetic field to the ferrimagnetic solid gadolinium iron garnet (GdIG) can yield an EDM-induced electric polarization of the sample, which is detectable in principle by ultra-sensitive charge measurement techniques. (27)

The chances are good that at least one of the many new experimental searches for $d_s$ will improve the existing limit by at least a factor of 10 in the relatively near future. The experimental searches employing paramagnetic molecules (YbF, PbO*, ...) are of particular interest because these molecules have very large $E_{\text{eff}}$ values.

5. The Muon EDM

In most theoretical models, including the standard model, the electron, muon and tau lepton EDMs are proportional or approximately proportional to their masses. Assuming this and given the present limit on $d_{\mu}$, one predicts $d_{\mu} < 3.3 \times 10^{-25} \text{e cm}$. However in some theoretical models $d_{\mu}$ could be larger than this by an order of magnitude or more (28) This provides motivation for $d_{\mu}$ searches at the $10^{-24} \text{e cm}$ level. The best current limit on the muon EDM: $|d_{\mu}| \leq 7 \times 10^{-19} \text{e cm}$, was obtained in an experiment at the CERN muon storage ring in the 1970's, the primary purpose of which was a precise measurement of the muon $g$-factor anomaly $a(\mu) = (g - 2)/2$. (29) Since then, storage ring technology has advanced considerably, resulting in a much more precise measurement of $a(\mu)$ in recent years at Brookhaven. (30) This has led to new proposals, not only to improve the limit on $d_{\mu}$ but also for storage ring searches for the proton, deuteron, and $^3\text{He}$ EDMs. The deuteron EDM $d_\text{D}$ appears especially promising (see §6).

In order to understand the main features of these experiments, we consider a relativistic particle moving with velocity $\beta$ in a horizontal plane, in electric and magnetic fields $\mathbf{E}$ and $\mathbf{B}$, where $\mathbf{B}$ is in the vertical direction (hence
\( \mathbf{B} \cdot \mathbf{B} = 0 \) and \( \mathbf{B} \cdot \mathbf{E} = 0 \) also. It can be shown\(^{31}\) that the angular velocity \( \omega \) of spin precession with respect to the particle momentum is:

\[
\omega = -\frac{e}{m} \left[ aB + \left( \frac{1}{\beta^2 - \gamma^2} - a \right) \hat{\beta} \times \mathbf{E} + \frac{1}{2} \eta (\mathbf{E} + \hat{\beta} \times \mathbf{B}) \right]
\]

(15)

Here, \( \gamma = (1 - \beta^2)^{-1/2} \), \( \eta = 2d(2mc^2eh)^{-1} \) where \( d \) is the EDM, and in eq. (15) we employ units where \( \hbar = c = 1 \). In the CERN and Brookhaven muon \( g - 2 \) experiments the muon energy was chosen so that \( \beta^2 \gamma^2 = a^{-1} \), and also \( \mathbf{E} \) was negligible. In this situation eq. (15) reduces to:

\[
\omega = \omega_a + \omega_e.
\]

(16)

where

\[
\omega_a = -\frac{e}{m} aB \quad \text{and} \quad \omega_e = -\frac{e}{2m} \eta \beta \times \mathbf{B}.
\]

In this case the spin precesses about \( \omega \) with frequency \( \omega = \sqrt{\omega_a^2 + \omega_e^2} \approx \omega_a \) and has a small oscillatory vertical component (see Fig. 2). In the CERN experiment this was searched for by observing the angular distribution of electrons emitted above and below the horizontal orbit plane in polarized muon decay. However the precision was limited because the vertical spin component was so very small, as well as oscillatory, owing to the presence of \( \omega_e \). In a newly proposed muon EDM search\(^{32}\) one chooses \( (\beta^2 - \gamma^2) - a \) and an electric field is applied of magnitude \( |\mathbf{E}| = \beta \gamma a |\mathbf{B}| \) in the direction \( \beta \times \mathbf{B} \). In other words \( \mathbf{E} \) is radial and in the orbit plane; see Fig. 3. In this case \( \omega_e \) is eliminated and thus \( \omega \) is directed along \( \mathbf{E} \) with magnitude:

\[
\omega = \frac{e}{2m} \eta \beta (1 + \gamma^2 a) \mathbf{B}
\]

(17)

Consequently starting from a horizontal orientation the spin precesses very slowly in the vertical plane and the vertical spin component increases approximately linearly with time, becoming much larger than in earlier experiments. With this new scheme it may be possible to extend the limit on \( d_{\mu} \) to \( \approx 10^{-24} \text{ cm} \).

6. The Deuteron EDM

In a recently proposed storage ring deuteron EDM search,\(^{33}\) polarized deuterons with momentum \( p = 1.5 \text{ GeV/c} \) are to circulate in a specially designed ring with a magnetic field \( B = 2 \text{ T} \) normal to the orbit plane, and with no applied electric field. In the instantaneous rest frame of the particle, the magnetic and electric fields are:

\[
\mathbf{B}' = \gamma \mathbf{B}
\]

(18)

and

\[
\mathbf{E}' = \gamma (\beta \times \mathbf{B})
\]

(19)

For \( p = 1.5 \text{ GeV/c} \) and \( B = 2 \text{ T} \), the rest frame electric field is \( \mathbf{E}' \approx 5 \times 10^9 \text{ V/m} \). In contrast to nuclear EDM searches employing neutral atoms, this very large electric field is applied directly to the deuteron without any “Schiff” screening. As usual, the component of precession angular velocity due to the magnetic moment is directed along \( \mathbf{B} \); in the laboratory frame this is described by the formula \( \omega_a = -\gamma (e/m) a \mathbf{B} \). However, in this experiment a novel feature is introduced: the beam velocity is modulated at frequency \( \omega_b \), (with \( \Delta \beta / \beta \approx \pm 1\% \)). As is evident from eq. (19) this produces a component of \( \mathbf{E}' \) that oscillates in the plane perpendicular to \( \mathbf{E} \) at the same frequency as the magnetic spin precession. The interaction of \( d \) with this oscillating component of \( \mathbf{E}' \) causes a spin reorientation analogous to that which occurs in conventional magnetic resonance. The net result in the laboratory frame is the development of a vertical component of spin polarization proportional to \( d_0 \) that is approximately linear in the time. Given the parameters we have mentioned, the rate of vertical polarization accumulation is \( \approx 10^{-29} \text{ rad/s} \) for \( d_0 = 10^{-29} \text{ cm} \).

Detection of the deuteron spin polarization could be achieved as follows. A thin gas jet causes Coulomb scattering of a small fraction of beam deuterons on each turn around the ring. The scattered deuterons strike a thick carbon target in the shape of an annulus surrounding the beam. Elastic scattering of \( D \) on carbon is spin-dependent.

---

Fig. 2. Schematic diagram (not to scale) showing orientation of vectors in original storage ring search for \( d_\mu \).\(^{29}\) Angular velocity \( \omega_a \) is collinear with \( \mathbf{B} \), while \( \omega_e \) is normal to \( \beta \mathbf{B} \) plane. Unit vector \( \hat{s} \) in direction of muon spin precesses in plane perpendicular to \( \omega = \omega_a + \omega_e \), and thus has an oscillating vertical component.

Fig. 3. Schematic diagram (not to scale) of vectors in proposed storage ring search for \( d_\mu \).\(^{33}\) Electric field \( \mathbf{E} \) in direction of \( \beta \times \mathbf{B} \) causes cancellation of \( \omega_e \). Thus \( \hat{s} \) precesses in \( \beta \mathbf{B} \) plane with very small angular frequency \( \omega_c \).
due to the spin–orbit interaction, and at 1.5 GeV/c, the analyzing power is known to be better than 30%. Downstream from the carbon target is an array of scintillation detectors, also in the form of an annulus and segmented into four quadrants (left, right, up, down). The left–right asymmetry provides the EDM signal, while the up-down asymmetry gives information on the $g - 2$ precession. Assuming an initial deuteron polarization of 95%, $10^{12}$ deuterons in the ring, and a polarization coherence time of roughly 1000 s, as well as other parameters previously mentioned, it appears possible to achieve a statistical uncertainty in $d_D$ of $10^{-23}$ e cm in several years of running. While such a small uncertainty is very impressive, it is important to note that systematic errors in this experiment could be significant and difficult to control, and a thorough analysis of such systematics has not yet been completed.

To a good approximation the deuteron EDM may be expressed as follows:

$$d_D = d_u + d_p + d_{\text{Nuc}}^\text{Noc},$$  

(20)

where $d_{\text{Nuc}}^\text{Noc}$ is due to the $P,T$-odd nucleon–nucleon interaction. Since $d_u \approx -d_p$ is expected, one has $d_D \approx d_{\text{Nuc}}^\text{Noc}$. Because of the simplicity of the deuteron, $d_{\text{Nuc}}^\text{Noc}$ can be estimated far more precisely than the corresponding $P,T$-odd NN interactions in complex nuclei such as $^{199}$Hg.

It is expected that the $\Theta$ contributions to $d_D, d_u, d_p, d_{3u}$ should be in the ratios:

$$d_D(\Theta) : d_u(\Theta) : d_p(\Theta) : d_{3u}(\Theta) \approx -1 : 3 : -3 : 3.$$  

(21)

Finally, $d_u$ and $d_p$ can be expressed in terms of the electric dipole moments $d^e_8$ and chromo-electric dipole moments $d^c_8$ of up and down quarks. It can be shown that:

$$d_u \approx 1.4(d_8 - 0.25d_0) + 0.27e(d^e_8 - d^c_8) + 0.83e(d^2_8 + d^2_8)$$  

(22)

and:

$$d_p \approx (d_8 + d_0) + 6e(d^2_8 - d^c_8) - 0.2e(d^2_8 + d^c_8)$$  

(23)

We note that the second term on the r.h.s. of eq. (23) is more than 20 times larger than the corresponding term in eq. (22). These various expressions show that comparison of the EDMs of $n$, $p$, $D$, and $^3\text{He}$ could yield much valuable information that would almost certainly be unobtainable from observation of EDMs of heavy nuclei in conventional atomic or molecular experiments.

7. The $\tau$ EDM and Weak Dipole Moment

Tau leptons have often been produced in $e^+e^-$ collisions at colliding beam accelerators:

$$e^+ + e^- \rightarrow \tau^+ + \tau^-$$  

(24)

In lowest order of perturbation theory two distinct amplitudes contribute to this reaction: single photon exchange (electromagnetic interaction) and single $Z^0$ exchange (neutral weak interaction), $P,T$-odd radiative corrections to the photon exchange amplitude introduce the possibility of a tau EDM $d_\tau$, while similar corrections to the $Z^0$ exchange amplitude involve an analogous weak dipole moment (WDM) $d$. Although the EDM and WDM are independent quantities, in most theoretical models they have comparable magnitudes. When the center-of-mass energy for reaction

$$|d| \leq 5.8 \times 10^{-18} \text{e cm.}$$  

(25)

This result together with plausible theoretical assumptions leads to the following limit on the EDM $d_\tau$:

$$|d_\tau| \leq 1.1 \times 10^{-17} \text{e cm.}$$  

(26)

8. Can a Neutrino Possess an EDM?

It is not yet known whether a neutrino and an antineutrino of the same mass eigenstate are distinct particles ("Dirac" neutrino and antineutrino) or whether a neutrino of given mass is self-conjugate ("Majorana" neutrino). Neutrino magnetic and electric dipole moments are described by 3 $\times$ 3 matrices $\mu_{ij}$, $d_{ij}$ respectively, where the diagonal elements $\mu_{ii}$, $d_{ii}$ refer to the static dipole moments of the $i$'th mass eigenstate. If neutrinos are of the Majorana type, the diagonal elements $\mu_{ii}$, $d_{ii}$ must be zero because under charge conjugation, the magnetic dipole and electric dipole operators change sign. Of course, no such restriction applies to Dirac neutrinos.

A neutrino EDM could cause anomalous ionization in a detector because of its interaction with atomic electrons. Making use of this, analysis(35) of an experiment carried out by Cowan and Reines in 1957 to detect $\bar{\nu}_e$ radiated from a reactor yielded the result $|d_{\nu,F}| \leq 2 \times 10^{-20} \text{e cm}$, where $F$ is a form factor.

9. Conclusion

After a half century of search, there is still no experimental evidence for an EDM of an elementary particle, nucleus, atom, or molecule. However, widespread appreciation of the significance of EDM searches for the general problem of CP violation and the development of new and refined experimental techniques now generate more intense interest in EDM searches than ever before. The present experimental upper limits for $d_u$, $d_p$, and $d_{\text{Nuc}}^{199}$Hg already provide serious constraints on various supersymmetric models of CP violation. Improvements of factors of 10–100 in the limits on $d_u$ and $d_p$, which may come in the relatively near future, should thus be very significant. Finally, success of the proposed deuteron storage ring experiment would bring the field of nuclear EDMs into an entirely new era.

Eugene D. Commins was born in New York, N.Y., U.S.A. in 1932. He obtained his B.A. degree in 1953, and his Ph. D. from Columbia University in 1958. He was an Instructor in Physics at Columbia from 1958 to 1960, and since then has been a member of the Physics Faculty at the University of California, Berkeley. His experimental research has mainly focused on space inversion symmetry or parity (P) and time reversal invariance (T), investigated with the tools of low-energy atomic and nuclear physics.
Higgs Particle: The Origin of Mass

Yasuhiro OKADA¹,²*

¹Theory Group, KEK, 1-1 Oho, Tsukuba, Ibaraki 305-0801
²The Graduate University for Advanced Studies (Sokendai), 1-1 Oho, Tsukuba, Ibaraki 305-0801

(Received June 18, 2007; accepted August 1, 2007; published November 12, 2007)

The Higgs particle is a new elementary particle predicted in the Standard Model of the elementary particle physics. It plays a special role in the theory of mass generation of quarks, leptons, and gauge bosons. In this article, theoretical issues on the Higgs mechanism are first discussed, and then experimental prospects on the Higgs particle study at the future collider experiments, LHC and ILC, are reviewed. The Higgs coupling determination is an essential step to establish the mass generation mechanism, which could lead to a deeper understanding of particle physics.

KEYWORDS: particle physics, Higgs mechanism, mass generation

DOI: 10.1143/JPSJ.76.111011

1. Introduction

Current understanding of the elementary particle physics is based on two important concepts, gauge invariance and spontaneous symmetry breaking. Out of four fundamental interactions of Nature, namely strong, weak and electromagnetic and gravity interactions, three of them except for gravity are described on the same footing in terms of gauge theory. The gauge group corresponding to the strong interaction is SU(3), and the weak and the electromagnetic interactions arise from SU(2) and U(1) groups and are called the electroweak interaction. Once quarks and leptons are assigned in proper representations of the three gauge groups, all properties of the three fundamental interactions are determined from the requirement of gauge invariance.

For more than thirty years, high energy experiments have been testing various aspects of gauge symmetry and have established gauge invariance as a fundamental principle of Nature. We have discovered gauge bosons mediating the three interactions, namely gluon for the strong interaction, W and Z bosons for the electroweak interaction. The couplings between quarks/leptons and gauge bosons have been precisely measured at the CERN LEP and SLAC SLC experiments, and we have confirmed the assignments of the gauge representations for quarks and leptons.

The gauge principle alone, however, cannot describe the known structure of the elementary particle physics. In the Standard Model of the elementary particle physics, all quarks, leptons and gauge bosons are first introduced as massless fields. In order to generate masses for these particles, the SU(2) × U(1) symmetries have to be broken spontaneously.

Spontaneous symmetry breaking itself is not new for particle physics.¹² The theory of the strong interaction, QCD, possesses an approximate symmetry among three light quarks called chiral symmetry. The vacuum of QCD corresponds to a state where quark and anti-quark pair is condensed, and the chiral symmetry is broken spontaneously. As a consequence, pseudo scalar mesons such as pions and kaons are light compared to the typical energy scale of the strong interaction since they behave approximately as Nambu–Goldstone bosons, a characteristic signature of spontaneous symmetry breaking.

In the case of the electroweak symmetry, it is shown theoretically that the Nambu–Goldstone bosons associated with spontaneous breakdown are absorbed by gauge bosons, providing the mass generation mechanism for gauge bosons (Higgs mechanism).³ Although we are now quite sure that this is the mechanism for gauge boson mass generation, we know little about how the symmetry breaking occurs, or what is dynamics behind the Higgs mechanism. Clearly, we need a new interaction other than four known fundamental forces, but we do not know what it is. The goal of the Higgs physics is to answer this question.

In this article, I would like to explain what are theoretical issues of the Higgs sector, what is expected at the future collider experiments, LHC and ILC, and what would be impacts of the Higgs physics on a deeper understanding of the particle physics.

2. Higgs Boson in the Standard Model

In the Standard Model, a single Higgs doublet field is included for the symmetry breaking of the SU(2) × U(1) gauge groups. This was introduced in S. Weinberg’s 1967 paper “A Model of Leptons”,⁴ and is the simplest possibility for generating the gauge boson masses.

The Higgs potential is given by

\[ V(\Phi) = -\mu^2|\Phi|^2 + \lambda|\Phi|^4, \] (1)

where the two component complex field is defined as

\[ \Phi(x) = \begin{pmatrix} \phi(x)^+ \\ \phi(x)^0 \end{pmatrix}. \] (2)

In order for the stability of the vacuum the parameter \( \lambda \) must be positive. The coefficient of the quadratic term, on the other hand, can be either sign. In fact, if the sign is negative, namely \( \mu^2 > 0 \), the origin of the potential is unstable, and the vacuum state corresponds to a non-zero value of the \( \Phi \) field. The states satisfying \( |\phi|^2 + |\phi|^2 = \mu^2/(2\lambda) \equiv v^2/2 \) are degenerate minimum of the potential. We can choose the vacuum expectation value in the \( \phi^0 \) direction, \( (\phi^0) = v/\sqrt{2}, \) and then there are three massless modes corresponding to the

¹E-mail: yasuhiro.okada@kek.jp
flat directions of the potential (Nambu–Goldstone modes). When the symmetry is a gauge symmetry, these massless particles disappear from the physical spectrum, and become longitudinal components of massive gauge bosons. This is seen most clearly if we take the “unitary gauge” where the Nambu–Goldstone modes are removed by an appropriate gauge transformation. The kinetic term of the scalar field is defined as

$$|D_\mu \Phi|^2 = \left| \left( \partial_\mu + \frac{g}{2} W^a_\mu + \frac{g'}{2} B_\mu \right) \Phi \right|^2,$$

and the gauge boson mass terms are obtained by substituting the vacuum expectation value into $\Phi(x)$.

Mass terms of quarks and leptons are also generated through interactions with the Higgs field. This follows from the chiral structure of quarks and leptons. Since the discovery of the parity violation in the weak interaction,\(^{25}\) chiral projected fermions (Weyl fermions) instead of Dirac fermions have been considered as building blocks of a particle physics model. In particular, only left-handed quarks and leptons are assigned as $SU(2)$ doublets because the weak interaction has a $V$ (vector)–$A$ (axial vector) current structure. Right-handed counterparts are singlet under the $SU(2)$ gauge group. This unbalance in the $SU(2)$ quantum number assignment forbids us to write direct mass terms for quarks and leptons: the only possible way to generate mass terms is to introduce Yukawa couplings with help of the $\Phi$ field such as $y_q \Phi \bar{d}_0 q_l$ where $q_l = (u_l, d_l)^T$. After replacing $\Phi(x)$ by its vacuum expectation value, this term generates a mass of $y_q v/\sqrt{2}$ for down-type quarks. Similar mechanism works for up-type quarks and charged leptons.

There is one important prediction of this model. Since we introduce a two-component complex field and three real degrees of freedom are absorbed by gauge bosons, one scalar particle appears in the physical spectrum, which is called the Higgs particle (= Higgs boson). In the Unitary gauge, interactions related to the Higgs boson can be obtained by replacing $v$ with $v + H(x)$ in the Lagrangian where $H(x)$ represents the Higgs boson. The mass of the Higgs boson is given by $m_H = \sqrt{2} \lambda v$, which means that the Higgs boson becomes heavier if the Higgs self-coupling gets larger. In fact, this is a general property of the particle mass generation mechanism due to the Higgs field: A stronger interaction leads to a heavier particle. The mass formula for the $W$, $Z$ bosons, quarks, leptons, and the Higgs boson at the lowest order approximation with respect to coupling perturbation (i.e., tree-revel) are summarized in Table I.

3. Naturalness and Physics beyond the Standard Model

Although the Higgs potential in eq. (1) is very simple and sufficient to describe a realistic model of mass generation, we think that this is not the final form of the theory but rather an effective description of a more fundamental theory. It is therefore important to know what is limitation of this description of the Higgs sector.

In renormalizable quantum field theories, the form of Lagrangian is specified by requirement for renormalizability. In the case of the Higgs potential, quadratic and quartic terms are only renormalizable interactions. We can then consider two kinds of corrections to the potential. One is a calculable higher order correction within the Standard Model. For instance the correction from the top Yukawa coupling constant can be evaluated up to a desired accuracy applying renormalization procedure of field theory. Another type of corrections comes from outside of the present model, presumably from physics at some high energy scale. We cannot really compute these corrections until we know the more fundamental theory. In this sense, the present theory is considered as an effective theory below some cutoff energy scale $\Lambda$.

Although the effective theory cannot include all physical effects, it is still useful because unknown correction is expected to be suppressed by $(E/\Lambda)^2$ where $E$ is a typical energy scale under consideration. Therefore, as long as the cutoff scale is somewhat larger than $E$, the theory can make fairly accurate predictions. For example, the correction is $10^{-4}$ when the cutoff scale is around $10$ TeV for physical processes in the $100$ GeV range. If the theory is valid up to the Planck scale ($\sim 10^{19}$ GeV) where the gravity interaction becomes as strong as the other gauge interactions, the correction becomes extremely small. In this way, an effective theory is useful description as long as we restrict ourselves to the energy regime below the cutoff scale.

Once we take a point of view that the Higgs sector of the Standard Model is an effective description of a more complete theory below $\Lambda$, naturalness with regard to parameter fine-tuning becomes a serious problem. In particular, the quadratic divergence of the Higgs mass radiative correction is problematic, and this has been one of main motivations to introduce various models beyond the Standard Model.

In the Higgs potential in eq. (1) the only mass parameter is $\mu^2$. At the tree level, this parameter is related to the vacuum expectation value $v$ by $\mu^2 = \lambda v^2$ where $v$ is known to be about $246$ GeV. $(v = (\sqrt{2}G_F)^{-1/2}$, where $G_F$ is the Fermi constant representing the coupling constant of the weak interaction.) If we include the radiative correction, $\mu^2$ becomes a sum of two contributions $\mu_0^2 + \delta \mu^2$ where $\mu_0^2$ is a bare mass term and $\delta \mu^2$ is the radiative correction. In the Standard Model, the top quark and gauge boson loop corrections are important and $\delta \mu^2$ from these sources are represented by a sum of terms of a form $C_i [g_i^2/(4\pi)^2] \Lambda^2$ where $g_i$ is the top Yukawa coupling constant, or $U(1)$ or $SU(2)$ gauge coupling constant and $C_i$ are $O(1)$ coefficients.

Since the radiative correction depends on the cutoff scale quadratically, the fine-tuning between the bare mass term and the radiative correction is necessary if the cutoff scale is much larger than $1$ TeV. Roughly speaking, the fine-tuning at 1% level is necessary for $\Lambda = 10$ TeV. If the cutoff scale is close to the Planck scale, the degree of the fine-tuning is enormous: A tuning of one out of $10^{32}$ is required. This is the naturalness problem of the Standard Model, and sometime also called the hierarchy problem. This problem suggests that the description of the Higgs sector by the simple

Table I. Mass formula for elementary particles, $g$, $g'$, $y$, and $A$ are the $SU(2)$ and the $U(1)$ gauge coupling constants, the Yukawa coupling constant for a fermion $f$, and the Higgs self-coupling constant.

<table>
<thead>
<tr>
<th>W boson</th>
<th>Z boson</th>
<th>Quarks, leptons</th>
<th>Higgs boson</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\frac{g}{\sqrt{2}}$</td>
<td>$\sqrt{g^2 + g'^2}$</td>
<td>$y \frac{v}{\sqrt{2}}$</td>
<td>$\sqrt{2} \lambda v$</td>
</tr>
</tbody>
</table>
potential in eq. (1) is not very satisfactory, and probably will be replaced by a more fundamental form at a higher energy scale.

Since the problem arises from the quadratic divergence in the renormalization of the Higgs mass terms, proposed solutions involve cancellation mechanism of such divergence. Supersymmetry is a unique symmetry that guarantees complete cancellation of the quadratic divergence in scalar field mass terms. This is a new symmetry between bosons and fermions and the cancellation occurs between loop diagrams of bosons and fermions. Particle physics models based on supersymmetry such as supersymmetric grand unified theory (SUSY GUT) have been proposed and studied since early 1980’s as a possible way out of the hierarchy problem. In a realistic model of a supersymmetric extension of the Standard Model, we need to introduce new particles connected by supersymmetry to ordinary quarks, leptons, gauge bosons, and Higgs particles. In 1990’s, precision studies on the Z boson were performed at LEP and SLC experiments, and it was pointed out that three precisely measured coupling constants are consistent with the prediction of SUSY GUT, although the gauge coupling unification fails badly without supersymmetric partner particles. Supersymmetry is also an essential ingredient of the superstring theory, a potential unified theory including gravity and gauge interactions. In this way, the supersymmetric model has become a promising candidate beyond the Standard Model. If supersymmetry realizes at or just above the TeV scale, it can provide a consistent and unified picture of the particle physics from the weak scale to the Planck scale.

A opposite idea for solution of the naturalness problem is considering that the cutoff scale is close to the electroweak scale. In particular, the Higgs field is considered to be a composite state of more fundamental objects at a relatively low energy scale. The simplest form of this model is called the technicolor model proposed in late 1970’s, in which the cutoff scale is about 1 TeV. The technicolor model is however strongly constrained from precision tests of electroweak theory later at the LEP and SLC experiments, but there have been continuous attempts to construct a phenomenologically viable model of a composite Higgs field. Little Higgs models are a recent proposal on this line, where the physical Higgs boson is dynamically formed by a new strong interaction around 10 TeV. An interesting feature of this model is that the quadratic divergence of the Higgs boson mass term is canceled by loop corrections due to new gauge bosons and a heavy partner of the top quark at one loop level. In this way the hierarchy problem between the electroweak scale and 10 TeV is nicely solved.

In addition to supersymmetry and little Higgs models, there have been many proposals for TeV scale physics. Motivations for many of them are solving the naturalness problem of the Standard Model or explaining the large (apparent) hierarchy between the weak scale and the gravity scale. Examples are models with large extra dimensions, models with warped extra-dimensions, the Higgsless model, the twin Higgs model, and the inert Higgs model, the split-supersymmetry model, etc. All of these proposals involve some characteristic signals around a TeV region. These signals are important to choose a correct model at the TeV scale and clarify the mechanism of the electroweak symmetry breaking.

4. Experimental Prospects of Higgs Physics

Higgs physics is expected to be the center of the particle physics in coming years starting from the commissioning of the CERN LHC experiment. The first step will be a discovery of a new particle which is a candidate of the Higgs boson. We then study its properties in detail and compare them with the prediction of the Standard Model Higgs boson. We may be able to confirm that the discovered particle is the Higgs boson responsible for the mass generation for elementary particles. Another possibility would be to find some deviation from the Standard Model Higgs boson. Deviation could be something like small difference of production cross section and decay branching ratios from the Standard Model predictions, or more drastic new signals such as discovery of several Higgs states. At the same time, we may also find other new particles predicted in extensions of the Standard Model, for example supersymmetric particles in the supersymmetric model or the heavy gauge bosons and the top partner in the little Higgs model. In order to accomplish these goals we probably need several steps in collider experiments including LHC and ILC experiments and possible upgrades for these facilities.

If we restrict ourselves to the Higgs boson in the Standard Model, all physical properties are determined by one parameter, the Higgs boson mass. Present experimental lower bound for the mass of the Standard Model Higgs boson is 114.4 GeV at the 95% confidence level, set by the direct Higgs boson search at LEP. It is remarkable that we can also draw an upper bound from a global fit of electroweak precision data. Although a heavy Higgs boson means a large self-coupling $\lambda$, we have not seen any evidence of such a large coupling in physical observables related to Z and W gauge boson processes. The upper limit of the Standard Model Higgs boson is 166 GeV at the 95% confidence level. This implies that a relatively light Higgs boson is favored. If the Higgs boson turns out to be heavier than 200 GeV, we would expect some additional new particles that have significant couplings to gauge bosons.

The decay branching ratios of the Higgs boson depends strongly on the Higgs boson mass, and therefore the discovery strategy for the Higgs boson at LHC differs for light and heavy Higgs bosons. The branching ratios for the Standard Model Higgs boson is shown in Fig. 1. Since the Higgs boson couples more strongly to a heavier particle, it tends to decay to heavier particles as long as kinematically allowed. For instance, the Higgs boson mostly decays into two gauge bosons if the Higgs boson mass is larger than 200 GeV, whereas the bottom and anti-bottom pair is the main decay mode for its mass less than 140 GeV. For this mass range, the Higgs boson search at LHC relies on other decay modes such as the loop-induced two photon decay mode, because two bottom modes are hidden by overwhelming QCD background processes. Detail simulation studies on the Higgs discovery at LHC have been performed, and it is shown that the Higgs boson can be found at LHC experiments within a few years for the entire mass region as long as the production and decay properties are similar to the
Standard Model Higgs boson.\textsuperscript{20,21} Furthermore, information on the Higgs couplings is obtained with a higher luminosity. Estimated precision for coupling ratios are typically \(0(10)\%\).\textsuperscript{22}

ILC is a future electron–positron linear collider project proposed in the international framework.\textsuperscript{23} One aspect of this facility is a Higgs factory. For instance, the number of produced Higgs bosons can be \(0(10^5)\) in the first stage of experiments with the center-of-mass collider energy of 500 GeV. Under clean environment of the \(e^+e^-\) collider, precise determinations on the mass, quantum numbers, and coupling constants of the Higgs boson are possible. Typical production and decay processes are shown Fig. 2. Precision of the coupling constant determination reaches a few \% level for Higgs-\(WW\), Higgs-\(ZZ\), and Higgs-\(bb\) couplings for the case of a relatively light Higgs boson. We can also measure the Higgs self-coupling from the double Higgs boson production and the top Yukawa coupling from the Higgs-\(tt\) production. Figure 3 shows precision of the Higgs coupling constant determination for various particles at ILC. The proportionality between coupling constants and particle masses is a characteristic feature of the one Higgs doublet model where the particle mass formulas involve only one vacuum expectation value. An important feature of ILC experiments is that absolute values of these coupling constants can be determined in a model-independent way. This is crucial in establishing the mass generation mechanism for elementary particles.

The precise determination of the Higgs coupling constants is also useful to explore physics beyond the Standard Model.

In some case, the Higgs boson coupling is modified from the Standard Model.

- The Higgs sector of supersymmetric models is different from the Standard Model. In any realistic supersymmetric model, the Higgs sector contains at least two sets of doublet fields. In the minimal supersymmetric standard model (MSSM), in particular, the Higgs sector is a two Higgs doublet model. Furthermore, there is a rather strict theoretical upper bound for the the lightest neutral Higgs boson,\textsuperscript{25} which is about 130 GeV. Since this light boson plays a role of the usual Higgs particle, this particle may be the only Higgs particle discovered at LHC. In such case, the branching ratio measurement for the lightest neutral Higgs boson is useful to obtain information on the masses of heavy Higgs bosons.\textsuperscript{26–28} In particular, the tau and bottom coupling constants show sizable enhancement if the heavy Higgs boson exists below 600 GeV. The ratio like \(B(H \rightarrow WW)/B(H \rightarrow \tau\tau)\) is useful to determine the heavy Higgs mass scale indirectly.

- In models with extra dimensions, there appears a scalar field called Radion, corresponding to the size of the extra space dimension. Since Radion is a neutral scalar field, it can mix with the Higgs field. It is pointed out that Radion-Higgs mixing in the warped extra dimension model could reduce the magnitude of Yukawa coupling constants and \(WWW\) and \(ZZZ\) constant in a universal way.\textsuperscript{29} In order to observe such effects, absolute coupling measurements at ILC are necessary.

- The two-gluon width of the Higgs boson is generated by loop diagrams, so that it can be a probe to virtual effects of new particles. The same is true for the two-photon width, the measurement of which is improved at the photon–photon collider option of ILC.\textsuperscript{30} There are many new physics models where such loop effects are sizable.
• Explaining the baryon number of the Universe is one of the most outstanding questions for particle physics in connection with cosmology. One possibility is the electroweak baryogenesis scenario, in which the baryon number was generated at the electroweak phase transition. For a successful electroweak baryogenesis, the Higgs sector has to be extended from that of the minimal Standard Model to realize a strong first-order phase transition. The change of the Higgs potential can lead to observable effects in the triple Higgs coupling measurement.

As we can see above examples, observations of new physics effects require precise determination of coupling constants. This will be an important goal of the future ILC experiment.

5. Conclusions

The Higgs sector is an unknown part of the particle physics model. Although a simple potential is assumed in the Standard Model, this description is supposed to be valid below a cutoff scale, beyond which the theory of the electroweak symmetry breaking takes a more fundamental form. If the cutoff scale is as low as 1 TeV, some direct signals on new physics are likely to appear at LHC. If the cutoff scale is much larger, the fine-tuning of the Higgs boson mass becomes a serious problem. Proposed solutions to this problem such as supersymmetry or little Higgs models also predict new physics signals at the TeV scale. These signals are targets of future collider experiments starting from LHC.

Experimental prospects for the Higgs physics are quite bright. The Higgs particle can be found and studied at LHC. At the proposed ILC, precise information on coupling constants between the Higgs boson and other particles will be obtained. These measurements are an essential step to establish the mass generation mechanism. At the same time, the precision measurement may reveal evidence of new force and/or new symmetry because these new physics is most probably related to the physics of electroweak symmetry breaking, i.e., the Higgs sector. In this way, the Higgs particle will play a special role in determining the future direction of the particle physics.

Acknowledgment

This work was supported in part by Grants-in-Aid for Science Research, Ministry of Education, Culture, Sports, Science and Technology Nos. 16081211 and 17540286.


Yasuhiro Okada was born in Fukuoka Prefecture, Japan in 1957. He obtained his B. Sc. (1980), M. Sc. (1982), and D. Sc. (1985) degrees from the University of Tokyo. After holding postdoctoral positions at KEK, the University of Michigan, and the University of North Carolina at Chapel Hill, he became an assistant professor in Tohoku University in 1989. He moved to KEK as an associate professor in 1992, and has been a professor since 2000. Currently, he is also a professor in the Graduate University for Advanced Studies (Sokendai). He has worked on theoretical elementary particle physics, especially in the areas of Higgs physics, flavor physics, and phenomenology of physics beyond the Standard Model.
Prospects of Physics Beyond the Standard Model: Supersymmetry

Masahiro YAMAGUCHI

Department of Physics, Tohoku University, Sendai 980-8578

(Received July 9, 2007; accepted July 24, 2007; published November 12, 2007)

Despite the success of the standard model of particle physics, there are reasons to expect new physics beyond it. Supersymmetric extension is a promising possibility, which solves the fine-tuning problem inherent to the origin of the electroweak scale in the standard model. It also provides a promising candidate for dark matter. The verification of low-energy supersymmetry in forthcoming experiments will lead us to a new paradigm of particle physics, in particular, it may open up the road to the unification of forces.

KEYWORDS: particle physics, standard model, supersymmetry

DOI: 10.1143/JPSJ.76.111012

1. Introduction

Although the standard model (SM) of particle physics successfully describes particle interactions at an energy scale below 100 GeV, there are some reasons to anticipate the existence of new physics beyond it. First, there is the fine-tuning problem (or naturalness problem) associated with the Higgs field. Second, the SM does not provide any candidate for the dark matter of the universe. Third, from the viewpoint of attempting to unify all forces and matter, the SM appears to be very far from its goal; the gauge structure is an elementary scalar field. The radiative correction to the square of the mass of such a scalar field is known to suffer a consequence of the spontaneous symmetry breakdown of this gauge symmetry, i.e., $SU(3)_c \times SU(2)_L \times U(1)_Y$ down to $SU(3)_C \times U(1)_Y$. In the SM, this is caused by the postulated Higgs scalar field.

The scalar potential of the Higgs field, $H$, is given as

$$V(H) = m_H^2 H^2 + \lambda (H^*)^2,$$

where $\lambda$ is the self-coupling of the Higgs field and $m_H^2$ is a mass parameter. The Higgs field develops a non vanishing vacuum expectation value (VEV) and the symmetry is spontaneously broken when $m_H^2 < 0$. It is easy to see that the VEV of the Higgs field is $\langle H \rangle = \sqrt{-m_H^2 / 2\lambda}$ and its mass is $\sqrt{-m_H^2}$. Thus, to obtain the correct weak scale of the order of 100 GeV, this mass $\sqrt{|m_H^2|}$ should also have the same order of magnitude.

The couplings of the $Z$ boson and those of the $W$ boson to quarks and leptons have been precisely measured in the LEP and Tevatron experiments, and are in perfect agreement with the SM prediction. Furthermore, the SM with a Higgs mass of less than roughly 200 GeV is consistent with the electroweak precision data.

Despite the success of the SM, it suffers from a fine-tuning problem. This is attributed to the fact that the SM Higgs field is an elementary scalar field. The radiative correction to the square of the mass of such a scalar field is known to suffer from quadratic divergence, and thus it is proportional to the square of the UV cut off $\Lambda$. Schematically, it is written

$$m_H^2 = m_{H,\text{phys}}^2 + \frac{c}{16\pi^2} \Lambda^2,$$

where $m_{H,\text{phys}}$ is the physical mass of order 100 GeV, while $m_{H,\text{bare}}$ is the mass parameter given in the bare Lagrangian. The second term of the right-hand side is the divergent quantum correction with $c$ being a number related to coupling constants. It is plausible that the UV cut off $\Lambda$ is around the Planck scale of $10^{18}$ GeV, the energy scale at which the quantum effects of the gravitational interactions become important, and presumably the description based on quantum field theory is no longer valid. In this case, a huge radiative correction proportional to a power of $\Lambda$ has to be canceled by the bare mass parameter with a precision of $10^{-18}$ to obtain the correct electroweak scale of 100 GeV. It is argued that such a fine tuning is unnatural, and there should be some way of solving this difficulty.

$^*$E-mail: masahiro.yamague@gmail.com
3. Supersymmetry

Supersymmetry is a symmetry that relates bosons with fermions. For the supersymmetry charges $Q_\alpha (\alpha = 1, 2)$ and $\tilde{Q}_\beta (\beta = 1, 2)$, in terms of two-component spinor notation, the supersymmetry algebra is represented as

$$[Q_\alpha, \tilde{Q}_\beta] = 2\sigma^\mu_{\alpha\beta} P_\mu.$$  \hspace{1cm} (3)

Here $P_\mu (\mu = 0, 1, 2, 3)$ on the right-hand side generates the four-dimensional space–time translation, and $\sigma^\mu$ is the Pauli matrix generalized to four dimensions, $\sigma^\mu = (-1, \sigma)$. Intuitively, the supersymmetry transformation is a square root of the translation; by applying the supersymmetry transformation twice, one obtains the space–time translation. A closer look at the Lorentz structure suggests that the supersymmetry charge $Q_\alpha$ should behave as a spinor, and thus should obey the Fermi–Dirac statistics. This is the reason why the above is written in the form of an anti-commutation relation, instead of a commutation relation. When $Q_\alpha$ is applied, bosons are transformed into fermions and vice versa.

Supersymmetry was discovered in the two-dimensional world sheet formulation of fermionic string theory. 1) The four-dimensional Lagrangian formalism was invented by Wess and Zumino. 2) One of the novel properties of supersymmetry is that it is the maximum symmetry that contains the space–time Poincaré symmetry (i.e., the symmetry generated by translations and Lorentz transformations), and is consistent with the desired properties of the S-matrix.

It is instructive to illustrate the simplest four-dimensional supersymmetric model given by Wess and Zumino. It has a complex scalar field $\phi$, a two component Weyl spinor $\psi$, and a complex auxiliary field $F$. Here we introduce the auxiliary field $F$ simply to make the supersymmetry transparent. As we will see, it does not have a physical degree of freedom, but can be eliminated, using the equation of motion. The supersymmetry transformation acts on these fields as

$$\delta_1 \phi = \sqrt{2} \xi \psi,$$

$$\delta_1 \psi = i 2 \sigma^\mu \tilde{\xi} \partial_\mu \phi + \sqrt{2} \xi F,$$

$$\delta_1 F = i \sqrt{2} \tilde{\xi} \sigma^\mu \partial_\mu \psi.$$  \hspace{1cm} (4)

As anticipated, under the supersymmetry, a boson and a fermion form a pair and transform into each other, exactly as under the isospin symmetry, a proton and a neutron are in the same multiplet and transform into each other. Furthermore, one can see that the number of bosonic degrees of freedom is the same as that of fermions.

The interaction of the Wess–Zumino model is determined by a single function called superpotential

$$W(\Phi) = \frac{1}{2} m \phi^2 + \frac{1}{3} \lambda \Phi^3,$$  \hspace{1cm} (5)

where $m$ represents the mass and $\lambda$ is the coupling constant, as we will see shortly. $\Phi$ is the superfield which contains $\phi$, $\psi$, and $F$ in a compact form. The Lagrangian of the Wess–Zumino model is written

$$\mathcal{L} = - \partial_\mu \phi^* \partial^\mu \phi + i \bar{\psi} \gamma^\mu \partial_\mu \psi + F^* F - F \partial W(\phi) - \frac{1}{2} \partial^2 W(\phi) \psi \psi + \text{h.c.}$$  \hspace{1cm} (6)

After eliminating the auxiliary field with the help of its equation of motion, one obtains

$$\mathcal{L} = - \partial_\mu \phi^* \partial^\mu \phi + i \bar{\psi} \gamma^\mu \partial_\mu \psi + m^2 \phi^2 = m \lambda (\phi^2 + \phi^* \phi) - \lambda^2 (\phi^2 \phi^2) - \frac{1}{2} m \psi \psi - \lambda \phi \psi \psi + \text{h.c.}$$  \hspace{1cm} (7)

It is straightforward to see that the Lagrangian in eq. (6) is invariant under the supersymmetry transformation up to the total derivative, which is irrelevant when it is integrated over the whole four-dimensional space–time to obtain the action. There are two important features that one can immediately see from the Lagrangian in eq. (7). First, the mass of the boson, $m$, is the same as that of its fermionic superpartner. Second, the coupling constant of the scalar quartic interaction, $\lambda^2$, is related to the Yukawa coupling, $\lambda$, of scalar-spinor-spinor interaction. These two novel properties are crucial in supersymmetry, and apply when more than one matter field is introduced and when gauge symmetry is considered.

The possible connection of supersymmetry with the real world was recognized in the early 80s 3) as a solution to the fine-tuning problem discussed above. Here we illustrate how the supersymmetry cancels the quadratic divergence of the scalar mass correction in the Wess–Zumino model.

In quantum field theory, all quantum fluctuations contribute to physical quantities. The quantum corrections to the square of the scalar mass in the Wess–Zumino model include the contribution where virtual states are bosons (boson loop) and also the contribution where virtual states consist of fermions (fermion loop). Each of the contributions is quadratically divergent, as we discussed in the previous section. In fact, the two contributions are the same in magnitude. However, because of the difference in statistics, the sign of the fermion loop contribution is opposite to that of the boson loop contribution, and thus the quadratic divergences of the two contributions cancel with each other, resulting in at most rather harmless logarithmic divergence.

This intuitive argument is further supported by the more rigorous non-renormalization theorem, which states that the superpotential is not renormalized at all orders in perturbative expansions. The radiative corrections to the scalar mass solely originate from the wave function renormalization, which has at most logarithmic divergence. The non-renormalization theorem is valid for more complicated models with more than one matter field and/or with gauge symmetry. It is thus the basis of the argument that the supersymmetric SM solves the fine-tuning problem associated with the Higgs boson mass.

4. Supersymmetry Phenomenology

As was discussed above, supersymmetry is a very attractive candidate for physics beyond the SM, since it solves the fine-tuning problem of the Higgs boson mass. The simplest supersymmetric extension is the minimal supersymmetric standard model (MSSM). Quarks and leptons in the SM are accompanied by their bosonic superpartners called squarks and sleptons. Associated with the gauge bosons corresponding to the SM gauge group $SU(3)_c \times SU(2)_L \times U(1)_Y$, there exist their fermionic superpartners,
gauginos. For the Higgs sector, the MSSM has two Higgs doublets with opposite hypercharge, and their superpartners, higgsinos. The extension of the Higgs sector is required to cancel the gauge anomaly and also to describe the Yukawa interactions of the quarks and leptons in the supersymmetric way. Particles in the MSSM are summarized in Table I.

One of the factors that favors low-energy supersymmetry when supersymmetry manifests itself down to the energy scale close to the electroweak scale is the argument of coupling unification. The gauge coupling constants run, namely, they change when the energy scale changes. How they run depends on the particle contents. Using the precisely measured values for the gauge coupling constants as the initial conditions and the renormalization group equations for the particle contents of the MSSM, the three gauge coupling constants meet at one energy scale of precisely measured values for the gauge coupling constants, namely, they change when the energy scale changes. How they run depends on the particle contents. Using the precisely measured values for the gauge coupling constants as the initial conditions and the renormalization group equations for the particle contents of the MSSM, the three gauge coupling constants meet at one energy scale of 10^{16} \text{ GeV}.^{4} This remarkable result may be interpreted as one of the first pieces of evidence for supersymmetry grand unification, in which the three gauge groups are unified into a single group such as SU(5) or SO(10).^{5} This fact was widely recognized in the early 90s. Together with the fact that most other proposals beyond the SM fail to reconcile the precision electroweak data,^{6} the low-energy supersymmetry has been enthusiastically anticipated as a prime candidate for physics beyond the SM.

With unbroken supersymmetry, a boson and a fermion in the same supersymmetry multiplet have the same mass. This is in a direct contradiction with experimental facts. For instance, the existence of a superpartner of the electron with the same mass as the electron is clearly excluded. This implies that the supersymmetry is broken in some way to allow sufficiently heavy superparticle masses.

Supersymmetry breaking and its mediation is one of the central issues of supersymmetry phenomenology, i.e., efforts to connect the supersymmetry to the real world. Introducing supersymmetry-breaking masses does not generically affect its high-energy behavior; thus the absence of the quadratic divergence is maintained, whereas it would be spoiled if supersymmetry was violated in the coupling relations. Such desirable masses that do not induce quadratic divergence are classified as soft supersymmetry-breaking masses. It is interesting to mention that all superpartners that have not yet been discovered can naturally become heavy without generating the unwanted quadratic divergence.

Supersymmetry is, on the other hand, probably a fundamental symmetry of nature. It is thus natural to consider that it is a local symmetry, in the same spirit as gauge symmetry. Local supersymmetry is called supergravity; it inevitably includes gravity because the gauge field associated with the local supersymmetry is a spin 3/2 field and its superpartner is the graviton with spin 2. The gauge field for the local supersymmetry is a fermion, called a gravitino.

In supergravity, the aforementioned soft supersymmetry-breaking masses ought not to be given by hand, but should arise as a consequence of the spontaneous breaking. In fact when supersymmetry is spontaneously broken in a sector different from the MSSM sector, it can be mediated to the MSSM sector, resulting in the soft supersymmetry-breaking masses.\(^7\) Different mediation mechanisms of supersymmetry breaking yield different superparticle mass spectra. Therefore, we can aim to discriminate between these spectra in future experiments when supersymmetry is discovered. There are some requirements, in general, that a successful mediation mechanism should satisfy. Firstly, the masses should not be very far from the electroweak scale, i.e., of the order 100 GeV. Second, the soft masses are chosen such that they do not cause too large flavor changing neutral current (FCNC). Third, the lightest superparticle (LSP) should be neutral when the R-parity is conserved and the LSP is stable. Furthermore the LSP can constitute (at least partly) the dark matter of the universe so that its expected abundance should not exceed the observed value of dark matter abundance.

The first of the three requirements is obvious because otherwise we would lose the very motivation to induce the weak scale supersymmetry in order to solve the fine-tuning problem of the Higgs mass. This results in our expectation that forthcoming energy frontier experiments such as the LHC\(^8\) will reveal this secret of nature by discovering the superparticles directly.

The second requirement, concerning the suppression of FCNC, requires further explanation. Here, let us first consider how the FCNC is suppressed in the SM. In the SM, the flavor (or generation) mixing processes occur due to the mismatch between the mass eigenstates and the weak-interaction eigenstates in the quarks. In fact, the FCNC contribution is absent at the lowest order in the perturbation expansion, and furthermore, it is suppressed by small quark masses and/or small generation mixings of quark mass matrices. This suppression mechanism is known as the Glashow–Iliopoulos–Maiani (GIM) mechanism. In a sense, it suppresses FCNC much more strongly than would be naively expected. When we consider, on the other hand, some extension of the SM, the GIM suppression mechanism would no longer be operative so that the resulting FCNC would exceed the experimental bounds by a few orders of magnitude. The supersymmetric extension of the SM suffers from this generic difficulty. Many mediation mechanisms have been proposed to solve this problem. The first and probably the most popular approach is gravity mediation,\(^9\) where non-renormalizable interaction suppressed by Planck mass of 10^{18} \text{ GeV} mediates the supersymmetry breaking that has taken in a hidden sector. This mediation mechanism is economical because such a non-renormalizable interaction is inevitable in supergravity. The suppression of the FCNC is not automatic in this case. For instance, the minimal supergravity assumes the universal scalar masses for the squarks and sleptons (at a high-energy scale where the soft terms are given), although the justification of this assumption is somewhat unclear. In gauge mediation,\(^10\) super-
symmetry breaking is mediated to messenger (s)quarks and (s)leptons, and then transmitted to the MSSM sector via the SM gauge interaction. In this case, the squarks and sleptons with the same gauge quantum numbers have generation-independent masses, guaranteeing the suppression of the FCNC contribution from supersymmetric particles. Another compelling mechanism is anomaly mediation\(^{11}\) where the supersymmetry breaking is transferred to superconformal anomaly. In this case, the soft masses are proportional to the beta functions. This mechanism is elegant, but suffers from tachyonic masses for the sleptons because both \(SU(2)_L\) and \(U(1)_Y\) are asymptotic non-free with positive beta functions, yielding negative contributions to the squares of the slepton masses. To circumvent this problem, one has to add another source of supersymmetry mediation. Recently, a string-inspired model has been considered, which exhibits an admixture of the anomaly mediation and the gravity mediation (or the moduli mediation) of supersymmetry breaking.\(^{12}\) It turns out that this solves the tachyonic slepton mass problem.

Given the soft supersymmetry-breaking mass parameters at the high-energy scale, the physical masses carry a lot of information at the mediation scale down to the electroweak scale. Thus the measurements of the superparticle masses are expected to provide a lot of information on high-energy physics, including the gauge structure at the high-energy scale as well as the flavor structure.

It is interesting to mention that the FCNC contribution from supersymmetry may be marginally suppressed to satisfy the present experimental constraints but may reappear in future experiments. In particular, lepton flavor violation processes are promising since the SM prediction is negligibly small, even if small neutrino masses and mixing are taken into account. Here, we would like to point out that a forthcoming experiment called MeG experiment\(^{13}\) to search for \(\mu \to e\gamma\) will start soon, and it has the potential to reveal the nature of new physics within a few years.

Searches for superparticles in energy frontier experiments have been carried out for a long time. Null results place lower bounds on the masses of superparticles.\(^{14}\) Gluino and squark searches at Tevatron constrain their masses to being heavier than about 300 GeV, but much of the interesting parameter region has not been explored yet. The forthcoming LHC experiment has much greater discovery potential; its reach of the gluino mass is far above 1 TeV (\(\approx 10^3\) GeV). One should also emphasize that the detailed study of the properties of the superparticles including the precise determination of their masses and couplings is equally important as the discovery of the superparticles. A lepton collider with, e.g., \(e^+e^-\) collisions is suitable for this purpose and the International Linear Collider (ILC) has been proposed. More details on the search for and study of superparticles at the LHC as well as at the ILC will be discussed elsewhere in this volume.

Another important prediction based on supersymmetry, in particular that of the MSSM, is that there exists at least one light Higgs boson whose mass is not far from the \(Z\) boson mass. This is because the quartic couplings of the Higgs potential are related to the gauge coupling constants in the MSSM. In fact, it was shown that, at the tree level, the lightest Higgs boson is lighter than the \(Z\) boson.\(^{15}\) It was recognized some time ago that a large quantum correction comes from top-stop contributions, which this bound.\(^{16}\) Approximately, the lightest Higgs boson mass, \(m_h\), is bounded as

\[
m_h^2 \lesssim m_Z^2 + \frac{3 G_F m_t^4}{2 \pi^2} \ln \frac{m_t^2}{m_0^2}
\]

(8)

The first term on the right-hand side is the classical contribution and the second term represents the quantum correction. \(m_t\) is the top mass, \(m_t\) is the stop mass, and \(G_F\) denotes the Fermi constant. As a result of this large quantum correction, the Higgs mass can easily exceed the tree-level bound of \(m_Z \approx 91 \text{ GeV}\) and it can be as large as about 130 GeV. Although this upper bound for the Higgs boson mass in the MSSM is beyond the capability of the Higgs search at LEP, it is still well within the reach of the LHC experiment. Thus, the Higgs boson search at the LHC will provide us with a clue to low-energy supersymmetry.

5. Connection with Cosmology

At this point, we briefly mention a possible connection of the supersymmetric models with cosmology.

In supersymmetry, \(R\)-parity is assigned to be even for ordinary particles that already exist in the SM and odd for their superpartners. The conservation of the \(R\)-parity is usually assumed since it forbids very rapid proton decay caused by dimension four operators. In this case, the LSP is stable and thus it should be colorless and neutral (with respect to electric charge). Among the superparticles in the MSSM, the lightest of the neutralinos often becomes the LSP in many supersymmetric models. Here a neutralino is a linear combination of neutral gauginos, \(\tilde{g}\), \(\tilde{Z}\), and neutral higgsinos, \(\tilde{H}_1^0\) and \(\tilde{H}_2^0\). The lightest neutralino, if it is the LSP, is in the category of weakly interacting massive particles (WIMP) and is indeed a good candidate for the dark matter of the universe.\(^{17}\) What we know about dark matter is merely the necessity of non-baryonic dark matter, which constitutes about 1/4 of the total energy of the universe. The direct study of the superparticles at the LHC and ILC and various searches for dark matter will enable us to identify the nature of dark matter and hopefully understand the early history of the universe such as when and how the dark matter was produced and remains until today.

Another cosmological implication of low-energy supersymmetry is related to gravitino production in the early universe.\(^{18}\) As was mentioned earlier, the gravitino is a superpartner of the graviton, and thus its interaction is very weak. If, for instance, the gravitino is not the LSP, then it is inevitably unstable but its lifetime is relatively long. For a gravitino mass of the order 1 TeV, the lifetime exceeds one second, and the decay takes place after the big-bang nucleosynthesis (BBN) commences. The gravitino abundance at its decay must be severely constrained so as not to spoil the success of the BBN, one of the triumphs of standard cosmology. It is known that the gravitinos are produced by scattering in the thermal bath immediately after the reheat of primordial inflation. Because the yield of the gravitino produced this way is proportional to the reheat temperature, the BBN constraint places a severe upper bound on the reheat temperature.\(^{19}\) Furthermore, it was recently pointed out that scalar fields, including inflaton, generically decay
into gravitino pairs with significant branching ratios. This places yet another severe constraint on the scenarios of inflation, baryogenesis, and also the mechanism of supersymmetry breaking and mediation.

6. Road to Unification

If low-energy supersymmetry is realized in nature, we will have a chance to approach the physics of the unification of forces and matter. In fact, as was mentioned earlier, the three running coupling constants of the SM gauge interactions, $SU(3)_C \times SU(2)_L \times U(1)_Y$, meet at one energy scale at around $10^{16}$ GeV if they follow the renormalization group flow of the MSSM. This strongly implies that the grand unification of forces is achieved at this scale. This is a typical example how to probe physics at the high-energy scale by studying the relations among the parameters in the Lagrangian. Another example is given by the relation between the bottom quark mass and the tau lepton mass. The observed value of this ratio is consistent with that implied by supersymmetric grand unification where quarks and leptons are also unified. In supersymmetry, there are new supersymmetry-breaking mass parameters, such as gaugino masses and squark/slepton masses. They will provide us with crucial hints on physics at the very high energy scale, and hopefully on grand unification. The flavor violation processes such as $\mu \rightarrow e\gamma$ also provide important information on the structure of generation mixing in high-energy physics.

Another way to probe physics around the grand unified scale is to seek rare processes. Proton decay is a typical example of this approach. Proton decay such as $p \rightarrow e^+\pi$ does not occur within the SM because the baryon number is accidentally conserved in the SM at the level of renormalizable interaction. This is no longer the case when we extend the SM to grand unified theories, where the baryon number is not conserved and the exchange of particles with GUT scale mass causes the proton decay. According to the uncertainty principle of quantum mechanics, the transition probability of the process with such an extremely energetic virtual state is highly suppressed, which results in the extreme longevity of the proton. Searches for proton decay have been made, which have already excluded many grand unified models. The search for this rare process in experiments at present or in the next generation experiments may reveal evidence of grand unification.

As we mentioned in the introduction, the SM does not include gravity. At present, superstring theory is a promising candidate for ultimate unified theory including gravity. Supergravity is then recognized as a low-energy effective theory of the superstring. The experimental verification of supersymmetry as well as theoretical developments of the superstring theory will shed new light on the understanding the physics around or beyond the Planckian scale.

8) LHC home page (http://lhc.web.cern.ch/lhc/).

Masahiro Yamaguchi was born in Tsuruoka, Japan, in 1963. He obtained his B. Sc. (1985), M. Sc. (1987), and D. Sc. (1990) from the University of Tokyo. He was a research associate (1992–1995), an associate professor (1995–2003), and is currently a professor (2003–) at Department of Physics, Tohoku University. He has worked on elementary particle theory, in particular physics beyond the standard model and connection with cosmology.
1. Introduction

The most urgent and important topics of contemporary particle physics are (1) to understand the origin of “Mass” (the Electroweak symmetry braking) and (2) to discover the physics beyond the Standard model (SM). These are the main purpose of the Large Hadron Collider (LHC)\(^1\) in which two protons collide with the center-of-mass energy of 14\,TeV. The first physics collision is expected in the summer of 2008 with the lower luminosities of about \(10^{32}\,\text{cm}^{-2}\,\text{s}^{-1}\). The design luminosity of \(10^{34}\,\text{cm}^{-2}\,\text{s}^{-1}\), which corresponds to 100\,fb\(^{-1}\) per year, will be achieved within several years.

The production cross-sections of the various high \(p_T\) and high mass elementary processes are expected to be large at LHC, since gluon inside protons can contribute remarkably. Furthermore, because the LHC provides the high luminosity of 10–100\,fb\(^{-1}\) per year, large numbers of the interesting events will be observed as summarized in Table I. LHC has an excellent potential to produce high mass particles, for example, the top quark, the Higgs boson and SUSY particles. I focus on the Higgs and Supersymmetry in this note, and refer to new gauge symmetry and extra space-dimensions only briefly as alternatives.

2. Detectors

Two general-purpose experiments have been constructed, ATLAS\(^2\) and CMS\(^3\) at the LHC. The ATLAS (A Toroidal LHC Apparatus) detector is illustrated in Fig. 1, and it measures 22\,m high, 44\,m long, and weights 7,000\,tons. The characteristics of the ATLAS detector are summarized as follows:\(^4\)

- Precision inner tracking system is made with pixel, strip of silicon and TRT (Transition Radiation Tracker) with 2\,T solenoid magnet. Good performance is expected on the \(B\)-tagging and the \(\gamma\)-conversion tagging.
- Liquid argon electromagnetic calorimeter has fine granularity for space resolution, and longitudinal segmentation to obtain fine angular resolution and excellent particle identifications. It has also good energy resolution of about \(1.3\%\) for 100\,GeV \(e^\pm\) and \(\gamma\).
- Large muon spectrometer with air core toroidal magnet will provide a precise measurement on muon momenta (about \(2\%\) for 100\,GeV\(-\mu^\pm\)) even in the forward region.

Table I. Production cross-section and event numbers for major high \(p_T\) and high mass processes with an integrated luminosity of 10\,fb\(^{-1}\).

<table>
<thead>
<tr>
<th>Process</th>
<th>(\sigma) (pb)</th>
<th>Event number at LHC</th>
</tr>
</thead>
<tbody>
<tr>
<td>(W^\pm \rightarrow e^\pm \nu)</td>
<td>(6.0 \times 10^4)</td>
<td>(\sim 10^6)</td>
</tr>
<tr>
<td>(Z^0 \rightarrow \ell^+ \ell^-)</td>
<td>(5.7 \times 10^3)</td>
<td>(\sim 10^6)</td>
</tr>
<tr>
<td>(t \bar{t})</td>
<td>830</td>
<td>(\sim 10^7)</td>
</tr>
<tr>
<td>(jj,p_T &gt; 200,\text{GeV})</td>
<td>(10^6)</td>
<td>(\sim 10^6)</td>
</tr>
<tr>
<td>SM Higgs (M = 115,GeV)</td>
<td>35</td>
<td>(\sim 10^6)</td>
</tr>
<tr>
<td>(88,\text{(M = 500,GeV)})</td>
<td>(\sim 100)</td>
<td>(\sim 10^6)</td>
</tr>
<tr>
<td>(\Phi,\Phi,\text{(M = 1,TeV)})</td>
<td>(3)</td>
<td>(\sim 10^4)</td>
</tr>
</tbody>
</table>

3. Higgs Physics

A discovery of one or several Higgs bosons will give a definite experimental proof of the breaking mechanism of the Electroweak gauge symmetry, and detail studies of the Yukawa couplings between the Higgs boson and various...
fermions will give insights on the origin of lepton and quark masses. The mass of the SM Higgs boson itself is not theoretically predicted, but it’s upper limit is considered to be about 1 TeV from the unitary bound of the $W^+W^-$ scattering amplitudes, or even 200 GeV (95% C.L.) from the Electroweak precision measurements. The lower limit of the Higgs boson mass is set at 114 GeV (95% C.L.) by the direct searches at LEP. The Higgs boson should exist in the narrow mass range of 114 – 200 GeV, and lighter than 130 – 150 GeV if the Supersymmetry exists.

The SM Higgs boson, $H_{SM}^0$, is produced at the LHC predominantly via gluon–gluon fusion (GF) and the second dominant process is vector boson fusion (VBF). The production cross-sections are summarized in the report. The $H_{SM}^0$ decays mainly into $b\bar{b}$ and $\tau^+\tau^-$ for the lighter case ($\lesssim 130$ GeV). On the other hand, it decays into $W^+W^-$ and $ZZ$ with a large branching fraction for the heavier case ($\gtrsim 140$ GeV). Although its decay into $\gamma\gamma$, via the one-loop process including top quark or W boson, is suppressed from the large irreducible background, the signal from the large irreducible background. Both the ATLAS and CMS detectors have excellent energy and position resolutions for photon, and the mass resolution of the $H_{SM}^0 \to \gamma\gamma$ process is expected to be 1.3 GeV (ATLAS) and 0.9 GeV (CMS). Sharp peak appears at Higgs boson mass over the smooth distribution of background events as shown in Fig. 3. This channel is promising for the light Higgs boson, whose mass is lighter than 140 GeV, and this mode indicates the spin of the Higgs boson candidate.

VBF provides additional signatures in which two high $p_T$ jets are observed in the forward region, and only the two photons from the decay of $H_{SM}^0$ will be observed in the wide rapidity gap between these jets. The rapidity gap (no jet activity in the central region) is expected because there is no color-connection between two out-going quarks. These signatures suppress the background contributions significantly and improve the signal-to-noise ratio as shown in Fig. 4. Although the signal statistics is limited, the background contributions are dramatically suppressed and the distribution of the background becomes flat. A significance of about 4.5 $\sigma$ is obtained from VBF processes with a integrated of 30 fb$^{-1}$ it gains a significance of the inclusive $H_{SM}^0 \to \gamma\gamma$ analysis.

### 3.1 $H_{SM}^0 \to \gamma\gamma$ in GF and VBF

Although the branching fraction of this decay mode is small and there is a large background processes via $q\bar{q} \to \gamma\gamma$, the distinctive features of the signal, high $p_T$ isolated two photons with a mass peak, allows us to separate the signal from the large irreducible background. Both the ATLAS and CMS detectors have excellent energy and position resolutions for photon, and the mass resolution of the $H_{SM}^0 \to \gamma\gamma$ process is expected to be 1.3 GeV (ATLAS) and 0.9 GeV (CMS). Sharp peak appears at Higgs boson mass over the smooth distribution of background events as shown in Fig. 3. This channel is promising for the light Higgs boson, whose mass is lighter than 140 GeV, and this mode indicates the spin of the Higgs boson candidate.

VBF provides additional signatures in which two high $p_T$ jets are observed in the forward region, and only the two photons from the decay of $H_{SM}^0$ will be observed in the wide rapidity gap between these jets. The rapidity gap (no jet activity in the central region) is expected because there is no color-connection between two out-going quarks. These signatures suppress the background contributions significantly and improve the signal-to-noise ratio as shown in Fig. 4. Although the signal statistics is limited, the background contributions are dramatically suppressed and the distribution of the background becomes flat. A significance of about 4.5 $\sigma$ is obtained from VBF processes with a integrated of 30 fb$^{-1}$ it gains a significance of the inclusive $H_{SM}^0 \to \gamma\gamma$ analysis.

### 3.2 $H_{SM}^0 \to \tau^+\tau^-$ in VBF

$H_{SM}^0 \to \tau^+\tau^-$ provides high $p_T \ell^\pm$, when $\tau$ decays leptonically, and it can makes a clear trigger. Momenta carried by $\nu$'s emitted from $\tau$ decays can be solved approximately by using the $E_T$ information, and the Higgs mass can be reconstructed. Figure 5 shows the mass distributions of the reconstructed tau-pair. The mass resolutions of about 10% can be achieved, and the signal can be separated from the Z-boson production background. The performance of the $E_T$ measurement is crucial in this analysis as the same as in the SUSY search. We can obtain a significance of about 4 $\sigma$ for this channel for $m_{H_{SM}} < 130$ GeV with a luminosity of 10 fb$^{-1}$, and also this channel provides a direct information on the coupling between the Higgs boson and a fermion, the tau lepton.
3.3 $H^0_{SM} \rightarrow ZZ \rightarrow \ell^+\ell^-\ell^+\ell^-$ in GF

Dominant decay modes of the heavier $H^0_{SM}$ are $ZZ$ and $W^+W^-$. The four-lepton channel ($H^0_{SM} \rightarrow ZZ \rightarrow \ell^+\ell^-\ell^+\ell^-$) is very clean and called “the gold-plated”. Although the branching fraction of $ZZ \rightarrow \ell^+\ell^-\ell^+\ell^-$ is small, a sharp mass peak is expected as shown in Fig. 6, if the Higgs boson is heavier than 140 GeV. Mass resolution of the four lepton system is typically 1%.11,12) As shown in Fig. 6, the main background process is $ZZ$ production, which gives continuous distribution above 200 GeV. Small contaminations come from $t\bar{t}$ and $Z^0bb$, in which semi-leptonic decays of bottom quark are identified as isolated leptons. In order to reduce these contaminations, leptons are required to be well isolated from hadron activities and their track impact parameters should be consistent with zero. The $\ell^+\ell^-\ell^+\ell^-$ channel has a good performance in the wide mass range from 130 to 800 GeV, except around 170 GeV, where the $W^+W^-$ is the dominate decay mode.

3.4 $H^0_{SM} \rightarrow W^+W^- \rightarrow \ell^+\ell^+\ell^-\ell^-$ in VBF

When the Higgs mass is around 170 GeV, the branching fraction of $H^0_{SM} \rightarrow W^+W^-$ is almost 100%. The whole mass range of 130–200 GeV is well covered by the analysis of VBF $H^0_{SM} \rightarrow W^+W^- \rightarrow \ell^+\ell^-\ell^+\ell^-$. The transverse mass, $M_T$, is defined as $\sqrt{2 E_T P_T (\ell^+\ell^- ) (1 - \cos \phi)}$, in which $\phi$ is the azimuthal angle between the $E_T$ and $P_T (\ell^+\ell^-)$. Figure 7 shows the $M_T$ distribution and a clear Jacobian peak is observed above smooth background distributions. The main background process is $t\bar{t}$, and this can be suppressed by using the azimuthal angle correlation between the dileptons. Since the Higgs boson is a spin zero particle, the helicities of the emitted $W$ bosons are opposite. The leptons are then emitted preferably in the same direction due to the 100% Parity violation in $W$ decays.

3.5 Overall discovery potential of $H^0_{SM}$

Discovery potential of $H^0_{SM}$ are summarized in Fig. 8 as a function of the Higgs mass with an integrated luminosity of $10 \text{ fb}^{-1}$. $H^0_{SM} \rightarrow \gamma\gamma$ in GF and VBF and $\tau^+\tau^-$ channels have good potential in the mass region lighter than 130 GeV. For the heavy mass case ($\geq 130 \text{ GeV}$), decay to $ZZ \rightarrow \ell^+\ell^-\ell^+\ell^-$ and $W^+W^-$ have an excellent performance even above 10$\sigma$. When both ATLAS and CMS significances are combined, we can perform to discover (5$\sigma$ C.L.) and exclude (98% C.L.) the Higgs boson up to 1 TeV mass with the integrated luminosities of 5 and 1 $\text{fb}^{-1}$, respectively. Therefore the crucial test of the Higgs mechanism of the symmetry braking can be performed within 2009.
3.6 Measurement of mass and couplings of the Higgs boson

Measurements on the properties of the discovered Higgs boson give further insights to the origin of masses. Higgs mass can be measured precisely in $H_\text{SM} \rightarrow \gamma \gamma$ and $H_\text{SM} \rightarrow ZZ (\rightarrow \ell^+ \ell^- \ell'^+ \ell'^-)$. Accuracy of less than 0.2% error can be achieved with $L = 100 \text{fb}^{-1}$, if the mass is smaller than 500 GeV. When the Higgs boson is heavier than 500 GeV, the resonance becomes too broad, and the precision becomes worse.

Measurements of the couplings between the Higgs boson and fermions/Gauge bosons will give the direct informations of the origin of “Mass”, and fermion’s couplings will give the first evidence of Yukawa couplings. Figure 9 shows the accuracy of the coupling measurements between the Higgs boson and fermion/Gauge bosons. We can measure without an assumption the relative magnitudes of the couplings normalized to the coupling between the $W$ and the Higgs boson. We can determine the coupling between the $Z$ and the Higgs boson precisely, where the accuracy of 5–10% can be achieved in all the mass region. Couplings between the Higgs and the 3rd generation fermions (top and tau) are determined with accuracies of 10–15%, but can not determine well for the bottom quark.

4. Supersymmetry

Supersymmetric (SUSY) standard models are most promising extensions of the SM, because the SUSY can naturally explain the weak boson mass scale. Furthermore, the SUSY models provide a natural candidate of the cold dark matter, and they have given a hint of the Grand Unification in which three gauge couplings of the SM are unified at around $2 \times 10^{16}$ GeV. In these theories, each elementary particle has a superpartner whose spin differs by 1/2 from that of the particle. Discovery of the SUSY particles should open a new epoch of the fundamental physics, which is another important purpose of the LHC project.

4.1 Introduction of Super-Gravity model

There are, in general, more than 100 free parameters to describe soft SUSY breaking, but there are strong constraints among them from the smallness of the flavor-changing neutral current. Among various models of the SUSY breaking that can naturally satisfy these constraints, Super-Gravity model, Gauge-mediated model, and Anomaly-mediated model are predictable and promising. Performance of the ATLAS experiments based on the Super-Gravity model is summarized in this section.

Minimal Super-Gravity Model (mSUGRA) is a special case of the Minimal Supersymmetric Mode (MSSM), in which the soft SUSY breaking terms are assumed to be communicated from the SUSY breaking sector by gravity only and that these terms are universal at the GUT scale. There are only five free parameters in this model; $m_0$ (universal mass of all scalar particles at the GUT scale), $m_{1/2}$ (universal mass of all gauginos at the GUT scale), $A_0$ (common scale of trilinear couplings at the GUT scale), $\tan \beta$ (ratio of VEV of two Higgs fields at the Electroweak scale) and the sign of $\mu$ (Higgsino mass term).

In this model, masses of supersymmetric particles are mainly determined by $m_{1/2}$ and $m_0$; $\tilde{g}$ becomes heavy due to large radiative corrections, and its mass is approximately 2.5 $m_{1/2}$ at the energy scale of the LHC. Higgsino mass ($|\mu|$) becomes larger than the weak Gaugino masses at the EW scale, except for the case of $m_0 \gg m_{1/2}$. Then the lighter states of the neutralinos, $\tilde{\chi}_1^0$ and $\tilde{\chi}_2^0$, become almost pure Bino- and Wino-states ($\tilde{\chi}_1^0 \sim \tilde{B}_1$, $\tilde{\chi}_2^0 \sim \tilde{W}_0$), and the lighter state of the charginos, $\tilde{\chi}_1^\pm$, is also Wino-like ($\tilde{\chi}_1^\pm \sim \tilde{W}_1^0$). Scalar lepton masses are determined mainly by $m_0$ and $m_{1/2}$.

On the other hand, scalar quark masses depend on both $m_0$ and $m_{1/2}$. Here is a typical spectrum of the SUSY particles:

- $m(\tilde{g}) \sim 2.5 m_{1/2}$
- $m(\tilde{t}_L) \sim 0.4 m_{1/2}$
- $m(\tilde{t}_R) \sim m(\tilde{\chi}_1^\pm) \sim 0.8 m_{1/2}$
- $m(\tilde{\ell}_R^\pm) \sim \sqrt{m_0^2 + 0.15m_{1/2}^2}$
- $m(\tilde{\ell}_R^0) \sim \sqrt{m_0^2 + 0.5m_{1/2}^2}$
- $m(\tilde{q}_L) \sim \sqrt{m_0^2 + 5m_{1/2}^2}$

Masses of the lighter state of third generation scalar fermions ($t_1$, $b_1$, and $t_1$) depend also on $A$ and $\tan \beta$ and they are generally lighter than the first and second generation scalar fermions because of the following two reasons. Firstly, one loop radiative corrections due to their Yukawa
4.2 Event topologies of SUSY events and discovery potential

The couplings of the colored Sparticles (Bino/Wino-eigenstates) are presented in the table. Each event contains a weak gaugino plus quark(s). In this case, Higgsino mass (μ) becomes much larger than gaugino masses at the EW scale, and the Higgsino component decouples from the lighter mass-eigenstates as explained in the text.

<table>
<thead>
<tr>
<th>m(˜g) &lt; m(˜q)</th>
<th>m(˜g) = m(˜q)</th>
<th>m(˜g) &lt; m(˜q)</th>
</tr>
</thead>
<tbody>
<tr>
<td>˜g → q ˜B (1/7)</td>
<td>˜g → q ˜W (2/7)</td>
<td>˜g → q ˜q</td>
</tr>
</tbody>
</table>

Fig. 10. Summary of decays of the colored Sparticles: Bino/Wino-eigenstates presented in the table can be regarded as mass-eigenstates, (B → ˜g, W → ˜q), when m_0 is not too larger than m_{1/2}. In this case, Higgsino mass (μ) becomes much larger than gaugino masses at the EW scale, and the Higgsino component decouples from the lighter mass-eigenstates as explained in the text.

Decay modes of ˜g and ˜q are controlled by the mass-relation between each other, and are summarized in Fig. 10. If kinematically possible, they decay into 2-body through the strong interactions. Otherwise, they decay into an electro-weak gaugino plus quark(s).

4.2 Event topologies of SUSY events and discovery potential

If kinematically possible, they decay into 2-body through the strong interactions. Otherwise, they decay into an electro-weak gaugino plus quark(s).

The following four SM processes can potentially have ˜E_τ event topology with jets:

- W ±(→ ℓν) + jets,
- Z(→ ℓν, ℓ+ℓ-) + jets,
- ττ + jets,
- QCD multi-jets with mis-measurement and semi-leptonic decays of bb and cc + jets

We require at least four jets with p_T ≥ 50 GeV, and 0.2 for the no-lepton mode to reduce the QCD background. Effective mass, M_eff = E_τ + ∑_jets p_T, is a good variable to distinguish the signal from the background listed above. Excess coming from the SUSY signals can be clearly seen as shown in Fig. 12 for both no-lepton and one-lepton modes. It is the dominant background process for the one-lepton mode, and all of the four processes listed above contribute to no-lepton mode. The QCD processes contribute only to the small E_τ and M_eff. The M_eff distribution has steeper slope for the SM background processes as shown in these figures. On the other hand, the distribution of the SUSY signals has a broad peak at large value (around 1.5 TeV in these figures), which is proportional to min(m(˜g), m(˜q)).

The SM background contributions can be suppressed dramatically by requiring di-leptons as shown in Fig. 13. They show the E_τ distributions for the OS- and SS-dilepton modes, where ℓ is the dominant background process for both modes. Although the background contributions is seriously suppressed (the SS dilepton mode is the almost background free), the statistics of the SUSY signal is also typically a few % of no-lepton mode. Discovery potentials of dilepton modes are limited in the early stage of the collision, but these modes are important to reconstruct the decay chain mentioned later.

Figure 14 show the 5σ-discovery potential in the m_{0}−m_{1/2} and the m_q−m_B plane for tan β = 10 with an integrated luminosity of 1 fb^{−1}. And the lowest shows the same figure.
with an integrated luminosity of 30 fb⁻¹. The results with three modes of no-lepton, one- and OS-dilepton are shown, and the hatched regions below each line show the systematic errors coming from the background estimations. The similar discovery potentials are obtained for no-lepton and one-lepton modes. As shown in these figures, g and q can be discovered up to the mass of ~1.5 TeV with a luminosity of 1 fb⁻¹, which corresponds to just one month run with 10³³ cm⁻² s⁻¹. The discovery reach does not depend strongly on tan β. The interesting region for the relic density of the dark matter is almost covered with just 1 fb⁻¹. g and q can be discovered up to ~2.5 TeV with a luminosity of 30 fb⁻¹. g and q, whose masses are about 2.7/3.0 TeV, can be discovered/excluded finally with a luminosity of 10⁰ fb⁻¹. This luminosity is corresponding to one year run with design detector are taken into account. Since two undetected LSP’s exist in each event, there are six unknown momentum components in addition to the \( \tilde{\chi}_0^0 \) mass. No mass peak is expected in general. However it is possible to use kinematic end points of various distributions as follows.¹¹,²⁴,²⁵

- Select specific decay chain exclusively. For example, the invariant masses of the SUSY particles are strongly correlated, accuracies of these masses are about 3–10% for \( m(\tilde{q}_L) = 800 \text{ GeV} \). Mass of the missing \( \tilde{\chi}_1^0 \) can be determined with an accuracy of about 10%, which is an extremely important result, since it can be the dark matter.

Reconstructions of the SUSY particle masses for the other decay-patterns are also possible¹¹,¹²,²⁵ by using the similar techniques. When we cannot identify a successive decay

\[
\tilde{q}_L \rightarrow \tilde{\chi}_1^0 q \rightarrow (\tilde{\ell}_R^\pm \tilde{\ell}_L^\mp) q \rightarrow ((\tilde{\ell}_R^\pm\tilde{\ell}_L^\mp)\ell^{\mp})q
\]

- Make various distributions of invariant masses and \( p_T \).
- Kinematic constraints are obtained from the end points of these distributions for the selected chain. These end points can be determined by the masses of the SUSY particles. SUSY events become background itself for detailed study, since there are many cascade decay patterns in \( \tilde{q} \) and \( \tilde{g} \). It is critical point that we can find out or not a useful decay-pattern in the SUSY events.

If there are three 2-body decay chains like the above example, which is a dominant mode in the parameter space of \( m_0 \leq 0.8 m_{1/2} \), full reconstruction of masses is possible model-independently. The invariant mass distributions of \( \ell \ell, \ell q \), jet, and \( \ell \ell \) + jet can be calculated, and the three kinematic end points and one production threshold of the 4-body system (\( \ell^+ \ell^- q \tilde{\chi}_1^0 \)) are obtained. On the other hand, there are four unknown masses (\( \tilde{q}_L, \tilde{\chi}_1^0, \tilde{\ell}_R^\pm, \text{and} \tilde{\chi}_1^0 \)). Then all the four unknown masses can be determined model-independently. Although the errors of the determined masses are strongly correlated, accuracies of these masses are about 3–10% for \( m(\tilde{q}_L) = 800 \text{ GeV} \). Mass of the missing \( \tilde{\chi}_1^0 \) can be determined with an accuracy of about 10%, which is an extremely important result, since it can be the dark matter.

Reconstructions of the SUSY particle masses for the other decay-patterns are also possible¹¹,¹²,²⁵ by using the similar techniques. When we cannot identify a successive decay
chain listed above, the number of the observed constraints is less than that of the unknown masses. Therefore, some assumptions on SUSY breaking pattern are necessary to determine the mass spectrum. All the massed of $\bar{g}-\bar{q}_{L/2}, \bar{\ell}^\pm-\bar{\chi}_L^2$, and $\bar{\chi}_R^0$ can be determined within the assumed model, and we will be able to test various SUSY models by using these reconstructed mass spectra.$^{29}$

5. New Gauge Bosons

Many theories beyond the SM predict the existence of additional U(1) or SU(2) gauge groups and a discovery of new gauge bosons, $Z'$ and $W'$, should be its clear evidence of extension of the SM. Current mass limits on $Z'$ and $W'$ are about 800 GeV obtained at the Tevatron.

$Z'$ and $W'$ are produced by the Drell–Yan ($q\bar{q}$ annihilation) process at the LHC, and they decay into fermion pairs. Figure 15 shows the transverse mass distribution between the $\mu^\pm$ and $E_T$ momenta, $M_T = \sqrt{2P_T(\mu^\pm)E_T(1-\cos\Delta\phi_{\mu missing})}$. Clear Jacobian peak will be observed at the mass of $W'$, and the background contributions are small in such a high $M_T$ region. The results are obtained by assuming that the $W'$ coupling are the same as the $W^\pm$ coupling and no decay mode into new particles. The $W'$ can be discovered or excluded up to mass of 4.5 and 5 TeV, respectively$^{12}$ with an integrated luminosity of 10 fb$^{-1}$, by combining the both $\mu\nu$ and $e\nu$ channels. $Z' \rightarrow \ell^+\ell^-$ makes a resonance peak in the high mass region, and $Z'$ can be discovered up to mass of 3.8 TeV with an integrated luminosity of 10 fb$^{-1}$.

6. Large Extra Space-Dimension and Black Holes

There is much recent theoretical interests that an extra-dimension exists in addition to the $3 + 1$ normal space–time dimensions, and it is compactified to the size of a few TeV, which is related to the true “Planck scale” of the complete theory. This new fundamental scale is denoted as “$M_P$”. The idea is introduced as a possible solution of the hierarchy problem of the SM. The scatter processes at a few TeV scale should be treated within multi-dimensions, and the gravity interaction becomes as strong as the other interactions. For example, gravitons (G) are emitted in the hard scatter of $gg \rightarrow gG$ and escapes the detection. This makes an event with a single jet and large $E_T$ (monojet). We have the discovery-potential up to $M_T = 7$ TeV (additional dimension $\leq 3$) with an integrated luminosity of 100 fb$^{-1}$. $^{12}$

If the true Plank scale is in the order of TeV, the Schwarzschild radius, $R_s$, is also sizable,

$$R_s = \frac{1}{\sqrt{\pi}M_P}\left[\frac{M_{BH}}{M_P}\left(8\pi\left(\frac{n+3}{2}\right)\frac{n+2}{n+1}\right)\right]^{1/(1+n)}$$

for $n$ extra space-dimensions. Mini black holes of a few TeV mass can be produced with parton–parton collisions within $R_s$ at the LHC. Large production cross-section of order $\pi R_s^2 \sim 100\text{ pb}$ is expected by classical arguments.$^{20}$ The produced black hole decays through Hawking evaporation with the temperature:

$$\frac{dN}{dt} \sim \frac{1}{\pi R_s^2}$$
Physics at the TeV scale. The SM Higgs boson ($H_0$).

**Summary**

The LHC is the first experiment that probes directly physics at the TeV scale. The SM Higgs boson ($H_0^{SM}$) should be discovered with an integrated luminosity of 10 fb$^{-1}$, if it exists. $H_0^{SM} \rightarrow \gamma\gamma$ in GF (gluon fusion) and VBF (vector-boson fusion) and $H_0^{SM} \rightarrow \tau^{+}\tau^{-}$ in VBF are important processes for the Higgs lighter than 140 GeV. The decays into $ZZ (\rightarrow \ell^{+}\ell^{-}\ell^{+}\ell^{-})$ and $W^{+}W^{-} (\rightarrow \ell\ell\nu\bar{\nu})$ play important role if it is heavier. The Higgs boson mass can be determined precisely, and the couplings between Higgs and fermions/Gauge bosons can be measured with accuracies of about 5–20%. We can perform a critical test on the Higgs mechanism and will understand the origin of the elementary particle masses.

Supersymmetry should be discovered at the LHC, if gluino ($\tilde{g}$) and squarks ($\tilde{q}$) are lighter than about 2.7 TeV. Signals will be found not only in the ($\ell\ell +$ jets) channel but also in [$\ell\ell +$ jets + lepton(s)] channels, where $\ell\ell$ stands for the missing transverse momentum carried away by the lightest supersymmetric particle (LSP). The SUSY particle masses, including that of the LSP, can be determined by exclusive studies model-independently, when a three successive two-body decay chain is identified. In more general cases, they can still be determined within various supersymmetry models.

The LHC can also discover new physics other than supersymmetry, such as new gauge boson symmetry and extra space dimensions.

Fig. 16. Expected black hole signal with a mass of 5 TeV (simulated with ATLAS detector). The central colored tracks show the observed charged particles, and the yellow towers in the green and orange regions show the energies deposited in the EM and hadron calorimeters, respectively. Eleven energetic partons are emitted from the black hole decay.

$$T_H = \frac{M_P}{M_{BH}} \left[ \frac{n + 2}{8\Gamma \left( \frac{n + 3}{2} \right)} \right]^{2/(n+1)}$$

As the produced black hole is heavier, the $T_H$ becomes smaller (but still higher than 100 GeV), and many particles are emitted as shown in Fig. 16. These particles are energetic as the same orders of $T_H$. High $p_T$ multi-particles are emitted spherically, and this event topology is quite different from the SM background processes. We can discover mini black holes up to mass of several TeV, but we need more theoretical studies on the production and decay processes of the mini black holes at the LHC.

1) http://lhcf.web.cern.ch/lhcf/
2) ATLAS technical proposal: CERN/LHCC/94-43.
6) http://lepewwg.web.cern.ch/LEPEWWG/plots/winter2007/
9) For a review, see, e.g., M. Spira: hep-ph/9705337.
13) $\nu$ is emitted along the momentum direction of the observed particles: Collinear approximation.

Shoji Asai was born in Ishikawa, Japan in 1967. He obtained D.Sc. (1995) degrees from the University of Tokyo. He solved the orthopositronium lifetime problem in his thesis. From 1995, Research Assistant of the University of Tokyo and his research interesting was SUSY search at LEP. From 2003, Associate Professor of the University of Tokyo. Research interests are on Physics at LHC (Susy and Higgs), and precision measurements of QED using positronium.
Physics at International Linear Collider (ILC)

Hitoshi YAMAMOTO*

Graduate School of Science, Tohoku University, Sendai 980-8578

(Received June 20, 2007; accepted August 27, 2007; published November 12, 2007)

International Linear Collider (ILC) is an electron–positron collider with the initial center-of-mass energy of 500 GeV which is upgradable to about 1 TeV later on. Its goal is to study the physics at TeV scale with unprecedented high sensitivities. The main topics include precision measurements of the Higgs particle properties, studies of supersymmetric particles and the underlying theoretical structure if supersymmetry turns out to be realized in nature, probing alternative possibilities for the origin of mass, and the cosmological connections thereof. In many channels, Higgs and leptonic sector in particular, ILC is substantially more sensitive than LHC, and is complementary to LHC overall. In this short article, we will have a quick look at the capabilities of ILC.

KEYWORDS: ILC, Higgs boson, New Physics, supersymmetry, extra dimensions, cosmology, LHC
DOI: 10.1143/JPSJ.76.111014

1. Introduction

The standard model is an astonishingly successful theory in describing what have been observed in the field of elementary particles. The Higgs particle, which gives mass to all massive particles, is at the core of the standard model, but so far has not been found. Furthermore, if one tries to calculate the radiative correction to the mass squared of Higgs, it diverges quadratically with the cut off energy, and if one assumes that the standard model is correct up to the Higgs, it diverges quadratically with the cut off energy, but so far has not been found. Furthermore, if one tries to include the gravitational force.

A theoretical solution to the fine-tuning problem is provided by supersymmetry (SUSY) which postulates that every particle in the standard model has its so-called super-partner (called a super particle or a s-particle) whose spin differs by one half from that of the original particle. Not only the inclusion of the super particles naturally cancels out the quadratic divergence of the Higgs mass correction, in SUSY the gauge coupling constants converges to a single value at the grand unification scale itself. Since precision measurements so far shows that the standard model Higgs should lie below ~200 GeV, this is only possible if the original mass and the correction are canceling out to an astonishing precision. This unpleasant situation is referred to as the fine-tuning problem, or the naturalness problem. The problem is in part caused by the large difference in energy scale from the Higgs mass to the grand unification scale, and in this context, it is referred to as the hierarchy problem. Also, the standard model does not include the gravitational force.

A theoretical solution to the fine-tuning problem is provided by supersymmetry (SUSY) which postulates that every particle in the standard model has its so-called super-partner (called a super particle or a s-particle) whose spin differs by one half from that of the original particle. Not only the inclusion of the super particles naturally cancels out the quadratic divergence of the Higgs mass correction, in SUSY the gauge coupling constants converges to a single value at the grand unification scale. Furthermore, the gravitational force can also be naturally incorporated in SUSY. As a bonus, SUSY has candidates for the dark matter which is thought to consist of unknown stable massive particles and accounts for one quarter of the energy of the universe.

Even though SUSY is an attractive theory with many merits for us, nature of course would not care about our conveniences. There are several alternative models that address the fine-tuning problem, and some of them may have connection to the reality of nature. Examples are the models with extra dimensions which postulate the existence of space dimensions more than our 3 (space) + 1 (time) dimensions and the little Higgs model where the Higgs particle is considered to be composite.

The physics potential of ILC has been extensively studied and documented. As we will see below, the standard model Higgs particle will have distinctive signals at ILC, and SUSY and other alternative models also have many possibilities of being found and studied at ILC. The advantage of ILC with respect to LHC is in the general cleanliness of the events where two elementary particles (an electron and a positron) with known kinematics and spin define the initial state, and the high resolutions of the detector that are made possible by the relatively low absolute rate of background events. The capability of ILC is further enhanced by the options such as the \( \gamma \gamma \) collision, \( e^- e^- \) collision, and Z-pole running (“Giga-Z”).

2. ILC Machine Parameters and Detectors

The basic parameters, such as energy and luminosity, of ILC are described in the parameter report. The baseline machine allows for a center-of-mass energy range between 200 and 500 GeV and luminosity of 500 fb\(^{-1}\) in the first four years of running not counting the year zero. The energy scan is possible at any energy within the range, and the electron polarization is at least 80%. Two detectors are expected which may be in a push–pull configuration.

For each of the two beams, a bunch is \( \sigma_y = 5.7 \text{ nm} \), \( \sigma_x = 655 \text{ nm} \), and \( \sigma_z = 300 \text{ mm} \), and contains \( 2 \times 10^9 \) particles. About 3000 bunches with 308 ns bunch separation form a train of about 1 ms which comes with 5 Hz repetition rate. The collision occurs with crossing angle of 14 mrad.

The highest priority beyond the baseline is the energy upgrade to approximately 1 TeV, and the upgraded machine should be able to collect 1 ab\(^{-1}\) in 3 to 4 years after the baseline running. The options include: running at 500 GeV to double the luminosity to 1 ab\(^{-1}\), \( e^- e^- \) collision, positron polarization of 50% or more, Z-pole running, WW threshold running, and \( \gamma \gamma \) and \( e^- \gamma \) collisions using backscattered laser
beams. The priorities of these options will depend on the results of LHC and the baseline ILC. In the following, the baseline machine with 200 to 500 GeV center-of-mass energy is assumed unless stated otherwise.

The physics of ILC is realized through synthesis of unprecedented performances of both machine and detectors. ILC detectors can take advantage of the relatively low rates and low radiation doses to achieve momentum resolution that is order of magnitude better, jet energy resolution factor of two better, and the vertex resolution several times better than those at the previous electron–positron colliders. As we will see below, these performances are not overkill; rather, they are needed to realize the physics potential of ILC.

## 3. Standard Model Particles

We start from the particles that are ingredients of the standard model. Their properties and interactions with other particles, however, may reveal physics beyond the standard model. The goal is to look at the behavior of the members of the standard model to see if there is any hint of new physics. The physics of ILC is realized through synthesis of these options; namely, polarization, respectively, and $500\,\text{fb}^{-1}$ at $\sqrt{s} = 500\,\text{GeV}$ and $1\,\text{ab}^{-1}$ at $\sqrt{s} = 800\,\text{GeV}$, anomalous couplings can be measured with typical errors of $10^{-3}$ relative. The $WW\gamma$ magnetic dipole coupling $\kappa_{\gamma}$, in particular, can be measured to $10^{-4}$, which is more than order of magnitude better than LHC with the same years of running. The triple gauge coupling $e^- e^+ \rightarrow V\gamma$ can also be studied by the single gauge boson productions $e^- e^+ \rightarrow e^- e^+ v\nu$, $v\nu Z$, and also by the $e\nu$ and $\gamma\nu$ options; namely, $e^- e^+ \rightarrow W^- \gamma$ and $W^- W^- \gamma$ where the ZZ coupling does not contribute.

If Higgs is not found at LHC or ILC, it may indicate that $W$-pair can form a bound state which could be found in the $WW$ scattering process $e^- e^+ \rightarrow VV$ where $V$ is $W$ or $Z$ and $f$ is $e$ or $\nu$, or by triple gauge boson productions $e^- e^+ \rightarrow VV\nu$. At ILC, one can tell the initial and final states of the gauge boson scatterings, which is often difficult at LHC.

If no Higgs or no new particles are found, precision measurements on $Z$ become important. The Giga-$Z$ option can collect 1 billion $Z$'s in a few months, and can improve by more than one order of magnitude those measurements that use $b$-tagging and/or beam polarizations. The improved $b$-tagging is realized by the excellent vertexing capability of ILC detectors.

Couplings of fermions and gauge boson can also be studied by $e^- e^+ \rightarrow f\bar{f}$ ($f$ stands for a fermion), where anomalous couplings may be parametrized by $(1/\Lambda_{ij}^2)(e^- e^+ e^- e^+ f f)/\sqrt{2}$ ($ij = L, R$). ILC is sensitive to $\Lambda_{ij}$ of typically 20 to 100 TeV.

The $e^- e^+ \rightarrow f\bar{f}$ modes are also sensitive to existence of an extra $Z$ boson ($Z'$) even when the mass of $Z'$ is above the CM energy. Such extra gauge bosons appear in many extensions of the standard model. Some examples are the $E_6$, $\chi$ model ($\chi$), left–right symmetric model (LR), Littlest Higgs model (LH), Simplest Little Higgs model (SLH), and model with extra dimensions where $Z$'s particles are actually spin-2 Kaluza–Klein excitations of gravitons (KK). The signatures appear in the forward-backward asymmetry of the $f\bar{f}$ production and in the dependence of the cross section on the beam polarization. The resolving power of ILC is in the two-dimensional space of $C_L^f$ and $C_R^f$ is shown in Fig. 2 for $e^- e^+ \rightarrow \mu^+ \mu^-$, where $C_L^f, R$ are the left- and right-handed $Z'\ell$ coupling coefficients where the lepton universality is assumed. Electron and positron polarizations of 80 and 60% respectively, are assumed. There are quadratic ambiguities due to the sign-independence of coupling coefficients. LHC may find a $Z'$ resonance, but it would take ILC to identify the underlying theory.

### 3.2 Top quark

The top quark is the heaviest elementary particle observed so far, and its mass $\sim 174\,\text{GeV}$ is in the range of the electroweak symmetry breaking. Its large mass indicates that it couples to Higgs strongly and thus should be sensitive to the structures in the Higgs sector, or whatever is responsible for creation of masses. In many models beyond the standard model, the Higgs mass strongly depends on the top mass. In MSSM (the minimal supersymmetric standard model), for example, an error in the top mass corresponds to a similar
error in the Higgs mass, which means that precision measurements of the top and Higgs masses serve as a stringent test of theoretical models. In some cases, non-standard top couplings may be the only area new physics can be found.

The top mass \( m_t \) is best measured by the \( e^+e^- \rightarrow t\bar{t} \) threshold scan, taking about 5 fb\(^{-1} \) each at several points of CM energy. Since the top quark decays before it hadronizes, the excitation curve, i.e., the cross section as a function of CM energy, around the threshold can reliably be calculated. It is affected by the beam energy spread, initial-state radiation, beamstrahlung (radiation from a beam particle under the coherent electromagnetic field of the incoming bunch), as well as the higher order corrections which has been performed up to including some of the next-to-next-to-leading logarithms (NNLL).\(^{15} \) The experimental and theoretical uncertainties are of the same order, and the resulting overall error on \( m_t \) is expected to be 100 to 200 MeV which can be compared to 1 to 2 GeV at LHC. The threshold scan also yield the top width to a few percent of its value which is around 1.5 GeV.

The production and decay of top quark in \( e^+e^- \rightarrow t\bar{t} \), \( t \rightarrow bW \) can be studied near the threshold, well above the threshold, or below the threshold (where one of the top quark is off-shell). The production is sensitive to \( t\bar{t}Z \) and \( tt\gamma \) couplings and the decay is sensitive to \( tbW \) coupling. Many beyond-the-standard models predict deviations in these couplings from the standard-model values. The models with fourth generation have both \( tbW \) coupling and the left-handed \( tbW \) coupling as well as the expected deviations for the top-flavor model and the little-Higgs model with T-parity, and the model with fourth generation. The numbers shown on the line for T-parity are the strength of the Higgs-top- (top partner) coupling and those on the line for the top flavor model are the mass of the extra Z boson. Furthermore, the KK mode of graviton with mass 10 to 100 TeV in Randall–Sundrum models\(^{18} \) may be indirectly detected as anomalous \( t\bar{t} \) production.

### 3.3 Higgs particle

Our current knowledge on the mass of the Higgs particle mainly comes from the LEP experiments.\(^{19} \) Within the framework of the standard model, Higgs mass \( m_H \) is bounded as 114.4 < \( m_H \) < 166 GeV at 95% confidence level, where the lower limit is from direct searches and the upper limit is by an overall fit of the standard model parameters to the data. On the other hand, Higgs in the MSSM is constrained to be less than 135 GeV, which is lower than the upper limit in the standard model. These Higgs particle, if they exist, will be found at LHC within the first few years of running. At ILC, even though the start would be many years later than LHC, the same level of discovery sensitivity can be obtained by one day of running at the design luminosity. With its clean initial and final states, and high resolutions of the ILC detectors, ILC will be able to perform measurements on spin and parity of the Higgs particle, and determine coupling strengths to various particles model-independent ways.

The primary production channels of the standard model Higgs are \( e^+e^- \rightarrow Z^+Z^- \) (Higgs-strahlung) and \( e^+e^- \rightarrow \nu\bar{\nu}H \) (WW fusion) as shown in Fig. 4. The Higgsstrahlung dominates at low CM energies (< 500 GeV) and the WW fusion dominates at high CM energies (~ 1 TeV). For \( m_H \) of 120 GeV, an integrated luminosity of 500 fb\(^{-1} \) at CM energy of 500 GeV will generate \((3.4 \times 10^3) \) Higgs particles in each of the two production channels. The decay branching fractions of Higgs are shown in Fig. 5. If the

\[ e^+ \rightarrow H^+ \]

\[ H 

\[ \rightarrow e^+ \]

\[ Z 

\[ \rightarrow e^- \]

\[ W 

\[ \rightarrow H \]

\[ \nu_e \]

\[ \bar{\nu}_e \]
Higgs mass is below around 140 GeV, it decays primarily to $b\bar{b}$ with a few % each for $c\bar{c}$, $\tau\bar{\tau}$, and $gg$ branching fractions. The width of Higgs in this mass range is less than 10 MeV. For $m_H$ larger than around 150 GeV, it decays primarily to $WW$ with the $ZZ$ channel following at 20% level. The $t\bar{t}$ final state opens for $m_H$ larger than around 350 GeV and peaks for $m_H \sim 500$ GeV at 20% branching fraction. At $m_H$ of around 500 GeV, the Higgs is quite broad with $\Gamma_H \sim 100$ GeV.

Figure 6 shows the recoil mass distribution for $e^+e^- \rightarrow ZZ$, $Z \rightarrow \mu\mu$ with 500 fb$^{-1}$ at CM energy of 300 GeV. Peaks corresponding to different values of $m_H$ are shown together with the background from $e^+e^- \rightarrow ZZ$ followed by one or both of the Z's decaying to $\mu\mu$. Since the Higgs particle is not reconstructed, the method is independent of the Higgs decay modes including the case where the decay is invisible. The range of detectable Higgs mass reaches close to the CM energy itself; more precisely, up to CM energy minus $m_Z$.

The Higgs mass is obtained from the recoil mass distribution itself. Under the same conditions used for Fig. 6, the error in $m_H$ is $\sim 70$ MeV which improves to $\sim 40$ MeV if hadronic decays of $Z$ are included. The spin and parity of the Higgs particle can be determined by the threshold excitation curve and the angular distribution of the Higgs production in the Higgs-strahlung process. If the rise of the cross section just above the threshold is $\sigma \propto \beta_H$, the $ZH$ pair is in a S-wave. Then the parity conservation in $Z^* \rightarrow ZH$ indicates that the parity of Higgs is plus. At all levels above threshold, $Z$ in the final state is mostly helicity 0. Since the intermediate $Z^*$ is polarized along the beam direction, the angular distribution of spin-0 Higgs is given by $|d\sigma_{HZZ}^{\mu\nu}(\theta)|^2 \propto \sin^2 \theta$. The spin parity of Higgs can also be checked in $e^+e^- \rightarrow ZH \rightarrow f\bar{f}f\bar{f}$ or in $H \rightarrow WW^\ast$, $ZZ\rightarrow f\bar{f}f\bar{f}$ where $f$ stands for a fermion. One can also study the spin correlation of the final state $\tau^\pm \tau^\mp$ to extract the $CP$ of Higgs.

The Higgs-strahlung process allows one to measure the $ZZH$ coupling independently of the Higgs decay modes. On the other hand, the $WW$ fusion process gives the $WWH$ coupling. At low CM energy, the $WW$ fusion process $e^+e^- \rightarrow \nu\bar{\nu}H$ has a substantial background coming from $e^+e^- \rightarrow ZH$, $Z \rightarrow \nu\bar{\nu}$ which can be removed by looking at the recoil mass of Higgs. Also, the $WW$ fusion process can be turned off and on by switching the beam polarizations to identify the contribution from the $WW$ fusion process. The $WWH$ coupling can also be extracted from the $H \rightarrow WW^\ast$ branching fraction. The statistical errors on $WWH$ and $ZZH$ couplings for $m_H$ of 120 GeV are 1–2%.

For the Higgs mass below 150 GeV, the couplings of Higgs to $b$, $c$, and $\tau$ are measured by reconstructing the Higgs decays to $b\bar{b}$, $c\bar{c}$, and $\tau^+\tau^-$ in the Higgs-strahlung process. Here, the branching ratios are proportional to the square of the fermion mass, and the excellent vertexing capability of ILC detectors is essential in separating $c\bar{c}$ from $b\bar{b}$.

The $t\bar{t}H$ Yukawa coupling is measured by $e^+e^- \rightarrow t\bar{t} \rightarrow t\bar{t}H$ at 1 TeV. The process $e^+e^- \rightarrow t\bar{t}$ is itself sensitive to the $t\bar{t}H$ coupling through the $H$-loop vertex correction. The gluonic decay $H \rightarrow gg$ as well as the decays $H \rightarrow \gamma\gamma$, $\gamma Z$ are sensitive to the $t\bar{t}H$ coupling though top loop, and also sensitive to new heavy particles that may contribute in the loop. For high Higgs masses, the gauge boson pair final states dominate. Still, with 1 fb$^{-1}$ at 1 TeV, the $b\bar{b}$ branching fraction can be measured to 12 and 28% for $m_H = 180$ and 220 GeV, respectively. Invisible final state can also be found by the recoil mass technique, with 5% confidence down to 2% branching fraction for $120 < m_H < 160$ GeV.

The total Higgs width for $m_H$ less than $\sim$200 GeV is too narrow to be measured directly, but can be indirectly measured by $\Gamma_H = \Gamma(H \rightarrow WW^\ast) / Br(H \rightarrow WW^\ast)$ where $Br(H \rightarrow WW^\ast)$ is directly measured and $\Gamma(H \rightarrow WW^\ast)$ is estimated from the measurement of the $WWH$ coupling by, say, the $WW$ fusion process. For $120 < m_H < 160$ GeV, the total Higgs width can be measured with an error of 4 to 13%.

The trilinear Higgs coupling, or the Higgs self coupling, can be measured by $e^+e^- \rightarrow ZZH \rightarrow ZHH$ or by $e^+e^- \rightarrow \nu\bar{\nu}H \rightarrow f\bar{f}f\bar{f}$. The cross section is quite small and the final state $(bb)(bb)(\ell^+\ell^-)$ challenges the capability of detector. Here, a superb vertexing resolution is critical for the $b$...
tagging, and an excellent jet energy reconstruction is needed for calculating the invariant masses of $b$ jet pairs. With 1 ab$^{-1}$ at 500 GeV CM energy and for $m_H = 200$ GeV, the error on the Higgs self coupling constant $\lambda_{HHH}$ is estimated to be about 20% using the $e^+e^- \rightarrow ZH \rightarrow ZHH$ mode only.\(^{24}\) If one combines $e^+e^- \rightarrow ZHH$ and $e^+e^- \rightarrow \nu\bar{\nu}HH$ at 1 TeV CM energy, the error in $\lambda_{HHH}$ becomes 12% for the same Higgs mass with 1 ab$^{-1}$ and 80% electron polarization.\(^{25}\)

Expected results for Higgs coupling measurements are plotted in Fig. 7 as functions of mass of the particle that Higgs couples to. Coupling constants of Higgs to fermions, weak bosons $W$ and $Z$, and Higgs itself are given by $m_f/a$, $g m_W$, $g m_Z/2\cos\theta_W$, and $m^2_H/2a$, respectively, where $g \sim 0.65$ is the $SU(2)$ coupling constant, and $a \sim 246$ GeV is the vacuum expectation value of Higgs. Thus, when properly normalized, the Higgs couplings of the standard model should be proportional to the mass of the particle it couples to. The pattern of deviation from the standard model serves as a powerful probe of the mechanism of mass generation. For example, for a two-Higgs-doublet model where up-type fermion masses are generated by one doublet and down-type fermion masses by another (so-called Type-II two-Higgs-doublet models), the Higgs couplings to all the up-type fermions are shifted by a factor, and those to all the down-type fermions are shifted by another factor. And in models with Radion-Higgs mixing, the Higgs couplings may be reduced uniformly with respect to the standard model values.

4. New Physics Particles

Among the extensions of the standard model, the SUSY models occupy a special place due to their theoretical virtues, the primary one of which is to make the Higgs mass stable in the weak scale. There are also other models that address the same problem, and these models usually contain particles that do not appear in the standard model. One should keep in mind, however, that Nature may have in store for us something that have nothing to do with any of these, and we may be lucky enough to encounter them at LHC/ILC.

4.1 SUSY particles

The MSSM is the most economical model with $R$-parity conservation which makes the lightest superparticle (LSP) stable. The LSP thus becomes a candidate for the dark matter. The two complex Higgs doublets and the four massless gauge bosons have 8 charged degrees of freedom and 8 neutral degrees of freedom. After breaking of SUSY and gauge symmetries, their super partners mix to form two chargeinos $\chi^0_{1,2}$ (8 degrees of freedom) and four neutralinos $\chi^0_{1,2,3,4}$ (8 degrees of freedom) all with spin 1/2. The neutralinos are self-conjugate; namely, they are Majorana particles. For each fermion $f$, there are two spin-0 superpartners corresponding to two helicities of the fermion: $f_R$ and $f_L$ which could in general mix, particularly for the third generation fermions. Since the actual masses of each particle and its super-partner are clearly different, the supersymmetry is broken by some mechanism. One popular model is a minimal model with gravity-mediated SUSY breaking (mSUGRA) in which there are only four free parameters and a sign, which may be taken as the mass parameters of scalers and winos: $m_0$ and $M_2$, the trilinear Higgs coupling $A_0$, the ratio of vacuum expectation values of the two Higgs doublets, tan $\beta$ and sign($\mu$) where $\mu$ is a Higgs mass parameter. For concreteness, we look for SUSY particles in this section with mSUGRA as a guide.

In many scenarios of SUSY, the super-partners of leptons (sleptons) are light enough to be produced at ILC. In addition, they tend to decay to the corresponding lepton plus the LSP neutralino. In the scenario called SPS1a of mSUGRA, all sleptons decay dominantly as $\tilde{e} \rightarrow \ell\tilde{\chi}^0_1$, where $\ell$ is a lepton and $\tilde{\ell}$ is its super-partner. Decays and interactions of right-handed sleptons are particularly simple since they are $SU(2)_L$ singlets and thus do not interact with $SU(2)_R$ gauge particles. Figure 8 demonstrates simultaneous mass determination of the right-handed smuon $\tilde{\mu}_R$ and the LSP neutralino $\tilde{\chi}^0_1$ in $e^+e^- \rightarrow \tilde{\mu}_R \tilde{\mu}_R$ followed by $\tilde{\mu}_R \rightarrow \mu^+ \tilde{\chi}^0_1$ and its charge conjugate mode. The data is taken well above the threshold with 100 fb$^{-1}$ at 350 GeV CM energy. The smuon and the LSP masses are assumed to be 142 and 118 GeV, respectively. The high and low end points of the
muon energy distribution gives both masses to a few $\times 10^{-3}$ of themselves. This is in contrast to the LHC case where the mass of LSP is difficult to measure directly. This mode also illustrates the effectiveness of beam polarization in background reduction. The muon acoplanarity distribution in $e^+e^- \rightarrow \mu_R \bar{\mu}_R$, $\mu_+^+ \bar{\mu}_R$ is shown in Fig. 9 for no electron polarization and with 90% electron polarization. Here, the acoplanarity angle is the angle between the muon pair projected to a plane perpendicular to the beam line. By polarizing the electron right-handedly, one can eliminate the background caused by $e^+e^- \rightarrow W^+W^-$, $W^+ \rightarrow \mu^+\nu$ and its charge conjugate. This is because for the s-channel the initial state $\bar{e}_R$ limits the intermediate state to $B$ (the gauge boson of hypercharge $Y$) which does not couple to $W$ in the final state, and the t-channel neutrino exchange is a $V-A$ interaction which does not couple to $e_R$.

The angular distribution of the smuon production should be $\sin^2 \theta$ since smuon is spin 0 and the intermediate $Z/\gamma$ state is polarized as $|1, \pm 1|$ along the beam line since the electron coupling to the intermediate state is a linear combination of vector and axial vector. The production angle can be reconstructed with a quadratic ambiguity where the wrong solution has a flat distribution that can be subtracted. The resulting angular distribution can be checked to be consistent with the expected shape.

The smuon mass can also be determined at the threshold, where an energy scan gives the threshold excitation curve which should rise slowly as $\beta^2_{\mu}$ due to the $P$ wave nature of the smuon pair.

A large mixing effect is expected for the stau sector and $\tilde{\tau}_R$ and $\tilde{\tau}_L$ would mix to form mass eigenstates $\tilde{\tau}_1$ and $\tilde{\tau}_2$ where $\tilde{\tau}_1$ is defined to be the lighter of the two. The mixing angle can be determined by two or more measurements of $e^+e^- \rightarrow \tilde{\tau}_1^\pm \tilde{\tau}_1^\mp$ with different beam polarizations. In the SPS1a scenario mentioned earlier, $\tilde{\tau}_1$ is the lightest of the sleptons with its mass around 100 GeV, and the dominant decay is $\tilde{\tau}_1 \rightarrow \mu \chi_1^0$. In this case, the mixing angle ($\cos 2\theta$) can be determined at the percent level.

The situation for the chargino pair production is similar to that of smuon pair production: $e^+e^- \rightarrow \chi_1^\pm \chi_1^- \bar{\chi}_1^\pm \bar{\chi}_1^-$ followed by $\chi_1^\pm \rightarrow W^\pm \chi_1^\mp$, where the energy distribution of $W^\pm$ simultaneously determines the masses of the chargino $\chi_1^\pm$ and the LSP neutralino. With the mass of the LSP obtained in the smuon study, the mass of the lightest chargino $\chi_1^\pm$ can be determined to 1% level.

For neutralinos, the invariant mass distribution of the lepton pair in $e^+e^- \rightarrow \chi_1^0 \chi_1^0$ followed by $\chi_2^0 \rightarrow \ell^+\ell^-$ can determine the mass difference between $\chi_2^0$ and $\chi_1^0$ to better than 1%. This mode may also demonstrate a sizable CP violation effects for some parameter space of MSSM. For example, the sign asymmetry of the $T$-odd triple product $p_{e^-} \times p_{e^+}$ can be as large as 20%. Similar $T$-odd triple products can be formed for other modes such as $e^+e^- \rightarrow \chi_1^+ \chi_1^-$. The ability to select the beam polarization allows us to probe into the structures of the SUSY models. For example, the charginos $\chi_{1,2}^\pm$ are the mass eigenstates of the system composed of the charged Higgsinos ($H_u^\pm, H_d^\pm$) and charged gauginos ($W^\pm$) where the mass matrix term can be written as

$$\left( \begin{array}{c} W^+ \\ H_u^- \end{array} \right) = \left( \begin{array}{cc} M_2 & \sqrt{2m_W} \cos \beta \\ \sqrt{2m_W} \sin \beta & \mu \end{array} \right) \left( \begin{array}{c} W^- \\ H_d^- \end{array} \right).$$

By using a right-handed electron beam for $e^+e^- \rightarrow \chi_1^+ \chi_1^-$, the intermediate s-channel state is purely $B$ which is the gauge boson for hypercharge $Y$. On the other hand, $B$ couples only to the Higgsino component of chargino; thus, one can obtain information on the mixing parameters of the charginos. Together with cross section measurements of $e^+e^- \rightarrow \tilde{\nu}_R \tilde{\nu}_R$ which is sensitive to the mass parameter $M_1$ which is the mass parameter for Bino (superpartner of $B$), one can perform a global fit to the parameters $(M_1, M_2, \mu, \tan \beta)$. Figure 10 shows the result of the global fit. If the masses $M_1$ and $M_2$ are to converge to a single value at the GUT scale, they would satisfy the GUT relation

$$M_1 = \frac{5}{3} \tan^2 \theta_w M_2,$$

which is tested in a highly model independent way.

In most SUSY scenarios, squarks are in general heavier than sleptons and many of them are beyond the reach of ILC even with the energy upgrade to 1 TeV. Still, due to the large mixing effects expected for the third generation squarks $\tilde{t}$ and $\tilde{b}$, the lighter ones, $\tilde{t}_1$ and $\tilde{b}_1$, can be within reach of ILC. When they can be pair produced as in

Fig. 9. (Color online) The muon acoplanarity distributions for the smuon production $e^+e^- \rightarrow \mu_R \bar{\mu}_R$, $\mu_+^+ \bar{\mu}_R$ with no electron polarization (left) and with 90% electron right-handed polarization.
$e^+e^- \rightarrow i_1i_1$, then multiple measurements of cross sections with different beam polarizations can determine the mixing angle just as in the case of the stau pair production with a similar precision.

4.2 KK mode gravitons

In the models with large extra dimensions where only gravitons can propagate in the extra dimensions, the fundamental gravity mass scale $M_D$ can be as small as the TeV scale. When the wave function of the graviton has a certain number of nodes in the direction of the extra dimensions (KK modes), it can have mass as a function of the number of nodes. When the number of the extra dimension $\delta$ is 2 to 6, the size of extra dimension can be very large and is around 0.1 mm to 1 fm, for which the KK mode graviton $G_{KK}$ has effectively a continuous mass spectrum. At ILC, one may search for emission of KK mode graviton in $e^+e^- \rightarrow \gamma G_{KK}$ where $G_{KK}$ escapes the detector and appear as missing energy. Here again, the beam polarization is a powerful handle to suppress the main background $e^+e^- \rightarrow \nu\bar{\nu}\gamma$. With 1 ab$^{-1}$ at 800 GeV and with the electron and positron beam polarizations of 80 and 60% respectively, the 95% confidence level lower limit of $M_D$ is 10 (3) TeV for $\delta$ of 2 (6). This is similar to the sensitivities at LHC. At ILC, however, one can utilize the angular distribution of $\gamma$ to verify the spin of $G_{KK}$ which should be two. In addition, the number of extra dimension $\delta$ can be measured at ILC by the energy dependence of the cross section, say at 500 GeV vs at 800 GeV, and the missing mass distribution.

4.3 Little Higgs models

In the Little Higgs models, the Higgs particle is composite, and there exist extra gauge bosons and top partners. Most new particles are too heavy to be directly detected at ILC, but indirect search for extra gauge bosons is possible with $e^+e^- \rightarrow f\bar{f}$ as described earlier (Fig. 2). Furthermore, in the model with T-parity, there could be a pseudo-scalar $\eta$ below 1 TeV. In such cases, $e^+e^- \rightarrow ZHH$ can be substantially enhanced by $ZH\eta$ coupling; $e^+e^- \rightarrow Z^+ \rightarrow \eta^+H$, $\eta^+ \rightarrow ZH$ which should be easily detectable with the TeV upgrade of the machine.

4.4 Cosmological connections

The WMAP satellite data indicates that the cold dark matter density of the universe is given by $\Omega_{DM}h^2 = 0.113 \pm 0.009$ and makes up about 1/4 of the energy of the universe. The error on $\Omega_{DM}$ will be reduced significantly by the Planck measurements expected around 2010. In the MSSM, the lightest neutralino $\chi^0_1$ serves as a candidate for the cold dark matter. In order to predict the relic density of the cold dark matter, however, all interactions contributing to $\chi^0_1$ annihilation should be known. Figure 11 shows the result of a study within the mSUGRA SPS1a scenario. The sensitivities of LHC and ILC in the two-dimensional space of $m_{\chi^0_1}$ and the estimated error on $\Omega_{DM}$ are shown together with the uncertainties on $\Omega_{DM}$ by WMAP and Planck. ILC can determine the mass of $\chi^0_1$ much more accurately than LHC, and the error on the estimate of $\Omega_{DM}$ is comparable to the error expected for the future measurement by Planck.

5. Options: $\gamma\gamma$ and $e^+e^-$ Colliders

The $e^+e^-$ mode of ILC can accommodate $e^+e^-$ and $\gamma\gamma$ colliders with relatively minor modifications. The $\gamma\gamma$ collider requires a pair of powerful lasers that are aimed at the interaction point from both sides along the beam line. The photons that are Compton back-scattered by incoming beams collide at the interaction point. The maximum CM energy of the $\gamma\gamma$ collision is only slightly lower than that of the $e^+e^-$ collision, and the luminosity is also comparable. The disrupted beams, however, need to be extracted without hitting the sensitive detector parts, and this necessitates a crossing angle greater than 25 mrad (compared to the nominal 14 mrad). Also, the original beams also collide on top of the $\gamma\gamma$ collisions, and this favors the $e^+e^-$ mode over the $e^+e^-$ mode which has larger total cross section. The $e^+e^-$ is suited for the $\gamma\gamma$ collision also because it is easier to produce polarized electrons than polarized positrons.

The Higgs particle can be produced by the s-channel $\gamma\gamma \rightarrow H$ process which involves loop diagrams of charged particles. It allows a precision measurement of the Higgs coupling to photon and is sensitive to new particles that can contribute in the loop. The Higgs mass reach is close to the
CM energy of the $e^+e^-$ beams itself. Higgs below 140 GeV would be detected in the $b\bar{b}$ final state. With 410 fb of $\gamma\gamma$ luminosity at the beam CM energy of 210 GeV, and for $m_H = 120$ GeV, $\Gamma(H \rightarrow \gamma\gamma) \times Br(H \rightarrow b\bar{b})$ can be determined to a statistical error of 2%.\textsuperscript{29} Even for heavy Higgs of 200 to 350 GeV, the two photon width can be determined with errors of 3 to 10%. The total Higgs decay width can be obtained by combining the $\Gamma(H \rightarrow \gamma\gamma)$ measurement with $Br(H \rightarrow \gamma\gamma)$ measured at the $e^+e^-$ collider at high CM energy. The expected error on the Higgs total width is about $m_H = 140$ to 200 GeV, the two photon width can be determined with errors of 3 to 10%. The total Higgs decay width can be obtained by combining the $\Gamma(H \rightarrow \gamma\gamma)$ measurement with $Br(H \rightarrow \gamma\gamma)$ measured at the $e^+e^-$ collider at high CM energy. The expected error on the Higgs total width is about $m_H = 120$ to 140 GeV, which is competitive with the method using $\Gamma_H = \Gamma(H \rightarrow WW^*)/Br(H \rightarrow WW^*)$ mentioned earlier.

The $e^+e^-$ collider can generate exotic charge two particles in s-channel. It is also sensitive to Majorana neutrino exchange in $e^+e^- \rightarrow W^+W^-$. The neutralino exchange interactions $e^-e^- \rightarrow \nu_L\bar{\nu}_L$ allows one to study the quantum numbers of selectrons through beam polarizations.

6. Summary

The clean environment and the well-defined initial state of $e^+e^-$ collision, including the spin states, as well as the superb resolutions of the ILC detectors make the ILC physics program very attractive. ILC can study the particles found at LHC in detail to uncover the underlying theoretical structures, and in some cases discover new particles and reactions that are buried in backgrounds at LHC.

Acknowledgment

The author would like to thank the editorial panel of the detector concept report (DCR) who has put together an excellent summary of the ILC physics studies performed so far, and all those who have contributed to these studies. He would like to thank, in particular, Professor Komamiya who has entrusted him to write this article. This work is supported in part by a Grant-in-Aid for Scientific Research 18GS0202 from the Japan Society for the Promotion of Science.

Hitoshi Yamamoto was born in Osaka Prefecture, Japan in 1955. He obtained his B. Sc. (1978) from Kyoto University, and Ph. D. (1985) from California Institute of Technology. He was a research associate at Stanford Linear Accelerator Center (1985–1986), a research associate (1986–1989) and a senior research associate (1989–1990) at Enrico Fermi Institute of University of Chicago, an assistant professor (1991–1993) and an associate professor (1993–1998) at Harvard University, and a professor at University of Hawaii (1998–2001). He has been a professor at Tohoku University since 2001. He has worked extensively on physics of electron–positron storage rings including B factories. His studies have also covered CP violation in neutral K meson system.

2) See, for example, S. Martin: hep-ph/9709356.
5) WWS DCR panel, the detector concept report (DCR) (http://www.linearcollider.org/wiki/doku.php).
12) W. Menges: LC-PHSM-2001-022.
14) S. Riemann: LC-TH-2001-007.
19) The LEP collaborations and the LEP electroweak working group: hep-ex/0612034.
Cosmic Acceleration, Dark Energy, and Fundamental Physics

Michael S. Turner and Dragan Huterer

1Kavli Institute for Cosmological Physics, The University of Chicago, Chicago, IL 60637, U.S.A.
2Department of Astronomy and Astrophysics, The University of Chicago, Chicago, IL 60637, U.S.A.
3Department of Physics, The University of Chicago, Chicago, IL 60637, U.S.A.

(Received April 17, 2007; accepted May 14, 2007; published November 12, 2007)

A web of interlocking observations has established that the expansion of the Universe is speeding up and not slowing, revealing the presence of some form of repulsive gravity. Within the context of general relativity the cause of cosmic acceleration is a highly elastic \( (p \sim -\rho) \), very smooth form of energy called “dark energy” accounting for about 75% of the Universe. The “simplest” explanation for dark energy is the zero-point energy density associated with the quantum vacuum; however, all estimates for its value are many orders-of-magnitude too large. Other ideas for dark energy include a very light scalar field or a tangled network of topological defects. An alternate explanation invokes gravitational physics beyond general relativity. Observations and experiments underway and more precise cosmological measurements and laboratory experiments planned for the next decade will test whether or not dark energy is the quantum energy of the vacuum or something more exotic, and whether or not general relativity can self consistently explain cosmic acceleration. Dark energy is the most conspicuous example of physics beyond the standard model and perhaps the most profound mystery in all of science.

KEYWORDS: cosmology, gravitation, early universe, cosmological parameters
DOI: 10.1143/JPSJ.76.111015

1. Quarks and the Cosmos

The final 25 years of the 20th century saw the rise of two highly successful mathematical models that describe the Universe at its two extremes, the very big and the very small. The standard model of particle physics (detailed in this volume) provides a fundamental description of almost all phenomena in the microscopic world. The standard hot big bang model describes in detail the evolution of the Universe from a fraction of a second after the beginning, when it was just a hot soup of elementary particles, to the present some 13.7 billion years later when it is filled with stars, planets, galaxies, clusters of galaxies and us. Both standard models are consistent with an enormous body of precision data, gathered from high-energy particle accelerators, telescopes and laboratory experiments. The standard model of particle physics and the hot big bang cosmology surely rank among the most important achievements of 20th century science (see Fig. 1).

Both models raise profound questions. Moreover, the “big questions” about the very small and the very large are connected, both in their asking and ultimately in their answering. This suggests that the deeper understanding that lies ahead will reveal even more profound connections between the quarks and the cosmos. The big questions include:

- How are the forces and particles of nature unified?
- What is the origin of space, time and the Universe?
- How are quantum mechanics and general relativity reconciled?
- How did the baryonic matter arise in the Universe?
- What is the destiny of the Universe?
- What is the nature of the dark matter that holds the Universe together and of the dark energy that is causing the expansion of the Universe to speed up?

The last question illustrates the richness of the connections between quarks and the cosmos: 96% of the matter and energy that comprises the Universe is still of unknown form, is crucial to its existence, and determines its destiny. Dark matter and dark energy are also the most concrete and possibly most important evidence for new physics beyond the standard model of particle physics.

The solution to the dark matter problem seems within reach: we have a compelling hypothesis, namely that it exists in the form of stable elementary particles left over from the big bang; we know that a small amount of dark matter exists in the form of massive neutrinos; we have two good candidates for the rest of it (the axion and neutralino) and an experimental program to test the particle dark matter hypothesis.

The situation with cosmic acceleration and dark energy is very different. While we have compelling evidence that the expansion of the Universe is speeding up, we are far from a working hypothesis or any significant understanding of cosmic acceleration. The solution to this profound mystery could be around the corner or very far away.

2. Evidence for Cosmic Acceleration

2.1 Cosmology basics

For mathematical simplicity Einstein assumed that the Universe is isotropic and homogeneous; today, we have good evidence that this is the case on scales greater than 100 Mpc (from the distribution of galaxies in the Universe) and that it was at early times on all scales (from the uniformity of the cosmic microwave background). Under this assumption, the expansion is merely a rescaling and is described by a single function, the cosmic scale factor, \( R(t) \). (By convention, the value of the scale factor today is set equal to 1.) The wavelengths of photons moving through the Universe scale...
with $R(t)$, and the redshift that light from a distant object suffers, $1 + z = t_{\text{emitted}}/t_{\text{emitted}}$, directly reveals the size of the Universe when that light was emitted: $1 + z = 1/R(t_{\text{emitted}})$.

The key equations of cosmology are

$$H^2 \equiv \left(\frac{\dot{R}}{R}\right)^2 = \frac{8\pi G\rho}{3} - \frac{k}{R^2} + \frac{\Lambda}{3},$$

(1)

$$\frac{\ddot{R}}{R} = -\frac{4\pi G}{3}(\rho + 3p) + \frac{\Lambda}{3},$$

(2)

$$\omega_i \equiv \frac{\rho_i}{\rho} \propto (1 + z)^{3(1+w_i)},$$

(3)

$$q(z) \equiv -\frac{\ddot{R}}{RHF^2} = \frac{1}{2}(1 + 3w),$$

(4)

where $\rho$ is the total energy density of the Universe (sum of matter, radiation, dark energy) and $p$ is the total pressure. For each component the ratio of pressure to energy density is the equation-of-state $w_i$, which, through the conservation of energy, $d(R^2\rho) = -pdR^3$, determines how the energy density evolves. For constant $w$, $\rho \propto (1 + z)^{3(1+w)}$. For matter ($w = 0$) $\rho_M \propto (1 + z)^3$ and for radiation ($w = 1/3$) $\rho_R \propto (1 + z)^3$. The first of these equations, known as the Friedmann equation, is the master equation of cosmology.

The quantity $k$ is the 3-curvature of the Universe and $R_{\text{curv}} = R/\sqrt{|k|}$ is the curvature radius; $k = 0$ corresponds to a spatially flat Universe, $k > 0$ a positively curved Universe and $k < 0$ a negatively curved Universe. Because of the evidence from the cosmic microwave background that the Universe is spatially flat (see Fig. 1), unless otherwise noted we shall assume $k = 0$.

$\Lambda$ is Einstein’s infamous cosmological constant; it is equivalent to a constant energy density, $\rho_\Lambda = \Lambda/8\pi G$, with pressure $p_\Lambda = -\rho_\Lambda$ ($w = -1$). The quantity $q(z)$ is the deceleration parameter, defined with a minus sign so that $q > 0$ corresponds to decelerating expansion.

The energy density of a flat Universe ($k = 0$), $\rho_C = 3H^2/8\pi G$, is known as the critical density. For a positively curved Universe, $\Omega_{\text{TOT}} \equiv \rho/\rho_C > 1$ and for a negatively curved Universe $\Omega_{\text{TOT}} < 1$. Provided the total pressure is greater than $-1/3$ times the total energy density, gravity slows the expansion rate, i.e., $\ddot{R} < 0$ and $q > 0$. Because of the $(\rho + 3p)$ term in the $\dot{R}$ equation (Newtonian gravity would only have $\rho$), the gravity of a sufficiently elastic form of energy ($p < -\rho/3$) is repulsive and causes the expansion of the Universe to accelerate. In Einstein’s static solution ($H = 0$, $q = 0$) the repulsive gravity of $\Lambda$ is balanced against the attractive gravity of matter, with $\rho_\Lambda = \rho_M/2$ and $R_{\text{curv}} = 1/\sqrt{8\pi G\rho_M}$. A cosmological constant that is larger than this results in accelerated expansion ($q < 0$); the observed acceleration requires $\rho_\Lambda \simeq (2-3)\rho_M$.

For an object of known intrinsic luminosity $L$, the measured energy flux $F$ defines the luminosity distance $d_L$ to the object (i.e., the distance inferred from the inverse square law). The luminosity distance is related to the cosmological model via

$$d_L(z) = \sqrt{L/4\pi F} = (1 + z)\int_0^z \frac{dz'}{H(z')}.$$

(5)

Astronomers determine the luminosity distance from the difference between the apparent magnitude $m$ of the object (proportional to the log of the flux) and the absolute magnitude $M$ (proportional to the log of the intrinsic luminosity), $m - M = 5\log_{10}(d_L/10\text{pc})$ (where 5 astronomical magnitudes correspond to a factor of 100 in flux or a factor of 10 in luminosity distance).

The use of “standard candles” (objects of known intrinsic luminosity $L$) and measurements of the energy flux $F$ constrain the cosmological model through this equation. In particular, the Hubble diagram (or magnitude-redshift diagram) is the simplest route to probing the expansion history. In terms of the deceleration parameter the equation is deceptively simple:

$$H_0d_L = z + \frac{1}{2}(1 - q_0)z^2 + \cdots$$

(6)

where the subscript “0” denotes the value today. While this Taylor expansion of eq. (5), valid for $z \ll 1$, is of historical significance and utility, it is not useful today since objects as distant as redshift $z \sim 2$ have been used to probe the expansion history. However, it does illustrate the general principle: the first term on the r.h.s. represents the linear Hubble expansion, and the deviation from a linear relation reveals the deceleration (or acceleration).

2.2 $\Lambda$’s checkered history

Before discussing the evidence for cosmic acceleration, we will recount some of the history of the cosmological constant. Realizing that there was nothing to forbid such a term and that it could be used to obtain an interesting solution (a static and finite Universe), Einstein introduced the cosmological constant in 1917. While his static solution was consistent with astronomical observations at that time, Hubble’s discovery of the expansion of the Universe in 1929 led Einstein to discard the cosmological constant in favor of expanding models without one, calling the cosmological constant “my greatest blunder”.

Fig. 1. (Color online) Multipole power spectrum of the CMB temperature fluctuations from WMAP and other CMB anisotropy experiments. Position of the first peak at $l \sim 200$ indicates the flatness of the Universe; height of the first peak determines the matter density, and the ratio of the first to second peaks determines the baryon density. Together with SDSS large-scale structure data, the CMB measurements have determined the shape and composition of the Universe: $\Omega_{\text{TOT}} = 1.003 \pm 0.015$, $\Omega_M = 0.24 \pm 0.02$, $\Omega_b = 0.042 \pm 0.002$, and $\Omega_c = 0.76 \pm 0.02$. The curve is the theoretical prediction of the “concordance cosmology”, with a band that indicates cosmic variance. Figure adopted from ref. 19.
In 1948, Bondi, Gold, and Hoyle put forth the “steady state cosmology”, with $\rho_A > 0$ and $\rho_M \approx 0$. The model was motivated by the aesthetics of an unchanging universe and a serious age problem (the measured value of the Hubble constant at the time, around 500 km s$^{-1}$ Mpc$^{-1}$ implied an expansion age of only 2 Gyr, less than the age of Earth). The redshift distribution of radio galaxies, the absence of quasars nearby and the discovery of the cosmic microwave background radiation in 1960s all indicated that we do not live in an unchanging Universe and ended this revival of a cosmological constant.

The cosmological constant was briefly resurrected in the late 1960s by Petrosian et al.\textsuperscript{3} to explain the preponderance of quasars at redshifts around $z \sim 2$ (as it turns out, this is a real effect: quasar activity peaks around $z \sim 2$). In 1975 weak evidence for a cosmological constant from a Hubble diagram of elliptical galaxies extending to redshifts of $z \sim 0.5$ was presented.\textsuperscript{4} Significant concerns about whether or not elliptical galaxies were good standard candles led to the demise of $\Lambda$ once again. Shortly thereafter came the rise of the standard cosmology with $\Lambda = 0$.

The current attempt at introducing a cosmological constant (or something similar), which is backed up by multiple lines of independent evidence, traces its roots to the inflationary universe scenario and its prediction of a spatially flat Universe. In the early 1980s when inflation was introduced, the best estimate of the average mass density fell short of the critical density by almost a factor of 10 ($\Omega_M \sim 0.1$); the saving grace for inflation was the large uncertainty associated with measuring the mean matter density. From 1980 to the mid 1990s, as measurement techniques took better account of dark matter, $\Omega_M$ rose to of order 0.5 or so. However, as the uncertainties got smaller, $\Omega_M$ began converging on a value of around 1/3, not 1. Moreover, the predictions of the cold dark matter scenario of structure formation matched observations if $\Omega_M$ was around 1/3, not 1.

Starting in 1984 and continuing to just before the discovery of cosmic acceleration, a number of papers suggested the solution to inflation’s “$\Omega$ problem”\textsuperscript{1} was a cosmological constant.\textsuperscript{5} Owing to its checkered history, there was not much enthusiasm for this suggestion at first. However, with time the indirect evidence for $\Lambda$ grew,\textsuperscript{6-8} and in 1998 when the supernova evidence for accelerated expansion was presented the cosmological constant was quickly embraced—this time, it was the missing piece of the puzzle that made everything work.

2.3 Discovery and confirmation

Two breakthroughs enabled the discovery that the Universe is speeding up and not slowing down. The first was the demonstration that type Ia supernovae (SNe Ia), the brightest of the supernovae and the ones believed to be associated with the thermonuclear explosions of 1.4 $M_\odot$ white-dwarf stars pushed over the Chandrasekhar mass limit by accretion, are (nearly) standard candles.\textsuperscript{10} The second breakthrough involved the use of large (of order 100 megapixel) CCD cameras to search big regions of the sky containing thousands of galaxies for these rare events (the SN Ia rate in a typical galaxy is of the order of one per 100 to 200 years). By comparing images of thousands of galaxies taken weeks apart the discovery of SNe could be reliably “scheduled” on a statistical basis.

Two teams working independently in the mid- to late-1990s took advantage of these breakthroughs to determine the expansion history of the Universe. They both found that distant SNe are dimmer than they would be in a decelerating Universe, indicating that the expansion has actually been speeding up for the past 5 Gyr,\textsuperscript{11,12} see Fig. 2. Analyzed for a Universe with matter and cosmological constant, their results provide evidence for $\Omega_M > 0$ at greater than 99% confidence; see Fig. 3.

Since this work, the two teams have discovered and studied more SNe, as have other groups.\textsuperscript{13-16} Not only has the new data confirmed the discovery, but it has also allowed measurements of the equation-of-state of dark energy $w = p/\rho$ (assuming constant $\omega$), and even constrains the time variation of $w$, with the parametrization $w = w_0 + w_a (1 - t/R)$.

Especially important in this regard are SNe with redshifts $z > 1$ which indicate that the universe was decelerating at earlier times (see Fig. 4), and hence that dark energy started its domination over the dark matter only recently, at redshift $z = (\Omega_M/\Omega_{DE})^{1/2} - 1 \approx 0.5$. This finding is an important reality check: without a long, matter-dominated, slowing phase, the Universe could not have formed the structure we see today.

Evidence for dark energy comes from several other independent probes. Measurements of the fraction of X-ray emitting gas to total mass in galaxy clusters, $f_{\text{gas}}$, also
indicates the presence of dark energy. Because galaxy clusters are the largest collapsed objects in the universe, the gas fraction in them is presumed to be constant and equal to the overall baryon fraction in the universe, $\Omega_B/\Omega_M$ (most of the baryons in clusters reside in the gas). Measurements of the gas fraction $f_{\text{gas}}$ depend not only on the observed X-ray flux, but also on the distance to the cluster; therefore, only the correct cosmology will produce distances which make the apparent $f_{\text{gas}}$ constant in redshift. Using data from the Chandra X-ray Observatory, Allen et al. have determined $\Omega_\Lambda$ to an accuracy of about $\pm0.2$; see Fig. 3.

Cosmic microwave background (CMB) anisotropies provide a record of the Universe at simpler time, before structure had developed and when photons were decoupling from baryons, $z \approx 1100$. The multipole power spectrum is dominated by the acoustic peaks that arise from gravitationally driven photon–baryon oscillations (see Fig. 1). The positions and amplitudes of the acoustic peaks encode much information about the Universe, today and at earlier times. In particular, they indicate that the Universe is spatially flat, with a matter density that accounts for only about a quarter of the critical density. However, the presence of a uniformly distributed energy density with large negative pressure which accounts for three-quarters of the critical density brings everything into good agreement, both with CMB data and the large-scale distribution of galaxies in the Universe. The CMB data of WMAP together with large-scale structure data of the Sloan Digital Sky Survey (SDSS) provides the following cosmic census:\cite{WMAP10} $\Omega_{\text{TOT}} = 1.003 \pm 0.010$, $\Omega_B = 0.24 \pm 0.02$, $\Omega_M = 0.042 \pm 0.002$, and $\Omega_\Lambda = 0.76 \pm 0.02$.

The presence of dark energy also affects the large-angle anisotropy of the CMB (the low multipoles) and leads to the prediction of a small correlation between the galaxy distribution and the CMB anisotropy. This subtle effect has been observed\cite{SDSS-BAO}, while not detected at a level of significance that could be called independent confirmation, its presence is a reassuring cross check.

Baryon acoustic oscillations (BAO), so prominent in the CMB anisotropy (see Fig. 1), leave a smaller characteristic signature in the clustering of galaxies that can be measured today and provide an independent geometric probe of dark energy. Measurements of the BAO signature in the correlation function of SDSS galaxies constrains the distance to redshift $z = 0.35$ to a precision of 5\%. While this alone does not establish the existence of dark energy, it serves as a significant complement to other probes, cf. Fig. 5.

Weak gravitational lensing\cite{WeakLensing} — slight distortions of galaxy shapes due to gravitational lensing by intervening large-scale structure — is a powerful technique for mapping dark matter and its clustering. Currently, weak lensing sheds light on dark energy by pinning down the combination $\sigma_8(\Omega_M/0.25)^{0.6} \approx 0.85 \pm 0.07$, where $\sigma_8$ is the amplitude of mass fluctuations on the 8 Mpc scale. Since other measurements put $\sigma_8$ at $\sim 0.9$, this implies that $\Omega_M \approx 0.25$, consistent with a flat Universe whose mass/energy density is dominated by dark energy. In the future, weak lensing will also be very useful in probing the equation-of-state of dark energy;\cite{SDSS-BAO} see §4.
Finally, because the time back to the big bang, \( t_0 = \int \frac{dz}{(1 + z)H(z)} \), depends upon the expansion history, the comparison of this age with other independent age estimates can be used to probe dark energy. The ages of the oldest stars in globular clusters constrain the age of the Universe: \( 11 \lesssim t_0 \lesssim 15 \) Gyr.\(^2\)\(^6\) CMB anisotropy is very sensitive to the expansion age, and WMAP data determine it accurately: \( t_0 = 13.84_{-0.30}^{+0.35} \) Gyr.\(^2\)\(^7\) Figure 6 shows that a consistent age is possible if \( -2 \lesssim w \lesssim -0.75 \). Agreement on the age of the Universe provides an important consistency check as well as confirmation of a key feature of dark energy, its large negative pressure.

3. Understanding Cosmic Acceleration

Sir Arthur Eddington is quoted as saying, “It is (also) a good rule not to put too much confidence in observational results until they are confirmed by theory”. While this may seem a bit paradoxical (or worse yet, an example of blatant theoretical arrogance), the point is well taken: science is not just a collection of facts, it is also understanding; if the understanding does not eventually follow new facts, perhaps there is something wrong with the facts.

Cosmic acceleration meets the Eddington criterion and at the same time presents a stunning opportunity for theorists: General relativity (GR) can accommodate accelerated expansion, but GR has yet to provide a deeper understanding of the phenomenon.

Within GR, a very elastic fluid has repulsive gravity, and, if present in sufficient quantity, can lead to the observed accelerated expansion. This then is the definition of dark energy: the mysterious, elastic and very smooth form of energy which is responsible for cosmic acceleration and is characterized by an equation-of-state \( w = p/\rho \sim -1 \).\(^8\)

Vacuum energy is a concrete example of dark energy. General covariance requires that the stress energy associated with the vacuum take the form of a constant times the metric tensor. This implies that it has a pressure equal to minus its energy density, is constant both in space and time, and is mathematically equivalent to a cosmological constant.

The stress energy associated with a homogeneous scalar field \( \phi \) can also be like dark energy. It takes the form of a perfect fluid with

\[
\rho = \frac{\dot{\phi}^2}{2} + V(\phi)
\]
\[
p = \frac{\dot{\phi}^2}{2} - V(\phi).
\]

where \( V(\phi) \) is the potential energy of the scalar field, \( \dot{\phi} \) denotes time derivative, and the evolution of the field \( \phi \) is governed by

\[
\ddot{\phi} + 3H\dot{\phi} + V(\phi) = 0.
\]

If the scalar field evolves slowly, that is \( \dot{\phi}^2 \ll V \), then \( p \approx -\rho \) and the scalar field behaves like a slowly varying vacuum energy.

While cosmic acceleration can be accommodated within the GR framework, the fundamental explanation could be new gravitational physics. With this as a prelude, we now briefly review the present theoretical situation.

(a) Vacuum energy. Vacuum energy is both the most plausible explanation and the most puzzling possibility. For almost 80 years we have known that there should in principle be an energy associated with the zero-point fluctuations of all quantum fields. Moreover, \( \rho_{\text{vac}} = -\rho_{\text{vac}} \). However, all attempts to compute the value of the vacuum energy lead to divergent results. The so-called cosmological constant problem was finally articulated about thirty years ago.\(^9\) However, because of the success of the standard hot big bang model (where \( \Lambda = 0 \) and the absence of good (or any) ideas, the problem was largely ignored. With the discovery of cosmic acceleration, the cosmological constant problem is now front and center and can no longer be ignored.

To be more quantitative, the energy density required to explain the accelerated expansion is about three quarters of the critical density or about \( 4 \times 10^{-47} \text{GeV}^4 \approx (3 \times 10^{-3} \text{eV})^4 \). This is tiny compared to energy scales in particle physics (with the exception of neutrino mass differences). Such a small energy precludes solving the problem by simply cutting off the divergent zero-point energy integral at some energy beyond which physics is not yet known. For example, a cutoff of 100 GeV would leave a 54 orders-of-magnitude discrepancy. If supersymmetry were an unbroken symmetry, fermionic and bosonic zero-point contributions would cancel. However, if supersymmetry is broken at a scale of order \( M \), one would expect that imperfect cancellations leave a finite vacuum energy of the order \( M^4 \), which for the favored value of \( M \sim 100 \text{ GeV} \) to \( 1 \text{ TeV} \), would leave a discrepancy of 50 or 60 orders-of-magnitude.

One approach to the cosmological constant problem involves the idea that the value of the vacuum energy is a random variable which can take on different values in different disconnected pieces of the Universe. Because a value much larger than needed to explain the observed cosmic acceleration would preclude the formation of galaxies, we could not find ourselves in such a region.\(^9\) This very anthropic approach finds a home in the landscape version of string theory.\(^9\)

(b) Scalar fields, etc. While introducing a new dynamical degree of freedom can also provide a very elastic form of energy density, it does not solve the cosmological constant problem. In order to roll slowly enough the mass

\[
\Omega_m = 1 - \Omega_{\text{DE}}
\]

![Graph](image-url)
of the scalar field must be very light, \( m \lesssim H_0 \sim 10^{-42} \text{ GeV} \), and its coupling to matter must be very weak to be consistent with searches for new long-range forces.\(^{31}\) Unlike vacuum energy, scalar-field energy clusters gravitationally, but only on the largest scales and with a very small amplitude.\(^{32}\)

The equation-of-state \( w \) for a scalar field can take on any value between \(-1\) and 1 and in general varies with time. (It is also possible to have \( w < -1 \), though at the expense of ghosts, by changing the sign of the kinetic energy term in the Lagrangian.) Scalar field models also raise new questions and possibilities: Is cosmic acceleration related to inflation? Is dark energy related to dark matter or neutrino mass? No firm or compelling connections have been made to either, although interesting possibilities have been suggested.

Scalar fields in a very different form can also explain cosmic acceleration. The topological solitons that arise in broken gauge theories, e.g., strings, walls, and textures, are very elastic, and tangled networks of such defects can on large scales behave like an elastic medium with \( w = -N/3 \), where \( N \) is the dimensionality of the network (\( N = 1 \) for strings, 2 for walls, and 3 for textures). In this case \( w \) is a fixed, rational number.

(c) Modified gravity. A very different approach holds that cosmic acceleration is a manifestation of new gravitational physics and not of dark energy. Assuming that our 4-d spacetime can still be described by a metric, the operational changes are twofold: (1) a new version of the Friedmann equation governing the evolution of the background spacetime; (2) modifications to the equations that govern the growth of the small matter perturbations that evolve into the structure seen in the Universe today. A number of ideas have been explored, from models motivated by higher-dimensional theories and string theory\(^{33,34}\) to generic modifications of the usual gravitational action.\(^{35}\)

An aside: One might be concerned that when the assumption of general relativity is dropped the evidence for accelerated expansion might disappear. This is not the case; using the deceleration \( q(z) \) as a kinematic description of the expansion, the SNe data still provide strong evidence for a period of accelerated expansion.\(^{36}\)

Changes to the Friedmann equation are easier to derive, discuss, and analyze. In order not to spoil the success of the standard cosmology at early times (from big-bang nucleosynthesis to the CMB anisotropy to the formation structure), the Friedmann equation must reduce to the GR form for \( z \gg 1 \). Because the matter term scales as \((1 + z)^3\) and the radiation term as \((1 + z)^4\), to be safe any modifications should decrease with redshift more slowly than this. As a specific example, consider the DGP model, which arises from a five-dimensional gravity theory,\(^{33}\) and has a 4-d Friedmann equation,

\[
H^2 = \frac{8\pi G \rho_M}{3} + \frac{H}{r_c},
\]

where \( r_c \) is an undetermined scale and \( \rho_M \) is the matter energy density. As \( \rho_M \to 0 \), there is a (self) accelerating solution, with \( H = 1/r_c \). The additional term in the Friedmann equation, \( H/r_c \), behaves just like dark energy with an equation-of-state that evolves from \( w = -1/2 \) (for \( z \gg 1 \)) to \( w = -1 \) in the distant future.

4. Prospects for Revealing the Nature of Dark Energy

We divide the probes of dark energy into three broad categories: kinematical and dynamical cosmological probes, and laboratory/astrophysical probes. Kinematical tests rely on the measurement of cosmological distances and volumes to constrain the evolution of the scale factor and thus the background cosmological model. Specific techniques include SNe Ia, CMB, and baryon acoustic oscillations.

The dynamical tests probe the effect of dark energy on perturbations of the cosmological model, including the evolution of the small inhomogeneities in the matter density that give rise to structure in the Universe. Specific techniques include the use of gravitational lensing to directly determine the evolution of structure in the dark matter and the study of the growth of the abundance of galaxy clusters to indirectly probe the growth of structure. A potential probe of dark energy, which at the present seems beyond reach, is to study the clustering of dark energy itself. Since vacuum energy does not cluster, detection of such would rule out vacuum energy as the explanation for cosmic acceleration.

In general relativity, for both kinematical and dynamical cosmological probes, the primary effect of dark energy enters through the Friedmann equation, cf. eq. (1),

\[
H(z)^2 = \frac{8\pi G}{3} \left[ \rho_M + \rho_{DE} \right] \\
= H_0^2 \left[ \Omega_M (1 + z)^3 + (1 - \Omega_M) (1 + z)^{3(w+1)} \right],
\]

where a flat Universe and constant \( w \) have been assumed. In turn, the expansion rate affects the luminosity distance, \( d_L = (1 + z) \int dz/H(z) \), the number of objects seen on the sky, \( d^2N/\text{d}\Omega \text{d}z = n(z)d_L^2/[1 + (1 + z)^3H(z)] \) \( n \) is the comoving density of objects, and the evolution of cosmic structure via the growth of small density perturbations. In GR the growth of small density perturbations in the matter, and on subhorizon scales, is governed by

\[
\ddot{\delta}_k + 2H \dot{\delta}_k - 4\pi G \rho_M \delta_k = 0,
\]

where density perturbations in the cold dark matter have been decomposed into their Fourier modes of wavenumber \( k \). Dark energy affects the growth through the “drag term”, \( 2H \delta_k \). The equations governing dark energy perturbations depend upon the specific dark energy model.

The kinematical and dynamical tests probe complementary aspects of the effect of dark energy on the Universe: the overall expansion of the Universe (kinematical) and the evolution of perturbations (dynamical). Together, they can test the consistency of the underlying gravity theory. In particular, different values of the dark energy equation-of-state obtained by the two methods would indicate an inconsistency of the underlying gravity theory.

The kinematical tests are easier to frame because they only depend upon knowing the effect of dark energy on the background cosmological model. Further, cosmological variables (such as \( q \)) can even be formulated without reference to a particular theory of gravity. The dynamical tests are both harder to frame—they require detailed knowledge of how dark energy clusters and affects the growth of density perturbations — and also harder to implement—they rely upon the details of describing and measuring the distribution of matter in the Universe. Both the
kinematical and dynamical tests have their greatest probative power at redshifts between about $z = 0.2$ and $z = 2$, for the simple reason that at higher redshifts dark energy becomes increasingly less important, $\rho_{\text{DE}}/\rho_{\text{M}} \propto (1 + z)^3w$. While the primary probes of dark energy are cosmological, laboratory experiments may be able to get at the underlying physics. If dark energy couples to matter there will be long-range forces that are in principle detectable; if it couples to electromagnetism, polarized light from distant astrophysical sources should suffer rotation. It is also possible that accelerator-based experiments will have something to say about dark energy. For example, if evidence for supersymmetry is found at the Large Hadron Collider, understanding how supersymmetry is broken could shed light on the vacuum energy puzzle.

Observations to date have established the existence of dark energy and have begun to probe its nature; e.g., by constraining $w = -1 \pm 0.1$. Future experiments will focus on testing whether or not it is vacuum energy and the consistency of GR to accommodate dark energy. The Supernova/Acceleration Probe (SNAP)\cite{38} proposed a space-based telescope to collect several thousand SNe out to $z \approx 2$, would significantly reduce uncertainties (both statistical and systematic) on dark-energy parameters. SNAP, together with the planned wide-field surveys from the ground, the Dark Energy Survey (DES)\cite{39} and the Large Synoptic Survey Telescope (LSST),\cite{40} would map the weak lensing signal from one arcminute out to the largest observable scales on the sky and accurately determine the effect of dark energy on the growth of structure. Large BAO surveys are also planned, both from the ground and space. The just-completed South Pole Telescope (SPT)\cite{41} and the Atacama Cosmology Telescope (ACT)\cite{42} will soon begin studying dark energy by determining the evolution of the abundance of galaxy clusters. In 2008, ESA’s Planck Surveyor CMB satellite\cite{43} will be launched and will extend precision measurements of CMB anisotropy to $l \sim 3000$ (i.e., down to angular scales of about 5 arcmin), more accurately pinning down the matter density and providing an important prior constraint for other dark energy measurements.\cite{44} Theoretical forecasts for future constraints on the parameters $(w_0, w_a)$ are shown in Fig. 7.

5. Dark Energy and Destiny

One of the first things one learns in cosmology is that geometry is destiny: a closed (positively curved) Universe eventually recollapses and an open (flat or negatively curved) Universe expands forever. Provided that the Universe only contains matter and $\Lambda = 0$, this follows directly from eq. (1). If $k > 0$, the Universe achieves a maximum size when $H^2$ is driven to zero by the inevitable cancellation of $\rho_M$ and $k/R^2$. If $k = 0$, $R$ always grows as $t^{2/3}$ and $q = 1/2$. For $k < 0$, the Universe ultimately reaches a coasting phase where $R$ grows as $t$ and $q = 0$. Adding radiation only changes the story at early times (see Fig. 8): because the radiation density increases as $(1 + z)^4$, for $z > 3 \times 10^3$ the Universe is radiation dominated and during this epoch, $R \propto t^{1/2}$ and $q = 1$ [this is a manifestation of the fact that gravity is sourced by $\rho + 3p = 2\rho$ for radiation, cf. eqs. (2) and (4)]. It is only during the matter-dominated phase that small density inhomogeneities are able to grow and form bound structures.

Dark energy provides a new twist: because the dark energy density varies slowly if at all, it eventually becomes the dominant form of matter/energy (around $z \sim 0.5$); see Fig. 8. After that, the expansion accelerates and structure formation ceases, leaving in place all the structure that has formed. The future beyond the present epoch of accelerated expansion is uncertain and depends upon understanding dark energy.

In particular, if dark energy is vacuum energy, acceleration will continue and the expansion will become exponential, leading inevitably to a dark Universe. (In a hundred billion years, the light from all but a few hundred nearby galaxies will be too redshifted to detect.) On the other hand, if dark energy is explained by a scalar field, then eventually the field relaxes to the minimum of its potential. If the minimum of the potential energy is zero, the Universe again
becomes matter dominated and returns to decelerated expansion. If the minimum of the scalar potential has negative energy density, the energy of dark matter and of scalar field energy will eventually cancel, leading to a recollapse. Finally, if the potential energy at the minimum is positive, no matter how small, accelerated expansion eventually ensues again.

Absent dark energy geometry and destiny are linked. The presence of dark energy severs this relation\(^7\) and links instead destiny to an understanding of dark energy.

### 6. Summary

We end our brief review with our list of the ten most important facts about cosmic acceleration

1. Independent of general relativity and based solely upon the SN Hubble diagram, there is very strong evidence that the expansion of the Universe has accelerated recently.\(^36\)

2. Within the context of general relativity, cosmic acceleration cannot be explained by any known form of matter or energy, but can be accommodated by a nearly smooth and very elastic \((\rho \sim -\rho)\) form of energy (“dark energy”) that accounts for about 75% of the mass/energy content of the Universe.

3. Taken together, current data (SNe, galaxy clustering, CMB and galaxy clusters) provide strong evidence for the existence of dark energy and constrain the fraction of critical density contributed by dark energy to be \(71 \pm 5\%\) and the equation-of-state to be \(w \approx -1 \pm 0.1\) (stat) \(\pm 0.1\) (sys), with no evidence for variation in \(w\). This implies that the Universe decelerated until \(z \approx 0.5\) and age \(\approx 10\) Gyr, when it began accelerating.

4. The simplest explanation for dark energy is the zero-point energy of the quantum vacuum, mathematically equivalent to a cosmological constant. In this case, \(w\) is precisely \(-1\), exactly uniformly distributed and constant in time. All extant data are consistent with a cosmological constant; however, all attempts to compute the energy of the quantum vacuum yield a result that is many orders-of-magnitude too large (or is infinite).

5. There is no compelling model for dark energy. However there are many intriguing ideas including a new light scalar field, a tangled network of topological defects, or the influence of additional spatial dimensions. It has also been suggested that dark energy is related to cosmic inflation, dark matter and neutrino mass.

6. Cosmic acceleration could be a manifestation of gravitational physics beyond general relativity rather than dark energy. While there are intriguing ideas about corrections to the usual gravitational action or modifications to the Friedmann equation that can give rise to the observed accelerated expansion, there is no compelling, self-consistent model for the new gravitational physics that explains cosmic acceleration.

7. Even assuming the Universe has precisely the critical density and is spatially flat, the destiny of the Universe depends crucially upon the nature of the dark energy. All three fates — recollapse or continued expansion with and without slowing — are possible.

8. Cosmic acceleration is arguably the most profound puzzle in physics. Its solution could shed light on or be central to unraveling other important puzzles, including the cause of cosmic inflation, the vacuum-energy problem, supersymmetry and superstrings, neutrino mass, new gravitational physics, and dark matter.

9. Today, the two most pressing questions about dark energy and cosmic acceleration are: Is dark energy something other than vacuum energy? Does general relativity self consistently describe cosmic acceleration? Establishing that \(w \not= -1\) or that it varies with time would rule out vacuum energy; establishing that the values of \(w\) determined by the kinematical and dynamical methods are not equal would indicate that GR cannot self consistently accommodate accelerated expansion.

10. Dark energy affects the expansion rate of the Universe, which in turn affects the growth of structure and the distances to objects. (In gravity theories other than GR, dark energy may have more direct effects on the growth of structure.) Upcoming ground- and space-based experiments should probe \(w\) at the percent level and its variation at the ten percent level. These measurements should dramatically improve our ability to discriminate between vacuum energy and something more exotic as well as testing the self consistency of general relativity. Laboratory- and accelerator-based experiments could also shed light on dark energy.

Because of its brevity, this review could not do justice to the extensive literature that now exists; for readers interested in a more thorough treatment of dark energy and/or a more extensive review, we refer them to ref. 45.

### References

19. WMAP web site: http://lambda.gsfc.nasa.gov/product/map/
41) South Pole Telescope web site: http://astro.uchicago.edu/spt
43) http://www.rssd.esa.int/index.php?project=PLANCK

Michael S. Turner was born in Los Angeles, CA in 1949. He obtained his B.S. (1971) in physics from Caltech and his M.S. (1973) and Ph. D. (1978) in physics from Stanford University. He was an Enrico Fermi Fellow (1978–1980) at The University of Chicago and joined the faculty in 1980. He is currently the Rauner Distinguished Service Professor in the Kavli Institute for Cosmological Physics at Chicago. Turner is one of the pioneers of the interdisciplinary field of particle cosmology, and has worked on many topics in early Universe cosmology, including inflation, dark matter, big bang nucleosynthesis, CMB, structure formation, and dark energy. His current research is focused on the mystery of cosmic acceleration.

Dragan Huterer was born in Sarajevo, Bosnia and Herzegovina (then Yugoslavia) in 1972. He obtained his B.S. (1996) from MIT and his Ph. D. (2001) in physics from the University of Chicago where Turner was his Ph. D. advisor. He was a postdoctoral fellow at Case Western Reserve University (2001–2004) and National Science Foundation fellow at the Kavli Institute for Cosmological Physics at Chicago (2004–2007). In 2007 Huterer will join the faculty of the Physics Department at the University of Michigan. His research is centered around understanding the mysterious dark energy, and makes use of the cosmic microwave background, type Ia supernovae, and large-scale structure as probes of the universe.
1. Introduction

Building from the earliest measurements of galaxy clusters, through decades of observational cosmology, ground-based observation of the cosmic microwave background (CMBR) and on in recent times to precision cosmology measurements with the CMBR satellite WMAP and observations of distant supernovae,1,2 the remarkable conclusion we have is that the Universe is geometrically flat ($\Omega \sim 1 \pm 0.04$) but contains only ~4% ordinary baryonic matter. Even of this 4% only about 1/10th is actually visible to us, as stars mainly, the rest is likely composed of cold gas, sub-solar mass “dead” stars and other forms of non-luminous baryonic matter. The great majority of the Universe then is found to be a mixture of mysterious dark energy ($\Omega_{\Lambda} \sim 73\%$), the nature of which is unknown, and non-luminous, non-baryonic dark matter ($\Omega \sim 23\%$).3,4

What form this dark matter takes is also so far unknown. However, a generic class of relic particles produced thermally in the early Universe and termed weakly interacting massive particles (WIMPs), has emerged as a leading possibility.5 Such, non-relativistic particles would constitute a cold dark matter (CDM) population that appears required to explain galaxy formation. The observed density required of them in the galaxy ($\Omega \sim 0.1$), is consistent with the freeze-out relic density appropriate if the mass and cross-section of the particles is determined by the weak scale.6,7 The CDM model itself provides significant motivation to search for such weakly interacting neutral particles, the appropriate mass range being several GeV to ~TeV. However, two leading theories in particle physics phenomenology greatly enhance this motivation by also, independently, predicting new particles with these features. Firstly, in supersymmetric (SUSY) extensions of the standard model,7 the lightest SUSY particle (LSP), stable in models where R-parity is conserved, provides the required population, known as neutralinos. Theories of universal extra dimensions in which Kaluza–Klein parity is conserved provides a second possible class known as the lightest Kaluza–Klein particles (LKP).8–10

As an alternative to WIMPs, axions form a further potential candidate, motivated here by extensions to the Standard Model through Peccei–Quinn symmetry as a solution to the strong CP problem.11 The PQ symmetry is spontaneously broken at a scale $f_a$, with the axion as the associated pseudo-Goldstone boson produced in the early Universe.12–15 Though outside the scope of this review we note the rapid progress being made by the ADMX axion search. This experiment is currently setting stringent limits, excluding at >90% c.l. the KSVZ halo axion mass of 1.9–3.3 eV and indicating that the local axion dark matter halo mass density is greater than 0.45 GeV cm$^{-3}$ for KSVZ DFSZ axions.16,17 For a full review on axions see for instance.18

In this rich context, research aimed at an explanation of non-baryonic dark matter encompasses a huge worldwide effort. This includes: searches for SUSY at accelerator experiments, such as the upcoming ATLAS/CMS at the Large Hadron Collider and elsewhere; indirect searches through attempts to detect the products of neutralino self-annihilation in astrophysical objects, such as neutrinos from the Galactic centre, the Sun and Earth; and direct particle searches for both axions and WIMPs using experiments in the laboratory. Focus will be placed here on the latter case—arguably the best motivated candidate studied by the best generic technique. That is, efforts toward detection of relic WIMPs in the galaxy via their direct interactions in detector target materials on Earth.19,20 For a recent review of indirect searches for neutralinos see for instance21 and for accelerators see for example.22,23 The firmest indirect neutralino limits from high energy neutrinos coming from the Sun are currently set by Super-Kamiokande.24

2. Requirements for Direct Detection

For direct detection our starting point is that the Galaxy contains a halo of WIMPs normally assumed to be of spherical isothermal form with local density 0.3
GeV cm$^{-2}$ cm$^{-3}$, an escape velocity of 650 km s$^{-1}$, with rms velocity 279 km s$^{-1}$ and relative Halo-Earth velocity of 235 km s$^{-1}$. The basis for detection is then elastic scattering of these neutral, non-relativistic particles, off target nuclei in a suitable detection medium, such that the energy transferred as the resulting nuclear recoil passes through the material can be observed, usually as either ionisation, scintillation or heat (phonons). Kinematics and the likely mass range and velocity of the particles implies a nuclear recoil spectrum with energy below $\sim 100$ keV, with exponential form rising to low energies and with no spectral features. This characteristic, together with the expected low interaction rate of likely $10^{-6}$ event kg$^{-1}$ d$^{-1}$, dictates three core requirements of WIMP detector technology: low energy threshold ($<10$ keV$_{\text{recoil}}$); potential for target masses of $>10$ kg; and low particle background of all types. The latter implies the need for a deep underground site to reduce cosmic ray muon-induced neutrons, that could otherwise produce nuclear recoils indistinguishable from WIMPs; use of additional passive and active gamma and neutron shielding; and detector construction using materials with greatly reduced radioactive U, Th, and K content.

The coupling of these non-relativistic WIMPs has two terms, a scalar, spin-independent (SI) part and an axial spin-dependent (SD) part. For most SUSY models SI provides the dominant coupling and hence highest rate. This is because although neutralino-nucleon cross-sections are mainly much smaller for the SI case, coherence across the nucleus results in constructive interference which greatly enhances the WIMP-nucleus elastic cross-section for high A targets. The opposite is true for SD where the axial coupling to nucleons with opposite spins interferes destructively. Effectively, sensitivity to SD interactions thus requires a target isotope with an unpaired nucleon, either proton or neutron. Although generally lower sensitivity is implied for the SD case this is not true for all neutralino models. SD targets are certainly required if the full WIMP parameter space is to be studied and the widest investigation of any signals undertaken.

The stringent requirement for low background, bearing in mind, for instance, that typical ambient environmental gamma fluxes can produce event rates $>10^5$ times higher than the expected WIMP signal rate in an unprotected detector, has focussed world attention on technologies that can actively reject electron recoil events, whilst maintaining high sensitivity to nuclear recoils. This is possible in principle because the latter have typically $\times 10$ higher dE/dx values. In practice, few technologies can make use of this physics, the prime ones being: (1) low temperature ionisation/phonon or scintillation/phonon detectors in which the ratio of event-produced ionisation or scintillation to phonons is measured in suitable cryogenic materials such as Ge or Si (ionisation) and CaWO$_4$ (scintillation); and (2) noble liquid gases, notably xenon and argon, in which scintillation and ionisation is measured simultaneously. A moderate level of discrimination can also be achieved in specific scintillators such as NaI(Tl), CsI(Tl), liquid Ar, and liquid Xe by pulse shape analysis (see §4).

Although recoil discrimination, and background reduction, appears feasible in such technologies there remains the issue, given the lack of spectral features in the recoil spectrum, of how to determine in a clear way whether any remaining counts are due to WIMPs from the galaxy and not either nuclear recoils from an unaccounted for background (such as neutrons or surface interactions), or a detector artefact. There are two prime possibilities for addressing this using galactic dynamics. Firstly, at least for the standard halo model, the Earth’s motion through the Galaxy implies an expected seasonal modulation in the recoil spectrum (flux and shape). This is because the component of the Earth’s solar orbital velocity in the direction of our galactic motion (orbital plane inclined at 60°) either adds to or subtracts from the galactic orbital velocity depending on the season. Secondly, thanks also to our galactic orbital motion ($\sim 235$ km s$^{-1}$), we would expect the direction of the WIMP-induced nuclear recoil tracks themselves within a target to be dominantly opposite to our direction of motion (in galactic coordinates) (see §8). Information may also be gleaned by comparing different targets, since WIMPs interact differently with target nuclei of different A (see §8) different technologies and different sites.

Unfortunately, the annual modulation effect is very small, typically a few %, requiring already at least tonne-scale detectors to obtain sufficient event statistics for a viable search. Such small effects are also at risk of being masked by detector characteristics always vulnerable to natural seasonal changes in the environment. The recoil direction effect is far more powerful, in principle. Only of order a few $10^3$ s of WIMP events may be required in such a recoil direction sensitive detector to identify them as of galactic origin (see §8). Furthermore, the angular distribution of WIMP-induced recoil tracks can not be mimicked by any terrestrial backgrounds since we would expect a sidereal (not daily) modulation of the WIMP-induced track directions and an average “washed-out” isotropic distribution of any background in the galactic frame. The challenge here, however, is the likely need to use low pressure gas detectors that would then need to be very large in volume ($100$ s m$^3$).

Built on these basics a wide variety of experiments have and are being run worldwide. Figure 1 provides a summary of results from recent key examples, given here as an exclusion plot of WIMP-nucleon cross-section vs WIMP mass for the SI case, assuming the standard halo model as above. Referring to this the following sections outline the current status and possible future scenarios. Note this is necessarily selective and likely tinged by personal bias—for a wider view we refer to recent workshop proceedings such as. Note also that some of these results are yet to be published. They are shown here for the sake of completeness as reported in workshops and preprints, leaving discussion of their validity out of the scope of this review.

3. Semiconductors

Ionization detectors, in the form of low background germanium (HPGe) and silicon diodes used for double beta decay searches, provided the first limits on WIMP interactions. Such experiments were vital to ruling out early candidates for WIMPs including, Cosmions and heavy Dirac neutrinos. However, as a technology they suffer from an inability to distinguish between gamma background events and the nuclear recoil events of interest. This is partly
compensated for by the possibility of high radio-purity in Ge which has allowed more recent experiments such as HDMS and IGEX to set interesting limits (see Fig. 1).

The IGEX experiment at Canfranc used 2.1 kg of purified Ge with a 20 cm thick Pb gamma shield inside a muon veto. The detector achieved an eventual background of 0.21 keV\(^{-1}\) kg\(^{-1}\) d\(^{-1}\) at 4–10 keV.

Next generation HPGe detectors aim at further reduction in activity, for instance by possibly \(\times 1000\) using novel techniques such as crystal growth underground to reduce cosmic-ray spallation activity. There is also prospect for active rejection of Compton scatter events using segmentation to provide position sensitivity and use of active coincidence Compton vetos. Key ideas have been proposed by GEDEON, following from IGEX; the GERDA experiment at Gran Sasso; and MAJORANA. All these are primarily aimed at neutrinoless double beta decay detection. The GERDA detector incorporates the novel prospect of using direct submersion in liquid argon or nitrogen (an idea tested by GENIUS-TF) with, for argon, use of the liquid as a possible active veto.

4. Scintillators

Pulse shape analysis (PSA) in certain organic and inorganic scintillators has been known to allow discrimination against low \(dE/dx\) events (electron recoils) for many decades. NaI in particular was turned to advantage for WIMP searches by the UKDM collaboration, using cooled undoped NaI and later NaI(Tl), and by BPRS/DAMA. Unfortunately, the light output \(\sim 40\) photons per keV\(_{\text{electron}}\) in NaI(Tl) is too low for event by event discrimination at low energy even though the quench factors \(\sim 9\%\) for I and \(\sim 25\%\) for Na) are relatively high. Statistical methods, combined with material purification to reduce intrinsic activity, can be used and were successfully implemented in, for instance, NAIAID to produce significant new limits. Nevertheless, the discrimination power with statistical analysis is limited. The DAMA experiment thus later turned to using NaI(Tl) in simple counting mode as a means of searching instead for an annual modulation signal, with no nuclear recoil identification applied. They have claimed evidence for a modulation, reporting the discovery of WIMPs in 1997.

The final DAMA result from a total of 107,731 kg day accumulated with 9 low background 9.7 kg NaI(Tl) crystals remains the only claimed direct observation of WIMPs, corresponding to a mass of \(\sim 52\) GeV and cross-section \(\sim 7.2 \times 10^{-6}\) pb (for standard halo model assumptions). However, the result appears in contradiction with several other experiments including the bolometric Ge experiments of EDELWEISS and CDMS, and the liquid xenon experiment ZEPLIN (see §5 and §6). This contradiction appears to hold regardless of the halo model or if SD interactions dominate, though there remains debate as to whether fine tuning of models can allow compatibility, particularly for the SD case. The DAMA group is now running an expanded array, the 250 kg LIBRA experiment. Following closure of NAIAD no direct test is being made of the result with an independent NaI-based detector, although the Zaragosa/Canfranc group is building a 107 kg NaI experiment, ANAIS, to address this gap and the KIMS experiment is producing competitive limits with CsI.

More recently there has been interest in other inorganic scintillators, notably CsI(Tl), CaF\(_2\)(Eu), and, for instance, CaWO\(_4\). The former, now developed for the KIMS experiment in South Korea, has intrinsically better pulse shape discrimination than NaI(Tl) but potentially higher intrinsic background, in particular due to \(^{137}\)Cs from nuclear fallout. Nevertheless, encouraging results have been obtained by taking care in material selection and purification. CaF\(_2\)(Eu) has relatively poor discrimination, while CaWO\(_4\) and similar compound inorganics are not efficient scintillators at room temperature but operate well at mK temperature. CaWO\(_4\) has become an integral part of the CRESST bolometric experiments in which scintillation light is measured simultaneously with heat (see §5). For recent measurements of the quench factors here see.

Fig. 1. (Color online) Summary of current spin-independent WIMP-nucleon limits (for references and details see text).
Certain organic crystal scintillators such as stilbene, plastics and liquids also demonstrate pulse shape discrimination.\textsuperscript{44} They have the advantage of potentially relatively low cost per kg and high radio-purity. However, their composition is dominated by H, C, and possibly F or other light elements. This results in poor quench factors, typically 2\% or less,\textsuperscript{63,69} and poor kinematic coupling to WIMPs leading to very poor sensitivity relative to the inorganics like NaI(Tl). Nevertheless, there has been significant interest in the organic crystals, in particular because some of these, for instance stilbene and anthracene, yield a response that is dependent on the direction of the contained recoiling nucleus, at least as measured using alpha particles. This yields a rare example of a technology relevant to the possibility of a direction sensitive WIMP detector\textsuperscript{70,71} (see §8).

Whilst the lack of powerful recoil discrimination is a disadvantage for experiments based purely on scintillation detection there is a potential advantage for SD sensitivity due to the greater possibility of using spin nuclei, particularly iodine (e.g., in NaI, CsI) and fluorine (e.g., in CaF\textsubscript{2}). This has allowed NAIAD to maintain competitive limits for WIMP-proton coupling.\textsuperscript{72} Finally, a remaining class of scintillators of interest is the liquid noble gases, notably liquid xenon and argon. These are covered in §6.

5. Bolometers

At low temperature the heat capacity of a dielectric crystal goes as $T^3$. Thus at mK temperatures the small energy deposition from a nuclear recoil can yield a measurable proportional increase in crystal temperature.\textsuperscript{20} Some of the earliest techniques investigated for WIMP dark matter detection were based on this, where energy released by particle interactions can be observed as phonons or quanta of lattice vibrations. Work started on this idea in the 1980’s (see for instance\textsuperscript{23}), the original motivation being in part the prospect of obtaining very low recoil energy thresholds and high energy resolution, due to the meV level of quantisation involved.\textsuperscript{74} However, it was soon demonstrated, first in Si\textsuperscript{25} and then in Ge,\textsuperscript{26} that phonon detection could be combined with simultaneous detection of ionisation to provide also a powerful means of discrimination against electron recoils, on an event by event basis. This arises because the proportion of energy observed in the two channels is dependent on the event $dE/dx$—a high $dE/dx$ event, such as a recoiling nucleus, produces proportionally more heat than ionisation (the ionisation is quenched). For instance, the ratio of ionization to recoil energy (the ionisation yield) for Ge recoils in Ge is $\sim$0.3 of the value for electron recoils above 20 keV.\textsuperscript{53}

Whilst bolometers without collection of ionisation have proven quite useful for dark matter searches, the hybrid technique of simultaneous ionisation and phonon collection with its capability for background rejection has been pushed harder. Most notable is the CDMS collaboration (at Soudan mine) and EDELWEISS-I (at Frejus)\textsuperscript{33,77-79} (see Fig. 1). The latter used 320 g Ge crystals operated at 17 mK with NTD-Ge thermometric sensors attached for the heat signal and Al electrodes used to collect the charge. 10 cm of Cu and 15 cm of Pb where used to shield the cryostat from rock gamma-ray background with an additional 7 cm Pb inside and a total of 30 cm paraffin outside the entire setup to reduce rock neutrons. A variety of detectors were tried in EDELWEISS-I with several runs completed from 2000 until March 2004. These yielded a total exposure of 62 kg days, the main results coming from three crystals with recoil energy threshold of 13 keV or better over 4 months of stable operation. Figure 1 shows the limits produced. After cuts a total of 40 nuclear recoil candidates were recorded in the range 15–200 keV with 3 events between 30 and 100 keV, most likely due to remaining background neutrons or surface electrons.

The CDMS experiment operates towers of Ge and Si crystals each 1 cm thick and respectively of mass 250 and 100 g. These are mounted in a dilution fridge and shielded mainly by 22.5 cm of external Pb and 50 cm of polyethylene. A 5 cm layer of plastic scintillator is used to veto any events coincident with cosmic muons (necessary here due to the relative shallowness of the Soudan site at 2080 m.w.e.). Charge electrodes are used for ionisation collection as in EDELWEISS but here athermal phonons are detected using superconducting transition edge sensors, applied by photolithography to the crystal surfaces. This design has the potential advantage of providing depth position sensitivity, via measurement of the phonon pulse risetime, and hence the possibility of rejecting surface electron events that could otherwise contaminate the signal region, as suspected in EDELWEISS-I. Two towers were operated in 2004 yielding an exposure for 10–100 keV\textsubscript{recoll} of 34 kg days Ge and 12 kg days Si. No events were observed in the Si and only one event, consistent with the expected surface event background in the Ge, yielding a 90\% c.l. SI upper limit in Ge of 1.6 $\times$ 10$^{-7}$ pb at 60 GeV c$^{-2}$ WIMP mass (see Fig. 1).

As an alternative ROSEBUD\textsuperscript{80} and CRESST\textsuperscript{66,81} have developed detectors in which scintillation light is measured in coincidence with heat, in particular using CaWO\textsubscript{4}.\textsuperscript{81} Here a silicon wafer of 30 $\times$ 30 $\times$ 0.4 mm$^3$ with tungsten thermometer is used to detect the photons and a 8 $\times$ 8 mm$^2$, 200-nm-thick superconducting evaporated film used as the heat sensor. Although only 1\% or less of the energy deposited is detected as photons this is much higher than feasible at room temperature and is sufficient to produce an energy resolution comparable to NaI(Tl). Results so far have been obtained with two 300 g crystals at the Gran Sasso underground laboratory with a total exposure of 20.5 kg days. This revealed 16 events in the range 12–40 keV consistent with the expected background from neutrons given that the experiment did not have a neutron shield. The resulting limit, due to W recoils (see Fig. 1), is comparable to others in the field, including EDELWEISS.

Notable in the pure cryogenic detector field is the work of the Milan group through the CUORE/COURICINO experiment at Gran Sasso.\textsuperscript{32,83} CUORE is designed primarily for neutrinoless double beta decay searches. CUORE demonstrates one particular advantage of pure cryogenic detectors over the hybrid types. The latter is essentially restricted to Ge and Si because only these are found to have sufficiently high electron–hole transport at mK temperatures. The non-hybrid technique, at the expense of throwing away recoil discrimination, is open to a much greater variety of target crystals, for instance TeO\textsubscript{2} for CUORE, LiF, Sapphire, and others have been demonstrated. CUORE aims to build an array of 988 TeO\textsubscript{2} cryogenic crystals with total mass...
~750 kg, building on the first stage CUORICINO experiment already operated with 62 (~40.7 kg) crystals. Although CUORE will have exceptionally high target mass, competitive WIMP limits will only come through more work to suppress intrinsic crystal backgrounds.84)

All these cryogenic experiments are now progressing towards significant upgrades. CDMS is proposing 25 kg and a possible move to the deeper SNOLAB site. EDELWEISS is progressing towards a more ambitious phase II with up to 120 detectors and CRESST is upgrading to allow 33 CaWO4 detectors, totalling 10 kg. However, as outlined in §7, it is likely that even greater target mass will be needed, possibly at the tonne-scale or larger.

6. Liquid Noble Gases

Whilst a large world effort has been devoted to cryogenic bolometers over many years, linked now to quite an industry in alternative applications, there has been recent rapid growth in liquid noble gas technology for WIMP searches. Most notable has been liquid xenon (LXe), started by DAMA/Xe,85,86 but also recently liquid neon87 and, in particular, liquid argon. A prime motivation here has been improved low background discrimination combined with prospects for tonne-scale target mass at reasonable cost (see §7). LXe has particularly good intrinsic properties for WIMP detection including: high mass (Z = 54, A = 131.3) yielding a good kinematic match to likely WIMP candidates; high scintillation and ionisation efficiency (~46 photons/keV at 178 nm); and high radiopurity, enhanced further by the availability of liquid gas purification techniques. However, of greater importance is the recoil discrimination achievable. This is possible firstly, as in NaI(Tl), by simple PSA of the scintillation light. This is the basis for the single phase LXe experiments in Japan88 and of ZEPLIN I.85 The latter detector, comprising 3.2 kg of active LXe viewed by 3 PMTs, accumulated 293 kg days during operation at Boulby mine until 2002, producing significant limits with this technology (see Fig. 1).

The XMASS group have also run a 100 kg prototype PSA detector and are aiming to achieve higher sensitivity by constructing an 800 kg experiment for operation in Kamioka mine. However, more powerful discrimination is feasible in LXe by recording also the ionisation produced and hence the ionisation/scintillation ratio. This arises because for nuclear recoils the ionisation signal (termed S2) is quenched significantly more than the primary scintillation (S1) relative to electron recoils of the same energy. This is being implemented by the ZEPLIN II/III, XENON 10/100 and XMASS II (two-phase) experiments, all aiming to achieve higher sensitivity with lower fiducial mass than likely required with PSA alone.89–93 Collection of event ionisation may not match that of bolometers. In this respect liquid argon (LAr) may provide better prospects. Some properties of LAr are inferior to LXe for WIMP searches, notably the lower Z, A (18, 40) and the need to use wavelength shifter for the VUV scintillation light (~40 photons/keV at 135 nm). However, both pulse shape discrimination and two phase primary/secondary discrimination are now known to be more powerful,94,95 capable in principle of combined discrimination factors up to ~105. LAr is also a factor ~×400 lower in cost. Based on this, the WARP collaboration has built and deployed a 3.2 kg LAr experiment at LNGS and recently reported a sensitivity near 10−8 pb at 100 GeV c−2 for an exposure of 96.5 kg days.96,97 WARP is currently constructing a larger 140 kg detector with a full active Compton veto. Other LAr experiments are also under construction including ArDM, CLEAN and DEAP.98–101 ArDM involves two-phase operation but with ionisation readout using direct collection by large electron multipliers (LEMs) in the gas phase. The other designs are based on single phase PSA plus self-shielding, akin to the XMASS concept with liquid xenon.

Although ZEPLIN I, II, XENON 10, and WARP have shown excellent progress, significant issues remain with
liquid noble gases. Firstly, the quench factor for nuclear recoils remains poorly determined. Measurements for LXe in zero field have indicated 0.13–0.23 (10–56 keV)\(^{102}\) but higher values have been claimed.\(^{103}\) For argon there are several conflicting results, for instance,\(^{97,100}\) Secondly, LXe has not yet demonstrated recoil discrimination competitive with cryogenic technology. Populations of events are observed to spread from the gamma region into the signal region. This may reflect the youthfulness of current detector designs but may be intrinsic. It is possible that spontaneous single electron emission occurs in the liquid, producing secondary electroluminescence with minimal ionisation signal and yielding events indistinguishable from nuclear recoils. The ZEPLIN III detector, currently being installed at Boulby, has improved light collection and higher drift fields than ZEPLIN II and will be used in part to investigate this prospect.\(^{91}\)

The background issue above may also be present in LAr. However, for argon\(^{104}\) there is a more important issue to resolve, the presence of radioactive \(^{39}\)Ar. Produced by cosmic ray spallation in the atmosphere, this yields in argon alpha background of \(\sim 1\) bq/lt. Discrimination of 10\(^{10}\) would be sufficient in principle to cope but then data acquisition deadtimes in a tonne-scale detector would likely be unmanageable. Calculations show that argon from deep gas wells, shielded from cosmic rays, could provide an economic source of, so called, dead-argon. Activation times on the surface are long enough that once brought to the surface there is sufficient time to construct and deploy an experiment.\(^{105}\)

7. Tonne-scale Concepts and Alternative Techniques

After over two decades of development, WIMP experiments with target masses of kg-scale are reaching sensitivities improved by about 4 orders of magnitude, probing well into SUSY favoured parameter space. This achievement has been accompanied by a continuing rise in the number of experimental scientists involved, now \(\sim 300\). There has been an expansion of interest in new and emerging technologies, not just liquid noble gases but others not detailed here, including superheated droplet detectors (SSDs), specifically SIMPLE and PICASSO, and the MACHE3 detector that uses superheated He.\(^{106-110}\) The SSD experiments, through use of Fluorine-loaded targets, show particular promise for SD sensitivity and are producing interesting limits, however, whilst all this activity together reflects substantial maturity, a crossroads has likely been reached in the field.

Firstly, it is pretty certain, setting aside claims by DAMA, that favoured SI coupled dark matter does not exist with cross-sections \(\sim 2 \times 10^{-7}\) pb (see Fig. 1). Meanwhile, theoretical predictions for both neutralinos and LKPs reach \(10^{-11}\) pb.\(^{8,111-113}\) Thus next generation experiments must not only achieve further background suppression but also be capable of tonne/multi-tonne masses, simply to ensure a statistically observable signal rate. This represents a major leap, implying significantly higher costs and likely larger collaborations. Secondly, for such large detectors it can be argued that though active gamma discrimination remains important, greater emphasis is needed on material purification, passive shielding of external backgrounds and on searches for additional features in the data to show that remaining events are non-terrestrial signals and not, in particular, neutrons.

The latter argument arises as follows: assuming next experiments are at sufficient depth to avoid muon-induced neutrons, then gammas and neutrons from U/Th chains in the environment and detector will dominate background. For the relevant energy range, \(\sim 200\) keV, such contamination produces typically \(10^5-10^6\) more gammas than neutron-induced nuclear recoils.\(^{20}\) The levels of detector sensitivity required for tonne-scale experiments now imply that gamma backgrounds must be comparable with or lower than the neutron rates, such as could be achieved by neutron/gamma discrimination of \(10^5-10^6\) [the rate for fast neutrons from the rock at Boulby, for instance, has recently been measured to be \(1.72 \pm 0.61\) (stat.) \(\pm 0.38\) (syst.) \(\times 10^{-16}\) cm\(^{-2}\) s\(^{-1}\) above 0.5 MeV].\(^{114}\) Thus neutron induced recoils, which can not be distinguished from WIMP interactions, naturally will be the dominant particle background. Detector position sensitivity may help, by allowing rejection of multi-scatter events.\(^{115}\) However, ultimately reliance will be needed on passive neutron shielding and material purification plus WIMP signal identification via: (1) use of at least two targets/technologies with different \(A\) and different systematics; and/or (2) correlation of events with Galactic motion by observation of annual modulation or a directional signal. The former relies on the different behaviour of WIMP and neutron scattering cross-section vs \(A\) to deduce that a signal is not neutrons. The latter allows direct identification of events as of extra-terrestrial origin.

Following these notions, similar to arguments adopted in neutrino physics by, for instance, Borexino and SNO,\(^{116,117}\) it is natural to consider larger WIMP detectors with (near) spherical design, a central fiducial zone containing minimal detector components other than the target material, and an integral passive outer shield. This is the basis of the XMASS (Xe), CLEAN (Ar and Ne), and DEEP (Ar) scale-up programmes (see §6). Single phase liquid noble gases are used here with photons recorded by photomultipliers in the outer region, pointing inwards (see for instance Fig. 3). This allows both position information (fiducialisation) and some recoil discrimination via PSA, but the dominant theme is bulk passive shielding for gammas and neutrons.

![Fig. 3.](image-url) (Color online) Schematic design for the proposed miniCLEAN 100 kg detector, a precursor to a potential tonne-scale experiment (for details see text and ref. 100).
The reliance on PSA and the presence of PMTs (with relatively high radioactivity) is a potential limitation for these experiments. Replacement of PMTs is a possibility by using an internal photocathode, such as CsI, to convert photons to electrons for collection by charge readout in the gas phase, for instance using micromegas or gas electron multipliers (GEMs). Current two-phase programmes, with greater discrimination potential, are also being developed for scale-up, for instance WARP, LUX, XENON100, and ArDM. The latter is already at the tonne-scale and uses LEM charge readout. However, use of gas-phase electroluminescence is not well suited to the benefits of a pure spherical concept due to the need for a top gas layer. An alternative hybrid design has been suggested, termed CORE, in which the ionisation signal is recorded directly in the liquid phase at a point gain region central to a sphere. Such a spherical TPC concept has in fact already been realised in the gas phase by NOSTOS. Use of high pressure noble gas, as developed by SIGN, may itself provide an alternative class of scale-up technique.

Cryogenic technology is not so well suited to the massive self-shielding spherical concepts above. Nevertheless, scale-up to tonne-scale is planned here also, making best use of the high discrimination power demonstrated notably by CDMS, EDELWEISS, and CRESST. Two particular efforts are foreseen, SuperCDMS and EURECA (European Underground Rare Event search with Calorimeter Array). The former will use Ge and Si ionisation/thermal technology like CDMS in a staged expansion from 27 to 145 kg and eventually to 1100 kg by 2015, either at the US DUSEL, if built, or SNOLAB in Canada. EURECA represents a merger of EDELWEISS, CRESST with further new groups to develop a 100–1000 kg array using various targets, possibly both ionisation/thermal and scintillation/thermal ideas. For both experiments a priority will be the need to develop improved detectors, in particular to allow better rejection of surface events, for instance through event position reconstruction or improved analysis, and to reduce unit costs.

8. Directional Detectors and Proof of a Galactic Signal

The scale-up programmes, assuming more than one becomes reality, in part address the signal identification issues noted in §7 by opening the way to examining the A-dependence of a potential WIMP signal. However, definitive proof that a signal is of galactic, and not terrestrial, origin can only be achieved by correlating in some way the events with our motion through the Galactic WIMP halo (see §1). This has been the objective of DAMA/LIBRA by making use of the small predicted annual modulation in flux or energy. However, a much more powerful, though technologically challenging, possibility is to correlate in 3D the physical direction of nuclear recoils in a target with our motion. This is the motivation behind the DRIFT, MIMAC, NEWAGE, and other low pressure gas time project chamber (TPC) r&d programmes. Calculations show that in principle only a few s of WIMP events are needed to prove a Galactic origin. Furthermore, a powerful sidereal day modulation of the signal is expected in the laboratory frame, impossible to be mimicked by any terrestrial background.

Much progress has been made here by the US-UK DRIFT collaboration using negative ion CS₂ TPCs at Bouly mine. The latest version, DRIFT II, comprises 3 units of 1 m³ of CS₂ at 40 Torr (each 70 g fiducial mass). The reduced pressure is needed so that nuclear recoil tracks are extended to a few mm, sufficient for observation by the multi-wire proportional counters (MWPC) readout used. Negative ion gas is used to minimise track diffusion without the need for expensive magnets, the CS₂ (and possibly additives) providing also a multi-A target. Each detector contains a 1 m³ central high voltage cathode plane and two back to back drift regions of 50 cm depth, each read out by a 1 m² MWPC comprising planes of 20 μm wires at 2 mm pitch.

DRIFT II demonstrated stable, neutron shielded, long-term running during 2005/6. Operation is by remote control at room temperature, with no cryogenics or complex services. The in-built sensitivity of the TPC technology to particle dE/dx (ionisation charge density) allows exceptional electron track rejection (>10⁶), sufficient that no gamma shielding is required for DRIFT II. More importantly analysis of event drift time, MWPC anode hits and induced signals on the orthogonal grid planes allows in principle full 3D reconstruction of ionisation tracks down to 300 NIPs (number of ionising pairs) or ~20 keV recoils. Figure 4 shows 3D reconstruction of a typical S-recoil event resulting from a neutron elastic scatter.

As currently the only known route to significant recoil direction sensitivity, TPC technology holds exceptional potential for WIMP physics and is possibly the only route to a definitive galactic signal. However, there are several challenges to address. The requirement for use of low pressure gas implies large volume detectors will eventually be needed, possibly 1000 s m³. For this, new charge readout technologies such as bulk Micromegas and GEM planes are under development to reduce spatial resolution and hence allow high pressure, lower volume, operation. A further issue is the desirability of achieving track head to tail discrimination. Orientation of a recoil track alone provides significant directional information but a factor ~×10 greater sensitivity can be achieved in principle if the head can be distinguished in form the tail. Whilst more careful data analysis is required to demonstrate feasibility it is clear from recent detailed simulations that this also may be possible. Shown in Fig. 5 is an example simulated track for a 100 keV S recoil in 40 Torr CS₂ together with dE/dx curves. Firstly, the predicted ranges for the simulated S tracks are found to agree to <10% with experimental data. The dE/dx for the observable electronic channel suggests, for the energy range of interest below 100 keV, a rapid decrease in ionisation per unit length toward the end of the track (we are well below the Bragg curve peak at these energies). This would indicate a head-tail asymmetry favouring more signal at the beginning of the track (head) than the end. However, simulations also show, as indicated in the example track, that straggling (ball-up) increases at the end. Therefore, a charge readout mechanism that involves projection of the track ionisation onto a single axis, for instance, will likely be sensitive to this topological issue, increasing the charge measured for the tail. Work is underway to determine whether any asymmetry
is observable in a practical detector set-up and how this depends on $dE/dx$ and the track topology in relation to readout systems. A new international cooperation, CYG-NUS, has formed to address this and to study the design challenge of building a very large directional dark matter detector.\cite{cyg-nus}

9. Conclusion

In summary, great progress has been made toward detection of particle dark matter in recent years, notably through development of cryogenic detectors but also liquid noble gas experiments now beginning to set the best limits. Much effort has been placed on producing technology with recoil discrimination against gammas. However, it is not necessarily clear that gamma discrimination alone will be sufficient to prove the presence of WIMPs. Proof that any remaining signal is in fact from extra-terrestrial dark matter and not neutrons or some other un-determined terrestrial background or artefact will be vital. Multi-tonne detectors with optimal passive shielding to achieve sufficient count rate are being developed but there is also a potential route toward the needed definitive galactic signal through new recoil direction sensitive technology.

Acknowledgments

The author wishes to thank E. Daw, V. Kudryavtsev, and P. Majewski (University of Sheffield) for many useful discussions. Thanks also to the ILIAS integrated activity (contract R113-CT-2004-506222) workpackage N3-LX, N3-AD, N3-BSNS; the INTAS programme (grant number 04-78-6744); and the UK’s Particle Physics and Astronomy Research Council (PPARC) for financial support.

Neil J. C. Spooner was born in 1960 in Redruth, England. He obtained a first class Physics degree from Southampton University and Ph. D. (1985) from Imperial College, London. After research in the nuclear industry he returned (1987) to Imperial and Rutherford Laboratory to pioneer technology for detection of weakly interacting massive particles. In 1989 he was key to establishing the Boulby Underground Laboratory, U.K. He moved to Oxford (1991) to develop NaI-based dark matter experiments and directional technology. In 1994 he moved to Sheffield University as Reader and in 1998 was appointed U.K. Dark Matter Consortium spokesman. In 1999 he was awarded a personal chair, became head of Particle Physics and a PPARC Senior Fellow (2002). He has been instrumental in expanding the Boulby Laboratory and in European Particle Astrophysics and Underground Science initiatives including two years as ILIAS Project Scientist. His current research interests are development of directional and tonne-scale technology for dark matter and proton decay.
Dark Matter and Particle Physics

Michael E. PESKIN*  
Stanford Linear Accelerator Center, Stanford University, Menlo Park, CA 94025, U.S.A.

(Received May 1, 2007; accepted June 7, 2007; published November 12, 2007)

Astrophysicists now know that 80% of the matter in the universe is "dark matter", composed of neutral and weakly interacting elementary particles that are not part of the Standard Model of particle physics. I will summarize the evidence for dark matter. I will explain why I expect dark matter particles to be produced at the CERN LHC. We will then need to characterize the new weakly interacting particles and demonstrate that they are the same particles that are found in the cosmos. I will describe how this might be done.

KEYWORDS: dark matter, WIMP, LHC, ILC
DOI: 10.1143/JPSJ.76.111017

1. Introduction

One of the themes of the history of physics has been the discovery that the world familiar to us is only a tiny part of an enormous and multi-faceted universe. From Copernicus, we learned that the earth is not the center of the universe, from Galileo, that there are other worlds. More recently, Hubble’s extragalactic astronomy taught us that our galaxy is a tiny part of an expanding universe, and the observation of the cosmic microwave background by Penzias and Wilson revealed an era of cosmology before the formation of structure. Over the past 10 years, astronomers have recognized another of these shifts of perspective. They have shown that the stuff that we are made of accounts for only 4% of the total content of the universe. As I will describe, we now know that about 20% of the energy in the universe takes the form of a new, weakly interacting form of matter, called “dark matter”. The remaining 75% of the energy of the universe is found in the energy content of empty space, “dark energy”.

Dark energy is the most mysterious of these components. Its story is described by Turner and Huterer. Dark matter, though, is the component that most worries the imaginations of particle physicists. What particle is this dark matter made of? Why have we not discovered it at our accelerators? How does it fit together with the quarks, leptons, and bosons that we have spent our lives studying?

And, conversely, dark matter is the component that most excites us by the possibility of its discovery. There are strong arguments that the next generation of particle accelerators, beginning next year with the Large Hadron Collider (LHC) at CERN, will produce the elementary particles of which dark matter is made. How can we recognize them? How can we prove that these particles are the ones that are present in the cosmos? And, finally, how can we use this knowledge to image the dark matter structure of the universe? I will address all of these questions in this article.

2. Evidence for Dark Matter

Although the astronomical picture of dark matter has become much clearer in the last ten years, the evidence for dark matter goes back to the early days of extragalactic astronomy. The evidence for dark matter is summarized in a beautiful 1988 review article by Virginia Trimble. I will describe the most telling elements here.

In 1933, Fritz Zwicky measured the mass of the Coma cluster of galaxies, one of the nearest clusters of galaxies outside of our local group. Zwicky’s technique was to measure the relative velocities of the galaxies in this cluster from their Doppler shift, use the virial theorem to infer the gravitational potential in which these galaxies were moving, and compute the mass that must generate the potential. He found this mass to be 400 times the mass of the visible stars in galaxies in the cluster. The observation was soon confirmed by similar measurements of the Virgo cluster by Smith.

We now know that most of the atoms in clusters of galaxies are not seen in observations with visible light. Because these clusters generate enormously deep gravitational potential wells, it is easy for hydrogen gas from the galaxies to leak out and fill the whole volume of the cluster. These atoms acquire large velocities and emit X-rays when they collide. X-ray images show the clusters as glowing balls of gas. This does not remove the mystery, however, The X-ray emitting gas accounts for at most 20% of the mass of the cluster and cannot explain the origin of the deep potential well. For this, we must postulate that the clusters are also filled with a new, invisible, weakly interacting form of matter.

In the 1970’s, astronomers began to systematically measure the rotational velocity profiles or rotation curves, for many galaxies. One would expect that the mass of a galaxy is concentrated in the region where the stars are visible. Then, outside this region, Kepler’s law would predict that the velocities should fall off as $1/r$. In fact, the velocities are seen to be constant or even slightly increasing. In the galaxy NGC 3067, using hydrogen gas lit up by a background quasar, Rubin, Thonnard, and Ford showed that the rotational velocity profile maintains its large value at a distance of 40 kpc ($10^3$ light-years) from the center of the galaxy, even though the visible stars become rare outside of 3 kpc. From measurements of the velocities of globular clusters, it was found that the rotation curve of our own galaxy is also flat out to distances of 100 kpc from the center.
Detailed measurements of cosmic microwave background, including not only the averaged intensity of this background radiation but also its fluctuation spectrum, give additional information on dark matter. The microwave background was emitted at the time of recombination, when the hydrogen filling the universe, at a temperature of about 1 eV, converted from an ionized plasma to a transparent neutral gas. From the Fourier spectrum of fluctuations of the background radiation, it is possible to measure the dissipation of this medium. The most recent measurements from the WMAP satellite require a medium in which only 20% of the matter is hydrogen gas and 80% is composed of a very weakly interacting species in nonrelativistic motion. These measurements can be converted to the current fractions of atomic and dark matter in the total energy of the universe, \( \Omega_i = \rho_i / \rho_{\text{tot}} \), \( \Omega_{\text{DM}} = 0.20 \pm 0.02 \) (1)

In all of these systems, dark matter is observed only through its gravitational influence. One might wonder, then, whether it is possible to explain the observations by modifying the law of gravity rather than by introducing a new form of matter.\(^{10}\) The interpretation in terms of a new form of matter was recently boosted by the observations shown in Fig. 1. These picture show three images of the galaxy cluster 1E0657-558.\(^{11}\) The first is the optical image, showing the galaxies that, as we have discussed, make up only a few percent of the mass of the cluster. The second picture shows the X-ray image from the Chandra telescope. This image shows where the bulk of the gas in the cluster is located. The superimposed contours show the total density of the mass in the cluster, as measured by gravitational lensing. It is remarkable that the peaks of the mass distribution occur where there are very few atoms. In this object, which probably arose from a collision of two clusters of galaxies, the atomic matter and the dark matter have become spatially separated. The observations cannot be explained by an altered law of gravity centered on the atoms. They require dark matter as a new and distinct component.

### 3. The WIMP Model of Dark Matter

Thus, dark matter exists. What is it made of? In the Standard Model of particle physics, we know no neutral heavy elementary particles that are stable for the lifetime of the universe. Let us postulate a new species of elementary particle to fill this role. Bahcall called this a Weakly Interacting Massive Particle (WIMP). I would like to add one more assumption: Although it is stable, the WIMP can be produced in pairs (perhaps with its antiparticle), and it was produced thermally at an early time when the temperature of the universe was very high. WIMPs must also annihilate in pairs. I will assume that these processes established a thermal equilibrium.

These assumptions lead to an attractive theory of dark matter whose consequences I will explore in the remainder of this article. There are other models of dark matter that do not fit into this paradigm. A comprehensive review of dark matter models that has recently been given by Bertone, Hooper, and Silk.\(^{12}\)

Using the WIMP model, we can build a quantitative theory of the density of dark matter in the universe. As the universe expanded and cooled, the reactions energetic enough to produce WIMPs became more rare. But at the same time, WIMPs had more difficulty finding partners to annihilate. Thus, at some temperature \( k_B T_i \), they dropped out of equilibrium. A small density of WIMPs was left over. At this era, the energy density of the universe was dominated by a hot thermal gas of quarks, gluons, leptons, and photons, with a total number of degrees of freedom \( g_* \approx 80 \). Using this thermal density to fix the expansion rate of the universe as a function of time, we can determine the evolution of the WIMP density by integrating the Boltzmann equation. It is convenient to normalize the WIMP density to the density of entropy, since in standard cosmology the universe expands approximately adiabatically. Then one finds

\[
\Omega_{\text{DM}} = \frac{s_0}{\rho_{\text{tot}}} \left( \frac{\pi}{45 g_*} \right)^{1/2} \frac{k_B T_i / m c^2}{m_p s / h^2} \langle \sigma v \rangle \quad ,
\]

where \( s_0 \) and \( \rho_{\text{tot}} \) are the current densities of entropy and energy in the universe, \( m_p \) is the Planck mass, equal to \( h c / \sqrt{G N} \), and \( \langle \sigma v \rangle \) is the thermally averaged annihilation cross section of WIMP pairs multiplied by their relative velocity. In the equation that determines \( T_i \), this temperature appears in a Boltzmann factor \( e^{-m c^2 / k_B T_i} \), where \( m \) is the WIMP mass. Taking the logarithm, one finds \( k_B T_i / m c^2 \approx 1/25 \) for a wide range of values of the annihilation cross section.

With this parameter determined, we know all of the terms in (2) except for the value of the cross section, and so we can solve for this factor. The result is

\[
\langle \sigma v \rangle = 1 \text{ pb}
\]
A particle physicist would recognize this value as the typical size of the production cross sections expected for new particles at the LHC. More generally, if we assume that the coupling constant in the WIMP interactions is roughly same size as the dimensionless coupling $g$ that gives the strength of weak and electromagnetic interactions, this cross section results from an interaction mediated by a particle whose mass is of the order of 100 GeV.

This result is remarkable for two reasons. First, it allows us to transform our astrophysical knowledge of the cosmic density of dark matter into a prediction of the mass of the dark matter particle. Second, that prediction picks out a value of the mass that is very close to the mass scale associated with the Higgs boson and the symmetry breaking in the weak interactions. In an earlier article in this volume, Okada has argued that we should expect to find new elementary particles at that mass scale.\(^{14}\) Perhaps these new particles are in some way responsible for the dark matter.

In fact, explicit models of symmetry breaking in electroweak interactions often provide a natural setting for dark matter. Supersymmetry, discussed by Yamaguchi,\(^{15}\) predicts a new boson for each known fermion in Nature, and vice versa. It is natural that the fermionic partner of the photon is its own antiparticle, so that it is stable but annihilates in pairs. This particle is then a perfect candidate for the WIMP. Other models of electroweak symmetry breaking also contain new neutral weakly-interacting particles that can be made stable by natural symmetry principles.

### 4. Production and Detection of WIMPs at the LHC

If the mass of the WIMP should be about 100 GeV, we should be able to produce WIMPs if we can build an accelerator that provides elementary particle collisions at energies higher than 100 GeV. However, it is not so straightforward. A WIMP, being as weakly interacting as a neutrino, passes through a typical elementary particle detector unseen. It is only from the properties of the other particles produced in association with the WIMP that we can recognize these events and select them for analysis. Particle physicists have analyzed in some detail how to do this. Most of the specific analysis has been done in models of supersymmetry, so, for concreteness, I will use that picture here. The general conclusions apply to WIMPs in many other models of weak interaction symmetry breaking.

Supersymmetry predicts many new elementary particles in addition to the WIMP. In particular, the gluon of QCD has a fermionic partner, the gluino, and the quarks have bosonic partners, called squarks. Gluinos and squarks carry the same conserved quantum number that keeps the WIMP stable. They are expected to be heavier than the WIMP and to decay to the WIMP by emitting quarks, leptons, and Standard Model bosons. Events with squark or gluino pair production, then, will have a characteristic form. Many energetic quarks and leptons will be emitted, but also each event will end with the production of two WIMPs that make no signal in the detector. The observable particles in the event will show an imbalance of total momentum. The missing momentum is that carried off by the WIMPs.

We have not yet seen events of this type at currently operating accelerators. The highest-energy accelerator now operating is the Tevatron collider at Fermilab, and the experiments there put lower limits of about 300 GeV on the masses of gluinos and squarks.\(^{16}\) In 2008, however, the LHC will begin operation with proton-proton collisions at a center of mass energy of 14000 GeV. Not all of this energy is available for production of supersymmetry particles. The proton, after all, is a bound state of quarks and gluons. Gluinos and squarks are produced in collisions of individual quarks and gluons, which typically carry 10% or less of the total energy of the proton. Still, we expect to see collisions with total energy above 2000 GeV at a significant rate. This implies that squark and gluon pair production, leading to events with WIMPs, can be seen over almost all of the parameter space of the model. Figure 2 shows a simulation of a characteristic event of this type, as it would be observed by the CMS detector at the LHC.\(^{17}\)

### 5. Recognizing the Mass of the WIMP

The discovery of events at the LHC with apparent unbalanced momentum will signal that this accelerator is producing weakly interacting massive particles. However, it would be far from clear that this particle is the same one that is the dominant form of matter in the universe. To demonstrate this, we would need to correlate properties of the WIMP that we observe at the LHC with astrophysical observations. This will probably first be done through measurements of the mass of the dark matter particle. Using detailed measurements of the kinematics of quarks and leptons in the LHC events, it is expected that the mass of the WIMP produced there will be measured to 10% accuracy.\(^{18}\) We then must compare this value with measurements of the mass of the cosmic WIMP. To do this, it is necessary to detect the dark matter in the galaxy, not as a distribution of gravitating mass, but as individual particles.

There are two strategies to make this detection. The first, reviewed by Spooner,\(^{19}\) is to place very sensitive detectors in ultra-low background environments and look for rare events in which a WIMP in our cosmic neighborhood falls to earth and scatters from an atomic nucleus in the detector.
The cross section for this process is expected to have the remarkably small value of 1–10 zeptobarns, but in the next few years semiconductor and liquid noble gas detectors in deep mines are expected to reach this level of sensitivity. The mean energy deposited in these events depends on the WIMP mass \( m \) and the target nucleus mass \( m_T \) roughly as

\[
(E) = \frac{2v^2 m_T}{(1 + m_T/m)^2}.
\]

Then, for a 100 GeV WIMP, detection of 100 scattering events would lead to a mass determination at roughly 20% accuracy.\(^{20,21}\)

The second strategy is to look for WIMP annihilations in our galaxy. Although the density of WIMPs is sufficiently small that WIMPs cannot annihilate frequently enough to affect the overall mass density of the universe, WIMPs still should annihilate at a low rate, especially in places where their density is especially high. Astrophysicists understand the formation of galaxies and larger structures in the universe as arising from the clumping of dark matter as a result of its gravitational attraction. So our galaxy, and especially the center of the galaxy, should be a place with a relatively high density of WIMPs and thus a higher rate of WIMP annihilations. In an individual WIMP annihilation, the two WIMPs produce two showers of quarks, which are observed mainly as pions and photons. The pions and other charged particles are bent by the galactic magnetic field. But the photons, energetic gamma rays, fly outward in a straight line from the annihilation point. A gamma ray telescope can observe these photons and measure their energy spectrum. The spectral shape is characteristic, with a sharp cutoff in energy at the mass of the WIMP.\(^{22}\) The galaxy is expected to contain clumps of dark matter that should show up as spots bright in gamma rays with no counterpart in optical radiation. These spots should be intense enough to be seen with the gamma ray telescope satellite GLAST, and, if the WIMP mass is greater than several hundred GeV, by new ground-based gamma ray telescopes. Measurement of the endpoint of the energy spectrum should give a second astrophysical determination of the WIMP mass to 20% accuracy.

If the mass of the WIMP seen at the LHC is the same as the mass from astrophysical detection experiments, this will provide strong evidence that the LHC is producing the true particle of dark matter.

6. Predicting the Properties of the WIMP

To provide additional evidence on the identity of the WIMP observed at the LHC, we would like to assemble enough data about this particle to predict its pair annihilation cross section. From (2), we see that knowledge of this cross section from particle physics would give a prediction of the cosmic density of dark matter. It will be very interesting to compare that prediction to the value of the dark matter density obtained from cosmic microwave background measurements. Agreement of these values would not only confirm the identity of the WIMP. It would also verify the standard picture of the early universe up to the temperature \( T_i \), corresponding to a time in the early universe about \( 10^{-9} \) s after the Big Bang.

It is quite a challenge to predict the WIMP pair annihilation cross section. At the minimum, this requires measuring the masses and couplings of the heavier particles that are exchanged in the process of WIMP annihilation. In supersymmetry, WIMP annihilation is often dominated by the exchange of the bosonic partners of leptons, which must be identified through their decays to leptons and missing momentum. An alternative mechanism for WIMP annihilation is the exchange of the fermionic partners of the weak interaction bosons \( W \) and \( Z \). These cross sections depend sensitively on the mixing angles that determine the exact mass eigenstates of these particles. If several different reactions can contribute, the parameters of each must be measured or bounded.

Detailed studies of this program in a variety of supersymmetry models show that it requires more precise knowledge of the parameters of the model than can be obtained from the LHC. Fortunately, there is another technique for producing and studying new elementary particles that is capable of achieving higher precision. Electron-positron annihilation at high energy can create pairs of the new particles in a controlled setting, through reactions that are much simpler than those that we expect at the LHC. This process will be studied at the future electron-positron collider ILC discussed in the contribution of Yamamoto.\(^{23}\) A simulated supersymmetry production event at the ILC is shown in Fig. 3.

Once we have measured the masses of supersymmetric particles with high precision and also measured the cross sections that determine their couplings and mixing angles, we will be able to put forward a prediction of the cosmic dark matter density from particle physics data that can be compared to astrophysical measurements. Recently, Baltz, Battaglia, Wizansky, and I discussed quantitatively how accurate such microscopic predictions could be. Starting from a set of supersymmetry models with a variety of different mechanisms for WIMP annihilation, we analyzed the accuracy of measurements on supersymmetric particles that could realistically be expected from the LHC and from the ILC and derived from these the accuracy of the prediction to be expected for the dark matter density.\(^{25}\)

![Fig. 3. (Color online) Simulated ILC event, with pair production of the supersymmetric partners of \( W \) bosons and subsequent decay to quarks, leptons, and WIMPs.\(^{24}\) Only the visible products are shown in the figure.](image)
Figure 4 shows our results for two of these models, expressed as the likelihood distribution for $\Omega_{\text{DM}}$ predicted from the collider data that would be expected from LHC, from ILC measurements at the design energy of 500 GeV, and from an upgraded ILC running at an energy of 1000 GeV. Other groups have found similar results for first of these models. These predictions will be compared to the cosmic microwave background results from the next-generation experiments, which should determine $\Omega_{\text{DM}}$ to the percent level. It will take some time to collect all of the data required, but eventually we will have this sharp test of the WIMP identity of dark matter.

7. The WIMP Profile of the Galaxy

Once we have established the identity and properties of the WIMP, these results should feed back into astronomy. I noted in §2 that it is possible to detect dark matter on cluster scales and to map its distribution using gravitation lensing. However, for dark matter in the galaxy, the gravitational bending of light is not a large enough effect to provide structure information. To see where the dark matter is in our galaxy, we need to map dark matter particles. The distribution of dark matter in the galaxy is still mysterious, and in fact is one of the most controversial questions in astrophysics. In the cold dark matter model of structure formation, a galaxy as large as ours must be built from the assembly of smaller clusters of dark matter. The smaller clusters merge through their gravitational interaction, disrupt one another tidally, and eventually smooth out to form the halo of the galaxy. The time required for this evolution is on the order of the current age of the universe. Thus, most cold dark matter theories predict that the halo of the galaxy is inhomogeneous. A model of the density distribution of dark matter in a model galaxy, based on the clustering model of Taylor and Babul is shown in Fig. 5. An especially large clustering of dark matter should occur at the center of the galaxy. Some models predict caustics with large, almost singular dark matter densities; other models predict smoothing of the dark matter below some scale. Understanding the true situation will bring us closer to understanding how our galaxy and the others in the universe were born and evolved.

Figure 4 shows our results for two of these models, expressed as the likelihood distribution for $\Omega_{\text{DM}}$ predicted from the collider data that would be expected from LHC, from ILC measurements at the design energy of 500 GeV, and from an upgraded ILC running at an energy of 1000 GeV. Other groups have found similar results for first of these models. These predictions will be compared to the cosmic microwave background results from the next-generation experiments, which should determine $\Omega_{\text{DM}}$ to the percent level. It will take some time to collect all of the data required, but eventually we will have this sharp test of the WIMP identity of dark matter.

7. The WIMP Profile of the Galaxy

Once we have established the identity and properties of the WIMP, these results should feed back into astronomy. I noted in §2 that it is possible to detect dark matter on cluster scales and to map its distribution using gravitation lensing. However, for dark matter in the galaxy, the gravitational bending of light is not a large enough effect to provide structure information. To see where the dark matter is in our galaxy, we need to map dark matter particles. The distribution of dark matter in the galaxy is still mysterious, and in fact is one of the most controversial questions in astrophysics. In the cold dark matter model of structure formation, a galaxy as large as ours must be built from the assembly of smaller clusters of dark matter. The smaller clusters merge through their gravitational interaction, disrupt one another tidally, and eventually smooth out to form the halo of the galaxy. The time required for this evolution is on the order of the current age of the universe. Thus, most cold dark matter theories predict that the halo of the galaxy is inhomogeneous. A model of the density distribution of dark matter in a model galaxy, based on the clustering model of Taylor and Babul is shown in Fig. 5. An especially large clustering of dark matter should occur at the center of the galaxy. Some models predict caustics with large, almost singular dark matter densities; other models predict smoothing of the dark matter below some scale. Understanding the true situation will bring us closer to understanding how our galaxy and the others in the universe were born and evolved.

The determination of the properties of the dark matter particle will give us the information that is needed to predict
the interaction rates of dark matter particles with ordinary matter and with one another. This in turn will allow us to interpret detection signals in terms of the absolute density of dark matter both here and elsewhere in the galaxy. By dividing the underground detection rate for dark matter by the interaction cross section determined from collider data, we will be able to measure the absolute flux of dark matter at our position in the galaxy. By measuring the luminosity of clumps of dark matter in the galaxy and dividing by the dark matter annihilation cross section determined from collider data, we will be able to map at least the largest clumps of dark matter in terms of their absolute density.

8. Conclusions

Today, dark matter is one of the great mysteries of physics and astronomy. But I have argued in this article that the time is approaching for its solution. I have motivated the idea that dark matter is composed of a new elementary particle, the WIMP, whose mass is about 100 GeV. If this is true, then over the next five years we should produce the WIMP at the LHC, and we should see signals of astrophysical WIMPs in several different detection experiments. This will set in motion a campaign to determine the properties of dark matter by measurements both in high-energy collider experiments and through mapping of astrophysical signals. Over the next fifteen years, we will learn the story of this major constituent of the universe, its identity, its properties, and its role in our cosmic origin.

Acknowledgment

I am grateful to Marco Battaglia, Ted Baltz, Jonathan Feng, Mark Trodden, Tommer Wizansky, and many others whose insights were essential in creating the picture explained here. I thank David Griffiths for instructive comments on the manuscript. This work was supported by the U.S. Department of Energy under contract DE-AC02-76SF00515.

17) http://cmsinfo.cern.ch/outreach/CMSdetectorInfo/NewPhysics/
24) I thank N. Graf for providing this figure.
30) I thank E. A. Baltz for providing this figure.

Michael E. Peskin was born in Philadelphia, Pennsylvania, U.S.A., in 1951. He obtained his A.B. (1973) from Harvard University and his Ph. D. (1978) from Cornell University. After holding postdoctoral positions at Harvard, CEN Saclay, and Cornell, he moved in 1982 to the Stanford Linear Accelerator Center where he is now a Professor of Theoretical Physics. He has worked on a wide variety of problems in theoretical high-energy physics, from formal aspects of lattice gauge theory and string theory to the phenomenology of technicolor, supersymmetric, and extra-dimensional models of physics beyond the Standard Model. With Daniel Schroeder, he has written the widely-used textbook An Introduction to Quantum Field Theory.
Generation of Cosmological Baryon Asymmetry  
—(B–L)-Genesis—

Motohiko YOSHIMURA*  

Department of Physics, Okayama University, 3-1-1 Tsushima-naka, Okayama 700-8530  
(Received February 2, 2007; accepted February 19, 2007; published November 12, 2007)  

The advent of the standard gauge theory of particle physics is a story of remarkable success. Yet it is the view of most particle physicists that it is not the end for the quest of the ultimate microscopic world. Direct experimental hints for physics beyond the standard theory have recently been provided by the discovery of neutrino oscillation. On the theoretical side, there are many unsolved problems, one of which we will explain in this article.

KEYWORDS: baryogenesis, leptogenesis, grand unified theory, cosmology, Majorana neutrino  
DOI: 10.1143/JPSJ.76.111018

Problem  
The big bang cosmology has undergone enormous developments via varieties of precision measurements. Along with a negative search, in our vicinity of the universe, for anti-matter (such as $\text{He}^-$ and annihilation $\gamma$ from $\pi^0$ decay of $\bar{NN}$ annihilation), the result pertaining to the matter content is summarized by saying that the present universe is baryon dominated and that the degree of the asymmetry is given by the important quantity, the baryon to the photon ratio, $n_B/n_\gamma \sim 6 \times 10^{-10}$. This value is consistent with the old value provided by comparing cosmic abundance observations with the calculation of the nucleosynthesis theory that predicts the primordial abundance of the elements $\text{He}$, $\text{D}$, $\text{Li}$.

With the expansion of the cosmic scale factor, this implies that the early universe prior to nucleosynthesis must have been baryon asymmetric; one excess out of $\approx 10^{10}$ pairs of nucleon + anti-nucleon created the present baryon dominated universe. As will be described shortly, this number is not small from the physics viewpoint, and it is a number physicists should explain.

Early history of baryogenesis theory in brief  
Immediately after the theoretical revolution related to the standard model in the early 70’s, there had been intriguing proposals of grand unified theories (GUT)$^2$ based on the gauge principle. These theories predict baryon number non-conservation, which was believed otherwise until those days. Proponents of these theories asserted that the gauge symmetry and its associated conservation law alone are exactly preserved, and other empirical conservation laws are doomed to be violated, although they may be suppressed by some inverse power of a new physics scale. Electric charge conservation belongs to the sacred class, but both the baryon and lepton numbers may be violated. To these proponents it is only the magnitude of violation that is questioned.

Experimental evidence confirming the detailed structure of the electroweak theory, one of the backbones of the standard theory, has accumulated towards 1978. It was in this atmosphere that the first idea$^3,4$ of applying baryon non-conservation to cosmology was put forward to explain the observed baryon asymmetry of the universe. Immediately after this suggestion, there appeared many alternative scenarios such as evaporating black hole as the seed of the asymmetry generation. It is particularly appropriate to mention that the possibility of baryogenesis at high temperatures in the electroweak theory was proposed$^5$ among these alternatives.

Basic ingredients  
Three necessary conditions for baryogenesis were written earlier.$^6$ These are (1) baryon number non-conservation, (2) C and CP violation, (3) departure from thermal and chemical equilibria. Intricate interplay of these conditions has been appreciated only much later.$^7$ Furthermore, the first condition was replaced by $B-L$ non-conservation, as will be explained later.

In most sensible extended unified theories, it is automatic to expect the first two conditions. Let us take an example of a heavy X-boson that violates the baryon number by having decay modes of different baryon numbers. Typically, a colored X-boson can simultaneously decay into two quarks and an anti-quark + anti-lepton, without violating the $SU(3)$ color gauge symmetry:

$$X \rightarrow q + \bar{q}, \quad X \rightarrow \bar{q} + q.$$  

(1)

with their respective rates $\gamma_q$ and $\gamma_{\bar{q}}$. The coexistence of these two channels that differ by 1 in the baryon number of final states is the origin of baryon non-conservation. Matter instability occurs due to the exchange of X-bosons, since the two quarks $u\bar{u}$ in proton can annihilate into $e^+\bar{d}$, making the decay $p \rightarrow e^+\bar{d}$ possible.

Suppose that equal amounts of the pair, $X$ and its anti-particle $\bar{X}$, existed in the early universe. When this pair decays independently, a non-vanishing baryon number $\Delta B$ may be created according to the rate formula,

$$\frac{d\Delta B}{dr} = \frac{2}{3} \gamma_q - \frac{1}{3} \gamma_{\bar{q}} - \frac{2}{3} \gamma_{\bar{q}} + \frac{1}{3} \gamma_q.$$  

(2)

On the other hand, the CPT theorem, an exact consequence of the local field theory, tells that the total decay rate $\gamma_{\bar{q}}$ is equal for particle $X$ and its anti-particle $\bar{X}$. In our context, this means that $\gamma_q = \gamma_{\bar{q}} + \gamma_{\bar{q}} = \gamma_q + \gamma_{\bar{q}}$. Hence $d\Delta B/dr = \gamma_q - \gamma_{\bar{q}}$. In other words, a non-vanishing asymmetry is generated if $\gamma_q \neq \gamma_{\bar{q}}$. Despite the equality of the total decay

---

*E-mail: yoshim@fphy.hep.okayama-u.ac.jp
rate, the partial decay rates of \( X \) and \( \tilde{X} \), \( \gamma_q \) and \( \tilde{\gamma}_q \), can differ when the C and CP symmetry is violated.

The third condition, the need for the arrow of time, also seems obvious, but its relation to unified theories is rather non-trivial. In thermal and chemical equilibria a process that can generate baryon asymmetry is precisely balanced by the inverse process that eliminates this asymmetry. The only possible way for finite asymmetry generation is to block the inverse process energetically. Suppose that this is realized, because when the decay of heavy X-bosons occurs, the universe is sufficiently cool to recreate the massive X-boson. Denoting the temperature of early hot universe by \( T \), this condition is expressed as

\[
T < m_X, \quad \text{at } H(T) = \gamma_X (\approx \alpha_X m_X),
\]

where \( H(T) \) is the Hubble expansion rate \( \approx \sqrt{N T^2/m_p} \), with \( m_p = 1/\sqrt{G} \approx 10^{19} \text{ GeV} \) the gravitational energy scale and \( N \) the number of particle species contributing to the energy density of the hot universe of \( O[100-200] \). We used the natural unit \( h = c = 1 \) along with the unit Boltzmann constant \( k_B = 1 \). The decay rate \( \gamma_X \) has a weak temperature dependence \( \propto T^3 \) due to the dilution effect, which was ignored here. The constant \( \alpha_X \) is the unified coupling constant at the energy scale of grand unification, and is typically \( O(1/45) \).

This out-of-equilibrium condition leads to a lower bound on the X-boson mass, \( m_X > O(\alpha_X m_p/\sqrt{N}) \). An experimental indication of coupling unification in supersymmetric theories is roughly consistent with this large unification mass scale.\(^{31}\)

An actual computation of baryon to photon ratio \( n_B/n_\gamma \) involves a numerical integration of a simplified set of Boltzmann equations for \( Y_i = n_i/n_\gamma \), the respective number densities \( n_i \) of \( X, \tilde{X}, B, \tilde{B} \) divided by the photon number density.\(^{9} \) These quantities are invariant with cosmic expansion unless physical processes change them, and for the baryon to the photon ratio \( Y_B = n_B/n_\gamma \),

\[
\frac{dY_B}{dt} = e\gamma_X (Y_+ - Y_B) + \gamma_X \delta Y_-, \quad (4)
\]

\( Y_+ = n_+ - n_B/n_\gamma \) with \( n_\pm = n_X \pm n_{\tilde{X}} \) obeying similar equations. Here, \( Y_B = \gamma_X(n_B/n_\gamma) \) refers to the thermal value. The fundamental asymmetry \( \epsilon \) is the baryon number created by the decay of the pair \( X \) and \( \tilde{X} \); \( \epsilon = (\gamma_X - \tilde{\gamma}_X)/(\gamma_q + \tilde{\gamma}_q) \), and \( \delta \) is a known quantity of the order of unity. The outcome of numerical integration may be summarized as

\[
\frac{n_B}{s} = O[10^{-2}] \frac{\epsilon}{1 + (16K)^{1/3}},
\]

where \( K = \gamma_X/H(m_X) \) is the ratio of the X-decay rate to the Hubble rate at the temperature \( m_X \), and the entropy density \( s \sim 7 n_\gamma \) in the present universe.

The important asymmetry parameter \( \epsilon \) that appears in this equation can be calculated using a perturbation theory of the grand unified theory. It is the sum over the phase space of differences in the rates of the form,

\[
\epsilon \propto \sum |g_1 f_1 + g_2 f_2 + \cdots|^2 - |g_1 f_1 + g_2 f_2 + \cdots|^2
\]

\[
\propto -4\alpha g_2 \alpha(f_1 f_2^*) + \cdots, \quad (6)
\]

where \( g_i \) is the product of coupling constants in basic Lagrangian, and \( f_i \) the amplitude determined by dynamics. The computed asymmetry is thus necessarily small due to higher order effects of interference terms of different perturbative orders.

Other important basic ingredients of baryogenesis have been summarized in my early review article.\(^{7} \)

**Broader perspective** Since the original suggestion of GUT baryogenesis,\(^3 \) many interesting works that bridge microphysics and the universe have appeared. Among them, the inflationary universe scenario\(^{10} \) became a paradigm of the modern time. Inflation has also given an important bonus to baryogenesis; the initial condition of the vanishing baryon number is automatically satisfied in an inflationary universe.

Inflation essentially depletes all conserved and non-conserved quantum numbers including the baryon number. What it left behind is a cold oscillation of the inflaton field. The hot big bang that has connection with our present universe is created when the inflaton field decays, and decay products interact fast enough to thermalize. How hot it must be after inflation will be discussed later with regard to the problem of gravitino overproduction.

The basic picture of inflation has been confirmed by recent observations including COBE and WMAP. They also posited a big condumnd of the dark energy, which faces the fundamental physics in the coming years.

**Electroweak baryon non-conservation** 't Hooft in his influential paper\(^{11} \) pointed out among other things that the baryon number may be violated in the standard electroweak theory. This is due to tunneling effect, which is suppressed by a factor like \( e^{-1/\alpha} \) with \( \alpha \sim 1/137 \), and hence completely negligible under a normal circumstance.

Already in their 1978 paper\(^5 \) Dimopoulos and Susskind suggested that this suppression might be removed at high temperatures of the early universe by barrier crossing instead of tunneling. The necessary ingredients for electroweak baryon non-conservations have steadily accumulated; (1) discovery of sphaleron solution\(^{12} \) which is an unstable object of half a baryon number made of Higgs and gauge fields, and has a mass of \( O[1 \text{ TeV}] \), (2) analysis of quark and lepton propagation under a non-trivial field configuration of sperhaleons and alike,\(^{13} \) and (3) the mechanism of how the baryon number is violated at high temperatures.\(^{14} \)

A concrete scenario of electroweak baryogenesis has been proposed.\(^{15} \) It utilizes a non-equilibrium first-order electroweak phase transition. This scenario has been criticized.\(^{16} \) Even disregarding the likely possibility of the second-order phase transition, a correct treatment gives an estimated \( n_B/n_\gamma \) as a combination of CP violation (CKM) parameters, which is much smaller than the observed \( n_B/n_\gamma \). It is our consensus that the standard theory cannot explain the problem of baryon asymmetry.

Electroweak processes, both at low and high temperatures, exactly conserve \( B - L \), the baryon minus lepton number. Although unsuccessful in generating baryon asymmetry, electroweak processes at high temperatures are very efficient in redistributing already existing \( B - L \) asymmetry to individual thermal \( B \) and \( L \) values. Assuming \( B - L \) generation at higher temperatures, the electroweak process gives finite baryon and lepton numbers according to the physics of thermal and chemical equilibria;
\[ B = a \times \Delta(B - L), \quad a = \frac{8n_g + 4n_H}{22n_g + 13n_H} = \frac{28}{79}. \]  

Here \( n_g = 3 \) is the number of generationa or families and \( n_H = 1 \) the number of Higgs doublets. \( B - L \) generation of amount \( \Delta(B - L) \) prior to electroweak baryon non-conserving processes thus becomes the prime objective of the microscopic theory irrespective of whether the first generation is \( B \) or \( L \).

**Leptogenesis**  
Baryogenesis according to GUT, however attractive it may appear, has an important missing element; observation of proton decay. On the other hand, leptogenesis\(^{17}\) is attractive since it might be related to the observed finiteness of neutrino masses confirmed by neutrino oscillation experiments. According to oscillation parameters, neutrino masses are fit by a formula of the seesaw mechanism,\(^{19}\) \( m_\nu = m^2_{\nu_i}/\mathcal{M} \), where \( m = O(100 \text{GeV}) \) is a typical quark and charged lepton mass, and \( \mathcal{M} \) is a new physics mass scale such as GUT.

Neutrinos are unique in the sense that they alone are charge-neutral fundamental fermions unlike all other quarks and leptons. Moreover, they participate in the weak interaction with a definite handedness; they have the left-handed chirality. This disparity between the left and the right is presumably the most important hint towards further unification. It is true that without interaction all charged fermions are described by the 4-component Dirac spinor. It is however possible to describe neutral fermions, even massive ones, by the 2-component spinor, and this is exactly what was proposed by Majorana, a long time ago in the 1930’s.

The Majorana particle without interaction is described by the Majorana equation,

\[ (i\partial - im \cdot \nabla)\psi = im\sigma_2\psi^*, \]  

where \( m \) is the neutrino mass. The most salient feature of this Majorana equation is that it contains \( m \) as well as its conjugate \( \psi^* \), thus the lepton number is not conserved, or more properly, one cannot define the lepton number. In the conventional 4-component (\( \psi \)) description the 2-component \( \psi = (1 - \gamma_5)\psi/2 \) is the left-handed chiral projection.

The right-handed partner of the left-handed neutrino, the other chiral projection, \( (1 + \gamma_5)\psi/2 \) may or may not exist in nature; indeed some GUT theory like the one based on the \( SU(5) \) gauge group lacks this partner. But most other theories such as those based on \( SO(10) \) have right-handed partners. The important point is that they are insensitive to quantum numbers of the low energy gauge symmetry \( SU(3) \times SU(2) \times U(1) \). It is then natural to expect that right-handed partners have masses, much larger than the electroweak scale of \( O(100 \text{GeV}) \) and necessarily of the Majorana type.

It is customary to write ordinary neutrinos by \( \nu_L \) and their chiral partners by \( N_R \). In the Lagrangian field theory, they are described by the diagonal Majorana \( N_R \) mass \( M \) and the off-diagonal Dirac mixing mass \( m_L \) as

\[ \frac{1}{2} (\nu_L N_R)^T i\sigma_2 \begin{pmatrix} m_L & m^2_{\nu} \end{pmatrix} \begin{pmatrix} \nu_L \end{pmatrix} \begin{pmatrix} N_R \end{pmatrix} + \text{h.c.}. \]  

Diagonalization leads to two Majorana types of masses of the order, \( M \) and \( m_L^2/M \).

The Majorana type neutral lepton \( N_R \) is a promising agent for generating cosmological lepton asymmetry. This replaces the role of X-boson in GUT baryogenesis. The asymmetry computation is quite similar to that in the GUT scenario.\(^{18}\)

**Problem to solve**  
Both X-boson and \( N_R \) lepton are very massive, much beyond ordinary electroweak energy scale. To create them with a large quantity in the early universe means that the universe must have been very hot right after inflation; the temperature after inflation \( T_R \gg M \), with \( M \) as the heavy particle mass.

One serious problem against a very hot universe exists; gravitino overproduction. In supersymmetric (SUSY) theories, the super-partner of the graviton has the spin 3/2 and is called gravitino denoted by \( g_{3/2} \). In popular SUSY models, the gravitino mass is of the order \( (0.1-1) \text{TeV} \), reflecting SUSY breaking at the TeV scale. The gravitino decays at the rate \( Gm^3_{3/2}/m^2_{3/2}/G \) (being the gravitational constant). Numerically, this lifetime is \( O(10^6) \sec(1 \text{TeV}/m_{3/2}) \). It is sufficiently long to cause concern over the destruction of light elements synthesized at \( \approx 3 \text{ min} \) after the big bang.

Gravitinos may be produced in the hot universe right after inflation. Its production rate is again dictated by the gravitational strength, and is compared with the Hubble rate \( \sqrt{N}T^2/m_{3/2} \). The produced thermal abundance relative to the photon density is then \( \approx T/(\sqrt{N}m_{3/2}) \). This abundance immediately tells how much entropy may be created when gravitinos decay.

The gravitino abundance is limited by the allowed destruction of light elements that proceeds by the quark jet produced by the gravitino decay.\(^{20}\) This argument gives an allowed maximum temperature after inflation of the order \( (10^6-10^8)\text{GeV} \). This appears too small compared with the \( N_R \) Majorana mass required by a detailed analysis of leptogenesis computation.\(^{18}\)

There are ways to overcome this difficulty; for instance, an entirely different baryogenesis idea exists that does not assume a very high temperature, known as the Affleck–Dine mechanism.\(^{21}\)

**Looking ahead for further experimental evidence**  
Whether it is B-generation or L-generation, it is important to verify their fundamental ingredients by experiments. We shall discuss experimental implications of L-generation and describe new types of experimental approach.\(^{22,23}\)

Leptogenesis requires \( 1 \) lepton number non-conservation along with the Majorana nature, (2) a new source of CP violation that is not related to a neutrino oscillation experiment. Examples of lepton number non-conserving processes are

\[ 2X \rightarrow 2 \gamma^+ + e^- + e^-, \quad 2X \rightarrow 2 \gamma^+ Y + e^+ + e^-. \]  

The first one is the neutrinoless double beta decay, the prime ongoing experimental target, and the second \( e^- \rightarrow e^+ \) conversion via laser irradiation has been discussed recently.\(^{22}\) These processes clearly violate the lepton number by 2 units, and it is natural to expect that their transition amplitudes should be proportional to the Majorana neutrino mass, or their weighted average of 3 mass eigenvalues. But the Majorana nature is conceptually independent of lepton

\[ 111018-3 \]
number non-conservation, and indeed it is possible to have other sources of lepton number non-conservation that give rise to eq. (10), although the main term of violation in many GUT models tends to be given by the Majorana mass term.

Whether neutral fermions belong to the Majorana class or the Dirac class is one of the most important questions facing fundamental science, much more important than any particular particle theory. A direct method of distinguishing the Majorana neutrino from the Dirac neutrino has been worked out recently. The Majorana nature may show up only in this way can one distinguish the Majorana particle from the Dirac particle. There are only few processes of this nature; neutrino pair emission from excited atoms.

A variety of energy differences between atomic levels are ideal for the determination of small neutrino masses indicated by the neutrino oscillation experiment. If a mass hierarchy is assumed, these masses are $\leq 50$ meV. Let us discuss how atomic transitions via a neutrino pair $\nu_1 \nu_2$ helps in the determination of the Majorana nature. The neutrino pair emission current (in the charge retention order after the Fierz transformation) is given as

$$\mathcal{F}(p_1, h_1, p_2, h_2) = -i \sqrt{\frac{E_2 - h_2}{E_2 + h_2}} u'(p_1, h_1) \sigma^2 u'(p_2, -h_2)$$

where $u(p, h)$’s are 2-component spinor solutions of the Majorana equation, of momentum $p$ and helicity $h$. The pair emission rate is proportional to $|\mathcal{F}(p_1, h_1, p_2, h_2)|^2$, the pair current squared, and this should be summed over the phase space. The interference of two terms in eq. (11) is due to two identical Majorana fermions, which is missing in the Dirac pair emission. It can be proved that the interference rate $\propto m_1 m_2$, the product of paired neutrino masses.

We hope that varieties of atomic transitions, when the rate is enhanced by laser or microwave irradiation, will open a new experimental method of mass spectroscopy. The method works both for the Majorana and the Dirac cases.


Motofuko Yoshimura was born in Shanghai, China in 1942. He obtained his B.S (1965) and M.S. (1967) degrees from University of Tokyo (UT), and his Ph. D (1971) from University of Chicago. He was a research associate at University of California, Berkeley (1970–1972), at University of Pennsylvania (1972–1974), and at Ecole Normale Superieure, Paris (1974–1975). He joined the faculty member at Tohoku University (TU) from 1975 and moved to National Laboratory for High Energy Physics as an associate professor in 1979 and was a professor since 1983. After serving as a professor at TU during 1988–2001, he became a professor and director of Institute for Cosmic Ray Research at TU during 2001–2004. After retirement from TU, he has been a professor at Okayama University since 2004. His research interest covers particle physics-phenomenology of the standard model and beyond that, and various areas of cosmology. He has initiated application of grand unified theories to the early universe and proposed a scenario of baryogenesis in 1978.
Frontiers of Elementary Particle Physics, the Standard Model and Beyond

Lattice QCD

Paul MACKENZIE*

Fermi National Accelerator Laboratory, Batavia, IL 60510, U.S.A.

(Received June 20, 2007; accepted July 23, 2007; published November 12, 2007)

Modern lattice gauge theory calculations are making it possible for lattice QCD to play an increasingly important role in the quantitative investigation of the Standard Model. The fact that QCD is strongly coupled at large distances has required the development of nonperturbative methods and large-scale computer simulations to solve the theory. The development of successful numerical methods for QCD calculations puts us in a good position to be ready for the possible discovery of new strongly coupled forces beyond the Standard Model in the era of the Large Hadron Collider.

KEYWORDS: quantum chromodynamics, QCD, lattice gauge theory, quarks, gluons, strong interactions
DOI: 10.1143/JPSJ.76.111019

1. The Standard Model and Beyond

The Standard Model is amazingly successful (maddeningly so), yet its many gaps and puzzles make it clear that it is simply the low energy manifestation of new, higher-energy physics yet to be discovered. Its many parameters are simply arbitrary, having their observed values as a result of as yet undiscovered physics at higher energies than obtained up until now. Its three similar generations of fermions and three similar forces are undoubtedly related in some way yet to be discovered. To help understand what lies beyond the Standard Model, the fundamental parameters of the Standard Model are being pinned down with greater and greater precision in heavy flavor experiments around the world. New particles and forces are being sought in very high energy experiments at the Tevatron and will be soon at the Large Hadron Collider (LHC).

Lattice calculations are essential to this program in two ways. First, they are required to extract properties of quarks from properties of hadrons (particles that contain quarks). Unlike leptons, such as the electron or neutrino, quarks cannot be observed directly, but are confined permanently within hadrons. Their properties must be inferred using lattice gauge theory calculations.

Secondly, lattice gauge theory calculations are essential to prepare for possible new nonperturbative phenomena in coming experiments. Lattice gauge theory is the first and only general tool for solving nonperturbative quantum field theories. Of the four interactions known to particle physics, only one (quantum electrodynamics) is known to be described by a perturbative theory, whose properties can be expressed as a power series in the electromagnetic coupling constant, \( \alpha_{\text{em}} \), over all energy scales. Strong interactions are known to be described by a nonperturbative theory, quantum chromodynamics or QCD. String theory, the current best candidate for a theory of gravity, must have nonperturbative effects in it, or it would produce a space-time quite different from the four dimensional world that we live in. In the theory of the weak interactions, consider the “Higgs”, the particle that generates particle masses. Is it

- an elementary, perturbative Higgs?
- a bound state of a new strong interactions (technicolor, topcolor, \ldots)?)?
- accompanied by very high energy gluino condensates (as in many models of supersymmetry with strongly coupled sectors)?

It is likely that whatever new physics is discovered by the LHC, it will contain some nonperturbative effects. QCD is providing an excellent test bed to sharpen our nonperturbative tools to prepare for such questions.

2. Quarks, Gluons, and Lattice QCD

Asymptotic freedom and quark confinement. In the early 60s, the classification properties of the observed hadrons led Gell-Mann and Zweig to note that the hadrons were arranged in multiplets as if they were composed of smaller particles, which Gell-Mann called quarks. In the same decade, deep inelastic scattering experiments at SLAC showed that in high energy electron–proton collisions, protons behaved as if they were composed of weakly interacting, almost-free constituents. Bjorken and Feynman called these entities partons. It was not immediately clear whether to regard quarks as actual particles, or whether they were merely a convenient classification tool. Furthermore, no one had ever seen a quark, so they seemed to be strongly confined inside hadrons. This seemed inconsistent with the weakly coupled nature observed in partons, so the relation between quarks and partons was not clear. Why should such almost-free constituents be permanently confined?

This paradox was resolved in 1973 with the discovery of the “asymptotic freedom” of QCD. The self-coupling of the gluons mediating the strong force caused the effective value of the strong coupling “constant” to become larger and larger at long distances (long compared with the proton radius), contrary to the well-known behavior of the electromagnetic coupling constant. This meant that even though the quarks were indeed weakly interacting at short distances and high energies, the force between them did not die off at long distances, leading to their permanent confinement. Gross, Politzer, and Wilczek shared the 2004 Nobel Prize for this discovery.

The consequence for particle physics is that, even though

*E-mail: mackenzie@fnal.gov
perturbation theory may be used to analyze quark-quark scattering at high energies, to infer the properties of quarks from the relatively low energy dynamics of hadron constituents, the nonperturbative methods of lattice QCD are required.

**Lattice gauge theory calculations.** Quantum field theories are defined by their path integrals. For gauge theory, this may be written schematically as

\[ Z = \int \mathcal{D}[A, \psi, \bar{\psi}] \exp(-S(A, \psi, \bar{\psi})) , \]  

(1)

where \( A \) and \( \psi \) are the field variables of the gluons and quarks, and \( S \) is the classical action of the theory. The quantum amplitude for a state of quarks and gluons at a given time to evolve into another state at a later time is obtained by integrating over all possible intervening classical field configurations. In principle, one integrates over independent fields defined at each space-time point. A quantum field theory is in principle defined by an infinite dimensional integral (not a very well-defined object). Quantum field theories must therefore be “regulated”.

A lattice quantum field theory regulates the continuum theory by defining the fields on a four-dimensional space-time lattice. Quarks are defined on the sites of the lattice, and gluons on the links. Continuum quantum field theory is obtained in the zero lattice spacing limit. This limit is computationally very expensive, which is why large-scale computer simulations are required.

Operationally, lattice QCD calculations consist of several steps. First, sets of gauge configurations are computed that form a representative sample of the infinite set of possible configurations. They are constructed in long Markov chains with Monte Carlo methods, such as the venerable Metropolis method, or the more modern hybrid Monte Carlo algorithm. Configurations are accumulated at several lattice spacings, and at several values of the masses of the light quarks in the fermi sea, which are heavier than the physical light quark masses. Final physics results must ultimately be extrapolated to the continuum and light quark mass limits.

Second, the propagation of quarks through the gauge configurations is calculated. This means solving the Dirac equation on each gauge configuration. On the lattice, this is a sparse-matrix problem, solved with relaxation methods, such as the biconjugate gradient algorithm. This step can consume compute power that approaches that of the first step if many different physical processes are analyzed.

Third, hadron correlation functions and amplitudes are computed from the quark propagators. This is a computationally cheap step, consisting mostly of I/O.

State-of-the-art price/performance for computing hardware for this type of calculation is currently under $1/MF. Larger projects are of order a few Teraflop-years. (That is, computing power of several delivered Teraflops, dedicated for a year.) Many types of large computers are used in lattice calculations,\textsuperscript{1} such as the purpose-built QCDOC at Brookhaven National Laboratory (Fig. 1), large clusters of commodity computers such as the ones at Fermilab (Fig. 2) and many other places, and the IBM Blue Gene at KEK (Fig. 3), currently the largest computer in the world predominantly dedicated to lattice QCD calculations.

Progress in numerical science comes from both larger computers and from improvement of methods. A methodological improvement that has been particularly important for the work I will discuss is improved discretization. Numerical analysis tells us that if a derivative is approximated by a discrete difference, the resulting discretization errors vanish as the square of the lattice spacing.
\[ \frac{\partial \psi(x_i)}{\partial x} = \Delta_x \psi(x_i) + \mathcal{O}(a^3), \]

where \[ \Delta_x \psi(x_i) \equiv \left[ \psi(x_i + a) - \psi(x_i - a) \right]/(2a). \] By incorporating next-to-nearest neighbor interactions, we can write down an approximation to the derivative whose errors vanish as a higher power of the lattice spacing:

\[ \frac{\partial \psi(x_i)}{\partial x} = \Delta_x \psi(x_i) - \frac{a^2}{6} \Delta_x^3 \psi + \mathcal{O}(a^4). \]

This allows control of discretization errors with far less computing power than the simpler derivative.

It is relatively unambiguous how to remove the \( \mathcal{O}(a^2) \) errors in the gluon action, and the various improvements in use are closely related to each other. The situation is dramatically different with lattice fermions. There are several families of discretization methods, that each have very different virtues and drawbacks. Staggered fermions can be calculated much more rapidly than the other methods. They have therefore been the first to produce reasonably precise unquenched results, and many of the results in Standard Model phenomenology in the next section use them. They have some ugly theoretical properties, however, that lead some physicists to look at alternatives. Wilson fermions were the original fermions used introduced by Ken Wilson. They break chiral symmetry badly, and for that reason have had trouble getting to quark masses as light as the ones in nature. Recent algorithmic advances have altered this situation much for the better. Domain wall fermions and overlap fermions have the nicest theoretical properties of all. They do not suffer the complications of staggered fermions, and have clean chiral structure unlike Wilson fermions. In the past, they have been by far the most expensive with which to calculate all of the methods, so phenomenological calculations are just beginning. Rapid algorithmic advance in the last few years have greatly sped up all the fermion methods, and it is not known at present which of these methods will ultimately prove superior. At present, lattice theorists around the world are hard at work on all of them, making sure that all methods give the same physical answers.

3. Lattice QCD Confronts Experiment

Progress in unquenched lattice QCD. In comparing QCD with experiment, we have several different types of tasks. One is comparing the results of QCD calculations with known experimental results and verifying that we can reproduce experiment. Since QCD is by now a very solidly established theory, this serves more to verify lattice gauge theory methods, rather than to test QCD. A second task is, where possible, to make predictions of physical results before they have been determined by experiment. Since hadron physics has being going on for decades, it is rare when opportunities can be found, but a few have been. A third, different type of task is to use verified lattice methods, combined with experiment, to obtain results that cannot be obtained by experiment alone. Examples of these include the quark masses and the strength of transformations of one quark to another under weak interactions (the Cabibbo–Kobayashi–Maskawa, or CKM, matrix elements). Since only hadrons are observed in experiment, and never quarks, these cannot be directly determined from experiment. Instead, quark masses and CKM matrix elements are used as parameters in lattice calculations and chosen so that the results of the calculations agree with experiment. These quark masses and CKM matrix elements are among the fundamental parameters of the Standard Model that must ultimately be postdicted by future Beyond-the-Standard-Model theories. Their determination is one of the most important tasks of lattice QCD as far as particle physics as a whole is concerned.

In the last few years, there has been dramatic progress in our ability to perform precise calculations of simple quantities. For twenty years after the first lattice Monte Carlo calculations appeared around 1980, almost all lattice QCD phenomenology was done in the quenched theory, meaning ignoring the effects of light quark-antiquark pairs. Although computationally much cheaper than correct unquenched calculations, this method introduces errors of unknown size into the results. The left-hand graph of Fig. 4 shows various combinations of particle masses and decay constants calculated in the quenched theory and shows around 10% discrepancies with experiment. There is no theory, however, that allows us to estimate in advance what size errors quenching introduces for any given quantity. Great strides, however, have been made in methods, in algorithms, and in computational power. They have now brought us to an era when unquenched calculations are becoming the norm, with all three light flavors of quark, \( u, d, \) and \( s \), included dynamically.

The calculations in the graph of Fig. 4 were performed with improved staggered fermions (called “asqtad” fermions in the jargon). The right-hand graph of Fig. 4 shows the same quantities as on the left, but unquenched and now showing good agreement with experiment at the few percent level. For these calculations, the masses of some quantities...
like the pion and kaon masses are used as inputs to fix the fundamental parameters of QCD, the quark masses and the strong coupling constant. Three different groups using this method, Fermilab, MILC, and HPQCD, then compared notes on their predictions for the simplest quantities they were calculating, with the results shown. These results are for the simplest physical quantities we know how to calculate in lattice QCD. The calculations are now being extended to more and more complicated quantities. Likewise, the results shown are obtained with staggered fermions, the least computationally costly of the fermion methods, and it will be interesting (and necessary) to verify that one obtains the same answers with more costly methods.

The prediction of a particle mass: the $B_c$. Most of the particle masses and other simple quantities that are to be "predicted" by lattice QCD have been well known for fifty years, so that only postdiction is possible. An exception has been the mass of the $B_c$ meson, a meson made of a bottom quark and a charm antiquark. Bottom quarks were discovered only in the 1970’s, and since they are rarely produced in association with charm quarks, $B_c$ mesons had not been observed as of a few years ago. Figure 5 shows the predictions of unquenched lattice calculations, before the observation of the $B_c$. In December of 2004, the CDF experiment at Fermilab announced the discovery of the $B_c$. Their result for the mass is shown in the gold bar across the graph, in good agreement with the lattice prediction.

The strong coupling constant. The effective coupling "constant", $\alpha_s(E)$, governs the strength of the strong interactions of QCD. It is one of the best measured parameters of QCD. Asymptotic freedom means that $\alpha_s(E)$ is small in collisions at high energy, $E$. Therefor, perturbation theory can be used to analyze high energy collisions in terms of a power series in $\alpha_s(E)$. The strong coupling constant can be measured in a large number of high-energy processes, some of which are shown in the plot in Fig. 6. One can also obtain the strong coupling constant with lattice methods. One should obtain the same results if lattice methods are correct. One obtains $\alpha_s$ on the lattice by using it as a parameter in particle spectroscopy calculations, as in ref. 12. One then uses perturbation theory to convert the lattice coupling constant to the form used in conventional continuum perturbation theory analyses. The result is shown in the next-to-bottom point in Fig. 6. It incorporates three-loop lattice perturbation theory. It agrees well with the world average and is the most precise individual determination.

The light quark masses. The value of the strong coupling constant was well known before lattice calculations. Its confirmation by lattice calculations is a welcome validation of lattice methods. By contrast, without the lattice, the values of the light quark masses can only be estimated approximately. The mass of the strange quark in particular plays an important role in analysis of weak interaction phenomenology, so a good determination of its value is a pressing concern. Quark models and a variety of phenomenological methods yielded conventional wisdom estimates of $m_s \sim 150$ MeV for the strange quark mass, and $m_u \sim 6$ MeV for the average of the up and down quark masses. That conventional wisdom, we now know, is far off the mark. With lattice QCD, we can determine these masses with first-principles calculations, for example, by tuning the quark masses to obtain the correct masses for pions and kaons. Last year, the MILC collaboration using improved staggered fermions reported

$$m_s = 90(6) \text{ MeV}, \quad m_l = 3.3(3) \text{ MeV}.$$  

A recent paper by the CP-PACS and JLQCD collaborations reported a result using $\mathcal{O}(a)$ improved Wilson fermions:

$$m_s = 91^{+15}_{-6} \text{ MeV}, \quad m_l = 3.5^{+0.6}_{-0.4} \text{ MeV}.$$  

The two results are very compatible, giving necessary evidence that the results of lattice calculations are not dependent on the quark method.
Golden quantities and the CKM matrix elements. Most of the results discussed so far are for a particularly simple kind of quantity for lattice QCD: stable mesons (that is, ones that only decay weakly and not hadronically), in processes with a single meson present at a time. These are golden quantities for lattice QCD, with uncertainties that are smaller and easier to understand than for most quantities. Although this is a restricted set, many of the most important tasks of lattice gauge theory can be accomplished with quantities of this type. In particular, almost all of the CKM matrix elements and quark masses can be determined with lattice calculations in this category.

CKM matrix elements are measured in decay processes in which a quark of one flavor turns into a quark of another flavor. Fig. 7 illustrates $B$ meson “semileptonic” decay, that is, decay into two leptons plus one or more hadrons. In the experimentally observed process, a $B$ meson decays into two leptons, for example the electron and a neutrino, plus hadrons (labeled $X$), for example a pion. The complicated strong interactions of gluons with quark-antiquark pairs responsible for confining the valence $b$ and $\pi$ quarks in the $B$ meson is represented schematically in the figure by the curly red lines (gluons) and green circles (quark = antiquark pairs). The experimental rate depends on a QCD amplitude, which must be supplied by lattice QCD, and on the CKM matrix element $V_{ub}$, which is the coupling between an “up” quark ($u$) and a “bottom” quark ($b$). Purely leptonic decays, such as a pion decaying into an electron plus a neutrino, are parameterized by decay constants such as $f_{\pi}$. Pion leptonic decay depends on the QCD amplitude $f_{\pi}$ and on $V_{ud}$, the CKM matrix element connecting up and down quarks. The amplitudes for mesons like $K$, $B$, and $B_s$ to mix with their antiparticles, $\bar{K}$, $\bar{B}$, $\bar{B}_s$, are proportional to other combinations of CKM matrix elements. In all, eight of the nine CKM matrix elements can be determined from relatively simple lattice QCD calculations combined with experiment, as shown in Table I.

Semileptonic decays. In semileptonic decay, the shape of the decay amplitude as a function of the momentum of the decay products is predicted by lattice QCD and can be measured in experiment. Figure 8 shows the form factor describing the semileptonic decay $D \to Kl\nu$, as a function of the momentum transfer $t = q^2$. The predicted shape (green circles) agrees well with the observed shape (blue diamonds).

Table I. (Color online) The Cabibbo–Kobayashi–Maskawa matrix elements, with particle processes by which they can be measured.

<table>
<thead>
<tr>
<th>$V_{ud}$</th>
<th>$V_{us}$</th>
<th>$V_{ub}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f_{\pi}$</td>
<td>$f_{K}$</td>
<td>$f_{K^*}$</td>
</tr>
<tr>
<td>$V_{cd}$</td>
<td>$V_{cs}$</td>
<td>$V_{cb}$</td>
</tr>
<tr>
<td>$f_{D}$</td>
<td>$f_{D^*}$</td>
<td>$f_{D^*}$</td>
</tr>
<tr>
<td>$V_{td}$</td>
<td>$V_{ts}$</td>
<td>$V_{tb}$</td>
</tr>
</tbody>
</table>

$V_{us}$, and $V_{ub}$ have two of which are called $\rho$ and $\eta$, $\rho$ and $\eta$ have the form $\rho = \eta \propto V_{ub}$. By determining these parameters in many different ways, one can test whether or not consistent results are obtained. Inconsistent results would be a signal of contributions to quark mixing from Beyond-the-Standard-Model theories, rather from the Standard Model alone.

$B\bar{B}$ and $B\bar{B}$ mixing. To illustrate the challenge ahead, consider the $\rho - \eta$ plane, shown in Fig. 9. In the Standard Model, the CKM matrix may be parameterized by four parameters, two of which are called $\rho$ and $\eta$. $\rho$ and $\eta$ have the form $\rho - \eta \propto V_{ub}$. By determining these parameters in many different ways, one can test whether or not consistent results are obtained. Inconsistent results would be a signal of contributions to quark mixing from Beyond-the-Standard-Model theories, rather from the Standard Model alone.

$\rho$ and $\eta$ parameterize the CP violation in the Standard Model. CP is a symmetry relating the properties of particles to those of their antiparticles. Understanding the source of CP violation in nature is key to understanding the abundance of matter over antimatter in the visible universe. The plot is one of the most famous graphs in particle physics at the moment, and reducing its uncertainties is an important goal of particle physics.

The plot shows the bounds on the $\rho - \eta$ arising from various physical processes, with the small red circle illustrating the combined bound. Several of the uncertainties in the plot arise from estimates of the uncertainties in lattice QCD calculations. For example, the bounds in the dark green curves, labeled $\epsilon_K$, arise from measuring the mixing between $K$ mesons and their antiparticles, analyzed with
lattice QCD. Similarly, the bounds in the yellow circles, labeled $/C_14$, arise from $B_B$ mixing. The orange circles show the constraint arising from the combination $B_B$ and $B_s$ mixing from before (Fig. 9) and after (Fig. 10) the discovery of $B_s$ mixing at the Tevatron last year. These constraints are only possible due to the existence of good lattice gauge theory calculations. The experimental errors on the mixings that have been measured are of order 1%. The 10 or 20% uncertainties in the quantities shown in the graph are estimates of the uncertainties of lattice calculations. The current round of calculations aims at reducing these to something of order 5%. Clearly, to profit fully from the experiments that have been done, one needs to aim at lattice uncertainties of around 1%. So challenging work remains ahead for lattice gauge theorists if the experimental results are to be fully exploited.

Decay constants. To show that accuracies of 1–2% in this type of calculation are not an unreasonable goal, we can consider a new paper from HPQCD on decay constants (which parameterize the amplitudes for decay of a meson into a pair of leptons). They employ several improvements over previous calculations. Most importantly, they use a staggered fermion actions for the quarks that is more highly improved than previously (HISQ, or Highly Improved Staggered Quarks\textsuperscript{25}). They employ several other improvements to reduce the size of the uncertainties. Their results are:

\begin{align*}
    f_D &= 241(3) \text{ MeV}, \\
    f_{D_s} &= 208(4) \text{ MeV}, \\
    f_K &/ f_D = 1.162(9), \\
    f_K &/ f_{D_s} = 1.189(7).
\end{align*}

The accuracies for the $D$ and $D_s$ decay constants are a factor of 4–5 improved over previous results,\textsuperscript{27} an impressive step forward.

Figure 11 shows their results as a function of light quark mass, extrapolated down to the physical light quark masses (black dashed lines), the results agree well with experiment (black circles at the left of the graphs). For the $D$ and $D_s$ mesons, the theory results are much more accurate than the experimental results.
contrast to the reverse situation in $B\bar{B}$ and $B_s\bar{B}_s$ mixing. The challenge now for lattice calculations is to extend this level of accuracy to many quantities.

4. The Future

I have emphasized a small set of well-done quantities that have a strong connection to particle physics experiment. However, the current reach of lattice QCD is much broader than this. It is possible to study processes with multiple hadrons present at the same time, although more difficult than for single-hadron processes. The case of $K \to \pi\pi$ has been worked out very clearly.\(^{20}\) Lattice calculations can investigate QCD in the realms of high temperatures and potentially of high densities that are of interest in neutron stars and the early universe.\(^{29}\) Well-developed investigations of nuclear structure are underway.\(^{30}\)

Calculations continue to become more and more powerful through improved methods, better algorithms, and more powerful computers. This is allowing us to improve the precision of existing calculations, to verify that all fermion formulations give the same answers, and to extend our reach to new QCD quantities.

More exciting times could await lattice gauge theory in Beyond-the-Standard-Model physics, depending on what is discovered at the LHC. Such new physics could present a variety of challenges. For theories that are like QCD, but with larger numbers of colors or flavors, the same methods that are proving successful for QCD can be used. For simple supersymmetric theories, promising methods are under development for their solutions.\(^{31}\) For other conceivable new theories, major algorithmic advances will be required, for example in the very interesting case of theories in which right and left handed quarks do not come in pairs with the same color charge. The era of the LHC and Beyond-the-Standard-Model physics is likely to prove as eventful for lattice gauge theory as the current one has been.

17) C. Bernard et al. (MILC Collaboration): hep-lat/0609053.

Paul Mackenzie was born in Illinois in 1950. He received his B.S. in physics and mathematics from the University of Illinois in 1975, and his Ph. D. in physics from Cornell University in 1981. He was a research associate at Fermilab and the Institute for Advanced Study. He has been a staff member at Fermilab since 1986. He has worked in various areas of perturbative and nonperturbative Quantum Chromodynamics (QCD). His recent research has focused on using nonperturbative QCD calculations to extract the fundamental parameters of the Standard Model from particle physics experiments.
String Theory

Tamiaki Yoneya

Institute of Physics, University of Tokyo, Komaba, Meguro, Tokyo 153-8902

(Received February 1, 2007; accepted February 15, 2007; published November 12, 2007)

This article is devoted to a nontechnical review on the present status of string theory towards an ultimate unification of all fundamental interactions including gravity. In particular, we emphasize the importance of string theory as a new theoretical framework in which the long-standing conflict between quantum theory and general relativity is resolved.

KEYWORDS: string theory, unified theory, quantum gravity, gravity-gauge correspondence

DOI: 10.1143/JPSJ.76.111020

1. Historical Background

The standard model has provided us a good understanding of the basic properties of present-day elementary particles, quarks, leptons and gauge bosons. Given various numerical data, we can in principle compute the probability amplitudes of every possible process involving these particles. However, it seems needless to emphasize that the model is still quite incomplete from a theoretical standpoint. In addition to the fact that the standard model has to assume a large number of input parameters, the model regarded as a fundamental theory of matter and its interactions is yet at a very unsatisfactory level, since it says nothing about quantum gravitational interaction of elementary particles.

As is well known, the mathematical framework of the standard model, gauge-field theory, has been developed in our endeavor towards unification of fundamental interactions: the structure of the standard model is governed by non-Abelian gauge symmetries. Even putting aside universal gravitational force, however, the standard model still has not really achieved desired unification between electroweak and strong nuclear forces. We often expect that the idea of unified gauge theory could be extended to a unification, “grand” unification, of these two fundamental forces. In regard to gravity, however, a majority of us now agree, after intensive efforts of many years, on that the ultimate unification of general relativity with quantum gauge-field theory would require a totally new mathematical framework.

The reason for this view is that there remains a deep conflict between quantum theory and general relativity. Quantum mechanics and special theory of relativity together are two basic physical laws on which the gauge-field theory and hence the standard model are based. But once we take into account general theory of relativity, we encounter serious troubles, both technically and conceptually. If we apply the standard method of quantization to general relativity, the difficulty of non-renormalizable infinities immediately mars our attempts. Even if gravitational field is left temporarily as a classical field, the quantum theory of matter with black hole backgrounds leads us into a conceptual problem of contradicting the principle of conservation of probabilities, namely unitarity, one of the basic principles of quantum mechanics, caused by the famous Hawking radiation. Of course, in terms of classical physics, the effect of gravity is negligible at present experimental scales when it is compared with other forces. However, the existence of such fundamental difficulties lying beneath the extremely successful framework of modern quantum physics should never be discarded. The situation is analogous to what physicists in the early 20th century were faced with in exploring microscopic laws of physics at atomic scale. The recent development of string theory strengths our hope that string theory contains crucial ingredients for achieving a reconciliaton between quantum theory and general relativity.

String theory has a quite curious history. It started out from something which was nothing to do with unified theories of interactions. From the 1950s to the 60s, even after a spectacular success of quantum electrodynamics, a large group of high-energy physicists at that time tended to believe that quantum field theory might not be the appropriate framework for describing strong nuclear force. Therefore the so-called “S-matrix approach” became a major stream during this period, and string theory actually emerged from this development in the late 60s. However, as our understanding on its nature was becoming deepened, various facets as an ideal theory of all interactions including gravity have gradually been uncovered. Even after almost 40 years since its first discovery, we are still in the midst of this process of exploring true meanings and new outcome of string theory. It is very important to recognize such peculiar evolution of our understanding in order to assess the present status of string theory impartially.

As another historical analogy, we may recall a stage in the development of non-Abelian (Yang–Mills) gauge theory from the mid 1950s to the whole 60s. During this period, precise method for quantizing non-Abelian gauge theory was not yet established, nor whether could its structure fit into observed properties of weak and strong nuclear forces. However, there have been many attempts proposing Yang–Mills type theories as attractive models for nuclear forces, on the ground of fundamental symmetry requirement which was inherited to theoretical physicists from the time of Einstein and Weyl. This attitude turned out to be right in the 70s, after appropriate understanding of its dynamics had been achieved, such as spontaneous symmetry breaking, Higgs mechanism, renormalizability, quark confinement, and quantum anomaly.

*E-mail: tam@hep1.c.u-tokyo.ac.jp
In the case of string theory, we have not yet arrived at any satisfactory non-perturbative definition of string theory, nor at primordial principles governing its structure. In spite of such an obscure status with regard to its ultimate fate, it seems fair to say that string theory has already provided us an entirely new perspective on how gravity could be unified with other interactions on the basis of quantum theory of strings and associated branes. It also suggested a new viewpoint on the dynamics of gauge-field theories in a way which has never been envisaged without unification with general relativity via string theory.

In this article, we try to convey the present situation of string theory to physicists who are working in other research areas than particle physics, explaining several key ingredients of string theory and reviewing some of important developments without technical details. For mathematical expressions, we use the natural units in which $h = 1$ and $c = 1$ throughout this article.

2. Perturbative Formulation of String Theory

2.1 Discovery of relativistic strings

String theory evolved from a proposal made in the late 60s for a particular $2 \leftrightarrow 2$ scattering amplitude, called the “Veneziano formula”, of mesons which satisfies a special symmetry requirement called $s$–$t$ “channel” duality. The latter demands that the amplitude is composed of elements such as the formula

$$V(s, t) = \int d\Omega x^{-\alpha' - \alpha_0 - 1}(1 - x)^{-\alpha' - \alpha_0 - 1}$$

(2.1)

which can equally be described by exchanges of particles between two interacting particles

$$V(s, t) = \sum_{n=0}^{\infty} \frac{r_n(s)}{t - m_n^2} = \sum_{n=0}^{\infty} \frac{r_n(t)}{s - m_n^2}$$

(2.2)

(first equality, “$s$-channel” description) or through formation of resonance-like states (second equality, “$s$-channel” description). Here, $s$ and $t$ are Lorentz invariant combinations of energy-momenta $s = -(p_1 + p_2)^2$, $t = -(p_2 + p_3)^2$, and $\alpha', \alpha_0$ are two parameters. It soon turned out that this amplitude and its various generalizations can be interpreted in terms of the dynamics of relativistic open strings propagating space–time, provided $\alpha_0 = 1$. The analogous amplitudes which correspond to closed strings were also constructed.

For example, the pole singularities at $s$ or $t = m_n^2 = (n - 1)/\alpha'$ are interpreted as representing possible states of strings with definite (mass)$^2$. There are an infinite number of them corresponding to various vibrational and rotational modes of strings. Actually, it also turned that for completely consistent formulations of quantum string theory, it is necessary that the space–time dimensions must be at some particular value (critical dimensions), 26, or if we want to include space–time (and world-sheet) fermions consistently, at 10. In the latter case we can eliminate the tachyonic ground state with negative (mass)$^2$ with $n = 0$, by demanding space–time supersymmetries. This is the origin of the naming, superstring theory. It was also understood that such closed strings can actually be generated by open strings, since one-loop amplitudes of open strings necessarily contain singularities corresponding to the propagation of closed strings. In other words, the $s$–$t$ channel duality extended to loop amplitudes of strings implies that closed strings are channel-dual to both open and closed strings.

2.2 World-sheet quantum mechanics of strings

We can formulate quantum string dynamics using a path-integral over all possible configurations of world sheets swept out by strings in space–time. In a symbolic and abbreviated notation, the amplitudes are expressed as

$$\sum_{\{\Sigma\}} g^{-\chi(\Sigma)} \int_{\mathcal{M}} [dX d\psi] \exp \left( -\frac{1}{4\pi \alpha'} S_{\Sigma}[X, \psi] \right)$$

(2.3)

where the symbol $\{\Sigma\}$ denotes the set of all in-equivalent (two-dimensional) Riemann surfaces, and $\mathcal{M}$ is the set of configurations of world sheets, described by fields $X, \psi, \ldots$ defined on the Riemann surface. The manner of how the constant $\alpha'$ appears in this expression shows that $1/\alpha'$ is essentially proportional to the energy, or tension, of the string per unit of length. As in the usual path integrals, we have to specify some boundary conditions corresponding to the initial and final states, which are suppressed in the present symbolic notation. The action $S_{\Sigma}$ is an integral over a given Riemann surface $\Sigma$ and takes the form

$$\int_{\Sigma} d^2 \xi L(X, \partial_\xi X, \psi, \partial_\xi \psi, \ldots)$$

with

$$L = g_{\mu\nu}(X) \partial_\xi X^\mu \partial_\xi X^\nu + \cdots$$

(2.4)

where $(\xi_1, \xi_2)$ with $z = \xi_1 + i\xi_2, \bar{z} = \xi_1 - i\xi_2$ are two-dimensional coordinates parametrizing the Riemann surface $\Sigma$. The space–time coordinates of strings are represented by fields $X^\mu(\xi) (\mu = 1, 2, \ldots, d - 1, 0$ with last index 0 being the time direction) on $\Sigma$, and $g_{\mu\nu}(X)$ is the metric tensor of target space–time. The additional field variable $\psi$ in (2.3) designates all other necessary fields, which are used to describe non-orbital degrees of freedom, such as spins, associated with strings. The constant $g_{\alpha\beta}$, called string coupling constant, specifies the weight of Riemann surfaces with various different topologies. It is well known that the topologies of Riemann surfaces are classified by the numbers of handles and boundaries, $(h$ and $b$ respectively). The symbol $\chi(\Sigma) \equiv 2 - 2h - b - p_c - p_o/2$ is the Euler number of Riemann surface fixed by topology, with additional information about the numbers, $p_c$ and/or $p_o$, of “punctures” inserted in the bulk of $\Sigma$ and/or on the boundaries, respectively. The punctures essentially amount to attaching infinite Riemann surfaces of cylinder topology ($p_c$) or of strip topology ($p_o$), which correspond to (initial and final) external states of closed or open strings, respectively, on their mass shell.

This description would look abstract at first sight, but it is not difficult to capture basic concept if one imagines an analogy with the notion of a particle propagator in ordinary quantum mechanics. In the latter case, one considers path integrals over all possible configurations of particle trajectories, world lines, in space–time. A world line is regarded as a one-dimensional base space, parametrized by a single coordinate $\tau$ for the particle coordinates $X^\mu(\tau)$. In string theory, the role of particle picture in local field theory is replaced by strings. The one-dimensional base spaces
composed of particle trajectories are elevated to special two-dimensional spaces, Riemann surfaces, and the particle coordinates are to fields on Riemann surfaces with their target space being the space–time. In particle theories which can be derived as perturbative rules of calculation in ordinary local-field theories, we have to sum up over all possible Feynman graphs corresponding to different topologies of particle trajectories, by combining propagators and vertices. Similarly, in string theory, we have to sum over all possible configurations of Riemann surfaces, on which string coordinates and their generalizations are defined. Just as the weight of each Feynman graph is essentially determined by the number of vertices and external lines, the string amplitudes are weighted by a single power of the string coupling constant. The outcome of particle Feynman rules is reduced to a series of expressions represented as multiple integrals with respect to Feynman’s auxiliary parameters, or equivalently Schwinger’s proper-time variables. The dimensions of such multiple integrals equal to the number of propagators involved. Similarly, the string amplitudes can be reduced to multiple integrals over the parameters called the “moduli” parameters of Riemann surfaces, which classify in-equivalent Riemann surfaces with given topology. The summation symbol \( \sum_{(E)} \) in (2.3) is meant to include such integrals too.

The Veneziano amplitude (2.1) corresponds to a single Riemann surface with one boundary (\( \sim \) disk) with four open-string punctures \( (h = 0, b = 1, p_e = 0, p_o = 4) \), where the four external states are the ground state corresponding to \( n = 0 \), which is tachyonic with a pure imaginary mass. The integration variable \( x \) appearing in (2.1) is the simplest example of the moduli parameters.

The choice of Riemann surfaces as the base space for world sheets of strings is a crucial ingredient for getting consistent theory, in that it leads to a mechanism by which the physical on-shell states of strings are interpreted in terms of a positive definite Hilbert space of infinite number of particle-like modes of strings, unitarity, and also to ultraviolet finiteness (also known as modular invariance\(^8\)) at each order with respect to expansion in terms of the string coupling \( g_s \). Riemann surfaces as the base space of generic two-dimensional field theories can equivalently be characterized by a symmetry requirement that we only allow conformally invariant theories. From the viewpoint of renormalization theory, conformally invariant field theories are very special, corresponding to fixed points of renormalization group. In particular, two-dimensional conformal group is an infinite dimensional group and consequently puts strong constraints. In string theory, one does not allow its violation even by quantum anomaly (conformal anomaly). Recall that in usual two-dimensional conformal field theories one allows the conformal anomaly. This stringent requirement of the world-sheet conformal invariance explains why in string theory we have many unusual constraints, such as critical space–time dimensions, possible gauge groups and so forth. Apart from the basic dimensional constant \( \alpha' \), there is no free continuous parameter.

### 2.3 Connection with local field theories

The similarity of these rules of world-sheet dynamics with particle Feynman rules is not a mere analogy. If we take the limit \( \alpha' \to 0 \), all of excitation modes other than the massless states corresponding to poles at \( s = 0 \) or \( t = 0 \) [with \( n = 1 \) in (2.2)] go infinity, and hence become invisible from the spectrum of strings. As the expression (2.3) clearly shows, the constant \( \alpha' \) controls the extendedness of strings. In this limit, only world sheets shrinking into infinitely thin tubes or strips contribute, and the string amplitudes reduce to sum of Feynman-graph amplitudes for massless particles of ordinary local-field theories. In particular, multiple integrals over the moduli parameters of Riemann surfaces reduce to those of Feynman–Schwinger parameters. The reduced theories are nothing but gauge theories\(^9\) and general relativity\(^10\) with suitable inclusion of massless fields other than ordinary vector gauge fields and tensor graviton. This discovery started the possibility of string theory as a unified theory as explicitly claimed first in ref. 11.

Basically, open strings can only lead to massless gauge fields and their cousins, while closed strings necessarily contain graviton as well as gauge fields, together with their supersymmetry partners. The mechanism why the string necessarily contains graviton and gauge bosons in closed and open string sector, respectively, owes again to conformal symmetry of the world-sheet dynamics. The infinite-dimensional nature of the conformal group serves the reduction of particle degrees of freedom, as required for the presence of massless spinning states. The graviton can be identified with the one represented by the space–time metric tensor \( g_{\mu\nu}(X) \) appearing in the world-sheet action in (2.3), through the standard state-operator correspondence in conformal field theories.

In the mid 80s, it was established that there are five consistent superstring theories in this sense. These five theories are classified depending on their spectra with respect to gauge groups, space–time supersymmetries, and chiralities of massless fermion states: type I with chiral gauge group SO(32), two “heterotic” string theories consisting only of closed strings with chiral gauge group SO(32) or \( E_8 \times E_8 \), all with one supersymmetry in 10-dimensional sense, and type IIA and IIB theories, with two supersymmetries, which again consist of only closed strings with chiral or non-chiral massless fermion states, respectively, but do not have gauge fields in 10 dimensions. The type I theory is unique in the sense that its world sheet must be regarded as non-orientable. The strange naming “heterotic” string is originated from the fact that strings in these cases are constructed by pairing bosonic strings with critical dimensions 26 with 10-dimensional fermionic strings. The difference \( 26 - 10 = 16 \) is responsible for the emergence of gauge groups SO(32) or \( E_8 \times E_8 \) of rank 16. In the low-energy limit \( \alpha' \to 0 \), these five perturbatively defined theories reduce to corresponding local field theories, called supergravities in 10-dimensional space–time. Construction of generic supergravity theories were started from the mid 70s, and its perturbative quantization has been studied vigorously in the early 80s. Even aside from the problem of non-renormalizability, it turned out that chiral supergravity theories other than those which were obtained as the low-energy limits of string theory suffered from difficulty of quantum anomalies with respect to gauge invariance and/or general coordinate invariance. Recognition of such status of
supergravity theories and its resolution\(^2\) in terms of strings strongly motivated to reconsider string theory as final serious model for the unification of gravity with gauge theories.

3. Duality Relations among Perturbative String Theories

3.1 T-duality

At present, we do not know any definition of string theory which goes beyond the perturbative rules. However, we believe that seemingly different perturbative string theories actually correspond to a set of particular local stationary points in the space of all classical solutions in some unknown nonperturbative formulation of a single deeper theory. Here, “classical” means an approximation where only tree-type \((h = 0\) and/or \(b = 1\)) contributions are kept. Each perturbative string theory is seeing only infinitesimally small oscillations around one of these classical solutions. There are many good reasons for this belief. Even in a semi-perturbative framework, there are links between different theories.

The oldest among such relations is known as T-duality. Owing to one-dimensional spatial extension of strings, there are excitations of strings which can never exist in particle theory. Imagine that one among 9 (or 25) spatial dimensions is curved into a circle of radius \(R\) (circle compactification). The center of motion of strings along this direction gives states (momentum modes) with quantized momenta \(p_n = n/R\) \((n = 0, \pm 1, \pm 2, \ldots)\), which contribute to the single-body energy as \(E = |n|/R\). We can also consider strings stretching along this direction such that they wind up the circle \(m\) times. This leads to states (winding modes) with single-body energy as \(E = |m|R/\alpha'\). It turned out\(^3\) that these two different sets of states play a symmetrical role in generic string amplitudes, and a string perturbation theory with a given radius \(R\) can be mapped into a string perturbation theory with the opposite radius \(R' = \alpha'/R\). The manner how this arises is essentially the same as in the well-known Kramers–Wannier duality of the Ising model in statistical mechanics. The radius \(R\) is the analogue of temperature \(T\) in the latter case.

In the case of bosonic string theory, this interchange of radius \(R \leftrightarrow R'\) is realized within the single theory. Hence, it is actually a symmetry of the theory. In the case of fermionic string theories, it relates different perturbative string theories. For example, type IIA theory is transformed into type IIB theory combining with the transformation of string coupling as \(g_s' = g_s\sqrt{\alpha'/R}\). Similarly, under the circle compactification, it is known that the SO(32) and \(E_8 \times E_8\) heterotic theories give the same physical spectrum in 9 \((= 8 + 1)\) dimensional sense. This implies that two different 10-dimensional heterotic theories should be regarded as different limits with infinitely large circle \(R \rightarrow \infty\) of a single theory.

3.2 S-duality

One of the surprising developments of string theory achieved from the mid 90s is the recognition that different perturbative string theories may be related to each other by interchanging the string coupling between strong and weak regimes. We have no reasonable definition of string theory beyond perturbation theory, and hence we cannot directly study strong coupling regimes in general. However, in the low-energy limit, it is not unreasonable to assume that supergravity theories provide effective descriptions even in strong-coupling regime at least when one restricts attention to some particular physics, which are not strongly affected by massive degrees of strings, especially their symmetries.

As a typical example, take the type IIB theory. The effective equations of motion for massless fields in this case turn out to be invariant under the transformation \(g_s \rightarrow g_s' = 1/g_s\), called by “S-duality” transformation,\(^4\) if one simultaneously redefines the space–time metric \(g_{\mu\nu}(X)\) of the world-sheet action by

\[
e^{-\Phi(X)} g_{\mu\nu}(X) = e^{-\Phi'(X)} g'_{\mu\nu}(X), \quad \Phi'(X) = -\Phi(X) \tag{3.1}\]

and \(\Phi(X)\) is a massless scalar field called “dilaton”. The string coupling is actually given by the vacuum expectation value of the dilaton as \(g_s = e^{(\Phi)}\). The transformation must be accompanied by some additional redefinition of other fields, most notably, the interchange of two second-rank anti-symmetric tensor fields as

\[
\left( \begin{array}{c}
B_{\mu\nu} \\
C_{\mu\nu}
\end{array} \right) = \left( \begin{array}{c}
-C_{\mu\nu} \\
B_{\mu\nu}
\end{array} \right). \tag{3.2}\]

The anti-symmetric tensor \(B_{\mu\nu} = -B_{\nu\mu}\) is a gauge field whose source is the strings themselves through an important term in the world-sheet action

\[
\int d^2\xi \; e^{\phi(X)} B_{\mu\nu}(X) \partial_\mu X^a \partial_\nu X^a, \tag{3.3}\]

which was suppressed for brevity in (2.4). This field together with the graviton and the dilaton belongs to a sector of closed strings called “N(eveu)S(chwarz)–NS” sector. However, the other anti-symmetric tensor field \(C_{\mu\nu}\) which is in a different category called “R(amond)–R(amond)” sector does not directly couple to strings. Two different sectors \(R\) and \(NS\) are originated from the different boundary conditions (periodic or anti-periodic) for the fermionic degrees of freedom living on the world sheet. The repetitions such as \(R-R\) or \(NS-NS\) correspond to the existence of two independent propagating modes (left and right) along the string. Clearly, in order for the S-duality transformation (3.2) to be a symmetry of the type IIB theory, we need an object which can be the source for fields in the \(R-R\) sector. It turned out that there are such objects which precisely play the required role.

3.3 Dirichlet-branes

They are known as Dirichlet (D)-branes.\(^5\) To understand the origin of the D-branes, we have to go back to T-duality. The open strings are defined by the free boundary condition \(\partial_\mu X^a(\xi) = 0\) at their end points, where \(\partial_\mu\) is the derivative along normal directions of world sheets. But T-duality transformation maps a normal derivative to a tangential derivative. This means that the open strings after T-duality obey essentially the boundary condition of Dirichlet type at their ends along the direction of T-duality. As we have already emphasized, the open strings can generate closed strings. In the case of open strings with Dirichlet boundary conditions, it turns out that closed strings generated in this way contain states in the \(R-R\)-sector. This implies that the...
new objects at which open strings can end serve as the source for R–R-gauge fields. They are the D-branes. The string coordinates at the ends of open strings with Dirichlet boundary conditions can be regarded as a sort of collective coordinates of the D-branes, and as such the dynamics of the D-branes can be formulated as collective motions of open strings, in an analogous way as various solitons are described in non-linear local field theories.

The D-branes can have various different dimensions with respect to spatial extension, depending on the number of remaining extended directions obeying the Neumann condition. If the number of such dimensions is $p$, we call the corresponding D-branes by “$Dp$-brane”. In type IIA or IIB theories, stable $Dp$-branes can exist only for even or odd $p$, respectively. Each time we make a T-duality transformation in a particular direction, the Neumann and Dirichlet boundary conditions are interchanged, and hence D-branes in respective theories are changed by one.

We can now go back to the S-duality in the type IIB theory. The additional second-rank anti-symmetric gauge field $C_{\mu\nu}$ has $D1$-branes as its source. In other words, by the S-duality, the original “$F$”(undamental)-string of the type IIB theory is actually transformed into the D1-brane (often called “$D$-string”). The collective coordinate interpretation of end points of open strings coupled to D1-branes leads to the fact that the tension of D-string is proportional to $1/g_s$, because the dependence on the string coupling of a simplest D-string amplitude is determined by $\chi = -1$ with $h = 0$, $b = 1$, $p_0 = 0 = p_c$. This is consistent with (3.1), since it requires that the tension of D1-string must be the same $1/\alpha'$ as that of the fundamental string in terms of the new unit using the transformed space–time metric $d\tau' \sim ds/g_s$. Other elements of the closed-string states in the $R$–$R$-sector and D-branes with other dimensions naturally fit into this interpretation of S-duality for D-branes. Thus we are convinced that the type IIB theory is self-dual with respect to S-duality, in a similar way that the bosonic string theory is self-dual with respect to T-duality. Although we cannot go into details, other D-branes also play their respective roles in S-duality. In particular, there exist D3-branes which actually be regarded as self-dual objects under S-duality.

In other theories than the type IIB, the S-duality in general gives various connections among different theories. Surprisingly, it turned out that the type IIA theory in the strong coupling regime actually leads to a new theory which is not listed among five possible perturbative string theories.

The lowest dimensional D-branes in the type IIA theory are the D0-branes, or D-particles, since the open strings associated to them obey the Dirichlet condition with respect to all spatial directions. They are sources for the 1-form gauge field $A_\mu$ in the RR-sector. From the viewpoint of low-energy effective theory, namely, 10-dimensional type IIA supergravity, it has long been known that this gauge field can nicely be interpreted geometrically if one starts from 11-dimensional supergravity theory and make a dimensional reduction to 10 dimensions. This is an application of the old idea of Kaluza–Klein theory. In 11 dimensions, the space–time metric $g_{\mu\nu}$ has $11 \times 12/2 = 66$ independent components. On the other hand, in 10 dimensions one has $10 \times 11/2 = 55$ components. Supposing that 10-dimensional space–time is a result of compactification of one of the spatial dimensions in 11-dimensional space–time along a circle of infinitely small radius $R_{11}$, one can regard the remaining $66 - 55 = 11$ components as composed of a 10-component vector field $R_{11\mu}$ and a scalar field $R_{11}$. Each of them can be related to the 1-form gauge field $A_\mu$ and the dilaton $\Phi$, respectively. In this interpretation, the charge associated to $A_\mu$ is identified with the momentum along the compactified direction. The excitations with the lowest unit of quantized momentum along this direction have a definite mass $1/R_{11}$ when interpreted in 10 dimensions. Under this 11-dimensional interpretation of the type IIA theory, the D0-branes as the source for $A_\mu$ should be regarded as such excitations along the compactified direction. This demands that

$$1/R_{11} = 1/g_s^4 \sqrt{\alpha'}.$$  \hspace{1cm} (3.4)

Thus we expect that the strong coupling regime, $R_{11} \rightarrow \infty$, of the type IIA theory should actually be formulated in 11-dimensional space–time.

It is interesting to note that with the help of T-duality between IIA and IIB, we can express the string coupling of the type IIB theory as $g_{s'} = R_{11}/R$ with $R$ being the radius of circle compactification along one of spatial directions in the remaining 10-dimensional space–time. This gives an attractive 11-dimensional interpretation of S-duality as the symmetry under the interchange, $R \leftrightarrow R_{11}$, of the two directions of circle compactifications.

### 3.4 M-theory and the unity of all string theories

The above observation strongly suggests that there exists some unknown theory as an extension of 11-dimensional supergravity such that under compactification to 10 (or lower) dimensions it reduces to type II string theories.\textsuperscript{17} The expected theory is called “M-theory”. There are various indications that the objects which then replace strings are (super)membranes (hence the naming “M”) whose spatial extensions are two. In fact, at least classically, it is easy to show that the action for such supermembrane reduces to that of the superstring when it is compactified along the circle in 11th dimension. The existence of D2-brane in type IIA theory is also consistent with this interpretation, since it should be the supermembrane of the M-theory immersed completely within 10 dimensions. Unfortunately, however, the quantum theory of (relativistic) membranes has long been an unsolved open problem, and it is difficult to confirm these ideas at the present stage of development.

Recognition of the existence of D-branes implies that the distinction between theories with only closed strings and type I theory is not quite fundamental. D-branes can exist in both theories, though there is some speciality in the case of the type I reflecting a characteristic that the world sheets of their open strings are non-orientable. The latter property after T-duality leads to special (hyper)planes known as “orientifolds” which play a sort of mirrors reversing both spatial directions and world-sheet parity simultaneously. Detailed studies of, say, D-strings in T-dualized type I theory show that the spectrum of oscillations around the ground state of a D-string coincides with that of a long heterotic strings with SO(32) gauge group. In the low-energy limit, the supergravity actions of the type I and SO(32) heterotic theories are connected to each other by relating the
metric tensors and dilatons as $g^{I\mu\nu} = e^{-\Phi^h} g^{I\mu\nu}$, $\Phi^I = -\Phi^h$, together with some relations of gauge fields on both sides. These observations indicate that the strong-coupling regime of the type I theory is equivalent to the weak coupling regime of the $SO(32)$ heterotic theory. Furthermore, the assumed M-theory can provide a new link including the $E_8 \times E_8$ heterotic theory. As we have already mentioned, with a circle compaction, two heterotic theories in nine dimensions are regarded as two different limits of a single theory. This together with the above link between the type I and $SO(32)$ heterotic theory makes us possible to connect the theory to the type IIA theory when the latter is considered with special compaction on a line $S^1/Z_2$ which is a circle divided by $Z_2$ reflection around a point. With the help of the relation between the type IIA and M theories, one finally arrives at a set of links by which all five perturbative string theories are connected owing to the assumed existence of the M-theory. In this way, we now believe that there must exist a single unified and background-independent framework of string/M theory from which all five string theories and M-theory in 11 dimensions are formulated as different limiting cases.\(^{18}\)

### 4. Facets of String Theory towards the Ultimate Unification

Thus far, we have explained what the perturbative formulation of string theory is and how we are trying to achieve the unity of different possibilities of the theories. The attentive readers would ask; “All these developments are indeed impressive as themselves, but why do we have to believe that this is the theory which you are seeking for? After all, the theory cannot yet provide any decisive experimental predictions by which one would check evidence for the theory directly…” Of course, only time will be able to answer this question.

#### 4.1 Merits of string theory

Before proceeding to issues relevant in making connections of string theory to the real world, let us summarize here several characteristic properties of string theory as a new framework toward the ultimate unification, which discriminate the string/M theory from any other attempts within the ordinary framework of local field theory.

**String theory**

1. **Encompasses almost all relevant ideas and/or methods devised in the past, towards unification of particle interactions:** Such ideas include gauge invariance, Kaluza–Klein mechanism, induced gravity, composite models, and various higher symmetries including supersymmetry. Methods include bootstrap, current algebra, topological excitations, various duality transformations, and so on. In a sense, string theory has already achieved a spectacular unification of ideas;

2. **Provides a conceptually satisfying scheme of unifying all interactions including gravity:** One of the merits of the world-sheet formulation of strings on the basis of Riemann surfaces is that interaction and motion become a completely unified concept. Note that locally a Riemann surface is always a single sheet. The interaction vertices of various different particle modes of strings automatically resulted from the difference of global topologies of the surfaces as an artifact of taking the low-energy limit. It automatically generates gravity and gauge interactions as inevitable consequences from its mathematical structure;

3. **Resolves the ultraviolet difficulty which is inherent to all perturbative theories of particle-field theories with local interactions:** Previous attempts to unify gravity suffered from the ultraviolet difficulty. Removal of the ultraviolet difficulty within the usual framework of local field theory or in an extended framework allowing non-local interactions usually induces violation of unitarity. In string theory, by contrast, the resolution of ultra-violet infinities is achieved by conformal invariance of Riemann surface in complete conformity with unitarity. The singular limit of a Riemann surface corresponds, in terms of space–time geometry, only to physical unitarity singularities of Feynman graphs at long distances, while, in local-field theories, Feynman graphs contain also unphysical singularities associated to short-distance limit which lead to the ultra-violet difficulty;

4. **Provides for the first time a microscopic explanation,\(^{19}\) albeit only for some special cases protected by supersymmetry, of Black hole entropy in terms of quantum statistical language:**

This is based upon the interpretation of the extremal and near extremal black holes in terms of D-branes;

5. **Provides several new perspectives for understanding the dynamics of ordinary gauge field theories:** The most recent and remarkable example of this is the so-called AdS/CFT correspondence (or more generally gravity-gauge correspondence).

Perhaps we should add to this list that string/M theory would also provide a possibility of explaining why our world has four dimensions, starting from the critical dimensions, provided that one could finally achieve a proper understanding on the dynamics of reduction of space–time dimensions.

Among others, the importance of resolving the difficulty of non-renormalizability which has been marring attempts quantum gravity cannot be overemphasized. For example, if one tries to compute the black-hole entropy using local field theories, one necessarily encounters ultraviolet infinities. Not only that, the renormalization also forces us to introduce infinitely many other dimensionful constants to the microscopic theory than the Newton constant. One might think that different approaches such as, say, loop-quantum gravity\(^{20}\) which uses instead of the metric tensor a set of variables defined on extended loop regions of space–time, would also be able to evade the ultraviolet infinities. In such attempts, however, there is unfortunately no microscopic principle in writing down the action. It seems to fair to say that they are useful only as possible languages for effective description of some particular aspects of quantum gravity. The importance of this aspect of string theory has been motivating various proposals on the interpretation of short-distance properties of string theory, such as a generalized uncertainty principle or a novel uncertainty principle of space–time itself.\(^{21}\)
As regards to the black-hole entropy, it is interesting to recall that in the genesis of quantum theory the statistical interpretation of the entropy of black body radiation played an indispensable role in identifying the correct microscopic degrees of freedom. The ultraviolet catastrophe which physicists had to struggle 100 years ago has never been completely resolved when gravity is taken into account. Certainly, string theory provides the first (and only known) promising direction toward its resolution.

4.2 Attempts towards non-perturbative formulations

We now briefly describe some representative attempts towards non-perturbative formulation of string/M theory. The leading among them is the string field theory. Non-perturbative effects of strings must involve various non-trivial processes of creation, annihilation and condensation of strings and D-branes. However, the world-sheet formulation explained in §2 is simply an extension of perturbative Feynman rules in ordinary field theory, and as such its applicability is in principle restricted to processes involving only a small number of strings. In order to treat more complicated phenomena without such restriction, it is natural to reformulate the world-sheet formalism into a field theory of strings by introducing string fields which create and annihilate strings. The string fields are then functional fields $\Psi[X(\sigma),\ldots]$ on the base space of all possible string configurations $[X^{\mu}(\sigma),\ldots]$. We can in principle write down action principle for the string fields, such that it reproduces the world-sheet string amplitudes (2.3) at each order with respect to power-series expansion in the string coupling constant $g_s$. If we represent the general string configuration in terms of Fourier expansion $X^{\mu}(\sigma) = x_0^\mu + \sum x_N^\mu \exp(i n \sigma)$ with $x_0^\mu$ being the center-of-mass coordinates of string, such a string field theory can be regarded as a special version of infinite-component field theories.

Construction of string field theory was started already from the mid 70s. The most successful among such attempts is the one proposed by Witten in the mid 80s for bosonic open string theory. But its applications to non-perturbative effects have become possible only in recent several years, owing mainly to technical obstacles. For example, we can discuss the problem of instability of bosonic string theory. The existence of the tachyonic ground state indicates that the usual vacuum of bosonic string theory is unstable against creation of strings in the tachyonic modes. If one treats the string field in classical approximation, one can discuss the fate of this unstable vacuum by studying the possibility of stable vacua which should manifest themselves as local minima of effective potential for the string field. Such stable vacua must correspond to non-trivial classical (static) solutions of the string-field equation of motion. Quite recently, an example of such exact solution has been constructed, and led to a proof of a conjecture that the unstable vacuum can be regarded in the classical approximation as the excitation of an unstable D-brane on the background of a newly found stable vacuum.

Unfortunately, however, fully satisfactory formulation of open-string field theory has never been achieved in the case of superstrings with fermionic strings. Also, the status of this approach for closed strings has been much more subtle. From the viewpoint of Riemann surface, the string field theory requires a decomposition of its moduli space such that each element in the decomposition corresponds to a single Feynman graph in the ordinary sense of particle theory. Taking the example of the Veneziano amplitude (2.1), the integration region $[0,1]$ of the variable $x$ is decomposed into two segments $[0,1/2]$ and $[1/2,1]$, related by a mapping $x \leftrightarrow 1-x, s \leftrightarrow t$, each of which leads to contribution with poles with respect only to either $s$ or $t$, corresponding to particle Feynman graphs with s- or t-channel particle exchange, respectively. A possible string field theory gives such a particular decomposition of the whole moduli space of Riemann surfaces. In the case of closed strings, however, to achieve this in a covariant way becomes an extremely complicated problem as one goes to higher orders in $g_s$. It should also be noted that, by so doing, various nice characteristics of string theory, especially the duality between open and closed strings, resulting from global properties of Riemann surfaces become largely invisible in terms of string fields.

4.3 Matrix models

In the framework of string field theory, dealing with multi-body processes of D-branes is not easy, since they are sort of collective excitations of strings. Here it is useful to recall that the low-energy approximations of open string field theories are (super) Yang–Mills gauge theories defined on $(p+1)$-dimensional base space–time, in which the size of gauge group corresponds to the number $N$ of Dp-branes. There are some situations where the Yang–Mills models can even be regarded exact in certain sense. For example, in the case of D0-brane (D-particles), large $N$ can be solved exactly in certain sense. For example, in the case of D0-brane (D-particles), large $N$ means large momentum $P \sim N/R_{11}$ along the compactified circle from the viewpoint of 11-dimensional M-theory. We can then go to an infinite-momentum frame which is boosted along this direction. All motions in this frame become infinitely slow and hence we can rely on a low-energy non-relativistic approximation. Motivated by this, it has been proposed that an appropriate large $N$ limit of $(p+1)$-dimensional super Yang–Mills theory (Yang–Mills quantum mechanics) could be a version, known as M(atrix) theory, of non-perturbative description of the M-theory. The Yang–Mills quantum mechanics for large $N$ can also be regarded as a particular regularized version of membrane theory. Unfortunately, the quantum dynamics of the theory for large $N$ is not accessible to our present technology, since the Yang–Mills quantum mechanics is necessarily strongly coupled in the infra-red regime. It is, however, remarkable that this theory in perturbation theory can reproduce non-linear general-relativistic effects of graviton self-interactions.

A similar proposal known as IIB matrix model has been given from the viewpoint of the type IIB string theory and its lowest dimensional D(1)-branes, which are essentially “instanton”-like objects whose open strings obey Dirichlet condition with respect to all directions including time direction. It is simply a large $N$ limit of 10-dimensional super Yang–Mills theory defined on one point, or dimensional reduction from 10 to 0, in close analogy with the so-called Eguchi–Kawai reduction of lattice gauge theories proposed in the early 80s.
In any of these and related proposals, we had not made truly new insight with respect to the most important characteristics of string/M theory such as the S-duality, and the (channel) duality between open and closed strings, unfortunately. It seems too early to assess whether the string field theory or its matrix model versions could be the right languages for non-perturbative formulation of fully quantized string/M theory.

4.4 Gravity-gauge correspondence

Perhaps, one of the most surprising and potentially useful outcome of string theory whose impact might go down to other research fields than fundamental physics is the existence of duality between gauge theory and general relativity. Of course, string theory unifies both theories: this is the main point of our discussions above. However, we usually regard these two theories as two different field theories formulated on the basis of entirely different fields, vector gauge fields and metric tensors, and on different principles, gauge principle and general covariance, respectively. In recent 10 years or so, it turned out that there can be special circumstances where these two theories embedded in string theory actually describe the same single physics using two different languages, in an analogous way as we learned in the wave-particle duality of quantum mechanics. The key ingredient to understand this remarkable phenomena is again the D-brane.

On one hand, D-branes can be formulated as collective modes of open strings attached to them. In an appropriate low-energy limit, they can be described by gauge field theory. On the other hand, they act as sources for closed strings, and thus their dynamics is reflected in the behavior of metric tensors and their partners including those in the R–R sector. If one can have some regions of space–time where massive string states on both theories are ignored, this implies that these two different field theories can describe one and the same physics. The most typical case of this conjecture is known as AdS/CFT correspondence between the four-dimensional super SU(N) Yang–Mills theory with maximum possible supersymmetry, known as $\mathcal{N} = 4$ supersymmetry, and supergravity corresponding to the type IIB theory around a special background geometry which is the product of a five-dimensional anti-deSitter space–time (AdS$_5$) and a five-dimensional (hyper) sphere ($S^5$). The former is the low-energy effective theory of N D3-branes, while the latter corresponds to the geometry produced by them in the near horizon region. The symbol CFT indicates that the gauge-theory side is a conformally invariant field theory in four dimensions. D-branes in general generate black-hole type geometries with event horizons. The AdS geometry appears when we study this black D3-brane geometry by approaching the horizon sufficiently closely, comparing to the string scale $\sqrt{\alpha'}$. The curvature radius in the near horizon limit of this geometry is of order $L \equiv (g_s N)^{1/4} \sqrt{\alpha'}/2$ for both five-dimensional elements of this product space–time, whose metric is

$$ds^2 = L^2(\varepsilon^{-2}(dx^2 + dx_5^2) + ds_{S^5}^2)$$

with $dx^2$ and $dx_5^2$ being the metrics of flat four-dimensional Minkowski space–time and a unit five-sphere, respectively. The low-energy approximation of supergravity is valid when $g_s N \gg 1$ and $g_s \ll 1$, the latter of which is necessary to suppress the quantum loop effects of closed strings. The Yang–Mills coupling constant $g_{YM}$ is related to the string coupling as $g_s \propto g_{YM}^{-2}$.

One might wonder how the correspondence could be realized between 4-dimensional and 10-dimensional field theories. The mapping between them is as follows. There are 10 bosonic fields on the Yang–Mills side. Among them, six scalar fields describe the collective motions of D3-branes along the six transverse spatial directions, while the remaining four-component vector gauge field corresponds to the lowest modes of open strings propagating along the base four-dimensional space–time of D3-branes. The former six directions are separated into one radial direction and five angles, since D3-branes are point-like with respect to these six directions. The radial direction and the four-dimensions of the D3-brane base space together are related to the five coordinates $(z, x_5^a)$ of AdS$_5$ geometry, while the five angles are to the $S^5$. The conformal invariance of the gauge theory is translated to the isometry of the metric (4.1) under SO(4, 2) coordinate transformations, the simplest among which is the scale transformation $\lambda(z, x_5^a) \rightarrow \lambda(z, x_5^a)$. Because of the common conformal symmetry together with supersymmetry, we can classify the spectra of both theories using conformal dimensions. In particular, there are certain classes of states, called BPS states, whose spectra do not depend on any continuous parameters. They can be reliably enumerated on both sides using weak-coupling analysis, and an explicit mapping of states and operators between them can be established. Non-BPS states are more difficult to make correspondence, since they necessarily depend on the strong effective coupling constant $g_{YM}^2N (\gg 1)$ of the large N gauge theory. In recent years, we are making important progress along this direction, including some explicit relations of correlation functions of the gauge theory and the corresponding amplitudes on the supergravity side. These new developments suggest how to interpret massive closed-string states in terms of the gauge-theory variables. This leads to a conjecture that the whole content of the type IIB closed string theory may be encoded in the gauge theory.

In the above correspondence, the four-dimensional gauge theory can be regarded as living on the four-dimensional space–time at the boundary $z \rightarrow 0$ of the five-dimensional bulk space–time, AdS$_5$. For example, correlation functions of the gauge theory can be interpreted as amplitudes of gravity theory observed at the boundary, examining responses from the bulk against small disturbances added at $z \sim 0$. In this sense, the AdS/CFT correspondence provides a concrete example of the idea of “Holography”. The latter has originated from an interpretation of the Bekenstein–Hawking entropy of black holes, which indicates that the information on micro-states of three-dimensional black holes is encoded in the degrees of freedom on the two-dimensional black-hole horizons. Some general considerations suggest that this decrease of physical dimensions is a universal feature of the correct quantum theory of gravity.

There are many attempts and proposals for possible extensions of the gravity-gauge correspondence. One can even hope that the complicated non-perturbative dynamics
of QCD may be rephrased holographically into some appropriate gravity theory in higher dimensions, generalizing the AdS/CFT correspondence to non-conformal and non-supersymmetric theories.

5. Roads to the Real World?

As emphasized above, we are still in the midst of the process of exploring hidden meanings of string theory, without knowing the real definition of the theory. Therefore, we cannot make any definite statement with respect to prospect on various issues related to scenarios how the theory gives predictions for the real four-dimensional world.

However, attempts to construct 4-dimensional string models by reducing 10 dimensions into products of 4-dimensional Minkowski space–time $M_4$ and small unobservable 6-dimensional spaces $K$ have been naturally a big enterprise of string physicists, ever since the first explosion of string theory in the mid 80’s. Such works were within the framework of five possible perturbative formulations. In these approaches, one requires in view of the hierarchy problem that reduced models should have $N = 1$ space–time supersymmetry in four dimensions, which sets some special geometrical constraints for the six-dimensional compactified manifolds $K$. A dominant role for satisfying this requirements is played by the so-called Calabi–Yau manifolds, and studies have been performed under fruitful interactions of string physicists with mathematicians. In particular, various duality relations on the string-theory side suggest useful insights on some unexpected relations, such as “mirror symmetry” which can be regarded as a cousin of T-duality, between mathematically different characterizations of these manifolds.

It is impossible here to summarize very extensive studies over 20 years along this direction. A general consensus is that although there are a small class of models which mimic the features of the standard model or grand-unified theories, there are a huge number of different possibilities of such schemes of compactifications. Even if one starts from some possibilities of standard model-like string vacua, they usually have many unwanted fields, such as dilatons and other perturbatively massless excitations called “moduli” fields corresponding to infinitesimal deformations of the geometry allowed under the constraints for the Calabi–Yau compactifications. These additional degrees of freedom must be lifted to massive fields in order to be consistent with various experimental and cosmological constraints. The issues together with the problem of breaking supersymmetry necessarily become non-perturbative.

Recently, with the advent of D-branes and related objects, there have been proposed various interesting partially non-perturbative mechanisms to resolve some of these problems. For example, we can assume presence of D-branes or their associated fluxes of R–R–gauge fields extending in the compactified manifolds. They can be used to generate some desirable masses for unwanted fields. Myriads (10$^{500}$ or more in the case of the type IIB model) of those possibilities are expected. Total space consisting of all such possible vacua is now called “String Landscape”. Majority of such vacua would have finite cosmological constant characterized by the string scale. There are various conceivable mechanisms to obtain vacua with different cosmological constant by considering excitations of D-branes and anti D-branes. Such arguments are however based on low-energy effective field theories. It is not clear whether these discussions of compactifications survive when we hopefully arrive at right stringy languages for non-perturbative string physics.

Finally, D-branes have also suggested a different scenario for approaching 4 dimensions. As we have explained above, the system of $N$ D$p$-branes in type II theories are described in a low-energy approximation by SU($N$) gauge field theories. For $p = 3$, we have thus four-dimensional gauge theories. This inspired the idea that our universe could actually be confined on a four-dimensional “world-brane”. The five-dimensional AdS part of the metric (4.1) provides simplest such a model. If one assumes that the universe sits at $y = \ell$ with $z = e^y$, there exist different length scales, separated exponentially by a “warp factor” $e^y$ from that of bulk (gravitational) physics, and one could hope that this might effectively explain the hierarchy of mass scales in unified models. This and related possibilities have been one of hottest topics in phenomenological approaches both in particle physics and cosmology in recent several years. We leave this subject to some of other articles in the present special issue.

Acknowledgments

The present work is supported in part by Grants-in-Aid for Scientific Research [No. 13135205 (Priority Areas) and No. 16340067 (B)] from the Ministry of Education, Culture, Sports, Science, and Technology.

1) The following bibliography is very incomplete, since we cannot mention all relevant original works in view of the nature of this article. The author apologizes anyone whose important works are not listed here. For extensive text-books accounts of string theory with more references, see M. B. Green, J. H. Schwarz, and E. Witten: Superstring Theory (Cambridge Univ. Press, 1987) Vols. 1 and 2; J. Polchinski: String Theory (Cambridge Univ. Press, 1998) Vols. 1 and 2; K. Becker, M. Becker, and J. Schwarz: String Theory and M-Theory: A Modern Introduction (Cambridge Univ. Press, 2007).
18) For more detailed discussions on the unity of perturbative string
theories, the reader should consult some recent textbooks cited in ref. 1.
20) For a review of other approaches to quantum gravity, see, e.g., A. Ashtekar and L. Lewandowski: Classical Quantum Gravity 21 (2004) R53.
33) For a review, see, e.g., N. Beisert: hep-th/0407277.

Tamiaki Yoneya was born in Hokkaido, Japan in 1947. He obtained his B. Sc. (1969), M. Sc. (1971), and D. Sc. (1974) from Hokkaido University. He was a research associate (1975–1980) at Faculty of Science, Hokkaido University, associate professor (1980–1991) at College of Arts and Sciences, University of Tokyo. Since 1991 he has been a professor at the Graduate School of Arts and Sciences, University of Tokyo. He has worked on various aspects of elementary particle theories, especially in non-perturbative gauge field theory, quantum gravity, and string theory. In his recent researches, a particular emphasis has been put on the unified theory of all fundamental forces including gravity on the basis of string theory.