α

β

γ

HEAVY FLAVOUR PHYSICS

Theory and Experimental Results in Heavy Quark Physics

EDITED BY
C T H DAVIES AND
S M PLAYFER

ROUTLEDGE

HEAVY FLAVOUR PHYSICS

THEORY AND EXPERIMENTAL RESULTS IN HEAVY QUARK PHYSICS AND CP VIOLATION

HEAVY FLAVOUR PHYSICS

THEORY AND EXPERIMENTAL RESULTS
IN HEAVY QUARK PHYSICS
AND CP VIOLATION

Proceedings of the Fifty-Fifth Scottish
Universities Summer School in Physics,
St. Andrews, 7 August – 23 August 2001.

A NATO Advanced Study Institute

Edited by

C Davies — University of Glasgow
S M Playfer — University of Edinburgh

Series Editor

P Osborne — University of Edinburgh

Copublished by
Scottish Universities Summer School in Physics &

SUSSP Proceedings

SUSSP Proceedings (continued)

Lecturers

Steve Abel	Institute for Particle Physics Phenomenology, Durham, UK
Gerhard Buchalla	CERN, Geneva, Switzerland
Christine Davies	University of Glasgow, UK
Konrad Kleinknecht	University of Mainz, Germany
Lawrence Krauss	Case Western University, Cleveland, USA
Peter Krizan	University of Llubljana, Slovenia
Tatsuya Nakada	University of Lausanne, Switzerland
Harry Nelson	University of California, Santa Barbara, USA
Yosef Nir	Weizmann Institute, Rehovoth, Israel
Ken Peach	University of Edinburgh & Rutherford Appleton Laboratory, UK
Jonathan Rosner	University of Chicago, USA
Klaus Schubert	Technical University, Dresden, Germany
Sheldon Stone	Syracuse University, USA

Organising Committee

Professor K J Peach	University of Edinburgh	*Co-Director*
Profssor Y Nir	Weizmann Institute	*Co-Director*
Mr A Walker	University of Edinburgh	*Secretary*
Dr C Froggatt	University of Glasgow	*Treasurer*
Professor C Davies	University of Glasgow	*Co-Editor*
Dr S M Playfer	University of Edinburgh	*Co-Editor*
Dr A Khan	University of Edinburgh	*Steward*
Dr P Soler	University of Glasgow	*Steward*
Ms L O'Donnell	University of Edinburgh	*Administration*

Preface

The timing of a school on Heavy Flavour Physics in August 2001 was particularly fortunate. Both BaBar and Belle had reported in the previous month new results with clear evidence for CP-violation in B_d decays. The Standard Model had been subjected to a qualitatively and quantitatively new test, and passed with flying colours. The result on $\sin 2\beta \simeq 0.7 \pm 0.2$ was in excellent agreement with the Standard Model expectations based on non-CP violating processes of $\sin 2\beta$ of about 0.7. The CP-violation observed more than 35 years ago in the neutral kaon system was confirmed as coming, at least predominantly, from the complex phase in the quark mixing matrix.

Of course, there is more, much more, to Heavy Flavour Physics than CP violation and the b-quark systems, and the course of lectures reflected this—'heavy flavour' was interpreted generously to include the c-quark and as appropriate the s-quark.

The subject was approached from a strongly phenomenological point of view. Jonathan Rosner gave an overview of the Standard Model and its development, Gerhard Buchalla introduced the topic of Heavy Quark Theory and Christine Davies explained how these ideas could be implemented on the lattice. Yosef Nir gave a thorough introduction to CP-violation, while Steve Abel introduced concepts beyond the current Standard Model. Sheldon Stone gave an overview of B-phenomenology, and the present and future experimental programme was reviewed by Klaus Schubert (B-factories), Peter Krizan (HERA and the Tevatron), and Tatsuya Nakada (LHCb and BTeV). Harry Nelson discussed the opportunities for observing physics beyond the Standard Model in the charm sector, and Konrad Kleinknecht discussed the status of CP-violation in the neutral kaon system. Two topics strictly outside the domain of heavy flavour, but with significant relevance were covered. Lawrence Krauss gave an introduction to cosmology, and Ken Peach discussed the possibility that CP-violation might one day be measurable in the neutrino sector at a neutrino factory.

We believe that this series of lectures will provide a thorough introduction to the phenomenology of heavy flavour physics, not only those working on the B-factories, LHCb or BTeV, but also those working on HERA, the Tevatron and eventually the LHC general purpose detectors. While the wealth of data from the B-factories and the Tevatron will lead to new discoveries and, we hope, clues to the physics beyond the Standard Model, it is already clear that the model provides a very good description of the data. Meanwhile, these lectures will be a useful guide to these processes, and will enable the student, whether theorist or experimentalist, to judge the significance of these developments as they unfold.

Summer schools are not just about science: they are also about dialogue, discussion, meeting people, and making friends. The school succeeded in this secondary aim, aided by full social programme and a friendly environment provided by the staff of the John Burnet Hall. Within this, the scientific discussions and personal interactions flourished.

The Organizing Committee acknowledge the support of NATO, PPARC (the UK Particle Physics and Astronomy Research Council), the National Science Foundation, the Scottish Universities Summer Schools in Physics and the Physics and Astronomy Departments of the Universities of Edinburgh, Glasgow and St. Andrews, without which the school would not have been possible.

We would also like to thank all of the lecturers and participants for their enthusiasm, for physics and for life, which helped make this a truly memorable school.

Ken Peach and Yossi Nir

Co-Directors, January 2002

Editors' Note

We regret that the lectures presented by Harry Nelson, Lawrence Krauss and Ken Peach are not included in this volume.

Christine Davies and Steve Playfer

Editors, April 2002

Contents

Standard Model

Heavy Quark Theory & CP Violation

Experimental Results

The Standard Model in 2001

Jonathan L Rosner

Enrico Fermi Institute and Department of Physics
University of Chicago, USA

1 Introduction

The "Standard Model" of elementary particle physics encompasses the progress that has been made in the past half-century in understanding the weak, electromagnetic, and strong interactions. The name was apparently bestowed by my doctoral thesis advisor, Sam Treiman, whose dedication to particle physics kindled the light for so many of his students during those times of experimental and theoretical discoveries. These lectures are dedicated to his memory.

As graduate students at Princeton in the 1960s, my colleagues and I had no idea of the tremendous strides that would be made in bringing quantum field theory to bear upon such a wide variety of phenomena. At the time, its only domain of useful application seemed to be in the quantum electrodynamics (QED) of photons, electrons, and muons.

Our arsenal of techniques for understanding the strong interactions included analyticity, unitarity, and crossing symmetry (principles still of great use), and the emerging SU(3) and SU(6) symmetries. The quark model (Gell-Mann 1964, Zweig 1964) was just beginning to emerge, and its successes at times seemed mysterious. The ensuing decade gave us a theory of the strong interactions, quantum chromo-dynamics (QCD), based on the exchange of self-interacting vector quanta. QCD has permitted quantitative calculations of a wide range of hitherto intractable properties of the *hadrons* (Lev Okun's name for the strongly interacting particles), and has been validated by the discovery of its force-carrier, the *gluon*.

In the 1960s the weak interactions were represented by a phenomenological (and unrenormalisable) four-fermion theory which was of no use for higher-order calculations. Attempts to describe weak interactions in terms of heavy boson exchange eventually bore fruit when they were unified with electromagnetism and a suitable mechanism for generation of heavy boson mass was found. This *electroweak theory* has been spectacularly successful, leading to the prediction and observation of the W and Z bosons and to precision tests which have confirmed the applicability of the theory to higher-order calculations.

In this introductory section we shall assemble the ingredients of the standard model —the quarks and leptons and their interactions. We shall discuss both the theory of the strong interactions, quantum chromo-dynamics (QCD), and the unified theory of weak and electromagnetic interactions based on the gauge group SU(2) $\otimes$ U(1). Since QCD is an unbroken gauge theory, we shall discuss it first, in the general context of gauge theories in Section 2. We then discuss the theory of charge-changing weak interactions (Section 3) and its unification with electromagnetism (Section 4). The unsolved part of the puzzle, the Higgs boson, is treated in Section 5, while Section 6 concludes.

These lectures are based in part on courses that I have taught at the University of Minnesota and the University of Chicago, as well as at summer schools (e.g., Rosner 1988, 1997). They owe a significant debt to the fine book by Quigg (1983).
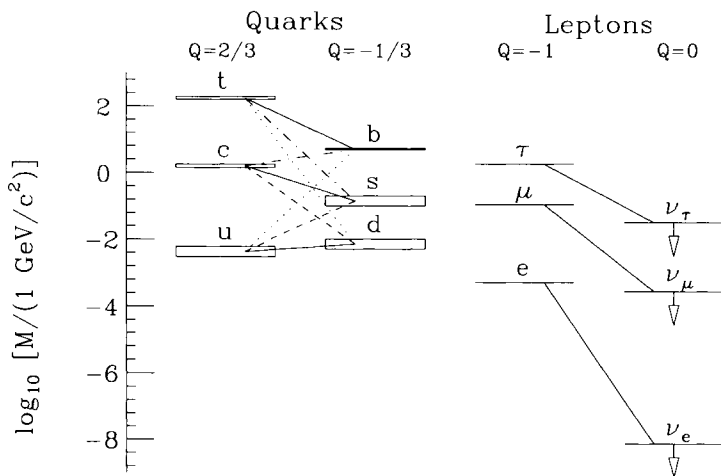
Figure 1. *Patterns of charge-changing weak transitions among quarks and leptons. The strongest inter-quark transitions correspond to the solid lines, with dashed, dot-dashed, and dotted lines corresponding to successively weaker transitions.*

1.1   Quarks and leptons

The fundamental building blocks of strongly interacting particles, the *quarks*, and the fundamental fermions lacking strong interactions, the *leptons*, are illustrated, along with their interactions, in Figure 1. Masses, as quoted by the Particle Data Group (2000), are summarised in Table 1. The relative strengths of the charge-current weak transitions between the quarks are summarised in Table 2.

The quark masses quoted in Table 1 are those which emerge when quarks are probed at distances short compared with 1 fm, the characteristic size of strongly interacting particles and the scale at which QCD becomes too strong to utilise perturbation theory. When regarded as constituents of strongly interacting particles, however, the u and d quarks act as quasi-particles with masses of about 0.3GeV. The corresponding "constituent-quark" masses of s, c, and b are about 0.5, 1.5, and 4.9GeV, respectively.

Quarks		Leptons	
Charge 2/3	Charge −1/3	Charge −1	Charge 0
Mass	Mass	Mass	Mass
u 0.001–0.005	d 0.003–0.009	e 0.000511	ν_e < 3 eV
c 1.15–1.35	s 0.075–0.175	μ 0.106	ν_μ < 190 keV
t 174.3 ± 5.1	b 4.0–4.4	τ 1.777	ν_τ < 18.2 MeV

Table 1. *The known quarks and leptons. Masses in GeV except where indicated otherwise. Here and elsewhere we take $c = 1$.*

Relative amplitude	Transition	Source of information (example)
~ 1	$u \leftrightarrow d$	Nuclear β-decay
~ 1	$c \leftrightarrow s$	Charmed particle decays
~ 0.22	$u \leftrightarrow s$	Strange particle decays
~ 0.22	$c \leftrightarrow d$	Neutrino prod. of charm
~ 0.04	$c \leftrightarrow b$	b decays
~ 0.003–0.004	$u \leftrightarrow b$	Charmless b decays
~ 1	$t \leftrightarrow b$	Dominance of $t \to Wb$
~ 0.04	$t \leftrightarrow s$	Only indirect evidence
~ 0.01	$t \leftrightarrow d$	Only indirect evidence

Table 2. *Relative strengths of charge-changing weak transitions.*

1.2   Colour and quantum chromo-dynamics

The quarks are distinguished from the leptons by possessing a three-fold charge known as "colour" which enables them to interact strongly with one another. (A gauged colour symmetry was first proposed by Nambu 1966.) We shall also speak of quark and lepton "flavour" when distinguishing the particles in Table 1 from one another. The experimental evidence for colour comes from several quarters.

   1. Quark statistics. One of the lowest-lying hadrons is a particle known as the Δ^{++}, an excited state of the nucleon first produced in $\pi^+ p$ collisions in the mid-1950s at the University of Chicago cyclotron. It can be represented in the quark model as uuu, so it is totally symmetric in flavour. It has spin $J = 3/2$, which is a totally symmetric combination of the three quark spins (each taken to be 1/2). Moreover, as a ground state, it is expected to contain no relative orbital angular momenta among the quarks.

   This leads to a paradox if there are no additional degrees of freedom. A state composed of fermions should be totally antisymmetric under the interchange of any two fermions, but what we have described so far is totally symmetric under flavour, spin, and space

interchanges, hence totally symmetric under their product. Colour introduces an additional degree of freedom under which the interchange of two quarks can produce a minus sign, through the representation $\Delta^{++} \sim \epsilon_{abc} u^a u^b u^c$. The totally antisymmetric product of three colour triplets is a colour singlet.

   2. Electron-positron annihilation to hadrons. The charges of all quarks which can be produced in pairs below a given centre-of-mass energy is measured by the ratio

$$R \equiv \frac{\sigma(e^+ e^- \to \text{hadrons})}{\sigma(e^+ e^- \to \mu^+ \mu^-)} = \sum_i Q_i^2 . \tag{1}$$

For energies at which only $u\bar{u}$, $d\bar{d}$, and $s\bar{s}$ can be produced, i.e., below the charmed-pair threshold of about 3.7 GeV, one expects

$$R = N_c \left[\left(\frac{2}{3}\right)^2 + \left(\frac{-1}{3}\right)^2 + \left(\frac{-1}{3}\right)^2 \right] = \frac{2}{3} N_c \tag{2}$$

for N_c "colours" of quarks. Measurements first performed at the Frascati laboratory in Italy and most recently at the Beijing Electron-Positron Collider (Bai *et al.* 2001; see Figure 2) indicate $R = 2$ in this energy range (with a small positive correction associated with the strong interactions of the quarks), indicating $N_c = 3$.
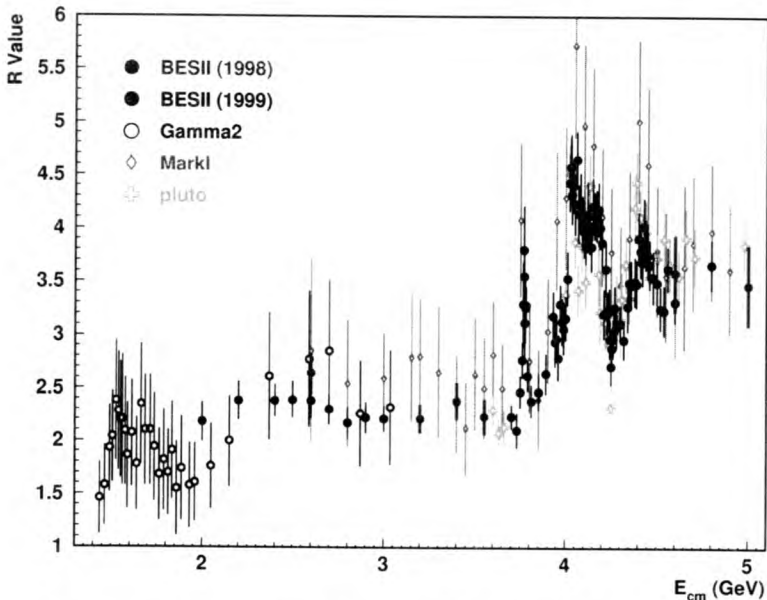
Figure 2. *Values of R measured by the BES Collaboration.*

   3. Neutral pion decay. The π^0 decay rate is governed by a quark loop diagram in which two photons are radiated by the quarks in $\pi^0 = (u\bar{u} - d\bar{d})/\sqrt{2}$. The predicted rate is

$$\Gamma(\pi^0 \to \gamma\gamma) = \frac{S^2 m_\pi^3}{8\pi f_\pi^2} \left(\frac{\alpha}{2\pi}\right)^2 , \tag{3}$$

where $f_\pi = 131$MeV and $S = N_c(Q_u^2 - Q_d^2) = N_c/3$. The experimental rate is 7.8 ± 0.6 eV, while (3) gives $7.6S^2$ eV, in accord with experiment if $S = 1$ and $N_c = 3$.

4. Triality. Quark composites appear only in multiples of three. Baryons are composed of qqq, while mesons are $q\bar{q}$ (with total quark number zero). This is compatible with our current understanding of QCD, in which only colour-singlet states can appear in the spectrum. Thus, mesons M and baryons B are represented by

$$M = \frac{1}{\sqrt{3}}(q^a \bar{q}'_a), \qquad B = \frac{1}{\sqrt{6}}(\epsilon_{abc} q^a q'^b q''^c). \qquad (4)$$

Direct evidence for the quanta of QCD, the gluons, was first presented in 1979 on the basis of extra "jets" of particles produced in electron-positron annihilation to hadrons. Normally one sees two clusters of energy associated with the fragmentation of each quark in $e^+ e^- \to q\bar{q}$ into hadrons. However, in some fraction of events an extra jet was seen, corresponding to the radiation of a gluon by one of the quarks.

The transformations which take one colour of quark into another are those of the group SU(3). We shall often refer to this group as SU(3)$_{\text{colour}}$ to distinguish it from the SU(3)$_{\text{flavour}}$ associated with the quarks u, d, and s.

1.3 Electroweak unification

The electromagnetic interaction is described in terms of photon exchange, for which the Born approximation leads to a matrix element behaving as $1/q^2$. Here q is the four-momentum transfer, and q^2 is its invariant square. The quantum electrodynamics of photons and charged point-like particles (such as electrons) initially encountered calculational problems in the form of divergent quantities, but these had been tamed by the late 1940s through the procedure known as *renormalisation*, leading to successful estimates of such quantities as the anomalous magnetic moment of the electron and the Lamb shift in hydrogen.

By contrast, the weak interactions as formulated up to the mid-1960s involved the point-like interactions of two currents, with interaction Hamiltonian $\mathcal{H}_W = G_F J_\mu J^{\mu\dagger}/\sqrt{2}$, where $G_F = 1.16637(1) \times 10^{-5}$GeV^{-2} the current value for the *Fermi coupling constant*. This interaction is very singular and cannot be renormalised. The weak currents J_μ in this theory were purely charge-changing. As a result of work by Lee and Yang, Feynman and Gell-Mann, and Marshak and Sudarshan in 1956–7 they were identified as having (vector)–(axial) or "$V - A$" form.

Hideki Yukawa (1935) and Oskar Klein (1938) proposed a boson-exchange model for the charge-changing weak interactions. Klein's model attempted a unification with electromagnetism and was based on a local isotopic gauge symmetry, thus anticipating the theory of Yang and Mills (1954). Julian Schwinger and others studied such models in the 1950s, but Glashow (1961) was the first to realise that a new *neutral* heavy boson had to be introduced as well in order to successfully unify the weak and electromagnetic interactions. The breaking of the electroweak symmetry (Weinberg 1967, Salam 1968) via the Higgs (1964) mechanism converted this phenomenological theory into one which could be used for higher-order calculations, as was shown by 't Hooft and Veltman in the early 1970s.

The boson-exchange model for charge-changing interactions replaces the Fermi interaction constant with a coupling constant g at each vertex and the low-q^2 limit of a propagator, $1/(M_W^2 - q^2) \to 1/M_W^2$, with factors of 2 chosen so that $G_F/\sqrt{2} = g^2/8M_W^2$. The q^2 term in the propagator helps the theory to be more convergent, but it is not the only ingredient needed, as we shall see.

The normalisation of the charge-changing weak currents J_μ suggested well in advance of electroweak unification that one regard the corresponding integrals of their time components (the so-called *weak charges*) as members of an SU(2) algebra (Gell-Mann and Lévy 1960, Cabibbo 1963). However, the identification of the neutral member of this multiplet as the electric charge was problematic. In the $V-A$ theory the W's couple only to left-handed fermions $\psi_L \equiv (1 - \gamma_5)\psi/2$, while the photon couples to $\psi_L + \psi_R$, where $\psi_R \equiv (1 + \gamma_5)\psi/2$. Furthermore, the high-energy behaviour of the $\nu\bar\nu \to W^+W^-$ scattering amplitude based on charged lepton exchange leads to unacceptable divergences if we incorporate it into the one-loop contribution to $\nu\bar\nu \to \nu\bar\nu$ (Quigg 1983).

A simple solution was to add a neutral boson Z coupling to W^+W^- and $\nu\bar\nu$ in such a way as to cancel the leading high-energy behaviour of the charged-lepton-exchange diagram. This relation between couplings occurs naturally in a theory based on the gauge group SU(2) $\otimes$ U(1). The Z leads to *neutral current interactions*, in which (for example) an incident neutrino scatters inelastically on a hadronic target without changing its charge. The discovery of neutral-current interactions of neutrinos and many other manifestations of the Z proved to be striking confirmations of the new theory.

Identifying the W^+ and W^- with raising and lowering operations in an SU(2), so that $W^\pm = (W^1 \mp iW^2)\sqrt{2}$, then left-handed fermions may be assigned to doublets of this "weak isospin", with $I_{3L}(u,c,t) = I_{3L}(\nu_e, \nu_\mu, \nu_\tau) = +1/2$ and $I_{3L}(d,s,b) = I_{3L}(e^-, \mu^-, \tau^-) = -1/2$. All the right-handed fermions have $I_L = I_{3L} = 0$. As mentioned, one cannot simply identify the photon with W^3, which also couples only to left-handed fermions. Instead, one must introduce another boson B associated with a U(1) gauge group. It will mix with the W^3 to form physical states consisting of the massless photon A and the massive neutral boson Z:

$$A = B\cos\theta + W^3\sin\theta, \qquad Z = -B\sin\theta + W^3\cos\theta. \tag{5}$$

The mixing angle θ appears in many electroweak processes. It has been measured to sufficiently great precision that one must specify the renormalisation scheme in which it is quoted. For present purposes we shall merely note that $\sin^2\theta \simeq 0.23$. The corresponding SU(2) and U(1) coupling constants g and g' are related to the electric charge e by $e = g\sin\theta = g'\cos\theta$, so that

$$\frac{1}{e^2} = \frac{1}{g^2} + \frac{1}{g'^2}. \tag{6}$$

The electroweak theory successfully predicted the masses of the $W^\pm$ and Z:

$$M_W \simeq 38.6 \text{ GeV}/\sin\theta \simeq 80.5 \text{ GeV}, \qquad M_Z \simeq M_W/\cos\theta \simeq 91.2 \text{ GeV}, \tag{7}$$

where we show the approximate *experimental* values. The detailed check of these predictions has reached the precision that one can begin to look into the deeper structure of the theory. A key ingredient in this structure is the *Higgs boson*, the price that had to be paid for the breaking of the electroweak symmetry.

1.4   Higgs boson

An unbroken $SU(2) \otimes U(1)$ theory involving the photon would require *all* fields to have zero mass, whereas the $W^\pm$ and Z are massive. The symmetry-breaking which generates W and Z masses must not destroy the renormalisability of the theory. However, a massive vector boson propagator is of the form

$$D_{\mu\nu}(q) = \frac{-g_{\mu\nu} + q_\mu q_\nu/M^2}{q^2 - M^2} \tag{8}$$

where M is the boson mass. The terms $q_\mu q_\nu$, when appearing in loop diagrams, will destroy the renormalisability of the theory. They are associated with longitudinal vector boson polarisations, which are only present for massive bosons. For massless bosons like the photon, there are only transverse polarisation states $J_z = \pm J$.

The *Higgs mechanism*, to be discussed in detail later in these lectures, provides the degrees of freedom needed to add a longitudinal polarisation state for each of W^+, W^-, and W^0. In the simplest model, this is achieved by introducing a doublet of complex Higgs fields:

$$\phi = \begin{bmatrix} \phi^+ \\ \phi^0 \end{bmatrix}, \qquad \phi^* = \begin{bmatrix} \bar{\phi}^0 \\ \phi^- \end{bmatrix}. \tag{9}$$

Here the charged Higgs fields $\phi^\pm$ provide the longitudinal component of $W^\pm$ and the linear combination $(\phi^0 - \bar{\phi}^0)/i\sqrt{2}$ provides the longitudinal component of the Z. The additional degree of freedom $(\phi^0 + \bar{\phi}^0)/\sqrt{2}$ corresponds to a physical particle, the *Higgs particle*, which is the subject of intense searches.

Discovering the nature of the Higgs boson is a key to further progress in understanding what may lie beyond the Standard Model. There may exist one Higgs boson or more than one. There may exist other particles in the spectrum related to it. The Higgs boson may be elementary or composite. If composite, it points to a new level of substructure of the elementary particles. Much of our discussion will lead up to strategies for the next few years designed to address these questions. First, we introduce the necessary topic of *gauge theories*, which have been the platform for all the developments of the past thirty years.

2   Gauge theories

2.1   Abelian gauge theories

The Lagrangian describing a free fermion of mass m is $\mathcal{L}_{\text{free}} = \bar{\psi}(i\,\slashed{\partial} - m)\psi$. It is invariant under the global phase change $\psi \to \exp(i\alpha)\psi$. (We shall always consider the fermion fields to depend on x.) Now consider independent phase changes at each point:

$$\psi \to \psi' \equiv \exp[i\alpha(x)]\psi. \tag{10}$$

Because of the derivative, the Lagrangian then acquires an additional phase change at each point: $\delta\mathcal{L}_{\text{free}} = \bar{\psi}i\gamma^\mu[i\partial_\mu\alpha(x)]\psi$. The free Lagrangian is not invariant under such changes of phase, known as *local gauge transformations*.

Local gauge invariance can be restored by the replacement $\partial_\mu \to D_\mu \equiv \partial_\mu + ieA_\mu$ in the free-fermion Lagrangian, which now becomes

$$\mathcal{L} = \overline{\psi}(i\,D\!\!\!\!/ - m)\psi = \overline{\psi}(i\,\partial\!\!\!/ - m)\psi - e\overline{\psi}\,A\!\!\!/(x)\psi\,. \tag{11}$$

The effect of a local phase in ψ can be compensated if we allow the *vector potential* A_μ to change by a total divergence, which does not change the electromagnetic field strength (defined as in Peskin and Schroeder 1995; Quigg 1983 uses the opposite sign)

$$F_{\mu\nu} \equiv \partial_\mu A_\nu - \partial_\nu A_\nu\,. \tag{12}$$

Indeed, under the transformation $\psi \to \psi'$ and with $A \to A'$ with A' yet to be determined, we have

$$\mathcal{L}' = \overline{\psi}'(i\,\partial\!\!\!/ - m)\psi' - e\overline{\psi}'A\!\!\!/'\psi' = \overline{\psi}(i\,\partial\!\!\!/ - m)\psi - \overline{\psi}[\partial\!\!\!/\alpha(x)]\psi - e\overline{\psi}A\!\!\!/'\psi\,. \tag{13}$$

This will be the same as $\mathcal{L}$ if

$$A'_\mu(x) = A_\mu(x) - \frac{1}{e}\partial_\mu\alpha(x)\,. \tag{14}$$

The derivative D_μ is known as the *covariant derivative*. One can check that under a local gauge transformation, $D_\mu\psi \to e^{i\alpha(x)}D_\mu\psi$.

Another way to see the consequences of local gauge invariance suggested by Yang (1974) and discussed by Peskin and Schroeder (1995, pp 482–486) is to define $-eA_\mu(x)$ as the local change in phase undergone by a particle of charge e as it passes along an infinitesimal space-time increment between x^μ and $x^\mu + dx^\mu$. For a space-time trip from point A to point B, the phase change is then

$$\Phi_{AB} = \exp\left(-ie\int_A^B A_\mu(x)dx^\mu\right)\,. \tag{15}$$

The phase in general will depend on the path in space-time taken from point A to point B. As a consequence, the phase Φ_{AB} is not uniquely defined. However, one can compare the result of a space-time trip along one path, leading to a phase $\Phi_{AB}^{(1)}$, with that along another, leading to a phase $\Phi_{AB}^{(2)}$. The two-slit experiment in quantum mechanics involves such a comparison; so does the Bohm-Aharonov effect in which a particle beam traveling past a solenoid on one side interferes with a beam traveling on the other side. Thus, phase differences

$$\Phi_{AB}^{(1)}\Phi_{AB}^{(2)*} = \Phi_C = \exp\left(-ie\oint A_\mu(x)dx^\mu\right)\,, \tag{16}$$

associated with *closed paths* in space-time (represented by the circle around the integral sign), are the ones which correspond to physical experiments. The phase Φ_C for a closed path C is independent of the phase convention for a charged particle at any space-time point x_0, since any change in the contribution to Φ_C from the integral up to x_0 will be compensated by an equal and opposite contribution from the integral departing from x_0.

The closed path integral (16) can be expressed as a surface integral using Stokes' theorem:

$$\oint A_\mu(x)dx^\mu = \int F_{\mu\nu}(x)d\sigma^{\mu\nu}\,, \tag{17}$$

where the electromagnetic field strength $F_{\mu\nu}$ was defined previously and $d\sigma^{\mu\nu}$ is an element of surface area. It is also clear that the closed path integral is invariant under changes (14) of $A_\mu(x)$ by a total divergence. Thus $F_{\mu\nu}$ suffices to describe all physical experiments as long as one integrates over a suitable domain. In the Bohm-Aharonov effect, in which a charged particle passes on either side of a solenoid, the surface integral will include the solenoid (in which the magnetic field is non-zero).

If one wishes to describe the energy and momentum of free electromagnetic fields, one must include a kinetic term $\mathcal{L}_K = -(1/4)F_{\mu\nu}F^{\mu\nu}$ in the Lagrangian, which now reads

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \overline{\psi}(i\,\partial\!\!\!/ - m)\psi - e\overline{\psi}\,A\!\!\!/\psi\,. \tag{18}$$

If the electromagnetic current is defined as $J_\mu^{\mathrm{em}} \equiv \overline{\psi}\gamma_\mu\psi$, then this Lagrangian leads to Maxwell's equations.

The local phase changes (10) form a U(1) group of transformations. Since such transformations commute with one another, the group is said to be *Abelian*. Electrodynamics, just constructed here, is an example of an *Abelian gauge theory*.

2.2 Non-Abelian gauge theories

One can imagine that a particle traveling in space-time undergoes not only phase changes, but also changes of identity. Such transformations were first considered by Yang and Mills (1954). For example, a quark can change in colour (red to blue) or flavour (u to d). In that case we replace the coefficient eA_μ of the infinitesimal displacement dx_μ by an $n \times n$ matrix $-g\mathbf{A}_\mu(x) \equiv -gA_\mu^i(x)\mathbf{T}_i$ acting in the n-dimensional space of the particle's degrees of freedom. (The sign change follows the convention of Peskin and Schroeder 1995.) For colours, $n = 3$. The $\mathbf{T}_i$ form a linearly independent basis set of matrices for such transformations, while the A_μ^i are their coefficients. The phase transformation then must take account of the fact that the matrices $\mathbf{A}_\mu(x)$ in general do not commute with one another for different space-time points, so that a *path-ordering* is needed:

$$\Phi_{AB} = \mathcal{P}\left[\exp\left(ig\int_A^B \mathbf{A}_\mu(x)dx^\mu\right)\right]\,. \tag{19}$$

When the basis matrices $\mathbf{T}_i$ do not commute with one another, the theory is *non-Abelian*.

We demand that changes in phase or identity conserve probability, i.e., that Φ_{AB} be *unitary*: $\Phi_{AB}^\dagger\Phi_{AB} = 1$. When Φ_{AB} is a matrix, the corresponding matrices $\mathbf{A}_\mu(x)$ in (19) must be Hermitian. If we wish to separate out pure phase changes, in which $\mathbf{A}_\mu(x)$ is a multiple of the unit matrix, from the remaining transformations, one may consider only transformations such that $\det(\Phi_{AB}) = 1$, corresponding to *traceless* $\mathbf{A}_\mu(x)$.

The $n \times n$ basis matrices $\mathbf{T}_i$ must then be Hermitian and traceless. There will be $n^2 - 1$ of them, corresponding to the number of independent SU(N) generators. (One can generalise this approach to other invariance groups.) The matrices will satisfy the commutation relations

$$[\mathbf{T}_i, \mathbf{T}_j] = ic_{ijk}\mathbf{T}_k\,, \tag{20}$$

where the c_{ijk} are *structure constants* characterising the group. For SU(2), $c_{ijk} = \epsilon_{ijk}$ (the Kronecker symbol), while for SU(3), $c_{ijk} = f_{ijk}$, where the f_{ijk} are defined in Gell-Mann

and Ne'eman (1964). A 3×3 representation in SU(3) is $\mathbf{T}_i = \lambda_i/2$, where $\lambda_i/2$ are the Gell-Mann matrices normalised such that Tr $\lambda_i\lambda_j = 2\delta_{ij}$. For this representation, then, Tr $\mathbf{T}_i\mathbf{T}_j = \delta_{ij}/2$.

In order to define the field-strength tensor $\mathbf{F}_{\mu\nu} = F^i_{\mu\nu}\mathbf{T}_i$ for a non-Abelian transformation, we may consider an infinitesimal closed-path transformation analogous to (16) for the case in which the matrices $\mathbf{A}_\mu(x)$ do not commute with one another. The result (see, e.g., Peskin and Schroeder 1995, pp 486–491) is

$$\mathbf{F}_{\mu\nu} = \partial_\mu\mathbf{A}_\nu - \partial_\nu\mathbf{A}_\nu - ig[\mathbf{A}_\mu, \mathbf{A}_\nu], \qquad F^i_{\mu\nu} = \partial_\mu A^i_\nu - \partial_\nu A^i_\mu + gc_{ijk}A^j_\mu A^k_\nu. \tag{21}$$

An alternative way to introduce non-Abelian gauge fields is to demand that, by analogy with (10), a theory involving fermions ψ be invariant under local transformations

$$\psi(x) \to \psi'(x) = U(x)\psi(x), \qquad U^\dagger U = 1, \tag{22}$$

where for simplicity we consider unitary transformations. Under this replacement we find that $\mathcal{L} \to \mathcal{L}'$, where

$$\begin{aligned}
\mathcal{L}' \equiv \overline{\psi}'(i\,\slashed{\partial} - m)\psi' &= \overline{\psi}U^{-1}(i\,\slashed{\partial} - m)U\psi \\
&= \overline{\psi}(i\,\slashed{\partial} - m)\psi + i\psi U^{-1}\gamma^\mu(\partial_\mu U)\psi.
\end{aligned} \tag{23}$$

As in the Abelian case, an extra term is generated by the local transformation. It can be compensated by replacing ∂_μ by

$$\partial_\mu \to \mathbf{D}_\mu \equiv \partial_\mu - ig\mathbf{A}_\mu(x). \tag{24}$$

In this case $\mathcal{L} = \overline{\psi}(i\,\slashed{D} - m)\psi$ and under the change (22) we find

$$\begin{aligned}
\mathcal{L}' \equiv \overline{\psi}'(i\,\slashed{D}' - m)\psi' &= \overline{\psi}U^{-1}(i\,\slashed{\partial} + g\,\slashed{A}' - m)U\psi \\
&= \mathcal{L} + \overline{\psi}[g(U^{-1}\slashed{A}'U - \slashed{A}) + iU^{-1}(\slashed{\partial}U)]\psi.
\end{aligned} \tag{25}$$

This is equal to $\mathcal{L}$ if we take

$$\mathbf{A}'_\mu = U\mathbf{A}_\mu U^{-1} - \frac{i}{g}(\partial_\mu U)U^{-1}. \tag{26}$$

This reduces to our previous expressions if $g = -e$ and $U = e^{i\alpha(x)}$.

The covariant derivative acting on ψ transforms in the same way as ψ itself under a gauge transformation: $\mathbf{D}_\mu\psi \to \mathbf{D}'_\mu\psi' = U\mathbf{D}_\mu\psi$. The field strength $\mathbf{F}_{\mu\nu}$ transforms as $\mathbf{F}_{\mu\nu} \to \mathbf{F}'_{\mu\nu} = U\mathbf{F}_{\mu\nu}U^{-1}$. It may be computed via $[\mathbf{D}_\mu, \mathbf{D}_\nu] = -ig\mathbf{F}_{\mu\nu}$; both sides transform as $U(\quad)U^{-1}$ under a local gauge transformation.

In order to obtain propagating gauge fields, as in electrodynamics, one must add a kinetic term $\mathcal{L}_K = -(1/4)F^i_{\mu\nu}F^{i\mu\nu}$ to the Lagrangian. Recalling the representation $\mathbf{F}_{\mu\nu} = F^i_{\mu\nu}$ in terms of gauge group generators normalised such that $\mathrm{Tr}(\mathbf{T}_i\mathbf{T}_j) = \delta_{ij}/2$, we can write the full Yang-Mills Lagrangian for gauge fields interacting with matter fields as

$$\mathcal{L} = -\frac{1}{2}\mathrm{Tr}(\mathbf{F}_{\mu\nu}\mathbf{F}^{\mu\nu}) + \overline{\psi}(i\,\slashed{D} - m)\psi. \tag{27}$$

We shall use Lagrangians of this type to derive the strong, weak, and electromagnetic interactions of the "Standard Model."

The interaction of a gauge field with fermions then corresponds to a term in the interaction Lagrangian $\Delta\mathcal{L} = g\overline{\psi}(x)\gamma^{\mu}\mathbf{A}_{\mu}(x)\psi(x)$. The $[\mathbf{A}_{\mu}, \mathbf{A}_{\nu}]$ term in $\mathbf{F}_{\mu\nu}$ leads to self-interactions of non-Abelian gauge fields, arising solely from the kinetic term. Thus, one has three- and four-field vertices arising from

$$\Delta\mathcal{L}_K^{(3)} = (\partial_\mu A_\nu^i)gc_{ijk}A^{\mu j}A^{\nu k}, \qquad \Delta\mathcal{L}_K^{(4)} = -\frac{g^2}{4}c_{ijk}c_{imn}A^{\mu j}A^{\nu k}A_\mu^m A_\nu^n. \tag{28}$$

These self-interactions are an important aspect of non-Abelian gauge theories and are responsible in particular for the remarkable *asymptotic freedom* of QCD which leads to its becoming weaker at short distances, permitting the application of perturbation theory.

2.3 Elementary divergent quantities

In most quantum field theories, including quantum electrodynamics, divergences occurring in higher orders of perturbation theory must be removed using charge, mass, and wave function renormalisation. This is conventionally done at intermediate calculational stages by introducing a cutoff momentum scale Λ or analytically continuing the number of space-time dimensions away from four. Thus, a vacuum polarisation graph in QED associated with external photon momentum k and a fermion loop will involve an integral

$$\Pi_{\mu\nu}(k) \sim \int \frac{d^4p}{(2\pi)^4} \text{Tr}\left(\frac{1}{\not{p}-m}\gamma_\mu \frac{1}{\not{p}+\not{k}-m}\gamma_\nu\right): \tag{29}$$

a self-energy of a fermion with external momentum p will involve

$$\Sigma(p) \sim \int \frac{d^4q}{(2\pi)^4}\frac{1}{q^2}\gamma_\mu\frac{1}{\not{p}+\not{q}-m}\gamma^\mu, \tag{30}$$

and a fermion-photon vertex function with external fermion momenta p, p' will involve

$$\Lambda_\mu(p',p) \sim \int \frac{d^4k}{(2\pi)^4}\frac{1}{k^2}\gamma_\nu\frac{1}{\not{p}'+\not{k}-m}\gamma_\mu\frac{1}{\not{p}+\not{k}-m}\gamma^\nu. \tag{31}$$

The integral (29) appears to be quadratically divergent. However, the gauge invariance of the theory translates into the requirement $k^\mu\Pi_{\mu\nu} = 0$, which requires $\Pi_{\mu\nu}$ to have the form

$$\Pi_{\mu\nu}(k) = (k^2 g_{\mu\nu} - k_\mu k_\nu)\Pi(k^2). \tag{32}$$

The corresponding integral for $\Pi(k^2)$ then will be only logarithmically divergent. The integral in (30) is superficially linearly divergent but in fact its divergence is only logarithmic, as is the integral in (31).

Unrenormalised functions describing vertices and self-energies involving n_B external boson lines and n_F external fermion lines may be defined in terms of a momentum cutoff Λ and a bare coupling constant g_0 (Coleman 1971, Ellis 1977, Ross 1978):

$$\Gamma^U_{n_B,n_F} \equiv \Gamma^U_{n_B,n_F}(p_i, g_0, \Lambda), \tag{33}$$

where p_i denote external momenta. *Renormalised* functions Γ^R may be defined in terms of a scale parameter μ, a renormalised coupling constant $g = g(g_0, \Lambda/\mu)$, and renormalisation constants $Z_B(\Lambda)$ and $Z_F(\Lambda)$ for the external boson and fermion wave functions:

$$\Gamma^R(p_i, g, \mu) \equiv \lim_{\Lambda\to\infty}[Z_B(\Lambda)]^{n_B}[Z_F(\Lambda)]^{n_F}\Gamma^U_{n_B,n_F}(p_i, g_0, \Lambda). \tag{34}$$

The scale μ is typically utilised by demanding that Γ^R be equal to some predetermined function at a Euclidean momentum $p^2 = -\mu^2$. Thus, for the one-boson, two-fermion vertex, we take

$$\Gamma^R_{1,2}(0,p,-p)\big|_{p^2=-\mu^2} = \lim_{\Lambda\to\infty} Z^2_F Z_B \Gamma^U_{1,2}(0,p,-p)\big|_{p^2=-\mu^2} \equiv g\,. \tag{35}$$

The unrenormalised function Γ^U is *independent* of μ, while Γ^R and the renormalisation constants $Z_B(\Lambda)$, $Z_F(\Lambda)$ will depend on μ. For example, in QED, the photon wave function renormalisation constant (known as Z_3) behaves as

$$Z_3 = 1 - \frac{\alpha_0}{3\pi} \ln \frac{\Lambda^2}{\mu^2}\,. \tag{36}$$

The bare charge e_0 and renormalised charge e are related by $e = e_0 Z_3^{1/2}$. To lowest order in perturbation theory, $e < e_0$. The vacuum behaves as a normal dielectric; charge is screened. It is the exception rather than the rule that in QED one can define the renormalised charge for $q^2 = 0$; in QCD we shall see that this is not possible.

2.4   Scale changes and the beta function

We differentiate both sides of (34) with respect to μ and multiply by μ. Since the functions Γ^U are independent of μ, we find

$$\left(\mu\frac{\partial}{\partial\mu} + \mu\frac{\partial g}{\partial\mu}\frac{\partial}{\partial g}\right)\Gamma^R(p_i,g,\mu) = \lim_{\Lambda\to\infty}\left(\frac{n_B}{Z_B}\mu\frac{\partial Z_B}{\partial\mu} + \frac{n_F}{Z_F}\mu\frac{\partial Z_F}{\partial\mu}\right)Z^{n_B}_B Z^{n_F}_F \Gamma^U\,, \tag{37}$$

or

$$\left[\mu\frac{\partial}{\partial\mu} + \beta(g)\frac{\partial}{\partial g} + n_B\gamma_B(g) + n_F\gamma_F(g)\right]\Gamma^R(p_i,g,\mu) = 0\,, \tag{38}$$

where

$$\beta(g) \equiv \mu\frac{\partial g}{\partial\mu}\,, \qquad \gamma_B(g) \equiv -\frac{\mu}{Z_B}\frac{\partial Z_B}{\partial\mu}\,, \qquad \gamma_F(g) \equiv -\frac{\mu}{Z_F}\frac{\partial Z_F}{\partial\mu}\,. \tag{39}$$

The behaviour of any generalised vertex function Γ^R under a change of scale μ is then governed by the universal functions (39).

Here we shall be particularly concerned with the function $\beta(g)$. Let us imagine $\mu \to \lambda\mu$ and introduce the variables $t \equiv \ln\lambda$, $\bar{g}(g,t) \equiv g(g_0,\Lambda/\lambda\mu)$, Then the relation for the beta-function may be written

$$\frac{d\bar{g}(g,t)}{dt} = \beta(\bar{g})\,, \qquad \bar{g}(g,0) = g(g_0,\Lambda/\mu) = g\,. \tag{40}$$

Let us compare the behaviour of $\bar{g}$ with increasing t (larger momentum scales or shorter distance scales) depending on the sign of $\beta(\bar{g})$. In general we will find $\beta(0) = 0$. We take $\beta(\bar{g})$ to have zeroes at $\bar{g} = 0$, g_1, g_2, $\ldots$. Then:

1. Suppose $\beta(\bar{g}) > 0$. Then $\bar{g}$ *increases* from its $t = 0$ value $\bar{g} = g$ until a zero g_i of $\beta(\bar{g})$ is encountered. Then $\bar{g} \to g_i$ as $t \to \infty$.

2. Suppose $\beta(\bar{g}) < 0$. Then $\bar{g}$ *decreases* from its $t = 0$ value $\bar{g} = g$ until a zero g_i of $\beta(\bar{g})$ is encountered.

In either case $\bar{g}$ approaches a point at which $\beta(\bar{g}) = 0$, $\beta'(\bar{g}) < 0$ as $t \to \infty$. Such points are called *ultraviolet fixed points*. Similarly, points for which $\beta(\bar{g}) = 0$, $\beta'(\bar{g}) > 0$ are *infrared fixed points*, and $\bar{g}$ will tend to them for $t \to -\infty$ (small momenta or large distances). The point $e = 0$ is an infrared fixed point for quantum electrodynamics, since $\beta'(e) > 0$ at $e = 0$.

It may happen that $\beta'(0) < 0$ for specific theories. In that case $\bar{g} = 0$ is an ultraviolet fixed point, and the theory is said to be *asymptotically free*. We shall see that this property is particular to non-Abelian gauge theories (Gross and Wilczek 1973, Politzer 1974).

2.5 Beta function calculation

In quantum electrodynamics a loop diagram involving a fermion of unit charge contributes the following expression to the relation between the bare charge e_0 and the renormalised charge e:

$$e = e_0 \left(1 - \frac{\alpha_0}{3\pi} \ln \frac{\Lambda}{\mu}\right), \tag{41}$$

as implied by (35) and (36), where $\alpha_0 \equiv e_0^2/4\pi$. We find

$$\beta(e) = \frac{e_0^3}{12\pi^2} \simeq \frac{e^3}{12\pi^2}, \tag{42}$$

where differences between e_0 and e correspond to higher-order terms in e. Thus $\beta(e) > 0$ for small e and the coupling constant becomes stronger at larger momentum scales (shorter distances).

We shall show an extremely simple way to calculate (42) and the corresponding result for a charged scalar particle in a loop. From this we shall be able to first calculate the effect of a charged vector particle in a loop (a calculation first performed by Khriplovich 1969) and then generalise the result to Yang-Mills fields. The method follows that of Hughes (1980).

When one takes account of vacuum polarisation, the electromagnetic interaction in momentum space may be written

$$\frac{e^2}{q^2} \to \frac{e^2}{q^2[1 + \Pi(q^2)]} \tag{43}$$

Here the long-distance $(q^2 \to 0)$ behaviour has been defined such that e is the charge measured at macroscopic distances, so $\Pi(0) = 0$. Following Sakurai (1967), we shall reconstruct $\Pi_i(q^2)$ for a loop involving the fermion species i from its imaginary part, which is measurable through the cross section for $e^+ e^- \to i\bar{i}$:

$$\text{Im } \Pi_i(s) = \frac{s}{4\pi\alpha}\sigma(e^+ e^- \to i\bar{i}), \tag{44}$$

where s is the square of the centre-of-mass energy and $\alpha \equiv e^2/4\pi$. For fermions f of charge e_f and mass m_f,

$$\text{Im } \Pi_f(s) = \frac{\alpha e_f^2}{3}\left(1 + \frac{2m_f^2}{s}\right)\left(1 - \frac{4m_f^2}{s}\right)^{1/2}\theta(s - 4m_f^2), \tag{45}$$

while for scalar particles of charge e_s and mass m_s,

$$\text{Im } \Pi_s(s) = \frac{\alpha e_s^2}{12}\left(1 - \frac{4m_s^2}{s}\right)^{3/2}\theta(s - 4m_s^2)\,. \tag{46}$$

The corresponding cross section for $e^+e^- \to \mu^+\mu^-$, neglecting the muon mass, is $\sigma(e^+e^- \to \mu^+\mu^-) = 4\pi\alpha^2/3s$, so one can define

$$R_i \equiv \sigma(e^+e^- \to i\bar{i})/\sigma(e^+e^- \to \mu^+\mu^-)\,, \tag{47}$$

in terms of which $\text{Im } \Pi_i(s) = \alpha R_i(s)/3$. For $s \to \infty$ one has $R_f(s) \to e_f^2$ for a fermion and $R_s(s) \to e_s^2/4$ for a scalar.

The full vacuum polarisation function $\Pi_i(s)$ cannot directly be reconstructed in terms of its imaginary part via the dispersion relation

$$\Pi_i(s) = \frac{1}{\pi}\int_{4m^2}^{\infty}\frac{ds'}{s' - s}\text{Im } \Pi_i(s')\,, \tag{48}$$

since the integral is logarithmically divergent. This divergence is exactly that encountered earlier in the discussion of renormalisation. For quantum electrodynamics we could deal with it by defining the charge at $q^2 = 0$ and hence taking $\Pi_i(0) = 0$. The *once-subtracted* dispersion relation for $\Pi_i(s) - \Pi_i(0)$ would then converge:

$$\Pi_i(s) = \frac{1}{\pi}\int_{4m^2}^{\infty}\frac{ds'}{s'(s' - s)}\text{Im } \Pi_i(s')\,. \tag{49}$$

However, in order to be able to consider cases such as Yang-Mills fields in which the theory is not well-behaved at $q^2 = 0$, let us instead define $\Pi_i(-\mu^2) = 0$ at some spacelike scale $q^2 = -\mu^2$. The dispersion relation is then

$$\Pi_i(s) = \frac{1}{\pi}\int_{4m^2}^{\infty}ds'\left[\frac{1}{s' - s} - \frac{1}{s' + \mu^2}\right]\text{Im } \Pi_i(s')\,. \tag{50}$$

For $|q^2| \gg \mu^2 \gg m^2$, we find

$$\Pi_i(q^2) \to -\frac{\alpha}{3\pi}R_i(\infty)\left[\ln\frac{-q^2}{\mu^2} + \text{const.}\right]\,, \tag{51}$$

and so, from (43), the "charge at scale q" may be written as

$$e_q^2 \equiv \frac{e^2}{1 + \Pi_i(q^2)} \simeq e^2\left[1 + \frac{\alpha}{3\pi}R_i(\infty)\ln\frac{-q^2}{\mu^2}\right]\,. \tag{52}$$

The beta-function here is defined by $\beta(e) = \mu(\partial e/\partial \mu)$ at fixed e_q. This expression gives $\beta(e) = -\beta_0 e^3/(16\pi^2) + \mathcal{O}(e^5)$ and one finds $\beta_0 = -(4/3)e_f^2$ for spin-1/2 fermions and $\beta_0 = -(1/3)e_s^2$ for scalars.

These results will now be used to find the value of β_0 for a single charged massless vector field. We generalise the results for spin 0 and 1/2 to higher spins by splitting contributions to vacuum polarisation into "convective" and "magnetic" ones. Furthermore, we take into account the fact that a closed fermion loop corresponds to an extra minus sign in $\Pi_f(s)$

(which is already included in our result for spin 1/2). The "magnetic" contribution of a particle with spin projection S_z must be proportional to S_z^2. For a massless spin-S particle, $S_z^2 = S^2$. We may then write

$$\beta_0 = \begin{cases} (-1)^{n_F}(aS^2 + b)(S = 0)\,, \\ (-1)^{n_F}(aS^2 + 2b)(S \neq 0)\,, \end{cases} \tag{53}$$

where $n_F = 1$ for a fermion, 0 for a boson. The factor of $2b$ for $S \neq 0$ comes from the contribution of each polarisation state $(S_z = \pm S)$ to the convective term. Matching the results for spins 0 and 1/2,

$$-\frac{1}{3} = b\,, \quad -\frac{4}{3} = -\left(\frac{a}{4} + 2b\right)\,, \tag{54}$$

we find $a = 8$ and hence for $S = 1$

$$\beta_0 = 8 - \frac{2}{3} = \frac{22}{3}\,. \tag{55}$$

The magnetic contribution is by far the dominant one (by a factor of 12), and is of opposite sign to the convective one. A similar separation of contributions, though with different interpretations, was obtained in the original calculation of Khriplovich (1969). The reversal of sign with respect to the scalar and spin-1/2 results is notable.

2.6   Group-theoretic techniques

The result (55) for a charged, massless vector field interacting with the photon is also the value of β_0 for the Yang-Mills group $SO(3) \sim SU(2)$ if we identify the photon with A_μ^3 and the charged vector particles with $A_\mu^\pm \equiv (A_\mu^1 \mp iA_\mu^2)/\sqrt{2}$. We now generalise it to the contribution of gauge fields in an arbitrary group G.

The value of $\beta_{0\text{ gauge fields}}$ depends on a sum over all possible self-interacting gauge fields that can contribute to the loop with external gauge field labels i and m:

$$\frac{\beta_0[G]}{\beta_0[SU(2)]} = \frac{c_{ijk}^G c_{mjk}^G}{c_{ijk}^{SU(2)} c_{mjk}^{SU(2)}}\,, \tag{56}$$

where c_{ijk}^G is the structure constant for G, introduced in (20). The sums in (56) are proportional to δ_{im}:

$$c_{ijk} c_{mjk} = \delta_{im} C_2(A)\,. \tag{57}$$

The quantity $C_2(A)$ is the *quadratic Casimir operator* for the adjoint representation of the group G.

Since the structure constants for $SO(3) \sim SU(2)$ are just $c_{ijk}^{SU(2)} = \epsilon_{ijk}$, one finds $C_2(A) = 2$ for $SU(2)$, so the generalisation of (55) is that $\beta_{0\text{ gauge fields}} = (11/3)C_2(A)$.

The contributions of arbitrary scalars and spin-1/2 fermions in representations R are proportional to $T(R)$, where

$$\text{Tr } (T_i T_j) \equiv \delta_{ij} T(R) \tag{58}$$

for matrices T_i in the representation R. For a single charged scalar particle (e.g., a pion) or fermion (e.g., an electron), $T(R) = 1$. Thus $\beta_{0\text{ spin }0} = -(1/3)T_0(R)$, while

β_0 spin 1/2 $= -(4/3)T_{1/2}(R)$, where the subscript on $T(R)$ denotes the spin. Summarising the contributions of gauge bosons, spin 1/2 fermions, and scalars, we find

$$\beta_0 = \frac{11}{3}C_2(A) - \frac{4}{3}\sum_f T_{1/2}(R_f) - \sum_s \frac{1}{3}T_0(R_s)\,. \tag{59}$$

One often needs the beta-function to higher orders, notably in QCD where the perturbative expansion coefficient is not particularly small. It is

$$\beta(\bar{g}) = -\beta_0\frac{\bar{g}^3}{16\pi^2} - \beta_1\frac{\bar{g}^5}{(16\pi^2)^2} + \cdots\,, \tag{60}$$

where the result for gauge bosons and spin 1/2 fermions (Caswell 1974) is

$$\beta_1 = \frac{2}{3}\left\{17[C_2(A)]^2 - 10T(R)C_2(A) - 6T(R)C_2(R)\right\}\,. \tag{61}$$

The first term involves loops exclusively of gauge bosons. The second involves single-gauge-boson loops with a fermion loop on one of the gauge boson lines. The third involves fermion loops with a fermion self-energy due to a gauge boson. The quantity $C_2(R)$ is defined such that

$$[T^i(R)T^i(R)]_{\alpha\beta} = C_2(R)\delta_{\alpha\beta}\,, \tag{62}$$

where α and β are indices in the fermion representation.

We now illustrate the calculation of $C_2(A)$, $T(R)$, and $C_2(R)$ for SU(N). More general techniques are given by Slansky (1981).

Any SU(N) group contains an SU(2) subgroup, which we may take to be generated by T_1, T_2, and T_3. The isospin projection I_3 may be identified with T_3. Then the I_3 value carried by each generator T_i (written for convenience in the fundamental N-dimensional representation) may be identified as shown below:

← 2 →		← N − 2 →		
0	1	1/2	$\cdots$	1/2
−1	0	−1/2	$\cdots$	−1/2
−1/2	1/2	0	$\cdots$	0
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
−1/2	1/2	0	$\cdots$	0

Since $C_2(A)$ may be calculated for any convenient value of the index $i = m$ in (57), we chose $i = m = 3$. Then

$$C_2(A) = \sum_{\text{adjoint}}(I_3)^2 = 1 + 1 + 4(N-2)\left(\frac{1}{2}\right)^2 = N\,. \tag{63}$$

As an example, the octet (adjoint) representation of SU(3) has two members with $|I_3| = 1$ (e.g., the charged pions) and four with $|I_3| = 1/2$ (e.g., the kaons).

For members of the fundamental representation of SU(N), there will be one member with $I_3 = +1/2$, another with $I_3 = -1/2$, and all the rest with $I_3 = 0$. Then again choosing $i = m = 3$ in (58), we find $T(R)|_{\text{fundamental}} = 1/2$. The SU(N) result for β_0 in the presence of n_f spin 1/2 fermions and n_s scalars in the fundamental representation then may be written

$$\beta_0 = \frac{11}{3}N - \frac{2}{3}n_f - \frac{1}{6}n_s \,. \tag{64}$$

The quantity $C_2(A)$ in (63) is most easily calculated by averaging over all indices $\alpha = \beta$. If all generators T^i are normalised in the same way, one may calculate the result for an *individual* generator (say, T_3) and then multiply by the number of generators $[N^2 - 1$ for SU(N)]. For the fundamental representation one then finds

$$C_2(R) = \frac{1}{N}(N^2 - 1)\left[\left(\frac{1}{2}\right)^2 + \left(-\frac{1}{2}\right)^2\right] = \frac{N^2 - 1}{2N} \,. \tag{65}$$

2.7   The running coupling constant

One may integrate (60) to obtain the coupling constant as a function of momentum scale M and a scale-setting parameter Λ. In terms of $\bar{\alpha} \equiv \bar{g}^2/4\pi$, one has

$$\frac{d\bar{\alpha}}{dt'} = -\beta_0 \frac{\bar{\alpha}^2}{4\pi} - \beta_1 \frac{\bar{\alpha}^3}{(4\pi)^2} \,, \qquad t' \equiv 2t = \ln\left(\frac{M^2}{\Lambda^2}\right) \,. \tag{66}$$

For large t' the result can be written as

$$\bar{\alpha}(M^2) = \frac{4\pi}{\beta_0 t'}\left[1 - \frac{\beta_1}{\beta_0^2}\frac{\ln t'}{t'}\right] + \mathcal{O}(t'^{-2}) \,. \tag{67}$$

Suppose a process involves p powers of $\bar{\alpha}$ to leading order and a correction of order $\bar{\alpha}^{p+1}$:

$$\Gamma = A\bar{\alpha}^p[1 + B\bar{\alpha} + \mathcal{O}(\bar{\alpha}^2)] \,. \tag{68}$$

If Λ is rescaled to $\lambda\Lambda$, then $t' \to t' - 2\ln\lambda = t'(1 - 2\ln\lambda/t')$, so

$$\bar{\alpha}^p \to \bar{\alpha}^p\left(1 + \frac{p\beta_0}{2\pi}\bar{\alpha}\ln\lambda\right) \,. \tag{69}$$

The coefficient B thus depends on the scale parameter used to define $\bar{\alpha}$.

Many prescriptions have been adopted for defining Λ. In one ('t Hooft 1973), the "minimal subtraction" or MS scheme, ultraviolet logarithmic divergences are parameterised by continuing the space-time dimension $d = 4$ to $d = 4 - \epsilon$ and subtracting pole terms $\int d^{4-\epsilon}/p^4 \sim 1/\epsilon$. In another (Bardeen *et al.* 1978) (the "modified minimal subtraction" or $\overline{\text{MS}}$ scheme) a term

$$\frac{1}{\hat{\epsilon}} = \frac{1}{\epsilon} + \frac{\ln 4\pi - \gamma_E}{2} \tag{70}$$

containing additional finite pieces is subtracted. Here $\gamma_E = 0.5772$ is Euler's constant, and one can show that $\Lambda_{\overline{\text{MS}}} = \Lambda_{\text{MS}}\exp[(\ln 4\pi - \gamma_E)/2]$. Many $\mathcal{O}(\bar{\alpha})$ corrections are quoted in the $\overline{\text{MS}}$ scheme. Specification of Λ in any scheme is equivalent to specification of $\bar{\alpha}(M^2)$.
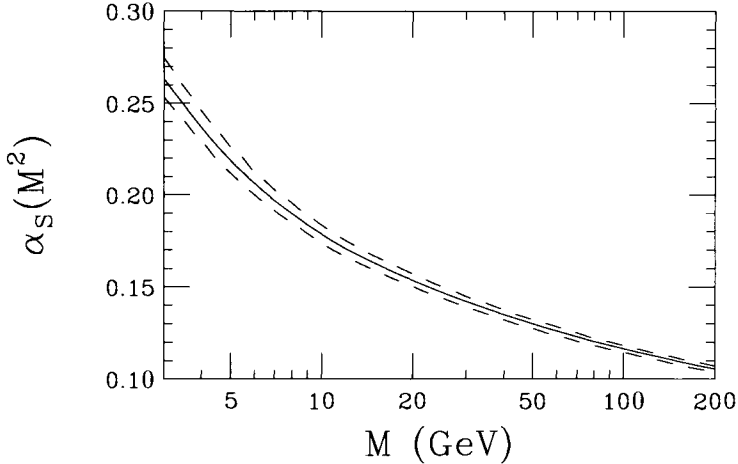
Figure 3. *Scale-dependence of the strong-coupling constant $\bar{\alpha}_S(M^2)$ subject to the constraint $\bar{\alpha}_S(M_Z^2) = 0.118 \pm 0.002$. The solid line shows the central value; dashed lines indicate $\pm 1\sigma$ limits.*

2.8   Applications to quantum chromodynamics

A "golden application" of the running coupling constant to QCD is the effect of gluon radiation on the value of R in $e^+ e^-$ annihilations. Since R is related to the imaginary part of the photon vacuum polarisation function $\Pi(s)$ which we have calculated for fermions and scalar particles, one calculates the effects of gluon radiation by calculating the correction to $\Pi(s)$ due to internal gluon lines. The leading-order result for colour-triplet quarks is $R(s) \to R(s)[1 + \bar{\alpha}(s)/\pi]$. There are many values of s at which one can measure such effects. For example, at the mass of the Z, the partial decay rate of the Z to hadrons involves the same correction, and leads to the estimate $\bar{\alpha}_S(M_Z^2) = 0.118 \pm 0.002$. The dependence of $\bar{\alpha}_S(M^2)$ satisfying this constraint on M^2 is shown in Figure 3. As we shall see in Section 5.1, the *electromagnetic* coupling constant also runs, but much more slowly, with α^{-1} changing from 137.036 at $q^2 = 0$ to about 129 at $q^2 = M_Z^2$.

A system which illustrates both perturbative and non-perturbative aspects of QCD is the bound state of a heavy quark and a heavy antiquark, known as *quarkonium* (in analogy with positronium, the bound state of a positron and an electron). We show in Figures 4 and 5 the spectrum of the $c\bar{c}$ and $b\bar{b}$ bound states (Rosner 1997). The charmonium ($c\bar{c}$) system was an early laboratory of QCD (Appelquist and Politzer 1975).

The S-wave ($L = 0$) levels have total angular momentum J, parity P, and charge-conjugation eigenvalue C equal to $J^{PC} = 0^{+-}$ and 1^{--} as one would expect for $^1S_0$ and $^3S_1$ states, respectively, of a quark and antiquark. The P-wave ($L = 1$) levels have $J^{PC} = 1^{+-}$ for the $^1P_1$, 0^{++} for the $^3P_0$, 1^{++} for the $^3P_1$, and 2^{++} for the $^3P_2$. The $J^{PC} = 1^{--}$ levels are identified as such by their copious production through single virtual photons in $e^+ e^-$ annihilations. The 0^{-+} level η_c is produced via single-photon emission from the J/ψ (so its C is positive) and has been directly measured to have J^P compatible with 0^-. Numerous studies have been made of the electromagnetic (electric dipole) transitions between the S-wave and P-wave levels and they, too, support the assignments shown.
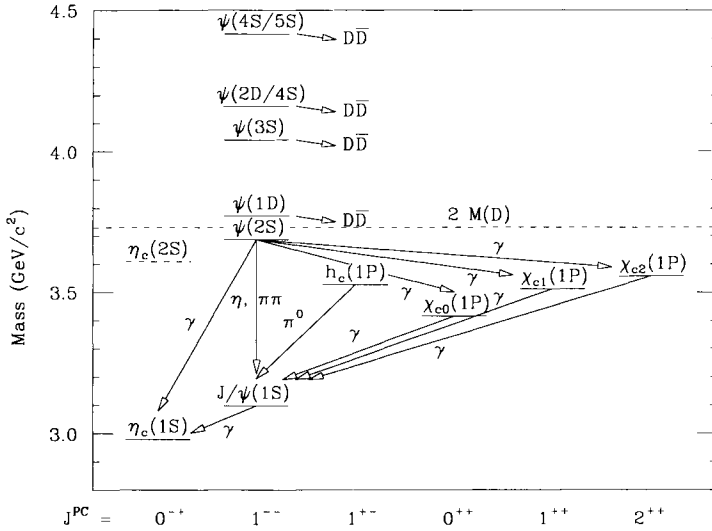
Figure 4. *Charmonium (cc̄) spectrum. Observed and predicted levels are denoted by solid and dashed horizontal lines, respectively. Arrows denote electromagnetic transitions (labeled by γ) and hadronic transitions (labeled by emitted hadrons).*
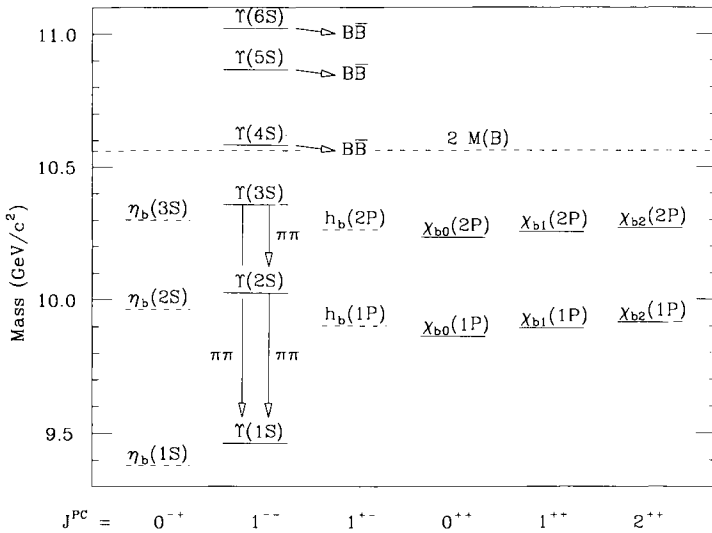
Figure 5. *Spectrum of bb̄ states. Observed and predicted levels are denoted by solid and dashed horizontal lines, respectively. In addition to the transitions labeled by arrows, numerous electric dipole transitions and decays of states below BB̄ threshold to hadrons containing light quarks have been seen.*
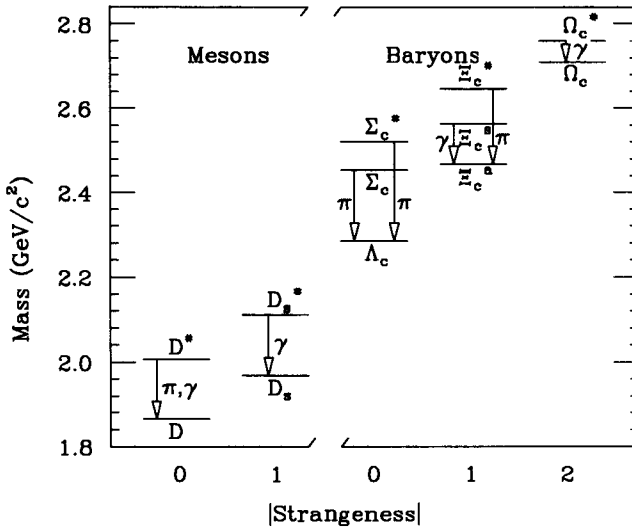
Figure 6. *Spectrum of lowest-lying states containing one charmed and one light quark. Observed and predicted levels are denoted by solid and broken horizontal lines, respectively.*
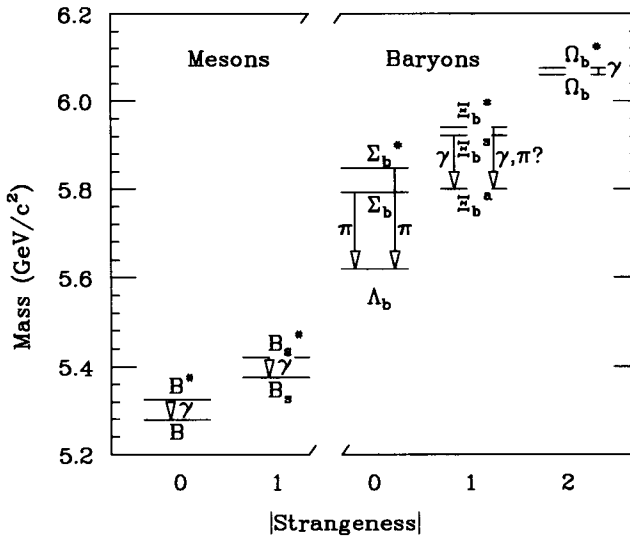
Figure 7. *Spectrum of lowest-lying states containing one bottom and one light quark. Observed and predicted levels are denoted by solid and broken horizontal lines, respectively.*

The $b\bar{b}$ and $c\bar{c}$ levels have a very similar structure, aside from an overall shift. The similarity of the $c\bar{c}$ and $b\bar{b}$ spectra is in fact an accident of the fact that for the interquark distances in question (roughly 0.2 to 1 fm), the interquark potential interpolates between short-distance Coulomb-like and long-distance linear behavior. The Coulomb-like behavior is what one would expect from single-gluon exchange, while the linear behavior is a particular feature of non-perturbative QCD which follows from Gauss' law if chromoelectric flux lines are confined to a fixed area between two widely separated sources (Nambu 1974). It has been explicitly demonstrated by putting QCD on a space-time lattice, which permits it to be solved numerically in the non-perturbative regime.

States consisting of a single charmed quark and light (u, d, or s) quarks or antiquarks are shown in Figure 6. Finally, the pattern of states containing a single b quark (Figure 7) is very similar to that for singly-charmed states, though not as well fleshed-out. In many cases the splittings between states containing a single b quark is less than that between the corresponding charmed states by roughly a factor of $m_c/m_b \simeq 1/3$ as a result of the smaller chromomagnetic moment of the b quark. Pioneering work in understanding the spectra of such states using QCD was done by De Rújula *et al.* (1975), building on earlier observations on light-quark systems by Zel'dovich and Sakharov (1966), Dalitz (1967), and Lipkin (1973).

3   W bosons

3.1   Fermi theory of weak interactions

The effective four-fermion Hamiltonian for the $V-A$ theory of the weak interactions is

$$\mathcal{H}_W = \frac{G_F}{\sqrt{2}} \left[\overline{\psi}_1 \gamma_\mu (1 - \gamma_5) \psi_2\right] \left[\overline{\psi}_3 \gamma^\mu (1 - \gamma_5) \psi_4\right] = 4\frac{G_F}{\sqrt{2}} \left(\overline{\psi}_{1L} \gamma_\mu \psi_{2L}\right) \left(\overline{\psi}_{3L} \gamma^\mu \psi_{4L}\right), \quad (71)$$

where G_F and ψ_L were defined in Section 1.3. We wish to write instead a Lagrangian for interaction of particles with charged W bosons which reproduces (71) when taken to second order at low momentum transfer. We shall anticipate a result of Section 4 by introducing the W through an SU(2) symmetry, in the form of a gauge coupling.

In the kinetic term in the Lagrangian for fermions,

$$\mathcal{L}_{Kf} = \overline{\psi}(i\,\partial\!\!\!/ - m)\psi = \overline{\psi}_L(i\,\partial\!\!\!/)\psi_L + \overline{\psi}_R(i\,\partial\!\!\!/)\psi_R - m\overline{\psi}\psi, \quad (72)$$

the $\partial\!\!\!/$ term does not mix ψ_L and ψ_R, so in the absence of the $\overline{\psi}\psi$ term one would have the freedom to introduce different covariant derivatives $\mathcal{D}$ acting on left-handed and right-handed fermions. We shall find that the same mechanism which allows us to give masses to the W and Z while keeping the photon massless will permit the generation of fermion masses even though ψ_L and ψ_R will transform differently under our gauge group. We follow the conventions of Peskin and Schroeder (1995, p 700 ff).

We now let the left-handed spinors be doublets of an SU(2), such as

$$\begin{bmatrix} \nu_e \\ e^- \end{bmatrix}_L, \quad \begin{bmatrix} \nu_\mu \\ \mu^- \end{bmatrix}_L, \quad \begin{bmatrix} \nu_\tau \\ \tau^- \end{bmatrix}_L. \quad (73)$$

(We will postpone the question of neutrino mixing until the last section.) The W is introduced via the replacement

$$\partial_\mu \to \mathbf{D}_\mu \equiv \partial_\mu - ig\mathbf{T}^i W^i_\mu, \qquad \mathbf{T}^i \equiv \tau^i/2, \tag{74}$$

where τ^i are the Pauli matrices and W^i_μ are a triplet of massive vector bosons. Here we will be concerned only with the $W^\pm$, defined by $W^\pm_\mu \equiv (W^1_\mu \mp iW^2_\mu)/\sqrt{2}$. The field W^+_μ annihilates a W^+ and creates a W^-, while W^-_μ annihilates a W^- and creates a W^+. Then $W^1_\mu = (W^+_\mu + W^-_\mu)/\sqrt{2}$ and $W^2_\mu = i(W^+_\mu - W^-_\mu)/\sqrt{2}$, so

$$\mathbf{T}^i W^i_\mu = \frac{1}{2} \begin{bmatrix} W^3_\mu & \sqrt{2}W^+_\mu \\ \sqrt{2}W^-_\mu & -W^3_\mu \end{bmatrix}. \tag{75}$$

The interaction arising from (72) for a lepton $l = e, \mu, \tau$ is then

$$\mathcal{L}^{(W^\pm)}_{\text{int}, l} = \frac{g}{\sqrt{2}} \left[\bar{\nu}_{lL} \gamma^\mu W^+_\mu l_L + \bar{l}_L \gamma^\mu W^-_\mu \nu_{lL} \right], \tag{76}$$

where we temporarily neglect the W^3_μ terms. Taking this interaction to second order and replacing the W propagator $(M^2_W - q^2)^{-1}$ by its $q^2 = 0$ value, we find an effective interaction of the form (71), with

$$\frac{G_F}{\sqrt{2}} = \frac{g^2}{8M^2_W}. \tag{77}$$

3.2   Charged-current quark interactions

The left-handed quark doublets may be written

$$\begin{bmatrix} u \\ d' \end{bmatrix}_L, \qquad \begin{bmatrix} c \\ s' \end{bmatrix}_L, \qquad \begin{bmatrix} t \\ b' \end{bmatrix}_L, \tag{78}$$

where d', s', and b' are related to the mass eigenstates d, s, b by a unitary transformation

$$\begin{bmatrix} d' \\ s' \\ b' \end{bmatrix} = V \begin{bmatrix} d \\ s \\ b \end{bmatrix}, \qquad V^\dagger V = 1. \tag{79}$$

The rationale for the unitary matrix V of Kobayashi and Maskawa (1973) will be reviewed in the next section when we discuss the origin of fermion masses in the electroweak theory. The interaction Lagrangian for W's with quarks then is

$$\mathcal{L}^{(W^\pm)}_{\text{int, quarks}} = \frac{g}{\sqrt{2}} (\bar{U}_L \gamma^\mu W^+_\mu V D_L) + \text{h.c.}, \qquad U \equiv \begin{bmatrix} u \\ c \\ t \end{bmatrix}, \qquad D \equiv \begin{bmatrix} d \\ s \\ b \end{bmatrix}. \tag{80}$$

A convenient parameterisation of V (conventionally known as the Cabibbo-Kobayashi-Maskawa matrix, or CKM matrix) suggested by Wolfenstein (1983) is

$$V \equiv \begin{bmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{bmatrix} = \begin{bmatrix} 1 - \frac{\lambda^2}{2} & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \frac{\lambda^2}{2} & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{bmatrix}. \quad (81)$$

Experimentally $\lambda \simeq 0.22$ and $A \simeq 0.85$. Present constraints on the parameters ρ and η are shown in Figure 8. The solid circles denote limits on $|V_{ub}/V_{cb}| = 0.090 \pm 0.025$ from
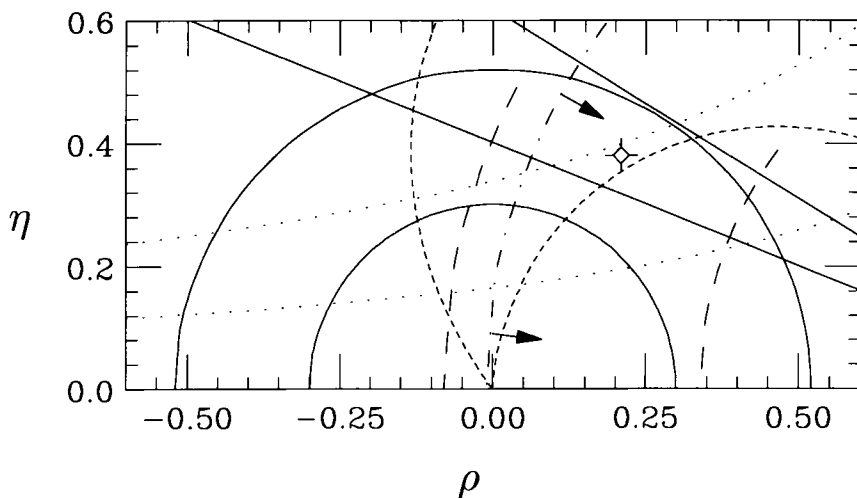
Figure 8. *Constraints on parameters of the Cabibbo-Kobayashi-Maskawa (CKM) matrix. The plotted point at $\rho = 0.21, \eta = 0.38$ lies in the middle of the allowed region. (See text.)*

charmless b decays. The dashed arcs are associated with limits on V_{td} from B^0-$\overline{B}^0$ mixing. The present lower limit on B_s-$\bar{B}_s$ mixing leads to a lower bound on $|V_{ts}/V_{td}|$ and the dot-dashed arc. The dotted hyperbolae arise from limits on CP-violating K^0-$\overline{K}^0$ mixing. The phases in the CKM matrix associated with $\eta \neq 0$ lead to CP violation in neutral kaon decays (Christenson *et al.* 1964) and, as recently discovered, in neutral B meson decays (Aubert *et al.* 2001a, Abe *et al.* 2001). These last results lead to a result shown by the two rays, $\sin(2\beta) = 0.79 \pm 0.10$, where $\beta = \mathrm{Arg}(-V_{cd}V_{cb}^*/V_{td}V_{tb}^*)$. The small dashed lines represent 1σ limits derived by Gronau and Rosner (2002) (see also Luo and Rosner 2001) on the basis of CP asymmetry data of Aubert *et al.* (2001b) for $B^0 \to \pi^+\pi^-$. Our range of parameters (confined by 1σ limits) is $0.10 \leq \rho \leq 0.32$, $0.33 \leq \eta \leq 0.43$. Similar plots are presented in several other lectures at this Summer School (see, e.g., Buchalla 2001, Nir 2001, Schubert 2001, Stone 2001), which may be consulted for further details, and an ongoing analysis of CKM parameters by Höcker *et al.* (2001) is now incorporating several other pieces of data.

3.3   Decays of the τ lepton

The τ lepton (Perl *et al.* 1975) provides a good example of "standard model" charged-current physics. The τ^- decays to a ν_τ and a virtual W^- which can then materialise into any kinematically allowed final state: $e^-\bar{\nu}_e$, $\mu^-\bar{\nu}_\mu$, or three colours of $\bar{u}d'$, where, in accord with (81), $d' \simeq 0.975d + 0.22s$.

Neglecting strong interaction corrections and final fermion masses, the rate for τ decay is expected to be

$$\Gamma(\tau^- \to \text{all}) = 5G_F^2 \frac{m_\tau^5}{192\pi^3} \simeq 2 \times 10^{-3} \text{ eV}, \tag{82}$$

corresponding to a lifetime of $\tau_\tau \simeq 3 \times 10^{-13}$ s as observed. The factor of $5 = 1 + 1 + 3$ corresponds to equal rates into $e^-\bar{\nu}_e$, $\mu^-\bar{\nu}_\mu$, and each of the three colours of $\bar{u}d'$. The branching ratios are predicted to be

$$\mathcal{B}(\tau^- \to \nu_\tau e^- \bar{\nu}_e) = \mathcal{B}(\tau^- \to \nu_\tau \mu^- \bar{\nu}_\mu) = \frac{1}{3}\mathcal{B}(\tau^- \to \nu_\tau \bar{u}d') = 20\%. \tag{83}$$

Measured values for the purely leptonic branching ratios are slightly under 18%, as a result of the enhancement of the hadronic channels by a QCD correction whose leading-order behavior is $1 + \alpha_S/\pi$, the same as for R in e^+e^- annihilation. The τ decay is thus further evidence for the existence of three colours of quarks.

3.4   W decays

We shall calculate the rate for the process $W \to f\bar{f}'$ and then generalise the result to obtain the total W decay rate. The interaction Lagrangian (76) implies that the covariant matrix element for the process $W(k) \to f(p)\bar{f}'(p')$ is

$$\mathcal{M}^{(\lambda)} = \frac{g}{2\sqrt{2}}\bar{u}_f(p)\gamma^\mu(1 - \gamma_5)v_{f'}(p')\epsilon_\mu^{(\lambda)}(k). \tag{84}$$

Here λ describes the polarisation state of the W. The partial width is

$$\Gamma(W^- \to f\bar{f}') = \frac{1}{2M_W}\frac{1}{3}\sum_{\text{pols}}|\mathcal{M}^{(\lambda)}|^2\frac{p^*}{4\pi M_W}, \tag{85}$$

where $(2M_W)^{-1}$ is the initial-state normalisation, $1/3$ corresponds to an average of W polarisations, the sum is over both W and lepton polarisations, and p^* is the final centre-of-mass (c.m.) 3-momentum. We use the identity

$$\sum_\lambda \epsilon_\mu^{(\lambda)}(k)\epsilon_\nu^{(\lambda)*}(k) = -g_{\mu\nu} + \frac{k_\mu k_\nu}{M_W^2} \tag{86}$$

for sums over W polarisation states. The result is that

$$\sum_{\text{pols}}|\mathcal{M}^{(\lambda)}|^2 = g^2\left[M_W^2 - \frac{1}{2}(m^2 + m'^2) - \frac{(m^2 - m'^2)^2}{2M_W^2}\right] \tag{87}$$

for any process $W \to f\bar{f}'$, where m is the mass of f and m' is the mass of f'. Recalling the relation between G_F and g^2, this may be written in the simpler form

$$\Gamma(W \to f\bar{f}') = \frac{G_F}{\sqrt{2}} \frac{M_W^3}{6\pi} \Phi_{ff'} , \qquad \Phi_{ff'} \equiv \frac{2p^*}{M_W} \frac{p^{*2} + 3EE'}{M_W^2} . \qquad (88)$$

Here $E = (p^{*2} + m^2)^{1/2}$ and $E' = (p^{*2} + m'^2)^{1/2}$ are the c.m. energies of f and f'. The factor $\Phi_{ff'}$ reduces to 1 as $m, m' \to 0$.

The present experimental average for the W mass (Kim 2001) is $M_W = 80.451 \pm 0.033\mathrm{GeV}$. Using this value, we predict $\Gamma(W \to e^-\bar{\nu}_e) = 227.8 \pm 2.3\mathrm{MeV}$. The widths to various channels are expected to be in the ratios

$$e^-\bar{\nu}_e : \mu^-\bar{\nu}_\mu : \tau^-\bar{\nu}_\tau : \bar{u}d' : \bar{c}s' = 1 : 1 : 1 : 3\left[1 + \frac{\alpha_S(M_W^2)}{\pi}\right] : 3\left[1 + \frac{\alpha_S(M_W^2)}{\pi}\right] , \quad (89)$$

so $\alpha_S(M_W^2) = 0.120 \pm 0.002$ leads to the prediction $\Gamma_{\mathrm{tot}}(W) = 2.10 \pm 0.02\mathrm{GeV}$. This is to be compared with a value (Drees 2001) obtained at LEP II by direct reconstruction of W's: $\Gamma_{\mathrm{tot}}(W) = 2.150 \pm 0.091\mathrm{GeV}$. Higher-order electroweak corrections, to be discussed in Section 5, are not expected to play a major role here. This agreement means, among other things, that we are not missing a significant channel to which the charged weak current can couple below the mass of the W.

3.5   W pair production

We shall outline a calculation (Quigg 1983) which indicates that the weak interactions cannot possibly be complete if described only by charged-current interactions. We consider the process $\nu_e(q) + \bar{\nu}_e(q') \to W^+(k) + W^-(k')$ due to exchange of an electron e^- with momentum p. The matrix element is

$$\mathcal{M}^{(\lambda,\lambda')} = \frac{G_F M_W^2}{\sqrt{2}} \bar{v}(q') \, \slashed{\epsilon}^{(\lambda')}(k')(1 - \gamma_5)\frac{\slashed{p}}{p^2} \, \slashed{\epsilon}^{(\lambda)}(k)u(q) . \qquad (90)$$

For a longitudinally polarised W^+, this matrix element grows in an unacceptable fashion for high energy. In fact, an inelastic amplitude for any given partial wave has to be bounded, whereas $\mathcal{M}^{(\lambda,\lambda')}$ will not be.

The polarisation vector for a longitudinal W^+ traveling along the z axis is

$$\epsilon_\nu^{(\lambda)}(k) = (|\mathbf{k}|, 0, 0, M_W) \simeq k_\nu/M_W , \qquad (91)$$

with a correction which vanishes as $|\mathbf{k}| \to \infty$. Replacing $\epsilon_\nu^{(\lambda)}(k)$ by k_ν/M_W, using $\slashed{k} = \slashed{q} - \slashed{p}$ and $\slashed{q}u(q) = 0$, we find

$$\mathcal{M}^{(\lambda,\lambda')} \simeq -\sqrt{2}G_F M_W \bar{v}(q') \, \slashed{\epsilon}^{(\lambda')}(k')u(q) , \qquad (92)$$

$$\sum_{\text{lepton pol.}} |\mathcal{M}^{(\lambda,\lambda')}|^2 = 2G_F^2 M_W^2 [8q' \cdot \epsilon^{(\lambda')} q \cdot \epsilon^{(\lambda')} - 4q \cdot q' \epsilon^{(\lambda')} \cdot \epsilon^{(\lambda')}] . \qquad (93)$$

This quantity contributes only to the lowest two partial waves, and grows without bound as the energy increases. Such behavior is not only unacceptable on general grounds

because of the boundedness of inelastic amplitudes, but it leads to divergences in higher-order perturbation contributions, e.g., to elastic $\bar{\nu}\nu$ scattering.

Two possible contenders for a solution of the problem in the early 1970s were (1) a neutral gauge boson Z^0 coupling to $\nu\bar{\nu}$ and W^+W^- (Glashow 1961, Weinberg 1967, Salam 1968), or (2) a left-handed heavy lepton E^+ (Georgi and Glashow 1972a) coupling to $\nu_e W^+$. Either can reduce the unacceptable high-energy behavior to a constant. The Z^0 alternative seems to be the one selected in nature. In what follows we will retrace the steps of the standard electroweak theory, which led to the prediction of the W and Z and all the phenomena associated with them.

4   Electroweak unification

4.1   Guidelines for symmetry

We now return to the question of what to do with the "neutral W" (the particle we called W^3 in the previous section), a puzzle since the time of Oskar Klein in the 1930s. The time component of the charged weak current

$$J_\mu^{(+)} = \bar{N}_L \gamma_\mu L_L + \bar{U}_L \gamma_\mu V D_L \,, \tag{94}$$

where N_L and L_L are neutral and charged lepton column vectors defined in analogy with U_L and D_L, may be used to define operators

$$Q^{(+)} \equiv \int d^3x \, J_0^{(+)} \,, \qquad Q^{(-)} \equiv Q^{(+)\dagger} \tag{95}$$

which are charge-raising and charge-lowering members of an SU(2) triplet. If we define $Q_3 \equiv (1/2)[Q^{(+)}, Q^{(-)}]$, the algebra closes: $[Q_3, Q^{(\pm)}] = \pm Q^{(\pm)}$. This serves to normalise the weak currents, as mentioned in the Introduction.

The form (94) (with unitary V) guarantees that the corresponding neutral current will be

$$J_\mu^{(3)} = \frac{1}{2}\left[\bar{N}_L \gamma_\mu N_L - \bar{L}_L \gamma_\mu L_L + \bar{U}_L \gamma_\mu U_L - \bar{D}_L \gamma_\mu D_L \right] \,, \tag{96}$$

which is *diagonal* in neutral currents. This can only succeed, of course, if there are equal numbers of charged and neutral leptons, and equal numbers of quarks of charge 2/3 and and quarks of charge $-1/3$.

It would have been possible to define an SU(2) algebra making use only of a doublet (Gell-Mann and Lévy 1960)

$$\begin{bmatrix} u \\ d' \end{bmatrix}_L = \begin{bmatrix} u \\ V_{ud}d + V_{us}s \end{bmatrix}_L \tag{97}$$

which was the basis of the Cabibbo (1963) theory of the charge-changing weak interactions of strange and non-strange particles. If one takes $V_{ud} = \cos\theta_C$, $V_{us} = \sin\theta_C$, as is assumed in the Cabibbo theory, the u, d, s contribution to the neutral current $J_\mu^{(3)}$ is

$$\begin{aligned} J_\mu^{(3)}\Big|_{u,d,s} = \ & \frac{1}{2}\Big[\bar{u}_L \gamma_\mu u_L - \cos^2\theta_C \bar{d}_L \gamma_\mu d_L - \sin^2\theta_C \bar{s}_L \gamma_\mu s_L \\ & - \sin\theta_C \cos\theta_C (\bar{d}_L \gamma_\mu s_L + \bar{s}_L \gamma_\mu d_L) \Big] \,. \end{aligned} \tag{98}$$

This expression contains *strangeness-changing neutral currents*, leading to the expectation of many processes like $K^+ \to \pi^+ \nu \bar{\nu}$, $K_L^0 \to \mu^+ \mu^-$, ..., at levels far above those observed. It was the desire to banish strangeness-changing neutral currents that led Glashow *et al.* (1970) to introduce the charmed quark c (proposed earlier by several authors on the basis of a quark-lepton analogy) and the doublet

$$\begin{bmatrix} c \\ s' \end{bmatrix}_L = \begin{bmatrix} c \\ V_{cd}d + V_{cs}s \end{bmatrix}_L . \tag{99}$$

In this four-quark theory, one assumes the corresponding matrix V is unitary. By suitable phase changes of the quarks, all elements can be made real, making V an orthogonal matrix with $V_{ud} = V_{cs} = \cos\theta_C$, $V_{us} = -V_{cd} = \sin\theta_C$. Instead of (98) one then has

$$J_\mu^{(3)}|_{u,d,s,c} = \frac{1}{2}[\bar{u}_L \gamma_\mu u_L + \bar{c}_L \gamma_\mu c_L - \bar{d}_L \gamma_\mu d_L - \bar{s}_L \gamma_\mu s_L], \tag{100}$$

which contains no flavour-changing neutral currents.

The charmed quark also plays a key role in higher-order charged-current interactions. Let us consider $K^0 - \overline{K}^0$ mixing. The CP-conserving limit in which the eigenstates are K_1 (even CP) and K_2 (odd CP) can be illustrated using a degenerate two-state system such as the first excitations of a drum head. There is no way to distinguish between the basis states illustrated in Figure 9(a), in which the nodal lines are at angles of $\pm 45°$ with respect to the horizontal, and those in Figure 9(b), in which they are horizontal and vertical.
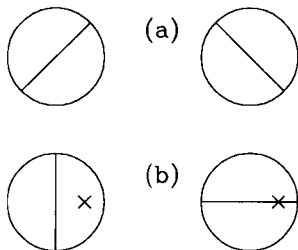
Figure 9. *Basis states for first excitations of a drum head. (a) Nodal lines at $\pm 45°$ with respect to horizontal; (b) horizontal and vertical nodal lines.*

If a fly lands on the drum-head at the point marked "×", the basis (b) corresponds to eigenstates. One of the modes couples to the fly; the other doesn't. The basis in (a) is like that of $(K^0, \overline{K}^0)$, while that in (b) is like that of (K_1, K_2). Neutral kaons are produced as in (a), while they decay as in (b), with the fly analogous to the $\pi\pi$ state. The short-lived state (K_1, in this CP-conserving approximation) has a lifetime of 0.089ns, while the long-lived state ($\simeq K_2$) lives ~ 600 times as long, for 52ns. Classical illustration of CP-violating mixing is more subtle but can be achieved as well, for instance in a rotating reference frame (Rosner and Slezak 2001, Kostelecký and Roberts 2001).

The shared $\pi\pi$ intermediate state and other low-energy states like π^0, η, and η' are chiefly responsible for CP-conserving K^0–$\overline{K}^0$ mixing. However, one must ensure that large *short-distance* contributions do not arise from diagrams such as those illustrated in Figure 10.
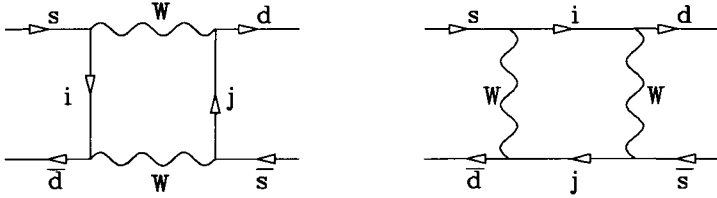
Figure 10. *Higher-order weak contributions to K^0–$\overline{K}^0$ mixing due to loops with internal u, c, t quarks.*

If the only charge 2/3 quark contributing to this process were the u quark, one would expect a contribution to Δm_K of order

$$\Delta m_K|_u \sim g^4 f_K^2 m_K \sin^2\theta_C \cos^2\theta_C / 16\pi^2 M_W^2 \sim G_F f_K^2 m_K (g^2/16\pi^2)\,, \qquad (101)$$

where f_K is the amplitude for $d\bar{s}$ to be found in a K^0, and the factor of $16\pi^2$ is characteristic of loop diagrams. This is far too large, since $\Delta m_K \sim \Gamma_{K_S} \sim G_F^2 f_K^2 m_K^3$. However, the introduction of the charmed quark, coupling to $-d\sin\theta_C + s\cos\theta_C$, cancels the leading contribution, leading to an additional factor of $[(m_c^2 - m_u^2)/M_W^2]\ln(M_W^2/m_c^2)$ in the above expression. Using such arguments Glashow *et al.* (1970) and Gaillard and Lee (1974) estimated the mass of the charmed quark to be less than several GeV. (Indeed, early candidates for charmed particles had been seen by Niu, Mikumo, and Maeda 1971.) The discovery of the J/ψ (Aubert *et al.* 1974, Augustin *et al.* 1974) confirmed this prediction; charmed hadrons produced in neutrino interactions (Cazzoli *et al.* 1975) and in e^+e^- annihilation (Goldhaber *et al.* 1976, Peruzzi *et al.* 1976) followed soon after.

An early motivation for charm relied on an analogy between quarks and leptons. Hara (1964), Maki and Ohnuki (1964), and Bjorken and Glashow (1964) inferred the existence of a charmed quark coupling mainly to the strange quark from the existence of the $\mu - \nu_\mu$ doublet:

$$\begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix} : \text{leptons} \Rightarrow \begin{pmatrix} c \\ s \end{pmatrix} : \text{quarks}\,. \qquad (102)$$

Further motivation for the quark-lepton analogy was noted by Bouchiat *et al.* (1972), Georgi and Glashow (1972b), and Gross and Jackiw (1972). In a gauge theory of the electroweak interactions, triangle anomalies associated with graphs of the type shown in Figure 11 have to be avoided. This cancellation requires the fermions f in the theory to contribute a total of zero to the sum over f of $Q_f^2 I_{3L}^f$. Such a cancellation can be achieved by requiring quarks and leptons to occur in complete families so that the terms

$$\text{Leptons} \quad : \quad (0)^2 \left(\frac{1}{2}\right) + (-1)^2 \left(-\frac{1}{2}\right) = -\frac{1}{2} \qquad (103)$$

$$\text{Quarks} \quad : \quad 3\left[\left(\frac{2}{3}\right)^2 \left(\frac{1}{2}\right) + \left(-\frac{1}{3}\right)^2 \left(-\frac{1}{2}\right)\right] = \frac{1}{2} \qquad (104)$$
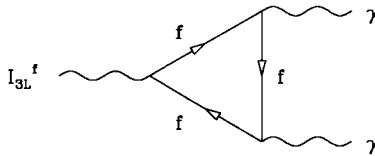
Figure 11. *Example of triangle diagram for which leading behaviour must cancel in a renormalisable electroweak theory.*

sum to zero for each family.

We are then left with a flavour-preserving neutral current $J_\mu^{(3)}$, given by (100), whose interpretation must still be given. It cannot correspond to the photon, since the photon couples to both left-handed and right-handed fermions. At the same time, the photon is somehow involved in the weak interactions associated with W exchange. In particular, the $W^\pm$ themselves are charged, so any theory in which electromagnetic current is conserved must involve a $\gamma W^+ W^-$ coupling. Moreover, the charge is sensitive to the third component of the SU(2) algebra we have just introduced. We shall refer to this as SU(2)$_L$, recognising that only left-handed fermions ψ_L transform non-trivially under it. Then we can define a *weak hypercharge* Y in terms of the difference between the electric charge Q and the third component I_{3L} of SU(2)$_L$ (*weak isospin*):

$$Q = I_{3L} + \frac{Y}{2}. \qquad (105)$$

Values of Y for quarks and leptons are summarised in Table 3.

Particle(s)	Q	I_{3L}	Y
ν_{eL}	0	$1/2$	-1
e_L^-	-1	$-1/2$	-1
u_L	$2/3$	$1/2$	$1/3$
d_L	$-1/3$	$-1/2$	$1/3$
e_R^-	-1	0	-2
u_R	$2/3$	0	$4/3$
d_R	$-1/3$	0	$-2/3$

Table 3. *Values of charge, I_{3L}, and weak hypercharge Y for quarks and leptons.*

If you find these weak hypercharge assignments mysterious, you are not alone. They follow naturally in unified theories (*grand unified theories*) of the electroweak and strong interactions. A "secret formula" for Y, which may have deeper significance (Pati and Salam 1973), is $Y = 2I_{3R} + (B - L)$, where I_{3R} is the third component of "right-handed" isospin, B is baryon number (1/3 for quarks), and L is lepton number (1 for leptons such as e^- and ν_e). The orthogonal component of I_{3R} and $B - L$ may correspond to a higher-mass, as-yet-unseen vector boson, an example of what is called a Z'. The search

for Z' bosons with various properties is an ongoing topic of interest; current limits are quoted by the Particle Data Group (2000).

The gauge theory of charged and neutral W's thus must involve the photon in some way. It will then be necessary, in order to respect the formula (105), to introduce an additional U(1) symmetry associated with weak hypercharge. The combined electroweak gauge group will have the form $SU(2)_L \otimes U(1)_Y$.

4.2   Symmetry breaking

Any unified theory of the weak and electromagnetic interactions must be broken, since the photon is massless while the W bosons (at least) are not. An explicit mass term in a gauge theory of the form $m^2 A^i_\mu A^{\mu i}$ violates gauge invariance. It is not invariant under the replacement (26). Another means must be found to introduce a mass. The symmetry must be broken in such a way as to preserve gauge invariance.

A further manifestation of symmetry breaking is the presence of fermion mass terms. Any product $\overline{\psi}\psi$ may be written as

$$\overline{\psi}\psi = (\overline{\psi}_L + \overline{\psi}_R)(\psi_L + \psi_R) = \overline{\psi}_L \psi_R + \overline{\psi}_R \psi_L \,, \tag{106}$$

using the fact that $\overline{\psi}_L = \overline{\psi}(1 + \gamma_5)/2$ and $\overline{\psi}_R = \overline{\psi}(1 - \gamma_5)/2$. Since ψ_L transforms as an $SU(2)_L$ doublet but ψ_R as an $SU(2)_L$ singlet, a mass term proportional to $\overline{\psi}\psi$ transforms as an overall $SU(2)_L$ doublet. Moreover, the weak hypercharges of left-handed fermions and their right-handed counterparts are different. Hence one cannot even have explicit *fermion* mass terms in the Lagrangian and hope to preserve local gauge invariance.

One way to generate a fermion mass without explicitly violating gauge invariance is to assume the existence of a complex scalar $SU(2)_L$ doublet ϕ coupled to fermions via a Yukawa interaction:

$$\mathcal{L}_Y = -g_Y(\overline{\psi}_L \phi \psi_R + \text{h.c.}), \qquad \phi \equiv \begin{bmatrix} \phi^+ \\ \phi^0 \end{bmatrix}. \tag{107}$$

Thus, for example, with $\overline{\psi}_L = (\bar{\nu}_e, e)_L$ and $\psi_R = e_R$, we have

$$\mathcal{L}_{Y,e} = -g_{Ye}(\bar{\nu}_{eL}\phi^+ e_R + \bar{e}_L \phi^0 e_R + \text{h.c.}). \tag{108}$$

If ϕ^0 acquires a vacuum expectation value, $\langle \phi^0 \rangle \neq 0$, this quantity will automatically break $SU(2)_L$ and $U(1)_Y$, and will give rise to a non-zero electron mass. A neutrino mass is not generated, simply because no right-handed neutrino has been assumed to exist. (We shall see in the last section how to generate the tiny neutrino masses that appear to be present in nature.) The gauge symmetry is not broken in the Lagrangian, but only in the solution. This is similar to the way in which rotational invariance is broken in a ferromagnet, where the fundamental interactions are rotationally invariant but the ground-state solution has a preferred direction along which the spins are aligned.

The d quark masses are generated by similar couplings involving $\overline{\psi}_L = (\bar{u}, \bar{d})_L$ and $\psi_R = d_R$, so that

$$\mathcal{L}_{Y,d} = -g_{Yd}(\bar{u}_L \phi^+ d_R + \bar{d}_L \phi^0 d_R + \text{h.c.}). \tag{109}$$

To generate u quark masses one must either use the multiplet

$$\tilde{\phi} \equiv \begin{bmatrix} \bar{\phi}^0 \\ -\phi^- \end{bmatrix} = i\tau^2 \phi^*, \tag{110}$$

which also transforms as an SU(2) doublet, or a separate doublet of scalar fields

$$\phi' = \begin{bmatrix} \phi'^0 \\ \phi'^- \end{bmatrix}. \tag{111}$$

With $\overline{\psi}_L = (\bar{u}, \bar{d})_L$ and $\psi_R = u_R$, we then find

$$\mathcal{L}_{Y,u} = -g_{Yu}(\bar{u}_L \bar{\phi}^0 u_L - \bar{d}_L \phi^- u_L + \text{h.c.}) \tag{112}$$

if we make use of $\tilde{\phi}$, or

$$\mathcal{L}_{Y,u} = -g_{Yu}(\bar{u}_L \phi'^0 u_L + \bar{d}_L \phi'^- u_L + \text{h.c.}) \tag{113}$$

if we use ϕ'. For present purposes we shall assume the existence of a single complex doublet, though many theories (notably, some grand unified theories or supersymmetry) require more than one.

4.3 Scalar fields and the Higgs mechanism

Suppose a complex scalar field of the form (107) is described by a Lagrangian

$$\mathcal{L}_\phi = (\partial_\mu \phi)^\dagger(\partial^\mu \phi) - \frac{\lambda}{4}(\phi^\dagger \phi)^2 + \frac{\mu^2}{2}\phi^\dagger \phi. \tag{114}$$

Note the "wrong" sign of the mass term. This Lagrangian is invariant under the symmetry $SU(2)_L \otimes U(1)_Y$. The field ϕ will acquire a constant vacuum expectation value which we calculate by asking for the stationary value of $\mathcal{L}_\phi$:

$$\frac{\partial \mathcal{L}_\phi}{\partial(\phi^\dagger \phi)} = 0 \Rightarrow \langle \phi^\dagger \phi \rangle = \frac{\mu^2}{\lambda}. \tag{115}$$

We still have not specified which component of ϕ acquires the vacuum expectation value. At this point only $\phi^\dagger \phi = |\phi^+|^2 + |\phi^0|^2$ is fixed, and (Re ϕ^+, Im ϕ^+, Re ϕ^0, Im ϕ^0) can range over the surface of a four-dimensional sphere. The Lagrangian (114) is, in fact, invariant under rotations of this four-dimensional sphere, a group SO(4) isomorphic to SU(2) $\otimes$ U(1). A lower-dimensional analogue of this surface would be the bottom of a wine bottle along which a marble rolls freely in an orbit a fixed distance from the centre.

Let us *define* the vacuum expectation value of ϕ to be a real parameter in the ϕ^0 direction:

$$\langle \phi \rangle = \begin{bmatrix} 0 \\ v/\sqrt{2} \end{bmatrix}. \tag{116}$$

The factor of $1/\sqrt{2}$ is introduced for later convenience. We then find, from the discussion in the previous section, that Yukawa couplings of ϕ to fermions ψ_i generate mass terms

$m_i = g_{Y_i} v/\sqrt{2}$. We must now see what such vacuum expectation values do to gauge boson masses. (For numerous illustrations of this phenomenon in simple field-theoretical models see Abers and Lee 1973, Quigg 1983, and Peskin and Schroeder 1995.)

In order to introduce gauge interactions with the scalar field ϕ, one must replace ∂_μ by $\mathbf{D}_\mu$ in the kinetic term of the Lagrangian (114). Here

$$\mathbf{D}_\mu = \partial_\mu - ig\frac{\tau^i W_\mu^i}{2} - ig'\frac{Y}{2}B_\mu\,, \tag{117}$$

where the $U(1)_Y$ interaction is characterised by a coupling constant g' and a gauge field B_μ, and we have written g for the $SU(2)$ coupling discussed earlier. It will be convenient to write ϕ in terms of four independent real fields $(\xi^i,\ \eta)$ in a slightly different form:

$$\phi = \exp\left(\frac{i\xi^i\tau^i}{2v}\right)\left[\begin{array}{c} 0 \\ \dfrac{v+\eta}{\sqrt{2}} \end{array}\right]. \tag{118}$$

We then perform an $SU(2)_L$ gauge transformation to remove the ξ dependence of ϕ, and rewrite it as

$$\phi = \left[\begin{array}{c} 0 \\ \dfrac{v+\eta}{\sqrt{2}} \end{array}\right]. \tag{119}$$

The fermion and gauge fields are transformed accordingly; we rewrite the Lagrangian for them in the new gauge. The resulting kinetic term for the scalar fields, taking account that $Y = 1$ for the Higgs field (107), is

$$\begin{aligned}
\mathcal{L}_{K,\phi} &= (\mathbf{D}_\mu\phi)^\dagger(\mathbf{D}^\mu\phi) \\
&= \left|\left\{\partial_\mu - \frac{ig}{2}\left[\begin{array}{cc} W_\mu^3 & W_\mu^1 - iW_\mu^2 \\ W_\mu^1 + iW_\mu^2 & -W_\mu^3 \end{array}\right] - \frac{ig'}{2}B_\mu\right\}\left[\begin{array}{c} 0 \\ \dfrac{v+\eta}{\sqrt{2}} \end{array}\right]\right|^2.
\end{aligned} \tag{120}$$

This term contains several contributions.

1. There is a kinetic term $\frac{1}{2}(\partial_\mu\eta)(\partial^\mu\eta)$ for the scalar field η.

2. A term $v\partial_\mu\eta$ is a total divergence and can be neglected.

3. There are $WW\eta$, $BB\eta$, $WW\eta^2$, and $BB\eta^2$ interactions.

4. The v^2 term leads to a mass term for the Yang-Mills fields:

$$\mathcal{L}_{m,YM} = \frac{v^2}{8}\left\{g^2[(W^1)^2 + (W^2)^2] + (gW^3 - g'B)^2\right\}. \tag{121}$$

The spontaneous breaking of the $SU(2)\otimes U(1)$ symmetry thus has led to the appearance of a mass term for the gauge fields. This is an example of the *Higgs mechanism* (Higgs 1964). An unavoidable consequence is the appearance of the scalar field η, the *Higgs field*. We shall discuss it further in Section 5.

The masses of the charged W bosons may be identified by comparing (121) and (75):

$$(gv)^2/8 = M_W^2/2\,, \qquad \text{or} \qquad M_W = gv/2\,. \tag{122}$$

Since the Fermi constant is related to g/M_W, one finds

$$\frac{G_F}{\sqrt{2}} = \frac{g^2}{8M_W^2} = \frac{1}{2v^2}, \qquad \text{or} \qquad v = 2^{-1/4}G_F^{-1/2} = 246 \text{ GeV}. \tag{123}$$

The combination $gW_\mu^3 - g'B_\mu$ also acquires a mass. We must normalise this combination suitably so that it contributes properly in the kinetic term for the Yang-Mills fields:

$$\mathcal{L}_{K,YM} = -\frac{1}{4}W_{\mu\nu}^i W^{\mu\nu i} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu}, \tag{124}$$

where

$$W_{\mu\nu}^i \equiv \partial_\mu W_\nu^i - \partial_\nu W_\mu^i + g\epsilon_{ijk}W_\mu^j W_\nu^k, \qquad B_{\mu\nu} \equiv \partial_\mu B_\nu - \partial_\nu B_\mu. \tag{125}$$

Setting

$$\cos\theta = \frac{g}{(g^2 + g'^2)^{1/2}}, \qquad \sin\theta = \frac{g'}{(g^2 + g'^2)^{1/2}}, \tag{126}$$

we may write the normalised combination $\sim gW_\mu^3 - g'B_\mu$ which acquires a mass as

$$Z_\mu \equiv W_\mu^3 \cos\theta - B_\mu \sin\theta. \tag{127}$$

The orthogonal combination does not acquire a mass. It may then be identified as the photon:

$$A_\mu = B_\mu \cos\theta + W_\mu^3 \sin\theta. \tag{128}$$

The mass of the Z is given by

$$\frac{(g^2 + g'^2)v^2}{8} = \frac{M_Z^2}{2}, \qquad \text{or} \qquad M_Z = \frac{M_W(g^2 + g'^2)^{1/2}}{g} = \frac{M_W}{\cos\theta}, \tag{129}$$

using (126) in the last relation. The W's and Z's have acquired masses, but they are not equal unless g' were to vanish. We shall see in the next subsection that both g and g' are nonzero, so one expects the Z to be heavier than the W.

It is interesting to stop for a moment to consider what has taken place. We started with four scalar fields ϕ^+, ϕ^-, ϕ^0, and $\bar{\phi}^0$. Three of them [ϕ^+, ϕ^-, and the combination $(\phi^0 - \bar{\phi}^0)/i\sqrt{2}$] could be absorbed in the gauge transformation in passing from (118) to (119), which made sense only as long as $(\phi^0 + \bar{\phi})/\sqrt{2}$ had a vacuum expectation value v. The net result was the generation of mass for three gauge bosons W^+, W^-, and Z.

If we had not transformed away the three components ξ^i of ϕ in (118), the term $\mathcal{L}_{K,\phi}$ in the presence of gauge fields would have contained contributions $W_\mu \partial^\mu \phi$ which mixed gauge fields and derivatives of ϕ. These can be expressed as

$$W_\mu \partial^\mu \phi = \partial^\mu (W_\mu \phi) - (\partial^\mu W_\mu)\phi \tag{130}$$

and the total divergence (the first term) discarded. One thus sees that such terms *mix longitudinal components of gauge fields* (proportional to $\partial^\mu W_\mu$) *with scalar fields*. It is necessary to redefine the gauge fields by means of a gauge transformation to get rid of such mixing terms. It is just this transformation that was anticipated in passing from (118) to (119).

The three "unphysical" scalar fields provide the necessary longitudinal degrees of freedom in order to convert the massless $W^\pm$ and Z to massive fields. Each massless field possesses only two polarisation states ($J_z = \pm J$), while a massive vector field has three ($J_z = 0$ as well). Such counting rules are extremely useful when more than one Higgs field is present, to keep track of how many scalar fields survive being "eaten" by gauge fields.

4.4   Interactions in the SU(2) ⊗ U(1) theory.

By introducing gauge boson masses via the Higgs mechanism, and letting the simplest non-trivial representation of scalar fields acquire a vacuum expectation value v, we have related the Fermi coupling constant to v, and the gauge boson masses to gv or $(g^2+g'^2)^{1/2}v$. We still have two arbitrary couplings g and g' in the theory, however. We shall show how to relate the electromagnetic coupling to them, and how to measure them separately.

The interaction of the fermions with the gauge fields is described by the kinetic term $\mathcal{L}_{K,\psi} = \bar{\psi}\,\slashed{D}\psi$. Here, as usual,

$$\slashed{D} = \slashed{\partial} - ig\frac{\tau^i\,\slashed{W}^i}{2} - ig'\frac{Y}{2}\,\slashed{B}\,. \tag{131}$$

The charged-W interactions have already been discussed. They are described by the terms (76) for leptons and (80) for quarks. The interactions of W^3 and B may be re-expressed in terms of A and Z via the inverse of (127) and (128):

$$W^3_\mu = Z_\mu \cos\theta + A_\mu \sin\theta\,, \qquad B_\mu = -Z_\mu \sin\theta + A_\mu \cos\theta\,. \tag{132}$$

Then the covariant derivative for neutral gauge bosons is

$$\slashed{D}|_{\text{neutral}} = \slashed{\partial} - igI_{3L}(\slashed{Z}\cos\theta + \slashed{A}\sin\theta) - ig'(Q - I_{3L})(-\slashed{Z}\sin\theta + \slashed{A}\cos\theta)\,. \tag{133}$$

Here we have substituted $Y/2 = (Q - I_{3L})$. We identify the electromagnetic contribution to the right-hand side of (133) with the familiar one $-ieQ\,\slashed{A}$, so that

$$e = g'\cos\theta = g\sin\theta\,. \tag{134}$$

The second equality, stemming from the demand that $I_{3L}\,\slashed{A}$ terms cancel one another in (133), is automatically satisfied as a result of the definition (126). Combining (126) and (134), we find

$$e = \frac{gg'}{\sqrt{g^2 + g'^2}} \qquad \text{or} \qquad \frac{1}{e^2} = \frac{1}{g^2} + \frac{1}{g'^2}\,, \tag{135}$$

the result advertised in the Introduction.

The interaction of the Z with fermions may be determined from Eq. (133) with the help of (126), noting that $g\cos\theta + g'\sin\theta = (g^2 + g'^2)^{1/2}$ and $g'\sin\theta = (g^2 + g'^2)^{1/2}\sin^2\theta$. We find

$$\slashed{D}|_{\text{neutral}} = \slashed{\partial} - ieQ\,\slashed{A} - i(g^2 + g'^2)^{1/2}(I_{3L} - Q\sin^2\theta)\,\slashed{Z}\,. \tag{136}$$

Knowledge of the weak mixing angle θ will allow us to predict the W and Z masses. Using $G_F/\sqrt{2} = g^2/8M_W^2$ and $g\sin\theta = e$, we can write

$$M_W = \left[\frac{\pi\alpha}{\sqrt{2}G_F}\right]^{1/2}\frac{1}{\sin\theta} = \frac{37.3\ \text{GeV}}{\sin\theta} \tag{137}$$

if we were to use $\alpha^{-1} = 137.036$. However, we shall see in the next section that it is more appropriate to use a value of $\alpha^{-1} \simeq 129$ at momentum transfers characteristic of the W mass. With this and other electroweak radiative corrections, the correct estimate is raised to $M_W \simeq 38.6\ \text{GeV}/\sin\theta$, leading to the successful predictions (7). The Z mass is expressed in terms of the W mass by $M_Z = M_W/\cos\theta$.

4.5 Neutral current processes

The interactions of Z's with matter,

$$\mathcal{L}_{\text{int},Z} = (g^2 + g'^2)^{1/2}\overline{\psi}(I_{3L} - Q\sin^2\theta)\not{Z}\psi\,, \tag{138}$$

may be taken to second order in perturbation theory, leading to an effective four-fermion theory for momentum transfers much smaller than the Z mass. In analogy with the relation between the W boson interaction terms (76) and (80) and the four-fermion charged-current interaction (71), we may write

$$\mathcal{H}_W^{NC} = 4G_F\sqrt{2}[\overline{\psi}_1(I_{3L} - Q\sin^2\theta)\gamma^\mu\psi_2][\overline{\psi}_3(I_{3L} - Q\sin^2\theta)\gamma_\mu\psi_4]\,, \tag{139}$$

where we have used the identity $(g^2 + g'^2)/8M_Z^2 = G_F/\sqrt{2}$ following from relations in the previous subsection.

Many processes are sensitive to the neutral-current interaction (139), but no evidence for this interaction had been demonstrated until the discovery in 1973 of neutral-current interactions on hadronic targets of deeply inelastically scattered neutrinos (Hasert *et al.* 1973; Benvenuti *et al.* 1974). For many years these processes provided the most sensitive measurement of neutral-current parameters. Other crucial experiments (see, e.g., reviews by Amaldi *et al.* 1987 and Langacker *et al.* 1992) included polarised electron or muon scattering on nucleons, asymmetries and total cross sections in $e^+e^- \to \mu^+\mu^-$ or $\tau^+\tau^-$, parity violation in atomic transitions, neutrino-electron scattering, coherent π^0 production on nuclei by neutrinos, and detailed measurements of W and Z properties. Let us take as an example the scattering of leptons on quarks to see how they provide a value of $\sin^2\theta$. In the next subsection we shall turn to the properties of the Z bosons, which are now the source of the most precise information.

One measures quantities

$$R_\nu \equiv \frac{\sigma(\nu A \to \nu + \ldots)}{\sigma(\nu A \to \mu^- + \ldots)}\,, \qquad R_{\bar\nu} \equiv \frac{\sigma(\bar\nu A \to \bar\nu + \ldots)}{\sigma(\bar\nu A \to \mu^+ + \ldots)}\,. \tag{140}$$

These ratios may be calculated in terms of the weak Hamiltonians (71) and (139). It is helpful to note that for states of the same helicity (L or R, standing for left-handed or right-handed) scattering on one another, the differential cross section is a constant:

$$\frac{d\sigma}{d\Omega}(RR \to RR) = \frac{d\sigma}{d\Omega}(LL \to LL) = \frac{\sigma_0}{4\pi}\,, \tag{141}$$

where σ_0 is some reference cross section, while for states of opposite helicity,

$$\frac{d\sigma}{d\Omega}(RL \to RL) = \frac{d\sigma}{d\Omega}(LR \to LR) = \frac{\sigma_0}{4\pi}\left(\frac{1 + \cos\theta_{\text{c.m.}}}{2}\right)^2\,. \tag{142}$$

Thus

$$\sigma(RR \to RR) = \sigma(LL \to LL) = 3\sigma(RL \to RL) = 3\sigma(LR \to LR)\,. \tag{143}$$

We first simplify the calculation by assuming the numbers of protons and neutrons are equal in the target nucleus, and neglecting the effect of antiquarks in the nucleon. (We shall use the shorthand $\nu = \nu_\mu$ and $\bar\nu = \bar\nu_\mu$.) Then

$$R_\nu = \frac{\sigma(\nu u \to \nu u) + \sigma(\nu d \to \nu d)}{\sigma(\nu d \to \mu^- u)}\,, \qquad R_{\bar\nu} = \frac{\sigma(\bar\nu u \to \bar\nu u) + \sigma(\bar\nu d \to \bar\nu d)}{\sigma(\bar\nu u \to \mu^+ d)}\,. \tag{144}$$

One can write the effective Hamiltonian (139) in the form

$$\mathcal{H}_{\nu q}^{NC} = \frac{G_F}{\sqrt{2}} [\bar{\nu}\gamma_\mu(1-\gamma_5)\nu] \left[\bar{u}\gamma^\mu(1-\gamma_5)u\epsilon_L(u) + \bar{u}\gamma^\mu(1+\gamma_5)u\epsilon_R(u) \right.$$
$$\left. + \bar{d}\gamma^\mu(1-\gamma_5)d\epsilon_L(d) + \bar{d}\gamma^\mu(1+\gamma_5)d\epsilon_R(d) \right], \quad (145)$$

where

$$\epsilon_L(u) = \frac{1}{2} - \frac{2}{3}\sin^2\theta, \qquad \epsilon_R(u) = -\frac{2}{3}\sin^2\theta, \quad (146)$$

$$\epsilon_L(d) = -\frac{1}{2} + \frac{1}{3}\sin^2\theta, \qquad \epsilon_R(d) = \frac{1}{3}\sin^2\theta. \quad (147)$$

Taking account of the relations (143), one finds

$$R_\nu = [\epsilon_L(u)]^2 + \frac{1}{3}[\epsilon_R(u)]^2 + [\epsilon_L(d)]^2 + \frac{1}{3}[\epsilon_R(d)]^2, \quad (148)$$

$$R_{\bar{\nu}} = [\epsilon_L(u)]^2 + 3[\epsilon_R(u)]^2 + [\epsilon_L(d)]^2 + 3[\epsilon_R(d)]^2, \quad (149)$$

where we have used the fact that $\sigma(\nu d \to \mu^- d) = 3\sigma(\bar{\nu}u \to \mu^+ d)$. The results are

$$R_\nu = \frac{1}{2} - \sin^2\theta + \frac{20}{27}\sin^4\theta, \qquad R_{\bar{\nu}} = \frac{1}{2} - \sin^2\theta + \frac{20}{9}\sin^4\theta. \quad (150)$$

If we consider also the antiquark content of nucleons, this result may be generalised (Llewellyn Smith 1983) by defining

$$r \equiv \frac{\sigma(\bar{\nu}N \to \mu^+ X)}{\sigma(\nu N \to \mu^- X)}. \quad (151)$$

Instead of (150) one then finds

$$R_\nu = \frac{1}{2} - \sin^2\theta + \frac{5}{9}(1+r)\sin^4\theta, \qquad R_{\bar{\nu}} = \frac{1}{2} - \sin^2\theta + \frac{5}{9}(1+\frac{1}{r})\sin^4\theta. \quad (152)$$

Some experimental values of R_ν, $R_{\bar{\nu}}$, and r are shown in Table 4 (Conrad *et al.* 1998). The relation between R_ν and $R_{\bar{\nu}}$ as a function of $\sin^2\theta$ is plotted in Figure 12. This result has a couple of interesting features.

Experiment	R_ν	$R_{\bar{\nu}}$	r
CHARM	0.3091 ± 0.0031	0.390 ± 0.014	0.456 ± 0.011
CDHS	0.3135 ± 0.0033	0.376 ± 0.016	0.409 ± 0.014
Average	0.3113 ± 0.0023	0.384 ± 0.011	0.429 ± 0.011

Table 4. *Neutrino neutral-current parameters.*

The observed $R_{\bar{\nu}}$ is very close to its minimum possible value of less than 0.4. Initially this made the observation of neutral currents quite challenging. Note that R_ν is even smaller. Its value provides the greatest sensitivity to $\sin^2\theta$. It is also more precisely measured than $R_{\bar{\nu}}$ (in part, because neutrino beams are easier to achieve than antineutrino beams). The effect of r on the determination of $\sin^2\theta$ is relatively mild.
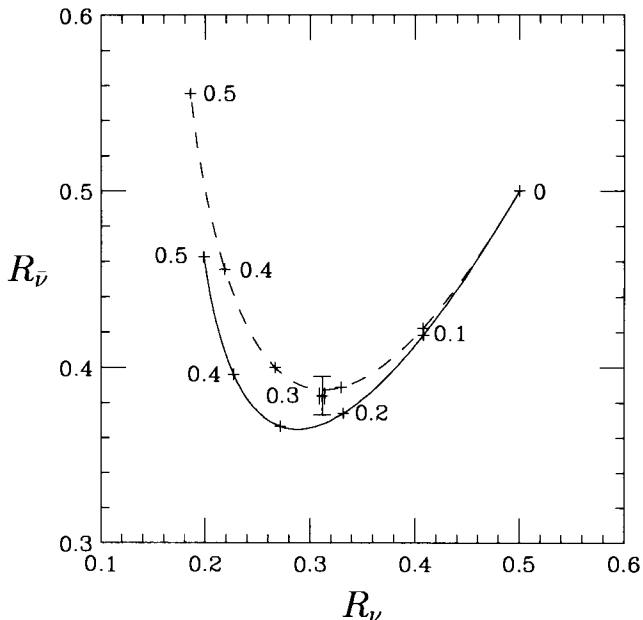
Figure 12. *The Weinberg-Salam "nose" depicting the relation between R_ν and $R_{\bar\nu}$. The solid line corresponds to $r = 0.429$, close to the actual situation; the dashed line corresponds to the idealised case $r = 1/3$ in which antiquarks in the nucleon are neglected. The plotted point with error bars corresponds to the average of measured values.*

A recent determination of $\sin^2\theta$ (Zeller *et al.* 1999), based on a method proposed by Paschos and Wolfenstein (1973), makes use of the ratio

$$R^- \equiv \frac{\sigma(\nu N \to \nu X) - \sigma(\bar\nu N \to \bar\nu X)}{\sigma(\nu N \to \mu^- X) - \sigma(\bar\nu N \to \mu^+ X)} = \frac{R_\nu - r R_{\bar\nu}}{1 - r} = \frac{1}{2} - \sin^2\theta. \qquad (153)$$

In these differences of neutrino and antineutrino cross sections, effects of virtual quark-antiquark pairs in the nucleon ("sea quarks," as opposed to "valence quarks") cancel one another, and an important systematic error associated with heavy quark production (as in $\nu s \to \mu^- c$) is greatly reduced. The result is

$$\sin^2\theta^{\text{ (on-shell)}} = 0.2253 \pm 0.0019(\text{stat.}) \pm 0.0010(\text{syst.}), \qquad (154)$$

which implies a W mass

$$M_W \equiv M_Z \cos\theta^{\text{ (on-shell)}} = 80.21 \pm 0.11 \text{ GeV}. \qquad (155)$$

The "on-shell" designation for $\sin^2\theta$ is necessary when discussing higher-order electroweak radiative corrections, which we shall do in the next section.

A more recent analysis by Zeller *et al.* (2001) finds

$$\sin^2\theta^{\text{ (on-shell)}} = 0.2277 \pm 0.0014(\text{stat.}) \pm 0.0008(\text{syst.}), \qquad (156)$$

equivalent to $M_W = 80.136 \pm 0.084$GeV. Incorporation of this result into the electroweak fits described in the next section is likely to somewhat relax constraints on the Higgs boson mass: See Rosner (2001).

4.6   Z and top quark properties

We have already noted the prediction and measurement of the W mass and width. The Z mass and width are very precisely determined by studying the shape of the cross section for electron-positron annihilation as one varies the energy across the Z pole. The results (LEP Electroweak Working Group [LEP EWWG] 2001) are

$$M_Z = 91.1875 \pm 0.0021 \text{ GeV}, \qquad \Gamma_Z = 2.4952 \pm 0.0023 \text{ GeV}. \qquad (157)$$

In much of the subsequent discussion we shall make use of the very precise value of M_Z as one of our inputs to the electroweak theory; the two others, which will suffice to specify all parameters at lowest order of perturbation theory, will be the Fermi coupling constant $G_F = 1.16637(1) \times 10^{-5}$ GeV^{-2} and the electromagnetic fine-structure constant, evolved to a scale M_Z^2: α^{-1} $^{(\overline{\text{MS}})}(M_Z^2) = 128.933 \pm 0.021$ (Davier and Höcker 1998). This last quantity depends for its determination upon a precise evaluation of hadronic contributions to vacuum polarisation, and is very much the subject of current discussion.

The relative branching fractions of the Z to various final states may be calculated on the basis of (138). One may write this expression as

$$\mathcal{L}_{Zf\bar{f}} = (g^2 + g'^2)^{1/2} \bar{f} \, \mathbb{Z}[(1 - \gamma_5)a_L + (1 + \gamma_5)a_R]f. \qquad (158)$$

The values of a_L and a_R for each fermion are shown in Table 5.

Channel	a_L	a_R	Partial width (MeV)	Number of channels	Subtotal (MeV)
$\nu\bar{\nu}$	$\frac{1}{4}$	0	165.9	3	498
$l\bar{l}$	$\frac{1}{2}\left(-\frac{1}{2} + \sin^2\theta\right)$	$\frac{1}{2}\sin^2\theta$	83.4	3	250
$u\bar{u}, \ c\bar{c}$	$\frac{1}{2}\left(\frac{1}{2} - \frac{2}{3}\sin^2\theta\right)$	$\frac{1}{2}\left(-\frac{2}{3}\sin^2\theta\right)$	296.5	2	593
$d\bar{d}, \ s\bar{s}$	$\frac{1}{2}\left(-\frac{1}{2} + \frac{1}{3}\sin^2\theta\right)$	$\frac{1}{2}\left(\frac{1}{3}\sin^2\theta\right)$	382.1	2	764
$b\bar{b}$	$\frac{1}{2}\left(-\frac{1}{2} + \frac{1}{3}\sin^2\theta\right)$	$\frac{1}{2}\left(\frac{1}{3}\sin^2\theta\right)$	372.8	1	373
Total					2478

Table 5. *Contributions to Γ_Z predicted in lowest-order electroweak theory (including leading-order QCD corrections to hadronic channels). Here we have taken $\sin^2\theta = 0.231$ and $\alpha_S(M_Z^2) = 0.12$.*

The partial width of Z into $f\bar{f}$ is

$$\Gamma(Z \to f\bar{f}) = \frac{4G_F}{3\pi\sqrt{2}} M_Z^3 (a_L^2 + a_R^2) n_c, \qquad (159)$$

where n_c is the number of colours of fermions f: 1 for leptons, 3 for quarks.

The predicted partial width for each $Z \to \nu\bar{\nu}$ channel is independent of $\sin^2\theta$:

$$\Gamma(Z \to \nu\bar{\nu}) = \frac{G_F}{\sqrt{2}} \frac{M_Z^3}{12\pi} = 165.9 \text{ MeV} \qquad (160)$$

using the observed value of M_Z. The partial decay rates to other channels are expected to be in the ratios

$$\nu\bar{\nu} : e^+e^- : u\bar{u} : d\bar{d} =$$

$$1 : 1 - 4\sin^2\theta + 8\sin^4\theta : 3 - 8\sin^2\theta + \frac{32}{3}\sin^4\theta : 3 - 4\sin^2\theta + \frac{8}{3}\sin^4\theta, \quad (161)$$

or 1: 0.503: 1.721: 2.218 for $\sin^2\theta = 0.231$. A small kinematic correction for the non-zero b quark mass leads to a suppression factor

$$\Phi_{b\bar{b}} = \left(1 - \frac{4m_b^2}{M_Z^2}\right)^{1/2} \left[f_V\left(1 + \frac{2m_b^2}{M_Z^2}\right) + f_A\left(1 - \frac{4m_b^2}{M_Z^2}\right)\right], \quad (162)$$

where f_V and $f_A = 1 - f_V$ are the relative fractions of the partial decay width proceeding via the vector ($\sim a_L + a_R$) and axial-vector ($\sim a_L - a_R$) couplings. For $\sin^2\theta = 0.23$, $f_V \simeq 1/3$, $f_A \simeq 2/3$, and $\Phi_{b\bar{b}} \simeq 0.988$. A further correction to $\Gamma(Z \to b\bar{b})$, important for the precise determinations in the next section, is associated with loop graphs associated with top quark exchange (see the review by Chivukula 1995), and is of the same size, about 0.988. Taking a correction factor $(1 + \alpha_S/\pi)$ with $\alpha_S(M_Z^2) = 0.12$ for the hadronic partial widths of the Z, we then predict the contributions to Γ_Z listed in Table 5. (The $t\bar{t}$ channel is, of course, kinematically forbidden.)

The measured Z width (157) is in qualitative agreement with the prediction, but above it by about 0.7%. This effect is a signal of higher-order electroweak radiative corrections such as loop diagrams involving the top quark and the Higgs boson. Similarly, the observed value of $\Gamma(Z \to e^+e^-)$, assuming lepton universality, is 83.984 ± 0.086 MeV, again higher by 0.7% than the predicted value of 83.4MeV. We shall return to these effects in the next section.

The width of the Z is sensitive to additional $\nu\bar{\nu}$ pairs. Clearly there is no room for an additional light pair coupling with full strength. Taking account of all precision data and electroweak corrections, the latest determination of the "invisible" width of the Z (see the compilations by the LEP EWWG 2001 and by Langacker 2001) fixes the number of "light" neutrino species as $N_\nu = 2.984 \pm 0.008$.

The Z is produced copiously in e^+e^- annihilation when the center-of-mass energy $\sqrt{s}$ is tuned to M_Z. The Stanford Linear Collider (SLC) and the Large Electron-Positron Collider at CERN (LEP) exploited this feature. The cross section of production of a final state f near the resonance, ignoring the effect of the virtual photon in the direct channel, should be

$$\sigma(e^+e^- \to Z^0 \to f) = 12\pi \frac{\Gamma(Z \to e^+e^-)\Gamma(Z \to f)}{(s - M_Z^2)^2 + M_Z^2\Gamma_Z^2}. \quad (163)$$

Setting $\mathcal{B}_{e^+e^-} \equiv \Gamma(Z^0 \to e^+e^-)/\Gamma_Z \simeq 3.37\%$ we see that at resonance the peak total cross section should be $\sigma_{\text{peak}} = 12\pi\mathcal{B}_{e^+e^-}/M_Z^2 \simeq 59.4$ nb, corresponding to

$$R_{\text{pk}} \equiv \frac{\sigma(e^+e^- \to Z^0 \to \text{all})}{\sigma(e^+e^- \to \mu^+\mu^-)} = \frac{9\mathcal{B}_{e^+e^-}}{\alpha^2} \simeq 5000. \quad (164)$$

This is a spectacular value of R, which is only a few units in the range of lower-energy e^+e^- colliders. Of course, not all of the cross section at the Z peak is visible: Nearly 12 nb goes into neutrinos! Another 6 nb goes into charged lepton pairs, leaving $\sigma_{\text{peak, hadrons}} = 41.541 \pm 0.037$ nb.

We close this subsection with a brief discussion of spin-dependent asymmetries at the Z. These are some of the most powerful sources of information on $\sin^2 \theta$. They have been measured both at LEP (through forward-backward asymmetries) and at SLC (through the use of polarised electron beams).

The discussion makes use of an elementary feature of vector- and axial-vector couplings. Processes involving such couplings to a real or virtual particle (such as the Z) always conserve chirality. In the direct-channel reactions $e^- e^+ \to Z \to f\bar{f}$ this means that a left- (right-)handed electron only interacts with a right- (left-) handed positron, and if the final fermion f is left- (right-)handed then the final anti-fermion $\bar{f}$ will be right- (left-) handed. Moreover, such reactions have characteristic angular distributions, with

$$\frac{d\sigma(e_L^- \to f_L)}{d\Omega} = \sigma_0 (a_L^e)^2 (a_L^f)^2 \left(\frac{1 + \cos\theta_{\text{c.m.}}}{2} \right)^2 ; \tag{165}$$

$$\frac{d\sigma(e_R^- \to f_R)}{d\Omega} = \sigma_0 (a_R^e)^2 (a_R^f)^2 \left(\frac{1 + \cos\theta_{\text{c.m.}}}{2} \right)^2 ; \tag{166}$$

$$\frac{d\sigma(e_L^- \to f_R)}{d\Omega} = \sigma_0 (a_L^e)^2 (a_R^f)^2 \left(\frac{1 - \cos\theta_{\text{c.m.}}}{2} \right)^2 ; \tag{167}$$

$$\frac{d\sigma(e_R^- \to f_L)}{d\Omega} = \sigma_0 (a_R^e)^2 (a_L^f)^2 \left(\frac{1 - \cos\theta_{\text{c.m.}}}{2} \right)^2 ; \tag{168}$$

where σ_0 is some common factor, and the $a_{L,R}$ are given in Table 5. Several asymmetries can be formed using these results.

The *polarised electron left-right asymmetry* compares the cross sections for producing fermions using right-handed and left-handed polarised electrons, as can be produced and monitored at the SLC. The cross section asymmetry is given by

$$
\begin{aligned}
A_{LR}^e(\text{hadrons}) &\equiv \frac{\sigma(e_L^- e^+ \to \text{hadrons}) - \sigma(e_R^- e^+ \to \text{hadrons})}{\sigma(e_L^- e^+ \to \text{hadrons}) + \sigma(e_R^- e^+ \to \text{hadrons})} \\
&= \frac{(a_L^e)^2 - (a_R^e)^2}{(a_L^e)^2 + (a_R^e)^2} = \frac{1 - 4\sin^2\theta}{1 - 4\sin^2\theta + 8\sin^4\theta} .
\end{aligned}
\tag{169}
$$

The measured value (LEP EWWG 2001) $A_{LR}^e(\text{hadrons}) = 0.1514 \pm 0.0022$ corresponds to $\sin^2\theta = 0.23105 \pm 0.00028$ using this formula. (We shall discuss small corrections in the next section.)

The *forward-backward asymmetry* in $e^+ e^- \to f\bar{f}$ uses the fact that

$$\left(\int_0^1 - \int_{-1}^0 \right) d\cos\theta (1 \pm \cos\theta)^2 = \pm 2 , \qquad \left(\int_0^1 + \int_{-1}^0 \right) d\cos\theta (1 \pm \cos\theta)^2 = \frac{8}{3} , \tag{170}$$

so that

$$
\begin{aligned}
A_{FB}^f &\equiv \frac{\sigma(e^- e^+ \to f\bar{f})_{\text{fwd}} - \sigma(e^- e^+ \to f\bar{f})_{\text{back}}}{\sigma(e^- e^+ \to f\bar{f})_{\text{fwd}} + \sigma(e^- e^+ \to f\bar{f})_{\text{back}}} \\
&= \frac{3}{4} \left(\frac{(a_L^e)^2 - (a_R^e)^2}{(a_L^e)^2 + (a_R^e)^2} \right) \left(\frac{(a_L^f)^2 - (a_R^f)^2}{(a_L^f)^2 + (a_R^f)^2} \right) = \frac{3}{4} A_{LR}^e A_{LR}^f
\end{aligned}
\tag{171}
$$

These quantities can be measured not only for charged leptons, but also for quarks such as the b, whose decays allow for a distinction to be made (at least on a statistical basis) between b and $\bar{b}$.

The discovery of the top quark by the CDF (1994) and D0 (1995) Collaborations culminated nearly two decades of detector and machine work at the Fermilab Tevatron. A ring of superconducting magnets was added to the 400GeV Fermilab accelerator, more than doubling its energy. Low-energy rings were added to accumulate and store antiprotons, which were then injected into the superconducting ring and made to collide with oppositely-directed protons at a centre-of-mass energy of 1.8TeV. The top quarks were produced in the reaction $p\bar{p} \to t\bar{t} + \dots$.

The top quark's mass is currently measured to be $m_t = 174.3 \pm 5.1$ GeV. It couples mainly to b, as expected in the pattern of couplings discussed in Section 3. One determination (see Gilman, Kleinknecht, and Renk 2000 for details) is that

$$\frac{|V_{tb}|^2}{|V_{td}|^2 + |V_{ts}|^2 + |V_{tb}|^2} = 0.99 \pm 0.29. \tag{172}$$

This result makes use of the measured fraction of the decays $t \to be^+\nu_e$ in top semileptonic decays.

The top quark is the only quark heavy enough to decay directly to another quark (mainly b) and a real W. Its decay width is

$$\Gamma(t \to W^+b) = \frac{G_F m_t^3}{8\pi\sqrt{2}} \left[\left(1 - \frac{M_W^2}{m_t^2}\right)^2 \left(1 + 2\frac{M_W^2}{m_t^2}\right) \right] \simeq 1.53 \text{ GeV}. \tag{173}$$

This is larger than the typical spacing between quarkonium levels (see Figures 4 and 5), and so there is not expected to be a rich spectroscopy of $t\bar{t}$ levels, but only a mild enhancement near threshold of the reaction $e^+e^- \to t\bar{t}$, associated with the production of the 1S level (Kwong 1991, Strassler and Peskin 1991). A good review of present and anticipated top quark physics is given by Willenbrock (2000).

5   Higgs boson and beyond

5.1   Searches for a standard Higgs boson

Let us assume that all quark and lepton masses and all W and Z masses arise from the vacuum expectation value of a single Higgs boson: $\langle\phi^0\rangle = v/\sqrt{2}$, where the strength of the Fermi coupling requires $v = 246$GeV. The Yukawa coupling g_{Yf} (107) for a fermion f is related to the fermion's mass: $g_{Yf} = \sqrt{2}m_f/v$. (It is a curious feature of the top quark's mass that, within present errors, $g_{Yt} = 1$. Since fermion masses "run" with scale μ, it is not clear how fundamental this relation is.) Those quarks with the greatest mass then are expected to have the greatest coupling to the physical Higgs boson $H = \sqrt{2}\phi^0 - v$. (Here we use H to denote the field represented by η in the previous section.)

The Higgs boson has a well-defined coupling to W's and Z's as a result of the discussion in the previous section. The term $(\mathbf{D}_\mu\phi)^\dagger(\mathbf{D}^\mu\phi)$ in the Lagrangian leads to

$$\mathcal{L}_{HWW} = gHM_W(W_\mu^- W^{\mu+}), \qquad \mathcal{L}_{HZZ} = \frac{(g^2 + g'^2)^{1/2}HM_Z}{2}(Z_\mu Z^\mu). \tag{174}$$

To lowest order, one find $\mathcal{L}_{HZ\gamma} = \mathcal{L}_{H\gamma\gamma} = 0$.

Processes involving the couplings (174) include $q\bar{q} \to W_{\text{virtual}} \to W + H$ and especially

$$e^+e^- \to Z_{\text{real or virtual}} \to Z_{\text{virtual or real}} + H\,, \qquad (175)$$

where the final Z^0 can be detected (for example) via its decay to e^+e^-, $\mu^+\mu^-$, or even its existence inferred from its $\nu\bar{\nu}$ decay. For a virtual intermediate and real final Z, the cross section (Quigg 1983) is

$$\sigma(e^+e^- \to ZH) = \frac{\pi\alpha^2(p^{*2} + 3M_Z^2)}{24\sin^4\theta\cos^4\theta(M_Z^2 - s)^2}\left(1 - 4\sin^2\theta + 8\sin^4\theta\right)\frac{2p^*}{\sqrt{s}}\,, \qquad (176)$$

where p^* is the final c.m. 3-momentum. This cross section behaves as $1/s$ for large s (as does any cross section for production of $q\bar{q}$, $\mu^+\mu^-$, ...), so that as $s \to \infty$,

$$\frac{\sigma(e^+e^- \to ZH)}{\sigma(e^+e^- \to \mu^+\mu^-)} \to \frac{1 - 4\sin^2\theta + 8\sin^4\theta}{128\sin^4\theta\cos^4\theta} \simeq \frac{1}{8}\,. \qquad (177)$$

At very high energies, the Higgs boson can be produced by means of W^+W^- and ZZ fusion; the (virtual) W's and Z's can be produced in either hadron-hadron or lepton-lepton collisions. A further proposal for producing Higgs bosons is by means of muon-muon collisions.

For Higgs bosons far above WW and ZZ threshold, one expects (Eichten *et al.* 1984)

$$\Gamma_H = \Gamma(H \to W^+W^-) + \Gamma(H \to ZZ) = \frac{3G_F}{16\pi\sqrt{2}}M_H^3 \simeq 60 \ \text{GeV} \left(\frac{M_H}{500 \ \text{GeV}}\right)^3\,, \qquad (178)$$

as one can show with the help of (174). The longitudinal degrees of freedom of the W and Z provide the dominant contribution to the decay width in this limit. For $M_H = 1\text{TeV}$, this relation implies that the Higgs boson's width will be nearly 500GeV. Such a broad object will be difficult to separate from background. However, mixed signals for a a much lighter Higgs boson have already been received at LEP.

At the very highest LEP energies attained, $\sqrt{s} \leq 209$ GeV, the four LEP collaborations ALEPH, DELPHI, L3, and OPAL have presented combined results (LEP Higgs Working Group 2001) which may be interpreted either as a lower limit on the Higgs boson mass of 114.1GeV, or as a weak signal for a Higgs boson of mass $M_H \simeq 115.6$GeV produced by the above process. This latter interpretation is driven in large part by the ALEPH data sample (Barate *et al.* 2001). The main decay mode of a Higgs boson in this mass range is expected to be $b\bar{b}$, with $\tau^+\tau^-$ taking second place.

LEP now has ceased operation in order to make way for the Large Hadron Collider (LHC), which will collide 7TeV protons with 7TeV protons and should have no problem producing such a boson. The LHC is scheduled to begin operation in 2006. In the meantime, the Fermilab Tevatron has resumed $p\bar{p}$ collider operation after a hiatus of 5 years. Its scheduled "Run II" is initially envisioned to provide an integrated luminosity of 2 fb^{-1}, which is thought to be sufficient to rival the sensitivity of the LEP search (Carena *et al.* 2000), making use of the subprocess $q\bar{q} \to W_{\text{virtual}} \to W + H$. With 10 fb^{-1} per detector, a benchmark goal for several years of running with luminosity improvements, it should be possible to exclude a Higgs boson with standard couplings nearly up to the ZZ

threshold of 182GeV, and to see a 3σ signal if $M_H \leq 125$ GeV. Other scenarios, including the potential for discovering the Higgs boson(s) of the Minimal Supersymmetric Standard Model (MSSM) are given by Carena *et al.* (2000). Meanwhile, we shall turn to the wealth of precise measurements of electroweak properties of the Z, W, top quark, and lighter fermions as indirect sources of information about the Higgs boson and other new physics.

5.2 Precision electroweak tests

We have calculated processes to lowest electroweak order in the previous section, with the exception that we took account of vacuum polarisation in the photon propagator, which leads to a value of α^{-1} closer to 129 than to 137.037 at the mass scale of the Z. The lowest-order description was found to be adequate at the percent level, but many electroweak measurements are now an order of magnitude more precise. As one example, we found that the predicted total and leptonic Z widths both fell short of the corresponding experimental values by about 0.7%. Higher-order electroweak corrections are needed to match the precision of the new data. These corrections can shed fascinating light on new physics, as well as validating the original motivation for the electroweak theory (which was to be able to perform higher-order calculations).

We shall describe a language introduced by Peskin and Takeuchi (1990) for precise electroweak tests which allows the constraints associated with nearly every observable to be displayed on a two-dimensional plot. The Standard Model implies a particular locus on this plot for every value of m_t and M_H, so one can see how observables can vary with m_t (not much, now that m_t is so well measured) and M_H. Moreover, one can spot at a glance if a particular measurement is at variance with others; this can either signify physics outside the purview of the two-dimensional plot, or systematic experimental error.

The corrections which fall naturally into the two-dimensional description are those known as *oblique corrections*. The name stems from the fact that they do not directly affect the fermions participating in the processes of interest, but appear as vacuum polarisation corrections in gauge boson propagators. In that sense processes which are sensitive to oblique corrections have a broad reach for discovering new physics, since they do not rely on a new particle's having to couple directly to the external fermion in question.

The oblique correction first identified by Veltman (1977), still the most important, is that due to top quarks in W and Z boson propagators. The large splitting between the top and bottom quarks' masses violates a *custodial SU(2)* symmetry (Sikivie *et al.* 1980) responsible for preserving the tree-level relation $M_W = M_Z \cos\theta$ mentioned in the previous section. As a result, an effect is generated which is equivalent to having a Higgs *triplet* vacuum expectation value.

For the photon, gauge invariance prohibits contributions quadratic in fermion masses, but for the W and Z, no such prohibition applies. The vacuum polarisation diagrams with $W^+ \to t\bar{b} \to W^+$ and $Z \to (t\bar{t}, b\bar{b}) \to Z$ lead to a modification of the relation between G_F, coupling constants, and M_Z for neutral-current exchanges:

$$\frac{G_F}{\sqrt{2}} = \frac{g^2 + g'^2}{8M_Z^2} \quad \Longrightarrow \quad \frac{G_F}{\sqrt{2}}\rho = \frac{g^2 + g'^2}{8M_Z^2}, \qquad \rho \simeq 1 + \frac{3G_F m_t^2}{8\pi^2\sqrt{2}}. \tag{179}$$

The Z mass is now related to the weak mixing angle by

$$M_Z^2 = \frac{\pi\alpha}{\sqrt{2}G_F\rho\sin^2\theta\cos^2\theta},$$ (180)

where we have omitted some small terms logarithmic in m_t. A precise measurement of M_Z now specifies θ only if m_t is known, so $\theta = \theta(m_t)$ and hence $M_W^2 = \pi\alpha/(\sqrt{2}G_F\sin^2\theta)$ is also a function of m_t.

The factor of ρ in (179) will multiply every neutral-current four-fermion interaction in the electroweak theory. Thus, for, example, cross sections for charge-preserving interactions of neutrinos with matter will be proportional to ρ^2, while parity-violating neutral-current amplitudes (to be discussed below) will be proportional to ρ. Partial decay widths of the Z, since they involve the combination $g^2 + g'^2$, will be proportional to ρ. A large part of the 0.7% correction mentioned previously is due to $\rho > 1$. The observed values of $M_W/M_Z = \rho\cos\theta$ and $\sin^2\theta$ also are much more compatible with each other for a value of ρ exceeding 1 by about a percent.

The W and Z propagators are also affected by virtual Higgs-boson states due to the couplings (174). Small corrections, logarithmic in M_H, affect all the observables, but notably ρ.

In order to display dependence of electroweak observables on such quantities as the top quark and Higgs boson masses m_t and M_H, we choose to expand the observables about "nominal" values calculated by Marciano (2000) for specific m_t and M_H. We thereby bypass a discussion of "direct" radiative corrections which are independent of m_t, M_H, and new particles. We isolate the dependence on m_t, M_H, and new physics arising from "oblique" corrections associated with loops in the W and Z propagators.

For $m_t = 174.3\,\mathrm{GeV}$, $M_H = 100\,\mathrm{GeV}$, the measured value of M_Z leads to a nominal expected value of $\sin^2\theta_{\mathrm{eff}} = 0.2314$. In what follows we shall interpret the effective value of $\sin^2\theta$ as that measured via leptonic vector and axial-vector couplings, namely $\sin^2\theta_{\mathrm{eff}} \equiv (1/4)(1 - [g_V^\ell/g_A^\ell])$.

Defining the parameter T by $\Delta\rho \equiv \alpha T$, we find

$$T \simeq \frac{3}{16\pi\sin^2\theta}\left[\frac{m_t^2 - (174.3\text{ GeV})^2}{M_W^2}\right] - \frac{3}{8\pi\cos^2\theta}\ln\frac{M_H}{100\text{ GeV}}.$$ (181)

The weak mixing angle θ, the W mass, and other electroweak observables now depend on m_t and M_H.

The weak charge-changing and neutral-current interactions are probed under a number of different conditions, corresponding to different values of momentum transfer. For example, muon decay occurs at momentum transfers small with respect to M_W, while the decay of a Z into fermion-antifermion pairs imparts a momentum of nearly $M_Z/2$ to each member of the pair. Small "oblique" corrections, logarithmic in m_t and M_H, arise from contributions of new particles to the photon, W, and Z propagators. Other (smaller) "direct" radiative corrections are important in calculating actual values of observables.

We may then replace the lowest-order relations between G_F, couplings, and masses by

$$\frac{G_F}{\sqrt{2}} = \frac{g^2}{8M_W^2}\left(1 + \frac{\alpha S_W}{4\sin^2\theta}\right), \qquad \frac{G_F\rho}{\sqrt{2}} = \frac{g^2 + g'^2}{8M_Z^2}\left(1 + \frac{\alpha S_Z}{4\sin^2\theta\cos^2\theta}\right),$$ (182)

where S_W and S_Z are coefficients representing variation with momentum transfer. Together with T, they express a wide variety of electroweak observables in terms of quantities sensitive to new physics. (The presence of such corrections was noted quite early by Veltman 1977.) The Peskin and Takeuchi (1990) variable U is equal to $S_W - S_Z$, while $S \equiv S_Z$.

Expressing the "new physics" effects in terms of deviations from nominal values of top quark and Higgs boson masses, we have the expression (181) for T, while contributions of Higgs bosons and of possible new fermions U and D with electromagnetic charges Q_U and Q_D to S_W and S_Z, in a leading-logarithm approximation, are (Kennedy and Langacker 1990)

$$S_Z = \frac{1}{6\pi} \left[\ln \frac{M_H}{100 \ \text{GeV}/c^2} + \sum N_C \left(1 - 4\overline{Q} \ln \frac{m_U}{m_D} \right) \right], \qquad (183)$$

$$S_W = \frac{1}{6\pi} \left[\ln \frac{M_H}{100 \ \text{GeV}/c^2} + \sum N_C \left(1 - 4Q_D \ln \frac{m_U}{m_D} \right) \right]. \qquad (184)$$

The expressions for S_W and S_Z are written for doublets of fermions with N_C colours and $m_U \geq m_D \gg m_Z$, while $\overline{Q} \equiv (Q_U + Q_D)/2$. The sums are taken over all doublets of new fermions. In the limit $m_U = m_D$, one has equal contributions to S_W and S_Z. For a single Higgs boson and a single heavy top quark, (183) and (184) become

$$S_Z = \frac{1}{6\pi} \left[\ln \frac{M_H}{100 \ \text{GeV}/c^2} - 2 \ln \frac{m_t}{174.3 \ \text{GeV}/c^2} \right],$$

$$S_W = \frac{1}{6\pi} \left[\ln \frac{M_H}{100 \ \text{GeV}/c^2} + 4 \ln \frac{m_t}{174.3 \ \text{GeV}/c^2} \right], \qquad (185)$$

where the leading-logarithm expressions are of limited validity for M_H and m_t far from their nominal values. (We shall plot contours of S and T for fixed m_t and M_H values without making these approximations.) A degenerate heavy fermion doublet with N_c colours thus contributes $\Delta S_Z = \Delta S_W = N_c/6\pi$. For example, in a minimal dynamical symmetry-breaking ("technicolour") scheme, with a single doublet of $N_c = 4$ fermions, one will have $\Delta S = 2/3\pi \simeq 0.2$. This will turn out to be marginally acceptable, while many non-minimal schemes, with large numbers of doublets, will be seen to be ruled out.

Many analyses of present electroweak data within the S, T rubric are available (e.g., Swartz 2001). We shall present a "cartoon" version after discussing possible extensions of the Higgs system. Meanwhile we note briefly a topic which will not enter that discussion.

The anomalous magnetic moment of the electron and muon have been measured ever more precisely. The latest measurement of the μ^+ (Brown *et al.* 2001), performed in a special storage ring at Brookhaven National Laboratory, gives

$$a_{\mu^+,\text{exp}} \equiv \left(\frac{g-2}{2} \right)_{\mu^+} = 11 \ 659 \ 202(14)(6) \times 10^{-10} \quad (1.3 \ \text{ppm}), \qquad (186)$$

The theoretical value (CPT invariance implies $a_{\mu^+} = a_{\mu^-}$) is

$$a_{\mu,\text{th}} \simeq 11 \ 659 \ 177(7) \times 10^{-10} \quad (0.6 \ \text{ppm}), \qquad (187)$$

being the sum of three terms:

$$
\begin{aligned}
a_{\mu,\text{QED}} &= 11\,658\,470.56(0.29) \times 10^{-10} \quad (0.025 \text{ ppm})\,, \\
a_{\mu,\text{weak}} &= 15.1(0.4) \times 10^{-10} \quad\quad\quad\quad (0.03 \text{ ppm})\,, \\
a_{\mu,\text{had}} &\simeq 691(7) \times 10^{-10} \quad\quad\quad\quad\quad (0.6 \text{ ppm})\,,
\end{aligned}
$$

where we have incorporated a recently-discovered sign change in the hadronic light-by-light scattering contribution (Knecht and Nyffeler 2001; Hayakawa and Kinoshita 2001). The difference,

$$
a_{\mu^+,\text{exp}} - a_{\mu^+,\text{th}} = 25(17) \times 10^{-10}\,, \tag{188}
$$

is not yet known precisely enough to test the expected weak contribution. Results of analysing a larger data sample are expected shortly.

5.3   Multiple Higgs doublets and Higgs triplets

There are several reasons for introducing a more complicated Higgs boson spectrum. Reasons for introducing separate Higgs doublets for u-type and d-type quarks include higher symmetries following from attempts to unify the strong and electroweak interactions, and supersymmetry. We examine the simplest model with more than one Higgs doublet, in which a single doublet couples to d-type quarks and charged leptons, and a different doublet couples to u-type quarks. This model turns out to naturally avoid flavour-changing neutral currents associated with Higgs exchange (Glashow and Weinberg 1977).

Let us denote by ϕ_u the Higgs boson coupling to u-type quarks and by ϕ_d the boson coupling to d-type quarks and charged leptons. We let

$$
\langle \phi_u \rangle = v_u/\sqrt{2}\,, \qquad \langle \phi_d \rangle = v_d/\sqrt{2}\,. \tag{189}
$$

The contribution of ϕ_u and ϕ_d to W and Z masses comes from

$$
\mathcal{L}_K + (\mathbf{D}_\mu \phi_u)^\dagger (\mathbf{D}^\mu \phi_u) + (\mathbf{D}_\mu \phi_d)^\dagger (\mathbf{D}^\mu \phi_d)\,. \tag{190}
$$

We find the same $W_\mu^3 - B_\mu$ mixing pattern as before, and in fact this pattern would remain the same no matter how many Higgs doublets were introduced. The parameters v_u and v_d may be related to the quantity $v = 246\text{GeV}$ introduced earlier by $v_u^2 + v_d^2 = v^2$, whereupon all previous expressions for M_W and M_Z remain valid. One would have $v^2 = \sum_i v_i^2$ for any number of doublets.

The quark and lepton couplings to Higgs doublets are enhanced if there are multiple doublets. Since $m_q = g_Y v_q/\sqrt{2}$ ($q = u$ or d) and $v_q < v$, one has larger Yukawa couplings than in the standard single-Higgs model. A more radical consequence, however, of multiple doublets in the $\text{SU}(2)_L$ gauge theory is that there are not enough gauge bosons to "eat" all the scalar fields. In a two-doublet model, five "uneaten" scalars remain: two charged and three neutral. The phenomenology of these is well-described by Gunion *et al.* (1990).

The prediction $M_Z = M_W/\cos\theta$ is specific to the assumption that only Higgs doublets of $\text{SU}(2)_L$ exist. [$\text{SU}(2)_L$ singlets which are neutral also have $Y = 0$, and do not affect W and Z masses.] If triplets or higher representations of $\text{SU}(2)$ exist, the situation is changed. We shall examine two cases of triplets: a complex triplet with charges $(++,+,0)$ and one with charges $(+,0,-)$.

Consider first a complex triplet of the form

$$
\Phi \equiv \begin{bmatrix} \Phi^{++} \\ \Phi^{+} \\ \Phi^{0} \end{bmatrix}, \qquad I_{3L} = \begin{cases} +1 \\ 0 \\ -1 \end{cases}. \tag{191}
$$

Since $Q = I_{3l} + (Y/2)$, one must have $Y = 2$ for this triplet. In calculating $|\mathbf{D}_\mu \Phi|^2$ we will need the triplet representation for weak isospin:

$$
I_3 = \begin{bmatrix} 1 & & \\ & 0 & \\ & & -1 \end{bmatrix}, \qquad I_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \qquad I_2 = \frac{i}{\sqrt{2}} \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & -1 \\ o & 1 & 0 \end{bmatrix}. \tag{192}
$$

The result, if $\langle \Phi^0 \rangle = V_{1,-1}/\sqrt{2}$, is that

$$
\langle |\mathbf{D}_\mu \Phi|^2 \rangle = \frac{V_{1,-1}^2}{2} \left\{ \frac{g^2}{2} [(W^1)^2 + (W^2)^2] + (-gW^3 + g'B)^2 \right\}. \tag{193}
$$

The same combination of W^3 and B gets a mass as in the case of one or more Higgs doublets, simply because we assumed that it was a neutral Higgs field which acquired a vacuum expectation value. Electromagnetic gauge invariance remains valid; the photon does not acquire a mass. However, the ratio of W and Z masses is altered. In the presence of doublets and this type of triplet, we find

$$
M_W^2 = \frac{g^2}{4}(v^2 + 2V_{1,-1}^2), \qquad M_Z^2 = \left(\frac{g^2 + g'^2}{4} \right)(v^2 + 4V_{1,-1}^2), \tag{194}
$$

so the ratio $\rho = (M_W/M_Z \cos\theta)^2$ is no longer 1, but becomes

$$
\rho = \frac{v^2 + 2V_{1,-1}^2}{v^2 + 4V_{1,-1}^2}. \tag{195}
$$

This type of Higgs boson thus leads to $\rho < 1$.

A complex triplet

$$
\Phi \equiv \begin{bmatrix} \Phi^{+} \\ \Phi^{0} \\ \Phi^{-} \end{bmatrix} \tag{196}
$$

is characterised by $Y = 0$. If we let $\langle \Phi^0 \rangle = V_{1,0}/\sqrt{2}$, we find by an entirely similar calculation, that

$$
M_W^2 = \frac{g^2}{4}(v^2 + 4V_{1,0}^2), \qquad M_Z^2 = \left(\frac{g^2 + g'^2}{4} \right) v^2. \tag{197}
$$

Here we predict

$$
\rho = 1 + \frac{4V_{1,0}^2}{v^2}, \tag{198}
$$

so this type of Higgs boson leads to $\rho > 1$.

We now examine a simple set of electroweak data (Rosner 2001), updating an earlier analysis (Rosner 1999) which may be consulted for further references. (See also Peskin and Wells 2001.) We omit some data which provide similar information but are less constraining. Thus, we take only the observed values of M_W as measured at the Fermilab Tevatron and LEP-II, the leptonic width of the Z, and the value of $\sin^2\theta_{\text{eff}}$ as measured in various asymmetry experiments at the Z pole in e^+e^- collisions. We also include parity violation in atoms, stemming from the interference of Z and photon exchanges between the electrons and the nucleus. The most precise constraint at present arises from the measurement of the *weak charge* (the coherent vector coupling of the Z to the nucleus), $Q_W = \rho(Z - N - 4Z\sin^2\theta)$, in atomic cesium. The prediction $Q_W(\text{Cs}) = -73.19 \pm 0.13$ is insensitive to standard-model parameters once M_Z is specified; discrepancies are good indications of new physics.

The inputs, their nominal values for $m_t = 174.3\text{GeV}$ and $M_H = 100\text{GeV}$, and their dependences on S and T are shown in Table 6. We do not constrain the top quark mass; we display its effect on S and T explicitly. Each observable specifies a pair of parallel lines in the S-T plane. The leptonic width mainly constrains T; $\sin^2\theta^{\text{eff}}$ provides a good constraint on S with some T-dependence; and M_W lies in between. Atomic parity violation experiments constrain S with almost no T dependence. Although the errors on S they entail are too large to have much impact, we include them for illustrative purposes. Since the slopes associated with constraints are very different, the resulting allowed region is an ellipse, shown in Figure 13. [Note added: Milstein and Sushkov (2001) have noted that a correction due to the strong nuclear field changes the central value of $Q_W(\text{Cs})$ in Table 6 to $\simeq -72.2$, while Dzuba *et al.* (2001) include this and further corrections to obtain $Q_W = -72.39 \pm 0.58$.]

Quantity	Experiment	Theory
M_W (GeV/c^2)	80.451 ± 0.033 [a]	80.385 [b] $- 0.29S + 0.45T$
$\Gamma_{\ell\ell}(Z)$ (MeV)	83.984 ± 0.086 [c]	84.011 [b] $- 0.18S + 0.78T$
$\sin^2\theta^{\text{eff}}$	0.23152 ± 0.00017 [c]	0.23140 [b] $+ 0.00362S - 0.00258T$
$Q_W(\text{Cs})$	-72.5 ± 0.8	$-73.19 - 0.800S - 0.007T$
$Q_W(\text{Tl})$	-115.0 ± 4.5	$-116.8 - 1.17S - 0.06T$

Table 6. *Electroweak observables described in fit. The superscripts refer to (a) Charlton (2001), (b) Marciano (2000) and (c) LEP EWWG (2001). References for atomic physics experiment and theory are given by Rosner (2001).*

Figure 13 also shows predictions by Peskin and Wells (2001) of the standard electroweak theory. Nearly vertical lines correspond, from left to right, to Higgs boson masses $M_H = 100, 200, 300, 500, 1000$ GeV; drooping curves correspond, from top to bottom, to $+1\sigma$, central, and -1σ values of $m_t = 174.3 \pm 5.1$ GeV.

In the standard model, the combined constraints of electroweak observables such as those in Table 6 and the top quark mass favour a very light Higgs boson, with most analyses favouring a value of M_H so low that the Higgs boson should already have been discovered. The efficacy of a small amount of triplet symmetry breaking has recently been
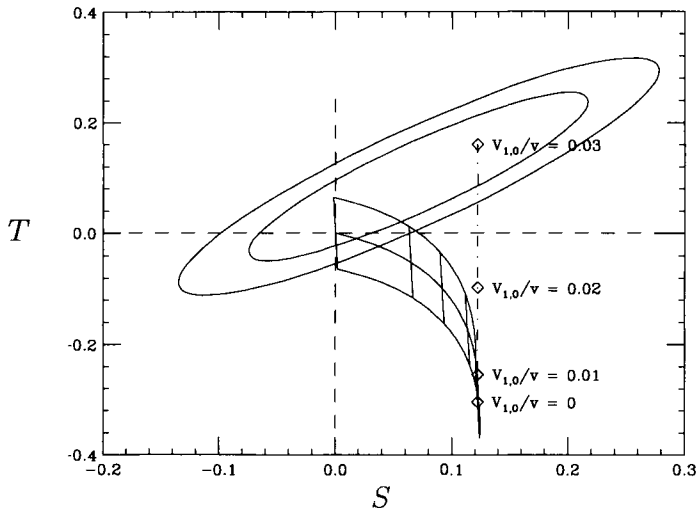
Figure 13. *Regions of 68% (inner ellipse) and 90% (outer ellipse) confidence level values of S and T based on the comparison of the theoretical and experimental electroweak observables shown in Table 6. Details are given in the text.*

stressed in a nice paper by Forshaw *et al.* (2001). It is also implied in the discussions of Dobrescu and Hill (1998), Collins *et al.* (2000), He *et al.* (2001), and Peskin (2001).

The standard model prediction for S and T curves down quite sharply in T as M_H is increased, quickly departing from the region allowed by the fit to electroweak data. (Useful analytic expressions for the contribution of a Higgs boson to S and T are given by Forshaw *et al.* 2001.) However, if a small amount of triplet symmetry breaking is permitted, the agreement with the electroweak fit can be restored. As an example, a value of $V_{1,0}/v = 0.03$ permits satisfactory agreement even for $M_H = 1\text{TeV}$, as shown by the vertical line in the Figure.

5.4 Supersymmetry, technicolour, and alternatives

What could lie beyond the standard model? The odds-on favourite among most theorists is *supersymmetry*, an extremely beautiful idea which may or may not be realised at the electroweak scale, but which almost certainly plays a role at the Planck scale at which space and time first acquire their meaning.

The simplest illustration of supersymmetry (in one time and no space dimensions!) does back to Darboux in 1882, who factored second-order differential operators into the product of two first-order operators. Dirac's famous treatment of the harmonic oscillator, writing its Hamiltonian as $H = \hbar\omega(a^\dagger a + \frac{1}{2})$, is an example of this procedure, which was generalised by Schrödinger in 1941 and Infeld and Hull in 1951. Some of this literature is reviewed by Kwong and Rosner (1986). The Hamiltonian is the generator of time translations, so this form of supersymmetry essentially amounts to saying that a time translation can be expressed as a composite of more fundamental operations.

Modern supersymmetry envisions both spatial and time translations as belonging to a super-algebra. The Lorentz group is isomorphic to $SU(2) \otimes SU(2)$ (with factors of i thrown in to account for the Minkowski metric); under this group space and time translations transform as $(1/2,1/2)$. The supercharges transform as $(1/2,0)$ and $(0,1/2)$, clearly more fundamental objects.

Electroweak-scale supersymmetry is motivated by several main points. You will hear further details in this lecture series from Abel (2001).

1. In any gauge theory beyond the standard $SU(3)_{\text{colour}} \otimes SU(2)_L$, if the scale Λ of new physics is very high, this scale tends to make its way into the Higgs sector through loop diagrams, leading to quadratic contributions $\sim g^2 \Lambda^2$ to the Higgs boson mass. Unless something cancels these contributions, one has to fine-tune counter-terms in the Lagrangian to exquisite accuracy, at each order of perturbation theory. This is known as the "hierarchy problem."

2. The very nature of a $\lambda(\phi^\dagger \phi)^2$ term in the Lagrangian is problematic when considered from the standpoint of scale changes. This is known as the "triviality problem."

3. In the simplest theory by Georgi and Glashow (1974) unifying $SU(3)_{\text{colour}} \otimes SU(2)_L$, based on the gauge group $SU(5)$, the coupling constants approach one another at high scale, but there is some "astigmatism." In a non-supersymmetric model, they do not all come together at the same scale. This is known as the "unification problem." It is cured in the simplest supersymmetric model, as a result of the different particle content in loop diagrams contributing to the running of the coupling constants. The model has a problem, however, in predicting too large a rate for $p \to K^+ \bar{\nu}$ (Murayama and Pierce 2001, Peskin 2001).

An alternative scheme for solving these problems, which has had a much poorer time constructing any sort of self-consistent theory, is *technicolour*, the notion that the Higgs boson is a bound state of more fundamental constituents in the same way that the pion is really a bound state of quarks. This mechanism works beautifully when applied to the generation of gauge boson masses, but fails spectacularly (and requires epicyclic patches!) when one attempts to describe fermion masses. The basic idea of technicolour is that there is no hierarchy problem because there is no hierarchy; a wealth of TeV-scale new physics awaits to be discovered in the simplest version (applied to gauge bosons) of the theory.

A further, even more radical notion, is that both Higgs bosons and fermions are composite. This scheme so far has run aground on the difficulty of constructing quarks and leptons, keeping their masses light by nearly preserving a chiral symmetry ('t Hooft 1980). One can make guesses as to quantum numbers of constituents (Rosner and Soper 1992), but a sensible dynamics remains completely elusive.

5.5   Fermion masses

We finessed the question of the origin of the Cabibbo-Kobayashi-Maskawa (CKM) matrix. It comes about in the following way.

The electroweak Lagrangian, before electroweak symmetry breaking, may be written

in flavour-diagonal form as

$$\mathcal{L}_{\text{int}} = -\frac{g}{\sqrt{2}}[\overline{U'}_L \gamma^\mu W_\mu^{(+)} D'_L + \text{h.c.}], \tag{199}$$

where $U' \equiv (u', c', t')$ and $D' \equiv (d', s', b')$ are column vectors describing *weak eigenstates*. Here g is the weak $SU(2)_L$ coupling constant, and $\psi_L \equiv (1 - \gamma_5)\psi/2$ is the left-handed projection of the fermion field $\psi = U$ or D.

Quark mixing arise because mass terms in the Lagrangian are permitted to connect weak eigenstates with one another. Thus, the matrices $\mathcal{M}_{U, D}$ in

$$\mathcal{L}_m = -[\overline{U'}_R \mathcal{M}_U U'_L + \overline{D'}_R \mathcal{M}_D D'_L + \text{h.c.}] \tag{200}$$

may contain off-diagonal terms. One may diagonalise these matrices by separate unitary transformations on left-handed and right-handed quark fields:

$$R_Q^+ \mathcal{M}_Q L_Q = L_Q^+ \mathcal{M}_Q^+ R_Q = \Lambda_Q. \tag{201}$$

where

$$Q'_L = L_Q Q_L; \qquad Q'_R = R_Q Q_R \qquad (Q = U, D). \tag{202}$$

Using the relation between weak eigenstates and mass eigenstates: $U'_L = L_U U_L$ and $D'_L = L_D D_L$, we find

$$\mathcal{L}_{\text{int}} = -\frac{g}{\sqrt{2}}[\overline{U}_L \gamma^\mu W_\mu V D_L + \text{h.c.}], \tag{203}$$

where $U \equiv (u, c, t)$ and $D \equiv (d, s, b)$ are the mass eigenstates, and $V \equiv L_U^+ L_D$. The matrix V is just the Cabibbo-Kobayashi-Maskawa matrix. By construction, it is unitary: $V^+ V = V V^+ = 1$. It carries no information about R_U or R_D. More information would be forthcoming from interactions sensitive to right-handed quarks or from a genuine theory of quark masses.

Quark mass matrices can yield the observed hierarchy in CKM matrix elements. As an example (Rosenfeld and Rosner 2001), the regularities of quark masses evolved to a common high mass scale can be reproduced by the choice

$$\mathcal{M}_Q = m_3 \begin{pmatrix} 0 & \epsilon^3 e^{i\phi} & 0 \\ \epsilon^3 e^{-i\phi} & \epsilon^2 & \epsilon^2 \\ 0 & \epsilon^2 & 1 \end{pmatrix}, \tag{204}$$

where m_3 denotes the mass eigenvalue of the third-family quark (t or b), and $\epsilon \simeq 0.07$ for u quarks, $\simeq 0.21$ for d quarks. Hierarchical descriptions of this type were first introduced by Froggatt and Nielsen (1979). The present ansatz is closely related to one described by Fritzsch and Xing (1995). This type of mass matrix leads to acceptable values and phases of CKM elements.

The question of neutrino masses and mixings has entered a whole new phase with spectacular results from neutrino observatories such as super-Kamiokande ("Super-K") in Japan and the Sudbury Neutrino Observatory (SNO) in Canada. These indicate that:

1. Atmospheric muon neutrinos oscillate in vacuum, probably to τ neutrinos, with near-maximal mixing and a difference in squared mass $\Delta m^2 \simeq 3 \times 10^{-3}$ eV^2.

2. Solar electron neutrinos oscillate, most likely in matter, to some combination of muon and τ neutrinos. All possible Δm^2 values are at most about 10^{-4} eV^2; several ranges of parameters are permitted, with large mixing favoured by some analyses.

In addition one experiment, the Liquid Scintillator Neutrino Detector (LSND) at Los Alamos National Laboratory, suggests $\bar{\nu}_\mu \rightarrow \bar{\nu}_e$ oscillations with $\Delta m^2 \simeq 0.1$ to 1 eV, with small mixing. This possibility is difficult to reconcile with the previous two, and a forthcoming experiment at Fermilab (Mini-BooNE) is scheduled to check the result. For late news on neutrinos see the Web page maintained by Goodman (2001).

A possible explanation of small neutrino masses (Gell-Mann, Ramond, and Slansky 1979, Yanagida 1979) is that they are Majorana masses of order $m_M = m_D^2/M_M$, where m_D is a typical Dirac mass and M_M is a large Majorana mass acquired by right-handed neutrinos. Such a mass term is invariant under SU(2)$_L$, and hence is completely acceptable in the electroweak theory. The pattern of neutrino Majorana and Dirac masses, and the mixing pattern, is likely to provide us with fascinating clues over the coming years as to the fundamental origin and nature of mass.

6   Summary

The Standard Model of electroweak and strong interactions has been in place for nearly thirty years, but precise tests have entered a phase that permits glimpses of physics beyond this impressive structure, most likely associated with the yet-to-be discovered Higgs boson. Studies of mixing between neutral kaons or neutral B mesons, covered by Stone (2001) in these lectures, are attaining impressive accuracy as well, and could yield cracks in the Standard Model at any time. It is time to ask what lies behind the pattern of fermion masses and mixings. This is an *input* to the Standard Model, characterised by many free parameters all of which await explanation.

Acknowledgments

References

Abachi S *et al.* (D0 Collaboration), 1995, *Physical Review Letters* **74** 2632.
Abe F (CDF Collaboration), 1994, *Physical Review Letters* **73** 225; 1995, *Physical Review* D **50** 2966.
Abe K *et al.* (Belle Collaboration), 2001, *Physical Review Letters* **87** 091802.
Abel S, 2001, lectures at Scottish Universities' Summer School in Physics (these Proceedings).
Abers E S and Lee B W, 1973, *Physics Reports* **9C** 1.
Amaldi U *et al.*, 1987, *Physical Review* D **36** 1385.

Appelquist T and Politzer H D, 1975, *Physical Review Letters* **34** 43.

Aubert B *et al.* (BaBar Collaboration), 2001a, *Physical Review Letters* **87** 091801; SLAC Report No. SLAC-PUB-9060, hep-ex/0201020, submitted to Phys. Rev. D.

Aubert B *et al.* (BaBar Collaboration), 2001b, SLAC-PUB-9012, hep-ex/0110062, submitted to Phys. Rev. D.

Aubert J J *et al.*, 1974, *Physical Review Letters* **33** 1404.

Augustin J J *et al.*, 1974, *Physical Review Letters* **33** 1406.

Bai J Z *et al.* (BES Collaboration), preprint hep-ex/0102003 v2, 31 May 2001.

Barate R *et al.* (ALEPH Collaboration), 2001, *Physics Letters* B **495** 1.

Bardeen W A, Buras A J, Duke D W, and Muta, T, 1978, *Physical Review* D **18** 3998.

Benvenuti A *et al.*, 1974, *Physical Review Letters* **32** 800.

Bjorken J D and Glashow S L, 1964, *Physics Letters* **11** 255.

Bouchiat C, Iliopoulos, J, and Meyer Ph, 1972, *Physics Letters* **38B** 519.

Brown H B *et al.* (Brookhaven E821 Collaboration), 2001, *Physical Review Letters* **86** 2227.

Buchalla G, 2001, lectures at Scottish Universities' Summer School in Physics (these Proceedings).

Cabibbo N, 1963, *Physical Review Letters* **10** 531.

Carena M, Conway J S, Haber H E, Hobbs J D *et al.*, 2000, "Report of the Higgs Working Group of the Tevatron Run 2 SUSY/Higgs Workshop," hep-ph/0010338.

Caswell W E, 1974, *Physical Review Letters* **33** 244.

Cazzoli E G *et al.*, 1975, *Physical Review Letters* **34** 1125.

Charlton D, 2001, plenary talk at International Europhysics Conference on High Energy Physics, Budapest, Hungary, July 12-18, Univ. of Birmingham report BHAM-HEP/01-02, hep-ex/0110086, to be published by JHEP.

Chivukula S in *The Albuquerque Meeting* (Proceedings of the 8th Meeting, Division of Particles and Fields, American Physical Society, Albuquerque, NM, August 2-6, 1994) editor Seidel S, 1995 (World Scientific), p 273.

Christenson J H, Cronin J W, Fitch V L, and Turlay R E, 1964, *Physical Review Letters* **13** 138.

Coleman S, 1971, in *Properties of the Fundamental Interactions* (1971 Erice Lectures), editor Zichichi, A (Editrice Compositori, Bologna), p 358.

Collins H, Grant A K, and Georgi H, 2000, *Physical Review* D **61** 055002.

Conrad J C, Shaevitz M H, and Bolton T, 1998 *Reviews of Modern Physics* **70** 1341.

Dalitz R H, 1967, *Proceedings of the XIII International Conference on High Energy Physics* (Berkeley, CA: Univ. of Calif. Press), p 215.

Davier M and Höcker A, 1998, *Physics Letters* B **419** 419; *Physics Letters* B **435** 427.

De Rújula A, Georgi H, and Glashow S L, 1975, *Physical Review* D **12** 147.

Dobrescu B and Hill C T, 1998, *Physical Review Letters* **81** 2634.

Drees J, XX International Symposium on Lepton and Photon Interactions at High Energies, Rome, Italy, July 23-27, 2001, hep-ex/0110077.

Dzuba V A, Flambaum V V, and Ginges J S M, hep-ph/0111019, submitted to Phys. Rev. A.

Eichten E *et al.*, 1984, *Reviews of Modern Physics* **56** 579; 1986, *Reviews of Modern Physics* **58** 1065(E).

Ellis J, 1977, in *Weak and Electromagnetic Interactions at High Energies* (1976 Les Houches Lectures), editors Balian R and Llewellyn Smith C H (North-Holland), p 1.

Forshaw J R, Ross D A, and White B E, University of Manchester report MC-TH-01/07, hep-ph/0107232.

Fritzsch H and Xing Z Z, 1995, *Physics Letters* B **353** 114; Xing Z Z, 1997, *Journal of Physics* G **23**, 1563.

Froggatt C D and Nielsen H B, 1979, *Nuclear Physics* **B147** 277.

Gaillard M and Lee B W, 1974, *Physical Review* D **10** 894.

Gell-Mann M, 1964, *Physics Letters* **8** 214.

Gell-Mann M and Lévy M, 1960, *Nuovo Cimento* **19** 705.

Gell-Mann M and Ne'eman Y, 1964, *The Eightfold Way* (Benjamin).

Gell-Mann M, Ramond P, and Slansky R, in *Supergravity*, editors van Nieuwenhuizen P and Freedman D Z, 1979 (North-Holland), p. 315. See also Gell-Mann M, Slansky R, and Stephenson G, 1979 (unpublished).

Georgi H and Glashow S L, 1972a, *Physical Review Letters* **28** 1494.

Georgi H and Glashow S L, 1972b, *Physical Review* D **6** 429.

Georgi H and Glashow S L, 1974, *Physical Review Letters* **32** 438.

Gilman F, Kleinknecht K, and Renk Z, 2000, in Particle Data Group (Groom D E *et al.*), 2000, *European Journal of Physics* C **15**, 1, pp 110-114.

Glashow S, 1961, *Nuclear Physics* **22** 579.

Glashow S L, Ilipoulos J, and Maiani L, 1970, *Physical Review* D **2** 1285.

Glashow S L and Weinberg S, 1977, *Physical Review* D **15** 1958.

Goldhaber G *et al.*, 1976, *Physical Review Letters* **37** 255.

Goodman M, 2001, http://www.hep.anl.gov/ndk/hypertext/nu_industry.html.

Gronau M and Rosner J L, 2001, Technion report TECHNION-PH-2001-35 and Enrico Fermi Institute Report No. 01-42, hep-ph/0109238, submitted to *Physical Review D*.

Gronau M and Rosner J L, 2002, *Physical Review* D **65** 013004.

Gross D J and Jackiw R, 1972, *Physical Review* D **6** 477.

Gross D J and Wilczek F, 1973, *Physical Review* D **8** 3633; 1974 *Physical Review* D **9** 980.

Gunion J, Haber H E, Kane G, and Dawson S, 1990, *The Higgs Hunter's Guide* (Addison-Wesley).

Hara Y, 1964, *Physical Review* **134** B701.

Hasert F J *et al.*, 1973, *Physics Letters* **46B** 121,138; 1974, *Nuclear Physics* **B73** 1.

Hayakawa M and Kinoshita M, KEK Report No. KEK-TH-793, hep-ph/0112102.

He H J, Polonsky N, and Su S, *Physical Review* D **64** 053004 2001; He H-J, Hill C T, and Tait T M P, Univ. of Texas Report No. UTEXAS-HEP-01-013, hep-ph/0108041 (unpublished).

Higgs P W, 1964, *Physics Letters* **12** 132; *Physics Letters* **13** 508; see also Englert F and Brout R, 1964, *Physical Review Letters* **13** 321; Guralnik G S, Hagen C R, and Kibble T W B, 1964, *Physical Review Letters* **13** 585.

Höcker A, Lacker H, Laplace S, and Le Diberder F, 2001, *European Journal of Physics* C **21**, 225; see http://www.slac.stanford.edu/ laplace/ckmfitter.html for periodic updates.

't Hooft G, 1973, *Nuclear Physics* **B61** 455.

't Hooft G, 1980, in *Recent Developments in Gauge Theories* (Cargèse Summer Institute, Aug. 26 - Sept. 8, 1979) 't Hooft G *et al.* editors (Plenum, New York), p 135.

Hughes R J, 1980, *Physics Letters* **97B** 246: 1981, *Nuclear Physics* **B186** 376.

Kennedy D C and Langacker P G, 1990, *Physical Review Letters* **65** 2967; 1991, *Physical Review Letters* **66** 395(E).

Khriplovich I B, 1969, *Yadernaya Fizika* **10** 409 [*Soviet Journal of Nuclear Physics* **10** 235].

Kim Y K, 2001, XX International Symposium on Lepton and Photon Interactions at High Energies, Rome, Italy, July 23-27, 2001.

Klein O, 1938, in *Les Nouvelles Théories de la Physique*, Paris, Inst. de Coöperation Intellectuelle (1939), p 81, reprinted in *Oskar Klein Memorial Lectures* vol 1, editor Ekspong G, 1991 (World Scientific, Singapore).

Knecht M and Nyffeler A, 2001, Centre de Physique Théorique (Marseille) Report No. CPT-2001/P.4253, hep-ph/0111058.

Kobayashi M and Maskawa T, 1973, *Progress of Theoretical Physics* (Kyoto) **49** 652.

Kostelecký V A and Roberts A, 2001, *Physical Review* D **63** 096002.

Kwong W, 1991, *Physical Review* D **43** 1488.

Kwong W and Rosner J L, 1986, *Progress of Theoretical Physics (Suppl.)* **86** 366.

Langacker P, "Precision Electroweak Data: Phenomenological Analysis," talk at Snowmass 2001 Workshop, Univ. of Pennsylvania report UPR-0959T, hep-ph/0110129.

Langacker P, Luo M, and Mann A K, 1992, *Reviews of Modern Physics* **64** 87.

LEP Electroweak Working Group, 2001: see http://lepewwg.web.cern.ch/LEPEWWG for periodic updates.

LEP Higgs Working Group, LHWG Note/2001-03, hep-ex/0107029, paper contributed to EPS '01 Conference, Budapest, and XX International Symposium on Lepton and Photon Interactions at High Energies, Rome, Italy, July 23–27, 2001.

Lipkin H J, 1973, *Physics Reports* **8** 173, and references therein.

Llewellyn Smith C H, 1983, *Nuclear Physics* **B228** 205.

Luo Z and Rosner J L, Enrico Fermi Institute Report No. 01-28, hep-ph/0108024, to be published in *Physical Review D*.

Maki Z and Ohnuki Y, 1964, *Progress of Theoretical Physics* (Kyoto) **32** 144.

Marciano W, Brookhaven National Laboratory report BNL-HET-00/04, hep-ph/0003181, published in *Proceedings of MuMu 99* (5th International Conference on Physics Potential and Development of $\mu^+\mu^-$ Colliders, San Francisco, CA, December 1999), editor Cline D, AIP Conference Proceedings v 542 (American Institute of Physics).

Milstein A I and Sushkov A P, 2001, Novosibirsk and Univ. of New South Wales preprint, hep-ph/0109257.

Murayama H and Pierce A, 2001, University of California at Berkeley preprint, hep-ph/0108104.

Nambu Y, 1966, in De-Shalit A, Feshbach H and Van Hove L eds *Preludes in Theoretical Physics in Honor of V. F. Weisskopf* (Amsterdam: North-Holland and New York: Wiley) pp 133-42.

Nambu Y, 1974, *Physical Review* D **10** 4262.

Nir Y, 2001, lectures at Scottish Universities' Summer School in Physics (these Proceedings).

Niu K, Mikumo E, and Maeda Y, 1971, *Progress of Theoretical Physics* (Kyoto) **46** 1644.

Paschos E A and Wolfenstein L, 1973, *Physical Review* D **7** 91.

Particle Data Group (Groom D E *et al.*), 2000, *European Journal of Physics* C **15**, 1. All experimental values are quoted from this source unless otherwise noted.

Pati J C and Salam A, 1973, *Physical Review Letters* **31** 661; 1974, *Physical Review* D **10** 275; see also Mohapatra R N and Pati J C, 1975, *Physical Review* D **11** 566, 2558.

Perl M L *et al.* (SLAC-LBL Collaboration), 1975, *Physical Review Letters* **35** 1489; 1976, *Physics Letters* **63B** 466.

Peruzzi I *et al.*, 1976, *Physical Review Letters* **37** 569.

Peskin M E, 2001, "Interpretation of Precision Electroweak Data, or Should We Really Believe There is a Light Higgs Boson?", seminar at Snowmass 2001 Workshop, July, 2001.

Peskin M E and Schroeder D V, 1995, *An Introduction to Quantum Field Theory* (Addison-Wesley).

Peskin M E and Takeuchi T, 1990, *Physical Review Letters* **65** 964; 1992, *Physical Review* D **46** 381.

Peskin M E and Wells J, 2001, *Physical Review* D **64** 093003.

Politzer H D, 1974, *Physics Reports* **14C** 129.

Quigg C, 1983, *Gauge Theories of the Strong, Weak, and Electromagnetic Interactions* (Benjamin–Cummings).

Rosenfeld R and Rosner J L, 2001, *Physics Letters* B **516** 408.

Rosner J, 1988, An Introduction to Standard Model Physics, in *The Santa Fe TASI-87* (Proceedings of the 1987 Theoretical Advanced Study Institute, St. John's College, Santa Fe, NM, July 6-10, 1987), editors Slansky R and West G (World Scientific), p 3.

Rosner J, 1997, Top Quark Mass, in *Masses of Fundamental Particles – Cargése 1996*, editors Lévy M *et al.* (Plenum), p 43.

Rosner J L, 1999, *Physical Review* D **61** 016006.

Rosner J L, 2001, Enrico Fermi Institute Report No. 01-43, hep-ph/0109239, to be published in *Physical Review D.*

Rosner J L and Slezak S, 2001, *American Journal of Physics* **69** 44.

Rosner J and Soper D E, 1992, *Physical Review* D **45** 3206.

Ross D A, 1978, *Acta Physica Polonica* **B10** 189.

Sakurai J J, 1967, *Advanced Quantum Mechanics* (Addison-Wesley), pp 282 ff.

Salam A, 1968, *Proceedings of the Eighth Nobel Symposium*, editor Svartholm N (Almqvist and Wiksell, Stockholm; Wiley, New York, 1978), p 367.

Schubert K, 2001, lectures at Scottish Universities' Summer School in Physics (these Proceedings).

Sikivie P, Susskind L, Voloshin M B, and Zakharov V, 1980, *Nuclear Physics* **B173** 189.

Slansky R, 1981, *Physics Reports* **79** 1.

Stone S, 2001, lectures at Scottish Universities' Summer School in Physics (these Proceedings).

Strassler M J and Peskin M E, 1991, *Physical Review* D **43** 1500.

Swartz M, lecture at Snowmass 2001 Workshop, see http://pha.jhu.edu/˜morris/higgs.pdf for transparencies.

Veltman M, 1977, *Nuclear Physics* **B123** 89; *Acta Physica Polonica* **B8** 475.

Weinberg S, 1967, *Physical Review Letters* **19** 1264.

Willenbrock S, 2000, *Reviews of Modern Physics* **72** 1141.

Wolfenstein L, 1983, *Physical Review Letters* **51** 1945.

Yanagida T, 1979, in *Proceedings of the Workshop on Unified Theory and Baryon Number in the Universe*, editors Sawada O and Sugamoto A, 1979, (Tsukuba, Japan, National Laboratory for High Energy Physics), p 95.

Yang C N, 1974, *Physical Review Letters* **33** 445.

Yang C N and Mills R L, 1954, *Physical Review* **96** 191.

Yukawa H, 1935, *Proceedings of the Physical and Mathematical Society of Japan* 17:48.

Zel'dovich Ya B and Sakharov A D, 1966, *Yadernaya Fizika* **4** 395 [1967 *Soviet Journal of Nuclear Physics* **4** 283].

Zeller G P *et al.* (NuTeV Collaboration), presented at 1999 Division of Particles and Fields Meeting, UCLA, 5–9 January 1999, hep-ex/9906024.

Zweig G, 1964, CERN Reports 8182/TH 401 and 8419/TH 412 (unpublished), second paper reprinted in *Developments in the Quark Theory of Hadrons*, 1980, editors Lichtenberg D B and Rosen S P (Hadronic Press), v 1, pp 22–101.

Heavy quark theory

Gerhard Buchalla

Theory Division, CERN, Switzerland

1   Introduction

The dedicated study of b-flavoured hadrons has developed into one of the most active and promising areas of experimental high-energy physics. The detailed investigation of b decays at the B-meson factories and at hadron colliders will probe the flavour physics of quarks with unprecedented precision. To fully exploit this rich source of data a systematic theoretical approach is necessary. The required field theoretical tools are the subject of these lectures.

We shall discuss the construction of effective weak Hamiltonians, introducing the operator product expansion (OPE) and the renormalisation group (RG) and presenting the effective Hamiltonians for nonleptonic $\Delta B = 1$ and $\Delta B = 2$ transitions as examples. The subsequent chapter is devoted to heavy-quark effective theory (HQET). It explains the basic formalism as well as the application to heavy-light and heavy-heavy currents, discussing the B-meson decay constant f_B and the exclusive semileptonic decay $B \to D^* l \nu$, respectively. Inclusive b decays and the heavy-quark expansion (HQE) are treated next, in particular the general formalism, the issue of quark-hadron duality, the theory of b-hadron lifetimes and of inclusive semileptonic decays. Finally, we discuss QCD factorisation for exclusive hadronic B decays, focusing on $B \to D\pi$ and $B \to \pi\pi$. We conclude with a short summary.

The effective-Hamiltonian framework is the oldest and most general of the methods we shall discuss. It dates back, more or less, to the beginnings of the standard model itself. HQET and HQE are later developments that have been established in the second half of the eighties and at the beginning of the nineties. The last topic, QCD factorisation for exclusive B decays is the most recent. It is the least well established among these methods and it continues to be studied and developed in further detail.

We would like to mention a very short selection of literature, which we hope will be helpful to obtain further information on the various subjects related to the contents of these lectures. Very useful resources are the BaBar Physics Book (Harrison & Quinn 1998) and the Fermilab B Physics Report (Anikeev et al. 2002). They collect nice reviews on both theoretical and experimental topics in B physics. A textbook more specifically directed towards theoretical heavy quark physics is the work of Manohar & Wise (2000). Review articles on particular subjects are (Buchalla, Buras & Lautenbacher 1996) on effective weak Hamiltonians, (Neubert 1994) on HQET and (Bigi, Shifman and Uraltsev 1997) on HQE.

2 Introduction and overview

2.1 Motivation

In the following chapters we will study the theoretical tools to compute weak decay properties of heavy hadrons. To put the formalism into perspective, we start by recalling the main motivation for this subject.

The central goal is the investigation of flavour physics, the most complicated sector in our understanding of fundamental interactions. A good example is give by particle-antiparticle mixing, as first studied with neutral kaons. The strong interaction eigenstates K^0 ($\bar{s}d$) and $\bar{K}^0$ ($\bar{d}s$) can be transformed into each other through second order weak interactions, which leads to a tiny off-diagonal entry M_{12} in the mass matrix (Figure 1).
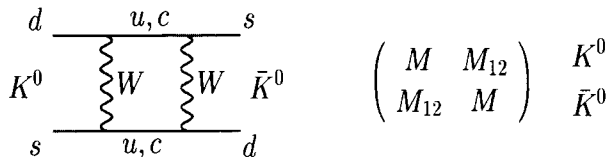
$$K_{L,S} = \frac{K^0 \pm \bar{K}^0}{\sqrt{2}}$$

Figure 1. K^0 – $\bar{K}^0$ *mixing.*

The physical eigenstates are $K_{L,S} = (K^0 \pm \bar{K}^0)/\sqrt{2}$ with a mass difference $\Delta m_K = m_L - m_S$ given by $2|M_{12}|$ so that

$$\frac{\Delta m_K}{m_K} \approx \frac{G_F^2 f_K^2 B_K}{6\pi^2} |V_{cs}V_{cd}|^2 m_c^2 = 7 \cdot 10^{-15}, \tag{1}$$

where the number on the right-hand side is the experimental value. The theoretical expression is derived neglecting the third generation of quarks and CP violation, which is a valid approximation for Δm_K. The factors $f_K^2 B_K$ ($B_K \approx 1$ for the purpose of a first estimate) account for the binding of the quarks into mesons. A crucial feature of (1) is the Glashow-Iliopoulos-Maiani (GIM) cancellation mechanism: the orthogonality of the quark mixing matrix V_{ij} ($i=u,c$; $j=d,s$) leads to a cancellation among the various contributions with intermediate up and charm quarks (symbolically the amplitude has the form $(uu) - (uc) - (cu) + (cc)$). For $m_u = m_c$ this cancellation would be complete, giving $\Delta m_K = 0$. Still, even for $m_u \neq m_c$, the contributions from virtual momenta $k \gg m_c, m_u$

cancel since both m_u and m_c are negligible in this case. What is left is a characteristic effect proportional to m_c^2, the up-quark contribution being subleading for $m_u \ll \Lambda_{\mathrm{QCD}} \ll m_c$. This circumstance allowed Gaillard and Lee in 1974 to correctly estimate the charm-quark mass $m_c \approx 1.5$ GeV, before charm was eventually discovered in the Fall of the same year.

In a similar way, the discovery of $B_d - \bar{B}_d$ mixing by the ARGUS collaboration (Albrecht et al. 1987) proved to be another milestone in flavour physics. In full analogy to $K - \bar{K}$ mixing we have

$$\frac{\Delta m_B}{m_B} \approx \frac{G_F^2 f_B^2 B_B}{6\pi^2} |V_{tb}V_{td}|^2 M_W^2 S\left(\frac{m_t^2}{M_W^2}\right) = 6 \cdot 10^{-14}, \tag{2}$$

where now the top-quark contribution is completely dominant. The unexpectedly large value observed by ARGUS provided the first evidence that the top-quark mass ($m_{t,\mathrm{pole}} = 176$ GeV) was comparable to the weak scale and in any case much heavier than anticipated at the time.

These examples show very nicely the "flavour" of flavour physics: precision observables are sensitive probes of high-energy scales and yield crucial insights into the fundamental structure of weak interactions. At the same time we see that hadronic effects manifest in quantities such as f_B, B_B, and strong interactions of the participating quarks in general, play an important role. Their understanding is necessary to reveal the underlying flavour dynamics and is the main subject of heavy quark theory.

$B - \bar{B}$ mixing, CP violation in B decays and other loop-induced reactions of b-flavoured hadrons are of great interest and continue to be pursued by numerous experiments. A central target is the Cabibbo-Kobayashi-Maskawa (CKM) matrix

$$V = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \simeq \begin{pmatrix} 1 - \frac{\lambda^2}{2} & \lambda & A\lambda^3(\varrho - i\eta) \\ -\lambda & 1 - \frac{\lambda^2}{2} & A\lambda^2 \\ A\lambda^3(1 - \varrho - i\eta) & -A\lambda^2 & 1 \end{pmatrix} \tag{3}$$

which parametrises charged-current weak interactions in the standard model. The second equality in (3) introduces the approximate form of the Wolfenstein parametrisation, where the four independent CKM quantities are λ, A, ϱ and η. The unitarity triangle, as defined in terms of ϱ and η is shown in Figure 2, indicating the CP violating angles α, β and γ.
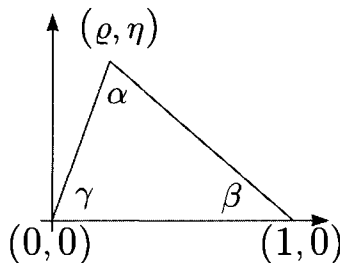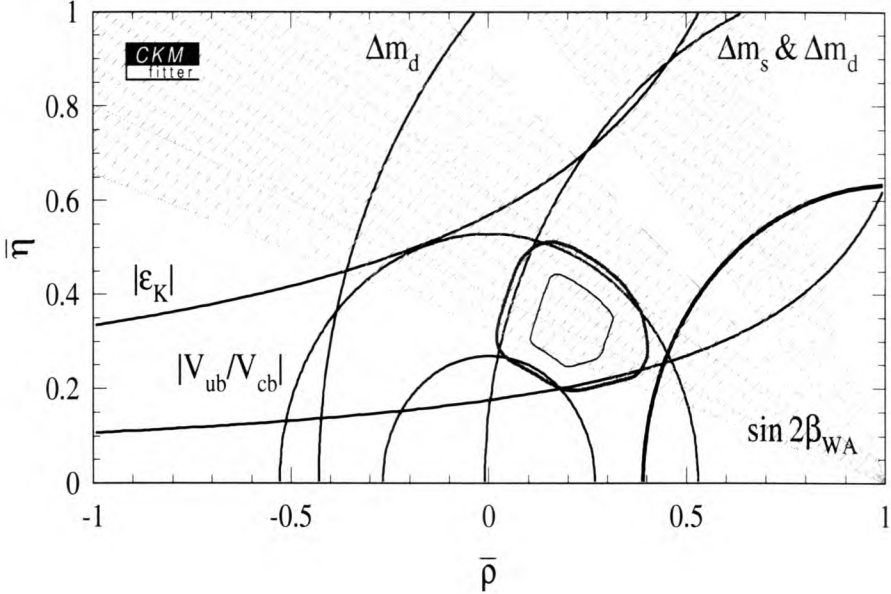

Figure 2. *Definition of the unitarity triangle.*

Figure 3. *Global fit of the unitarity triangle (darker shaded region), without* $\sin 2\beta$ *from CP violation in* $B \rightarrow J/\Psi K_S$. *The constraint from the world average* $\sin 2\beta$ *(hatched areas) is overlaid (Höcker et al. 2001).*

The status of the unitarity triangle (in terms of $\bar{\varrho} = \varrho(1 - \lambda^2/2)$ and $\bar{\eta} = \eta(1 - \lambda^2/2)$) is displayed in Figure 3, with input from CP violation in the kaon sector (ε_K) and from B decays ($|V_{ub}/V_{cb}|$, Δm_{B_d}, Δm_{B_s} and $\sin 2\beta$).

2.2 B decays – overview

A vast number of different B-decay observables is available to further test the standard scenario of flavour dynamics. One may distinguish the following broad classes.

- Dominant decays:
$$b \rightarrow c\bar{u}d, \quad b \rightarrow c\bar{c}s, \quad b \rightarrow cl\bar{\nu}$$
$$\bar{B} \rightarrow D\pi, \quad \bar{B} \rightarrow \Psi K, \quad \bar{B} \rightarrow D^{(*)}l\bar{\nu}$$

- Rare decays:
$$b \rightarrow u\bar{u}d, \quad b \rightarrow u\bar{u}s, \quad b \rightarrow ul\bar{\nu}$$
$$\bar{B} \rightarrow \pi\pi, \quad \bar{B} \rightarrow \pi K, \quad \bar{B} \rightarrow \pi l\bar{\nu}, \quad \bar{B} \rightarrow l\bar{\nu}$$

- Rare and radiative (loop induced) decays:
$$b \rightarrow s(d)\gamma, \quad b \rightarrow s(d)l^+l^-, \quad b \rightarrow s(d)\nu\bar{\nu}$$
$$\bar{B} \rightarrow K^{(*)}\gamma, \quad \bar{B} \rightarrow \rho\gamma, \ldots$$

- $\Delta B = 2$ oscillations:
$$B_d - \bar{B}_d, \quad B_s - \bar{B}_s \text{ mixing}$$

Other obvious classifications are between inclusive and exclusive processes or between hadronic and (semi)leptonic decays.

A few key properties of *b*-hadrons enhance considerably the possibilities in *B* physics both experimentally and theoretically. First, the smallness of $V_{cb} = 0.04$ leads to a long lifetime of $\tau_B \approx 1.6$ ps. In addition the *b*-quark mass is large compared with the QCD scale

$$m_b \gg \Lambda_{\text{QCD}} \approx 0.3 \text{GeV}. \tag{4}$$

The exact value of m_b depends on the definition. In particular, the running $\overline{MS}$ mass $\bar{m}_b(\bar{m}_b) = 4.2 \pm 0.1$GeV, the pole mass $m_{b,pole} \approx 4.8$GeV, whereas the mass of the lightest *b*-hadron is $m_B = 5.28$GeV. The smallness of Λ_{QCD}/m_b provides us with a useful expansion parameter. Together with the property of asymptotic freedom of QCD and $\alpha_s(m_b) \ll 1$, this opens the possibility of systematic approximations, which are exploited in the various applications of heavy quark theory.

3 Effective weak Hamiltonians

Computing weak decays of hadrons is a complicated problem in quantum field theory. Two typical cases, the first-order nonleptonic process $\bar{B}^0 \to \pi^+\pi^-$, and the loop-induced, second-order weak transition $B^- \to K^-\nu\bar{\nu}$ are illustrated in Figure 4. The dynamics
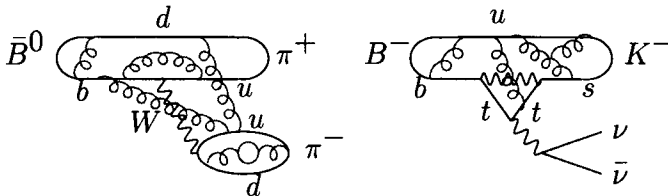

Figure 4. *QCD effects in weak decays.*

of the decays is determined by a nontrivial interplay of strong and electroweak forces, which is characterised by several energy scales of very different magnitude, the *W* mass, the various quark masses and the QCD scale: $m_t, M_W \gg m_b, m_c \gg \Lambda_{\text{QCD}} \gg m_u, m_d, (m_s)$. While it is usually sufficient to treat electroweak interactions to lowest non-vanishing order in perturbation theory, it is necessary to consider all orders in QCD. Asymptotic freedom still allows us to compute the effect of strong interactions at short distances perturbatively. However, since the participating hadrons are bound states with light quarks, confined inside the hadron by long-distance dynamics, it is clear that also non-perturbative QCD interactions enter the decay process in an essential way.

To deal with this situation, we need a method to disentangle long- and short-distance contributions to the decay amplitude in a systematic fashion. A basic tool for this purpose is provided by the operator product expansion (OPE).

3.1   Operator product expansion

We will now discuss the basic concepts of the OPE for B meson decay amplitudes. These concepts are of crucial importance for the theory of weak decay processes, not only in the case of B mesons, but also for kaons, mesons with charm, light or heavy baryons and weakly decaying hadrons in general. Consider, for instance, the basic W-boson exchange process shown on the left-hand side of Figure 5. This diagram mediates the decay of
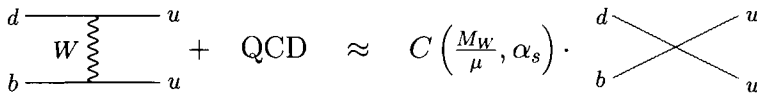
Figure 5. *OPE for weak decays.*

a b quark and triggers the nonleptonic decay of a B meson such as $\bar{B}^0 \rightarrow \pi^+\pi^-$. The quark-level transition shown is understood to be dressed with QCD interactions of all kinds, including the binding of the quarks into the mesons. To simplify this problem, we may look for a suitable expansion parameter, as we are used to do in theoretical physics. Here, a key feature is provided by the fact that the W mass M_W is very much heavier than the other momentum scales p in the problem (m_b, $\Lambda_{\rm QCD}$, m_u, m_d, m_s). We can therefore expand the full amplitude A, schematically, as follows

$$A = C\left(\frac{M_W}{\mu}, \alpha_s\right) \cdot \langle Q \rangle + \mathcal{O}\left(\frac{p^2}{M_W^2}\right),\tag{5}$$

which is sketched in Figure 5. Up to negligible power corrections of $\mathcal{O}(p^2/M_W^2)$, the full amplitude on the left-hand side is written as the matrix element of a local four-quark operator Q, multiplied by a Wilson coefficient C. This expansion in $1/M_W$ is called a (short-distance) operator product expansion because the nonlocal product of two bilinear quark-current operators $(\bar{d}u)$ and $(\bar{u}b)$ that interact via W exchange, is expanded into a series of local operators. Physically, the expansion in Figure 5 means that the exchange of the very heavy W boson can be approximated by a point-like four-quark interaction. With this picture the formal terminology of the OPE can be expressed in a more intuitive language by interpreting the local four-quark operator as a four-quark interaction vertex and the Wilson coefficient as the corresponding coupling constant. Together they define an effective Hamiltonian $\mathcal{H}_{\rm eff} = C \cdot Q$, describing weak interactions of light quarks at low energies. Ignoring QCD, the OPE reads explicitly (in momentum space)

$$\begin{aligned}A &= \frac{g_W^2}{8} V_{ud}^* V_{ub} \frac{i}{k^2 - M_W^2} (\bar{d}u)_{V-A}(\bar{u}b)_{V-A}\\ &= -i\frac{G_F}{\sqrt{2}} V_{ud}^* V_{ub} C \cdot \langle Q \rangle + \mathcal{O}\left(\frac{k^2}{M_W^2}\right),\end{aligned}\tag{6}$$

with $C = 1$, $Q = (\bar{d}u)_{V-A}(\bar{u}b)_{V-A}$ and

$$\mathcal{H}_{\rm eff} = \frac{G_F}{\sqrt{2}} V_{ud}^* V_{ub}(\bar{d}u)_{V-A}(\bar{u}b)_{V-A}.\tag{7}$$

As we will demonstrate in more detail below after including QCD effects, the most important property of the OPE in (5) is the *factorisation* of long- and short-distance

contributions: all effects of QCD interactions above some factorisation scale μ (short distances) are contained in the Wilson coefficient C. All the low-energy contributions below μ (long distances) are collected into the matrix elements of local operators $\langle Q \rangle$. In this way the short-distance part of the amplitude can be systematically extracted and calculated in perturbation theory. The problem to evaluate the matrix elements of local operators between hadron states remains. This task requires in general nonperturbative techniques, as for example lattice QCD or QCD sum rules, but it is considerably simpler than the original problem of the full standard-model amplitude. In some cases also the approximate flavour symmetries of QCD (isospin, $SU(3)$) can help to determine the non-perturbative input. This is true in general for hadronic weak decays. A decisive advantage of *heavy* hadrons is the fact that the heavy-quark mass itself is still large in comparison to Λ_{QCD}. The limit $\Lambda_{\mathrm{QCD}}/m_b \ll 1$ can then be exploited, which is achieved, depending on the application, by using heavy quark effective theory, heavy quark expansion or QCD factorisation for exclusive nonleptonic decays.

The short-distance OPE that we have described, the resulting effective Hamiltonian, and the factorisation property are fundamental for the theory of B decays. However, the concept of factorisation of long- and short-distance contributions reaches far beyond these applications. In fact, the idea of factorisation, in various forms and generalisations, is the key to essentially all applications of perturbative QCD, including the important areas of deep-inelastic scattering and jet or lepton pair production in hadron-hadron collisions. The reason is the same in all cases: Perturbative QCD is a theory of quarks and gluons, but those never appear in isolation and are always bound inside hadrons. Nonperturbative dynamics is therefore always relevant to some extent in hadronic reactions, even if these occur at very high energy or with a large intrinsic mass scale. Thus, before perturbation theory can be applied, nonperturbative input has to be isolated in a systematic way, and this is achieved by establishing the property of factorisation. It turns out that the weak effective Hamiltonian for nonleptonic B decays provides a nice example to demonstrate the general idea of factorisation in simple and explicit terms.

We would therefore like to discuss the OPE for B decays in more detail, including the effects of QCD, and illustrate the calculation of the Wilson coefficients. A diagrammatic representation for the OPE is shown in Figure 6. The key to calculating the coefficients

Figure 6. *Calculation of Wilson coefficients of the OPE.*

C_i is again the property of factorisation. Since factorisation implies the separation of all long-distance sensitive features of the amplitude into the matrix elements of $\langle Q_i \rangle$, the short-distance quantities C_i are, in particular, independent of the external states. This means that the C_i are always the same, no matter whether we consider the actual physical amplitude where the quarks are bound inside mesons, or any other, unphysical amplitude with on-shell or even off-shell external quark lines. Thus, even though we are ultimately interested in, e.g., $B \to \pi\pi$ amplitudes, for the perturbative evaluation of C_i we are free to choose any treatment of the external quarks according to our calculational convenience. A convenient choice that we will use below is to take all external quarks massless and

with the same off-shell momentum p $(p^2 \neq 0)$. The computation of the C_i in perturbation theory then proceeds in the following steps:

- Compute the amplitude A in the full theory (with W propagator) for arbitrary external states.

- Compute the matrix elements $\langle Q_i \rangle$ with the same treatment of external states.

- Extract the C_i from $A = C_i \langle Q_i \rangle$.

We remark that with the off-shell momenta p for the quark lines the amplitude is even gauge dependent and clearly unphysical. However, this dependence is identical for A and $\langle Q_i \rangle$ and drops out in the coefficients. The actual calculation is most easily performed in Feynman gauge. To $\mathcal{O}(\alpha_s)$ there are four relevant diagrams, the one shown in Figure 6 together with the remaining three possibilities to connect the two quark lines with a gluon. Gluon corrections to either of these quark currents need not be considered, they are the same on both sides of the OPE and drop out in the C_i. The operators that appear on the right-hand side follow from the actual calculations. Without QCD corrections there is only one operator of dimension 6
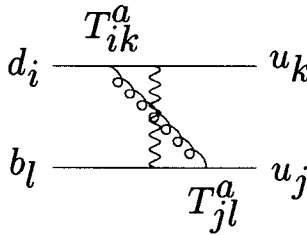
$$Q_1 = (\bar{d}_i u_i)_{V-A}(\bar{u}_j b_j)_{V-A}, \tag{8}$$

where the colour indices have been made explicit. To $\mathcal{O}(\alpha_s)$ QCD generates another operator

$$Q_2 = (\bar{d}_i u_j)_{V-A}(\bar{u}_j b_i)_{V-A}, \tag{9}$$

which has the same Dirac and flavour structure, but a different colour form. Its origin is illustrated in Figure 7, where we recall the useful identity for $SU(N)$ Gell-Mann matrices

$$(\bar{d}_i T^a_{ik} u_k)(\bar{u}_j T^a_{jl} b_l) = -\frac{1}{2N}(\bar{d}_i u_i)(\bar{u}_j b_j) + \frac{1}{2}(\bar{d}_i u_j)(\bar{u}_j b_i). \tag{10}$$

Figure 7. *QCD correction with colour assignment.*

It is convenient to employ a different operator basis, defining

$$Q_\pm = \frac{Q_1 \pm Q_2}{2}. \tag{11}$$

The corresponding coefficients are then given by

$$C_\pm = C_1 \pm C_2. \tag{12}$$

If we denote by $S_\pm$ the spinor expressions that correspond to the operators $Q_\pm$ (in other words: the tree-level matrix elements of $Q_\pm$), the full amplitude can be written as

$$A = \left(1 + \gamma_+ \alpha_s \ln \frac{M_W^2}{-p^2}\right) S_+ + \left(1 + \gamma_- \alpha_s \ln \frac{M_W^2}{-p^2}\right) S_-. \tag{13}$$

Here we have focused on the logarithmic terms and dropped a constant contribution (of order α_s, but nonlogarithmic). Further, p^2 is the virtuality of the quarks and $\gamma_\pm$ are numbers that we will specify later on. We next compute the matrix elements of the operators in the effective theory, using the same approximations, and find

$$\langle Q_\pm \rangle = \left(1 + \gamma_\pm \alpha_s \left(\frac{1}{\varepsilon} + \ln \frac{\mu^2}{-p^2}\right)\right) S_\pm. \tag{14}$$

The divergence that appears in this case has been regulated in dimensional regularisation ($D = 4 - 2\varepsilon$ dimensions). Requiring

$$A = C_+ \langle Q_+ \rangle + C_- \langle Q_- \rangle, \tag{15}$$

we obtain

$$C_\pm = 1 + \gamma_\pm \alpha_s \ln \frac{M_W^2}{\mu^2}, \tag{16}$$

where the divergence has been subtracted in the minimal subtraction scheme. The effective Hamiltonian we have been looking for then reads

$$\mathcal{H}_{\text{eff}} = \frac{G_F}{\sqrt{2}} V_{ud}^* V_{ub} \left(C_+(\mu) Q_+ + C_-(\mu) Q_-\right), \tag{17}$$

with the coefficients $C_\pm$ determined in (16) to $\mathcal{O}(\alpha_s \log)$ in perturbation theory. The following points are worth noting:

- The $1/\varepsilon$ (ultraviolet) divergence in the effective theory (14) reflects the $M_W \to \infty$ limit. This can be seen from the amplitude in the full theory (13), which is finite, but develops a logarithmic singularity in this limit. Consequently, the renormalisation in the effective theory is directly linked to the $\ln M_W$ dependence of the decay amplitude.

- We observe that although A and $\langle Q_\pm \rangle$ both depend on the long-distance properties of the external states (through p^2), this dependence has dropped out in $C_\pm$. Here we see explicitly how factorisation is realised. Technically, to $\mathcal{O}(\alpha_s \log)$, factorisation is equivalent to splitting the logarithm of the full amplitude according to

$$\ln \frac{M_W^2}{-p^2} = \ln \frac{M_W^2}{\mu^2} + \ln \frac{\mu^2}{-p^2}. \tag{18}$$

  Ultimately the logarithms stem from loop momentum integrations and the range of large momenta, between M_W and the factorisation scale μ, is indeed separated into the Wilson coefficients.

- To obtain a decay amplitude from $\mathcal{H}_{\text{eff}}$ in (17), the matrix elements $\langle f|Q_\pm|\bar{B}\rangle(\mu)$ have to be taken, normalised at a scale μ. An appropriate value for μ is close to the b-quark mass scale in order not to introduce an unnaturally large scale into the calculation of $\langle Q \rangle$.

- The factorisation scale μ is unphysical. It cancels between Wilson coefficient and hadronic matrix element, to a given order in α_s, to yield a scale independent decay amplitude. The mechanism of this cancellation to $\mathcal{O}(\alpha_s)$ is clear from the above example (13-16).

- In the construction of $\mathcal{H}_{\text{eff}}$ the W-boson is said to be "integrated out", that is, removed from the effective theory as an explicit degree of freedom. Its effect is still implicitly contained in the Wilson coefficients. The extraction of these coeffcients is often called a "matching calculation", matching the full to the effective theory by "adjusting" the couplings $C_\pm$.

- If we go beyond the leading logarithmic approximation $\mathcal{O}(\alpha_s \log)$ and include the finite corrections of $\mathcal{O}(\alpha_s)$ in (13), (14), an ambiguity arises when renormalising the divergence in (14) (or, equivalently, in the Wilson coefficients $C_\pm$). This ambiguity consists in what part of the full (non-logarithmic) $\mathcal{O}(\alpha_s)$ term is attributed to the matrix elements, and what part to the Wilson coefficients. In other words, coefficients and matrix elements become *scheme dependent*, that is, dependent on the renormalisation scheme, beyond the leading logarithmic approximation. The scheme dependence is unphysical and cancels in the product of coefficients and matrix elements. Of course, both quantities have to be evaluated in the same scheme to obtain a consistent result. The renormalisation scheme is determined in particular by the subtraction constants (minimal or non-minimal subtraction of $1/\varepsilon$ poles), and also by the definition of γ_5 used in $D \neq 4$ dimensions in the context of dimensional regularisation.

- Finally, the effective Hamiltonian (17) can be considered as a modern version of the old Fermi theory for weak interactions. It is a systematic low-energy approximation to the standard model for b-hadron decays and provides the basis for any further analysis.

3.2   Renormalisation group

Let us have a closer look at the Wilson coefficients, which read explicitly

$$C_\pm = 1 + \frac{\alpha_s(\mu)}{4\pi} \frac{\gamma_\pm^{(0)}}{2} \ln \frac{\mu^2}{M_W^2} \qquad \gamma_\pm^{(0)} = \left\{ \begin{array}{c} 4 \\ -8 \end{array} \right. , \tag{19}$$

where we have now specified the exact form of the $\mathcal{O}(\alpha_s \log)$ correction. Numerically the factor $\alpha_s(m_b)\gamma_\pm^{(0)}/(8\pi)$ is about $+3.5\%$ (-7%), a reasonable size for a perturbative correction (we used $\alpha_s(\mu = 4.2\,\text{GeV}) = 0.22$). However, this term comes with a large logarithmic factor of $\ln(\mu^2/M_W^2) = -6$, for an appropriate scale of $\mu = 4.2\,\text{GeV}$. The total correction to $C_\pm = 1$ in (19) is then -21% (42%)! The presence of the large logarithm spoils the validity of a straightforward perturbative expansion, despite the fact that the coupling constant itself is still reasonably small. This situation is quite common in renormalisable quantum field theories. Logarithms appear naturally and can become very large when the problem involves very different scales. The general situation is indicated in the following table, where we display the form of the correction terms in higher orders,

denoting $\ell \equiv \ln(\mu/M_W)$:

$$
\begin{array}{ccc}
\text{LL} & \text{NLL} & \\
\alpha_s \ell & \alpha_s & \\
\alpha_s^2 \ell^2 & \alpha_s^2 \ell & \alpha_s^2 \\
\alpha_s^3 \ell^3 & \alpha_s^3 \ell^2 & \alpha_s^3 \ell \quad \alpha_s^3 \\
\downarrow & \downarrow & \\
\mathcal{O}(1) & \mathcal{O}(\alpha_s) &
\end{array}
\tag{20}
$$

In ordinary perturbation theory the expansion is organised according to powers of α_s alone, corresponding to the rows in the above scheme. This approach is invalidated by the large logarithms since $\alpha_s \ell$, in contrast to α_s, is no longer a small parameter, but a quantity of order 1. The problem can be resolved by re-summing the terms $(\alpha_s \ell)^n$ to all orders n. The expansion is then reorganised in terms of columns of the above table. The first column is of $\mathcal{O}(1)$ and yields the leading logarithmic approximation, the second column gives a correction of relative order α_s, and so forth. Technically the reorganisation is achieved by solving the renormalisation group equation (RGE) for the Wilson coefficients. The RGE is a differential equation describing the change of $C_\pm(\mu)$ under a change of scale. To leading order this equation can be read off from (19)

$$
\frac{d}{d\ln\mu} C_\pm(\mu) = \frac{\alpha_s}{4\pi} \gamma_\pm^{(0)} \cdot C_\pm(\mu).
\tag{21}
$$

The factors $(\alpha_s/4\pi)\gamma_\pm^{(0)}$ are called the anomalous dimensions of $C_\pm$. To understand the term "dimension", compare with the following relation for the quantity μ^n, which has (energy) dimension n:

$$
\frac{d}{d\ln\mu}\mu^n = n \cdot \mu^n.
\tag{22}
$$

The analogy is obvious. Of course, the $C_\pm(\mu)$ are dimensionless numbers in the usual sense; they can depend on the energy scale μ only because there is another scale, M_W, present under the logarithm in (19). Their "dimension" is therefore more precisely called a scaling dimension, measuring the rate of change of $C_\pm$ with a changing scale μ. The nontrivial scaling dimension derives from $\mathcal{O}(\alpha_s)$ loop corrections and is thus a genuine quantum effect. Classically the coefficients are scale invariant, $C_\pm \equiv 1$. Whenever a symmetry that holds at the classical level is broken by quantum effects, we speak of an "anomaly". Hence, the $\gamma_\pm^{(0)}$ represent the anomalous (scaling) dimensions of the Wilson coefficients.

We can solve (21), using

$$
\frac{d\alpha_s}{d\ln\mu} = -2\beta_0 \frac{\alpha_s^2}{4\pi}, \qquad \beta_0 = \frac{33 - 2f}{3}, \qquad C_\pm(M_W) = 1,
\tag{23}
$$

and find

$$
C_\pm(\mu) = \left[\frac{\alpha_s(M_W)}{\alpha_s(\mu)}\right]^{\gamma_\pm^{(0)}/2\beta_0} = \left[1 + \beta_0 \frac{\alpha_s(\mu)}{4\pi} \ln \frac{M_W^2}{\mu^2}\right]^{-\gamma_\pm^{(0)}/2\beta_0}
\tag{24}
$$

This is the solution for the Wilson coefficients $C_\pm$ in leading logarithmic approximation, that is to leading order in RG improved perturbation theory. The all-orders re-summation of α_s log terms is apparent in the final expression in (24).

3.3   The $\Delta B = 1$ effective Hamiltonian

In this section we will complete the discussion of the $\Delta B = 1$ effective Hamiltonian. So far we have considered the operators

$$Q_1^p = (\bar{d}_i p_i)_{V-A}(\bar{p}_j b_j)_{V-A}, \tag{25}$$

$$Q_2^p = (\bar{d}_i p_j)_{V-A}(\bar{p}_j b_i)_{V-A}, \tag{26}$$

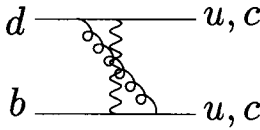which come from the simple W-exchange graph and the corresponding QCD corrections (Figure 8). We have slightly generalised our previous notation, allowing for the cases $p = u, c$.
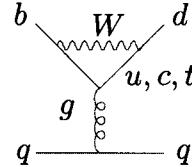

Figure 8. *Correction to W exchange.*   **Figure 9.** *QCD-penguin diagram.*

In addition, there is a further type of diagram at $\mathcal{O}(\alpha_s)$, which we have omitted until now: the QCD-penguin diagram shown in Figure 9. It gives rise to the four new operators

$$Q_3 = (\bar{d}_i b_i)_{V-A} \sum_q (\bar{q}_j q_j)_{V-A}, \tag{27}$$

$$Q_4 = (\bar{d}_i b_j)_{V-A} \sum_q (\bar{q}_j q_i)_{V-A}, \tag{28}$$

$$Q_5 = (\bar{d}_i b_i)_{V-A} \sum_q (\bar{q}_j q_j)_{V+A}, \tag{29}$$

$$Q_6 = (\bar{d}_i b_j)_{V-A} \sum_q (\bar{q}_j q_i)_{V+A}. \tag{30}$$

Two structures appear when the light-quark current $(\bar{q}q)_V$ from the bottom end of the diagram is split into $V - A$ and $V + A$ parts. In turn, each of those comes in two colour forms in a way similar to Q_1 and Q_2. Finally, one further gauge-invariant operator of dimension six appears in the matching procedure, the chromomagnetic penguin operator

$$Q_{8g} = -\frac{g}{8\pi^2} m_b \, \bar{d}_i \sigma^{\mu\nu} (1 + \gamma_5) T_{ij}^a b_j \, G_{\mu\nu}^a. \tag{31}$$

This operator corresponds to the diagrams in Figure 9 with the lower quark line omitted. The gluon is thus an external field, represented in (31) by the field-strength tensor $G_{\mu\nu}^a$. Note that the characteristic tensor current necessitates a helicity flip in the $b \to d$ transition, which is accompanied by a factor of the quark mass m_b (the effect of m_d is neglected). The contribution of Q_{8g} would be very small for the Hamiltonian of K decays, which only involves light external quarks, but it is unsuppressed for b decays.

The operators $Q_1, \ldots, Q_6, Q_{8g}$ mix under renormalisation, that is the RGE for their Wilson coefficients is governed by a matrix of anomalous dimensions, generalising (21). In this way the RG evolution of $C_{1,2}$ affects the evolution of $C_3, \ldots, C_6, C_{8g}$. On the other

hand $C_{1,2}$ remain unchanged in the presence of the penguin operators $Q_3, \ldots, Q_6, Q_{8g}$, so that the results for $C_{1,2}$ derived above are still valid.

The construction of the effective Hamiltonian follows the principles we have discussed in the previous sections. First the Wilson coefficients $C_i(\mu_W)$, $i = 1, \ldots, 6, 8g$, are determined at a large scale $\mu_W = \mathcal{O}(M_W, m_t)$ to a given order in perturbation theory. In this step both the W boson and the heavy top quark are integrated out. Since the renormalisation scale is chosen to be $\mu_W = \mathcal{O}(M_W, m_t)$, no large logarithms appear and straightforward perturbation theory can be used for the matching calculation. The anomalous dimensions are computed from the divergent parts of the operator matrix elements, which correspond to the UV-renormalisation of the Wilson coefficients. Solving the RGE the C_i are evolved from μ_W to a scale $\mu = \mathcal{O}(m_b)$ in a theory with $f = 5$ active flavours $q = u, d, s, c, b$. The terms taken into account in the RG improved perturbative evaluation of $C_i(\mu)$ are, schematically:

$$
\text{LO:} \quad \left(\alpha_s \ln \tfrac{M_W}{\mu} \right)^n, \qquad \qquad \text{NLO:} \quad \alpha_s \left(\alpha_s \ln \tfrac{M_W}{\mu} \right)^n,
$$

at leading and next-to-leading order, respectively.

The final result for the $\Delta B = 1$ effective Hamiltonian can be written as

$$
\mathcal{H}_{\text{eff}}^{\Delta B=1} = \frac{G_F}{\sqrt{2}} \sum_{p=u,c} \lambda_p \left[C_1 Q_1^p + C_2 Q_2^p + \sum_{i=3,\ldots,6,8g} C_i Q_i \right] + \text{h.c.}, \tag{32}
$$

where $\lambda_p \equiv V_{pd}^* V_{pb}$. In principle there are three different CKM factors, λ_u, λ_c and λ_t, corresponding to the different flavours of up-type quarks that can participate in the charged-current weak interaction. Using CKM unitarity, one of them can be eliminated. If we eliminate λ_t, we arrive at the CKM structure of (32).

The Hamiltonian in (32) is the basis for computing nonleptonic b-hadron decays within the standard model (to lowest order in electroweak interactions) with $\Delta B = 1$ and ΔS, $\Delta C = 0$. The Hamiltonian for b-decays with different flavour quantum numbers of the light quarks has a completely analogous form. For instance, $\Delta B = 1$ transitions with a simultaneous change in strangeness, $\Delta S = 1$, are simply described by (32) after the replacement $d \to s$. When new physics is present at some higher energy scale, the effective Hamiltonian can be derived in an analogous way. The matching calculation at the high scale μ_W will give new contributions to the coefficients $C_i(\mu_W)$, the initial conditions for the RG evolution. In general, new operators may also be induced. The Wilson coefficients C_i are known in the standard model at NLO. A more detailed account of $\mathcal{H}_{\text{eff}}^{\Delta B=1}$ and information on the technical aspects of the necessary calculations can be found in (Buchalla, Buras & Lautenbacher 1996) and (Buras 1998).

3.4 $B - \bar{B}$ mixing at NLO

In the following section we present the effective Hamiltonian for a $\Delta B = 2$ transition, which is relevant for $B - \bar{B}$ mixing. In this case only a single operator contributes. The form of the Hamiltonian is therefore particularly simple. We use this example to illustrate the structure of Wilson coefficients at next-to-leading order. The mass difference Δm_B in the $B - \bar{B}$ system is related to the effective Hamiltonian $\mathcal{H}_{\text{eff}}^{\Delta B=2}$ by

$$
\Delta m_B = 2|M_{12}|_B = \frac{1}{m_B} \left| \langle \bar{B} | \mathcal{H}_{\text{eff}}^{\Delta B=2} | B \rangle \right|. \tag{33}
$$

In order to construct $\mathcal{H}_{\text{eff}}^{\Delta B=2}$, the full standard model amplitude for $\Delta B = 2$ transitions is matched onto the effective theory amplitude at the matching scale $\mu_t = \mathcal{O}(m_t) = \mathcal{O}(M_W)$. This is sketched in Figure 10.
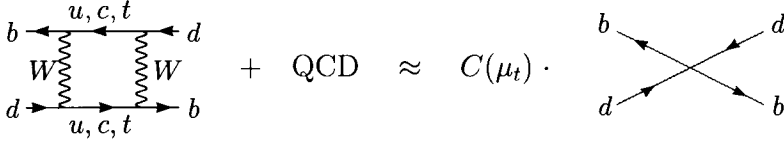
Figure 10. *OPE for* B – $\bar{B}$ *mixing.*

There is only one local operator

$$Q = (\bar{b}d)_{V-A}(\bar{b}d)_{V-A}. \tag{34}$$

The Wilson coefficient, up to next-to-leading order, can be written as

$$C(\mu_t) = C^{(0)}(\mu_t) + \frac{\alpha_s}{4\pi}C^{(1)}(\mu_t), \tag{35}$$

where $C^{(0)}$ is the lowest order result and $C^{(1)}$ comes from the corrections with one-gluon exchange. The RG evolution from the high scale μ_t down to a scale $\mu = \mathcal{O}(m_b)$ has the form

$$C(\mu) = \left[1 + \frac{\alpha_s(\mu) - \alpha_s(\mu_t)}{4\pi}J_5\right] \cdot \left[\frac{\alpha_s(\mu_t)}{\alpha_s(\mu)}\right]^{6/23} \cdot C(\mu_t). \tag{36}$$

The second factor on the right-hand side is familiar from the leading logarithmic approximation (the only difference is that at NLO the two-loop expression for $\alpha_s(\mu)$ has to be used). The first factor represents the next-to-leading order correction. Here J_5 is a scheme-dependent constant, which in the usual, so-called NDR, scheme reads $J_5 = 5165/3174$.

We now have the ingredients to write the effective Hamiltonian up to NLO precision

$$
\begin{aligned}
\mathcal{H}_{\text{eff}}^{\Delta B=2} &= \frac{G_F^2 M_W^2}{16\pi^2}(V_{tb}^*V_{td})^2 \cdot C(\mu)\, Q \\
&= \frac{G_F^2 M_W^2}{16\pi^2}(V_{tb}^*V_{td})^2 S_0(x_t)\eta_B[\alpha_s(\mu)]^{-6/23}\left[1 + \frac{\alpha_s(\mu)}{4\pi}J_5\right] Q.
\end{aligned}
\tag{37}
$$

The result is entirely dominated by the top-quark contribution. It is common practice to separate the coefficient $C(\mu)$ into the function $S_0(x_t)$ (with $x_t = m_t^2/M_W^2$), which would be the coefficient in the absence of QCD effects, into the terms that depend on $\alpha_s(\mu)$, and the remainder, which is defined as the QCD-correction factor η_B. This has been done in the second equation in (37). Taking the matrix element of $\mathcal{H}_{\text{eff}}^{\Delta B=2}$ between the B and the $\bar{B}$ state and using (33) gives

$$\Delta m_B = \frac{G_F^2 M_W^2}{6\pi^2}|V_{tb}^*V_{td}|^2 S_0(x_t)\eta_B\, B_B f_B^2 m_B. \tag{38}$$

One encounters the hadronic matrix element of Q, which is written as

$$\langle\bar{B}|Q|B\rangle(\mu) \equiv \frac{8}{3}f_B^2 m_B^2 B_B(\mu), \tag{39}$$

defining the (scale and scheme dependent) hadronic parameter $B_B(\mu)$. The combination

$$B_B \equiv B_B(\mu)[\alpha_s(\mu)]^{-6/23}\left[1 + \frac{\alpha_s(\mu)}{4\pi}J_5\right], \tag{40}$$

is formally scale and scheme independent and has been used in (38). The parameter $B_B(\mu)$ is a nonperturbative quantity and has to be determined e.g. by lattice calculations. At present the value of B_B is still very uncertain, in contrast to the short-distance QCD corrections, which are precisely known. A numerical illustration is given in Table 1, where we have put $B_B(\mu) = 0.9$ as an example.

Table 1.	0.846			0.9	= 0.761
	η_B	$[\alpha_s(\mu)]^{-6/23}$	$\left[1 + \frac{\alpha_s(\mu)}{4\pi}J_5\right]$	$B_B(\mu)$	
	0.551	1.38			= 0.761

Two different definitions of a short-distance QCD factor can be considered, depending on where the terms with $\alpha_s(\mu)$ are included. One possibility is to include them with η_B into a Wilson coefficient (=0.846), which is to be multiplied by the hadronic matrix element $B_B(\mu) = 0.9$. The other possibility is the formally scheme independent separation into $\eta_B = 0.551$ and $B_B = 1.38$ (for η_B this is the precise result; $B_B = 1.38$ is only true in our example). The purpose of this exercise is to remind us that different definitions are sometimes employed for the parameter B_B and care has to be taken which one is being used, in order to combine it with the appropriate short-distance corrections. We can also see that the large deviation of the QCD correction factor η_B from 1 is merely a consequence of pulling out the large factor $[\alpha_s(\mu)]^{-6/23}$. It is somewhat artificial and does certainly not indicate a problem for perturbation theory. In fact, the coefficient 0.846 is the one that has the proper limit, approaching 1 as $\alpha_s \to 0$. It is indeed much closer to unity in accordance with the expectation for a perturbative correction factor. Still, the use of $\eta_B = 0.551$ is often adopted due to its formally scheme invariant definition.

An important application is the ratio of the mass differences for B_d and B_s mesons, for which (38) implies

$$\frac{\Delta m_{B_d}}{\Delta m_{B_s}} = \left|\frac{V_{td}}{V_{ts}}\right|^2 \frac{m_{B_d} f_{B_d}^2 B_{B_d}}{m_{B_s} f_{B_s}^2 B_{B_s}}. \tag{41}$$

This quantity is a very useful measure of $|V_{td}/V_{ts}|$. All other short-distance physics (top-dependence, η_B) has dropped out. Hadronic uncertainties are reduced in the ratio of matrix elements, which is 1 in the limit of unbroken $SU(3)$ flavour symmetry. The cancellation of the short-distance contribution is a direct consequence of the factorisation property of the OPE. Lattice calculations give for the ratio of matrix elements (Höcker et al. 2001, and references therein)

$$\frac{f_{B_s}\sqrt{B_{B_s}}}{f_{B_d}\sqrt{B_{B_d}}} = 1.16 \pm 0.06. \tag{42}$$

The ratio $\Delta m_{B_d}/\Delta m_{B_s}$ is a very powerful constraint for the unitarity triangle, as can be seen in Figure 3.

4   Heavy quark effective theory

4.1   Basic formalism

Heavy quark effective theory (HQET) is an effective field theory designed to systematically exploit the simplifications of QCD interactions in the heavy-quark limit for the case of hadrons containing a single heavy quark. The HQET Lagrangian can be derived as follows. We start from the usual QCD Lagrangian for a heavy-quark field Ψ with mass m

$$\mathcal{L} = \bar{\Psi} i \not{D} \Psi - m \bar{\Psi} \Psi, \tag{43}$$

with the covariant derivative

$$D_\mu = \partial_\mu - ig T^a A^a_\mu. \tag{44}$$

The heavy-quark momentum can be decomposed as

$$p = mv + k, \tag{45}$$

where v is the 4-velocity of the heavy *hadron*. Once mv, the large kinematical part of the momentum is singled out, the remaining component k is determined by soft QCD bound state interactions, and thus $k = \mathcal{O}(\Lambda_{\mathrm{QCD}}) \ll m$. We next decompose the quark field Ψ into

$$h_v(x) \equiv e^{imv \cdot x} \frac{1 + \not{v}}{2} \Psi(x), \tag{46}$$

$$H_v(x) \equiv e^{imv \cdot x} \frac{1 - \not{v}}{2} \Psi(x), \tag{47}$$

which implies

$$\Psi(x) = e^{-imv \cdot x} \left(h_v(x) + H_v(x) \right). \tag{48}$$

The expressions $(1 \pm \not{v})/2$ are projection operators. Their action represents the covariant generalisation of decomposing Ψ into upper and lower components. Using the standard representation for γ-matrices, this is evident in the rest frame where $\not{v} = \gamma^0$. Note also that the equation of motion with respect to the large momentum components, $m(\not{v} - 1)h_v = 0$, is manifest for h_v.

The exponential factor $\exp(imv \cdot x)$ in (46), (47) removes the large-frequency part of the x-dependence in $\Psi(x)$ resulting from the large momentum mv. Consequently, the x-dependence of h_v, H_v is only governed by the small residual momentum and derivatives acting on h_v and H_v count as $\mathcal{O}(\Lambda_{\mathrm{QCD}})$. (Our sign conventions are appropriate for a heavy *quark*. To describe the case of a heavy *anti-quark*, similar definitions are valid with the sign of v reversed.)

Multiplying the QCD equation of motion $(i\not{D} - m)\Psi = 0$ with the projectors $(1 - \not{v})/2$ and $(1 + \not{v})/2$, and using (46) – (48) and the definition

$$D^\mu_\perp \equiv D^\mu - v^\mu v \cdot D, \tag{49}$$

we obtain the coupled system of equations

$$iv \cdot D h_v = -i \not{D}_\perp H_v, \tag{50}$$

$$(iv \cdot D + 2m) H_v = i \not{D}_\perp h_v. \tag{51}$$

They represent the equation of motion in terms of h_v and H_v. The second equation implies that $H_v = \mathcal{O}(\Lambda_{\mathrm{QCD}}/m)h_v$ by power counting. Hence H_v is suppressed with respect to h_v in the heavy-quark limit. In other words, h_v contains the large components, H_v the small components of Ψ.

The HQET Lagrangian is obtained starting from (43), expressing Ψ in terms of h_v, H_v and eliminating H_v using (51). We find

$$\mathcal{L} = \bar{h}_v iv\cdot D h_v + \bar{h}_v i\slashed{D}_\perp \frac{1}{iv\cdot D + 2m} i\slashed{D}_\perp h_v. \tag{52}$$

Alternatively, H_v as obtained from (51) in terms of h_v can be inserted into (50) to yield the equation of motion for h_v. This equation is just the equation of motion implied by (52) (upon variation with respect to $\bar{h}_v$, i.e. $\delta\mathcal{L}/\delta\bar{h}_v = 0$). The Lagrangian may thus be written down immediately given the equation of motion for the field h_v.

The second term in (52) contains the nonlocal operator $(iv\cdot D + 2m)^{-1}$. It can be expanded in powers of Λ_{QCD}/m to yield a series of local operators. Keeping only the leading-power correction we can simply replace $(iv\cdot D + 2m)^{-1}$ by $(2m)^{-1}$ and get

$$\mathcal{L} = \bar{h}_v iv\cdot D h_v + \frac{1}{2m}\bar{h}_v(iD_\perp)^2 h_v + \frac{g}{4m}\bar{h}_v \sigma^{\mu\nu} G_{\mu\nu} h_v. \tag{53}$$

We now discuss some important aspects of this result.

The first term on the right hand side of (53) is the basic, lowest-order Lagrangian of HQET. It describes the "residual" QCD dynamics of the heavy quark once the kinematic dependence on m is separated out. Since there is no longer any reference to the mass m, the only parameter to distinguish quark flavours, this term is flavour symmetric: The dynamics is the same for b and c quarks in the static limit. Since the operator $v\cdot D$ contains no γ-matrices, which would act on the spin degrees of freedom, the leading HQET Lagrangian also exhibits a spin symmetry. This corresponds to the decoupling of the heavy-quark spin in the $m \to \infty$ limit. Together, we have the famous spin–flavour symmetries of HQET (Isgur & Wise 1989). They lead to relations among different heavy-hadron form factors.

From the Lagrangian $\bar{h}_v iv\cdot D h_v$ the Feynman rules for HQET can be read off. The propagator is

$$\frac{i}{v\cdot k} \frac{1 + \slashed{v}}{2}, \tag{54}$$

where the projector $(1 + \slashed{v})/2$ appears since h_v is a constrained spinor (see Equation 46). The interaction of the heavy-quark field h_v with gluons is given by the vertex

$$igv^\mu T^a. \tag{55}$$

These Feynman rules enter in the computation of QCD quantum corrections.

The remaining terms in (53) are the leading power corrections. They have an intuitive interpretation. In the first term one recognises the operator for the nonrelativistic kinetic energy $\mathbf{p}^2/(2m)$, which describes the residual motion of the heavy quark recoiling against the light degrees of freedom inside the heavy hadron. The last term represents the chromomagnetic interaction of the heavy-quark spin with the gluon cloud. Both effects violate flavour symmetry, the chromomagnetic term also spin symmetry, but they are power suppressed.

So far we have only considered QCD interactions. Weak interactions introduce external currents, which can also be incorporated in HQET. A generic heavy-light transition current $\bar{q}\Gamma\Psi$, arising for instance in semileptonic decays, can be represented as

$$\bar{q}\,\Gamma\,\Psi = \bar{q}\,\Gamma\,h_v + \mathcal{O}(\frac{1}{m}), \tag{56}$$

replacing the heavy-quark field Ψ by the HQET field h_v using (48).

4.2   Theory of heavy-hadron masses

Before considering HQET in the context of weak decays, let us discuss a first application of the basic HQET Lagrangian (53) in the spectroscopy of heavy hadrons. To be specific, we shall analyse the masses of the ground-state mesons B and B^*. These mesons constitute a doublet that arises because the spin 1/2 of the heavy quark couples with the *total spin* 1/2 of the light degrees of freedom in their ground state to form a spin-0 and a spin-1 meson, the pseudo-scalar B and the vector B^*, respectively. Because the b-quark spin decouples in the heavy-quark limit, the state of the light cloud is identical for B and B^* to leading order, and the angular-momentum coupling described above is the appropriate scheme. If we neglect the power corrections in (53), we can immediately write down the composition of the meson masses

$$m_B^{(0)} = m_{B^*}^{(0)} = m_b + \bar{\Lambda}. \tag{57}$$

Evidently the meson mass has a component m_b from the heavy quark. In addition it has a term $\bar{\Lambda} = \mathcal{O}(\Lambda_{\rm QCD})$ from the energy of the light constituents. The latter is determined only by the interactions among the light degrees of freedom and their interaction with the static b-quark (h_v) through the first term in (53). It is therefore independent of m_b. The sum of m_b and $\bar{\Lambda}$ is a physical quantity, however, separately both parameters are dependent on the scheme used to define them.

In order to include the first power corrections, we treat the $1/m$ terms in (53) as perturbations to the lowest-order HQET dynamics. To first order in perturbation theory the corrections to (57) are then simply given by the expectation values of the $1/m$ terms. The proper normalisation is obtained as follows. If $\mathcal{H} = -\mathcal{L}_{1/m}$ is the Hamiltonian (density) corresponding to the correction term $\mathcal{L}_{1/m}$ in (53), and $H = \int d^3x\,\mathcal{H}$ is the Hamilton operator, the mass correction due to $\mathcal{H}$ is just

$$\delta m_B = \langle B_1|H|B_1\rangle, \tag{58}$$

where $|B_1\rangle$ is the B-meson state normalised to one, $\langle B_1|B_1\rangle = 1$. Using the conventionally normalised states with $\langle B|B\rangle = 2m_B V$, we can write

$$\delta m_B = \frac{1}{2m_B V}\langle B|\int d^3x\,\mathcal{H}(\mathbf{x})|B\rangle = \frac{1}{2m_B V}\int d^3x\,\langle B|\mathcal{H}(0)|B\rangle = \frac{\langle B|\mathcal{H}(0)|B\rangle}{2m_B}, \tag{59}$$

where we have used the translation invariance of $\mathcal{H}$ and $\int d^3x = V$. Defining

$$\lambda_1 \equiv \frac{\langle B|\bar{h}(iD)^2 h|B\rangle}{2m_B} \qquad \lambda_2 \equiv \frac{1}{6}\frac{\langle B|\bar{h}g\sigma\cdot Gh|B\rangle}{2m_B}, \tag{60}$$

we obtain

$$\delta m_B = -\frac{\lambda_1 + 3\lambda_2}{2m_b}. \tag{61}$$

Note that we may replace D^2 by $D_\perp^2$ in the definition of λ_1, up to higher order corrections (see (49), (50)). The parameter λ_1 corresponds to (minus) the expectation value of the momentum squared of the heavy quark, $\lambda_1 = -\langle \mathbf{p}_h^2 \rangle = \mathcal{O}(\Lambda_{\text{QCD}}^2)$. This gives a positive correction in (61) representing the (small) kinetic energy of the heavy-quark. The λ_2-correction to the mass reflects the interaction energy of the heavy-quark spin with its hadronic environment, as already discussed in the previous section. While the λ_1-term is independent of the heavy-quark spin and identical for B and B^*, the chromomagnetic correction $\sim \lambda_2 = \mathcal{O}(\Lambda_{\text{QCD}}^2)$ is different for B^*. We have

$$\delta m_{B^*} = -\frac{\lambda_1 - \lambda_2}{2m_b}. \tag{62}$$

Including (57) we arrive at the following expansion for the meson masses

$$m_B = m_b + \bar{\Lambda} - \frac{\lambda_1 + 3\lambda_2}{2m_b}, \tag{63}$$

$$m_{B^*} = m_b + \bar{\Lambda} - \frac{\lambda_1 - \lambda_2}{2m_b}, \tag{64}$$

where the dependence on m_b is explicit order by order.

If we apply the heavy-quark limit to D mesons, we obtain analogous relations

$$m_D = m_c + \bar{\Lambda} - \frac{\lambda_1 + 3\lambda_2}{2m_c}, \tag{65}$$

$$m_{D^*} = m_c + \bar{\Lambda} - \frac{\lambda_1 - \lambda_2}{2m_c}, \tag{66}$$

with the same $\bar{\Lambda}$, λ_1 and λ_2 as before.

These results have a few interesting consequences. First, λ_2 parametrises the spin-splitting between the pseudo-scalar and the vector mesons:

$$m_{B^*} - m_B = \frac{2\lambda_2}{m_b} = 46\,\text{MeV}, \tag{67}$$

$$m_{D^*} - m_D = \frac{2\lambda_2}{m_c} = 141\,\text{MeV}. \tag{68}$$

HQET predicts that the spin-splitting scales inversely proportional to the heavy-quark mass. This is seen to be quite well fulfilled given that $m_b \approx 3m_c$. Relation (67) can be used to determine the nonperturbative quantity λ_2 from experiment

$$\lambda_2 = \frac{1}{4}(m_{B^*}^2 - m_B^2) = 0.12\,\text{GeV}^2. \tag{69}$$

On the other hand, the quantity λ_1 has to be estimated theoretically. Finally one may introduce the spin-averaged masses

$$\bar{m}_B \equiv \frac{m_B + 3m_{B^*}}{4} = m_b + \bar{\Lambda} - \frac{\lambda_1}{2m_b}, \tag{70}$$

$$\bar{m}_D \equiv \frac{m_D + 3m_{D^*}}{4} = m_c + \bar{\Lambda} - \frac{\lambda_1}{2m_c}. \tag{71}$$

This eliminates λ_2 and yields the useful result

$$m_b - m_c = (\bar{m}_B - \bar{m}_D)\left(1 - \frac{\lambda_1}{2\bar{m}_B\bar{m}_D}\right). \tag{72}$$

Since the λ_1-correction is fairly small, the quark-mass difference is rather well determined, much better than individual quark masses.

Exercise

Derive the relative factor between the chromomagnetic correction to the mass of the B and the B^* meson.

Solution: Denote the heavy-quark spin by **s**, the total spin of the light degrees of freedom by **j** and the total spin of the meson by $\mathbf{J} = \mathbf{s} + \mathbf{j}$. The chromomagnetic field of the light cloud has to be proportional to **j**. Hence the energy of the interaction between this field and **s** is proportional to $\langle \mathbf{s}\cdot\mathbf{j}\rangle$. The angular momentum algebra then gives $\langle 2\mathbf{s}\cdot\mathbf{j}\rangle = J(J+1) - s(s+1) - j(j+1)$, which is $(-3/2)$ for B and $(1/2)$ for B^*, hence the relative factor $(-1/3)$ of the λ_2-term in (62) with respect to (61).

4.3   Heavy-light currents and f_B

The B-meson decay constant f_B is defined by the matrix element

$$\langle 0|A_\mu|B(p)\rangle = -if_B m_B v_\mu, \tag{73}$$

of the heavy-light axial vector current

$$A_\mu \equiv \bar{q}\gamma_\mu\gamma_5\Psi. \tag{74}$$

Here q is the light-quark, Ψ the heavy-quark field in full QCD, with $\Psi = b$ in the present case. The B-meson momentum is $p = m_B v$.

Let us analyse A_μ in HQET, including QCD corrections. The expansion of A_μ in HQET to leading order in $1/m$, but allowing for QCD effects, has the form

$$A = C_1(\mu)\tilde{A}_1 + C_2(\mu)\tilde{A}_2 + \mathcal{O}\left(\frac{1}{m}\right) \tag{75}$$

$$\tilde{A}_1 = \bar{q}\gamma_\mu\gamma_5 h_v \qquad \tilde{A}_2 = \bar{q}v_\mu\gamma_5 h_v. \tag{76}$$

The matching conditions at the b-quark mass scale $\mu = m_b$ are

$$C_1(m_b) = 1 + \mathcal{O}(\alpha_s) \qquad C_2(m_b) = \mathcal{O}(\alpha_s). \tag{77}$$

To leading order in QCD only $\tilde{A}_1$ is present in HQET, with coefficient one. Radiative corrections at $\mathcal{O}(\alpha_s)$ modify C_1 and generate a new operator $\tilde{A}_2$. Note that the matching calculation of the full-QCD current A onto HQET, leading to (75), is completely analogous to the OPE procedure of constructing the effective weak Hamiltonian from the W-exchange amplitude in the full standard model, which we have discussed in Section 3. The difference is only that a $1/M_W$ expansion is performed in the latter case, and a $1/m_b$

expansion in the case of HQET. The basic philosophy is essentially the same. In partic-
ular, a factorisation of long and short-distance contributions is obtained: contributions
from large scales greater than μ, including the m_b-dependence, are again contained in the
coefficient functions $C_{1,2}$. Soft scales less than μ are factorised into the hadronic matrix
elements of $\tilde{A}_{1,2}$.

In contrast to the full-QCD current A, the HQET currents do have an anomalous
dimension, reflecting a logarithmic dependence of f_B on the heavy-quark mass at $\mathcal{O}(\alpha_s)$.
The logarithms can be re-summed by renormalisation group methods, again in full analogy
to the procedure in Section 3. In leading logarithmic approximation (LLA) C_2 can be
neglected and C_1 acquires the familiar form

$$C_1(\mu) = \left[\frac{\alpha_s(m_b)}{\alpha_s(\mu)}\right]^{-2/\beta_0}. \tag{78}$$

Here the LLA assumes the hierarchy $\alpha_s(m_b) \ll 1$, $\alpha_s \ln(m_b/\mu) = \mathcal{O}(1)$, which holds in
the heavy-quark limit (m_b large, $\mu = \mathcal{O}(1\,\text{GeV})$).

To express f_B in HQET via (73), (75) and (78), we need the matrix element of $\tilde{A}_1$

$$\langle 0|\tilde{A}_1|B(p)\rangle = -i\tilde{f}(\mu)\sqrt{m_B}v_\mu. \tag{79}$$

Since the dynamics of HQET is independent of m_b, the reduced decay constant $\tilde{f}(\mu)$ is
m_b-independent. The only m_b-dependence in (79) enters through a trivial factor $\sqrt{m_B}$
from the normalisation of the B-meson state, which in the usual convention is given by

$$\langle B|B\rangle = 2m_B V. \tag{80}$$

Collecting the ingredients, (75) yields

$$f_B = \frac{\tilde{f}(\mu)}{\sqrt{m_B}}\left[\frac{\alpha_s(m_b)}{\alpha_s(\mu)}\right]^{-2/\beta_0}. \tag{81}$$

This expression for f_B is true to leading order in the HQET expansion in Λ_{QCD}/m_b
and in leading logarithmic approximation in QCD. The factor $\tilde{f}(\mu)$ in (81) is still a
nonperturbative quantity to be determined by other methods. However, the dependence
of f_B on the heavy-quark mass is now explicit. Equation (81) implies the scaling behaviour
$f_B \sim 1/\sqrt{m_B}$, up to a calculable logarithmic dependence on m_b. In principle such a
relation can be used to relate f_B to the analogous quantity f_D for heavy mesons with
charm. In practice, it turns out that the leading order scaling result for f_B is not very
well fulfilled even for the b-mass scale and that subleading power corrections are important
in this case. Nevertheless the result in (81) is of conceptual interest and can serve as a
simple example of an application of HQET.

4.4   Heavy-heavy currents: $\bar{B}\rightarrow D^{(*)}l\bar{\nu}$ and V_{cb}

One of the most important results of HQET is the extraction of V_{cb} from exclusive semilep-
tonic $\bar{B} \rightarrow D^*l\bar{\nu}$ decay. We will here give a short outline of the main steps in this analysis.
The starting point is the differential decay rate

$$\frac{d\Gamma(\bar{B} \rightarrow D^*l\bar{\nu})}{dw} = |V_{cb}|^2\,\mathcal{K}(w)\,\mathcal{F}^2(w) \tag{82}$$

in the kinematical variable $w = v \cdot v'$, where v and v' are the 4-velocities of $\bar{B}$ and D^*, respectively. The dependence of (82) on $|V_{cb}|$, the quantity of interest, is obvious, and $\mathcal{K}(w)$ is a known kinematical function. Finally, $\mathcal{F}(w)$ contains the nontrivial QCD dynamics encoded in the $\bar{B} \to D^*$ transition form factors. The corresponding matrix elements of the weak currents can be written in the heavy-quark limit as

$$\frac{1}{\sqrt{m_{D^*} m_B}} \langle D^*(v', \epsilon)|\bar{c}\gamma_\mu b|\bar{B}(v)\rangle = i\,\xi(w)\,\varepsilon(\mu, \epsilon, v', v), \tag{83}$$

$$\frac{1}{\sqrt{m_{D^*} m_B}} \langle D^*(v', \epsilon)|\bar{c}\gamma_\mu \gamma_5 b|\bar{B}(v)\rangle = \xi(w)\left[(1+w)\epsilon_\mu - (\epsilon \cdot v)v'_\mu\right]. \tag{84}$$

In the heavy-quark limit, that is to lowest order in HQET, all hadronic dynamics is expressed in a single function $\xi(w)$, the Isgur–Wise function (Isgur & Wise 1989). In this limit we further have

$$\mathcal{F}(w) = \xi(w). \tag{85}$$

Moreover, ξ is absolutely normalised at the no-recoil point

$$\xi(1) = 1. \tag{86}$$

The no-recoil point $w = 1$ corresponds to the kinematical situation where the D^* meson stays at rest in the rest frame of the decaying $\bar{B}$ ($v' = v \Rightarrow w = 1$). Measuring $d\Gamma/dw$ at $w = 1$, $|V_{cb}|$ can then be determined from (82) since all ingredients are known. Because $w = 1$ is at the edge of phase space, an extrapolation is necessary to find $d\Gamma/dw|_{w=1}$ from the measured spectrum.

For a realistic analysis corrections to the heavy-quark limit need to be considered. An important property of $\bar{B} \to D^* l \bar{\nu}$ is that linear power corrections in HQET are absent, $\delta_{1/m} = 0$, where m can be either m_c or m_b. Consequently the leading corrections enter only at second order and are thus greatly reduced. This result is known as Luke's theorem. The absence of linear corrections does not hold for $\bar{B} \to D l \bar{\nu}$ decays, hence the particular importance of $\bar{B} \to D^* l \bar{\nu}$. Including corrections, the lowest order approximation $\mathcal{F}(1) = \xi(1) = 1$ is modified to

$$\mathcal{F}(1) = \eta_A(1 + \delta_{1/m^2}), \tag{87}$$

where δ_{1/m^2} are the second order power corrections and η_A is a correction from perturbative QCD. To first order in α_s it reads

$$\eta_A = 1 + \frac{\alpha_s}{\pi}\left(\frac{m_b + m_c}{m_b - m_c}\ln\frac{m_b}{m_c} - \frac{8}{3}\right). \tag{88}$$

A detailed numerical analysis yields (Harrison & Quinn 1998)

$$\mathcal{F}(1) = 0.913 \pm 0.042, \tag{89}$$

which gives (Höcker et al. 2001)

$$V_{cb} = 0.0409 \pm 0.0014_{\text{exp}} \pm 0.0019_{\text{th}}. \tag{90}$$

To summarise the crucial points for the extraction of V_{cb} from $\bar{B} \to D^* l \bar{\nu}$ decay:

- Heavy-quark symmetry relates the various semileptonic form factors (four different functions V, A_0, A_1, A_2 in full QCD) to a single quantity $\xi(w)$, the Isgur–Wise function.

- The function ξ is absolutely normalised, $\xi(1) = 1$. This property has an intuitive reason: At the kinematical point $w = 1$ the decaying b-quark at rest is transformed into a c-quark, also at rest. Since both quarks are treated in the static approximation (m_b, $m_c \to \infty$, m_b/m_c fixed), the light hadronic cloud doesn't notice the flavour change $b \to c$ and is transferred from the $\bar{B}$ to a D meson with probability one. The function ξ is identical for $\bar{B} \to D$ and $\bar{B} \to D^*$ transitions, because these are related by heavy-quark spin symmetry.

- HQET provides a framework for systematic corrections to the strict heavy-quark limit governed by $\xi(w)$. Luke's theorem guarantees the absence of first-order corrections in $1/m$ for $\bar{B} \to D^* l \bar{\nu}$.

4.5 HQET: conclusions

We would finally like to summarise the basic ideas and virtues of HQET, and to re-emphasise the salient points.

- HQET describes the static approximation for a heavy quark, covariantly formulated as an effective field theory and allowing for a systematic inclusion of power corrections.

- Order by order in the expansion in Λ_{QCD}/m HQET achieves a *factorisation* of hard, perturbative contributions (momentum scales between m and a factorisation scale μ) and soft, nonperturbative contributions (scales below μ). The former are contained in Wilson coefficients, the latter in the matrix elements of HQET operators.

- The procedure of matching full QCD onto HQET is analogous to the construction of the effective weak Hamiltonian $\mathcal{H}_{\mathrm{eff}}$. The difference lies in the massive degrees of freedom that are being integrated out: the W boson (mass M_W) for $\mathcal{H}_{\mathrm{eff}}$, the lower-component spinor field H_v (mass $2m$) for HQET. The perturbative matching can be supplemented by RG resummation of logarithms, $\ln(M_W/\mu)$ in the former case, $\ln(m/\mu)$ in the latter.

- The usefulness of HQET is based on two important features: The spin-flavour symmetry of HQET relates form factors in the heavy-quark limit and thus reduces the number of unknown hadronic quantities. The dependence on the heavy-quark masses is made explicit (scaling, power corrections).

We conclude with briefly mentioning another field, called large energy effective theory (LEET), which has some similarities with HQET. LEET is needed for matrix elements of the form $\langle M | \bar{q}\, \Gamma\, b | \bar{B} \rangle$ at *large recoil* of the light meson $M = \pi$, ρ, $K^{(*)}$, etc. Then HQET is not sufficient in this situation because not only soft but also collinear infrared singularities need to be factorised. The latter occur due to the light-like kinematics of the fast and energetic light quark emitted from the weak current. To define LEET the

usual heavy-quark limit can be considered for the B meson with velocity v. The large-energy limit is taken for the light meson M with light-like momentum vector En. Here $E = \mathcal{O}(m_b)$ is the energy of M and n is a light-like 4-vector with $n^2 = 0$ and $v \cdot n = 1$. The momentum of the energetic light quark q is written as $p_q = En + k$, with a residual momentum $k = \mathcal{O}(\Lambda_{\mathrm{QCD}})$. In formal analogy to the fields h_v and H_v in HQET, the new light-quark fields

$$q_n(x) = e^{iEn \cdot x} \frac{\not{n} \not{v}}{2} q(x) \qquad Q_n(x) = e^{iEn \cdot x} \frac{\not{v} \not{n}}{2} q(x), \tag{91}$$

can be defined and used in the construction of LEET (Dugan & Grinstein 1991, Charles et al. 1999, Beneke & Feldmann 2001, Bauer et al. 2001a). As a consequence of the LEET limit the ten form factors needed to describe all matrix elements $\langle M | \bar{q} \, \Gamma \, b | \bar{B} \rangle$ of bilinear heavy-light currents can be reduced to only three independent functions. LEET has received increasing interest quite recently and is still under active development.

5   Inclusive decays and the heavy quark expansion

5.1   Basic formalism and theory of lifetimes

The heavy-quark limit, $m \gg \Lambda_{\mathrm{QCD}}$, proves to be extremely useful also for the computation of *inclusive* decay rates of heavy hadrons (Chay et al. 1990, Bigi et al. 1992, 1997). The specific technique appropriate for this application is distinct from HQET and goes by the name of heavy quark expansion (HQE). Consider the total decay rate Γ_H of a heavy hadron H. The starting point for the HQE is the following representation of Γ_H

$$\Gamma_H = \frac{1}{2m_H} (H | \mathcal{T} | H) \equiv \langle \mathcal{T} \rangle, \tag{92}$$

where the transition operator $\mathcal{T}$ is defined as

$$\mathcal{T} = \mathrm{Im}\, i \int d^4 x \, T \, \mathcal{H}_{\mathrm{eff}}(x) \mathcal{H}_{\mathrm{eff}}(0), \tag{93}$$

with $\mathcal{H}_{\mathrm{eff}}$ the effective weak Hamiltonian. Equations (92), (93) express the total decay rate as the absorptive part of the forward scattering amplitude $H \to H$ under the action of $\mathcal{H}_{\mathrm{eff}}$. This expression is referred to as the optical theorem by analogy to a similar relation in optics. One may rewrite (92), (93) in a more directly understandable form by inserting a complete set of states $|X\rangle\langle X|$ between the two factors of $\mathcal{H}_{\mathrm{eff}}$ in (93) and removing the T-product by explicitly taking the absorptive part. This yields

$$\Gamma_H \sim \langle H | \mathcal{H}_{\mathrm{eff}} | X \rangle \langle X | \mathcal{H}_{\mathrm{eff}} | H \rangle, \tag{94}$$

where one immediately recognises the decay rate as the modulus squared of the decay amplitude (summed over all final states X). The reason to introduce (93) is that the T-product, by means of Wick's theorem, allows for a direct evaluation in terms of Feynman diagrams.

In order to compute Γ_H an operator product expansion is applied to (93), resulting in a series of local operators of increasing dimension. The coefficients of these operators are

correspondingly suppressed by increasing powers of $1/m_b$. The series has the form

$$\mathcal{T} = \Gamma_b \bar{b}b + \frac{z_G}{m_b^2} \bar{b}g\sigma\cdot Gb + \sum \frac{z_{qi}}{m_b^3} \bar{b}\Gamma_i q \; \bar{q}\Gamma_i b + \ldots, \tag{95}$$

where we have written the first few operators of dimension three $(\bar{b}b)$, five $(\bar{b}g\sigma\cdot Gb)$ and six $(\bar{b}\Gamma_i q \; \bar{q}\Gamma_i b)$. The matrix elements of the operators contain the soft, nonperturbative physics, their Wilson coefficients Γ_b, z_k the hard contributions, which are calculable in perturbation theory. Again, the coefficients are determined by an appropriate matching calculation between (93) and the r.h.s. of (95). The Feynman diagrams for the three terms in (95) are shown in Figure 11 The two weak-interaction vertices in these diagrams
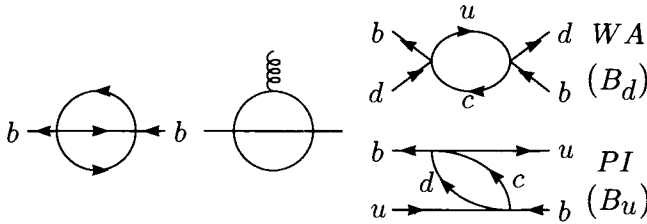
Figure 11. *Heavy quark expansion for the total decay rate of b-hadrons.*

correspond to the two factors of $\mathcal{H}_{\text{eff}}$ in the definition of $\mathcal{T}$ in (93) (the absorptive part of the diagrams is understood).

Obviously, the heavy quark expansion is different from HQET. However, we may still use HQET in conjunction with (95) in order to further analyse the hadronic matrix elements. An important example is the leading dimension-three operator $\bar{b}b$. Its matrix element between heavy-hadron states H can be expanded in HQET as

$$\langle \bar{b}b \rangle = 1 + \frac{1}{2m_b^2} \langle \bar{h}(iD)^2 h \rangle + \frac{1}{4m_b^2} \langle \bar{h}g\sigma Gh \rangle, \tag{96}$$

where $\langle \ldots \rangle \equiv \langle H| \ldots |H\rangle/(2m_H)$.

Equations (92), (95) and (96) imply that to leading order in the HQE $\Gamma_H = \Gamma_b$, that is the total decay rate of all b-flavoured hadrons is equal to the rate of free b-quark decay. Pictorially this can be seen from the first diagram in Figure 11, which represents essentially the amplitude squared for the partonic decay of a b-quark. Note also that perturbative QCD corrections to Γ_b can consistently be taken into account. The gluonic corrections to inclusive b-quark decay are infrared safe, as required for Γ_b in its role as a Wilson coefficient of the HQE. Also, corrections proportional to powers of $\alpha_s(m_b) \sim 1/\ln(m_b/\Lambda)$ are only suppressed by inverse powers of $\ln m_b$ in the heavy-quark limit, and hence formally leading in comparison to higher corrections in the HQE, which are suppressed by powers of Λ/m_b. The calculation of heavy-quark decay in the parton picture has been used since the beginnings of heavy-quark physics as an approximation for inclusive decays of the corresponding heavy hadrons. As we have seen, the HQE gives a formal justification for this approach and provides us with a theoretical framework to compute nonperturbative corrections.

The first correction term in (96) depends on the expectation value of the momentum squared $\langle \mathbf{p}^2 \rangle$ of the heavy quark inside the hadron. This matrix element is non-zero

because the heavy quark is recoiling against the light degrees of freedom through gluonic interactions in the hadronic bound state. This term has a very intuitive interpretation. It corresponds to a correction factor $1 - \langle \mathbf{p}^2 \rangle/(2m_b^2) = 1 - \langle \mathbf{v}_b^2 \rangle/2$, which is just the reduction of the free decay rate from time dilatation due to the recoil motion of the heavy quark. The second correction comes from interactions of the light hadronic cloud with the heavy-quark spin. We have

$$\langle \bar{h} g \sigma G h \rangle = \begin{cases} \frac{3}{2}(m_{B^*}^2 - m_B^2) & H = B \\ 0 & H = \Lambda_b \end{cases}. \tag{97}$$

The result is zero for the Λ_b baryon since the light degrees of freedom are in a state of zero total angular momentum. Note that the spin interaction enters twice in (95), explicitly with coefficient z_G and via the expansion of $\langle \bar{b}b \rangle$.

The leading nonperturbative corrections start only at second order. There is no correction linear in $1/m_b$. This is because there is no gauge-invariant operator of dimension four that could appear in the HQE.

At order $1/m_b^3$ contributions appear where the spectator quark participates directly in the weak interactions. For b-mesons they can be interpreted as the effect of weak annihilation (WA) of the b-quark with the valence $\bar{d}$-quark (for $\bar{B}_d$) and as the effect of Pauli interference (PI) (for $\bar{B}_u$). The latter phenomenon occurs because in the nonleptonic decay of a $\bar{B}_u$, $b(\bar{u}) \to c\bar{u}d(\bar{u})$, two identical $\bar{u}$-quarks are present in the final state. These corrections distinguish, in particular, between B_d and B_u mesons and are responsible for their lifetime difference. Despite the suppression by three powers of m_b these effects can be relatively important due to their two-body kinematics, which brings a phase-space enhancement factor of $16\pi^2$ in comparison to the leading three-body decay.

As one of the possible applications, the HQE provides us with a theory of heavy-hadron lifetimes. The deviations of lifetime ratios from unity probes the power corrections. At present there are still sizeable theoretical uncertainties due to the hadronic matrix elements $\langle \bar{b} \Gamma q \, \bar{q} \Gamma b \rangle$. They can in principle be computed with the help of lattice gauge theory. Table 2 shows a comparison of theoretical predictions and experimental results (see for instance (Ligeti 2001)).

Table 2.

	Experiment	Theory
$\tau(B^+)/\tau(B_d^0)$	1.068 ± 0.016	$1 - 1.1$
$\bar{\tau}(B_s)/\tau(B_d)$	0.947 ± 0.038	$0.99 - 1.01$
$\tau(\Lambda_b)/\tau(B_d)$	0.795 ± 0.053	$0.9 - 1.0$

5.2   Local quark-hadron duality

A systematic uncertainty within the HQE framework, which is often debated in the literature, arises from the issue of quark-hadron duality. In this paragraph we give a brief and heuristic discussion of the basic idea behind this topic.

The theoretical prediction for an inclusive decay rate obtained from the HQE has the
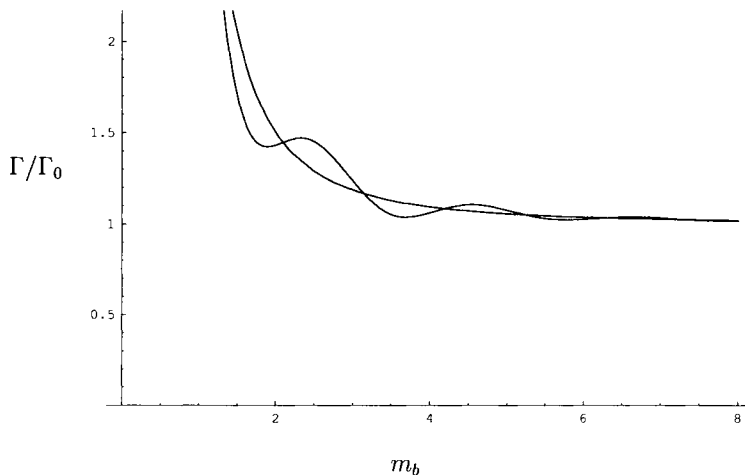
Figure 12. Γ/Γ_0 *as function of* m_b *(arbitrary units).*

form

$$\Gamma/\Gamma_0 = 1 + \sum_{n=2}^{\infty} z_n \left(\frac{\Lambda}{m_b}\right)^n, \qquad (98)$$

where we have denoted the leading, free-quark result by Γ_0. Let us consider the decay rate as a function of m_b, keeping $\Lambda = \Lambda_{\mathrm{QCD}}$ constant. Then the quantity Γ/Γ_0, to any finite order in (Λ/m), is a simple polynomial expression in this variable. This is sketched as the monotonic curve in Figure 12 showing Γ/Γ_0 as function of m_b (in arbitrary units). Now since, by construction, the HQE for Γ/Γ_0 yields a power expansion in (Λ/m), any term of the form

$$\exp\left(-\left(\frac{m_b}{\Lambda}\right)^k\right) \sin\left(\frac{m_b}{\Lambda}\right)^k, \qquad (99)$$

for example, present in the true result for Γ/Γ_0 would be missed by the HQE. This is due to the exponential suppression in the expansion parameter. In fact, the function $\exp(-1/x)$ is non-analytic. Its power expansion around $x = 0$ gives identically zero. However, such (or similar) terms are expected to be part of the true Γ/Γ_0 on general grounds. The corresponding complete result for Γ/Γ_0, including such a term, is sketched as the oscillating graph in Figure 12. This true curve represents the physical result for the decay rate Γ/Γ_0, which consists of the inclusive sum over all the different exclusive decay channels. It is intuitively understandable that the true m_b-dependence will have such a damped oscillating behaviour: if we imagine continually increasing m_b, Γ/Γ_0 will undergo a small jump whenever it reaches a value at which the presence of a further higher hadronic resonance in the final state becomes kinematically allowed. Since the excited hadrons have finite widths, the threshold behaviour will be smoothed out, resulting in the pattern of damped oscillations.

The term *quark-hadron duality* refers to the idea that the inclusive rate as the sum over all exclusive *hadronic* decay channels and the inclusive rate as predicted by the heavy *quark* expansion are dual to each other. This means they are both valid representations of the same quantity using different descriptions, the hadron level or the quark level. The

term *local* refers to the fact that the energy scale m_b is a fixed quantity, as opposed to e.g. the centre-of-mass energy in e^+e^- annihilation, which can be averaged to obtain suitably defined "global" quantities. In principle, the hadronic description gives the true result, measured in experiment. The problem is, however, that we would have to compute all exclusive rates first, which is far beyond our current control of nonperturbative QCD. On the other hand, the HQE calculation can be performed, within some uncertainties, but it is clear that the result need not be identical to the true answer. A deviation between the latter and the HQE (including power corrections) is referred to as a violation of quark-hadron duality. Indeed, contributions violating quark-hadron duality are expected (see(99)), but the numerical size of these terms cannot be strictly computed at present. Conceptually this is no problem because they are formally subleading in comparison to power corrections, so that the HQE still makes sense even at higher orders. The remaining question is how large can violations of quark-hadron duality be numerically. While there are at the moment, within the uncertainties intrinsic to HQE, no established cases in inclusive B decays where duality is violated, the issue clearly needs further investigation, both theoretically and phenomenologically.

A more detailed account of the status of quark-hadron duality can be found in the papers by Blok, Shifman & Zhang (1998), Shifman (2000) and Bigi & Uraltsev (2001).

5.3   Inclusive semileptonic decays: V_{ub} and V_{cb}

The HQE cannot only be applied to the total decay rates, but also to inclusive rates with specific flavour quantum numbers in the final state, such as semileptonic processes. Furthermore one can analyse differential decay rates.

An example of special interest is the inclusive decay $\bar{B} \to X_c l \bar{\nu}$, which can be used to extract V_{cb}. The HQE for the integrated rate has the form

$$\Gamma(B \to X_c l \nu) = \frac{G_F^2 m_b^5}{192\pi^3} |V_{cb}|^2 \left[z_3 \left(1 + \frac{\lambda_1 + 3\lambda_2}{2m_b^2} \right) + z_5 \frac{6\lambda_2}{m_b^2} + \dots \right], \qquad (100)$$

with

$$\langle \bar{b}b \rangle = 1 + \frac{\lambda_1 + 3\lambda_2}{2m_b^2} \qquad \langle \bar{b}\sigma Gb \rangle = 6\lambda_2 = \frac{3}{2}(m_{B^*}^2 - m_B^2). \qquad (101)$$

The Wilson coefficients read

$$z_3 = 1 - 8x + 8x^3 - x^4 - 12x^2 \ln x + \mathcal{O}(\alpha_s), \qquad (102)$$
$$z_5 = -(1-x)^4, \qquad (103)$$

where $x = (m_c/m_b)^2$.

A major source of theoretical uncertainty for the determination of $|V_{cb}|$ using (100) is the b-quark mass. This appears to be especially problematic since m_b comes with the fifth power in (100). Fortunately, however, the actual situation is not as bad. Taking into account the phase-space function z_3, one finds that the combined dependence on m_b and m_c shows the approximate behaviour

$$\Gamma(B \to X_c l \nu) \sim m_b^{2.3} (m_b - m_c)^{2.7}. \qquad (104)$$

Since the difference $m_b - m_c$ is better known than the individual quark masses, the corresponding uncertainty is reduced. The quark-mass difference is in fact constrained by HQET, which gives (72)

$$
\begin{aligned}
m_b - m_c &= (\bar{m}_B - \bar{m}_D)\left(1 - \frac{\lambda_1}{2\bar{m}_B \bar{m}_D}\right) \\
&= 3.40 \pm 0.03 \pm 0.03 \text{GeV},
\end{aligned}
\tag{105}
$$

where $\bar{m}_B \equiv (m_B + 3m_{B*})/4$.

The QCD corrections to z_3 are known to $\mathcal{O}(\alpha_s)$ and partly at $\mathcal{O}(\alpha_s^2)$. The special class of corrections $\mathcal{O}(\beta_0^{n-1}\alpha_s^n)$ has been calculated to all orders n.

Numerically the inclusive method gives (Höcker et al. 2001)

$$
V_{cb} = 0.04076 \pm 0.00050_{\text{exp}} \pm 0.00204_{\text{th}},
\tag{106}
$$

which can be compared with the result from the exclusive determination via $\bar{B} \to D^* l \bar{\nu}$ (90).

One can also try to extract $|V_{ub}|$ from $B \to X_u l \nu$ decays. This is more difficult since the very large background from semileptonic $b \to c$ transitions requires kinematical cuts (in the lepton energy, the hadronic or the dilepton invariant mass), which renders the HQE less reliable and introduces larger uncertainties. A recent discussion has been given by Bauer et al. (2001b). The HQE has further useful applications, for instance in the case of the inclusive rare decays $B \to X_{s,d}\gamma$, $B \to X_{s,d}l^+l^-$, or $B \to X_{s,d}\nu\bar{\nu}$.

Exercise

Show that quark-hadron duality is exactly fulfilled for the semileptonic $b \to c$ transition rate in the Shifman-Voloshin (small-velocity, or SV) limit $m_b, m_c \gg m_b - m_c \gg \Lambda_{\text{QCD}}$. This holds with only two exclusive channels on the hadronic side of the duality relation, that is the inclusive rate is saturated as $\Gamma(B \to X_c l\nu) \equiv \Gamma(B \to Dl\nu) + \Gamma(B \to D^* l\nu)$ in this limit.

Solution: We start from the exclusive differential decay rates in the *heavy-quark limit*. They read (see e.g. Harrison & Quinn 1998):

$$
\frac{d\Gamma(B \to Dl\nu)}{dw} = \frac{G_F^2 |V_{cb}|^2}{48\pi^3}(m_B + m_D)^2 m_D^3 (w^2 - 1)^{3/2}\,\xi^2(w)
\tag{107}
$$

$$
\begin{aligned}
\frac{d\Gamma(B \to D^* l\nu)}{dw} = {}& \frac{G_F^2 |V_{cb}|^2}{48\pi^3}(m_B - m_D^*)^2 m_{D*}^3 \sqrt{w^2 - 1}(w + 1)^2 \\
&\times \left(1 + \frac{4w}{w+1}\frac{m_B^2 - 2wm_B m_{D*} + m_{D*}^2}{(m_B - m_{D*})^2}\right)\xi^2(w).
\end{aligned}
\tag{108}
$$

In the strict SV limit we have

$$
m_B = m_b, \qquad m_{D*} = m_D = m_c, \qquad m_c = m_b(1 - \epsilon),
\tag{109}
$$

where $\epsilon \equiv (m_b - m_c)/m_b$ is a small parameter. The variable w is related to the dilepton invariant mass q^2 through

$$
q^2 = m_B^2 + m_D^2 - 2m_B m_D w.
\tag{110}
$$

The kinematic limits of q^2 are easily identified as

$$q_{max}^2 = (m_b - m_c)^2, \qquad q_{min}^2 = 0. \tag{111}$$

The corresponding limits of w are

$$w_{min} = 1, \qquad w_{max} = \frac{m_b^2 + m_c^2}{2 m_b m_c}. \tag{112}$$

Defining $s \equiv w - 1$ we have $0 \leq s \leq \epsilon^2/2$, where the upper limit is valid to leading order in ϵ. Expanded to leading order in ϵ, (107) gives

$$\Gamma(B \to Dl\nu) = \frac{G_F^2 |V_{cb}|^2}{6\pi^3} m_b^5 \sqrt{2} \int_0^{\epsilon^2/2} s^{3/2}\, ds = \frac{G_F^2 |V_{cb}|^2}{60\pi^3} (m_b - m_c)^5, \tag{113}$$

which is the decay rate in the SV limit. In this derivation we have made use of the fact that $\xi(w) = \xi(1) + \mathcal{O}(\epsilon^2)$, which can be approximated by $\xi(1) = 1$. In this way any dependence on nontrivial hadronic input has disappeared. Similarly we can expand the integral over (108) in ϵ to extract the leading contribution in the SV limit. We obtain

$$\Gamma(B \to D^* l\nu) = \frac{G_F^2 |V_{cb}|^2}{20\pi^3} (m_b - m_c)^5. \tag{114}$$

We also observe that higher D-meson resonances and hadronic multiparticle states have wave functions of the light degrees of freedom that are orthogonal to the ground state wave function of the light cloud (identical for D, D^* and B) in the SV limit. There is therefore no overlap of those higher excitations with the initial B and the corresponding rates vanish.

Finally, we need to take the SV limit of the inclusive rate as obtained from the heavy quark expansion in (100). In this limit the second-order power corrections and perturbative QCD corrections disappear, and we only have to expand the phase space function z_3 in the small-ϵ limit. We find $z_3 = 64\epsilon^5/5 + \mathcal{O}(\epsilon^6)$ and

$$\Gamma(B \to X_c l\nu) = \frac{G_F^2 |V_{cb}|^2}{15\pi^3} (m_b - m_c)^5. \tag{115}$$

We see that indeed the inclusive HQE result (115) is saturated by the sum of just the two exclusive rates (113) and (114). Clearly, the SV limit is a very special situation. Nevertheless, it is an interesting example of exact (local) quark-hadron duality. Moreover, the semileptonic rates into D and D^* measured in experiment account for roughly two thirds of the inclusive rate, indicating that the SV limit is not even entirely unrealistic.

# 6   QCD factorisation in exclusive hadronic B decays

6.1   Introduction

Decay amplitudes for exclusive nonleptonic B decays, such as $B \to \pi\pi$, can be computed starting from the effective weak Hamiltonian discussed in Section 3.3. Whereas the Wilson coefficients C_i are well understood, the main problem is posed by the hadronic matrix elements of the operators Q_i. In some cases this problem can be circumvented (CP asymmetry in $B \to J/\Psi K_S$), or at least reduced using SU(2) or SU(3) flavour symmetries

and an appropriate combination of various channels. However, an improved understanding of the QCD dynamics in exclusive hadronic B decays would greatly enhance our capability to extract from these processes the underlying flavour physics.

Indeed, it turns out that the heavy-quark limit leads to substantial simplifications also in the problem of hadronic two-body decays of heavy hadrons. Again the main feature is the factorisation of short-distance and long-distance contributions. In the case of the matrix elements of four-quark operators Q_i the factorisation takes the form

$$\langle \pi(p')\pi(q)|Q_i|\bar{B}(p)\rangle = f^{B\to\pi}(q^2)\int_0^1 du\, T_i^I(u)\Phi_\pi(u)$$

$$+ \int_0^1 d\xi du dv\, T_i^{II}(\xi, u, v)\Phi_B(\xi)\Phi_\pi(u)\Phi_\pi(v). \quad (116)$$

This *factorisation formula* is valid up to corrections of relative order $\Lambda_{\rm QCD}/m_b$. Here $f^{B\to\pi}(q^2)$ is a $B \to \pi$ form factor evaluated at $q^2 = m_\pi^2 \approx 0$, and Φ_π (Φ_B) are leading-twist light-cone distribution amplitudes "wave functions") of the pion (B meson). These objects contain the long-distance dynamics. The short-distance physics, dominated by scales of order m_b, is described by the hard-scattering kernels $T_i^{I,II}$, which are calculable in perturbation theory. T_i^I starts at $\mathcal{O}(\alpha_s^0)$, T_i^{II} at $\mathcal{O}(\alpha_s^1)$ (see Figure 13). In (116) long-
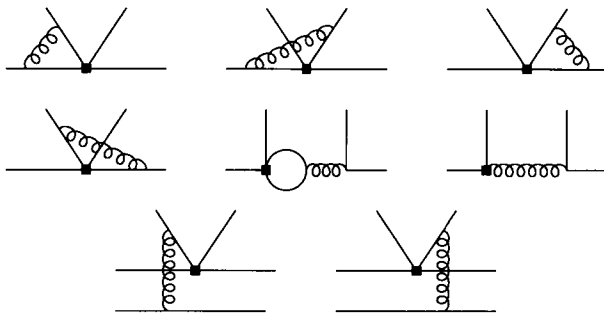
Figure 13. *Order α_s corrections to the hard scattering kernels T_i^I (first two rows) and T_i^{II} (last row). In the case of T_i^I, the spectator quark does not participate in the hard interaction and is not drawn. The two lines directed upwards represent the two quarks forming the emitted pion.*

and short-distance contributions are thus systematically disentangled, that is *factorised*. The long-distance sensitive quantities (form factors and wave functions) still need to be determined by other means, but they are universal quantities and much simpler than the original full $B \to \pi\pi$ matrix elements we started with. They could in principle be calculated by nonperturbative methods or extracted experimentally from other observables. In any case (116) represents a substantial simplification of our problem.

The general expression (116) further simplifies when we neglect perturbative α_s-corrections. The T^{II} term is then absent and the kernel T^I becomes a constant in u, such that the pion distribution amplitude integrates to the pion decay constant. The matrix element of operator Q_1^u, for instance, reduces to

$$\langle \pi^+\pi^-|(\bar{u}b)_{V-A}(\bar{d}u)_{V-A}|\bar{B}\rangle \to \langle \pi^+|(\bar{u}b)_{V-A}|\bar{B}\rangle \cdot \langle \pi^-|(\bar{d}u)_{V-A}|0\rangle = im_B^2 f^{B\to\pi}(0)f_\pi. \quad (117)$$

This procedure, termed "naive factorisation" has long been used in phenomenological application, but the justification had been less clear. An obvious issue is the scheme

and scale dependence of the matrix elements of four-quark operators, which is needed to cancel the corresponding dependence in the Wilson coefficients. This dependence is lost in naive factorisation as the factorised currents are scheme independent objects. In QCD factorisation (116) the proper scale and scheme dependence is recovered by the inclusion of $\mathcal{O}(\alpha_s)$ corrections as we will see explicitly below.

A qualitative justification for (117) had been given by Bjorken (1989). It is based on the *colour transparency* of the hadronic environment for the highly energetic pion emitted in B decay (the π^- in the above example, which is being created from the vacuum). This is related to the decoupling of soft gluons from the small-size colour-singlet object that the emitted pion represents. The QCD factorisation approach as encoded in (116) may be viewed as a consistent formalisation and generalisation of Bjorken's colour transparency argument. This treatment of hadronic B decays is based on the analysis of Feynman diagrams in the heavy quark limit, utilising consistent power counting to identify the leading contributions. The framework is very similar in spirit to more conventional applications of perturbative QCD in exclusive hadronic processes with a large momentum transfer, as the pion electromagnetic form factor (see the article by Sterman and Stoler (1997) for a recent review). It justifies and extends the ansatz of naive factorisation. In particular the method includes, for $B \to \pi\pi$, the hard nonfactorisable spectator interactions, penguin contributions and re-scattering effects (Figure 13). As a corollary, one finds that strong re-scattering phases are either of $\mathcal{O}(\alpha_s)$, and calculable, or power suppressed. In any case they vanish therefore in the heavy quark limit. QCD factorisation is valid for cases where the emitted particle (the meson created from the vacuum in the weak process, as opposed to the one that absorbs the b-quark spectator) is a small size colour-singlet object, e.g. either a fast light meson (π, ϱ, K, K^*) or a J/Ψ.

Note that the term *factorisation* is used here for two a priori entirely different things. In the case of QCD factorisation (116), it refers to the factorisation of short-distance and long-distance contributions. In the sense of the phenomenological approach of naive factorisation (117), it simply denotes the separation of the hadronic matrix element of a four-quark operator into two factors of matrix elements of bilinear currents. It is a nontrivial result that the latter, naive factorisation is obtained as the lowest order approximation of QCD factorisation. To avoid confusion, it is useful to keep the distinction in mind. For example, the hard gluon exchange corrections between the two quark currents in Figure 13 are "nonfactorisable" in the sense of naive factorisation, although they are a consistent ingredient of (116), hence "factorisable" in the sense of QCD.

In the following we shall discuss QCD factorisation in some detail using the example of $B \to D\pi$ decays. In this case the $b \to u$ transition current is replaced by a heavy-heavy $b \to c$ current. This case is somewhat simpler than $B \to \pi\pi$ since the spectator interaction (the T^{II} term, bottom line of Figure 13) does not contribute to leading power. This is because for a heavy-to-heavy transition the spectator quark, and hence the gluon attached to it, is always soft. This leads to a suppression, according to the colour transparency argument, when this gluon couples to the emitted pion. Also penguin contributions are absent for $B \to D\pi$. We shall illustrate explicitly how factorisation emerges at the one-loop order in this specific case, and in the heavy-quark limit, defined as m_b, $m_c \gg \Lambda_{\mathrm{QCD}}$ with m_c/m_b fixed.

Further details on QCD factorisation in B decays and additional literature can be found in the articles by Beneke et al. (1999, 2000 & 2001).

6.2 $B \to D\pi$: factorisation to one-loop order

6.2.1 Preliminaries

The effective Hamiltonian relevant for $B \to D\pi$ can be written as

$$\mathcal{H}_{\text{eff}} = \frac{G_F}{\sqrt{2}} V_{ud}^* V_{cb} \left(C_0 O_0 + C_8 O_8 \right), \tag{118}$$

with the operators

$$O_0 = \bar{c}\gamma^\mu (1 - \gamma_5) b \, \bar{d}\gamma_\mu (1 - \gamma_5) u, \tag{119}$$

$$O_8 = \bar{c}\gamma^\mu (1 - \gamma_5) T^a b \, \bar{d}\gamma_\mu (1 - \gamma_5) T^a u. \tag{120}$$

Here we have chosen to write the two independent operators in the singlet-octet basis, which is most convenient for our purposes, rather than in the more conventional bases of Q_1, Q_2 or Q_+, Q_- (see the discussion in Section 3; because all four quark flavours are different in (118), penguin operators are absent). The Wilson coefficients C_0, C_8 have been calculated at next-to-leading order in renormalisation-group improved perturbation theory (Altarelli et al. 1981, Buras & Weisz 1990) and are given by

$$C_0 = \frac{N_c + 1}{2N_c} C_+ + \frac{N_c - 1}{2N_c} C_-, \qquad C_8 = C_+ - C_-, \tag{121}$$

where

$$C_\pm(\mu) = \left(1 + \frac{\alpha_s(\mu)}{4\pi} B_\pm \right) \bar{C}_\pm(\mu), \tag{122}$$

$$\bar{C}_\pm(\mu) = \left[\frac{\alpha_s(M_W)}{\alpha_s(\mu)} \right]^{d_\pm} \left[1 + \frac{\alpha_s(M_W) - \alpha_s(\mu)}{4\pi} (B_\pm - J_\pm) \right]. \tag{123}$$

(The coefficients C_0, C_8 are related to the ones of the standard basis by $C_0 = C_1 + C_2/3$ and $C_8 = 2C_2$.) We employ the next-to-leading order expression for the running coupling,

$$\alpha_s(\mu) = \frac{4\pi}{\beta_0 \ln(\mu^2/\Lambda_{\text{QCD}}^2)} \left[1 - \frac{\beta_1}{\beta_0^2} \frac{\ln \ln(\mu^2/\Lambda_{\text{QCD}}^2)}{\ln(\mu^2/\Lambda_{\text{QCD}}^2)} \right], \tag{124}$$

$$\beta_0 = \frac{11N_c - 2f}{3}, \qquad \beta_1 = \frac{34}{3}N_c^2 - \frac{10}{3}N_c f - 2C_F f, \qquad C_F = \frac{N_c^2 - 1}{2N_c}, \tag{125}$$

where N_c is the number of colours, and f the number of light flavours. $\Lambda_{\text{QCD}} \equiv \Lambda_{\overline{\text{MS}}}^{(f)}$ is the QCD scale in the $\overline{\text{MS}}$ scheme with f flavours. Next we have

$$d_\pm = \frac{\gamma_\pm^{(0)}}{2\beta_0}, \qquad \gamma_\pm^{(0)} = \pm 12 \frac{N_c \mp 1}{2N_c}, \qquad B_\pm = \pm \frac{N_c \mp 1}{2N_c} B. \tag{126}$$

The general definition of $J_\pm$ may be found in (Buchalla, Buras & Lautenbacher 1996). Numerically, for $N_c = 3$ and $f = 5$

$$d_\pm = \begin{cases} \dfrac{6}{23}, \\[2mm] -\dfrac{12}{23}, \end{cases} \qquad B_\pm - J_\pm = \begin{cases} \dfrac{6473}{3174}, \\[2mm] -\dfrac{9371}{1587}. \end{cases} \tag{127}$$
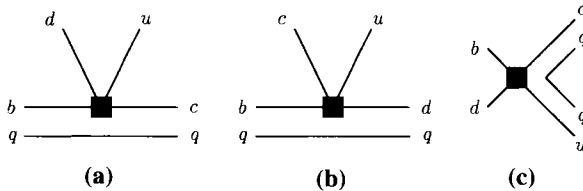
Figure 14. *Basic quark-level topologies for $B \to D\pi$ decays $(q = u, d)$: (a) class-I, (b) class-II, (c) weak annihilation. $\bar{B}_d \to D^+\pi^-$ receives contributions from (a) and (c), $\bar{B}_d \to D^0\pi^0$ from (b) and (c), and $B^- \to D^0\pi^-$ from (a) and (b). Only (a) contributes in the heavy-quark limit.*

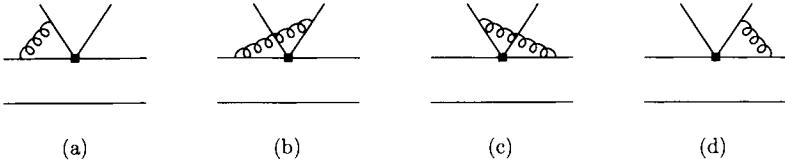The quantities β_0, β_1, $d_\pm$, $B_\pm - J_\pm$ are scheme independent. The scheme dependence of the coefficients at next-to-leading order is parametrised by $B_\pm$ in (122). In the naive dimensional regularisation (NDR) and 't Hooft-Veltman (HV) schemes, this scheme dependence is expressed in a single number B with $B_{\mathrm{NDR}} = 11$ and $B_{\mathrm{HV}} = 7$. The dependence of the Wilson coefficients on the renormalisation scheme and scale is cancelled by a corresponding scale and scheme dependence of the hadronic matrix elements of the operators O_0 and O_8.

Before continuing with a discussion of these matrix elements, it is useful to consider the flavour structure for the various contributions to $B \to D\pi$ decays. The possible quark-level topologies are depicted in Figure 14. In the terminology generally adopted for two-body non-leptonic decays, the decays $\bar{B}_d \to D^+\pi^-$, $\bar{B}_d \to D^0\pi^0$ and $B^- \to D^0\pi^-$ are referred to as class-I, class-II and class-III, respectively. In both $\bar{B}_d \to D^+\pi^-$ and $B^- \to D^0\pi^-$ decays the pion can be directly created from the weak current. We may call this a class-I contribution, following the above terminology. In addition, in the case of $\bar{B}_d \to D^+\pi^-$ there is a contribution from weak annihilation and a class-II amplitude contributes to $B^- \to D^0\pi^-$, see Figure 14. The important point is that the spectator quark goes into the light meson in the case of the class-II amplitude. This amplitude is suppressed in the heavy-quark limit, as is the annihilation amplitude. The amplitude for $\bar{B}_d \to D^0\pi^0$, receiving only class-II and annihilation contributions, is therefore subleading compared with $\bar{B}_d \to D^+\pi^-$ and $B^- \to D^0\pi^-$, which are dominated by the class-I topology. The treatment of this leading class-I mechanism will be the main subject of the following sections. We shall use the one-loop analysis for $\bar{B}_d \to D^+\pi^-$ as a concrete example on which we will illustrate explicitly how the factorisation formula can be derived.

6.2.2   Soft and collinear cancellations at one-loop

In order to demonstrate the property of factorisation for $\bar{B}_d \to D^+\pi^-$, we have to analyse the "non-factorisable" one-gluon exchange contributions (Figure 15) to the $b \to c\bar{u}d$ transition. We consider the leading, valence Fock state of the emitted pion. This is justified since higher Fock components only give power-suppressed contributions to the decay amplitude in the heavy-quark limit. The valence Fock state of the pion can be written as

$$|\pi(q)\rangle = \int \frac{du}{\sqrt{u\bar{u}}} \frac{d^2 l_\perp}{16\pi^3} \frac{1}{\sqrt{2N_c}} \left(a_\uparrow^\dagger(l_q)b_\downarrow^\dagger(l_{\bar{q}}) - a_\downarrow^\dagger(l_q)b_\uparrow^\dagger(l_{\bar{q}}) \right) |0\rangle \, \Psi(u, \mathbf{1}_\perp), \qquad (128)$$

(a)   (b)   (c)   (d)

Figure 15. *"Non-factorisable" vertex corrections.*

where $a_s^\dagger$ ($b_s^\dagger$) denotes the creation operator for a quark (antiquark) in a state with spin $s = \uparrow$ or $s = \downarrow$, and we have suppressed colour indices. This representation of the pion state is adequate for a leading-power analysis. The wave function $\Psi(u, \mathbf{l}_\perp)$ is defined as the amplitude for the pion to be composed of two on-shell quarks, characterised by longitudinal momentum fraction u and transverse momentum $l_\perp$. The on-shell momenta ($l_{q,\bar{q}}^2 = 0$) of the quark (l_q) and the antiquark ($l_{\bar{q}}$) are given by

$$l_q = uq + l_\perp + \frac{\mathbf{l}_\perp^2}{4uE} n_-, \qquad l_{\bar{q}} = \bar{u}q - l_\perp + \frac{\mathbf{l}_\perp^2}{4\bar{u}E} n_-. \tag{129}$$

In these expressions $q = E(1, 0, 0, 1)$ is the pion momentum, $E = p_B \cdot q/m_B$ the pion energy and $n_- = (1, 0, 0, -1)$. Furthermore $l_\perp \cdot q = l_\perp \cdot n_- = 0$. For the purpose of power counting $l_\perp \sim \Lambda_{\rm QCD} \ll E \sim m_b$. Note that the invariant mass of the valence state is $(l_q + l_{\bar{q}})^2 = \mathbf{l}_\perp^2/(u\bar{u})$, which is of order $\Lambda_{\rm QCD}^2$ and hence negligible in the heavy-quark limit, unless u is in the vicinity of the endpoints (0 or 1). In this case the invariant mass of the quark-antiquark pair becomes large and the valence Fock state is no longer a valid representation of the pion. However, in the heavy-quark limit the dominant contributions to the decay amplitude come from configurations where both partons are hard (u and $\bar{u}$ both of order 1) and the two-particle Fock state yields a consistent description. The suppression of the soft regions (u or $\bar{u} \ll 1$) is related to the endpoint behaviour of the pion wave function. We will provide an explicit consistency check of this important feature later on.

As a next step we write down the amplitude

$$\langle \pi(q) | u(0)_\alpha \bar{d}(y)_\beta | 0 \rangle = \int du \frac{d^2 l_\perp}{16\pi^3} \frac{1}{\sqrt{2N_c}} \Psi^*(u, \mathbf{l}_\perp) (\gamma_5 \slashed{q})_{\alpha\beta} \exp\{il_q \cdot y\}, \tag{130}$$

which appears as an ingredient of the $B \to D\pi$ matrix element. The right-hand side of (130) follows directly from (128). Using (130) it is straightforward to write down the one-gluon exchange contribution to the $B \to D\pi$ matrix element of the operator O_8 (Figure 15). We have

$$(D^+\pi^- | O_8 | \bar{B}_d)_{1-\text{gluon}} = \tag{131}$$

$$ig_s^2 \frac{C_F}{2} \int \frac{d^4 k}{(2\pi)^4} (D^+ | \bar{c} A_1(k) b | \bar{B}_d) \frac{1}{k^2} \int_0^1 du \frac{d^2 l_\perp}{16\pi^3} \frac{\Psi^*(u, \mathbf{l}_\perp)}{\sqrt{2N_c}} \operatorname{tr}[\gamma_5 \slashed{q} A_2(l_q, l_{\bar{q}}, k)],$$

where

$$A_1(k) = \frac{\gamma^\lambda(\slashed{p}_c - \slashed{k} + m_c)\Gamma}{2p_c \cdot k - k^2} - \frac{\Gamma(\slashed{p}_b + \slashed{k} + m_b)\gamma^\lambda}{2p_b \cdot k + k^2}, \tag{132}$$

$$A_2(l_q, l_{\bar{q}}, k) = \frac{\Gamma(\slashed{l}_{\bar{q}} + \slashed{k})\gamma_\lambda}{2l_{\bar{q}} \cdot k + k^2} - \frac{\gamma_\lambda(\slashed{l}_q + \slashed{k})\Gamma}{2l_q \cdot k + k^2}. \tag{133}$$

Here $\Gamma = \gamma^\mu(1-\gamma_5)$ and p_b, p_c are the momenta of the b quark and the c quark, respectively. Note that this expression holds in an arbitrary covariant gauge. The gauge-parameter dependent part of the gluon propagator gives no contribution to (131), as can be seen from (132) and (133). There is no correction to the matrix element of O_0 at order α_s, because in this case the $(d\bar{u})$ pair is necessarily in a colour-octet configuration and cannot form a pion.

In (131) the pion wave function $\Psi(u, l_\perp)$ appears separated from the $B \to D$ transition. This is merely a reflection of the fact that we have represented the pion state by (128). It does not, by itself, imply factorisation, since the right-hand side of (131) involves still nontrivial integrations over $l_\perp$ and gluon momentum k, and long- and short-distance contributions are not yet disentangled. In order for (131) to make sense we need to show that the integral over k receives only subdominant contributions from the region of small k^2. This is equivalent to showing that the integral over k does not contain infrared divergences at leading power in $1/m_b$.

To demonstrate infrared finiteness of the one-loop integral

$$J \equiv \int d^4k \, \frac{1}{k^2} \, A_1(k) \otimes A_2(l_q, l_{\bar{q}}, k) \tag{134}$$

at leading power, the heavy-quark limit and the corresponding large light-cone momentum of the pion are again essential. First note that when k is of order m_b, $J \sim 1$ for dimensional reasons. Potential infrared divergences could arise when k is soft or when k is collinear to the pion momentum q. We need to show that the contributions from these regions are power suppressed in m_b. (Note that we do not need to show that J is infrared finite. It is enough that logarithmic divergences have coefficients that are power suppressed.)

We treat the soft region of integration first. Here all components of k become small simultaneously, which we describe by scaling $k \sim \lambda$. Counting powers of λ ($d^4k \sim \lambda^4$, $1/k^2 \sim \lambda^{-2}$, $1/p \cdot k \sim \lambda^{-1}$) reveals that each of the four diagrams (corresponding to the four terms in the product in (134)) is logarithmically divergent. However, because k is small, the integrand can be simplified. For instance, the second term in A_2 can be approximated as

$$\frac{\gamma_\lambda(\slashed{l}_q + \slashed{k})\Gamma}{2l_q \cdot k + k^2} = \frac{\gamma_\lambda \left(u\slashed{q} + \slashed{l}_\perp + \left(\dfrac{l_\perp^2}{4uE}\right)\slashed{n}_- + \slashed{k}\right)\Gamma}{2uq \cdot k + 2l_\perp \cdot k + \left(\dfrac{l_\perp^2}{2uE}\right)n_- \cdot k + k^2} \simeq \frac{q_\lambda}{q \cdot k}\Gamma, \tag{135}$$

where we used the fact that $\slashed{q}$ to the extreme left or right of an expression gives zero due to the on-shell condition for the external quark lines. We get exactly the same expression but with an opposite sign from the other term in A_2 and hence the soft divergence cancels out when adding the two terms in A_2. More precisely, we find that the integral is infrared finite in the soft region when $l_\perp$ is neglected. When $l_\perp$ is not neglected, there is a divergence from soft k which is proportional to $l_\perp^2/m_b^2 \sim \Lambda_{\text{QCD}}^2/m_b^2$. In either case, the soft contribution to J is of order Λ_{QCD}/m_b or smaller and hence suppressed relative to the hard contribution. This corresponds to the standard soft cancellation mechanism, which is a technical manifestation of colour transparency.

Each of the four terms in (134) is also divergent when k becomes collinear with the light-cone momentum q. It is convenient to introduce, for any four-vector v, the light-cone

components

$$v_\pm = \frac{v^0 \pm v^3}{\sqrt{2}}. \tag{136}$$

The collinear region is then described by the scaling

$$k^+ \sim \lambda^0, \quad k_\perp \sim \lambda, \quad k^- \sim \lambda^2. \tag{137}$$

Then $d^4k \sim dk^+ dk^- d^2k_\perp \sim \lambda^4$ and $q \cdot k = q^+ k^- \sim \lambda^2$, $k^2 = 2k^+ k^- + k_\perp^2 \sim \lambda^2$. The divergence is again logarithmic and it is thus sufficient to consider the leading behaviour in the collinear limit. Writing $k = \alpha q + \ldots$ we can now simplify the second term of A_2 as

$$\frac{\gamma_\lambda (\slashed{l}_q + \slashed{k})\Gamma}{2l_q \cdot k + k^2} \simeq q_\lambda \frac{2(u + \alpha)\Gamma}{2l_q \cdot k + k^2}. \tag{138}$$

No simplification occurs in the denominator (in particular $l_\perp$ cannot be neglected), but the important point is that the leading-power contribution is proportional to q_λ. Therefore, substituting $k = \alpha q$ into A_1 and using $q^2 = 0$, we obtain

$$q_\lambda A_1 \simeq \frac{\slashed{q}(\slashed{p}_c + m_c)\Gamma}{2\alpha p_c \cdot q} - \frac{\Gamma(\slashed{p}_b + m_b)\slashed{q}}{2\alpha p_b \cdot q} = 0, \tag{139}$$

employing the equations of motion for the heavy quarks. Hence the collinearly divergent region is seen to cancel out via the standard collinear Ward identity. This completes the proof of the absence of infrared divergences at leading power in the hard-scattering kernel for $\bar{B}_d \to D^+ \pi^-$ to one-loop order. In other words, we have shown that the "non-factorisable" diagrams of Figure 15 are dominated by hard gluon exchange. The proof at two loops has been given by Beneke et al. (2000) and a proof to all orders by Bauer et al. (2001c).

Since we have now established that the leading contribution to J arises from k of order m_b ("hard" k), and since $|l_\perp| \ll E$, we may expand A_2 in $|l_\perp|/E$. To leading power the expansion simply reduces to neglecting $l_\perp$ altogether, which implies $l_q = uq$ and $l_{\bar{q}} = \bar{u}q$ in (133). As a consequence, we may perform the $l_\perp$ integration in (131) over the pion wave function. Defining

$$\int \frac{d^2 l_\perp}{16\pi^3} \frac{\Psi^*(u, \mathbf{l}_\perp)}{\sqrt{2N_c}} \equiv \frac{if_\pi}{4N_c} \Phi_\pi(u), \tag{140}$$

the matrix element of O_8 in (131) becomes

$$\langle D^+ \pi^- | O_8 | \bar{B}_d \rangle_{1-\text{gluon}} = \tag{141}$$

$$-g_s^2 \frac{C_F}{8N_c} \int \frac{d^4k}{(2\pi)^4} \langle D^+ | \bar{c} A_1(k) b | \bar{B}_d \rangle \frac{1}{k^2} f_\pi \int_0^1 du \, \Phi_\pi(u) \, \text{tr}[\gamma_5 \slashed{q} A_2(uq, \bar{u}q, k)].$$

Putting y on the light-cone in (130), $y^+ = y_\perp = 0$, hence $l_q \cdot y = l_q^+ y^- = uqy$, we see that the $l_\perp$-integrated wave function $\Phi_\pi(u)$ in (140) is precisely the light-cone distribution amplitude of the pion in the usual definition. Indeed, the leading-twist light-cone distribution amplitude for pseudoscalar mesons (P) reads

$$\langle P(q) | \bar{q}(y)_\alpha q'(x)_\beta | 0 \rangle \big|_{(x-y)^2 = 0} = \frac{if_P}{4} (\slashed{q}\gamma_5)_{\beta\alpha} \int_0^1 du \, e^{i(\bar{u}qx + uqy)} \Phi_P(u, \mu). \tag{142}$$

Here it is understood that the operator on the left-hand side is a colour singlet. We use the "bar"-notation, i.e. $\bar{v} \equiv 1 - v$ for any longitudinal momentum fraction variable. The parameter μ is the renormalisation scale of the light-cone operators on the left-hand side. The distribution amplitude is normalised as $\int_0^1 du\, \Phi_P(u, \mu) = 1$. One defines the asymptotic distribution amplitude as the limit in which the renormalisation scale is sent to infinity. The asymptotic form is

$$\Phi_P(u, \mu) \stackrel{\mu \to \infty}{=} 6u\bar{u}. \tag{143}$$

The decay constant appearing in (142) refers to the normalisation in which $f_\pi = 131$ MeV. There is a path-ordered exponential that connects the two quark fields at different positions and makes the light-cone operators gauge invariant. In (142) we have suppressed this standard factor.

This derivation demonstrates the relevance of the light-cone wave function for the factorisation formula. Note that the collinear approximation for the quark-antiquark momenta emerges automatically in the heavy-quark limit.

After the k integral is performed, the expression (141) can be cast into the form

$$\langle D^+ \pi^- | O_8 | \bar{B}_d \rangle_{1-\text{gluon}} \sim F_{B \to D}(0) \int_0^1 du\, T_8(u, z) \Phi_\pi(u), \tag{144}$$

where $z = m_c/m_b$, $T_8(u, z)$ is the hard-scattering kernel, and $F_{B \to D}(0)$ the form factor that parametrises the $\langle D^+ | \bar{c}[\ldots]b | \bar{B}_d \rangle$ matrix element. The result for $T_8(u, z)$ will be given in the following section.

6.2.3 Matrix elements at next-to-leading order

As we have seen above, the $\bar{B}_d \to D^+ \pi^-$ amplitude factorises in the heavy-quark limit into a matrix element of the form $\langle D^+ | \bar{c}[\ldots]b | \bar{B}_d \rangle$ for the $B \to D$ transition and a matrix element $\langle \pi^- | \bar{d}(x)[\ldots]u(0) | 0 \rangle$ with $x^2 = 0$ that gives rise to the pion light-cone distribution amplitude. Leaving aside power-suppressed contributions, the essential requirement for this conclusion was the absence of both soft and collinear infrared divergences in the gluon exchange between the $(\bar{c}b)$ and $(\bar{d}u)$ currents. This gluon exchange is therefore calculable in QCD perturbation theory. We now present these corrections explicitly to order α_s.

The effective Hamiltonian (118) can be written as

$$\mathcal{H}_{\text{eff}} = \frac{G_F}{\sqrt{2}} V_{ud}^* V_{cb} \left\{ \left[\frac{N_c + 1}{2N_c} \bar{C}_+(\mu) + \frac{N_c - 1}{2N_c} \bar{C}_-(\mu) + \frac{\alpha_s(\mu)}{4\pi} \frac{C_F}{2N_c} BC_8(\mu) \right] O_0 \right.$$
$$\left. + [C_8(\mu)] O_8 \right\}, \tag{145}$$

where the scheme-dependent terms in the coefficient of the operator O_0, proportional to the constant B defined after (127), have been written explicitly.

Schematically, the matrix elements of both O_0 and O_8 can be expressed in a form analogous to (144). Because the light-quark pair has to be in a colour singlet to produce the pion in the leading Fock state, only O_0 gives a contribution to zeroth order in α_s. Similarly, to first order in α_s only O_8 can contribute. The result of computing the diagrams

in Figure 15 with an insertion of O_8 can be written in a form that holds simultaneously for $H = D, D^*$ and $L = \pi, \rho$, using only that the $(\bar{u}d)$ pair is a colour singlet and that the external quarks can be taken on-shell. One obtains ($z = m_c/m_b$)

$$\langle H(p')L(q)|O_8|\bar{B}_d(p)\rangle = \frac{\alpha_s}{4\pi}\frac{C_F}{2N_c} if_L \int_0^1 du\,\Phi_L(u) \tag{146}$$

$$\times \left[-\left(6\ln\frac{\mu^2}{m_b^2}+B\right)(\langle J_V\rangle - \langle J_A\rangle) + F(u,z)\,\langle J_V\rangle - F(u,-z)\,\langle J_A\rangle\right],$$

where

$$\langle J_V\rangle = \langle H(p')|\bar{c}\slashed{q}\,b|\bar{B}_d(p)\rangle, \qquad \langle J_A\rangle = \langle H(p')|\bar{c}\slashed{q}\gamma_5 b\,|\bar{B}_d(p)\rangle. \tag{147}$$

In obtaining (146) we have used the equations of motion for the quarks to reduce the operator basis to J_V and J_A. The form of (146) is identical for pions and longitudinally polarised ρ mesons. The production of transversely polarised ρ mesons is power suppressed in Λ_{QCD}/m_b.

In the case of a distribution amplitude $\Phi_L(u)$ that is symmetric under $u \leftrightarrow \bar{u}$, which is relevant for $L = \pi, \rho$, the function $F(u, z)$ appearing in (146) can be compactly written as

$$F(u,z) = 3\ln z^2 - 7 + f(u,z) + f(u,1/z), \tag{148}$$

with

$$f(u,z) = -\frac{u(1-z^2)[3(1-u(1-z^2)) + z]}{[1-u(1-z^2)]^2}\ln[u(1-z^2)] - \frac{z}{1-u(1-z^2)}. \tag{149}$$

The contribution of $f(u, z)$ in (148) comes from the first two diagrams in Figure 15 with the gluon coupling to the b quark, whereas $f(u, 1/z)$ arises from the last two diagrams with the gluon coupling to the charm quark. The absorptive part of the amplitude, which is responsible for the occurrence of strong rescattering phases, arises from $f(u, 1/z)$ and can be obtained by recalling that z^2 is $z^2 - i\epsilon$ with $\epsilon > 0$ infinitesimal. We then have

$$\frac{1}{\pi}\text{Im}\,F(u,z) = -\frac{(1-u)(1-z^2)[3(1-u(1-z^2)) + z]}{[1-u(1-z^2)]^2}. \tag{150}$$

As mentioned above, (146) is applicable to all decays of the type $\bar{B}_d \to D^{(*)+}L^-$, where L is a light hadron such as a pion or a (longitudinally polarised) ρ meson. Only the operator J_V contributes to $\bar{B}_d \to D^+L^-$, and only J_A contributes to $\bar{B}_d \to D^{*+}L^-$. (Due to helicity conservation the vector current $B \to D^*$ matrix element contributes only in conjunction with a transversely polarised ρ meson and hence is power suppressed in the heavy-quark limit.) Our final result can therefore be written as

$$\langle D^+L^-|O_{0,8}|\bar{B}_d\rangle = \langle D^+|\bar{c}\gamma^\mu(1-\gamma_5)b|\bar{B}_d\rangle \cdot if_L q_\mu \int_0^1 du\, T_{0,8}(u,z)\,\Phi_L(u), \tag{151}$$

where $L = \pi, \rho$, and the hard-scattering kernels are

$$T_0(u,z) = 1 + O(\alpha_s^2), \tag{152}$$

$$T_8(u,z) = \frac{\alpha_s}{4\pi}\frac{C_F}{2N_c}\left[-6\ln\frac{\mu^2}{m_b^2} - B + F(u,z)\right] + O(\alpha_s^2). \tag{153}$$

When the D meson is replaced by a D^* meson, the result is identical except that $F(u, z)$ in (153) must be replaced by $F(u, -z)$. Since no order α_s corrections exist for O_0, the matrix element retains its leading-order factorised form

$$\langle D^+ L^- | O_0 | \bar{B}_d \rangle = i f_L q_\mu \langle D^+ | \bar{c} \gamma^\mu (1 - \gamma_5) b | \bar{B}_d \rangle \qquad (154)$$

to this accuracy. From (149) it follows that $T_8(u, z)$ tends to a constant as u approaches the endpoints ($u \to 0, 1$). Therefore the contribution to (151) from the endpoint region is suppressed, both by phase space and by the endpoint suppression intrinsic to $\Phi_L(u)$, which can be represented as

$$\Phi_L(u) = 6u(1 - u) \left[1 + \sum_{n=1}^{\infty} \alpha_n^L(\mu) \, C_n^{3/2}(2u - 1) \right], \qquad (155)$$

vanishing $\sim u$ ($\sim (1 - u)$) at the endpoints. Here $C_1^{3/2}(x) = 3x$, $C_2^{3/2}(x) = \frac{3}{2}(5x^2 - 1)$, etc. are Gegenbauer polynomials. The Gegenbauer moments $\alpha_n^L(\mu)$ are nonperturbative quantities. They are multiplicatively renormalised and approach zero as $\mu \to \infty$. (The scale dependence of these quantities enters the results for the coefficients only at order α_s^2, which is beyond the accuracy of a NLO calculation.)

As a consequence of the endpoint suppression the emitted light meson is indeed dominated by energetic constituents, as required for the self-consistency of the factorisation formula (151).

Combining (145), (151), (152) and (153), we obtain our final result for the class-I, non-leptonic $\bar{B}_d \to D^{(*)+} L^-$ decay amplitudes in the heavy-quark limit, and at next-to-leading order in α_s. The results can be compactly expressed in terms of the matrix elements of a "transition operator"

$$\mathcal{T} = \frac{G_F}{\sqrt{2}} V_{ud}^* V_{cb} \left[a_1(DL) \, Q_V - a_1(D^*L) \, Q_A \right], \qquad (156)$$

where

$$Q_V = \bar{c} \gamma^\mu b \otimes \bar{d} \gamma_\mu (1 - \gamma_5) u, \qquad Q_A = \bar{c} \gamma^\mu \gamma_5 b \otimes \bar{d} \gamma_\mu (1 - \gamma_5) u, \qquad (157)$$

and hadronic matrix elements of $Q_{V,A}$ are understood to be evaluated in factorised form, i.e.

$$\langle DL | j_1 \otimes j_2 | \bar{B} \rangle \equiv \langle D | j_1 | \bar{B} \rangle \langle L | j_2 | 0 \rangle. \qquad (158)$$

Equation (156) defines the quantities $a_1(D^{(*)}L)$, including the leading "non-factorisable" corrections, in a renormalisation-scale and -scheme independent way. To leading power in Λ_{QCD}/m_b these quantities should not be interpreted as phenomenological parameters (as is usually done), because they are dominated by hard gluon exchange and thus calculable in QCD. At next-to-leading order we get

$$\begin{aligned} a_1(DL) &= \frac{N_c + 1}{2N_c} \bar{C}_+(\mu) + \frac{N_c - 1}{2N_c} \bar{C}_-(\mu) \\ &\quad + \frac{\alpha_s}{4\pi} \frac{C_F}{2N_c} C_8(\mu) \left[-6 \ln \frac{\mu^2}{m_b^2} + \int_0^1 du \, F(u, z) \, \Phi_L(u) \right], \qquad (159) \\ a_1(D^*L) &= \frac{N_c + 1}{2N_c} \bar{C}_+(\mu) + \frac{N_c - 1}{2N_c} \bar{C}_-(\mu) \\ &\quad + \frac{\alpha_s}{4\pi} \frac{C_F}{2N_c} C_8(\mu) \left[-6 \ln \frac{\mu^2}{m_b^2} + \int_0^1 du \, F(u, -z) \, \Phi_L(u) \right]. \qquad (160) \end{aligned}$$

These expressions generalise the well-known leading-order formula

$$a_1^{\text{LO}} = \frac{N_c + 1}{2N_c} C_+^{\text{LO}}(\mu) + \frac{N_c - 1}{2N_c} C_-^{\text{LO}}(\mu). \tag{161}$$

We observe that the scheme dependence, parametrised by B, is cancelled between the coefficient of O_0 in (145) and the matrix element of O_8 in (151). Likewise, the μ dependence of the terms in brackets in (159) and (160) cancels against the scale dependence of the coefficients $\bar{C}_\pm(\mu)$, ensuring a consistent physical result at next-to-leading order in QCD.

The coefficients $a_1(DL)$ and $a_1(D^*L)$ are seen to be non-universal, i.e. they are explicitly dependent on the nature of the final-state mesons. This dependence enters via the light-cone distribution amplitude $\Phi_L(u)$ of the light emission meson and via the analytic form of the hard-scattering kernel ($F(u, z)$ vs. $F(u, -z)$). However, the non-universality enters only at next-to-leading order.

Exercise

Verify that the coefficients $a_1(D^{(*)}L)$ in (159) and (160) are independent of the unphysical renormalisation scale μ up to terms of $\mathcal{O}(\alpha_s^2)$.

6.2.4 Phenomenological applications for $B \to D\pi$

An important test of QCD factorisation can be performed by comparing the hadronic decays $\bar{B}_d \to D^{(*)+}L^-$ with the semileptonic processes $\bar{B}_d \to D^{(*)+}l^-\nu$. It is useful to introduce the ratios

$$R_L^{(*)} = \frac{\Gamma(\bar{B}_d \to D^{(*)+}L^-)}{\left.\dfrac{d\Gamma(\bar{B}_d \to D^{(*)+}l^-\bar{\nu})}{dq^2}\right|_{q^2 = m_L^2}} = 6\pi^2 |V_{ud}|^2 f_L^2 |a_1(D^{(*)}L)|^2 X_L^{(*)}, \tag{162}$$

where $X_\rho = X_\rho^* = 1$ for a vector meson (because the production of the lepton pair via a $V - A$ current in semi-leptonic decays is kinematically equivalent to that of a vector meson with momentum q), whereas X_π and X_π^* deviate from 1 only by (calculable) terms of order m_π^2/m_B^2, which numerically are below 1%. The main virtue of (162) is that the $B \to D^{(*)}$ form factors cancel in the ratio. The theoretical prediction for the QCD coefficients, based on QCD factorisation to leading power and at NLO in perturbative QCD, is $|a_1(D^{(*)}L)| = 1.05$. The uncertainty of this leading-power result is small, about ± 0.01. The prediction is to be compared with the experimental results, extracted from (162), which read $|a_1(D^*\pi)| = 1.08 \pm 0.07$, $|a_1(D^*\rho)| = 1.09 \pm 0.10$ and $|a_1(D^*a_1)| = 1.08 \pm 0.11$. These values show fair agreement with the theoretical number, albeit within experimental uncertainties that are still large.

Another interesting consideration concerns the comparison of class-I modes with those of class II and III. For $B \to D^{(*)}\pi$, all three decay modes, which are related by isospin, have been measured. A nice discussion of the present experimental status and its interpretation in the context of QCD factorisation has been given by Neubert & Petrov (2001). Let us briefly discuss here a few important aspects. The experimental status is

	$\bar{B} \to D\pi$	$\bar{B} \to D^*\pi$
$\bar{B}^0 \to D^{(*)+}\pi^-$	3.0 ± 0.4	2.76 ± 0.21
$\bar{B}^0 \to D^{(*)0}\pi^0$	0.27 ± 0.05	0.17 ± 0.05
$B^- \to D^{(*)0}\pi^-$	5.3 ± 0.5	4.6 ± 0.4
$|C - A|/|T + A|$	0.42 ± 0.05	0.35 ± 0.05
δ_{TC}	$(56 \pm 20)^\circ$	$(51 \pm 20)^\circ$

Table 3. *Experimental data for $\bar{B} \to D^{(*)}\pi$ branching ratios (in units of 10^{-3}) and extracted parameters $|C - A|/|T + A|$, δ_{TC} (see Neubert & Petrov (2001)).*

summarised in Table 3. Denoting the basic topologies from Figures 14(a,b,c) by T, C and A, respectively (where the notation refers to "tree", "colour-suppressed tree" and "annihilation"), we have

$$\mathcal{A}(\bar{B}_d \to D^+\pi^-) = T + A, \tag{163}$$
$$\sqrt{2}\,\mathcal{A}(\bar{B}_d \to D^0\pi^0) = C - A, \tag{164}$$
$$\mathcal{A}(B^- \to D^0\pi^-) = T + C. \tag{165}$$

For later use we may further define the spectator-interaction contribution to T, T_{spec} (see bottom row of Figure 13). A similar decomposition holds for the $\bar{B} \to D^*\pi$ modes. Isospin symmetry is reflected in the amplitude relation $\mathcal{A}(\bar{B}_d \to D^+\pi^-) + \sqrt{2}\,\mathcal{A}(\bar{B}_d \to D^0\pi^0) = \mathcal{A}(B^- \to D^0\pi^-)$, which is manifest in the parametrisation of (163) – (165). This means that there are only two independent amplitudes, which we can take to be $(T + A)$ and $(C - A)$. These amplitudes are complex due to strong phases from final-state interactions. Only the relative phase of the two independent amplitudes is an observable and we define δ_{TC} to be the relative phase of $(T + A)$ and $(C - A)$. The phase can then be extracted from the data through the relation

$$\cos\delta_{TC} = \frac{\left(\dfrac{\tau(\bar{B}^0)}{\tau(B^-)}\right) B(B^- \to D^0\pi^-) - B(\bar{B}^0 \to D^+\pi^-) - 2B(\bar{B}^0 \to D^0\pi^0)}{2\sqrt{2}\sqrt{B(\bar{B}^0 \to D^+\pi^-)\,B(\bar{B}^0 \to D^0\pi^0)}} \tag{166}$$

In the usual heavy-quark limit, $m_b \sim m_c \gg \Lambda_{\text{QCD}}$, only T is calculable. T_{spec}, C and A are not, but they are power suppressed. It is instructive to consider the alternative limit $m_b \gg m_c \gg \Lambda_{\text{QCD}}$, which allows us to distinguish between m_b and m_c (Beneke et al. 2000). In this case, due to $m_c \ll m_b$, also the D becomes a "light" meson and in that respect the process is similar to $B \to \pi\pi$, where both T_{spec} and C are calculable. However, since $m_c \gg \Lambda_{\text{QCD}} \equiv \Lambda$, the D-meson wave function $\Phi_D(u)$ is highly asymmetric and strongly peaks at $\bar{u} \equiv (1 - u) \sim \Lambda/m_c$, where u is the c-quark momentum fraction. These properties can be used to derive the scaling rules

$$\frac{A}{T} \sim \frac{\Lambda}{m_b}, \qquad \frac{T_{\text{spec}}}{T} \sim \frac{\Lambda}{m_c}, \qquad \frac{C}{T} \sim \frac{\Lambda}{m_c}. \tag{167}$$

The amplitude A is still not calculable in this scheme, while T, T_{spec} and C are. Note that from (167) we can recover the two standard scenarios we have been discussing: In the heavy-quark limit $m_c \sim m_b$ (167) reduces to a simple power suppression $\sim \Lambda/m_b$ for A, T_{spec} and C compared to T. On the other hand, if m_c becomes a truly light quark, corresponding to the case of $B \to \pi\pi$, we count $m_c \sim \Lambda$ and see that both T_{spec} and C are of the same order as T, while A is still power suppressed.

The general scenario $m_b \gg m_c \gg \Lambda_{QCD}$ allows us to interpret the experimental results in Table 3. We can even make numerical estimates for T, T_{spec} and C based on the factorisation formula for light-light final states (116). These are somewhat model dependent because Φ_D is not known at present. It is not too difficult to accommodate substantial values $|C - A|/|T + A| \sim 0.2 - 0.3$ and $\delta_{TC} \sim 40°$, in qualitative agreement with Table 3. Given the special role of the charm quark (not light, but also not too heavy), the current experimental situation is not in contradiction with QCD factorisation in the large-m_b limit. For a comparison with experiment it is useful to keep in mind that, according to (167), the suppression of C over T is only $\sim \Lambda/m_c$ (not Λ/m_b) and that δ_{TC} is naturally expected to be of order one.

6.3   CP violation in $B \to \pi^+\pi^-$ decay

Hadronic B decays into a pair of light mesons, such as $B \to \pi K$ or $B \to \pi\pi$, have a very rich phenomenology. Their main interest lies in their sensitivity to short-distance flavour physics, including CKM parameters, CP violation and the search for new physics. By way of an outlook we mention here the important example of CP violation in $B \to \pi^+\pi^-$ decay. The starting point for computing the required decay amplitudes is the effective Hamiltonian in (32). The needed matrix elements of the operators Q_i can be analysed within QCD factorisation using (116). We will not go into the technical details of such an analysis and the discussion of limitations of the approach, in particular from power corrections in Λ_{QCD}/m_b. These can be found in (Beneke et al. 2001). Here we just want to present the phenomenological motivation and to illustrate that a theoretical approach towards a direct dynamical calculation of hadronic matrix elements will be very valuable, even if it necessitates some approximations.

The observable of interest is the time-dependent CP asymmetry in the decays of B^0 and $\bar{B}^0$ to $\pi^+\pi^-$; it is sensitive to the B_d-$\bar{B}_d$ mixing phase $e^{-2i\beta}$. We define

$$A_{CP}^{\pi\pi}(t) = \frac{B(B^0(t) \to \pi^+\pi^-) - B(\bar{B}^0(t) \to \pi^+\pi^-)}{B(B^0(t) \to \pi^+\pi^-) + B(\bar{B}^0(t) \to \pi^+\pi^-)}$$
$$= -S_{\pi\pi} \sin(\Delta m_B t) + C_{\pi\pi} \cos(\Delta m_B t) \tag{168}$$

where

$$S_{\pi\pi} = \frac{2 \, \text{Im} \, \lambda_{\pi\pi}}{1 + |\lambda_{\pi\pi}|^2}, \quad C_{\pi\pi} = \frac{1 - |\lambda_{\pi\pi}|^2}{1 + |\lambda_{\pi\pi}|^2}, \quad \lambda_{\pi\pi} = e^{-2i\beta} \frac{e^{-i\gamma} + P_{\pi\pi}/T_{\pi\pi}}{e^{i\gamma} + P_{\pi\pi}/T_{\pi\pi}}. \tag{169}$$

The amplitudes $T_{\pi\pi}$ ("tree") and $P_{\pi\pi}$ ("penguin") are the components of the $B \to \pi^+\pi^-$ amplitude corresponding to the terms in (32) involving λ_u and λ_c, respectively. In the standard phase conventions λ_c is real and λ_u has a weak phase $-\gamma$, which has been factored out above. The coefficient $C_{\pi\pi}$, which is a function of γ, represents direct CP violation
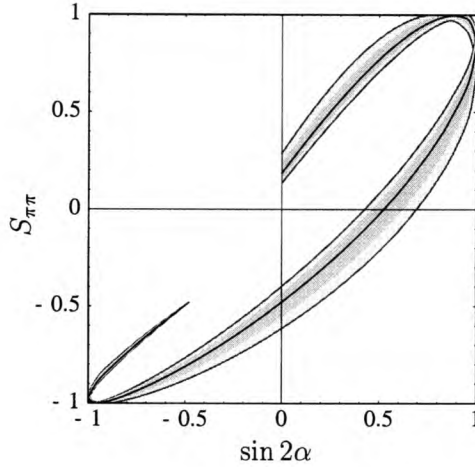
Figure 16. *Relation between* $\sin 2\alpha$ *and the mixing-induced CP asymmetry* $S_{\pi\pi}$, *assuming* $\sin 2\beta = 0.48$. *The dark band reflects parameter variations of the first kind, the light band shows the total theoretical uncertainty. The lower portion of the band refers to values* $45° < \alpha < 135°$, *the upper one to* $0 < \alpha < 45°$ *(right branch) or* $135° < \alpha < 180° - \beta$ *(left branch).*

and is expected to be small. We shall not discuss it further here. The mixing-induced asymmetry $S_{\pi\pi}$ depends on γ and β. In fact, in the limit where $P_{\pi\pi}/T_{\pi\pi}$ is set to zero it follows that $\lambda_{\pi\pi} = e^{-2i(\beta+\gamma)} = e^{2i\alpha}$, and hence $S_{\pi\pi} = \sin 2\alpha$. In this limit $\lambda_{\pi\pi}$ is just the relative weak phase between the direct amplitude $B \to \pi^+\pi^-$ and the one with mixing $B \to \bar{B} \to \pi^+\pi^-$. All dependence on hadronic input has cancelled in this situation. In practice, however, $P_{\pi\pi}/T_{\pi\pi}$ is not fully negligible. It is here that information on hadronic dynamics becomes crucial. QCD factorisation predicts that $P_{\pi\pi}/T_{\pi\pi}$ is suppressed (either by α_s or by powers of $\Lambda_{\rm QCD}/m_b$), because $T_{\pi\pi}$ can arise at tree level, $P_{\pi\pi}$ only through loops. Estimates within this framework give values of about 0.25–0.3.

To illustrate the effect of penguin amplitudes, we first assume that $|V_{ub}/V_{cb}|$ and the weak phase β have been determined accurately. Then using $\gamma = 180° - \alpha - \beta$ the expression for $\lambda_{\pi\pi}$ in (169) becomes a function of α and our prediction for the penguin-to-tree ratio $P_{\pi\pi}/T_{\pi\pi}$. If we further assume that the unitarity triangle lies in the upper half of the $(\bar\rho, \bar\eta)$ plane, then a measurement of $S_{\pi\pi}$ determines $\sin 2\alpha$ with at most a two-fold discrete ambiguity. Figure 16 shows the relation between the two quantities for the particular case where $|V_{ub}/V_{cb}| = 0.085$ and $\beta = 14.3°$, corresponding to $\sin 2\beta = 0.48$. The dark band shows the theoretical uncertainty due to input parameter variations, whereas the light band indicates the total theoretical uncertainty including estimates of the effect of power corrections. We observe that for negative values $\sin 2\alpha$ as preferred by the global analysis of the unitarity triangle, a measurement of the coefficient $S_{\pi\pi}$ could be used to determine $\sin 2\alpha$ with a theoretical uncertainty of about ± 0.1. Interestingly, for such values of $\sin 2\alpha$ the "penguin pollution" effect enhances the value of the mixing-induced CP asymmetry, yielding values of $S_{\pi\pi}$ between -0.5 and -1. Such a large asymmetry should be relatively easy to observe experimentally.
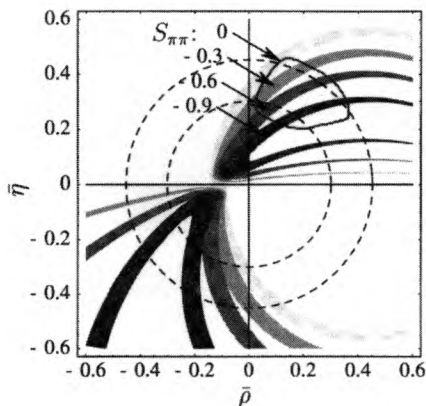
Figure 17. *Allowed regions in the $(\bar{\rho}, \bar{\eta})$ plane corresponding to constant values of the mixing-induced asymmetry $S_{\pi\pi}$ assuming the Standard Model. The widths of the bands reflect the total theoretical uncertainty. The corresponding bands for positive values of $S_{\pi\pi}$ are obtained by a reflection about the $\bar{\rho}$ axis. The light circled area in the upper-right quadrant shows the allowed region obtained from the standard global fit of the unitarity triangle (Höcker et al. 2001).*

Although it illustrates nicely the effect of "penguin pollution" on the determination of $\sin 2\alpha$, Figure 16 is not the most appropriate way to display the constraint on the unitarity triangle implied by a measurement of $S_{\pi\pi}$. In general, there is a four-fold discrete ambiguity in the determination of $\sin 2\alpha$, which we have reduced to a two-fold ambiguity by assuming that the triangle lies in the upper half-plane. Next, and more importantly, we have assumed that $|V_{ub}/V_{cb}|$ and β are known with precision, whereas α is undetermined. However, in the Standard Model $|V_{ub}/V_{cb}|$ and the angles α, β, γ are all functions of the Wolfenstein parameters $\bar{\rho}$ and $\bar{\eta}$. It is thus more appropriate to represent the constraint implied by a measurement of $S_{\pi\pi}$ as a band in the $(\bar{\rho}, \bar{\eta})$ plane. To this end, we write

$$e^{\mp i\gamma} = \frac{\bar{\rho} \mp i\bar{\eta}}{\sqrt{\bar{\rho}^2 + \bar{\eta}^2}}, \qquad e^{-2i\beta} = \frac{(1-\bar{\rho})^2 - \bar{\eta}^2 - 2i\bar{\eta}(1-\bar{\rho})}{(1-\bar{\rho})^2 + \bar{\eta}^2}, \qquad \frac{P_{\pi\pi}}{T_{\pi\pi}} = \frac{r_{\pi\pi}\, e^{i\phi_{\pi\pi}}}{\sqrt{\bar{\rho}^2 + \bar{\eta}^2}}, \qquad (170)$$

where $r_{\pi\pi}\, e^{i\phi_{\pi\pi}}$ is independent of $\bar{\rho}$ and $\bar{\eta}$. We now insert these relations into (169) and draw contours of constant $S_{\pi\pi}$ in the $(\bar{\rho}, \bar{\eta})$ plane. The result is shown by the bands in Figure 17. The widths of the bands reflect the total theoretical uncertainty (including power corrections). For clarity we show only bands for negative values of $S_{\pi\pi}$; those corresponding to positive $S_{\pi\pi}$ values can be obtained by a reflection about the $\bar{\rho}$ axis (i.e., $\bar{\eta} \to -\bar{\eta}$). Note that even a rough measurement of $S_{\pi\pi}$ would translate into a rather narrow band in the $(\bar{\rho}, \bar{\eta})$ plane, which intersects the ring representing the $|V_{ub}/V_{cb}|$ constraint at almost a right angle. In a similar way, the constraint is also quite robust against hadronic uncertainties. Even the approximate knowledge of hadronic matrix elements, as provided by QCD factorisation, will therefore be very valuable and can lead to powerful constraints on the Wolfenstein parameters.

7 Summary

In these lectures we have discussed the theory of heavy quarks, focusing on the important case of B physics. (All methods relying on the heavy-quark limit could in principle be applied to charmed hadrons as well, but they are in most cases much less reliable due to the smaller value of the charm mass.) We shall conclude by summarising the key points.

- A crucial and general idea for dealing with QCD effects is the *factorisation of short-distance and long-distance dynamics*. We have encountered this principle in many different forms and applications:

  - The OPE to construct the effective weak Hamiltonians ($\mathcal{H}_{\text{eff}}^{\Delta B=1,2}$) factorises the short-distance scales of order M_W, m_t from the scales of order m_b.
  - The heavy-quark scale m treated as a short-distance scale can be factorised further from the intrinsic long-distance scale of QCD, Λ_{QCD}. This leads to a systematic expansion of observables simultaneously in $1/m$ and $\alpha_s(m)$ with often very important simplifications. The precise formulation of this class of factorisation depends on the physical situation and can take the form of HQET, LEET, HQE or QCD factorisation in exclusive hadronic B decays.

- HQET exhibits the spin-flavour symmetry of QCD in the heavy-quark limit, which allows us to relate different form factors, and makes the m_Q dependence explicit. Examples of typical applications are $B \to D^{(*)}l\nu$ or f_B.

- HQE is a theory for inclusive B decays. It justifies the "parton model" and allows us to study the nonperturbative power corrections. This is of great use for processes as $B \to X_{c,u}l\nu$, $B \to X_s\gamma$, $B \to X_s l^+l^-$, and the lifetimes of b-flavoured hadrons.

- QCD factorisation, finally, refers to a framework for analysing exclusive hadronic B decays with a fast light meson as for instance $B \to D\pi$, $B \to \pi\pi$, $B \to \pi K$ and $B \to V\gamma$.

With these tools at hand we are in a good position to make full use of the rich experimental results in the physics of heavy flavours. We can determine fundamental parameters of the flavour sector, such as V_{ub}, V_{cb}, V_{td}/V_{ts}, η and $\sin 2\alpha$, and probe electroweak dynamics at the quantum level through $b \to s\gamma$ or $B - \bar{B}$ mixing. This will enable us to thoroughly test the standard model and to learn about new structures and phenomena that are yet to be discovered.

Acknowledgments

References

Albrecht H et al. [ARGUS Collaboration], 1987, Observation of B^0 - $\bar{B}^0$ mixing, *Phys. Lett. B* **192** 245.

Altarelli G, Curci G, Martinelli G and Petrarca S, 1981, QCD nonleading corrections to weak decays as an application of regularization by dimensional reduction, *Nucl. Phys. B* **187** 461.

Anikeev K et al., 2002, *B Physics at the Tevatron: Run II and Beyond*, FERMILAB-Pub-01/197, [*hep-ph*/0201071].

Bauer C W, Fleming S, Pirjol D and Stewart I W, 2001a, An effective field theory for collinear and soft gluons: heavy to light decays, *Phys. Rev. D* **63** 114020 [*hep-ph*/0011336].

Bauer C W, Ligeti Z and Luke M E, 2001b, Precision determination of $|V_{ub}|$ from inclusive decays, *Phys. Rev. D* **64** 113004 [*hep-ph*/0107074].

Bauer C W, Pirjol D and Stewart I W, 2001c, A proof of factorization for $B \to D\pi$, *Phys. Rev. Lett.* **87** 201806 [*hep-ph*/0107002].

Beneke M, Buchalla G, Neubert M and Sachrajda C T, 1999, QCD factorization for $B \to \pi\pi$ decays: Strong phases and CP violation in the heavy quark limit, *Phys. Rev. Lett.* **83** 1914 [*hep-ph*/9905312].

Beneke M, Buchalla G, Neubert M and Sachrajda C T, 2000, QCD factorization for exclusive, non-leptonic B meson decays: General arguments and the case of heavy-light final states, *Nucl. Phys. B* **591** 313 [*hep-ph*/0006124].

Beneke M, Buchalla G, Neubert M and Sachrajda C T, 2001, QCD factorization in $B \to \pi K$, $\pi\pi$ decays and extraction of Wolfenstein parameters, *Nucl. Phys. B* **606** 245 [*hep-ph*/0104110].

Beneke M and Feldmann T, 2001, Symmetry-breaking corrections to heavy-to-light B meson form factors at large recoil, *Nucl. Phys. B* **592** 3 [*hep-ph*/0008255].

Bigi I I, Shifman M A and Uraltsev N, 1997, Aspects of heavy quark theory, *Ann. Rev. Nucl. Part. Sci.* **47** 591 [*hep-ph*/9703290].

Bigi I I and Uraltsev N, 2001, A vademecum on quark hadron duality, *Int. J. Mod. Phys. A* **16** 5201 [*hep-ph*/0106346].

Bigi I I, Uraltsev N G and Vainshtein A I, 1992, Nonperturbative corrections to inclusive beauty and charm decays: QCD versus phenomenological models, *Phys. Lett. B* **293** 430; *Erratum-ibid. B* **297** (1993) 477 [*hep-ph*/9207214].

Bjorken J D, 1989, Topics in B physics, *Nucl. Phys. Proc. Suppl.* **11** 325.

Blok B, Shifman M A and Zhang D X, 1998, An illustrative example of how quark-hadron duality might work, *Phys. Rev. D* **57** 2691; *Erratum-ibid. D* **59** 019901 [*hep-ph*/9709333].

Buchalla G, Buras A J and Lautenbacher M E, 1996, Weak decays beyond leading logarithms, *Rev. Mod. Phys.* **68** 1125 [*hep-ph*/9512380].

Buras A J, 1998, Weak Hamiltonian, CP violation and rare decays, [*hep-ph*/9806471].

Buras A J and Weisz P H, 1990, QCD nonleading corrections to weak decays in dimensional regularization and 't Hooft-Veltman schemes, *Nucl. Phys. B* **333** 66.

Charles J, Le Yaouanc A, Oliver L, Pène O and Raynal J-C, 1999, Heavy-to-light form factors in the final hadron large energy limit of QCD, *Phys. Rev. D* **60** 014001 [*hep-ph*/9812358].

Chay J, Georgi H and Grinstein B, 1990, Lepton energy distributions in heavy meson decays from QCD, *Phys. Lett. B* **247** 399.

Dugan M J and Grinstein B, 1991, QCD basis for factorization in decays of heavy mesons, *Phys. Lett. B* **255** 583.

Gaillard M K and Lee B W, 1974, Rare decay modes of the K - mesons in gauge theories, *Phys. Rev. D* **10** 897.

Harrison P F and Quinn H R [BABAR Collaboration], 1998, *The BaBar Physics Book: Physics at an Asymmetric B Factory*, SLAC-R-0504.

Höcker A, Lacker H, Laplace S and Le Diberder F, 2001, A new approach to a global fit of the CKM matrix, *Eur. Phys. J. C* **21** 225 [*hep-ph*/0104062]; http://ckmfitter.in2p3.fr

Isgur N and Wise M B, 1989, Weak decays of heavy mesons in the static quark approximation, *Phys. Lett. B* **232** 113; 1990, Weak transition form-factors between heavy mesons, *Phys. Lett. B* **237** 527.

Ligeti Z, 2001, New results on flavor physics, *hep-ph*/0112089.

Manohar A V and Wise M B, 2000, *Heavy Quark Physics*, Cambridge Monogr. Part. Phys. Nucl. Phys. Cosmol., Vol. **10**. Cambridge University Press.

Neubert M, 1994, Heavy quark symmetry, *Phys. Rept.* **245** 259 [*hep-ph*/9306320].

Neubert M and Petrov A A, 2001, Comments on color suppressed hadronic B decays, *Phys. Lett. B* **519** 50 [*hep-ph*/0108103].

Shifman M A, 2000, Quark-hadron duality, *hep-ph*/0009131.

Sterman G and Stoler P, 1997, Hadronic form factors and perturbative QCD, *hep-ph*/9708370.

Lattice QCD

Christine Davies

University of Glasgow, UK

1   Introduction

Lattice QCD was invented, way ahead of its time, in 1974. It really became a useful technique in the 1990s when a huge amount of progress was made in the understanding and reduction of systematic errors. Now, we are poised to start a second lattice revolution with the onset of Teraflop supercomputing around the world and further improvements in methodology. This will enable calculations using lattice QCD to reach errors of a few percent, over the next five years. At this level, lattice results, where they exist, will be the theoretical calculations of choice for the experimental community.

It seems, then, a good time to review the fundamentals of lattice QCD, for an audience of experimental particle physicists. As 'consumers' of lattice calculations, it is important to be aware of how these calculations are done so that a critical assessment of different results can be made. I have tried to keep technical details to a minimum in what follows but it is necessary to understand some of them, to appreciate the significance and the limitations of the lattice results that you might want to use. For a more detailed discussion see, for example (Gupta, 1998) or (Di Pierro, 2001). This school is largely concerned with CP violation and heavy quark physics, so in Section 4 I concentrate on lattice results relevant to these areas.

2   Lattice QCD formalism and methods

2.1   The path integral approach

Lattice QCD is just QCD, no more and no less. We take the theory, express it in Feynman Path Integral language, and calculate the integral as well as we can. We would like to be able to do this in the continuous space-time of the real world, but this is not possible. Instead, we must break space-time up into a 4-d grid of points, i.e. a lattice (Figure 1), and evaluate the Feynman Path Integral by Monte Carlo methods on a computer. It turns out to be a calculation that requires a huge amount of computing power and tests the fastest supercomputers that we have.
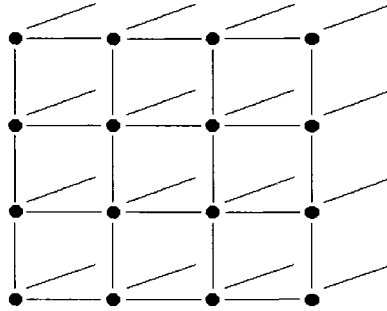
Figure 1. *A 2-dimensional rendition of a 3-dimensional cubic lattice. Lattice QCD calculations use a 4-dimensional grid.*

In the Feynman Path Integral approach, we first express the quantity that we want to calculate as the matrix element in the vacuum of an operator, $\mathcal{O}$, which will be a product of quark and gluon fields so that, for example, $\mathcal{O} = (\bar{\psi}\psi)_y(\bar{\psi}\psi)_x$. creates a hadron at a point x and destroys it at a point y. We will discuss later other forms that $\mathcal{O}$ might take to calculate useful quantities. Then:

$$\langle 0|\mathcal{O}|0\rangle = \frac{\int [d\psi]\,[d\bar{\psi}]\,[dA_\mu]\,\mathcal{O}[\psi,\bar{\psi},A]e^{-S}}{\int [d\psi]\,[d\bar{\psi}]\,[dA_\mu]e^{-S}} \tag{1}$$

where S is the action, the integral of the Lagrangian:

$$S = \int \mathcal{L}\,d^4x. \tag{2}$$

We are using Euclidean space here (imaginary time) so that the integrand doesn't contain the oscillatory e^{iS}, but the more easily integrated e^{-S}. The integral of Equation 1 can then be evaluated numerically if we can convert it to a finite-dimensional problem.

Currently the integral runs over all values of the quark and gluon fields ψ and A at every point in space-time. We need to make the number of space-time points (and therefore field variables) finite and we do this by taking a 4-d box of space-time and discretising it into a cubic grid, or lattice. It is then a relatively simple matter to transcribe the continuous theory onto the lattice, and we use the standard methods used for discretising e.g. differential equations for numerical solution. Continuous space-time (x, t) becomes a grid of labelled points, (x_i, t_i) or $(n_i a, n_t a)$ where a is the spacing between the points, called the lattice spacing. The fields are then associated only with the sites, $\psi(x, t) \to \psi(n_i, n_t)$. The action must also be discretised, but this is also straightforward. The Lagrangian typically contains fields and derivatives of fields. The fields are replaced with fields at the lattice sites and the derivatives replaced with finite differences of these fields. The integral over space-time of the Lagrangian becomes a sum over all lattice sites: ($\int d^4x \to \sum_n a^4$). There are inevitably discretisation errors associated with this procedure (just as there are for differential equations) because the lattice Lagrangian only matches the continuum Lagrangian at $a = 0$. At non-zero a there are effectively additional unwanted terms in the lattice Lagrangian that are proportional to powers of a. We will discuss this further later. Another view of the lattice is that it provides an ultra-violet cut-off on the theory

in momentum space, since no momenta larger than π/a make sense (the wavelength is then smaller than a). In this way it is an alternative regularisation of QCD.

As an illustration of the simplicity of the discretisation procedure, let us consider a scalar field theory with Lagrangian

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \phi)^2 + \frac{1}{2}m^2\phi^2 + \lambda\phi^4. \tag{3}$$

The lattice action, S is then

$$S = \sum_n a^4 \left(\frac{1}{2} \sum_{\mu=1}^{4} \left[\frac{\phi(n+1_\mu) - \phi(n-1_\mu)}{2a} \right]^2 + \frac{1}{2}m^2\phi^2(n) + \lambda\phi^4(n) \right). \tag{4}$$

The point $n+1_\mu$ is one lattice point up from the point n in the μ direction. We are always free to rescale parameters and fields and we do this on the lattice, rescaling by powers of the lattice spacing, so that the parameters and fields we work with are dimensionless. Everything is then said to be in 'lattice units'. In the scalar theory above we rescale to primed quantities where $\phi' = \phi a$, $m' = ma$, $\lambda' = \lambda$. Then

$$S = \sum_n \left(\phi'^2(n) \left[2 + \frac{1}{2}m'^2 \right] + \lambda'\phi'^4 - \frac{1}{4}\sum_\mu \phi'(n+1_\mu)\phi'(n-1_\mu) \right). \tag{5}$$

The rescaling has the effect of removing the lattice spacing explicitly from the action. A lattice calculation is done then without input of any value for the lattice spacing, or even knowing what it is. We will discuss later converting results back from lattice units to physical units, so that we can compare results to the real world. Equation 5 has in addition been rearranged to collect similar lattice terms together, using $\sum_n$ to move the space-time indices. It now looks very like a spin model, revealing a deep connection between lattice field theory and the statistical mechanics of spin systems.

2.2 Lattice gauge theories for gluons

To discretise gauge theories such as QCD onto a lattice requires a little additional thought because of the paramount importance of local gauge invariance. The rôle of the gluon (gauge) field in QCD is to transport colour from one place to another so that we can rotate our colour basis locally. It should then seem natural for the gluon fields to 'live' on the links connecting lattice points, if the quark fields 'live' on the sites.

The gluon field is also expressed somewhat differently on the lattice to the continuum. The continuum A_μ is an 8-dimensional vector, understood as a product of coefficients A_μ^b times the 8 matrices, T_b, which are generators of the SU(3) gauge group for QCD. On the lattice it is more useful to take the gluon field on each link to be a member of the gauge group itself i.e. a special (determinant = 1) unitary 3×3 matrix. The lattice gluon field is denoted $U_\mu(n_i, n_t)$, where μ denotes the direction of the link, n_i, n_t refer to the lattice point at the beginning of the link, and the color indices are suppressed. We will often just revert to continuum notation for space-time, as in $U_\mu(x)$. The lattice and continuum fields are then related exponentially,

$$U_\mu = e^{iagA_\mu} \tag{6}$$
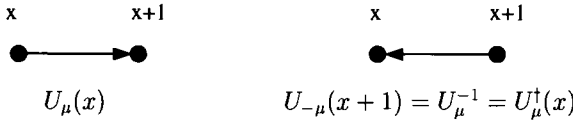
Figure 2. *The gluon field on the lattice.*

where the a in the exponent makes it dimensionless, and we include the coupling, g, for convenience. If $U_\mu(x)$ is the gluon field connecting the points x and $x + 1_\mu$ (see Figure 2), then the gluon field connecting these same points but in the downwards direction must be the inverse of this matrix, $U_\mu^{-1}(x)$. Since the U fields are unitary matrices, satisfying $U^\dagger U = 1$, this is then $U_\mu^\dagger(x)$.

This form for the gluon field makes it possible to maintain exact local gauge invariance on a lattice. To apply a gauge transformation to a set of gluon fields we must specify an SU(3) gauge transformation matrix at each point. Call this $G(x)$. Then the gluon field $U_\mu(x)$ simply gauge transforms by the (matrix) multiplication of the appropriate G at both ends of its link. The quark field (a 3-dimensional colour vector) transforms by multiplication by G at its site.

$$
\begin{aligned}
U_\mu^{(g)}(x) &= G(x)U_\mu(x)G^\dagger(x + 1_\mu) \\
\psi^{(g)}(x) &= G(x)\psi(x) \\
\overline{\psi}^{(g)}(x) &= \overline{\psi}(x)G^\dagger(x).
\end{aligned}
\tag{7}
$$

To understand how this relates to continuum gauge transformations try the exercise of setting $G(x)$ to a simple U(1) transformation, $e^{i\alpha(x)}$, and show that Equation 7 is equivalent to the QED-like gauge transformation in the continuum, $A_\mu^g = A_\mu - \partial_\mu\alpha$.
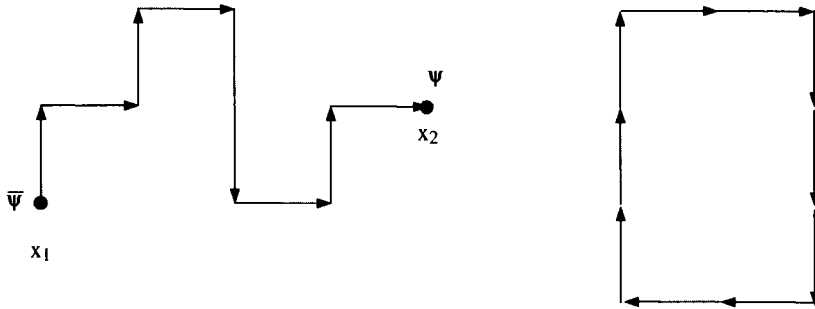
Figure 3. *A string of gluon fields connecting quark and antiquark fields (left) and a closed loop of gluon fields (right).*

Gauge-invariant objects can easily be made on the lattice out of closed loops of gluon fields or strings of gluon fields (Figure 3) with a quark field at one end and an anti-quark field at the other, e.g. $\overline{\psi}(x_1)U_\mu(x_1)U_\nu(x_1 + 1_\mu)\dots U_\epsilon(x_2 - 1_\epsilon)\psi(x_2)$. Under a gauge transformation the G matrix at the beginning of one link 'eats' the $G^\dagger$ at the end of the previous link, since $G^\dagger G = 1$. The G matrices at x_1 and x_2 are 'eaten' by those transforming the quark and anti-quark fields, if we sum over quark and antiquark colors.

The same thing happens for any closed loop of Us, provided that we take a trace over color indices. Then the G at the beginning of the loop and the $G^\dagger$ at the end of the loop, the same point for a closed loop, can 'eat' each other. (Try this as an exercise, remembering that U fields going in the downward direction are $U^\dagger$s and, from Equation 7, $U_\mu^{\dagger,(g)}(x) = G(x + 1_\mu)U_\mu^\dagger(x)G^\dagger(x)$.)

The purely gluonic piece of the continuum QCD action is

$$S_{\text{cont}} = \int d^4 x \frac{1}{4g^2} \operatorname{Tr} F_{\mu\nu} F^{\mu\nu} \tag{8}$$

and the simplest lattice discretisation of this is the so-called Wilson plaquette action:

$$S_{\text{latt}} = \beta \sum_p \left(1 - \frac{1}{3}\operatorname{Re}\left\{\operatorname{Tr} U_p\right\}\right); \quad \beta = \frac{6}{g^2}. \tag{9}$$

U_p is the closed 1×1 loop called the plaquette, an SU(3) matrix formed by multiplying
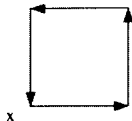
Figure 4. *A plaquette on the lattice.*

4 gluon links together in a sequence. For the plaquette with corner x in the i, j plane we have (Figure 4):

$$U_p(x) = U_i(x)U_j(x + 1_i)U_i^\dagger(x + 1_j)U_j^\dagger(x) \tag{10}$$

Tr in S_{latt} denotes taking the trace of U_p i.e. the sum of the 3 diagonal elements. S_{latt} sums over all plaquettes of all orientations on the lattice. β is a more convenient version for the lattice of the QCD bare coupling constant, g^2. This is the single input parameter for a QCD calculation (whether on the lattice or not) involving only gluon fields. Notice that the lattice spacing is not explicit anywhere, and we do not know its value until *after* the calculation. The value of the lattice spacing depends on the bare coupling constant. Typical values of β for current lattice calculations using the Wilson plaquette action are $\beta \approx 6$. This corresponds to $a \approx 0.1$fm. Smaller values of β give coarser lattices, larger ones, finer lattices. Other improved discretisations of the gluon action are also used. In these the bare coupling constant appears in a different way and so comparison of the bare coupling constant between different gluon lattice actions is meaningless. The only comparison which makes sense is that of the resulting values for the lattice spacing. That S_{latt} of Equation 9 is a discretisation of S_{cont} is not obvious, and we will not demonstrate it here. It should be clear, however, from Equations 6 and 10 that S_{latt} does contain terms of the form $\partial_\mu A_\nu$.

S_{latt} is gauge-invariant, as will be clear from our earlier discussion. Thus lattice QCD calculations do *not* require gauge fixing or any discussion of different gauges or ghost terms. We calculate the appropriate Feynman Path Integral using S_{latt}. Since we are only describing calculations for gluons at this stage, $\mathcal{O}$ will be some gauge-invariant product of U fields, for example the closed loop of Figure 3. Such a calculation is fully non-perturbative since the Feynman Path Integral includes all possible interactions in

the matrix element that we are evaluating. In contrast to the real world, however, the calculations are done with a non-zero value of the lattice spacing and a non-infinite volume. In principle we must take $a \to 0$ and $V \to \infty$ by extrapolation. In practice it suffices to demonstrate, with calculations at several values of a and V, that the a and V dependence of our results is small, and understood, and include a systematic error for this in our result.

2.3   Algorithms

The Feynman Path Integral (Equation 1) for gluons only becomes

$$\langle 0|\mathcal{O}|0\rangle = \frac{\int [dU]\, \mathcal{O} e^{-S}}{\int [dU]\, e^{-S}}. \tag{11}$$

To evaluate this integral we can generate random sets of U fields on the lattice and work out the result:

$$\langle 0|\mathcal{O}|0\rangle = \frac{\sum_\alpha \mathcal{O}_\alpha e^{-S_\alpha}}{\sum_\alpha e^{-S_\alpha}}. \tag{12}$$

$\{U\}_\alpha$ is a set of U matrices, one for each link of the lattice, and is called a *configuration*. $\mathcal{O}_\alpha$ is the value of $\mathcal{O}$ on that configuration (e.g. the trace of a closed loop of Us). A set of configurations is an *ensemble*.

This is a very inefficient way of working. If S_α is large for a particular configuration it contributes very little to the result. Instead it is better to generate the configurations with probability e^{-S}. This is called 'importance sampling' since we preferentially choose configurations with a large contribution to the integral. If we have a set of configurations so distributed then

$$\langle 0|\mathcal{O}|0\rangle = \langle \mathcal{O}\rangle = \frac{1}{N}\sum_{\alpha=1}^{N} \mathcal{O}_\alpha, \tag{13}$$

i.e. the result simply becomes the ensemble average of the value of the operator $\mathcal{O}$ evaluated on each configuration. The calculation then has a statistical uncertainty associated with it, which varies with the ensemble size, N, as $1/\sqrt{N}$.

Several algorithms exist to generate an ensemble of configurations with distribution e^{-S}. The Metropolis algorithm is the earliest and simplest, but shares several features with later more sophisticated algorithms. The first step is to generate a starting configuration, $\{U\}_1$, e.g. by setting all the U matrices to the unit 3×3 matrix or by generating random SU(3) matrices. The algorithm then sweeps round the configuration, one U matrix at a time. For each U matrix a small change is proposed, i.e. a random matrix close to the unit matrix is generated which could multiply U. The change in S is calculated if this change to U were to happen. If S is reduced, the change is accepted; if not, it is accepted with probability $e^{-\Delta S}$ (by comparing $e^{-\Delta S}$ to a random number between 0 and 1). Once this has been done for every $U \in \{U\}_1$ we have a new configuration, $\{U\}_2$. We then repeat to obtain $\{U\}_3$ etc. Once we have an ensemble we can do any number of different calculations (often called 'measurements') on it for different operators $\mathcal{O}$. Ensembles are the equivalent of experimental data sets created by collaborations of theorists. They are

often stored for years and re-used many times. Some ensembles are publicly available - see http://qcd.nersc.gov/ and http://www.ph.ed.ac.uk/ukqcd/.

An important point to note is that each member of an ensemble is generated from a previous member. The ensemble therefore has a (computer) time history. We have to worry about the 'equilibration time' and the 'decorrelation' (autocorrelation) time of the ensemble. The equilibration time is the number of sweeps required to reach a configuration typical of the distribution e^{-S} that we are trying to create, i.e a configuration which has 'forgotten' the starting configuration. The autocorrelation time is the number of sweeps it takes to generate a sufficiently different configuration that results can be considered statistically independent. The autocorrelation time can be determined from the sequence of results for $\mathcal{O}$ and will depend on $\mathcal{O}$. In general if $\mathcal{O}$ is an operator with large extent, e.g. a closed loop of U fields over many lattice sites, it will have a longer autocorrelation time than if $\mathcal{O}$ is a small loop. This is because the changes to a configuration spread out randomly from a point, one step per sweep. As we try to reach smaller values of a, closer to the continuous space-time of the real world, we expect a phenomenon called 'critical slowing-down'. This is because a given physical distance, say the size of a hadron, takes up many more lattice sites as a gets smaller. For an ensemble to decorrelate on this physical distance scale then requires more sweeps. This makes the numerical cost of reducing the lattice spacing at fixed physical volume far worse than the naïve a^4 (see Figure 5).
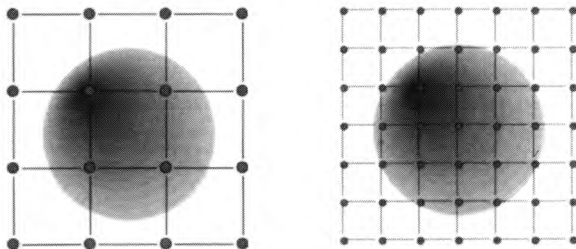
Figure 5. *A given physical distance requires more lattice points to cover it as a is reduced.*

2.4   Quarks on the lattice

2.4.1   The fermion doubling problem

The inclusion of quarks in the lattice QCD action causes several difficulties related to their fermionic nature and makes lattice QCD calculations very costly in computer time.

The so-called 'fermion doubling' problem is apparent even for free quarks, in the absence of any interaction with the gluon field. The continuum action for a single flavor of free fermions is

$$S_f = \int d^4x \, \bar{\psi}(\gamma^\mu \partial_\mu + m)\psi. \tag{14}$$

The obvious (so-called naïve) lattice discretisation gives

$$S_f^{\text{latt,naive}} = a^4 \sum_x \left[\bar{\psi}_x \sum_{\mu=1}^{4} \gamma_\mu \frac{\psi_{x+1_\mu} - \psi_{x-1_\mu}}{2a} + m\bar{\psi}_x \psi_x \right]. \tag{15}$$
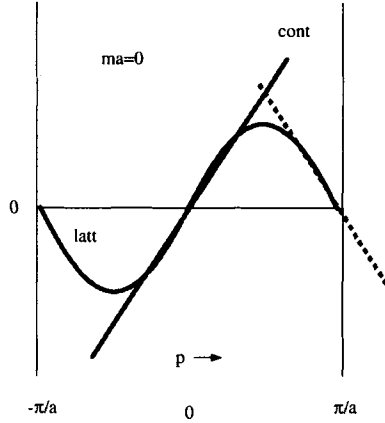
Figure 6. *The doubling problem for lattice fermions. The sine curve shows the lattice quark inverse propagator in 1-d. The straight lines through p = 0 (solid) and through p = π/a (dotted) are those for a continuum quark.*

The problems become evident when we Fourier transform this and compare the lattice inverse propagator:

$$G_{\text{latt,naive}}^{-1}(p) = i\gamma_\mu \frac{\sin p_\mu a}{a} + m \tag{16}$$

to that obtained in the continuum from Equation 14,

$$G_{\text{cont}}^{-1}(p) = i\gamma_\mu p_\mu + m. \tag{17}$$

The two are plotted for a massless quark in one-dimension in Figure 6 over one lattice Brillouin zone (momenta beyond $\pm\pi/a$ are equivalent to those in this range). The lattice result looks continuum-like around $p \approx 0$, where the inverse propagator is close to zero. The lattice inverse propagator is also close to zero around $p \approx \pi/a$, however. Since π/a and $-\pi/a$ are periodically connected on the lattice, another continuum-like line can be drawn at this point (with opposite slope to the one at the origin). Thus in one-dimension, our lattice fermion contains two continuum-like fermions rather than one! On a 4-dimensional lattice we have 2^4 fermions instead of one. The 15 excess fermions are called doublers. The doubling problem is clearly a consequence of the fact that the sine function appears in Equation 16 and this is because of the single derivatives in the Dirac action for a relativistic fermion, Equation 14. For a scalar particle (Equation 5) we would have a cosine instead, and no difficulty.

2.4.2   Wilson quarks

There are several approaches to the doubling problem. The most severe in terms of its effects, but currently the most popular for a lot of applications, is the Wilson quark action. In this the doublers are entirely removed, by adding a 'Wilson' term to the action which gives them a much larger mass than ma, so that they drop out of the physics. The term added is a double derivative so appears with an extra power of a (a^5) in order to have the

same dimensions as the other terms in S_f^{latt} (Equation 15):

$$S_f^W = S_f^{\text{naive}} - \frac{r}{2} a^5 \sum_x \bar{\psi}_x \Box \psi_x, \tag{18}$$

$$\Box \psi_x = \sum_{\mu=1}^4 \frac{\psi_{x+1_\mu} - 2\psi_x + \psi_{x-1_\mu}}{a^2}, \tag{19}$$

where r is the Wilson parameter (almost always set to 1). The extra power of a in Equation 18 means that the correspondence between the lattice and continuum actions as $a \to 0$ is not changed. However, if we look at the inverse propagator again, there is a difference.

$$G_W^{-1}(p) = G_{\text{naive}}^{-1} + \frac{2r}{a^2} \sum_{\mu=1}^4 \sin^2(p_\mu a/2). \tag{20}$$

If we substitute for G_{naive}^{-1} from Equation 16 and expand out the sin function around $p \approx 0$ we get

$$G_W^{-1}(p) = i\gamma_\mu p_\mu + m + \frac{ra}{2} \sum_{\nu=1}^4 p_\nu^2. \tag{21}$$

Comparing this to the continuum form (Equation 17) as $a \to 0$, the r term will disappear and a fermion of mass m will have the right form. If instead we look at the doublers, we must expand around $p \approx \pi/a$. If we call $\tilde{p}$ the momentum difference between p and π/a and consider the case where p has only one component close to π/a, and the others are close to zero, then

$$G^{-1}(\tilde{p}) = i\gamma_\mu \tilde{p}_\mu + m + \frac{2r}{a} + \cdots . \tag{22}$$

Now as $a \to 0$, the mass of the doubler, $m + 2r/a \to \infty$. The doublers at other corners of the Brillouin zone pick up masses of $4r/a$, $6r/a$, $8r/a$: check this as an exercise. Thus we are assured that our quark action describes only the one quark that we intended, but there is a price for this, as we shall see below.

The Wilson quark action is converted to dimensionless units by a rescaling $a^{3/2}\psi \to \psi$, leaving the quark mass parameter as a mass in lattice units, ma (previously called m').

$$S_f^W = \sum_x \left\{ \bar{\psi}_x \sum_\mu \left[(\gamma_\mu - r)\psi_{x+1_\mu} - (\gamma_\mu + r)\psi_{x-1_\mu} \right] + (ma + 4r)\bar{\psi}_x \psi_x \right\}. \tag{23}$$

It is conventional to define a 'hopping parameter' called κ which is $1/(2ma + 8r)$ and so $1/\kappa$ plays the rôle of the quark mass. ψ is conventionally rescaled by $\sqrt{2\kappa}$ so that κ moves to multiply the terms connecting the ψ field on different sites (thus allowing 'hops'). If we now couple in a gluon field, the ψ field will become a 3(color)×4(spin) dimensional vector on each site. The gluon field must be included in such a way as to keep the action gauge-invariant. From our earlier discussion it is then obvious that U matrices must be inserted as a link between the ψ and $\bar{\psi}$ fields when they are on neighbouring sites. The Wilson quark action is then conventionally written:

$$S_f^W = \sum_x \left\{ \kappa \left[\sum_\mu \bar{\psi}_x (\gamma_\mu - r) U_\mu(x)\psi_{x+1_\mu} - \bar{\psi}_{x+1_\mu}(\gamma_\mu + r)U_\mu^\dagger(x)\psi_x \right] + \bar{\psi}_x \psi_x \right\}. \tag{24}$$

The price we pay for using the Wilson quark action is that we break explicitly the chiral symmetry of continuum QCD. This is a symmetry of the derivative terms in S_f (Equation 14) which allows us to rotate separately right- and left-handed components of the quark field. The spontaneous breaking of this symmetry gives us a massless pseudoscalar meson called the pion as a Goldstone boson and has other important consequences for particle physics. Chiral symmetry is broken explicitly by a quark mass (so that the real pion is not actually massless) but also, more seriously for the lattice, by the Wilson term. As $a \to 0$, chiral symmetry will be recovered, but for real lattice calculations at non-zero a, the lack of chiral symmetry can cause difficulties for some calculations.

One surprising feature of Wilson quarks is that it is still possible to get a massless pion even at non-zero a, when chiral symmetry is broken. However, we have to search for the value of $1/\kappa$ at which it occurs—it is not simply the point $1/\kappa = 8r$, as it would be in the free theory, above. Lattice calculations of the mass of the pseudoscalar meson ($M_{\rm PS}$) must be done at various input values of κ (see Section 3) for a given ensemble. A plot of $M_{\rm PS}^2$ against $1/\kappa$ is then extrapolated to the point where $M_{\rm PS}$ is zero. The value of κ at this point is called $\kappa_{\rm critical}$ and is the point at which the bare quark mass in the interacting theory is zero (but matrix elements will not necessarily show chirally symmetric behaviour). The bare quark mass in lattice units, ma, at other values of κ can then be taken to be $(1/2\kappa - 1/2\kappa_{\rm critical})$.

Another problem for the Wilson quark action is the presence of large discretisation errors. The naïve quark action has discretisation errors proportional (at lowest power) to a^2 because (see Equations 16 and 17) $\sin(pa)/a = p(1 - p^2a^2/6 + \cdots)$. In the measurement of a hadron mass, the terms proportional to p^2a^2 in the action will induce an error proportional to Λ^2a^2 where Λ is some typical momentum scale inside the hadron in question, say 300MeV. For lattice spacing values we can reach, around 0.1fm $(= (2{\rm GeV})^{-1}$ when $\hbar = c = 1)$, this gives an expected error of order 2%. The Wilson term (Equation 18) that we added, however, is proportional to a, so that $S_f^W = S_f^{\rm cont} + O(a)$. Now hadron masses will have an error of typical size Λa, which could be 15% at $a = 0.1$fm. One can extrapolate this error away by doing calculations at several values of a but the size of the extrapolation adds uncertainty.

Instead, we can 'improve' the quark action, by adding additional terms to counteract the errors at any order in a. This is equivalent to a higher order discretisation scheme for differential equations. For the Wilson quark action we can add the so-called clover term, making the clover, or Sheikholeslami-Wohlerti, action:

$$S_f^{\rm clover} = S_f^W - \frac{iac_{\rm sw}\kappa r}{4} \sum_x \bar{\psi}_x \sigma_{\mu\nu} F_{\mu\nu} \psi_x. \tag{25}$$

The standard discretisation of $aF_{\mu\nu}$ is as a set of 4 plaquettes arranged in a clover-leaf shape. If the clover coefficient, $c_{\rm sw}$, is chosen correctly then the clover action has leading order errors proportional to a^2 again. It is in the correct choice of this coefficient that the difficulties of discretising a field theory, as opposed to a standard differential equation, appear. We are trying to match QCD with an ultraviolet momentum cut-off of π/a to QCD with an infinite momentum cut-off. Gluonic interactions with gluon momenta between π/a and ∞ in the continuum must be accounted for on the lattice by a renormalisation of coefficients in the action. Thus the naïve (tree-level) value of 1 for $c_{\rm sw}$ is renormalised by an amount which depends on the QCD coupling constant at some momentum scale around π/a. This momentum scale is typically quite large (for $a = 0.1$fm

it is 6GeV) so that a perturbative calculation of c_{sw} can work well. $c_{sw} = 1 + c_1\alpha_s(\pi/a) +$ $c_2\alpha_s^2(\pi/a) + \cdots$. In fact it has been shown that a lot of the perturbative correction can be absorbed into a renormalisation of the U field by a factor called u_0, and this is called tadpole-improvement (Lepage, 1993). Alternatively c_{sw} can be determined within the lattice calculation itself (i.e. non-perturbatively) by insisting that some continuum relationship, broken by the discretisation errors, works on the lattice (Sommer, 1998). For c_{sw} we can impose Ward identities from chiral symmetry, for example. This improvement programme for the lattice action can be carried further at the cost of introducing more coefficients that have to be determined by a match to continuum QCD. However, this must be compared to the cost of *not* improving the action, which requires calculations on very fine lattices to achieve small enough discretisation errors for the accuracy we require and is generally prohibitive.

2.4.3 Staggered quarks

Here we return to the naïve quark action and ask, what was so bad about having 16 quarks instead of 1? If we had 16 flavors of quarks of the same mass in Nature, the naïve action might be fine. In fact we only have two quarks that might be considered degenerate, u and d. They both have masses of a few MeV. Although we do not believe that their masses are the same, the difference is much smaller than any other mass, and they are treated as degenerate in most lattice calculations at present.

We can 'thin' the degrees of freedom of the naïve lattice quark action by removing the 4 spin degrees of freedom (which can be shown to be multiple copies of the same thing). The quark field, χ, then becomes a 3(colors)×1(spin) component object on a site and the staggered (Kogut-Susskind) fermion action is:

$$S_f^S = \sum_x \bar\chi_x \left\{ \frac{1}{2} \sum_\mu \eta_{x,\mu} \left(U_\mu(x)\chi_{x+1_\mu} - U_\mu^\dagger(x - 1_\mu)\chi_{x-1_\mu} \right) + ma\chi_x \right\}. \tag{26}$$

η_x is ± 1 according to the formula $\eta_{x,\mu} = (-1)^k$ where $k = \sum_{\nu<\mu} x_\nu$. This action describes $16/4 = 4$ quarks, now much closer to the real world, if we want to interpret the doublers as flavors. We might hope that if the 4 flavors do behave as 4 copies of the same thing we can reduce their effect by a factor of two or four (depending on how many degenerate flavors we want to simulate) by multiplication with the required factor at appropriate points (as we could in QCD perturbation theory). The 4 spin degrees of freedom for the 4 flavors are made from the 16 components of the χ field on a 2^4 hypercube, which is a complication if we need to separate out the flavors. The staggered action, however, has a remnant of chiral symmetry which ensures the very desirable feature that the quark mass (and the associated Goldstone boson pion mass) vanish at $ma = 0$. This behaviour gives the added benefit of making staggered quarks rather better behaved and computationally much faster to work with than Wilson-type quarks.

The down-side of staggered quarks is again the discretisation errors. These are formally $O(a^2)$, just as for naïve quarks, but some of the errors induce flavor-changing interactions and so are rather dangerous. In practice they produce a larger than expected effect for simple a^2 errors. A quark with momentum around 0 can be scattered to one with momentum around π/a i.e a doubler, and therefore a different flavor, by the interaction of Figure 7. One of the results of this is that the 16 different pions (for 4 flavors) no
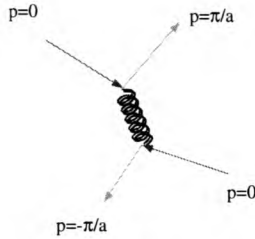
Figure 7. *A flavor-changing interaction for staggered quarks on the lattice.*

longer have the same mass and only one of them has a mass which vanishes as $ma \to 0$. Improvement terms have recently been developed which can be added to the action to reduce these interactions to a much lower level, and the masses of the different pions are then much closer together (Bernard, 2001, MILC collaboration). This makes the prospects for working with staggered quarks in lattice QCD calculations much better, and a lot more work with these quarks will certainly be done.

2.4.4   Ginsparg-Wilson quarks

A recent development has been a set of quark actions which maintain chiral symmetry of the action while still describing only one quark flavor, but at the cost of a very complicated lattice discretisation of the continuum derivative. This is then costly to implement. For example, in the domain-wall formulation an additional 5th dimension is required whose length, in principle, must go to infinity. A lot of work is being done to develop algorithms for these quark actions which may make them feasible in the long-term. In the meanwhile, they are already being used for calculations that really need chiral symmetry at finite lattice spacing, such as that of the CP-violating parameter in the K system, ϵ'.

2.5   Algorithms for quarks

Another problem with handling quarks in lattice QCD is that they are fermions, obeying the Pauli Exclusion Principle, and therefore cannot be represented by ordinary numbers in a computer. We must do the quark functional integral by hand:

$$\int [dU] \, [d\psi] \, [d\bar\psi] \, e^{-S_g + \bar\psi M \psi} = \int [dU] \det M e^{-S_g} \tag{27}$$

where the form for the matrix M depends on the quark formulation and can be derived from the forms given above for the quark action (Equations 24, 25 and 26). The QCD action then becomes

$$S = \beta \sum_p \left(1 - \frac{1}{N_c} \mathrm{Re} \, \mathrm{Tr} \, (U_p) \right) - \ln \left(\det M \right). \tag{28}$$

We now generate ensembles of gluon fields (only) with importance sampling based on this action. The standard algorithm for doing this is called Hybrid Monte Carlo . The second term is a very expensive one to include, because it requires frequent calculations of M^{-1}

(various algorithms, such as Conjugate Gradient exist to do this) and M is a large matrix (4(forWilson) $\times$ 3 $\times$ V $\approx$ 2 $\times$ 10^6 on a side). If this term is missed out for expediency (so that the action is just S_g) then we talk of using the 'quenched approximation'. Most calculations in the past have been quenched (and most of the results I discuss later will be in the quenched approximation) but recently calculations using the full QCD action ('unquenched' or 'with dynamical/sea quarks') have been attempted and in the future we hope that the quenched approximation will become redundant. We can think of the $\ln(\det M)$ term as giving rise to a sea of quark/anti-quark pairs appearing and disappearing in the vacuum. For every quark flavor for which we have a separate matrix M we should in principle include a term of the form $\ln(\det M)$ in the dynamical quark action. However, it is only the production of light (u, d, s) quark/anti-quark pairs that we envisage having a significant effect for most of the quantities that we calculate. Dynamical lattice calculations are then done with $N_f = 2$ for u, d dynamical quarks or 2+1 if s is included.

Quarks must also be integrated out of the operators, $\mathcal{O}$. For $\mathcal{O} = (\overline{\psi}\psi)_y(\overline{\psi}\psi)_x$, the form mentioned earlier, which creates a meson at the point x and destroys it at the point y, then

$$\int [dU][d\psi][d\bar{\psi}]\bar{\psi}_y^{u,a}\psi_y^{d,a}\bar{\psi}_x^{d,b}\psi_x^{u,b}e^{-S} = \int [dU](M_{x,y}^{-1,u}[U])^{ab}(M_{y,x}^{-1,d}[U])^{ba}\det M e^{-S_g}. \quad (29)$$

M^{-1} is the quark propagator from x to y on a given gluon configuration, obtained by solving $Mx = b$ where b is a vector with a 1 at x (and a certain color and spin index) and 0 everywhere else. We have been explicit here about the flavor indices, which we have taken as u and d, although lattice calculations usually then assume that u and d are degenerate and therefore the two M^{-1} factors are the same. However, if the hadron actually does contain two quarks of the same flavor then 'disconnected' pieces containing $M_{x,x}^{-1}$ will appear, as well as the 'connected' pieces above. The color indices, a and b, are also explicit (and summed over) and make $\mathcal{O}$ gauge-invariant. The sums over spin indices have not been made explicit because in this case they follow the color indices (but see Section 2.6). On an importance-sampled ensemble (either quenched or unquenched) for this example we then have to calculate $\text{Tr}_{\text{color,spin}}(M_{x,y}^{-1,u})(M_{y,x}^{-1,d})$ on every configuration and average over configurations.

Calculating M^{-1} is computationally expensive and gets harder as M develops small eigenvalues, which happens as $ma \to 0$ (for staggered quarks) or $\kappa \to \kappa_{\text{crit}}$ (for Wilson or clover quarks). Thus, even in the quenched approximation, we cannot actually calculate with quark masses close to those of real u and d quarks. Instead we work with heavier quarks and perform so-called chiral extrapolations to the chiral limit where u and d quarks would be (almost) massless.

2.6 Relating lattice results to physics

Above we have given an example for $\mathcal{O}$, which includes the creation of a valence quark and anti-quark at the point x and their destruction at the point y. This is a so-called hadron correlator or 2-point function on the lattice since it simply has a source and a sink, and is one of the simplest quantities to calculate. It is shown pictorially at the left of Figure 8, where the solid lines indicate the valence quark propagators, and the blobs
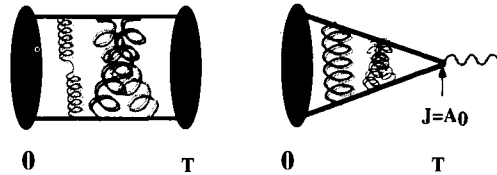
Figure 8. *A graphical representation of two types of 2-point functions calculated on the lattice. Left, for a hadron mass calculation; right, for a decay constant.*

at the two ends indicate the creation and annihilation of the meson. A baryon would of course have 3 valence quark propagator lines. Usually we project onto specific values of **p** for the hadron, so in Figure 8 we have suppressed spatial indices at the source and sink and just refer to the time index, 0 at the source and T at the sink. The figure shows how, as the valence quarks propagate, they interact any number of times by exchange of gluons. This is a pictorial representation of the fully non-perturbative nature of a lattice QCD calculation. The interactions include the production of dynamical quark/anti-quark pairs if a dynamical calculation is being done.

The calculation of this 2-point function will enable the extraction of the hadron mass (see below), for the hadron corresponding to the J^{PC} quantum numbers of $\mathcal{O}$. We make different quantum numbers by inserting γ matrices between the ψ and $\overline{\psi}$ fields in each piece of $\mathcal{O}$. For example, $(\overline{\psi}\gamma_5\psi)_x$ creates a pseudoscalar meson (such as π) and $\overline{\psi}\gamma_i\psi$ a vector (such as ρ). When the quark functional integral is done, as in Equation 29, γ matrices will appear between the two M^{-1} factors and appropriate sums over spin indices will have to be done.

The blobs in Figure 8 indicate that we can use more complicated forms for $\mathcal{O}$ for a given hadron, e.g. the ψ and $\overline{\psi}$ fields do not both need to taken at the point x. We can separate them spatially, either by inserting U fields to keep $\mathcal{O}$ gauge-invariant, or by fixing a gauge to allow spatial separation without including U fields. This enables us to feed in information, or prejudice, about the relative spatial distribution of the quarks in the hadron, i.e. its 'wavefunction'. Each piece of $\mathcal{O}$ takes the form $\overline{\psi}_{x+r}\phi(r)\psi_x$ (suppressing the U fields) where ϕ is some function of the separation between ψ and $\overline{\psi}$: it is known as the 'smearing' function and $\mathcal{O}$ is then a smeared operator. When the quark functional integral is done, factors of ϕ will appear between the M^{-1} factors. The factor of ϕ is absorbed at the source by solving $Mx = \phi$ for, say, the quark (making a 'smeared quark propagator') and $Mx = \delta$ for the anti-quark (a 'local quark propagator'). The two propagators are then put together with an explicit insertion of ϕ at the sink. Often calculations measure separately hadron correlators with several different smearing functions at both source and sink, enabling a more precise determination of the hadron mass.

Another type of 2-point function is shown on the right of Figure 8. In this case we create the hadron with a smeared operator and destroy it with a local operator. This is a 'smeared-local' or 'smeared-current' correlator, since the quantity that we can extract from this is the matrix element of the appropriate current operator, J, between the vacuum and the hadron. For example, this is used to calculate the decay constant, f_π, related to the vacuum to π matrix element of the axial vector current (denoted by its time component,

A_0, in Figure 8). This couples to the W particle and mediates the purely leptonic decay of a π meson. See the Lagrangian for the weak interactions in (Rosner, 2002), but note that the W particle is not included explicitly in lattice QCD calculations. $\mathcal{O}$ in this case then takes the form $(\overline{\psi}_{x+r}\gamma_5\phi(r)\psi_x)(\overline{\psi}\gamma_0\gamma_5\psi)_y$, where the first factor creates the pion with a smeared operator at x and the second destroys it with the time component of the local axial vector current. The quark functional integral converts this to the same type of quantity, with two factors of M^{-1}, that we discussed above.
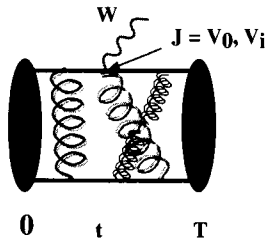
Figure 9. *A graphical representation of a 3-point function (for semileptonic decay) calculated on the lattice.*

Figure 9 shows a lattice 3-point function appropriate to the semi-leptonic decay of a hadron. One of the valence quark lines emits a W and changes to a different flavor. A new hadron is then formed with the spectator quark. The emission of the W can be represented by the insertion of a current on one of the valence quark lines. The Figure shows a vector current (with temporal component V_0 and spatial component V_i) which contributes to the decay of a pseudoscalar meson to a pseudoscalar meson (e.g. $B \to D$). We then have a (smeared) source and sink at 0 and T, and a (local) current insertion at t, i.e. 3 points. When the quark functional integral is done there will be 3 factors of M^{-1}, one for the original valence quark which decays (from 0 to t), one for the final valence quark (from t to T) and one for the spectator (from 0 to T). In fact the most efficient way to do this calculation is to solve for the final valence quark propagator from T to t, taking as a source the spectator quark propagator from 0 to T.

3 Lattice QCD calculations

3.1 The steps of a typical lattice calculation

Step 1

A volume and a rough lattice spacing are chosen. A volume of $(3\text{fm})^3$ is considered to be large enough not to 'squeeze', and therefore distort, typical hadrons placed on it. The time extent is usually taken as twice the spatial size since masses etc are extracted from the time dependence of hadron correlators (see below). The selection of the lattice spacing is a trade-off between getting close to the continuum limit (and therefore small discretisation errors) and the cost of the calculation, which grows as some large power of a^{-1}. Improvement of the action, discussed above, helps here by giving small discretisation errors on coarser lattices. Lattice spacings around 0.1fm are reasonable on both counts. From experience we know roughly what value of the bare QCD coupling constant to take

in the gluon part of the QCD action to achieve various values of a (determined *after* the calculation, see below). However, the quark contribution to the action affects this also, and we have much less experience with this. A $(3\text{fm})^3 \times 6\text{fm}$ lattice with $a \approx 0.1\text{fm}$ requires $(30)^3 \times 60$ sites.

Step 2

A quark formulation, number of quark flavors, and masses in lattice units, ma, are chosen for the quark part of the QCD action. Again we have a trade-off between trying to take realistically small masses for the u and d quarks, and the cost. Again we do not know what the quark mass actually is until after the calculation, when we have calculated the masses of hadrons containing that quark. Recent calculations have been able to take dynamical quark masses down to the s quark mass and some have gone further; future calculations need to reach much smaller masses than this. Extrapolations to u and d quark masses will continue to be necessary, however (see step 8). Some interpolation will always be necessary too since the masses chosen will inevitably not be exactly correct, e.g. for the physical strange quark mass.

Step 3

An ensemble of gluon configurations must then be generated using importance sampling with e^{-S}. As discussed above, dynamical quarks appear implicitly through the quark determinant.

Step 4

Quark propagators are calculated on each gluon configuration of the ensemble by inverting the quark matrix, M, to make the 'valence' quarks inside the hadron. Where they are supposed to have the same flavor as the dynamical quarks, they should have the same mass in lattice units, ma. However, we can also calculate valence quark propagators for quarks with different mass from the dynamical quarks, and perform separate extrapolations in valence and dynamical quark masses. This is sometimes useful and particularly so if there is a very limited set of dynamical quark masses. It is known as the partially quenched approximation (PQA).

Step 5

The quark propagators are then put together in various combinations to form hadron correlators (see the discussion of the form taken for operators, $\mathcal{O}$, above) which are then averaged over all the configurations in the ensemble. We are concentrating here on operators $\mathcal{O}$ which are related to quark-based hadrons but gluonic operators can also be measured on the ensemble and averaged in the same way.

Step 6

The hadron correlators are fitted to their expected theoretical form to extract hadron masses and matrix elements. For the 2-point function described for the spectrum, the ensemble average of the product of smeared quark propagators described above gives us the vacuum expectation value of a hadron correlator, $\langle 0|H^\dagger(T)H(0)|0\rangle$ (see Equation 1). The hadron creation(destruction) operator can create(destroy) from the vacuum all the hadron states which have the same J^{PC} quantum numbers as the operator. For example, if the operator has the quantum numbers of a pseudoscalar meson containing u and d quarks, the π and all its radial excitations can be created(destroyed). The amplitude, A, with which a particular state is created or destroyed depends on the overlap with that
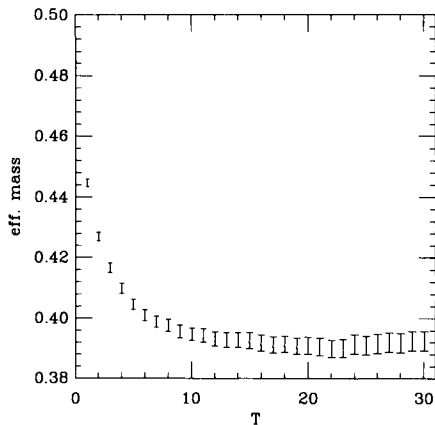
Figure 10. *The effective mass of a* 1^{--} $b\bar{b}$ *(Υ) correlator, calculated from a lattice 2-point function with local source and sink.*

state of the operator used, i.e. the smearing function, at source or sink. Thus we obtain

$$\langle 0|H^\dagger(T)H(0)|0\rangle = \sum_n \frac{A_{\mathrm{src},n}A_{\mathrm{snk},n}}{2E_n}e^{-E_nT} \tag{30}$$

where the factor e^{-E_nT} arises because the two hadron operators are offset by a time distance T in Euclidean space, and E_n is the energy of the nth state. The states which dominate the fit, especially at large values of T, are those with lowest energy; if a projection on zero momentum has been done, these will be the states with lowest mass. Often we are interested in the one state with lowest mass, the ground state (the π in the example above), and then try to design a good smearing function to have large overlap with that state, and very small overlap with its radial excitations. In that case fits can be done in which data at small values of T are thrown away and only a single exponential is used in the fit. The extent to which this works can be gauged by plotting the 'effective mass', the log of the correlator at time t divided by t. If one state completely dominates the fit, a constant result is obtained as a function of t - the effective mass is said to 'plateau'. The plateau value is the ground state mass. Figure 10 shows the result for the effective mass of the correlator for an Υ particle (see Section 4) calculated on the lattice. A clear plateau is seen but only for $t > 15$. For smaller t the correlator clearly contains excitations of higher mass, because no smearing was used in this case.

The best calculations use several different smearing functions at source and sink and perform simultaneous multi-exponential fits of the type in Equation 30. If the masses of several states can be obtained from the fit the reliability of the ground state mass is increased. It should also be pointed out that correlated fitting techniques must be used since the correlators at adjacent times are not statistically independent of each other.

For the 2-point function used to calculate decay constants, the amplitude with which the hadron is destroyed at the sink is the vacuum to hadron matrix element of the current.

$$\langle 0|J(T)H(0)|0\rangle = \sum_n \frac{A_{\mathrm{src},n}\langle 0|J|n\rangle}{2E_n}e^{-E_nT} \tag{31}$$

and $\langle 0|A_0|\pi(\mathbf{p} = 0)\rangle = f_\pi m_\pi$. To isolate the part proportional to the decay constant requires dividing the total amplitude of the ground state exponential by $A_{\mathrm{src},n=g.s.}$. This can be obtained from a fit of the type in Equation 30, if the same smearing function is used at source and sink so that $A_{\mathrm{src},n} = A_{\mathrm{snk},n}$.

For the 3-point function we have two sets of hadrons with different flavor quarks, separated by a current insertion.

$$\langle 0|H'^{\dagger}(T)J(t)H(0)|0\rangle = \sum_n \sum_m \frac{A_{\mathrm{src},n}A_{\mathrm{snk},m}\langle m|J|n\rangle}{2E_n 2E_m}e^{-E_n t}e^{-E_m(T-t)} \qquad (32)$$

where n runs over hadrons with the quantum numbers of the operator at 0 and m, those of the operator at T. Again the matrix element of interest, that of the current between two hadrons (usually for the ground states in the two cases), can be obtained by dividing out the amplitudes at source and sink from two separate 2-point fits for the two different hadrons of the kind in Equation 30.

Step 7

It is now possible to determine what the lattice spacing was in the simulation. This then sets the single dimensionful scale so that everything can be converted to physical units (GeV) from lattice units. The lattice spacing is determined by requiring one dimensionful quantity to take its real world value. Usually a hadron mass is chosen, because these are easiest to determine on the lattice, but it should not be one whose mass depends strongly on valence quark masses to be determined in the next step (see below) otherwise a complicated iterative tuning procedure will result. The most popular quantity to use at present is known as r_0, a parameter associated with the potential between two infinitely heavy quarks. It is extracted from the energy exponent of a gluonic operator (the closed loop of Figure 3), so can be precisely determined and does not contain any valence quark masses. The only problem is that it is not an experimentally accessible quantity, and we rely on potential model results to give a phenomenological value, estimated to be 0.5fm. Another quantity frequently used is the mass of the ρ meson, obtained by chiral extrapolation to the point where the π meson mass, and therefore the u,d quark mass, is (almost) zero. The chiral extrapolation, however, can produce large errors. A better quantity is the orbital excitation energy, i.e the splitting between P states and S states, in $b\bar{b}$ or $c\bar{c}$ systems, since these don't contain light quarks and this splitting is even insensitive to the heavy quark mass. (The treatment of heavy quarks on the lattice will be discussed in Section 4.)

Step 8

The step above yields all hadron masses in GeV. However, before we can compare to experiment we must tune the quark masses. This requires calculations at several different values of the bare quark masses in an appropriate region. For each quark mass we then select a hadron whose mass will be used for tuning (and is therefore not predicted). For that hadron we interpolate/extrapolate the results to find the bare quark mass at which that hadron mass is correct. The masses of other hadrons containing that quark are then predicted if we interpolate/extrapolate those masses to the same quark mass, or combination of quark masses. In the process we learn about the dependence of hadron masses on the quark mass and this can be useful theoretical information. The hadrons used for tuning should be low-lying states with accurate experimental masses which can be calculated precisely on the lattice. The π mass is usually used to fix the u, d mass

(taken to be the same), although sometimes the approximation $m_\pi = 0$ is used. The mass of the K, K^*, or ϕ can be used to fix the s quark mass. The K or K^* obviously require the u and d masses to have been fixed. The dimensionless ratio of the K^* to the K mass can also be used, and this is then less dependent on the quantity used to fix the lattice spacing. For the $c(b)$ quarks, the $D(B)$, $D_s(B_s)$ or $\psi(\Upsilon)$ systems are convenient ones.

The interpolation/extrapolation of hadron masses as a function of bare quark masses is a relatively simple procedure in the quenched approximation. Then there is no feedback from the quark sector into the gluon sector. We can create gluon field configurations at a fixed value of the lattice spacing (as determined, for example, from a purely gluonic quantity such as r_0) and measure hadron masses at many different quark masses on those configurations. The issues are then the correlations between results at different quark masses that must be taken into account and the spurious non-analytic behaviour in quark mass that can arise in the quenched approximation in extrapolations to u and d masses ('quenched chiral logarithms').

When we include dynamical quarks in the calculation, the effects of the quark determinant at a particular quark mass feed into the gluon field configurations. Results at different dynamical quark masses then represent a completely new calculation, generating a new ensemble of gluon configurations with statistically independent results. The interpolations/extrapolations in quark mass take on a new dimension and there are subtleties associated with how to do this. Some groups have chosen to generate configurations at fixed bare coupling constant and various dynamical bare quark masses. Then the lattice spacing will vary with quark mass and extrapolations in quark mass must be done in lattice units, before fixing the lattice spacing at the end. I believe a more satisfactory approach from a physical perspective is to adjust the bare coupling constant at different bare quark masses so that the lattice spacing remains approximately the same (as determined from r_0, for example). This then allows interpolations/extrapolations for physical hadron masses, and a better picture of the physical dependence of quantities on the presence of dynamical quarks. Several groups have also carried out this procedure.

In all of these approaches we must extrapolate to reach the physical u/d mass region, and so we need to know the appropriate functional form for this extrapolation. This can be derived for light enough u/d mass using an effective theory of Goldstone pions called chiral perturbation theory. This shows that logarithmic behaviour of quantities as a function of the π mass (the variable representing the u/d quark mass) should be present in general as well as simple power-law behaviour. These 'chiral logarithms' will only show up at rather small quark masses $(m_{u,d} \lesssim m_s/4)$ and so it is important for dynamical simulations to reach quark masses low enough to be able to match on to this behaviour and extrapolate down.

Step 9

The calculation needs to be repeated at several values of the lattice spacing to check that the dependence of physical results on the lattice spacing is at an acceptable level and/or to extrapolate to the continuum limit $a = 0$. Extrapolations again obviously require knowledge of an appropriate functional form.

Step 10

Compare to experiment or give a prediction for experiment!

Concluding remark

Above we have described an ideal situation. Lack of computer power has meant compromising on one or more aspects in existing calculations. A lot of calculations have used the quenched approximation. More recent dynamical calculations have used heavy dynamical masses on rather coarse and sometimes rather small lattices. These difficulties should be overcome in the next few years and this will represent a huge improvement in the reliability of lattice results.

3.2   Control of lattice systematic errors

We aim for errors of a few percent from future lattice calculations. This requires both improved statistical errors in general and good control of systematic errors. Improved statistical accuracy is obtained by generating larger ensembles of configurations with a cost proportional to the square of the improvement. Improved systematic accuracy requires theoretical understanding of the sources of error and how to remove them. It is this understanding, described below, that has been responsible for the development of good lattice techniques and the convergence of lattice results in the quenched approximation through the late 1990s. This must be carried further in the next phase of dynamical simulations to reach the goal of providing quantitative tests of QCD and input to experiment.

3.2.1   Discretisation errors

As discussed earlier, these arise from errors in the lattice form of the Lagrangian, and operators $\mathcal{O}$, compared to the continuum versions. Lattice results, even when converted to physical units, have some dependence on a. This will be as a power series in a, starting at a^n. As discussed earlier, $n = 1$ if the Wilson quark action is used, 2 for the clover quark action and 2 for the staggered quark action. n is also 2 for the Wilson plaquette gluon action of Equation 9. We expect the size of the a dependence to be controlled by a typical momentum scale relevant to the quantity being calculated. Quantities sensitive to shorter distances than others will be more susceptible to discretisation errors, even though the value of n depends only on the action used. Improved gluonic and quark actions are available in which higher order terms are added to $\mathcal{L}$ to increase n, and therefore reduce the a dependence, and these can be tested for their efficacy in the quenched approximation. The systematic improvement method is known as Symanzik improvement (Gupta, 1998).

Figure 11 shows a scaling plot of the vector meson mass (the ρ, except that the quark mass is heavier than the real u, d mass) in GeV versus the lattice spacing for various quark actions (Toussaint, 2002). Some of the calculations use an improved gluon action, with discretisation errors reduced beyond $O(a^2)$, but others use the Wilson plaquette action. There is very little difference between these (compare fancy diamonds and squares) so that most of the difference arises from the quark action used. A variant of r_0, called r_1, is used to set the lattice spacing so the vector mass and scale are given in units of r_1. The plot shows results for clover quarks (improved Wilson quarks), staggered quarks, improved staggered quarks and Ginsparg-Wilson (domain wall) quarks. The last two formulations, which are both improved to remove $O(a^2)$ errors show an impressively flat line, i.e. very little a dependence for this quantity. The clover quarks shown here have a clover improvement coefficient (see Section 2.4.1) chosen using tadpole-improvement.
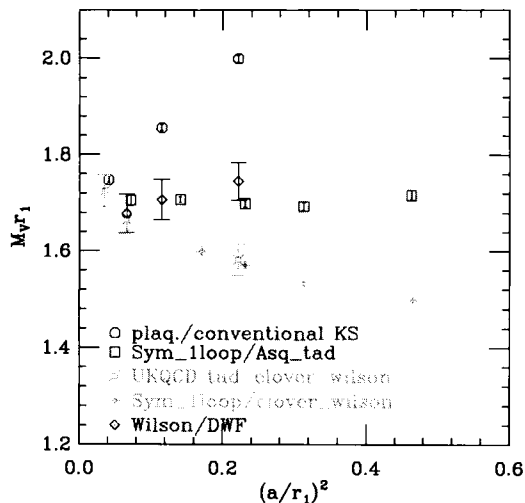
Figure 11. *A scaling plot in the quenched approximation for the vector meson at a quark mass such that the pseudoscalar meson has mass $m_\pi r_1 = 0.807$. The vector meson mass is given in units of r_1, a variant of r_0, where $r_1 \approx 0.35fm = 0.57\text{GeV}^{-1}$. It is plotted versus the square of the lattice spacing, also given in units of r_1. The squares and fancy diamonds use an improved gluon action; the others use the Wilson plaquette action. The quark actions used are: circles, staggered (Kogut-Susskind); squares, improved staggered; fancy squares and fancy diamonds, tadpole-improved Wilson (clover); diamonds, Ginsparg-Wilson (domain wall). (Toussaint, 2002)*

This reduces the a dependence of Wilson quarks to $\alpha_s a$ but it is clearly still visible. A non-perturbative determination of the clover improvement coefficient can reduce the a dependence further to $O(a^2)$, and then this formulation looks rather better. Notice the large discretisation errors visible for unimproved staggered quarks, despite the fact that the errors are $O(a^2)$ (and results therefore lie on a straight line in the Figure). Provided that all the different quark formulations have been fixed to the same physical quark mass, all the results for the vector meson mass should agree in the $a \to 0$ limit. This does seem to be true, within the statistical errors shown.

3.2.2 Finite volume

Lattice results will be distorted if the space-time box in which the calculation is done is too small to adequately represent the infinite space-time volume of the real world. For large enough volumes the error should be exponential in the lattice size, $\propto e^{-ML}$, for a lattice of size L in physical units. This means that it is possible to reduce finite volume errors rapidly to zero by taking large enough volumes. The lightest particle is the π, so this sets the volume required as we reduce the u, d quark masses to their physical values. For u, d quark masses of $m_s/4$, $m_\pi L > 5$ for $L > 3$fm, giving a finite volume error of less than 1%. Most recent lattice calculations have used volumes of this size, although there has been little systematic dependence of the volume dependence of results.

3.2.3   Matching hadronic matrix elements to the continuum

The calculation of hadronic matrix elements of various currents, J, on the lattice is discussed for 2- and 3-point functions in Section 2.6. An important point is that these depend in general on how QCD has been regularised and a finite renormalisation is then required to convert lattice results to those appropriate to a continuum scheme (such as $\overline{MS}$). Since lattice QCD and continuum QCD differ in the ultra-violet (for momenta greater than π/a), this renormalisation can be calculated in perturbation theory, by matching the matrix elements of J between quark states. We usually need several lattice currents to make up the continuum current and a mixing and matching calculation must be done.

$$
\begin{aligned}
J_{\text{cont}} &= Z_0 J_{\text{latt}}^{(0)} + a Z_1 J_{\text{latt}}^{(1)} + \cdots \\
Z_i &= 1 + c_i^{(1)} \alpha_s(2/a) + c_i^{(2)} \alpha_s^2(2/a) + \cdots
\end{aligned}
\tag{33}
$$

Lattice perturbation theory is done in the same way as continuum perturbation theory, in terms of the field A_μ and including gauge-fixing and ghost terms, if necessary. Relatively little lattice perturbation theory has been done up to now and few results exist beyond $O(\alpha_s)$. This leaves errors of $O(\alpha_s^2)$, 5–10% if we take a scale for α_s of $2/a$ at $a=0.1$fm. Higher order calculations will be required to reduce this to the required level of 2–3%, and techniques are being developed to do this. It is also sometimes possible to fix the normalisation of lattice currents non-perturbatively using symmetry arguments or to match numerically between lattice and continuum MOM-type schemes. In whatever way it is done, the matching of lattice matrix elements to the continuum is a lot of work and an area where improvements are still necessary.

3.2.4   Unquenching

The neglect of dynamical quarks in the quenched approximation is obviously wrong, but how wrong? For many years systematic errors from the quenched approximation were obscured by the size of the statistical and discretisation errors. Now improved quenched calculations are showing internal inconsistencies and disagreement with experiment which we believe will be removed once realistic dynamical calculations can be done.

One effect expected in the quenched approximation is the incorrect (too fast) running of the coupling constant from one scale to another because of the absence of $g \to \bar{q}q \to g$ pieces in the vacuum polarisation to give quark screening of the color charge. From this we might expect that the determination of the lattice spacing would depend on the quantity used to fix it, since different quantities will be sensitive to different distance/momentum scales and these will not be connected correctly by the running of α_s in the quenched approximation. (Using a quantity to fix a is equivalent to fixing the QCD coupling constant at the momentum scale relevant to that quantity). This is indeed found and illustrated by the quenched point in Figure 12. Likewise hadron masses depend on the hadron used to fix the quark mass. Then if a set of hadron masses is studied, sensitive to a range of scales and containing different combinations of quarks, errors will show up (see Figure 13 (Aoki, 2000, CP-PACS collaboration)).

The quenched approximation also does not allow the decay of particles where this requires the production of a $\bar{q}q$ pair from the vacuum, e.g $\rho \to \pi\pi$. Once dynamical quarks are light enough for this to happen, it will in fact be difficult to determine m_ρ
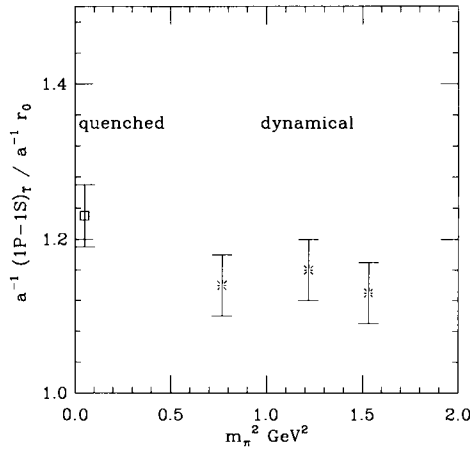
Figure 12. *The ratio of inverse lattice spacings, a^{-1}, obtained from the orbital excitation energy, the splitting between 1P and 1S states, in the Υ system and from r_0. Results are given for quenched simulations and for dynamical simulations using two flavors of dynamical quarks at three different values of the quark mass, all heavier than m_s, indicated by the square of the corresponding pion mass along the x axis. (Marcantonio, 2001, UKQCD collaboration)*
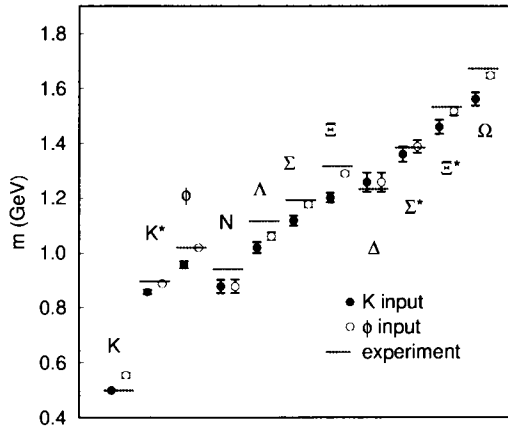
Figure 13. *The spectrum of light mesons and baryons obtained in the quenched approximation after extrapolation to u, d quark masses and to the continuum limit. The ρ and π masses are missing since they were used to fix the lattice spacing and u, d masses. Results are compared using the K or the ϕ to fix the strange quark mass and disagreement between the two is seen. The size of the discrepancy with experiment depends on this and varies between hadrons, but is at the level of 10%. (Aoki, 2000, CP-PACS collaboration)*
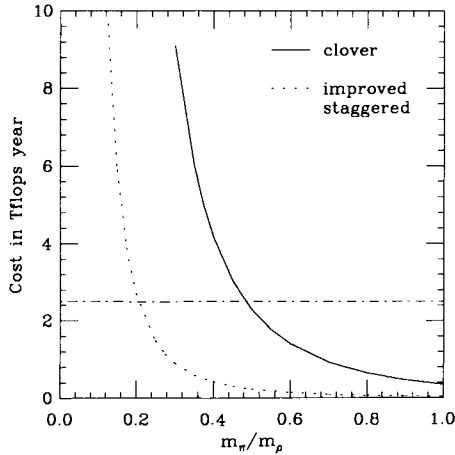
Figure 14. *The computer cost in teraflop-years of generating 500 $30^3 \times 60$ configurations with $a = 0.1fm$ and a dynamical quark mass which gives the ratio of pseudoscalar to vector meson masses along the x axis. Clover and improved staggered quarks are compared, assuming the same scaling behaviour as $(m_\pi/m_\rho)^3$. The straight line shows what is possible in 6 months on a 5Tflops computer.*

since we will obtain instead the lighter mass of the two-pion system. It is then important in dynamical simulations to use hadrons which are stable in QCD, or have very narrow widths, to fix the quark masses in the QCD action.

It has been stressed that the numerical cost of unquenched calculations is very high. It increases very rapidly as a is reduced at fixed physical volume and as $m_{u,d}$ is reduced, although the exact scaling behaviour is not completely clear. Figure 14 estimates the cost of generating an ensemble of 500 gluon configurations on an $L^3 \times T$ lattice with $L = 3fm$ and $T = 2L$ at a lattice spacing, $a = 0.1fm$, as a function of the u, d dynamical quark mass. The x axis is plotted as the ratio m_π/m_ρ where the π and ρ are the pseudoscalar and vector mesons made with valence quarks of the same mass as the dynamical quarks. The real world has $m_\pi/m_\rho = 0.2$. For $m_{u,d} = m_s$ the ratio is 0.7, for $m_{u,d} = m_s/2$, 0.55 and for $m_s/4$, 0.4. For $m_s/2$ the ratio is obtained from the K and K^* masses. For m_s and $m_s/4$ some arguments must be made about the scaling of hadron masses with quark masses because, for example, no pure $s\bar{s}$ pseudoscalar meson exists. The cost varies here as $(m_\pi/m_\rho)^3$, which is based on estimates from simulations. Figure 14 compares the cost for clover quarks and improved staggered quarks, again based on simulations at one quark mass, and using the same scaling formula. The cost advantage of improved staggered quarks is clear on this plot. One disadvantage is that the algorithm generally used for two flavors of dynamical staggered quarks is not exact, unlike that for clover. This means that there are systematic errors, rather like discretisation errors, which increase with the computer time step, ϵ, which is used to generate one gluon configuration from the previous one. Checks must to be done to make sure that this systematic error is at an acceptable level and/or an extrapolation to $\epsilon = 0$ must be done.

Recent unquenched calculations, albeit with rather heavy dynamical quark masses,
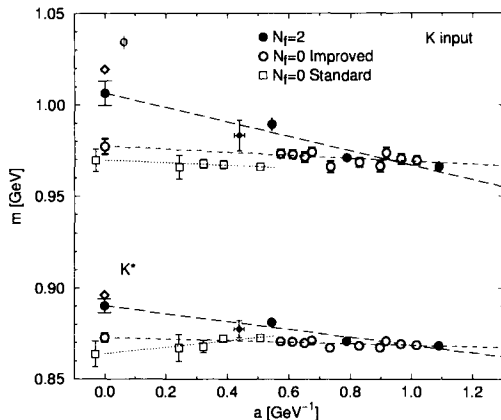
Figure 15. *The masses of the ϕ and K^* mesons as a function of lattice spacing, a, for quenched simulations and those using two flavors of dynamical quarks. The K meson is used to fix the strange quark mass. The experimental results are indicated by diamonds at $a = 0$. The Iwasaki improved gluon action was used with a clover quark action. (Ali Khan, 2002, CP-PACS collaboration)*

have shown encouraging signs that systematic errors from the quenched approximation are being overcome. Figure 12 shows that the ratio of a^{-1} values obtained from two different quantities is closer to 1 on dynamical configurations (using two flavors of dynamical quarks with a mass around m_s) than it was on quenched configurations (Marcantonio, 2001, UKQCD collaboration). From this we can hope that with 2 dynamical light quarks and a dynamical strange quark there will be only one value of the lattice spacing, corresponding to the one dimensionful scale of QCD in the continuum.

Figure 15 compares results for the masses of ϕ and K^* mesons on quenched and unquenched configurations as a function of a. The K meson is used to fix m_s and gives poor results for the K^* and the ϕ in the quenched approximation, as described earlier. For two flavors of dynamical quarks, the K^* and ϕ masses are much closer to experiment, at least after a continuum extrapolation (Ali Khan, 2002, CP-PACS collaboration). One worrying feature of this plot is the size of discretisation errors in the unquenched case, implying that the improved action used does not work very well unquenched.

Figure 16 shows another quantity from light hadron physics that gives a problem in the quenched approximation. This is the difference of the squared vector and pseudoscalar masses for given quark combinations. Experimentally the result is very flat as a function of quark mass, being $\approx 0.55 \text{GeV}^2$ from the π, ρ to the D, D^*. In the quenched approximation this quantity has a pronounced downward slope as the quark mass is increased. Recent results from the MILC collaboration with 2 $(m_s/4) + 1$ (m_s) flavors of dynamical improved staggered quarks show qualitatively different behaviour, much closer to that of experiment (Bernard, 2001, MILC collaboration). This is the strongest indication yet that calculations with dynamical quarks will overcome the disagreements between the quenched approximation and experiment.
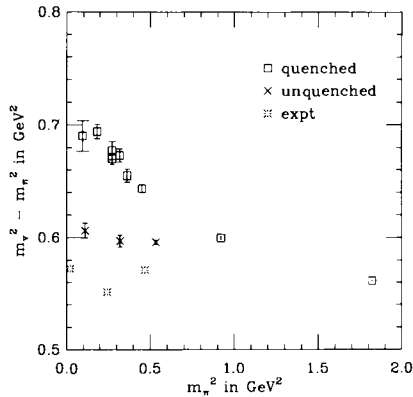
Figure 16. *The difference of the squares of the vector and pseudoscalar masses for various light hadrons, obtained with quenched and dynamical lattice QCD. The dynamical results have 2+1 flavors of dynamical quarks with masses $\approx m_s/4$ and m_s. The dynamical results are given only for valence quark masses equal to the dynamical ones. Experimental results are given by the bursts, using an estimated mass for the pseudoscalar $s\bar{s}$ meson. The lattice spacing has been obtained using $r_0 = 0.5fm$. Errors do not include errors from fixing the lattice spacing (Bernard, 2001, MILC collaboration).*

4   Lattice QCD results

The Proceedings of each year's lattice conference provide a useful summary of current results and world averages. See (LAT2000, LAT2001). Almost all lattice papers can be found on the hep-lat archive, http://arXiv.org/hep-lat/. I have deliberately chosen to refer to reviews where possible and these should be consulted for fuller access to the literature.

4.1   Methods for heavy quarks

Bottom and charm quarks are known as heavy quarks since they have masses much greater than the typical QCD scale, Λ_{QCD}, of a few hundred MeV. Top quarks are also heavy, of course, but do not have interesting bound states so are not studied by lattice QCD. b and c quarks could be treated in the same way as u, d, or s quarks on the lattice except that, with current lattice spacings of about 0.1fm, we have $m_b a$ in the interval 2–3 and $m_c a$ in 0.5–1. If ma is not small then discretisation errors of the form ma, $(ma)^2$ etc. will not be small either and such an approach will not give accurate results. Relativistic momenta, $p \approx m$, can also not be well simulated if pa is not small: pa of $O(1)$ corresponds to wavelengths which are in danger of being small enough to 'fall through' the holes in the lattices.

To reach the very fine lattices that would be required to give $m_b a \ll 1$ and accurate simulations for b quarks would require an amount of computing power way beyond our current hardware even in the quenched approximation. Luckily the physics of heavy

quark systems in the real world means that we do not have to do this; indeed, it would be largely a waste of computer power. b and c quarks are non-relativistic in their bound states, so that m and $p \approx m$ are irrelevant dynamical scales. The non-relativistic nature is evident from the hadron spectrum. There are heavy-heavy bound states in which both the valence quark and anti-quark are heavy (Υ, ψ and B_c) and heavy-light bound states in which the heavy (anti-)quark is bound to a light partner (B, B_s, D, D_s) or partners, in the case of baryons (Λ_b, Λ_c). In all cases the mass difference (splitting) between excitations of these quark systems is much less than the mass of the hadrons. For example $m(\Upsilon') - m(\Upsilon) = 560\text{MeV}$, $m(\Upsilon) = 9.46\text{GeV}$. The internal dynamics, which controls these splittings, operates with scales much smaller than the quark mass. Instead the important scales are the typical momentum carried by the quark inside the bound state, mv, and the typical kinetic energy, $\frac{1}{2}mv^2$. That these scales are small compared to m implies that $v/c \ll 1$. The use of non-relativistic techniques on the lattice is then a good match to the physics of b and c systems as well as providing an efficient way to handle them numerically on the lattice.

There are several ways to proceed, and it is important when reading the lattice literature to understand which method has been used. In the remainder of this section we consider three methods in particular: (a) static quarks, (b) NRQCD (a non-relativistic version of QCD) and (c) heavy relativistic quarks.

Static quarks

This is the $m = \infty$ limit of heavy quarks. In this limit Heavy Quark Symmetry holds and quarks become static sources of colour charge with no spin or flavor. This is evident on the lattice as the quark propagator becomes simply a string of gluon fields along the time direction (Eichten, 1990). Obviously no real quarks have infinite mass but this is a useful limit for studying heavy-light systems. Corrections away from the infinite mass limit are the subject of Heavy Quark Effective Theory (Buchalla, 2002).

NRQCD

NRQCD is a non-relativistic version of QCD (Lepage, 1992). The Lagrangian for heavy quarks is the non-relativistic expansion of the Dirac Lagrangian:

$$\mathcal{L}_Q = \overline{\psi}(D_t - \frac{\mathbf{D}^2}{2m_Q a} - c\frac{\boldsymbol{\sigma} \cdot \mathbf{B}}{2m_Q a} + \cdots)\psi \tag{34}$$

where additional terms can be added to go to higher order in v/c. ψ is now a 2-component spinor since the quark and anti-quark fields of the Dirac fields decouple from each other. D is a covariant derivative, including coupling to the gluon field. $\mathbf{B}$ is the chromomagnetic field, related to space-space components of the field strength tensor, $B_i = \epsilon_{ijk}F_{jk}$. m_Q is the quark mass; heavy quarks are frequently generically denoted Q in contrast to the q used for light quarks. Notice that the quark mass term $\overline{\psi}m_Q a\psi$ has been dropped. This simply redefines the zero of energy so that the energies of all hadrons determined on the lattice are less than 1.

The NRQCD Lagrangian can be discretised onto a lattice and leads to much simpler and faster numerical algorithms for calculating the quark propagator than for light quarks.

Instead of having to explicitly invert a matrix using an expensive iterative procedure such as Conjugate Gradient, the propagator is simply calculated by stepping through the lattice in time and calculating the propagator at time t from that at time $t - 1$. This is simply illustrated if we look at the Lagrangian in the infinite mass limit, where it becomes the Lagrangian for static quarks. Only the first term above contributes and we have:

$$S_Q = \sum_x \overline{\psi}(x) \left(U_t(x)\psi(x + 1_t) - \psi(x) \right) \qquad (35)$$

M is then an upper triangular matrix, using the notation of Equation 27, and the quark propagator is given by:

$$(M_{0,t+1}^{-1,Q}) = U_t^\dagger (M_{0,t}^{-1,Q}). \qquad (36)$$

The general start and end points, x and y, are simply denoted here by their t co-ordinates, 0 for the origin and t for the end point. To move from end point t to $t + 1$ just requires multiplication by the appropriate U field in the time direction, so M^{-1} does not change spatially and becomes a string of U fields as described for static quarks above. For NRQCD with non-infinite masses, the evolution equation in t for the propagator is not as simple and does contain spatial variations (e.g. from the spatial covariant derivatives in Equation 34) but the same principles apply. A smearing function, ϕ, is chosen at the time origin and then the propagator calculated from 0 to later times by an evolution equation from one t to the next. This makes NRQCD numerically very attractive. Heavy quark propagators, once calculated, can be combined together or with a light quark propagator to make 2- and 3-point functions for heavy hadrons as described for light hadrons earlier. As described there also, the value for the bare heavy quark mass in lattice units, $m_Q a$, is adjusted, given a value for a, until a heavy hadron mass is correct in GeV. The energies of heavy hadrons calculated on the lattice do not in fact equate directly to their masses because the mass term was removed from the Lagrangian. Instead, for one heavy hadron we have to calculate an energy-momentum dispersion relation and derive its mass from the momentum dependence ($E \propto \mathbf{p}^2/2M$).

NRQCD is an effective theory, containing the right physics for low momentum heavy quarks. Adding more relativistic corrections to the Lagrangian can make this more accurate. These higher order terms appear with coefficients (such as c in equation 34) which must be determined by matching to relativistic QCD. These coefficients represent the effect of relativistic momenta missing from NRQCD and they are governed by α_s at this high momentum scale and so are perturbative. High momenta for both quarks and gluons are missing anyway on the lattice because of the discretisation of space-time. We described earlier how a better match between lattice QCD and QCD is made by adding terms to the lattice QCD Lagrangian which are higher order in a, with a coefficient which depends on the strong coupling constant at the lattice cut-off scale. That the two procedures are very similar is not an accident; indeed, the same higher dimension operators appear in both cases. In this case NRQCD is simply making a virtue of the existence of the lattice cut-off. The difference is, however, that in the NRQCD case the operators appear with inverse powers of $m_Q a$ (in a dimensionless lattice notation) and so $m_Q a$, and therefore a, cannot be taken to zero in this approach. NRQCD has no continuum limit, but this does not prevent physical results being obtained at finite lattice spacing. It is just necessary to show that the results are sufficiently independent of a over a range of values of a.

Heavy relativistic quarks

This method looks very different from NRQCD, but has a lot of features in common. The use of a relativistic action, such as the Wilson/clover action, for heavy quarks on a lattice does not have to be incorrect if the results are interpreted carefully (El-Khadra, 1997). The main point to realise is that the existence of a large value for ma breaks the symmetry between space and time. The inverse quark propagator in momentum space has an energy at zero momentum very different from its mass (e.g. for a free Wilson quark, $E(\mathbf{p} = 0) = \ln(1 + ma)$) but its momentum dependence for small momenta is correct (i.e. as $\mathbf{p}^2/ma$). Thus, we can ignore the ma errors in the energy if we fix masses from the energy-momentum relation as for NRQCD. For more precision we must add higher order discretisation/relativistic corrections. These will appear with coefficients chosen to match continuum relativistic QCD. As we have seen the coefficients are a power series in α_s at the cut-off scale and they will depend on ma. For small ma the coefficients will be those of a discretisation correction to the action; for large ma they will go over to the NRQCD coefficients. For example, the $\sigma_{\mu\nu}F_{\mu\nu}$ clover term corrects for an $O(a)$ error in the Wilson action for light quarks; for heavy quarks, it becomes the relativistic correction which couples the quark spin and the chromomagnetic field. In this way an action can be developed that smoothly interpolates between heavy and light quark physics, at the numerical cost of having to handle heavy quarks in the same way as light ones. This method is sometimes known as the Fermilab method, since it was pioneered there.

The charm quark mass is not very heavy on the finest of current quenched lattices, and some groups have taken the standard relativistic approach in this case. To reach the b quarks then requires an extrapolation jointly in the heavy quark mass and the lattice spacing (Maynard, 2002, UKQCD collaboration) to avoid confusing discretisation and relativistic corrections. Such an extrapolation inevitably has rather large errors. A better approach is to consider a formalism which explicitly breaks space-time symmetry in order to restore the relativistic energy-momentum relation for heavy quarks. For example, you can take an anisotropic lattice which has a much finer spacing in the time direction than in the space directions. ma_t is then small and the heavy quark looks like a light one, at the cost of having many more timeslices on the lattice, and having to determine the lattice spacing in both directions (Chen, 2001).

4.2 The heavy hadron spectrum

The spectrum of heavy-heavy states has largely been the province of NRQCD (Davies, 1998). Figure 17(a) shows the radial and orbital excitations of the $b\bar{b}$ Υ system, obtained both on quenched gluon configurations and those with two flavors of dynamical quarks (Marcantonio, 2001, UKQCD collaboration). For these results the lattice spacing has been fixed by demanding that the splitting between the $\Upsilon(1S)$ and the spin-average of the P-wave (χ_b) states is correct. The b quark mass has been fixed by requiring that the $\Upsilon(1S)$ mass be correct. It is only the $2S$ (Υ'), $3S$ (Υ'') and $2P$ (χ_b') states that are predicted from this calculation, and they have rather large statistical errors at present. It is a general feature of lattice calculations that ground state masses are more precise than excited state masses. For both excited and ground states the noise is controlled by the ground state mass. For excited states the signal/noise ratio is then much worse and becomes exponentially bad at large T.
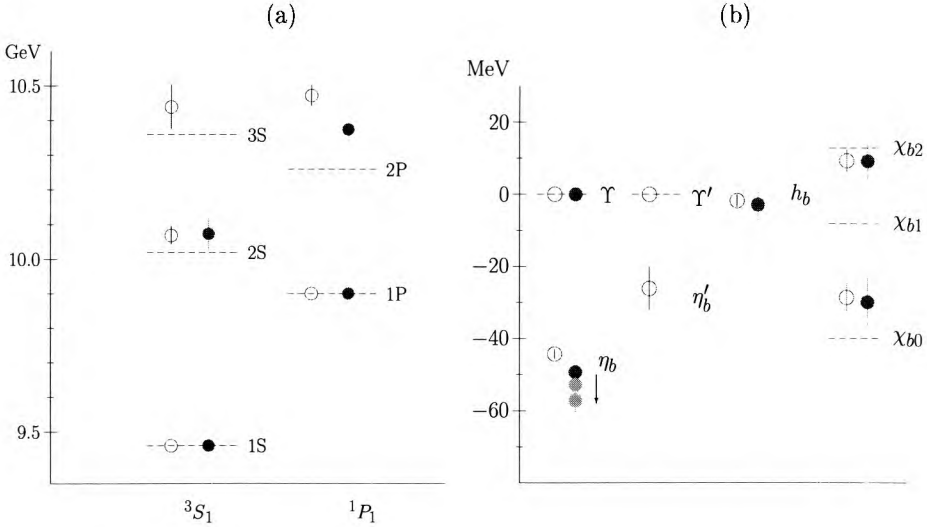
Figure 17. *(a) The radial and orbital excitations in the $b\bar{b}$ system, as calculated in lattice QCD using NRQCD for the b quarks (Marcantonio, 2001, UKQCD collaboration). (b) The fine structure of low-lying $b\bar{b}$ states. Legend: the horizontal dashed lines are experimental values; open circles show the quenched approximation; solid circles correspond to 2 flavors of clover dynamcal quarks with mass m_s. (The lowest cluster of points on the right show an extrapolation to lighter dynamical mass and to N_f=3.*

Of more immediate interest is the fine structure of the low-lying S and P states, shown in Figure 17(b). These can be determined very precisely on the lattice, particularly the 'hyperfine' splitting between the spin-parallel vector Υ state and the not-yet-seen spin-antiparallel pseudoscalar η_b. A comparison with experiment, when it exists, for this splitting will provide a very good test of lattice QCD and our b quark action, which will be important for the lattice predictions of B matrix elements described in Section 4.3.

The accuracy of the NRQCD, or other lattice action, for heavy-heavy bound states can be estimated by working out what order in an expansion in powers of v/c is represented by each term. e.g. the first two terms in the NRQCD action of Equation 34, i.e. the time derivative and the kinetic energy term, are both $O(v^2/c^2)$. This is because the 'potential energy' and kinetic energy terms are roughly equal for two heavy particles. These terms give rise to the radial and orbital splittings, and the ratio of these (≈ 500MeV) to half the Υ mass gives an estimate of $v^2/c^2 \approx 0.1$ for b quarks in an Υ. Higher relativistic corrections, such as the $\mathbf{D}^4/8m_Q^3$ term, are $O(v^4/c^4)$ and should give roughly a 10% correction to these splittings. These terms were included here, but not the v^6/c^6 corrections, so an error of roughly 1% remains. The $\boldsymbol{\sigma} \cdot \mathbf{B}$ term of Equation 34 is the first spin-dependent term and is $O(v^4/c^4)$. It gives rise to the hyperfine splitting and a similar term of the same order, proportional to $\boldsymbol{\sigma} \cdot \mathbf{D} \times \mathbf{E}$, gives rise to the P fine structure. The fine structure is indeed roughly 10% of the radial and orbital splittings. Including only these terms in the NRQCD action, as was done here, implies an error of roughly 10% in these splittings. A more precise calculation, necessary to test this action against
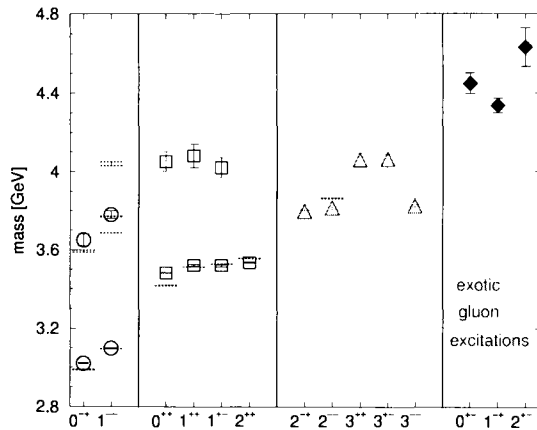
Figure 18. *The spectrum of cc̄ states, as calculated in lattice QCD using anisotropic quenched configurations (Chen, 2001).*

experiment, will require the v^6/c^6 spin-dependent terms and the $\alpha_s v^4/c^4$ terms implied by calculating the coefficient c in front of the $\sigma \cdot \mathbf{B}$ term in equation 34. This is now being done. Figure 17(b) does show, however, that the hyperfine splitting increases when two flavors of dynamical quarks are included, and continues to increase as the dynamical quark mass is reduced towards real u and d quark masses. We expect the Υ to see also s quarks in the vacuum and extrapolating the number of dynamical flavors to three increases the splitting further.

The charmonium, ψ, system is more relativistic that the Υ system and correspondingly less well-suited to NRQCD. Estimates as above give $v^2/c^2 \approx 0.3$. Figure 18 shows the charmonium spectrum obtained from anisotropic relativistic clover quarks in the quenched approximation (Chen, 2001). The lattice spacing and charm quark mass were fixed in the analogous way to that described above, except that the spin average of the vector $J\psi$ and pseudoscalar η_c masses was used to fix m_c. Since the η_c mass is known experimentally this gives improved precision since the spin-average is not sensitive to any inaccuracies in spin-dependent terms. The spectrum given in Figure 18 includes some gluonic excitations of the $c\bar{c}$ system, i.e. $c\bar{c}g$ states, called hybrids. Their existence is expected simply from the non-Abelian nature of QCD which allows gluons themselves to carry color charge. Some of these hadrons have exotic quantum numbers not available to mesons made purely of valence quarks, and the prediction of their masses will be important for their experimental discovery.

Figure 19 shows the spectrum of mesons made from one b quark and one light $(u/d$ or $s)$ anti-quark in the quenched approximation (Hein, 2000). NRQCD was used for the b quark, and the clover action for the light quark. In this case the lattice spacing was fixed using a quantity from the light hadron spectrum, m_ρ, because heavy-light systems are more similar in terms of internal momentum scales to light hadrons than heavy-heavy ones. See the comments in Section 3.2 on how the lattice spacing in the quenched approximation depends on the quantity used to fix it. The u/d and s quark masses were fixed using the π and K masses. The b quark mass was fixed from the spin-average of the B and B^* meson masses. Taking a spin-average, as above for charmonium, avoids
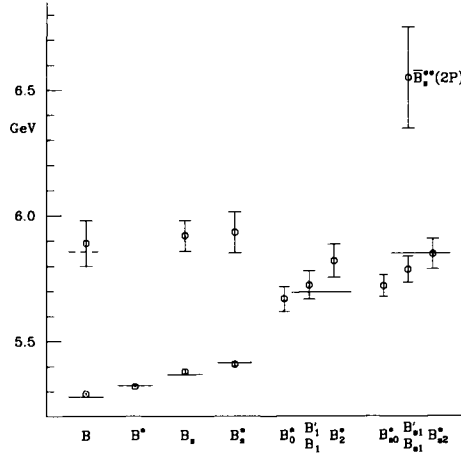
Figure 19. *The spectrum of bound states of a b quark with a light anti-quark as calculated in lattice QCD in the quenched approximation using NRQCD for the b quark (Hein, 2000).*

any errors from spin-dependent terms in the action. The b quark mass obtained this way differs from that obtained above from the Υ system, and is another feature of the quenched approximation. In the 'real world' there is only one lattice spacing and one set of quark masses and parameters fixed from the Υ system will be used to predict the entire B spectrum.

The power counting in v/c for terms in the Lagrangian works rather differently in heavy-light systems compared to heavy-heavy ones. Now there is one quark that carries almost all the mass of the heavy-light system and it sits in the centre surrounded by the swirling light quark cloud. This picture makes sense even in the limit in which the heavy quark has infinite mass when the Lagrangian would contain only the covariant temporal derivative D_t (static quarks). The higher order terms in the Lagrangian can then be ordered in terms of the inverse powers of the heavy quark mass that they contain. This is equivalent to an expansion in powers of v/c. The typical momentum of a heavy quark in a heavy-light system is $O(\Lambda_{QCD})$ (as is that of the light quark) and so $v/c \approx \Lambda_{QCD}/m_Q$. This gives $v/c \approx 10\%$ for the B and 30% for the D.

Again the power counting exercise enables us to understand the approximate relative sizes of different mass splittings in the spectrum and the accuracy of our lattice QCD calculation to a given order in v/c. The leading spin-independent term in the action is D_t giving rise to the orbital and radial excitations of ≈ 500MeV. The kinetic energy term, $D^2/2m_Q$ gives a Λ_{QCD}/m_Q correction to this, which depends on the quark mass and, therefore flavor. This explains why these excitation energies are so similar for B and D systems; the similarity between ψ and Υ is more accidental. The leading spin-dependent term is $\sigma \cdot B/2m_Q$, which gives rise to fine structure such as the splitting between the pseudoscalar B and vector B^*. This splitting should then be smaller by a factor of Λ_{QCD}/m_Q compared to the spin-independent splittings and this is indeed observed. To calculate this splitting precisely on the lattice requires the inclusion of higher order terms in the Lagrangian, as well as a better matched coefficient c for the $\sigma \cdot B$ term and this
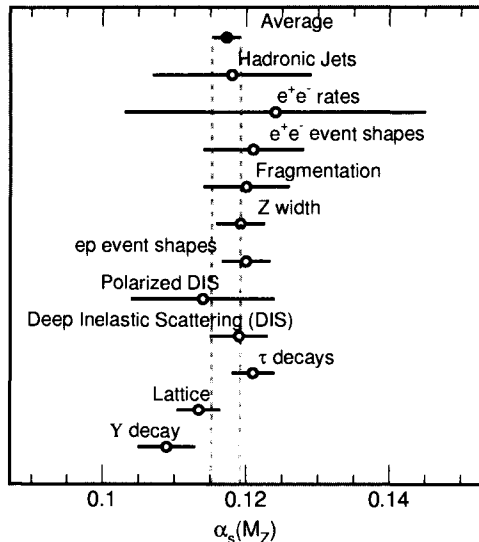
Figure 20. *A comparison of determinations of the strong coupling constant, expressed as $\alpha_s^{\overline{MS}}(M_Z)$ (PDG, 2001).*

will be done in future calculations.

We have stressed that lattice QCD is simply a way of handling QCD. It has the same a priori unknown parameters as QCD, the overall scale (equivalent to the coupling constant) and the quark masses. These parameters come from a deeper theory and must simply be fixed in the QCD Lagrangian using experiment and the results from a calculation in QCD. As described in Section 3, Lattice QCD provides the most direct way of doing this. The values for the parameters obtained are then useful input to other theoretical techniques.

Determination of the lattice spacing at a given lattice bare coupling constant, is equivalent to (and can be converted into) a determination of the renormalised coupling constant, α_s, at a physical scale in GeV. To compare to other determinations of α_s, this can be converted to the $\overline{MS}$ scheme and run to M_Z. Figure 20 shows a comparison of different determinations of α_s from the Particle Data Group (PDG, 2001). It is clear that the lattice result is one of the most precise.

All methods for determining α_s have three components:

1. Theoretical input: a perturbative expansion in α_s, for some quantity.

2. A value for that quantity.

3. An energy scale.

Most methods use an experimental result for stage 2, where the lattice uses a non-perturbative evaluation on the lattice of the vacuum expectation value of a simple short-distance gluonic operator (such as the plaquette). This avoids the problems of hadronisation etc which reduce the precision of methods based on the experimental determination

of jet shapes or cross-sections. All methods use experiment for stage 3, and here the lattice-based determination needs an experimental result to fix the lattice spacing. A good quantity to use here is the orbital excitation energy $(1P-1S)$ in, say, the Υ system since this is well-determined on the lattice and directly measured experimentally.

Quark masses are also well determined on the lattice. Since quarks are not freely available to be weighed, as an electron would be, care must be taken in defining what exactly is meant by the quark mass. The bare mass in the lattice QCD Lagrangian for a particular action, determined by the requirement to get a particular hadron mass correct and converted to physical units, is a well-defined quantity but not very convenient. We can convert it perturbatively into, say, the running quark mass in the $\overline{MS}$ scheme. The best current determination of the b quark mass is in fact from the static approximation in which b quarks have infinite mass. There is no bare b quark mass in that case; instead the binding energy $m_B - m_b$ is calculated, and from that, m_b is determined. The binding energy is small compared to m_B and has only weak dependence on the b quark mass, so for this quantity the static approximation is a good one. The b quark mass obtained in this way is 4.30(10)GeV in the quenched approximation, with some indications that it is slightly lighter when dynamical quarks are included (Lubicz, 2001).

4.3   Heavy hadron matrix elements

Precise lattice calculations of matrix elements for B decay are essential to the experimental B factory programme (Stone, 2002). This aims to test the internal consistency of the Standard Model in which CP violation occurs through the Cabibbi-Kobayashi-Maskawa matrix. The weak decays of the b quark are particularly useful in giving us access to poorly known elements of this matrix. The unitarity of the CKM matrix can be represented by a triangle; the position of the upper vertex being constrained by the determination of angles and sides, see Figure 21. The angles are determined directly by measurement of asymmetries. The determination of the sides requires both the experimental measurement
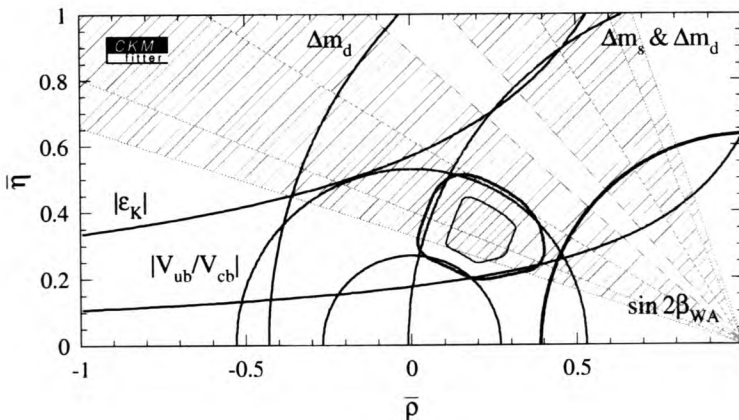
Figure 21. *The unitarity triangle with constraints on the upper vertex obtained from different quantities (Hocker, 2001). The lower vertices are at $\bar{\eta} = 0$, $\bar{\rho} = 0$ and 1.*

of a decay rate and its theoretical calculation.This allows the magnitude of one of the CKM elements to be extracted. Below we describe the lattice calculation of the matrix elements most important for this programme. The extent to which the unitarity triangle can be tested depends on both the experimental and the theoretical errors. It is critical to reduce the errors from lattice calculations to a few percent, otherwise they will dominate the uncertainties from experiment.

The simplest 2-point matrix element that can be calculated on the lattice is that for the decay constant of the charged pseudoscalar heavy-light mesons (see Figure 8). For the B this is known as f_B and it is obtained from the vacuum to B matrix element of the axial vector current which couples to the W.

$$\langle 0|A_\mu|B\rangle = p_\mu f_B. \tag{37}$$

The purely leptonic decay rate of the B meson is then proportional to f_B^2 times kinematic factors times the square of the CKM element which multiplies the appropriate axial vector current in the Lagrangian, in this case $\bar{u}\gamma_\mu\gamma_5 b$ (Rosner, 2002). In principle an experimental determination of the leptonic decay rate could be combined with the lattice calculation to yield V_{ub}, but in practice the experiment is very hard to do because the rate is so low. For other heavy-light mesons, it may be possible. f_{D_s} has been measured experimentally, but not very precisely as yet. It can be used, with lattice calculations, to give V_{cs}.

It is important to realise that, although we are discussing the weak decay of a b or c quark, the calculations are done in lattice QCD. The quark cannot decay in isolation, but must be bound into a hadron by the confinement property of QCD. The determination of the decay matrix element must then take into account all the QCD interactions inside the hadron (see Figure 8) and this requires lattice QCD. We do not put the W boson on the lattice. As far as QCD is concerned the B meson annihilates into the vacuum. The virtual W boson decay to leptons is put in by hand when we calculate the decay rate.

Lattice calculations of f_B improved markedly through the 1990s (this has been true of most lattice calculations) as we got to grips with the systematic errors. Figure 22 shows a timeline of results in the quenched approximation. It shows both that lattice calculations have markedly improved and that early calculations had very unreliable estimates of their
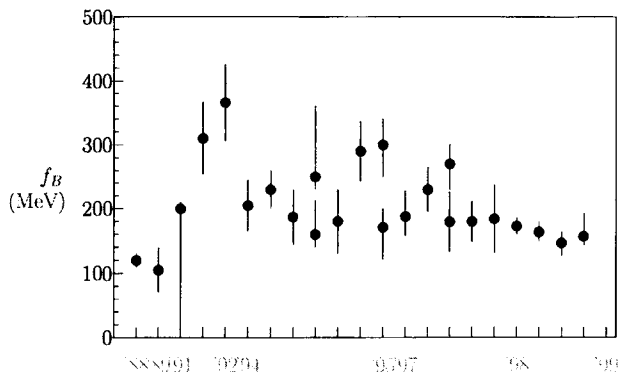
Figure 22. *A timeline of results for the B meson decay constant, f_B, calculated in lattice QCD in the quenched approximation.*

errors. The most recent and best calculations do a careful job of matching the lattice representation of the axial vector current to the continuum. For heavy-light mesons we have to be careful both about relativistic (Λ_{QCD}/m_Q) corrections and discretisation corrections to the leading order lattice current. Since $m_Q a$ is a dimensionless number, these two corrections in fact appear together and can be considered simultaneously. The matching between lattice and continuum is currently done only to $O(\alpha_s)$ and this is the major source of error in the quenched approximation. Table 1 shows a typical 'error budget' for such a calculation. We need a more precise matching, either to α_s^2 or non-perturbatively (both of which can be done with a lot of hard work), to improve the errors beyond the 10% level.

Source	percent error
statistical + fitting	3
discretisation $O((a\Lambda)^2)$	4
perturbative $O(\alpha_s^2, \alpha_s^2/(aM))$	7
NRQCD $O((\Lambda/M)^2, \alpha_s\Lambda/M)$	2
light quark mass	4
$a^{-1}(m_\rho)$	4
Total	**10**

Table 1. *Source of error in a typical lattice calculation of f_B using NRQCD for the heavy quark in the quenched approximation. $a \approx 0.1fm$, M is the b quark mass, Λ a typical QCD scale of a few hundred MeV and α_s is evaluated at $2/a$.*

Recent reviews of lattice results (Ryan, 2002), (Bernard, 2001) have given the following 'world averages' for lattice results in the quenched approximation:

- $f_B = 173 \pm 23 \text{MeV}$
- $f_{D_s} = 230 \pm 14 \text{MeV}$
- $f_{B_s}/f_B = 1.15(3)$; $f_{D_s}/f_D = 1.12(2)$.

(Note that the B_s does not decay purely leptonically but the calculation of the appropriate matrix element can still be done in lattice QCD and yields useful information on its dependence on the light quark mass.) Large-scale calculations on dynamical configurations are only just beginning, so unquenched results are still unclear. It seems likely that decay constants will be 10–20% larger unquenched.

A more important quantity from the point of view of the B factory programme is the mixing amplitude for neutral B mesons, B^0 and B_s. This mixing gives rise to a difference in mass between the CP-eigenstates, Δm, which can be measured experimentally through oscillations between particle and anti-particle (Stone, 2002). The mixing amplitude is given by the 'box diagram' (see Figure 23) in which the b quark and light anti-quark convert to a b anti-quark and light quark through the mediation of virtual Ws and (preferentially) t quarks. The mixing amplitude is then proportional to the matrix element of the box between, say, a B^0 and a $\overline{B}^0$ multiplied by the product of CKM elements $V_{tb}^* V_{td}$. The current determination of $|V_{td}|^2$ from experiment and theory gives a curve
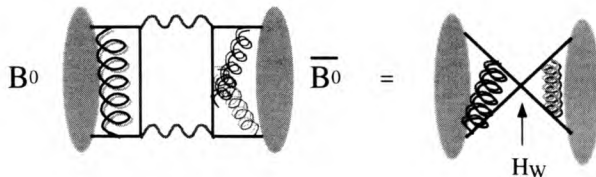
Figure 23. *The B box diagram, related to that of a 4-quark operator.*

on the unitarity triangle plot (marked Δm_d on Figure 21). Future experiments will be able to see oscillations of the B_s and then ratios of $\Delta m_{B_s}/\Delta m_B$ will allow a more precise determination of $|V_{ts}/V_{td}|^2$, since some of the systematic errors will cancel out.

As explained earlier, W bosons do not appear in lattice QCD calculations. The matrix element of the box diagram is calculated in lattice QCD by replacing it with the equivalent 4-quark operator which appears in the effective (low-energy) weak Hamiltonian (Rosner, 2002). Conventionally this matrix element for the B is set equal to $(8/3)f_B^2 M_B^2 B_B$, giving a definition of the parameter confusingly called B_B. B_B is the amount by which the matrix element differs from the result that would be obtained by saturating the H_W vertex of Figure 23 with the vacuum (comparing this to the right hand picture of Figure 8 we can see that this would be f_B^2). B_B is generally expected to be roughly 1, and this explains why lattice calculations originally concentrated on calculating f_B. To calculate B_B is harder, but is now being done. It requires, as for f_B, a careful matching between the lattice and the continuum, and this has again been done to $O(\alpha_s)$ so far.

Recent world averages for the renormalisation-group-invariant definition of B_B in the quenched approximation have been given as (Ryan, 2002), (Bernard, 2001):

- $\hat{B}_{B_d} = 1.30(12)(13)$
- $f_{B_d}\sqrt{\hat{B}_{B_d}} = 230(40)\text{MeV}$
- $\hat{B}_{B_s}/\hat{B}_{B_d} = 1.00(4)$
- $f_{B_s}\sqrt{\hat{B}_{B_s}}/f_{B_d}\sqrt{\hat{B}_{B_d}} = 1.15(6)$

A lot of the matching errors cancel out in the ratios between B_s and B_d so that the errors in these ratios are less than 10%. The ratio may also not be significantly affected by unquenching.

Heavy-light mesons decay semi-leptonically through a diagram in which the heavy quark changes flavor, emitting a virtual W, and the other (spectator) quark in the meson combines with the new quark flavor to make a new meson. In this way B mesons can decay to D or D^* mesons if $b \to c$ and to π or ρ mesons if $b \to u$. In each case the appropriate CKM element appears at the current vertex in the three-point diagram (see Figure 9) and can therefore be determined by a comparison of the experimental exclusive rate to the theoretical one. The ratio V_{ub}/V_{cb} gives an important circular constraint in the unitarity triangle (see Figure 21).

The calculation of the matrix element for B semi-leptonic decay on the lattice requires the calculation and simultaneous fitting of the 3-point function of Figure 9 and the appropriate 2-point functions necessary to isolate the matrix element. It is therefore

significantly harder than a simple 2-point calculation. In addition the matrix element depends on q^2, the squared difference of 4-momenta between the initial and final meson. This can take a range of values, because the decay is a three-body one. The matrix element can then be written as a combination of form-factors which are q^2 dependent, in contrast to the two-body leptonic decay which was parameterised by a single number, f_B. For example the pseudoscalar to pseudoscalar transition (e.g. B to D or π) proceeds only through the vector current and has two form factors, f_+ and f_0:

$$\langle P'(p')|V_\mu|P(p)\rangle = f_+(q^2)\left[(p+p')_\mu - \frac{M_P^2 - M_{P'}^2}{q^2}q_\mu\right] + f_0(q^2)\frac{M_P^2 - M_{P'}^2}{q^2}q_\mu \qquad (38)$$

The differential decay rate is proportional to the square of f_+ because the leptonic current L_μ coupling to the W has $q^\mu L_\mu = 0$ for massless leptons. The pseudoscalar to vector transition proceeds through both the vector and axial currents and has 5 form factors, 3 of which appear in the decay rate.

To explore different values of q^2 for semi-leptonic decay on the lattice it is easiest to insert different 3-momenta at the final meson and at the current, and then work out the resulting 3-momentum of the initial meson. We are restricted to values of 3-momentum allowed on the lattice, i.e. the components of **p** have the form $p_j a = n_j 2\pi/L$, where $n_j = 0, 1, 2 \ldots$ and L is the number of lattice sites in the j direction. The smallest non-zero value of p_j is then $2\pi/(La)$ where La is the physical size of the lattice in a spatial direction. A big physical volume is then required to achieve a fine discretisation of momentum space and avoid a large jump from one momentum to the next. In general results at higher momenta are much noisier than those at small momenta (this is for the same reason that excited states are noisier than ground states, discussed above) and calculations tend to be restricted to a few of the smallest possible momenta. Discretisation errors will also be larger at larger values of pa, so systematic errors will be higher.

For the matrix element for B to $D^{(*)}$ semi-leptonic decay it is useful to consider both the b quark and the c quark in the heavy quark limit. In that limit, as discussed above, the Lagrangian for heavy quarks becomes insensitive to the heavy quark spin or flavour (Buchalla, 2002). The light quark cloud in the meson cannot tell whether it is surrounding a b or a c quark or one whose spin is pointing parallel or anti-parallel to its spin. Thus the form factors for $B \to D$ and $B \to D^*$ will become identical (or vanish) and the same as the $B \to B$ elastic form factor, provided they are viewed as a function of the right variable. This is not q^2 but $v \cdot v'$ where v is the 4-velocity (p_μ/m) of the initial meson and v' is the 4-velocity of the final meson. $v \cdot v'$ is often given the symbol w. In the notation of Equation 38 $w = (M_P^2 + M_{P'}^2 - q^2)/(2M_P M_{P'})$. The limit $w = 1$ is known as the 'zero-recoil' limit because this corresponds to the kinematic point where the B meson at rest decays to, say, a D meson at rest and the decay products of the W come out back-to-back. This point has maximum $q^2 = (M_P - M_{P'})^2$.

The $B \to B$ elastic form factor takes the form

$$\langle B(v')|V_\mu|B(v)\rangle = M_B \xi(w)(v + v') \qquad (39)$$

in the limit of infinite b quark mass, where $\xi(w) = f_+(q^2)$, $f_- = 0$. $\xi(w)$ is known as the Isgur-Wise function. $\xi(1) = 1$ is an absolute normalisation in the continuum because $\bar{b}\gamma_\mu b$ is a conserved current. The lattice current is not a conserved one (except for the NRQCD/static actions) but if we are interested only in the shape of $\xi(w)$ we can
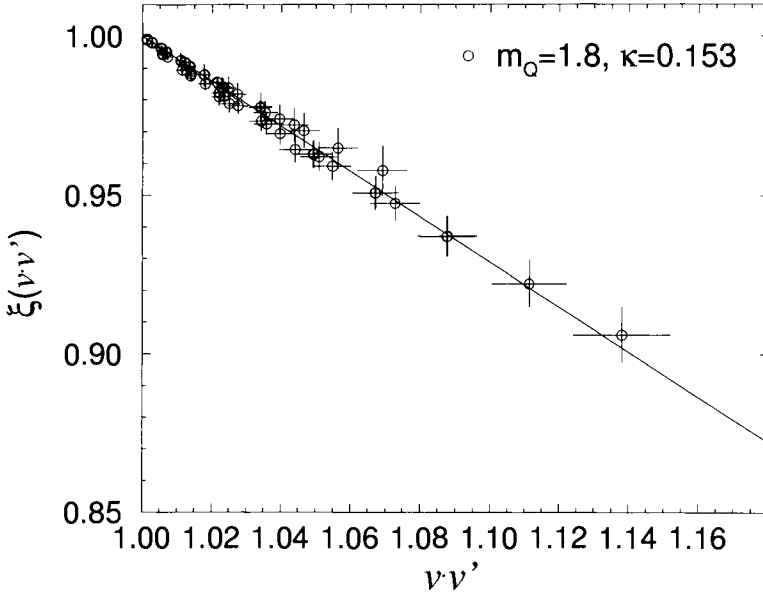
Figure 24. *The $B \to B$ elastic form factor calculated in lattice QCD using NRQCD for the b quark and plotted versus w (Hashimoto, 1996).*

renormalise to match 1 at $w = 1$. There have been several calculations of the $B \to B$ form factor on the lattice, for various heavy quark masses. Figure 24 shows such a calculation using NRQCD with a mass close to that for the b quark (Hashimoto, 1996).

The interest in calculating the Isgur-Wise function is that, in the Heavy Quark Symmetry picture described above, it is also applicable to $B \to D$ and $B \to D^*$ decays. In these cases, however, there is an additional overall perturbative renormalisation because $\bar{b}\gamma_\mu c$ is not a conserved current, and there are corrections which appear as differences of inverse powers of the b and c quark masses. For kinematic reasons, $B \to D^*$ is experimentally easier to measure in the $w \to 1$ region. The differential decay rate is

$$\frac{d\Gamma}{dw} = |V_{cb}|^2 K(w) \mathcal{F}^2(w) \tag{40}$$

where V_{cb} is the CKM element that we want to determine, $K(w)$ is a kinematic factor and $\mathcal{F}$ is the form factor for the decay. Figure 25 shows results from the CLEO collaboration (CLEO, 2000) for $\mathcal{F}(w)|V_{cb}|$. The lighter hashed curve is the result from the lattice shown in Figure 24 rescaled by a constant to match at $w = 1$. Given lattice results for $B \to D^*$ rather than $B \to B$, the constant required for rescaling would be $|V_{cb}|$ which would then be determined.

In fact, a number of simplifications can be made to the lattice calculation at the $w = 1$ point and so it is currently better to perform a phenomenological extrapolation of the experimental data to $w = 1$ and divide the extrapolated result by the lattice result for $\mathcal{F}(1)$. The Fermilab group, using heavy relativistic (Fermilab) quarks and $O(\alpha_s)$ matching
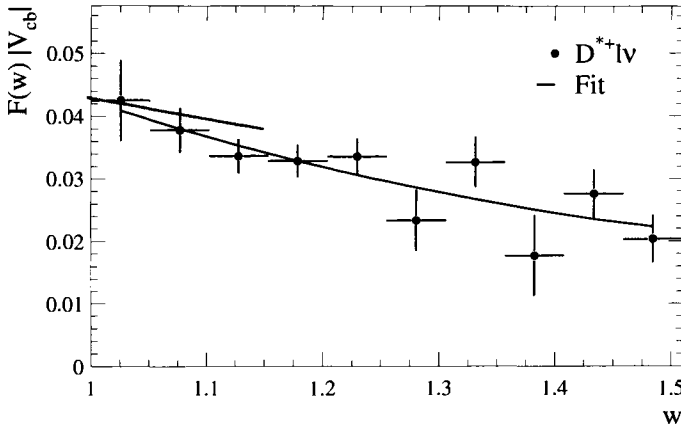
Figure 25. $|V_{cb}|F(w)$ *extracted from the experimental* $B \to D^*$ *decay rate plotted as a function of* w *(CLEO, 2000). The shorter curve on* $1<w<1.15$ *is a rescaled version of the curve in Figure 24.*

to the continuum, give the most precise result so far:

$$\mathcal{F}_{B \to D^*}(1) = 0.913^{+0.024+0.017}_{-0.017-0.030} \tag{41}$$

in which the first error comes from statistics and fitting and the second from systematic errors, including the effect of using the quenched approximation (Hashimoto, 2001). The resulting value of $|V_{cb}|$ extracted depends on which experiment's value for $|V_{cb}|\mathcal{F}(1)$ is taken. Using an average result (Stone, 2002) of $37.8 \pm 1.4 \times 10^{-3}$ gives a value for V_{cb} of $41.4 \pm 1.5 \pm 1.7 \times 10^{-3}$ where the final theoretical error comes from adding the lattice errors in quadrature. The lattice and experimental errors are currently of about the same size. The lattice error can be improved further in an unquenched calculation with a higher order matching of the lattice current to the continuum.

$B \to$ light (π, ρ) semi-leptonic decay is rather harder to calculate on the lattice. In many ways it is more important to do, however, because continuum techniques, such as HQET, can give very little useful input. One difficulty is that lattice systematic errors are smallest where the B and, say, the π lattice momenta are smallest, around the zero-recoil point discussed above, but there is very little experimental data there. Most experimental data occurs at relatively low q^2 values $(q^2 < 16\text{GeV}^2)$ when the zero recoil point has $q^2 = q^2_{\text{max}} = (m_B - m_\pi)^2 = 26\text{GeV}^2$. A comparison of lattice results for the form factors for $B \to \pi$ decay is shown in Figure 26 (Bernard, 2001). Different lattice results are shown covering a range of q^2. The reason that some results are at smaller q^2 than others is because some use relativistic quarks (marked NPclover) at a mass around the c quark mass rather than the b. For reasons discussed earlier, none of the lattice calculations can be done at the physical u, d quark masses and so must be chirally extrapolated to that point. This is done in a different way by different groups and has led to very different final results, even though the intermediate data does not show very different behaviour (see Figure 26). A better understanding of how the chiral extrapolation should be done will be required before precise lattice results will be available. Good experimental results in the q^2 region that the lattice can reach will then allow a determination of V_{ub}.
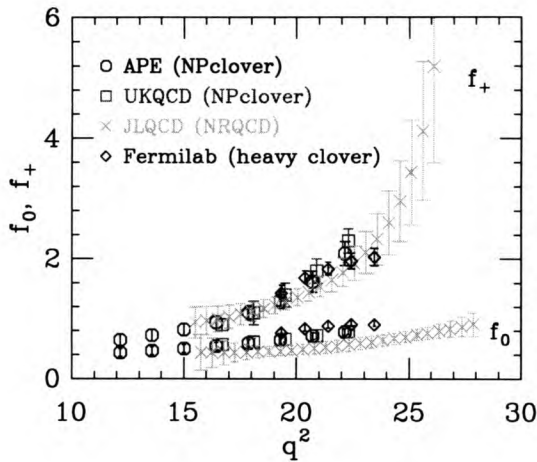
Figure 26. *Lattice results for the form factors for B→π decay (Bernard, 2001).*

5   Conclusions

Lattice QCD has come a long way from the original calculations of the 1970s. The original idea that we could solve a simple discretisation of QCD numerically by 'brute force' has been replaced by a more sophisticated approach. Unfortunately, to the uninitiated, this can look like cookery. I have tried to describe some of the calculational and technical details so that non-practitioners feel able to make an informed judgement about lattice calculations, and see where progress will be made in the future. There is no doubt, for example, that precise lattice calculations are needed to obtain maximum benefit from the huge experimental investment in *B* physics. In the next few years such calculations will become possible, at least for some quantities, and this will mark the 'coming of age' of the lattice QCD approach at last.

Acknowledgments

References

Ali Khan A et al, 2002, CP-PACS collaboration, *Light hadron spectroscopy with two flavors of dynamical quarks on the lattice*, Phys. Rev. **D65** 054505, hep-lat/0105015.
Aoki S et al, 2000, CP-PACS collaboration, *Quenched light hadron spectrum*, Phys. Rev. Lett. **84**, 238; hep-lat/9904012.

Bernard C, 2001, *Heavy quark physics on the lattice*, Nucl. Phys. B (Proc. Suppl. **94**) 159; hep-lat/0011064

Bernard C et al, 2001, MILC collaboration, *The QCD spectrum with three quark flavors*, Phys. Rev. D**64** 054506; hep-lat/0104002.

Buchalla G, 2002, *Heavy quark theory*, this volume; hep-ph/0202106.

Chen P, Liao X and Manke T, 2001, *Relativistic quarkonia from anisotropic lattices*, Nucl. Phys. B (Proc. Suppl. **94**) 342; hep-lat/0010069.

CLEO collaboration, 2000, *Determination of the $B \rightarrow D^*l\nu$ decay width and $|V_{cb}|$*, hep-ex/0007052.

Davies C, 1998, *The heavy hadron spectrum* in *Computing Particle Properties*, Springer LectureNotes in Physics, Eds Gausterer, Lang; hep-ph/9710394.

Di Pierro M, 2001, *From Monte Carlo Integration to Lattice Quantum Chromodynamics*, hep-lat/0009001

Eichten E and Hill B, *An effective field theory for the calculation of matrix elements involving heavy quarks*, Phys. Lett. B**243**, 511.

El-Khadra A, Kronfeld A and Mackenzie P, 1997, *Massive fermions in lattice gauge theory*, Phys. Rev. D**55**, 3933; hep-lat/9604004.

Gupta R, 1998, *Introduction to Lattice QCD*, hep-lat/9807028

Hashimoto S and Matsufuru H, 1996, *Lattice heavy quark effective theory and the Isgur-Wise function*, Phys. Rev D**54** 4578; hep-lat/9511027.

Hashimoto S et al, 2001, *Lattice calculation of the zero-recoil form factor of $B \rightarrow D^*l\nu$: towards a model independent determination of $|V_{cb}|$*, hep-lat/0110253

Hein J et al, 2000, *Scaling of the B and D meson spectrum in lattice QCD*, Phys. Rev. D**62** 074503; hep-ph/0003130.

Hocker A, 2001, *A new approach to gloabl fit of the CKM matrix*, Eur. J. Phys. C**21** 25; hep-ph/0104062; http://www.slac.stanford.edu/~laplace/ckmfitter.html.

LAT2000, Nucl. Phys. B (Proc. Suppl. **94**) 2001

LAT2001, Nucl. Phys. B (Proc. Suppl. **106**) 2002

Lepage P and Mackenzie P, 1993, *On the viability of lattice perturbation theory*, Phys. Rev. D**48**, 2250; hep-lat/9209022.

Lepage P et al, 1992, *Improved nonrelativistic QCD for heavy quark physics*, Phys. Rev. D**46**, 4052; hep-lat/9205007.

Maynard C et al, 2002, UKQCD collaboration, *Heavy-light decay constants on the lattice*, Nucl. Phys. B (Proc. Suppl. **106**), 388; hep-lat/0109026.

Lubicz V, 2001, *Quark masses on the lattice, light and heavy*, Nucl. Phys. B (Proc. Suppl. **94**) 116; hep-lat/0012003.

Marcantonio L et al, 2001, UKQCD collaboration, *The unquenched Υ spectrum*, Nucl. Phys. B (Proc. Suppl. **94**), 363; hep-lat/0011053.

PDG, 2001, http://pdg.lbl.gov/.

Rosner J, 2002, *The Standard Model in 2001*, this volume; hep-ph/0108195.

Ryan S, 2002, *Heavy quark physics from lattice QCD*, Nucl. Phys. B (Proc. Suppl. **106**) 86; hep-lat/0111010.

Sommer R, 1998, *Non-perturbative renormalisation of QCD* in *Computing Particle Properties*, Springer Lecture Notes in Physics, Eds Gausterer, Lang; hep-ph/9711243.

Stone S, 2002, *B Phenomenology*, this volume; hep-ph/0112008.

Toussaint D, 2002, *Spectrum results with Kogut-Susskind quarks*, Nucl. Phys. B (Proc. Suppl. **106**) 111; hep-lat/0110010.

CP violation — a new era

Yosef Nir

Weizmann Institute of Science, Israel

1 Introduction

The Standard Model predicts that the only way that CP is violated is through the Kobayashi-Maskawa mechanism [1]. Specifically, the source of CP violation is a *single* phase in the mixing matrix that describes the charged current weak interactions of quarks. In the introductory chapter, we briefly review the present evidence that supports the Kobayashi-Maskawa picture of CP violation, as well as the various arguments against this picture.

1.1 Why believe the Kobayashi-Maskawa mechanism?

Experiments have measured to date three independent CP violating observables:

- Indirect CP violation in $K \to \pi\pi$ decays [2] and in $K \to \pi\ell\nu$ decays is given by:

$$\varepsilon_K = (2.28 \pm 0.02) \times 10^{-3} \; e^{i\pi/4}. \tag{1}$$

- Direct CP violation in $K \to \pi\pi$ decays [3, 4, 5, 6, 7] is given by

$$\frac{\varepsilon'}{\varepsilon} = (1.72 \pm 0.18) \times 10^{-3}. \tag{2}$$

- CP violation in $B \to \psi K_S$ decay and other, related modes has been measured [8, 9, 10, 11, 12]:

$$a_{\psi K_S} = 0.79 \pm 0.10. \tag{3}$$

All three measurements are consistent with the Kobayashi-Maskawa picture of CP violation. In particular, the two recent measurements of CP violation in B decays [11, 12] have provided the first precision test of CP violation in the Standard Model. Since the model has passed this test successfully, we are able, for the first time, to make the following statement: *The Kobayashi-Maskawa phase is, very likely, the dominant source of CP violation in low-energy flavor-changing processes.*

In contrast, various alternative scenarios of CP violation that have been phenomeno-
logically viable for many years are now unambiguously excluded. Two important examples
are the following:

- The superweak framework [13], that is, the idea that CP violation is purely indirect,
  is excluded by the evidence that $\varepsilon'/\varepsilon \neq 0$.

- Approximate CP, that is, the idea that all CP violating phases are small, is excluded
  by the evidence that $a_{\psi K_S} = \mathcal{O}(1)$.

The experimental result (3) and its implications for theory signify a new era in the study
of CP violation. In this series of lectures we will explain these recent developments and
their significance.

1.2   Why doubt the Kobayashi-Maskawa mechanism?

1.2.1   The baryon asymmetry of the Universe

Cosmology shows that the Kobayashi-Maskawa phase cannot be the only source of CP
violation: baryogenesis, that is, the history of matter and antimatter in the Universe,
cannot be accounted for by the Kobayashi-Maskawa mechanism.

To understand this statement, let us provisionally switch off all CP violation. Then,
for every process that occurs in Nature, the corresponding CP conjugate process proceeds
with a precisely equal rate. Let us further assume that the initial conditions are such
that the number density of quarks and the number density of the matching antiquarks
are equal. Then, CP invariance guarantees that the number densities remain equal to
each other along the history of the Universe. In other words, the baryon asymmetry,
$\eta \equiv (n_B - n_{\bar{B}})/n_\gamma$, is guaranteed to remain zero. Two particularly significant processes
are proton-antiproton annihilation and production. While the first would happen at any
temperature, the latter is allowed only if the energy of the photons is large enough to
produce a proton-antiproton pair. At high enough temperatures, $T \gtrsim 2m_p$, annihilation
and production will keep the protons and antiprotons in equilibrium and their number
densities would be (precisely equal to each other and) similar to the photon number den-
sity, $n_B = n_{\bar{B}} \approx n_\gamma$. But at temperatures well below GeV, proton-antiproton production
slows down until it practically stops. Since annihilation continues to take place, the num-
ber densities of protons and antiprotons (remain equal to each other but) decrease, and
at present there would be practically neither matter nor antimatter. This is, of course,
inconsistent with observations.

Now let us switch on CP violation. That allows a different rate for a process and its
CP conjugate. Such a situation would have relevant consequences if two more conditions
are met [14]: there is a departure from thermal equilibrium and baryon number can be vi-
olated. When all three conditions are satisfied, a difference between the number densities
of quarks and of antiquarks can be induced. We assume that the number of quarks be-
comes slightly larger than the number of antiquarks. This scenario is called *baryogenesis*.
At the electroweak phase transition (temperatures of order a few hundred GeV, $t \sim 10^{-11}$
seconds) baryon number violating processes become highly suppressed, and the baryon
number cannot change any longer. The history of matter and antimatter in the Universe

proceeds along the same lines as described in the previous paragraph. In particular, at temperatures well below GeV the number densities of protons and antiprotons decrease. There is however an important difference: at some time, practically all antiprotons would disappear. But the small surplus of protons have no matching antiprotons to annihilate with. It remains there forever. The resulting picture of the present Universe is then as follows: there is no antimatter. There is a small amount of matter, with the present ratio $(n_B/n_\gamma)_0$ reflecting the baryon asymmetry, $[(n_q - n_{\bar{q}})/n_\gamma]_{\mathrm{BG}}$, induced by baryogenesis. This picture is qualitatively consistent with observations. Thus we have good reasons to think that we understand the general mechanism of baryogenesis.

The important point for our purposes is that baryogenesis is a consequence of CP violating processes. Therefore the present baryon number, which is accurately deduced from nucleosynthesis constraints (for a recent analysis, see [15]),

$$\frac{n_B}{n_\gamma} = (5.5 \pm 0.5) \times 10^{-10}, \tag{4}$$

is essentially a CP violating observable! It can be added to the list of known CP violating observables, Equations (1), (2) and (3). Within a given model of CP violation, one can check for consistency between the data from cosmology, Equation (4), and those from laboratory experiments.

The surprising point is that the Kobayashi-Maskawa mechanism for CP violation fails to account for (4). It predicts present baryon number density that is many orders of magnitude below the observed value [16, 17, 18]. This failure is independent of other aspects of the Standard Model: the suppression of n_B/n_γ from CP violation is much too strong, even if the departure from thermal equilibrium is induced by mechanisms beyond the Standard Model. This situation allows us to make the following statement: *There must exist sources of CP violation beyond the Kobayashi-Maskawa phase.*

Three important examples of viable models of baryogenesis are the following:

1. **GUT baryogenesis** (for a recent review see [19]): the source of the baryon asymmetry is in CP violating decays of heavy bosons related to grand unified theories. In general, baryon number is not a conserved quantity in GUTs. Departure from thermal equilibrium is provided if the lifetime of the heavy boson is long enough that it decays when the temperature is well below its mass. The relevant CP violating parameters are not expected to affect low energy observables.

2. **Leptogenesis** (for a recent review see [20]): lepton asymmetry is induced by CP violating decays of heavy fermions that are singlets of the Standard Model gauge group (sterile neutrinos). Departure from thermal equilibrium is provided if the lifetime of the heavy neutrino is long enough that it decays when the temperature is below its mass. $B + L$-violating processes are fast before the electroweak phase transition and convert the lepton asymmetry into a baryon asymmetry. The CP violating parameters may be related to CP violation in the mixing matrix for the light neutrinos (but this is a model dependent issue [21]).

3. **Electroweak baryogenesis** (for a review see [22]): the source of baryon asymmetry is the interactions of top (anti)quarks with the Higgs field during the electroweak phase transition. CP violation is induced, for example, by supersymmetric interactions. Sphaleron configurations provide baryon number violating interactions.

Departure from thermal equilibrium is provided by the wall between the false vacuum ($\langle\phi\rangle = 0$) and the expanding bubble with the true vacuum, where electroweak symmetry is broken.

1.2.2   The strong CP problem

Non-perturbative QCD effects induce an additional term in the SM Lagrangian,

$$\mathcal{L}_\theta = \frac{\theta_{\text{QCD}}}{32\pi^2}\epsilon_{\mu\nu\rho\sigma}F^{\mu\nu a}F^{\rho\sigma a}. \tag{5}$$

This term violates CP. In particular, it induces a neutron electric dipole moment (EDM). The leading contribution in the chiral limit is given by [23]

$$\begin{aligned} d_N &= \frac{g_{\pi NN}\bar{g}_{\pi NN}}{4\pi^2 M_N}\ln(M_N/m_\pi) \\ &\approx 5\times10^{-16}\,\theta_{\text{QCD}}\,\text{e cm}, \end{aligned} \tag{6}$$

where M_N is the nucleon mass, and $g_{\pi NN}$ ($\bar{g}_{\pi NN}$) is the pseudoscalar coupling (CP-violating scalar coupling) of the pion to the nucleon. (The leading contribution in the large N_c limit was calculated in the Skyrme model [24] and leads to a similar estimate.) The experimental bound on d_N is given by

$$d_N \le 6.3\times10^{-26}\,\text{e cm}\quad[25]. \tag{7}$$

It leads to the following bound on θ_{QCD}:

$$\theta_{\text{QCD}} \lesssim 10^{-10}. \tag{8}$$

Since θ_{QCD} arises from non-perturbative QCD effects, it is impossible to calculate it. Yet, there are good reasons to expect that these effects should yield $\theta_{\text{QCD}} = \mathcal{O}(1)$ (for a clear review of this subject, see [26]). Within the SM, a value as small as (8) is unnatural, since setting θ_{QCD} to zero does not add symmetry to the model. [In particular, as we will see below, CP is violated by $\delta_{\text{KM}} = \mathcal{O}(1)$.] Understanding why CP is so small in the strong interactions is the strong CP problem.

It seems then that the strong CP problem is a clue to new physics. Among the solutions that have been proposed are a massless u-quark (for a review, see [27]), the Peccei-Quinn mechanism [28, 29] and spontaneous CP violation. As concerns the latter, it is interesting to note that in various string theory compactifications, CP is an exact gauge symmetry and must be spontaneously broken [30, 31].

1.2.3   New physics

Another motivation to measure CP violating processes is that almost any extension of the Standard Model provides new sources of CP violation. These sources often allow for significant deviations from the Standard Model predictions. Moreover, various CP violating observables can be calculated with very small hadronic uncertainties. Consequently, CP violation provides an excellent probe of new physics.

1.3   Will new CP violation be observed in experiments?

The SM picture of CP violation is testable because the Kobayashi-Maskawa mechanism is unique and predictive. These features are mainly related to the fact that there is a single phase that is responsible to all CP violation. As a consequence of this situation, one finds two classes of tests:

- Correlations: many independent CP violating observables are correlated within the SM. For example, the SM predicts that the CP asymmetries in $B \to \psi K_S$ and in $B \to \phi K_S$, which proceed through different quark decay processes, are equal to each other. Another important example is the strong SM correlation between CP violation in $B \to \psi K_S$ and in $K \to \pi \nu \bar{\nu}$.

- Zeros: since the KM phase appears in flavor-changing, weak-interaction couplings of quarks, and only if all three generations are involved, many CP violating observables are predicted to be negligibly small. For example, the SM predicts no CP violation in the lepton sector, practically no CP violation in flavor-diagonal processes (*i.e.* a tiny electric dipole moment for the neutron) and very small CP violation in tree level D decays.

In addition, some of the CP violating observables can be calculated with very small hadronic uncertainties.

The strongest argument that new sources of CP violation must exist in Nature comes from baryogenesis. Whether the CP violation that is responsible for baryogenesis would be manifest in measurements of CP asymmetries in B decays depends on two issues:

- The scale of the new CP violation: if the relevant scale is very high, such as in GUT baryogenesis or leptogenesis, the effects cannot be signalled in these measurements. To estimate the limit on the scale, the following three facts are relevant: First, the Standard Model contributions to CP asymmetries in B decays are $\mathcal{O}(1)$. Second, the expected experimental accuracy would reach in some cases the few percent level. Third, the contributions from new physics are expected to be suppressed by $(\Lambda_{\rm EW}/\Lambda_{\rm NP})^2$. The conclusion is that, if the new source of CP violation is related to physics at $\Lambda_{\rm NP} \gg 1$ TeV, it cannot be signalled in B decays. Only if the true mechanism is electroweak baryogenesis, can it potentially affect B decays.

- The flavor dependence of the new CP violation: if it is flavor diagonal, its effects on B decays would be highly suppressed. It can still manifest itself in other, flavor diagonal CP violating observables, such as electric dipole moments.

We conclude that new measurements of CP asymmetries in meson decays are particularly sensitive to new sources of CP violation that come from physics at (or below) the few TeV scale and that are related to flavor changing couplings. This is, for example, the case in certain supersymmetric models of baryogenesis [32, 33]. The search for electric dipole moments can reveal the existence of new flavor diagonal CP violation. Of course, there could be new flavor physics at the TeV scale that is not related to the baryon asymmetry and may give signals in B decays. The best motivated extension of the SM where this situation is likely is that of supersymmetry. We will discuss supersymmetric CP violation in the last chapter.

2   The Kobayashi-Maskawa mechanism

2.1   Yukawa interactions are the source of CP violation

A model of elementary particles and their interactions is defined by three ingredients:
(1) the symmetries of the Lagrangian, (2) the representations of fermions and scalars and
(3) the pattern of spontaneous symmetry breaking. For the Standard Model (SM) the
corresponding ingredients are as follows:

- The gauge symmetry is

$$G_{\rm SM} = SU(3)_{\rm C} \times SU(2)_{\rm L} \times U(1)_{\rm Y}. \tag{9}$$

- There are three fermion generations, each consisting of five representations of $G_{\rm SM}$:

$$Q_{Li}^I(3,2)_{+1/6}, \quad U_{Ri}^I(3,1)_{+2/3}, \quad D_{Ri}^I(3,1)_{-1/3}, \quad L_{Li}^I(1,2)_{-1/2}, \quad E_{Ri}^I(1,1)_{-1}. \tag{10}$$

  Our notation means that, for example, left-handed quarks, Q_L^I, are triplets of
  $SU(3)_{\rm C}$, doublets of $SU(2)_{\rm L}$ and carry hypercharge $Y = +1/6$. The super-index I
  denotes interaction eigenstates. The sub-index $i = 1, 2, 3$ is the flavor (or generation)
  index.

  There is also a single scalar representation,

$$\phi(1,2)_{+1/2}. \tag{11}$$

- The scalar ϕ assumes a vacuum expectation value (VEV),

$$\langle \phi \rangle = \begin{pmatrix} 0 \\ \frac{v}{\sqrt{2}} \end{pmatrix}, \tag{12}$$

  so that the gauge group is spontaneously broken,

$$G_{\rm SM} \to SU(3)_{\rm C} \times U(1)_{\rm EM}. \tag{13}$$

The Standard Model Lagrangian, $\mathcal{L}_{\rm SM}$, is the most general renormalisable Lagrangian
that is consistent with the gauge symmetry (9). It can be divided to three parts:

$$\mathcal{L}_{\rm SM} = \mathcal{L}_{\rm kinetic} + \mathcal{L}_{\rm Higgs} + \mathcal{L}_{\rm Yukawa}. \tag{14}$$

We consider each of these terms in turn,

Kinetic terms

For the kinetic terms, to maintain gauge invariance, one has to replace the derivative with
a covariant derivative:

$$D^\mu = \partial^\mu + i g_s G_a^\mu L_a + i g W_b^\mu T_b + i g' B^\mu Y. \tag{15}$$

Here G_a^μ are the eight gluon fields, W_b^μ the three weak interaction bosons and B^μ the single hypercharge boson. The L_a's are $SU(3)_C$ generators (the 3×3 Gell-Mann matrices $\lambda_a/2$ for triplets, 0 for singlets), the T_b's are $SU(2)_L$ generators (the 2×2 Pauli matrices $\tau_b/2$ for doublets, 0 for singlets), and Y are the $U(1)_Y$ charges. For example, for the left-handed quarks Q_L^I, we have

$$\mathcal{L}_{\text{kinetic}}(Q_L) = i\overline{Q_{Li}^I}\gamma_\mu \left(\partial^\mu + \frac{i}{2}g_s G_a^\mu \lambda_a + \frac{i}{2}g W_b^\mu \tau_b + \frac{i}{6}g' B^\mu\right) Q_{Li}^I, \tag{16}$$

while for the left-handed leptons L_L^I, we have

$$\mathcal{L}_{\text{kinetic}}(L_L) = i\overline{L_{Li}^I}\gamma_\mu \left(\partial^\mu + \frac{i}{2}g W_b^\mu \tau_b - ig' B^\mu\right) L_{Li}^I. \tag{17}$$

These parts of the interaction Lagrangian are always CP conserving.

The Higgs potential

The Higgs potential, which describes the scalar self interactions, is given by:

$$\mathcal{L}_{\text{Higgs}} = \mu^2 \phi^\dagger \phi - \lambda(\phi^\dagger \phi)^2. \tag{18}$$

For the Standard Model scalar sector, where there is a single doublet, this part of the Lagrangian is also CP conserving. For an extended scalar sector, such as that of a two Higgs doublet model, $\mathcal{L}_{\text{Higgs}}$ can be CP violating. Even in the case that it is CP symmetric, it may lead to spontaneous CP violation.

Quark Yukawa interactions

The quark Yukawa interactions are given by

$$-\mathcal{L}_{\text{Yukawa}}^{\text{quarks}} = Y_{ij}^d \overline{Q_{Li}^I}\phi D_{Rj}^I + Y_{ij}^u \overline{Q_{Li}^I}\tilde{\phi}U_{Rj}^I + \text{h.c.}. \tag{19}$$

This part of the Lagrangian is, in general, CP violating. More precisely, CP is violated if and only if [34]

$$\text{Im}\left\{\det[Y^d Y^{d\dagger}, Y^u Y^{u\dagger}]\right\} \neq 0. \tag{20}$$

An intuitive explanation of why CP violation is related to *complex* Yukawa couplings goes as follows. The hermiticity of the Lagrangian implies that $\mathcal{L}_{\text{Yukawa}}$ has its terms in pairs of the form

$$Y_{ij}\overline{\psi_{Li}}\phi\psi_{Rj} + Y_{ij}^*\overline{\psi_{Rj}}\phi^\dagger\psi_{Li}. \tag{21}$$

A CP transformation exchanges the operators

$$\overline{\psi_{Li}}\phi\psi_{Rj} \leftrightarrow \overline{\psi_{Rj}}\phi^\dagger\psi_{Li}, \tag{22}$$

but leaves their coefficients, Y_{ij} and Y_{ij}^*, unchanged. This means that CP is a symmetry of $\mathcal{L}_{\text{Yukawa}}$ if $Y_{ij} = Y_{ij}^*$.

Lepton Yukawa interactions

The lepton Yukawa interactions are given by

$$-\mathcal{L}_{\text{Yukawa}}^{\text{leptons}} = Y_{ij}^e \overline{L_{Li}^I} \phi E_{Rj}^I + \text{h.c.}. \tag{23}$$

It leads, as we will see in the next section, to charged lepton masses but predicts massless neutrinos. Recent measurements of the fluxes of atmospheric and solar neutrinos provide evidence for neutrino masses. That means that $\mathcal{L}_{\text{SM}}$ cannot be a complete description of Nature. The simplest way to allow for neutrino masses is to add dimension-five (and, therefore, non-renormalisable) terms, consistent with the SM symmetry and particle content:

$$-\mathcal{L}_{\text{Yukawa}}^{\text{dim}-5} = \frac{Y_{ij}^\nu}{M} L_i L_j \phi\phi + \text{h.c.}. \tag{24}$$

The parameter M has dimension of mass. The dimensionless couplings Y_{ij}^ν are symmetric ($Y_{ij}^\nu = Y_{ji}^\nu$). We will refer to the SM extended to include the terms $\mathcal{L}_{\text{Yukawa}}^{\text{dim}-5}$ of Equation (24) as the "extended SM" (ESM):

$$\mathcal{L}_{\text{ESM}} = \mathcal{L}_{\text{kinetic}} + \mathcal{L}_{\text{Higgs}} + \mathcal{L}_{\text{Yukawa}} + \mathcal{L}_{\text{Yukawa}}^{\text{dim}-5}. \tag{25}$$

The inclusion of nonrenormalisable terms is equivalent to postulating that the SM is only a low energy effective theory, and that new physics appears at the scale M.

How many independent CP violating parameters are there in $\mathcal{L}_{\text{Yukawa}}^{\text{quarks}}$? Each of the two Yukawa matrices Y^q ($q = u, d$) is 3×3 and complex. Consequently, there are 18 real and 18 imaginary parameters in these matrices. Not all of them are, however, physical. One can think of the quark Yukawa couplings as spurions that break a global symmetry,

$$U(3)_Q \times U(3)_D \times U(3)_U \ \rightarrow \ U(1)_B. \tag{26}$$

This means that there is freedom to remove 9 real and 17 imaginary parameters [the number of parameters in three 3×3 unitary matrices minus the phase related to $U(1)_B$]. We conclude that there are 10 quark flavor parameters: 9 real ones and a single phase. This single phase is the source of CP violation in the quark sector.

How many independent CP violating parameters are there in the lepton Yukawa interactions? The matrix Y^e is a general complex 3×3 matrix and depends, therefore, on 9 real and 9 imaginary parameters. The matrix Y^ν is symmetric and depends on 6 real and 6 imaginary parameters. Not all of these 15 real and 15 imaginary parameters are physical. One can think of the lepton Yukawa couplings as spurions that break (completely) a global symmetry,

$$U(3)_L \times U(3)_E. \tag{27}$$

This means that 6 real and 12 imaginary parameters are not physical. We conclude that there are 12 lepton flavor parameters: 9 real ones and three phases. These three phases induce CP violation in the lepton sector.

2.2 CKM mixing is the (only!) source of CP violation in the quark sector

Upon the replacement $\text{Re}(\phi^0) \to (v+H^0)/\sqrt{2}$ [see Equation (12)], the Yukawa interactions (19) give rise to mass terms:

$$-\mathcal{L}_M^q = (M_d)_{ij}\overline{D_{Li}^I}D_{Rj}^I + (M_u)_{ij}\overline{U_{Li}^I}U_{Rj}^I + \text{h.c.,} \tag{28}$$

where

$$M_q = \frac{v}{\sqrt{2}}Y^q, \tag{29}$$

and we decomposed the $SU(2)_L$ quark doublets into their components:

$$Q_{Li}^I = \begin{pmatrix} U_{Li}^I \\ D_{Li}^I \end{pmatrix}. \tag{30}$$

The mass basis corresponds, by definition, to diagonal mass matrices. We can always find unitary matrices V_{qL} and V_{qR} such that

$$V_{qL}M_qV_{qR}^\dagger = M_q^{\text{diag}} \quad (q = u, d), \tag{31}$$

with M_q^{diag} diagonal and real. The quark mass eigenstates are then identified as

$$q_{Li} = (V_{qL})_{ij}q_{Lj}^I, \quad q_{Ri} = (V_{qR})_{ij}q_{Rj}^I \quad (q = u, d). \tag{32}$$

The charged current interactions for quarks [that is the interactions of the charged $SU(2)_L$ gauge bosons $W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2)$], which in the interaction basis are described by (16), have a complicated form in the mass basis:

$$-\mathcal{L}_{W^\pm}^q = \frac{g}{\sqrt{2}}\overline{u_{Li}}\gamma^\mu(V_{uL}V_{dL}^\dagger)_{ij}d_{Lj}W_\mu^+ + \text{h.c..} \tag{33}$$

The unitary 3×3 matrix,

$$V_{\text{CKM}} = V_{uL}V_{dL}^\dagger, \quad (V_{\text{CKM}}V_{\text{CKM}}^\dagger = 1), \tag{34}$$

is the Cabibbo-Kobayashi-Maskawa (CKM) *mixing matrix* for quarks [35, 1]. A unitary 3×3 matrix depends on nine parameters: three real angles and six phases. The form of this matrix is not unique and in fact there are two distinct freedoms which must be fixed.

1. There is freedom in defining V_{CKM} in that we can permute between the various generations. This freedom is fixed by ordering the up quarks and the down quarks by their masses, *i.e.* $(u_1, u_2, u_3) \to (u, c, t)$ and $(d_1, d_2, d_3) \to (d, s, b)$. The elements of V_{CKM} are written as follows:

$$V_{\text{CKM}} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}. \tag{35}$$

2. There is further freedom in the phase structure of V_{CKM}. Let us define P_q ($q = u, d$) to be diagonal unitary (phase) matrices. Then, if instead of using V_{qL} and V_{qR} for the rotation (32) to the mass basis, we use $\tilde{V}_{qL}$ and $\tilde{V}_{qR}$, defined by $\tilde{V}_{qL} = P_q V_{qL}$ and $\tilde{V}_{qR} = P_q V_{qR}$, we still maintain a legitimate mass basis since M_q^{diag} remains unchanged by such transformations. However, V_{CKM} does change:

$$V_{\text{CKM}} \to P_u V_{\text{CKM}} P_d^*. \tag{36}$$

This freedom is fixed by demanding that V_{CKM} has the minimal number of phases. In the three generation case V_{CKM} has a single phase. (There are five phase differences between the elements of P_u and P_d and, therefore, five of the six phases in the CKM matrix can be removed.) This is the Kobayashi-Maskawa phase δ_{KM} which is the single source of CP violation in the quark sector of the Standard Model [1].

As a result of the fact that V_{CKM} is not diagonal, the $W^{\pm}$ gauge bosons couple to quark (mass eigenstates) of different generations. Within the Standard Model, this is the only source of *flavor changing* quark interactions.

2.3   The three phases in the MNS mixing matrix

The leptonic Yukawa interactions (23) and (24) give rise to mass terms:

$$-\mathcal{L}_M^\ell = (M_e)_{ij}\overline{e_{Li}^I}e_{Rj}^I + (M_\nu)_{ij}\nu_{Li}^I\nu_{Lj}^I + \text{h.c.}, \tag{37}$$

where

$$M_e = \frac{v}{\sqrt{2}}Y^e, \quad M_\nu = \frac{v^2}{2M}Y^\nu, \tag{38}$$

and we decomposed the $SU(2)_{\text{L}}$ lepton doublets into their components:

$$L_{Li}^I = \begin{pmatrix} \nu_{Li}^I \\ e_{Li}^I \end{pmatrix}. \tag{39}$$

We can always find unitary matrices V_{eL} and V_ν such that

$$V_{eL}M_eM_e^\dagger V_{eL}^\dagger = \text{diag}(m_e^2, m_\mu^2, m_\tau^2), \quad V_\nu M_\nu^\dagger M_\nu V_\nu^\dagger = \text{diag}(m_1^2, m_1^2, m_3^2). \tag{40}$$

The charged current interactions for leptons, which in the interaction basis are described by (17), have the following form in the mass basis:

$$-\mathcal{L}_{W^\pm}^\ell = \frac{g}{\sqrt{2}}\overline{e_{Li}}\gamma^\mu(V_{eL}V_\nu^\dagger)_{ij}\nu_{Lj}W_\mu^- + \text{h.c.}. \tag{41}$$

The unitary 3×3 matrix,

$$V_{\text{MNS}} = V_{eL}V_\nu^\dagger, \tag{42}$$

is the Maki-Nakagawa-Sakata (MNS) *mixing matrix* for leptons [36]. As for the CKM matrix, the form of the MNS matrix is not unique: there are the same two freedoms but they are fixed by different conventions:

1. We can permute between the various generations. This freedom is usually fixed in the following way. We order the charged leptons by their masses, *i.e.* $(e_1, e_2, e_3) \rightarrow (e, \mu, \tau)$. For the neutrinos, one takes into account that the interpretation of atmospheric and solar neutrino data in terms of two-neutrino oscillations implies that $\Delta m^2_{\text{ATM}} \gg \Delta m^2_{\text{SOL}}$. It follows that one of the neutrino mass eigenstates is separated in its mass from the other two, which have a smaller mass difference. The convention is to denote this separated state by ν_3. For the remaining two neutrinos, ν_1 and ν_2, the convention is to call the heavier state ν_2. In other words, the three mass eigenstates are defined by the following conventions:

$$|\Delta m^2_{3i}| \gg |\Delta m^2_{21}|, \quad \Delta m^2_{21} > 0. \tag{43}$$

Note in particular that ν_3 can be either heavier or lighter than $\nu_{1,2}$. The elements of V_{MNS} are written as follows:

$$V_{\text{MNS}} = \begin{pmatrix} V_{e1} & V_{e2} & V_{e3} \\ V_{\mu1} & V_{\mu2} & V_{\mu3} \\ V_{\tau1} & V_{\tau2} & V_{\tau3} \end{pmatrix}. \tag{44}$$

2. There is further freedom in the phase structure of V_{MNS}. (In the MNS paper [36] there is no reference to CP violation.) One can change the charged lepton mass basis by the transformation $e_{(L,R)i} \rightarrow e'_{(L,R)i} = (P_e)_{ii} e_{(L,R)i}$, where P_e is a phase matrix. There is, however, no similar freedom to redefine the neutrino mass eigenstates: From Equation (37) one learns that a transformation $\nu_L \rightarrow P_\nu \nu_L$ will introduce phases into the diagonal mass matrix. This is related to the Majorana nature of neutrino masses, assumed in Equation (24). The allowed transformation modifies V_{MNS}:

$$V_{\text{MNS}} \rightarrow P_e V_{\text{MNS}}. \tag{45}$$

This freedom is fixed by demanding that V_{MNS} will have the minimal number of phases. Out of six phases of a unitary 3×3 matrix, the multiplication by P_e can be used to remove three. We conclude that the three generation V_{MNS} matrix has three phases. One is the analog of the Kobayashi-Maskawa phase: it is the only source of CP violation in processes that conserve lepton number, such as neutrino flavor oscillations. The other two phases can affect lepton number changing processes.

With $V_{\text{MNS}} \neq 1$, the $W^\pm$ gauge bosons couple to lepton (mass eigenstates) of different generations. Within the ESM, this is the only source of *flavor changing* lepton interactions.

2.4 The flavor parameters

Examining the quark mass basis, one can easily identify the flavor parameters. In the quark sector, we have six quark masses and four mixing parameters: three mixing angles and a single phase. The fact that there are only three real and one imaginary physical parameters in V_{CKM} can be made manifest by choosing an explicit parameterisation. For example, the standard parameterisation [37], used by the Particle Data Group, is given by

$$V_{\text{CKM}} = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix}, \tag{46}$$

where $c_{ij} \equiv \cos\theta_{ij}$ and $s_{ij} \equiv \sin\theta_{ij}$. The three $\sin\theta_{ij}$ are the three real mixing parameters while δ is the Kobayashi-Maskawa phase. Another, very useful, example is the Wolfenstein parameterisation, where the four mixing parameters are (λ, A, ρ, η) with $\lambda = |V_{us}| = 0.22$ playing the role of an expansion parameter and η representing the CP violating phase [38]:

$$V_{\rm CKM} = \begin{pmatrix} 1 - \frac{\lambda^2}{2} & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \frac{\lambda^2}{2} & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}(\lambda^4). \qquad (47)$$

Various parameterisations differ in the way that the freedom of phase rotation, Equation (36), is used to leave a single phase in $V_{\rm CKM}$. One can define, however, a CP violating quantity in $V_{\rm CKM}$ that is independent of the parameterisation [34]. This quantity, $J_{\rm CKM}$, is defined through

$$\mathrm{Im}[V_{ij}V_{kl}V_{il}^*V_{kj}^*] = J_{\rm CKM} \sum_{m,n=1}^{3} \epsilon_{ikm}\epsilon_{jln}, \quad (i,j,k,l = 1,2,3). \qquad (48)$$

In terms of the explicit parameterisations given above, we have

$$J_{\rm CKM} = c_{12}c_{23}c_{13}^2 s_{12}s_{23}s_{13}\sin\delta \simeq \lambda^6 A^2 \eta. \qquad (49)$$

It is interesting to translate the condition (20) to the language of the flavor parameters in the mass basis. One finds that the following is a necessary and sufficient condition for CP violation in the quark sector of the SM:

$$\Delta m_{tc}^2 \Delta m_{tu}^2 \Delta m_{cu}^2 \Delta m_{bs}^2 \Delta m_{bd}^2 \Delta m_{sd}^2 J_{\rm CKM} \neq 0. \qquad (50)$$

Here

$$\Delta m_{ij}^2 \equiv m_i^2 - m_j^2. \qquad (51)$$

Equation (50) puts the following requirements on the SM in order that it violates CP:

- Within each quark sector, there should be no mass degeneracy;

- None of the three mixing angles should be zero or $\pi/2$;

- The phase should be neither 0 nor π.

For the lepton sector of the ESM, the flavor parameters are the six lepton masses, and six mixing parameters: three mixing angles and three phases. One can parameterise $V_{\rm MNS}$ in a convenient way by factorising it into $V_{\rm MNS} = VP$. Here P is a diagonal unitary matrix that depends on two phases, e.g. $P = \mathrm{diag}(e^{i\phi_1}, e^{i\phi_2}, 1)$, while V can be parameterised in the same way as (46). The advantage of this parameterisation is that for the purpose of analyzing lepton number conserving processes and, in particular, neutrino flavor oscillations, the parameters of P are usually irrelevant and one can use the same Chau-Keung parameterisation as is being used for $V_{\rm CKM}$. (An alternative way to understand these statements is to use a single-phase mixing matrix and put the extra two phases in the neutrino mass matrix. Then it is obvious that the effects of these 'Majorana-phases' always appear in conjunction with a factor of the Majorana mass that is a lepton

number violating parameter.) On the other hand, the Wolfenstein parameterisation (47) is inappropriate for the lepton sector: it assumes $|V_{23}| \ll |V_{12}| \ll 1$, which does not hold here.

In order that the CP violating phase δ in V would be physically meaningful, *i.e.* there would be CP violation that is not related to lepton number violation, a condition similar to (50) should hold:

$$\Delta m_{\tau\mu}^2 \Delta m_{\tau e}^2 \Delta m_{\mu e}^2 \Delta m_{32}^2 \Delta m_{31}^2 \Delta m_{21}^2 J_{\text{MNS}} \neq 0. \tag{52}$$

2.5 The unitarity triangles

A very useful concept is that of the *unitarity triangles*. We will focus on the quark sector, but analogous triangles can be defined in the lepton sector. The unitarity of the CKM matrix leads to various relations among the matrix elements, *e.g.*

$$V_{ud}V_{us}^* + V_{cd}V_{cs}^* + V_{td}V_{ts}^* = 0, \tag{53}$$
$$V_{us}V_{ub}^* + V_{cs}V_{cb}^* + V_{ts}V_{tb}^* = 0, \tag{54}$$
$$V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0. \tag{55}$$

Each of these three relations requires the sum of three complex quantities to vanish and so can be geometrically represented in the complex plane as a triangle. These are "the unitarity triangles", although the term "unitarity triangle" is usually reserved for the relation (55) only. It is a surprising feature of the CKM matrix that all unitarity triangles are equal in area: the area of each unitarity triangle equals $|J_{\text{CKM}}|/2$ while the sign of J_{CKM} gives the direction of the complex vectors around the triangles.

The rescaled unitarity triangle is derived from (55) by (a) choosing a phase convention such that $(V_{cd}V_{cb}^*)$ is real, and (b) dividing the lengths of all sides by $|V_{cd}V_{cb}^*|$. Step (a) aligns one side of the triangle with the real axis, and step (b) makes the length of this side 1. The form of the triangle is unchanged. Two vertices of the rescaled unitarity triangle are thus fixed at $(0,0)$ and $(1,0)$. The coordinates of the remaining vertex correspond to the Wolfenstein parameters (ρ, η). The area of the rescaled unitarity triangle is $|\eta|/2$.

Depicting the rescaled unitarity triangle in the (ρ, η) plane, the lengths of the two complex sides are

$$R_u \equiv \left|\frac{V_{ud}V_{ub}}{V_{cd}V_{cb}}\right| = \sqrt{\rho^2 + \eta^2}, \quad R_t \equiv \left|\frac{V_{td}V_{tb}}{V_{cd}V_{cb}}\right| = \sqrt{(1-\rho)^2 + \eta^2}. \tag{56}$$

The three angles of the unitarity triangle are defined as follows [39, 40]:

$$\alpha \equiv \arg\left[-\frac{V_{td}V_{tb}^*}{V_{ud}V_{ub}^*}\right], \quad \beta \equiv \arg\left[-\frac{V_{cd}V_{cb}^*}{V_{td}V_{tb}^*}\right], \quad \gamma \equiv \arg\left[-\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*}\right]. \tag{57}$$

They are physical quantities and can be independently measured by CP asymmetries in B decays [41, 42, 43, 44, 45]. It is also useful to define the two small angles of the unitarity triangles (54) and (53):

$$\beta_s \equiv \arg\left[-\frac{V_{ts}V_{tb}^*}{V_{cs}V_{cb}^*}\right], \quad \beta_K \equiv \arg\left[-\frac{V_{cs}V_{cd}^*}{V_{us}V_{ud}^*}\right]. \tag{58}$$
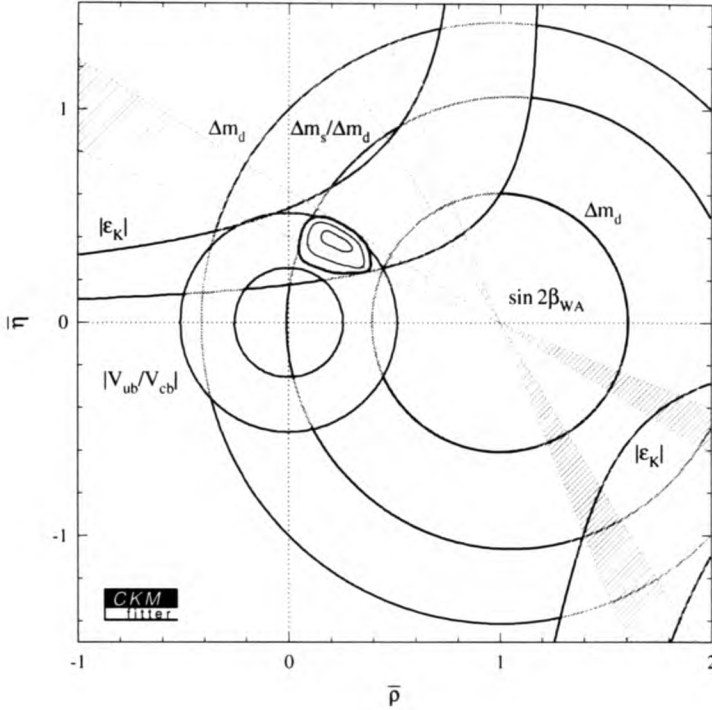
Figure 1. *Present Standard Model constraints and the result from the global CKM fit.*

To make predictions for future measurements of CP violating observables, we need to find the allowed ranges for the CKM phases. There are three ways to determine the CKM parameters (see *e.g.* [46]):

- **Direct measurements** are related to SM tree level processes. At present, we have direct measurements of $|V_{ud}|$, $|V_{us}|$, $|V_{ub}|$, $|V_{cd}|$, $|V_{cs}|$, $|V_{cb}|$ and $|V_{tb}|$.

- **CKM Unitarity** $(V_{CKM}^\dagger V_{CKM} = 1)$ relates the various matrix elements. At present, these relations are useful to constrain $|V_{td}|$, $|V_{ts}|$, $|V_{tb}|$ and $|V_{cs}|$.

- **Indirect measurements** are related to SM loop processes. At present, we constrain in this way $|V_{tb}V_{td}|$ (from Δm_B and Δm_{B_s}) and δ_{KM} or, equivalently, η or β (from ε_K and $a_{\psi K_S}$).

When all available data are taken into account, one finds [47]:

$$\lambda = 0.2221 \pm 0.0021, \quad A = 0.827 \pm 0.058, \tag{59}$$
$$\rho = 0.23 \pm 0.11, \quad \eta = 0.37 \pm 0.08, \tag{60}$$
$$\sin 2\beta = 0.77 \pm 0.08, \quad \sin 2\alpha = -0.21 \pm 0.56, \quad 0.43 \lesssim \sin^2 \gamma \leq 0.91. \tag{61}$$

Of course, there are correlations between the various parameters. The full information in the (ρ, η) plane is given in Figure 1 [47].

2.6   Uniqueness of the standard model picture of CP violation

In the previous subsections, we have learnt several features of CP violation as explained by the Standard Model. It is important to understand that various reasonable (and often well-motivated) extensions of the SM provide examples where some or all of these features do not hold. Furthermore, until a few years ago, none of the special features of the Kobayashi-Maskawa mechanism of CP violation was experimentally tested. This situation has dramatically changed recently. Let us survey some of the SM features, how they can be modified with new physics, and whether experiment has shed light on these questions.

1. δ_{KM} *is the only source of CP violation in meson decays.* This is arguably the most unique feature of the SM and gives the model a strong predictive power. It is violated in almost any low-energy extension. For example, in the supersymmetric extension of the SM there are 44 physical CP violating phases, many of which affect meson decays. The measured value of $a_{\psi K_S}$ is consistent with the correlation between K and B decays that is predicted by the SM. It is therefore very likely that δ_{KM} is indeed the dominant source of CP violation in meson decays.

2. *CP violation is small in* $K \to \pi\pi$ *decays because of flavor suppression and not because CP is an approximate symmetry.* In many (though certainly not all) supersymmetric models, the flavor suppression is too mild, or entirely ineffective, requiring approximate CP to hold. The measurement of $a_{\psi K_S} = \mathcal{O}(1)$ confirms that not all CP violating phases are small.

3. *CP violation appears in both* $\Delta F = 1$ *(decay) and* $\Delta F = 2$ *(mixing) amplitudes.* Superweak models suggest that CP is violated only in mixing amplitudes. The measurement of ε'/ε confirms that there is CP violation in $\Delta S = 1$ processes.

4. *CP is not violated in the lepton sector.* Models that allow for neutrino masses, such as the ESM framework presented above, predict CP violation in leptonic charged current interactions. The data from neutrino oscillation experiments makes it very likely that charged current weak interactions violate CP also in the lepton sector.

5. *CP violation appears only in the charged current weak interactions and in conjunction with flavor changing processes.* Here both various extensions of the SM (such as supersymmetry) and non-perturbative effects within the SM (θ_{QCD}) allow for CP violation in other types of interactions and in flavor-diagonal processes. In particular, it is difficult to avoid flavor-diagonal phases in the supersymmetric framework. The fact that no electric dipole moment has been measured yet poses difficulties to many models with diagonal CP violation (and, of course, is responsible to the strong CP problem within the SM).

6. *CP is explicitly broken.* In various extensions of the scalar sector, it is possible to achieve spontaneous CP violation. It will be very difficult to test this question experimentally.

This situation, where the Standard Model has a very unique and predictive description of CP violation and the number of experimentally measured CP violating observables is

very limited (ε_K, ε'/ε and $a_{\psi K_S}$), is the basis for the strong interest, experimental and theoretical, in CP violation. There are two types of unambiguous tests concerning CP violation in the Standard Model: First, since there is a single source of CP violation, all observables are correlated with each other. For example, the CP asymmetries in $B \rightarrow \psi K_S$ and in $K \rightarrow \pi \nu \bar{\nu}$ are strongly correlated [48, 49, 50]. Second, since CP violation is restricted to flavor changing fermion processes, it is predicted to be highly suppressed in the lepton sector and practically vanish in flavor diagonal processes. For example, the transverse lepton polarisation in semileptonic meson decays, CP violation in $t\bar{t}$ production, and (assuming $\theta_{\text{QCD}} = 0$) the electric dipole moment of the neutron are all predicted to be orders of magnitude below the (present and near future) experimental sensitivity. We conclude that it is very important to search for CP violation in many different systems.

3   Meson decays

In the previous section, we explained how CP violation arises in the Standard Model. In the next three sections, we would like to understand the implications of this theory for the phenomenology of CP violation in K, D and B decays. To do so, we first present a model-independent analysis of CP violation in meson decays. We distinguish between three different types of CP violation in meson decays:

- CP violation in mixing, which occurs when the two neutral mass eigenstate admixtures cannot be chosen to be CP-eigenstates;

- CP violation in decay, which occurs in both charged and neutral decays, when the amplitude for a decay and its CP-conjugate process have different magnitudes;

- CP violation in the interference of decays with and without mixing, which occurs in decays into final states that are common to B^0 and $\bar{B}^0$.

3.1   Notations and formalism

To define these three types and to discuss their theoretical calculation and experimental measurement, we first introduce some notation and formalism. We refer specifically to B meson mixing and decays, but most of our discussion applies equally well to K, B_s and D mesons.

A B^0 meson is made from a b-type antiquark and an d-type quark, while the $\bar{B}^0$ meson is made from a b-type quark and an d-type antiquark. Our phase convention for the CP transformation law of the neutral B mesons is defined by

$$\text{CP}|B^0\rangle = \omega_B|\bar{B}^0\rangle, \quad \text{CP}|\bar{B}^0\rangle = \omega_B^*|B^0\rangle, \quad (|\omega_B| = 1). \tag{62}$$

Physical observables do not depend on the phase factor ω_B.

The light, B_L, and heavy, B_H, mass eigenstates can be written as linear combinations of B^0 and $\bar{B}^0$:

$$\begin{aligned} |B_L\rangle &= p|B^0\rangle + q|\bar{B}^0\rangle, \\ |B_H\rangle &= p|B^0\rangle - q|\bar{B}^0\rangle, \end{aligned} \tag{63}$$

with

$$|q|^2 + |p|^2 = 1. \tag{64}$$

The mass difference Δm_B and the width difference $\Delta\Gamma_B$ are defined as follows:

$$\Delta m \equiv M_H - M_L, \quad \Delta\Gamma \equiv \Gamma_H - \Gamma_L. \tag{65}$$

The average mass and width are given by

$$m_B \equiv \frac{M_H + M_L}{2}, \quad \Gamma_B \equiv \frac{\Gamma_H + \Gamma_L}{2}. \tag{66}$$

It is useful to define dimensionless ratios x and y:

$$x \equiv \frac{\Delta m}{\Gamma}, \quad y \equiv \frac{\Delta\Gamma}{2\Gamma}. \tag{67}$$

The time evolution of the mass eigenstates is simple:

$$\begin{aligned}
|B_H(t)) &= e^{-iM_H t} e^{-\Gamma_H t/2} |B_H\rangle, \\
|B_L(t)) &= e^{-iM_L t} e^{-\Gamma_L t/2} |B_L\rangle.
\end{aligned} \tag{68}$$

The time evolution of the strong interaction eigenstates is complicated and obeys a Schrödinger-like equation,

$$i\frac{d}{dt}\begin{pmatrix} B \\ \bar{B} \end{pmatrix} = \left(M - \frac{i}{2}\Gamma\right)\begin{pmatrix} B \\ \bar{B} \end{pmatrix}, \tag{69}$$

where M and Γ are 2×2 Hermitian matrices. The off-diagonal terms in these matrices, M_{12} and Γ_{12}, are particularly important in the discussion of mixing and CP violation. M_{12} is the dispersive part of the transition amplitude from B^0 to $\bar{B}^0$, while Γ_{12} is the absorptive part of that amplitude. Solving the eigenvalue equation gives

$$(\Delta m)^2 - \frac{1}{4}(\Delta\Gamma)^2 = (4|M_{12}|^2 - |\Gamma_{12}|^2), \qquad \Delta m \Delta\Gamma = 4\mathrm{Re}(M_{12}\Gamma_{12}^*), \tag{70}$$

$$\frac{q}{p} = -\frac{2M_{12}^* - i\Gamma_{12}^*}{\Delta m - \frac{i}{2}\Delta\Gamma} = -\frac{\Delta m - \frac{i}{2}\Delta\Gamma}{2M_{12} - i\Gamma_{12}}. \tag{71}$$

In the B system, $|\Gamma_{12}| \ll |M_{12}|$ (see discussion below), and then, to leading order in $|\Gamma_{12}/M_{12}|$, Equations (70) and (71) can be written as

$$\Delta m_B = 2|M_{12}|, \qquad \Delta\Gamma_B = \frac{2\mathrm{Re}(M_{12}\Gamma_{12}^*)}{|M_{12}|}, \tag{72}$$

$$\frac{q}{p} = -\frac{M_{12}^*}{|M_{12}|}. \tag{73}$$

To discuss CP violation in mixing, it is useful to write Equation (71) to first order in $|\Gamma_{12}/M_{12}|$ [rather than to zeroth order as in (73)]:

$$\frac{q}{p} = -\frac{M_{12}^*}{|M_{12}|}\left[1 - \frac{1}{2}\mathrm{Im}\left(\frac{\Gamma_{12}}{M_{12}}\right)\right]. \tag{74}$$

To discuss CP violation in decay, we need to consider decay amplitudes. The CP transformation law for a final state f is

$$\text{CP}|f\rangle = \omega_f|\bar{f}\rangle, \quad \text{CP}|\bar{f}\rangle = \omega_f^*|f\rangle, \quad (|\omega_f| = 1). \tag{75}$$

For a final CP eigenstate $f = \bar{f} = f_{\text{CP}}$, the phase factor ω_f is replaced by $\eta_{f_{\text{CP}}} = \pm 1$, the CP eigenvalue of the final state. We define the decay amplitudes A_f and $\bar{A}_f$ according to

$$A_f = \langle f|\mathcal{H}_d|B^0\rangle, \quad \bar{A}_f = \langle f|\mathcal{H}_d|\bar{B}^0\rangle, \tag{76}$$

where $\mathcal{H}_d$ is the decay Hamiltonian.

CP relates A_f and $\bar{A}_{\bar{f}}$. There are two types of phases that may appear in A_f and $\bar{A}_{\bar{f}}$. Complex parameters in any Lagrangian term that contributes to the amplitude will appear in complex conjugate form in the CP-conjugate amplitude. Thus their phases appear in A_f and $\bar{A}_{\bar{f}}$ with opposite signs. In the SM these phases occur only in the mixing matrices that parameterise the charged current weak interactions, hence these are often called "weak phases". The weak phase of any single term is convention dependent. However the difference between the weak phases in two different terms in A_f is convention independent because the phase rotations of the initial and final states are the same for every term. A second type of phase can appear in scattering or decay amplitudes even when the Lagrangian is real. Such phases do not violate CP and they appear in A_f and $\bar{A}_{\bar{f}}$ with the same sign. Their origin is the possible contribution from intermediate on-shell states in the decay process, that is an absorptive part of an amplitude that has contributions from coupled channels. Usually the dominant re-scattering is due to strong interactions and hence the designation "strong phases" for the phase shifts so induced. Again only the relative strong phases of different terms in a scattering amplitude have physical content, an overall phase rotation of the entire amplitude has no physical consequences. Thus it is useful to write each contribution to A in three parts: its magnitude A_i; its weak phase term $e^{i\phi_i}$; and its strong phase term $e^{i\delta_i}$. Then, if several amplitudes contribute to $B \to f$, we have

$$\left|\frac{\bar{A}_{\bar{f}}}{A_f}\right| = \left|\frac{\sum_i A_i e^{i(\delta_i - \phi_i)}}{\sum_i A_i e^{i(\delta_i + \phi_i)}}\right|. \tag{77}$$

To discuss CP violation in the interference of decays with and without mixing, we introduce a complex quantity λ_f defined by

$$\lambda_f = \frac{q}{p} \frac{\bar{A}_f}{A_f}. \tag{78}$$

We further define the CP transformation law for the quark fields in the Hamiltonian (a careful treatment of CP conventions can be found in [51]):

$$q \to \omega_q \bar{q}, \quad \bar{q} \to \omega_q^* q, \quad (|\omega_q| = 1). \tag{79}$$

The effective Hamiltonian that is relevant to M_{12} is of the form

$$H_{\text{eff}}^{\Delta b=2} \propto e^{+2i\phi_B} \left[\bar{d}\gamma^\mu(1 - \gamma_5)b\right]^2 + e^{-2i\phi_B} \left[\bar{b}\gamma^\mu(1 - \gamma_5)d\right]^2, \tag{80}$$

where $2\phi_B$ is a CP violating (weak) phase. (We use the SM $V-A$ amplitude, but the results can be generalised to any Dirac structure.) For the B system, where $|\Gamma_{12}| \ll |M_{12}|$, this leads to

$$\frac{q}{p} = \omega_B \omega_b^* \omega_d e^{-2i\phi_B}. \tag{81}$$

(We implicitly assumed that the vacuum insertion approximation gives the correct sign for M_{12}. In general, there is a sign(B_B) factor on the right hand side of Equation (81) [52].) To understand the phase structure of decay amplitudes, we take as an example the $b \to q\bar{q}d$ decay ($q = u$ or c). The decay Hamiltonian is of the form

$$H_d \propto e^{+i\phi_f} \left[\bar{q}\gamma^\mu(1 - \gamma_5)d\right] \left[\bar{b}\gamma_\mu(1 - \gamma_5)q\right] + e^{-i\phi_f} \left[\bar{q}\gamma^\mu(1 - \gamma_5)b\right] \left[\bar{d}\gamma_\mu(1 - \gamma_5)q\right], \quad (82)$$

where ϕ_f is the appropriate weak phase. (Again, for simplicity we use a $V-A$ structure, but the results hold for any Dirac structure.) Then

$$\frac{\bar{A}_{\bar{f}}}{A_f} = \omega_f \omega_B^* \omega_b \omega_d^* e^{-2i\phi_f}. \quad (83)$$

Equations (81) and (83) together imply that for a final CP eigenstate,

$$\lambda_{f_{\mathrm{CP}}} = \eta_{f_{\mathrm{CP}}} e^{-2i(\phi_B + \phi_f)}. \quad (84)$$

3.2 The three types of CP violation in meson decays

3.2.1 CP violation in mixing

$$|q/p| \neq 1. \quad (85)$$

This type of CP violation results from the mass eigenstates being different from the CP eigenstates, and requires a relative phase between M_{12} and Γ_{12}. For the neutral B system, this effect could be observed through the asymmetries in semileptonic decays:

$$a_{\mathrm{SL}} = \frac{\Gamma(\bar{B}^0_{\mathrm{phys}}(t) \to \ell^+\nu X) - \Gamma(B^0_{\mathrm{phys}}(t) \to \ell^-\nu X)}{\Gamma(\bar{B}^0_{\mathrm{phys}}(t) \to \ell^+\nu X) + \Gamma(B^0_{\mathrm{phys}}(t) \to \ell^-\nu X)}. \quad (86)$$

In terms of q and p,

$$a_{\mathrm{SL}} = \frac{1 - |q/p|^4}{1 + |q/p|^4}. \quad (87)$$

CP violation in mixing has been observed in the neutral K system (Re $\varepsilon_K \neq 0$).

In the neutral B system, the effect is expected to be small, $\lesssim \mathcal{O}(10^{-2})$. The reason is that, model independently, the effect cannot be larger than $\mathcal{O}(\Delta\Gamma_B/\Delta m_B)$. The difference in width is produced by decay channels common to B^0 and $\bar{B}^0$. The branching ratios for such channels are at or below the level of 10^{-3}. Since various channels contribute with differing signs, one expects that their sum does not exceed the individual level. Hence, we can safely assume that $\Delta\Gamma_B/\Gamma_B = \mathcal{O}(10^{-2})$. On the other hand, it is experimentally known that $\Delta m_B/\Gamma_B \approx 0.7$.

To calculate a_{SL}, we use (87) and (74), and get:

$$a_{\mathrm{SL}} = \mathrm{Im}(\Gamma_{12}/M_{12}). \quad (88)$$

To predict it in a given model, one needs to calculate M_{12} and Γ_{12}. This involves large hadronic uncertainties, in particular in the hadronisation models for Γ_{12}.

3.2.2 CP violation in decay

$$|\bar{A}_{\bar{f}}/A_f| \neq 1. \tag{89}$$

This appears as a result of interference among various terms in the decay amplitude, and will not occur unless at least two terms have different weak phases and different strong phases. CP asymmetries in charged B decays,

$$a_{f^\pm} = \frac{\Gamma(B^+ \to f^+) - \Gamma(B^- \to f^-)}{\Gamma(B^+ \to f^+) + \Gamma(B^- \to f^-)}, \tag{90}$$

are purely an effect of CP violation in decay. In terms of the decay amplitudes,

$$a_{f^\pm} = \frac{1 - |\bar{A}_{f^-}/A_{f^+}|^2}{1 + |\bar{A}_{f^-}/A_{f^+}|^2}. \tag{91}$$

CP violation in decay has been observed in the neutral K system (Re $\varepsilon'_K \neq 0$).

To calculate $a_{f^\pm}$, we use (91) and (77). For simplicity, we consider decays with contributions from two weak phases and with $A_2 \ll A_1$. We get:

$$a_{f^\pm} = -2(A_2/A_1) \sin(\delta_2 - \delta_1) \sin(\phi_2 - \phi_1). \tag{92}$$

The magnitude and strong phase of any amplitude involve long distance strong interaction physics, and our ability to calculate these from first principles is limited. Thus quantities that depend only on the weak phases are much cleaner than those that require knowledge of the relative magnitudes or strong phases of various amplitude contributions, such as CP violation in decay.

3.2.3 CP violation in interference between decays with and without mixing

$$\text{Im}\,\lambda_{f_{\rm CP}} \neq 0. \tag{93}$$

This effect is the result of interference between a direct decay amplitude and a first-mix-then-decay path to the same final state. For the neutral B system, the effect can be observed by comparing decays into final CP eigenstates of a time-evolving neutral B state that begins at time zero as B^0 to those of the state that begins as $\bar{B}^0$:

$$a_{f_{\rm CP}}(t) = \frac{\Gamma(\bar{B}^0_{\rm phys}(t) \to f_{\rm CP}) - \Gamma(B^0_{\rm phys}(t) \to f_{\rm CP})}{\Gamma(\bar{B}^0_{\rm phys}(t) \to f_{\rm CP}) + \Gamma(B^0_{\rm phys}(t) \to f_{\rm CP})}. \tag{94}$$

This time dependent asymmetry is given, in general, by

$$a_{f_{\rm CP}}(t) = -\frac{1 - |\lambda_{f_{\rm CP}}|^2}{1 + |\lambda_{f_{\rm CP}}|^2} \cos(\Delta m_B t) + \frac{2\text{Im}\lambda_{f_{\rm CP}}}{1 + |\lambda_{f_{\rm CP}}|^2} \sin(\Delta m_B t). \tag{95}$$

In decays with $|\lambda_{f_{\rm CP}}| = 1$, (93) is the only contributing effect:

$$a_{f_{\rm CP}}(t) = \text{Im}\lambda_{f_{\rm CP}} \sin(\Delta m_B t). \tag{96}$$

We often use

$$a_{f_{\rm CP}} \equiv \frac{2\text{Im}\lambda_{f_{\rm CP}}}{1 + |\lambda_{f_{\rm CP}}|^2}. \tag{97}$$

CP violation in the interference of decays with and without mixing has been observed for the neutral K system (Im $\varepsilon_K \neq 0$) and for the neutral B system ($a_{\psi K_S} \neq 0$). In the latter, it is an effect of $\mathcal{O}(1)$. For such cases, the contribution from CP violation in mixing is clearly negligible. For decays that are dominated by a single CP violating phase (for example, $B \to \psi K_S$ and $K_L \to \pi^0 \nu \bar{\nu}$), so that the contribution from CP violation in decay is also negligible, $a_{f_{\mathrm{CP}}}$ is cleanly interpreted in terms of purely electroweak parameters. Explicitly, Im$\lambda_{f_{\mathrm{CP}}}$ gives the relative phase between the $B - \bar{B}$ mixing amplitude and the relevant decay amplitudes [see Equation (84)]:

$$\mathrm{Im}\lambda_{f_{\mathrm{CP}}} = -\eta_{f_{\mathrm{CP}}} \sin[2(\phi_B + \phi_f)]. \tag{98}$$

3.2.4 Direct and indirect CP violation

The terms indirect CP violation and direct CP violation are commonly used in the literature. While various authors use these terms with different meanings, the most useful definition is the following:

- **Indirect CP violation** refers to CP violation in meson decays where the CP violating phases can all be chosen to appear in $\Delta F = 2$ (mixing) amplitudes.

- **Direct CP violation** refers to CP violation in meson decays where some CP violating phases necessarily appear in $\Delta F = 1$ (decay) amplitudes.

Examining Equations (85) and (71), we learn that CP violation in mixing is a manifestation of indirect CP violation. Examining Equations (89) and (76), we learn that CP violation in decay is a manifestation of direct CP violation. Examining Equations (93) and (78), we learn that the situation concerning CP violation in the interference of decays with and without mixing is more subtle. For any single measurement of Im$\lambda_f \neq 0$, the relevant CP violating phase can be chosen by convention to reside in the $\Delta F = 2$ amplitude [$\phi_f = 0$, $\phi_B \neq 0$ in the notation of Equation (84)], and then we would call it indirect CP violation. Consider, however, the CP asymmetries for two different final CP eigenstates (for the same decaying meson), f_a and f_b. Then, a non-zero difference between Imλ_{f_a} and Imλ_{f_b} requires that there exists CP violation in $\Delta F = 1$ processes ($\phi_{f_a} - \phi_{f_b} \neq 0$), namely direct CP violation.

Experimentally, both direct and indirect CP violation have been established. Below we will see that ε_K signifies indirect CP violation while ε'_K signifies direct CP violation.

Theoretically, most models of CP violation (including the Standard Model) have predicted that both types of CP violation exist. There is, however, one class of models, that is *superweak models*, that predict only indirect CP violation. The measurement of $\varepsilon'_K \neq 0$ has excluded this class of models.

4 K decays

Measurements of CP violation have played an enormous role in particle physics. First, the measurement of ε_K in 1964 provided the first evidence that CP is not a symmetry of Nature. This discovery revolutionised the thinking of particle physicists and was essential

for understanding baryogenesis. Second, the measurement in 1988 of ε'_K provided the first evidence for direct CP violation and excluded the superweak scenario. In the future, the search for CP violation in $K \to \pi\nu\bar\nu$ decays will add significantly to our understanding of CP violation.

4.1 ε_K and ε'_K

Historically, a different language from the one used by us has been employed to describe CP violation in $K \to \pi\pi$ and $K \to \pi\ell\nu$ decays. In this section we 'translate' the language of ε_K and ε'_K to our notation. Doing so will make it easy to understand which type of CP violation is related to each quantity.

The two CP violating quantities measured in neutral K decays are

$$\eta_{00} = \frac{\langle \pi^0\pi^0 | \mathcal{H} | K_L \rangle}{\langle \pi^0\pi^0 | \mathcal{H} | K_S \rangle}, \quad \eta_{+-} = \frac{\langle \pi^+\pi^- | \mathcal{H} | K_L \rangle}{\langle \pi^+\pi^- | \mathcal{H} | K_S \rangle}. \tag{99}$$

Define, for $(ij) = (00)$ or $(+-)$,

$$A_{ij} = \langle \pi^i\pi^j | \mathcal{H} | K^0 \rangle, \quad \bar A_{ij} = \langle \pi^i\pi^j | \mathcal{H} | \bar K^0 \rangle, \quad \lambda_{ij} = \left(\frac{q}{p}\right)_K \frac{\bar A_{ij}}{A_{ij}}. \tag{100}$$

Then

$$\eta_{00} = \frac{1 - \lambda_{00}}{1 + \lambda_{00}}, \quad \eta_{+-} = \frac{1 - \lambda_{+-}}{1 + \lambda_{+-}}. \tag{101}$$

The η_{00} and η_{+-} parameters get contributions from CP violation in mixing ($|(q/p)|_K \neq 1$) and from the interference of decays with and without mixing ($\text{Im}\lambda_{ij} \neq 0$) at $\mathcal{O}(10^{-3})$ and from CP violation in decay ($|\bar A_{ij}/A_{ij}| \neq 1$) at $\mathcal{O}(10^{-6})$.

There are two isospin channels in $K \to \pi\pi$ leading to final $(2\pi)_{I=0}$ and $(2\pi)_{I=2}$ states:

$$\begin{aligned} \langle \pi^0\pi^0 | &= \sqrt{1/3}\langle (\pi\pi)_{I=0} | - \sqrt{2/3}\langle (\pi\pi)_{I=2} |, \\ \langle \pi^+\pi^- | &= \sqrt{2/3}\langle (\pi\pi)_{I=0} | + \sqrt{1/3}\langle (\pi\pi)_{I=2} |. \end{aligned} \tag{102}$$

The fact that there are two strong phases allows for CP violation in decay. The possible effects are, however, small (on top of the smallness of the relevant CP violating phases) because the final $I = 0$ state is dominant (this is the $\Delta I = 1/2$ rule). Define

$$A_I = \langle (\pi\pi)_I | \mathcal{H} | K^0 \rangle, \quad \bar A_I = \langle (\pi\pi)_I | \mathcal{H} | \bar K^0 \rangle, \quad \lambda_I = \left(\frac{q}{p}\right)_K \left(\frac{\bar A_I}{A_I}\right). \tag{103}$$

Experimentally, $|A_2/A_0| \approx 1/20$. Instead of η_{00} and η_{+-} we may define two combinations, ε_K and ε'_K, in such a way that the possible effects of direct (indirect) CP violation are isolated into ε'_K (ε_K).

The experimental definition of the ε_K parameter is

$$\varepsilon_K \equiv \frac{1}{3}(\eta_{00} + 2\eta_{+-}). \tag{104}$$

The experimental value is given by Equation (1). To zeroth order in A_2/A_0, we have $\eta_{00} = \eta_{+-} = \varepsilon_K$. However, the specific combination (104) is chosen in such a way that the following relation holds to *first* order in A_2/A_0:

$$\varepsilon_K = \frac{1 - \lambda_0}{1 + \lambda_0}. \tag{105}$$

Since, by definition, only one strong channel contributes to λ_0, there is indeed no CP violation in decay in (105). It is simple to show that Re $\varepsilon_K \neq 0$ is a manifestation of CP violation in mixing while Im $\varepsilon_K \neq 0$ is a manifestation of CP violation in the interference between decays with and without mixing. Since experimentally $\arg \varepsilon_K \approx \pi/4$, the two contributions are comparable. It is also clear that $\varepsilon_K \neq 0$ is a manifestation of indirect CP violation: it could be described entirely in terms of a CP violating phase in the M_{12} amplitude.

The experimental definition of the ε'_K parameter is

$$\varepsilon'_K \equiv \frac{1}{3}(\eta_{+-} - \eta_{00}). \tag{106}$$

The quantity that is actually measured in experiment is

$$1 - \left|\frac{\eta_{00}}{\eta_{+-}}\right|^2 = 6\mathrm{Re}(\varepsilon'/\varepsilon). \tag{107}$$

The world average is given in Equation (2). The theoretical expression is

$$\varepsilon'_K \approx \frac{1}{6}(\lambda_{00} - \lambda_{+-}). \tag{108}$$

Obviously, any type of CP violation which is independent of the final state does not contribute to ε'_K. Consequently, there is no contribution from CP violation in mixing to (108). It is simple to show that Re $\varepsilon'_K \neq 0$ is a manifestation of CP violation in decay while Im $\varepsilon'_K \neq 0$ is a manifestation of CP violation in the interference between decays with and without mixing. Following our explanations in the previous section, we learn that $\varepsilon'_K \neq 0$ is a manifestation of direct CP violation: it requires $\phi_2 - \phi_0 \neq 0$ [where ϕ_I is the CP violating phase in the A_I amplitude defined in (103)].

4.1.1 The ε_K parameter in the Standard Model

An approximate expression for ε_K, that is convenient for calculating it, is given by

$$\varepsilon_K = \frac{e^{i\pi/4}}{\sqrt{2}} \frac{\mathrm{Im} M_{12}}{\Delta m_K}. \tag{109}$$

A few points concerning this expression are worth emphasising:

- Equation (109) is given in a specific phase convention, where A_2 is real. Within the SM, this is a phase convention where $V_{ud}V_{us}^*$ is real, a condition fulfilled in both the standard parameterisation of Equation (46) and the Wolfenstein parameterisation of Equation (47).

- The phase of $\pi/4$ is approximate. It is determined by hadronic parameters and therefore is independent of the electroweak model. Specifically,

$$\arg(\varepsilon_K) \approx \arctan(-2\Delta m_K/\Delta\Gamma_K) \approx \pi/4. \tag{110}$$

- A term of order $2\dfrac{\operatorname{Im} A_0}{\operatorname{Re} A_0}\dfrac{\operatorname{Re} M_{12}}{\operatorname{Im} M_{12}} \lesssim 0.02$ is neglected when (109) is derived.

- There is a large hadronic uncertainty in the calculation of M_{12} coming from long distance contributions. There are, however, good reasons to believe that the long distance contributions are important in $\operatorname{Re} M_{12}$ (where they could be even comparable to the short distance contributions), but negligible in $\operatorname{Im} M_{12}$. To avoid this uncertainty, one uses $\operatorname{Im} M_{12}/\Delta m_K$ with the experimentally measured value of Δm_K, instead of $\operatorname{Im} M_{12}/2\operatorname{Re} M_{12}$ with the theoretically calculated value of $\operatorname{Re} M_{12}$.

- The matrix element $\langle \bar{K}^0|(\bar{s}d)_{V-A}(\bar{s}d)_{V-A}|K^0\rangle$ is yet another source of hadronic uncertainty. If both $\operatorname{Im} M_{12}$ and $\operatorname{Re} M_{12}$ were dominated by short distance contributions, one would use the ratio $\operatorname{Im} M_{12}/\operatorname{Re} M_{12}$ where the matrix element cancels out. However, as explained above, this is not the case.

Within the Standard Model, $\operatorname{Im} M_{12}$ is accounted for by box diagrams. We follow here the notations of reference [53], where precise definitions, numerical values and appropriate references are given. One obtains:

$$\varepsilon_K = e^{i\pi/4} C_\varepsilon B_K \operatorname{Im}(V_{ts}^* V_{td}) \left\{ \operatorname{Re}(V_{cs}^* V_{cd})[\eta_1 S_0(x_c) - \eta_3 S_0(x_c, x_t)] - \operatorname{Re}(V_{ts}^* V_{td})\eta_2 S_0(x_t) \right\}, \tag{111}$$

where $C_\varepsilon \equiv \dfrac{G_F^2 f_K^2 m_K m_W^2}{6\sqrt{2}\pi^2 \Delta m_K}$ is a well known parameter, the η_i are QCD correction factors, S_0 is a kinematic factor, and B_K is the ratio between the matrix element of the four quark operator and its value in the vacuum insertion approximation.

We would like to emphasise the following points:

- CP violation was discovered through the measurement of ε_K. Hence this measurement played a significant role in the history of particle physics.

- For a long time, ε_K has been the only measured CP violating parameter. Roughly speaking, this measurement set the value of δ_{KM} (and, by requiring $\delta_{KM} = \mathcal{O}(1)$, made the KM mechanism plausible) but could not serve as a test of the KM mechanism. (More precisely, a value of $|\varepsilon_K| \gg 10^{-3}$ would have invalidated the KM mechanism, but any value $|\varepsilon_K| \lesssim 10^{-3}$ was acceptable.) It is only the combination of the new measurement of $a_{\psi K_S}$ with ε_K that provides the first precision test of the KM mechanism.

- Within the SM, the smallness of ε_K is not related to suppression of CP violation but rather to suppression of flavor violation. Specifically, it is the smallness of the ratio $|(V_{td}V_{ts})/(V_{ud}V_{us})| \sim \lambda^4$ that explains $|\varepsilon_K| \sim 10^{-3}$.

- Until recently, the measured value of ε_K provided a unique type of information on the CKM phase. For example, the measurement of $\operatorname{sign}(\operatorname{Re} \varepsilon_K) > 0$ tells us that $\eta > 0$ and excludes the lower half of the $\rho - \eta$ plane. Such information cannot be obtained from any CP conserving observable.

- The ε_K constraint gives hyperbolae in the $\rho - \eta$ plane. It is shown in fig. 1. The measured value is consistent with all other CKM-related measurements and further narrows the allowed region.

- The main sources of uncertainty are in the B_K parameter, $B_K = 0.85 \pm 0.15$, and in the $|V_{cb}|^4$ dependence.

- ε_K is an extremely powerful probe of new physics. Its small value poses a problem to any model of new physics where the flavor suppression is less efficient than the GIM mechanism [54] of the SM. For example, the construction of viable supersymmetric models is highly constrained by the requirement that they do not give contributions that are orders of magnitude higher than the experimental value.

4.1.2 The ε'_K parameter in the Standard Model

Direct CP violation in $K \to \pi\pi$ decays was first measured in 1988 [3]. Two recent measurements achieved impressive accuracy:

$$\frac{\varepsilon'}{\varepsilon} = \begin{cases} (20.7 \pm 2.8) \times 10^{-4} & \text{KTeV [6]}, \\ (15.3 \pm 2.6) \times 10^{-4} & \text{NA48 [7]}. \end{cases} \tag{112}$$

In combination with previous results [4, 5], the present world average has an accuracy of order 10% [see Equation (2)].

A convenient approximate expression for ε'_K is given by:

$$\varepsilon'_K = \frac{i}{\sqrt{2}} \left| \frac{A_2}{A_0} \right| e^{i(\delta_2 - \delta_0)} \sin(\phi_2 - \phi_0). \tag{113}$$

Note that:

- The approximations used in (113) are $|q/p| = 1$ and $|A_2/A_0| \ll 1$.

- The phase of ε'_K is determined by hadronic parameters and is, therefore, model independent: $\arg(\varepsilon'_K) = \pi/2 + \delta_2 - \delta_0 \approx \pi/4$. The fact that, accidentally, $\arg(\varepsilon_K) \approx \arg(\varepsilon'_K)$, means that

$$\text{Re}(\varepsilon'/\varepsilon) \approx \varepsilon'/\varepsilon. \tag{114}$$

- $\text{Re}\,\varepsilon'_K \neq 0$ requires $\delta_2 - \delta_0 \neq 0$, consistent with our statement that it is a manifestation of CP violation in decay. $\varepsilon'_K \neq 0$ requires $\phi_2 - \phi_0 \neq 0$, consistent with our statement that it is a manifestation of direct CP violation.

The calculation of ε'/ε within the Standard Model suffers from large hadronic uncertainties. A very naive order of magnitude estimate gives $\varepsilon'/\varepsilon \sim (A_2/A_0)(A_0^{\text{penguin}}/A_0^{\text{tree}}) \sim 10^{-3}$. Note that ε'/ε is not small because of small CP violating parameters but because of hadronic parameters.

The value of the phase β_K cancels in the ratio ε'/ε and therefore did not affect our estimate. In actual calculations, one usually uses the experimental value of ε_K and the theoretical expression for ε'_K. Then the expression for ε'/ε depends on the CP violating phase.

The detailed calculation of ε'/ε is complicated. There are several comparable contributions with differing signs. The final result can be written in the form (for details and references, see [53]):

$$
\begin{aligned}
\frac{\varepsilon'}{\varepsilon} &= \operatorname{Im}(V_{td}V_{ts}^*)\left[P^{(1/2)} - P^{(3/2)}\right] \\
&\approx 13\,\operatorname{Im}(V_{td}V_{ts}^*)\left(\frac{110\,MeV}{m_s(2\,GeV)}\right)^2\left(\frac{\Lambda_{\overline{MS}}^{(4)}}{340\,MeV}\right) \\
&\quad \times \left[B_6^{(1/2)}(1-\Omega_{\eta+\eta'}) - 0.4B_8^{(3/2)}\left(\frac{m_t}{165\,GeV}\right)^{2.5}\right].
\end{aligned}
\tag{115}
$$

We have omitted here a phase factor using the approximation $\arg(\varepsilon_K) = \arg(\varepsilon'_K)$. $P^{(1/2)}$, which is dominated by QCD penguins, gives the contributions from $\Delta I = 1/2$ transitions, while $P^{(3/2)}$, which is dominated by electroweak penguins, gives the contributions from $\Delta I = 3/2$ transitions. The $B_6^{(1/2)}$ and $B_8^{(3/2)}$ factors parameterise the corresponding hadronic matrix elements. The QCD penguin contributions are suppressed by isospin breaking effects ($m_u \neq m_d$), parameterised by $\Omega_{\eta+\eta'}$. The resulting estimates vary in the range [53]:

$$
\operatorname{Re}(\varepsilon'/\varepsilon)^{\text{SM}} = (0.5 - 4) \times 10^{-3}.
\tag{116}
$$

We would like to emphasise the following points:

- Direct CP violation was discovered through the measurement of ε'.

- The SM range (116) is consistent with the experimental result (2).

- The main sources of uncertainties lie then in the parameters m_s, $B_6^{(1/2)}$, $B_8^{(3/2)}$, $\Omega_{\eta+\eta'}$ and $\Lambda_{\overline{MS}}^{(4)}$. The importance of these uncertainties is increased because of the cancellation between the two contributions in (115).

- The large hadronic uncertainties make it difficult to use the experimental value of ε'/ε to constrain the CKM parameters. Still, a negative value or a value much smaller than 10^{-4} would have been very puzzling in the context of the SM.

- The experimental result is useful in probing and constraining new physics.

4.2   CP violation in $K \to \pi\nu\bar{\nu}$

Observing CP violation in the rare $K \to \pi\nu\bar{\nu}$ decays would be experimentally very challenging and theoretically very rewarding. It is very different from the CP violation that has been observed in $K \to \pi\pi$ decays which is small and involves theoretical uncertainties. Similar to the CP asymmetry in $B \to \psi K_S$, it is predicted to be large and can be cleanly interpreted. Furthermore, observation of the $K_L \to \pi^0\nu\bar{\nu}$ decay at the rate predicted by the Standard Model will provide further evidence that CP violation cannot be attributed to mixing ($\Delta S = 2$) processes only, as in superweak models.

Define

$$
A_{\pi^0\nu\bar{\nu}} = \langle \pi^0\nu\bar{\nu}|\mathcal{H}|K^0\rangle, \quad \bar{A}_{\pi^0\nu\bar{\nu}} = \langle \pi^0\nu\bar{\nu}|\mathcal{H}|\bar{K}^0\rangle, \quad \lambda_{\pi\nu\bar{\nu}} = \left(\frac{q}{p}\right)_K \frac{\bar{A}_{\pi^0\nu\bar{\nu}}}{A_{\pi^0\nu\bar{\nu}}}.
\tag{117}
$$

The ratio between the neutral K decay rates is then

$$\frac{\Gamma(K_L \to \pi^0 \nu \bar{\nu})}{\Gamma(K_S \to \pi^0 \nu \bar{\nu})} = \frac{1 + |\lambda_{\pi\nu\bar{\nu}}|^2 - 2\mathrm{Re}\lambda_{\pi\nu\bar{\nu}}}{1 + |\lambda_{\pi\nu\bar{\nu}}|^2 + 2\mathrm{Re}\lambda_{\pi\nu\bar{\nu}}}. \tag{118}$$

We learn that the $K_L \to \pi^0 \nu \bar{\nu}$ decay rate vanishes in the CP limit ($\lambda_{\pi\nu\bar{\nu}} = 1$), as expected on general grounds [55]. (The CP conserving contributions were explicitly calculated within the Standard Model [56] and within its extensions with massive neutrinos [57] and with extra scalars [58] and found to be negligible.)

CP violation in decay and in mixing are expected to be negligibly small, of order 10^{-5} and 10^{-3}, respectively. Consequently, $\lambda_{\pi\nu\bar{\nu}}$ is, to an excellent approximation, a pure phase. Defining $2\theta_K$ to be the relative phase between the $K-\bar{K}$ mixing amplitude and twice the $s \to d\nu\bar{\nu}$ decay amplitude, namely $\lambda_{\pi\nu\bar{\nu}} = e^{2i\theta_K}$, we get from (118):

$$\frac{\Gamma(K_L \to \pi^0 \nu \bar{\nu})}{\Gamma(K_S \to \pi^0 \nu \bar{\nu})} = \tan^2 \theta_K. \tag{119}$$

Using the isospin relation $A(K^0 \to \pi^0 \nu \bar{\nu})/A(K^+ \to \pi^+ \nu \bar{\nu}) = 1/\sqrt{2}$, we get

$$a_{\pi\nu\bar{\nu}} \equiv \frac{\Gamma(K_L \to \pi^0 \nu \bar{\nu})}{\Gamma(K^+ \to \pi^+ \nu \bar{\nu})} = \sin^2 \theta_K. \tag{120}$$

The present experimental searches give

$$\begin{aligned}
\mathcal{B}(K^+ \to \pi^+ \nu \bar{\nu}) &= (1.5^{+3.4}_{-1.2}) \times 10^{-10} \quad [59], \\
\mathcal{B}(K_L \to \pi^0 \nu \bar{\nu}) &< 5.9 \times 10^{-7} \quad [60].
\end{aligned} \tag{121}$$

Equation (120) implies that $a_{\pi\nu\bar{\nu}} \leq 1$. This inequality is based on isospin considerations only. Consequently a measurement of $\Gamma(K^+ \to \pi^+ \nu \bar{\nu})$ can be used to set a model independent upper limit on $\Gamma(K_L \to \pi^0 \nu \bar{\nu})$ [61]:

$$\mathcal{B}(K_L \to \pi^0 \nu \bar{\nu}) < 4.4 \, \mathcal{B}(K^+ \to \pi^+ \nu \bar{\nu}). \tag{122}$$

From the range given in (121) for the K^+ decay, the isospin bound on the K_L decay is $\mathcal{B}(K_L \to \pi^0 \nu \bar{\nu}) < 2.6 \times 10^{-9}$, which is more than two orders of magnitude below the direct bound.

Within the Standard Model, the $K \to \pi \nu \bar{\nu}$ decays are dominated by short distance Z-penguins and box diagrams and can be expressed in terms of ρ and η (see [53] for details and references)

$$\begin{aligned}
\mathcal{B}(K^+ \to \pi^+ \nu \bar{\nu}) &= 4.11 \times 10^{-11}[X(x_t)]^2 A^4 \left[\eta^2 + (\rho_0 - \rho)^2\right], \\
\mathcal{B}(K_L \to \pi^0 \nu \bar{\nu}) &= 1.80 \times 10^{-10}[X(x_t)]^2 A^4 \eta^2.
\end{aligned} \tag{123}$$

Here $\rho_0 = 1 + \dfrac{P_0(X)}{A^2 X(x_t)}$, and $X(x_t)$ and $P_0(X)$ represent the electroweak loop contributions in NLO for the top quark and for the charm quark, respectively.

We would like to emphasise the following points:

- The $K \to \pi \nu \bar{\nu}$ decays are theoretically clean. The main theoretical uncertainty in the K^+ decay is related to the strong dependence of the charm contribution on the renormalisation scale and the QCD scale, $P_0(X) = 0.42 \pm 0.06$. The K_L decay has hadronic uncertainties smaller than a percent.

- In the future, these decays will provide excellent $\rho - \eta$ constraints.

- Present constraints on the CKM parameters give the SM predictions [47]:

$$
\begin{aligned}
\mathcal{B}(K^+ \to \pi^+ \nu \bar{\nu}) &= (7.0 \pm 1.9) \times 10^{-11}, \\
\mathcal{B}(K_L \to \pi^0 \nu \bar{\nu}) &= (2.9 \pm 1.1) \times 10^{-11}.
\end{aligned}
\tag{124}
$$

The experimental range for the K^+ decay (121) is then consistent with the SM but not yet accurate enough to constrain it, while the experimental bound on the K_L decay is still four orders of magnitude above the SM range.

- The CP violations in $K \to \pi \nu \bar{\nu}$ and in $B \to \psi K_S$ are strongly correlated and can provide the most stringent test of the Kobayashi-Maskawa mechanism.

- The $K \to \pi \nu \bar{\nu}$ decays are interesting probes of CP violation related to new physics.

5 D decays

Within the Standard Model, $D - \bar{D}$ mixing is expected to be well below the experimental bound. Furthermore, effects related to CP violation in $D - \bar{D}$ mixing are expected to be negligibly small since this mixing is described to an excellent approximation by physics of the first two generations. An experimental observation of $D - \bar{D}$ mixing close to the present bound or, more strongly, of related CP violation, will then be evidence for New Physics.

To explain how $D - \bar{D}$ mixing is searched for and how CP violation can be signalled, we use notation similar to that of the B system. We thus use Equation (63) to define the two mass eigenstates $|D_{1,2}\rangle$, Equation (66) to define the average width Γ, Equation (67) to define the width and mass differences y and x, Equation (76) to define the decay amplitudes A_f and $\bar{A}_f$ and Equation (78) to define λ_f.

5.1 $D \to K\pi$ and $D \to KK$ decays

Processes that are relevant to the most sensitive measurements at present are the doubly-Cabibbo-suppressed $D^0 \to K^+ \pi^-$ decay, the singly-Cabibbo-suppressed $D^0 \to K^+ K^-$ decay, the Cabibbo-favored $D^0 \to K^- \pi^+$ decay, and the three CP-conjugate decay processes. We follow here the analysis presented in reference [62]. We write down approximate expressions for the time-dependent decay rates that are valid for times $t \lesssim 1/\Gamma$. We take into account the experimental information that x, y and $\tan \theta_c$ are small. In particular, the smallness of $\tan \theta_c$ implies that

$$
|\lambda_{K^+\pi^-}^{-1}| \ll 1; \quad |\lambda_{K^-\pi^+}| \ll 1.
\tag{125}
$$

We expand each of the rates only to the order that is relevant to present measurements:

$$\Gamma[D^0(t) \to K^+\pi^-] = e^{-\Gamma t}|\bar{A}_{K^+\pi^-}|^2 |q/p|^2$$
$$\times \left\{ |\lambda^{-1}_{K^+\pi^-}|^2 + [\mathrm{Re}(\lambda^{-1}_{K^+\pi^-})y + \mathrm{Im}(\lambda^{-1}_{K^+\pi^-})x]\Gamma t + \frac{1}{4}(y^2+x^2)(\Gamma t)^2 \right\},$$

$$\Gamma[\overline{D^0}(t) \to K^-\pi^+] = e^{-\Gamma t}|A_{K^-\pi^+}|^2 |p/q|^2$$
$$\times \left\{ |\lambda_{K^-\pi^+}|^2 + [\mathrm{Re}(\lambda_{K^-\pi^+})y + \mathrm{Im}(\lambda_{K^-\pi^+})x]\Gamma t + \frac{1}{4}(y^2+x^2)(\Gamma t)^2 \right\} \quad (126)$$

$$\Gamma[D^0(t) \to K^+K^-] = e^{-\Gamma t}|A_{K^+K^-}|^2 \left\{ 1 + [\mathrm{Re}(\lambda_{K^+K^-})y - \mathrm{Im}(\lambda_{K^+K^-})x]\Gamma t \right\},$$

$$\Gamma[\overline{D^0}(t) \to K^+K^-] = e^{-\Gamma t}|\bar{A}_{K^+K^-}|^2 \left\{ 1 + [\mathrm{Re}(\lambda^{-1}_{K^+K^-})y - \mathrm{Im}(\lambda^{-1}_{K^+K^-})x]\Gamma t \right\}, \quad (127)$$

$$\Gamma[D^0(t) \to K^-\pi^+] = e^{-\Gamma t}|A_{K^-\pi^+}|^2,$$

$$\Gamma[\overline{D^0}(t) \to K^+\pi^-] = e^{-\Gamma t}|\bar{A}_{K^+\pi^-}|^2. \quad (128)$$

Within the Standard Model, the physics of $D^0-\overline{D^0}$ mixing and of the tree level decays is dominated by the first two generations and, consequently, CP violation can be safely neglected. In almost all 'reasonable' extensions of the SM, the six decay modes of Equations (126), (127) and (128) are still dominated by the SM CP conserving contributions [63, 64]. On the other hand, there could be new short distance, possibly CP violating, contributions to the mixing amplitude M_{12}. Allowing for only such effects of new physics, the picture of CP violation is simplified since there is no direct CP violation. The effects of indirect CP violation can be parameterised in the following way [65]:

$$\begin{aligned} |q/p| &= R_m, \\ \lambda^{-1}_{K^+\pi^-} &= \sqrt{R}\, R_m^{-1}\, e^{-i(\delta+\phi_D)}, \\ \lambda_{K^-\pi^+} &= \sqrt{R}\, R_m\, e^{-i(\delta-\phi_D)}, \\ \lambda_{K^+K^-} &= -R_m\, e^{i\phi_D}. \end{aligned} \quad (129)$$

Here R and R_m are real and positive dimensionless numbers. CP violation in mixing is related to $R_m \neq 1$ while CP violation in the interference of decays with and without mixing is related to $\sin\phi_D \neq 0$. The choice of phases and signs in (129) is consistent with having $\phi_D = 0$ in the SM and $\delta = 0$ in the $SU(3)$ limit. We further define

$$\begin{aligned} x' &\equiv x\cos\delta + y\sin\delta, \\ y' &\equiv y\cos\delta - x\sin\delta. \end{aligned} \quad (130)$$

With our assumption that there is no direct CP violation in the processes that we study, and using the parameterisations (129) and (67), we can rewrite Equations (126), (127) and (128) as follows:

$$\Gamma[D^0(t) \to K^+\pi^-] = e^{-\Gamma t}|A_{K^-\pi^+}|^2$$
$$\times \left[R + \sqrt{R}R_m(y'\cos\phi_D - x'\sin\phi_D)\Gamma t + \frac{R_m^2}{4}(y^2+x^2)(\Gamma t)^2 \right],$$

$$\Gamma[\overline{D^0}(t) \to K^-\pi^+] = e^{-\Gamma t}|A_{K^-\pi^+}|^2$$
$$\times \left[R + \sqrt{R}R_m^{-1}(y'\cos\phi_D + x'\sin\phi_D)\Gamma t + \frac{R_m^{-2}}{4}(y^2+x^2)(\Gamma t)^2 \right] \quad (131)$$

$$\Gamma[D^0(t) \to K^+K^-] = e^{-\Gamma t}|A_{K^+K^-}|^2\left[1 - R_m(y\cos\phi_D - x\sin\phi_D)\Gamma t\right],$$
$$\Gamma[\overline{D^0}(t) \to K^+K^-] = e^{-\Gamma t}|A_{K^+K^-}|^2\left[1 - R_m^{-1}(y\cos\phi_D + x\sin\phi_D)\Gamma t\right], \tag{132}$$
$$\Gamma[D^0(t) \to K^-\pi^+] = \Gamma[\overline{D^0}(t) \to K^+\pi^-] = e^{-\Gamma t}|A_{K^-\pi^+}|^2. \tag{133}$$

Of particular interest is the linear term in Equation (131) which is potentially CP violating [66, 67]. It is useful to define a CP violating quantity $a_{D\to K\pi}$ which depends on the six measurable coefficients in (131):

$$
\begin{aligned}
a_{D\to K\pi} &= \frac{\mathrm{Re}(\lambda_{K^-\pi^+})y + \mathrm{Im}(\lambda_{K^-\pi^+})x}{2|\lambda_{K^-\pi^+}|\sqrt{x^2+y^2}} - \frac{\mathrm{Re}(\lambda_{K^+\pi^-}^{-1})y + \mathrm{Im}(\lambda_{K^+\pi^-}^{-1})x}{2|\lambda_{K^+\pi^-}^{-1}|\sqrt{x^2+y^2}} \\
&= \frac{x'}{\sqrt{x^2+y^2}}\sin\phi_D. \tag{134}
\end{aligned}
$$

Observing $a_{D\to K\pi} \neq 0$ would be the most convincing evidence for new physics in $D-\bar{D}$ mixing.

The CLEO measurement [68] gives the coefficient of each of the three terms $[1, \Gamma t$ and $(\Gamma t)^2]$ in the doubly-Cabibbo suppressed decays (131). Such measurements allow a fit to the parameters R, R_m, $x'\sin\phi$, $y'\cos\phi$, and $x^2 + y^2$. Fit A of reference [68] quotes the following one sigma ranges:

$$
\begin{aligned}
R &= (0.48 \pm 0.13) \times 10^{-2}, \\
y'\cos\phi_D &= (-2.5^{+1.4}_{-1.6}) \times 10^{-2}, \\
x' &= (0.0 \pm 1.5) \times 10^{-2}, \tag{135} \\
A_m &= 0.23^{+0.63}_{-0.80}, \\
\sin\phi_D &= 0.0 \pm 0.6.
\end{aligned}
$$

It is assumed here that R_m is not very different from one and can be parameterised by a small parameter A_m,

$$R_m^{\pm 2} = 1 \pm A_m. \tag{136}$$

We would like to make two further comments in this regard:

- The experimental results in Equation (135) do not show any signal of CP violation, that is, both $\sin\phi_D$ and A_m are consistent with zero. Consequently, there is no hint of new physics in the present results.

- To test models of new physics, it would be useful to know the value of the strong phase δ. Such an estimate is a difficult theoretical task [69, 70, 71] but experimental data on related channels would be useful [72, 73].

For the singly-Cabibbo suppressed modes of Equation (132), several experiments fit the time dependent decay rates to pure exponentials. We define $\hat{\Gamma}$ to be the parameter that is extracted in this way. More explicitly, for a time dependent decay rate with $\Gamma[D(t) \to f] \propto e^{-\Gamma t}(1 - z\Gamma t + \cdots)$, where $|z| \ll 1$, we have $\hat{\Gamma}(D \to f) = \Gamma(1 + z)$. The above equations imply the following relations:

$$
\begin{aligned}
\hat{\Gamma}(D^0 \to K^+K^-) &= \Gamma\left[1 + R_m(y\cos\phi_D - x\sin\phi_D)\right], \\
\hat{\Gamma}(\overline{D^0} \to K^+K^-) &= \Gamma\left[1 + R_m^{-1}(y\cos\phi_D + x\sin\phi_D)\right], \tag{137} \\
\hat{\Gamma}(D^0 \to K^-\pi^+) &= \hat{\Gamma}(\overline{D^0} \to K^+\pi^-) = \Gamma.
\end{aligned}
$$

Note that deviations of $\hat{\Gamma}(D \to K^+ K^-)$ from Γ do not require that $y \neq 0$. They can be accounted for by $x \neq 0$ and $\sin \phi_D \neq 0$, but then they have a different sign in the D^0 and $\overline{D^0}$ decays. Combining the two $D \to K^+ K^-$ modes, one obtains the CP conserving quantity $y_{\rm CP}$:

$$y_{\rm CP} \equiv \frac{\hat{\Gamma}(D \to K^+ K^-)}{\hat{\Gamma}(D^0 \to K^- \pi^+)} - 1 = y \cos \phi_D - \frac{A_m}{2} x \sin \phi, \tag{138}$$

where we made the approximations of zero production asymmetry and small A_m [62]. The one sigma ranges measured by various experiments are given by

$$y_{\rm CP} = \begin{cases} (3.4 \pm 1.6) \times 10^{-2} & \text{FOCUS [74]} \\ (0.8 \pm 3.1) \times 10^{-2} & \text{E791 [75]} \\ (-1.1 \pm 2.9) \times 10^{-2} & \text{CLEO [76]} \\ (0.5 \pm 1.3) \times 10^{-2} & \text{BELLE [77]} \end{cases} \tag{139}$$

giving a world average of

$$y_{\rm CP} = (1.3 \pm 0.9) \times 10^{-2}. \tag{140}$$

Finally, we note that direct CP violation has been searched for in the Cabibbo-favored [78], singly-Cabibbo-suppressed [79, 80, 81] and doubly-Cabibbo-suppressed [68] decays with all results consistent with zero.

We conclude that at present there is no evidence for mixing and certainly not for CP violation in the neutral D system. These results are consistent with the SM and constrain models of new physics. If evidence is found in the future, the $D \to K\pi$ and $D \to KK$ decays will provide rich enough information that we will be able to point out the origin of the signals in much detail.

6   *B* decays

6.1   CP violation in mixing

CP violation in mixing is related to a non-zero value for the following quantity [see (74)]:

$$1 - \left|\frac{q}{p}\right| \simeq \frac{1}{2} \text{Im} \left(\frac{\Gamma_{12}}{M_{12}}\right). \tag{141}$$

The effect can be isolated by measuring the asymmetry in semileptonic decays [see (87)]:

$$a_{\rm SL} \simeq 2(1 - |q/p|) \simeq \text{Im}(\Gamma_{12}/M_{12}). \tag{142}$$

This has been searched for in several experiments, with sensitivity at the level of 10^{-2}:

$$a_{\rm SL} = \begin{cases} (1.4 \pm 4.2) \times 10^{-2} & \text{CLEO [82]} \\ (0.4 \pm 5.7) \times 10^{-2} & \text{OPAL [83]} \\ (-1.2 \pm 2.8) \times 10^{-2} & \text{ALEPH [84]} \\ (0.48 \pm 1.85) \times 10^{-2} & \text{BABAR [85]} \end{cases} \tag{143}$$

giving a world average of

$$a_{\rm SL} = (0.2 \pm 1.4) \times 10^{-2}. \tag{144}$$

As explained above, in the B_d system we expect, in a model independent way, that $|\Gamma_{12}/M_{12}| \ll 1$. Within any given model we can actually calculate the two quantities from quark diagrams. Within the SM, M_{12} is given by box diagrams. For both the B_d and B_s systems, the long distance contributions are expected to be negligible and the calculation of these diagrams with a high loop momentum is a very good approximation. Γ_{12} is calculated from a cut of box diagrams [86]. Since the cut of a diagram always involves on-shell particles and thus long distance physics, the calculation is, at best, a reasonable approximation to Γ_{12}. (For $\Gamma_{12}(B_s)$ it has been shown that local quark-hadron duality holds exactly in the simultaneous limit of small velocity and large number of colors. We thus expect an uncertainty of $\mathcal{O}(1/N_C) \sim 30\%$ [87, 88]. For $\Gamma_{12}(B_d)$ the small velocity limit is not as good an approximation but an uncertainty of order 50% still seems a reasonable estimate [89].)

Within the Standard Model, M_{12} is dominated by top-mediated box diagrams (see [53] for details and references):

$$M_{12} = \frac{G_F^2}{12\pi^2} m_B m_W^2 \eta_B B_B f_B^2 (V_{tb} V_{td}^*)^2 S_0(x_t), \tag{145}$$

where $S_0(x_t)$ is a kinematic factor, η_B is a QCD correction, and $B_B f_B^2$ parameterises the hadronic matrix element. For Γ_{12}, we have [91, 92, 93]

$$\begin{aligned} \Gamma_{12} = {} & -\frac{G_F^2}{24\pi} m_B m_b^2 B_B f_B^2 (V_{tb} V_{td}^*)^2 \\ & \times \left[\frac{5}{3} \frac{m_B^2}{(m_b+m_d)^2} \frac{B_S}{B_B} (K_2 - K_1) + \frac{4}{3}(2K_1 + K_2) + 8(K_1 + K_2) \frac{m_c^2}{m_b^2} \frac{V_{cb} V_{cd}^*}{V_{tb} V_{td}^*} \right], \end{aligned} \tag{146}$$

where $K_1 = -0.39$ and $K_2 = 1.25$ [93] are combinations of Wilson coefficients and B_S parameterises the $(S - P)^2$ matrix element. New physics usually takes place at a high energy scale and is relevant to the short distance part only. Therefore, the SM estimate in Equation (146) remains valid model independently. Combining (145) and (146), we learn that $|\Gamma_{12}/M_{12}| = \mathcal{O}(m_b^2/m_t^2)$, which confirms our model independent order of magnitude estimate, $|\Gamma_{12}/M_{12}| \lesssim 10^{-2}$. For the imaginary part of this ratio, we have

$$a_{\mathrm{SL}} = \mathrm{Im} \frac{\Gamma_{12}}{M_{12}} \approx -1.4 \times 10^{-3} \frac{\eta}{(1-\rho)^2 + \eta^2}. \tag{147}$$

The suppression by a factor of $\mathcal{O}(10)$ of a_{SL} compared to $|\Gamma_{12}/M_{12}|$ comes from the fact that the leading contribution to Γ_{12} has the same phase as M_{12}. Consequently we have $a_{\mathrm{SL}} = \mathcal{O}(m_c^2/m_t^2)$. The CKM factor does not give any further significant suppression, $\mathrm{Im}\left(\frac{V_{cb} V_{cd}^*}{V_{tb} V_{td}^*}\right) = \mathcal{O}(1)$. In contrast, for the B_s system, where the same expression holds except that V_{cd}/V_{td} is replaced by V_{cs}/V_{ts}, there is an additional CKM suppression from $\mathrm{Im}\left(\frac{V_{cb} V_{cs}^*}{V_{tb} V_{ts}^*}\right) = \mathcal{O}(\lambda^2)$.

In the SM and in most of its reasonable extensions, both Γ_{12} and $b \to c\bar{c}s$ transitions are dominated by SM tree level decays. Consequently, new physics affects a_{SL} and $a_{\psi K_S}$ only through its contributions to M_{12}. This leads to interesting correlations between a_{SL} and $a_{\psi K_S}$ that can be used to probe flavor parameters [94, 95]. Conversely, one can use the measured value of $a_{\psi K_S}$ to give model independent predictions for a_{SL} [96, 97].

6.2 Penguin pollution

In purely hadronic B decays, CP violation in decay and in the interference of decays with and without mixing is $\geq \mathcal{O}(10^{-2})$. We can therefore safely neglect CP violation in mixing in the following discussion and use

$$\frac{q}{p} = \frac{V_{tb}^* V_{td}}{V_{tb} V_{td}^*} \omega_B. \tag{148}$$

(From here on we omit the convention-dependent quark phases ω_q defined in Equation (79). Our final expressions for physical quantities are of course unaffected by such omission.)

A crucial aspect of our discussion is the number of relevant weak phases for a given decay process:

- If there is a single weak phase that dominates the decay, CP violation in decay will be small and difficult to observe. On the other hand, CP asymmetries in neutral B decays into final CP eigenstates are subject to clean theoretical interpretation: we will either have precise measurements of CKM parameters or be provided with unambiguous evidence for new physics.

- If there are two (or more) weak phases that contribute comparably, hadronic uncertainties will appear in the theoretical interpretation of CP violation in the interference of decays with and without mixing. On the other hand, if there are also large strong phase differences, CP violation in decay can be observed in the corresponding charged and neutral B decays.

In many cases of interest, different weak phases are carried by tree and penguin contributions. The difficulties arising from hadronic uncertainties related to comparable tree and penguin contributions have become known as "penguin pollution."

To illustrate the problem, we will consider two relevant CP asymmetries. First, the CP asymmetry in $B \to \psi K_S$ is an example of a case where the penguin pollution is negligibly small and a theoretically very clean interpretation of the experimental measurement is possible. Second, the CP asymmetry in $B \to \pi\pi$ is an example of a case where penguin pollution cannot be a-priori ignored. We also list various ways in which the problem might be overcome.

6.3 $B \to \psi K_S$

The first evidence for CP violation outside K decays has been provided by the recent BaBar and Belle measurements of the CP asymmetry in $B \to \psi K_S$,

$$a_{\psi K_S} = \begin{cases} 0.59 \pm 0.15 & \text{Babar [11]} \\ 0.99 \pm 0.15 & \text{Belle [12]} \end{cases} \tag{149}$$

These results in combination with previous ones [8, 9, 10] give the world average quoted in Equation (3). The process $B \to \psi K_S$ is one where the penguin contribution is harmless and the CP asymmetry is subject to an impressively clean theoretical interpretation.

The decay is mediated by the quark transition $\bar{b} \to \bar{c}c\bar{s}$. It gets contributions from a tree level diagram and from penguin diagrams with intermediate u, c and t quarks. Using the unitarity relation (54), we can write the various contributions in terms of two CKM combinations:

$$A(\bar{b} \to \bar{c}c\bar{s}) = (T_{c\bar{c}s} + P_s^c - P_s^t)V_{cb}^* V_{cs} + (P_s^u - P_s^t)V_{ub}^* V_{us}. \tag{150}$$

The second term is suppressed by two factors. First, there is the ratio between penguin and tree contributions,

$$r_{PT}^{\psi K} \equiv \frac{P_{\psi K}}{T_{\psi K}} \equiv \frac{P_s^u - P_s^t}{T_{c\bar{c}s} + P_s^c - P_s^t} \approx \left[\frac{\alpha_s}{12\pi} \ln \frac{m_t^2}{m_b^2}\right] \frac{\langle \psi K_S | \bar{b} \gamma^\mu T^a s \bar{c} \gamma_\mu T^a c | B^0 \rangle}{\langle \pi^+ \pi^- | \bar{b}_L \gamma^\mu c_L \bar{c}_L \gamma_\mu s_L | B^0 \rangle}. \tag{151}$$

The term in the square brackets is $\mathcal{O}(0.03)$ but the ratio of matrix elements may partially compensate for this suppression. Secondly, there is the ratio of CKM elements, $|(V_{ub}^* V_{us})/(V_{cb}^* V_{cs})| \sim \lambda^2$. We conclude that the second term is suppressed by a factor of $r_{PT}^{\psi K}\lambda^2 \lesssim 10^{-2}$ and we can safely neglect $P_{\psi K_S}$. Thus the $B \to \psi K$ decay is dominated by a single weak phase, that is, $\arg(V_{cb}^* V_{cs})$.

Neglecting $P_{\psi K_S}$ means that, to a very good approximation, we have $|\lambda_{\psi K_S}| = 1$,

$$a_{\psi K_S} = \mathrm{Im}\lambda_{\psi K_S}, \tag{152}$$

and that the experimental value of $a_{\psi K_S}$ [Equation (3)] can be cleanly interpreted in terms of a CP violating phase.

A new ingredient in the analysis is the effect of $K - \bar{K}$ mixing. For decays with a single K_S in the final state, $K - \bar{K}$ mixing is essential because $B^0 \to K^0$ and $\bar{B}^0 \to \bar{K}^0$, and interference is possible only due to $K - \bar{K}$ mixing. This adds a factor of

$$\left(\frac{p}{q}\right)_K = \frac{V_{cs}V_{cd}^*}{V_{cs}^* V_{cd}}\omega_K^* \tag{153}$$

into $(\bar{A}/A)$:

$$\frac{\bar{A}_{\psi K_S}}{A_{\psi K_S}} = \eta_{\psi K_S}\left(\frac{V_{cb}V_{cs}^*}{V_{cb}^* V_{cs}}\right)\left(\frac{V_{cs}V_{cd}^*}{V_{cs}^* V_{cd}}\right)\omega_B^*. \tag{154}$$

The CP-eigenvalue of the state is $\eta_{\psi K_S} = -1$. Combining Equations (148) and (154), we find

$$\lambda(B \to \psi K_S) = -\left(\frac{V_{tb}^* V_{td}}{V_{tb}V_{td}^*}\right)\left(\frac{V_{cb}V_{cs}^*}{V_{cb}^* V_{cs}}\right)\left(\frac{V_{cd}^* V_{cs}}{V_{cd}V_{cs}^*}\right), \tag{155}$$

which leads to

$$a_{\psi K_S} = \sin 2\beta. \tag{156}$$

What we have learned above is that Equation (156) is clean of hadronic uncertainties to $\mathcal{O}(r_{PT}^{\psi K}\lambda^2) \lesssim 10^{-2}$. This means that the measurement of $a_{\psi K_S}$ can give the theoretically cleanest determination of a CKM parameter, even cleaner than the determination of $|V_{us}|$ from $K \to \pi\ell\nu$. [If BR$(K_L \to \pi\nu\bar{\nu})$ is measured, it will give a comparably clean determination of η.]

Taking into account all the constraints on the CKM parameters *except* for the $a_{\psi K_S}$ measurements, the SM prediction is [47]

$$\sin 2\beta = 0.68 \pm 0.18, \tag{157}$$

consistent with the experimental result (3). This consistency has important implications. In particular,

- the Kobayashi-Maskawa mechanism has successfully passed its first precision test;

- models of approximate CP which, by definition, predict $|a_{\psi K_S}| \ll 1$, are excluded.

6.4 $B \to \pi\pi$

The CP asymmetry in the $B \to \pi^+\pi^-$ mode has the form

$$a_{\pi\pi}(t) = -\frac{1 - |\lambda_{\pi\pi}|^2}{1 + |\lambda_{\pi\pi}|^2} \cos \Delta mt + \frac{2\mathrm{Im}\lambda_{\pi\pi}}{1 + |\lambda_{\pi\pi}|^2} \sin \Delta mt. \tag{158}$$

Recently, the BaBar collaboration presented the first constraints on this asymmetry [99]:

$$\frac{2\mathrm{Im}\lambda_{\pi\pi}}{1 + |\lambda_{\pi\pi}|^2} = 0.03^{+0.54}_{-0.57},$$

$$\frac{1 - |\lambda_{\pi\pi}|^2}{1 + |\lambda_{\pi\pi}|^2} = -0.25^{+0.47}_{-0.49}. \tag{159}$$

The results are not yet precise enough to give useful constraints. But we discuss this mode to show how penguin pollution arises and how it complicates the analysis.

The decay is mediated by the quark transition $\bar{b} \to \bar{u}ud$. It gets contributions from a tree level diagram and from penguin diagrams with intermediate u, c and t quarks. Using the unitarity relation (55), we can write the various contributions in terms of two CKM combinations:

$$A(\bar{b} \to \bar{u}ud) = (T_{u\bar{u}d} + P_d^u - P_d^c)V_{ub}^*V_{ud} + (P_d^t - P_d^c)V_{tb}^*V_{td}. \tag{160}$$

The ratio between magnitudes of the second and first terms is given by $r_{PT}^{\pi\pi} \left| \dfrac{V_{tb}^* V_{td}}{V_{ub}^* V_{ud}} \right|$. Since both $|V_{ub}V_{ud}^*|$ and $|V_{tb}^*V_{td}|$ are of $\mathcal{O}(\lambda^3)$, the second term is suppressed only by the factor $r_{PT}^{\pi\pi}$, where

$$r_{PT}^{\pi\pi} \equiv \frac{P_{\pi\pi}}{T_{\pi\pi}} \equiv \frac{P_d^t - P_d^c}{T_{u\bar{u}d} + P_d^u - P_d^c}. \tag{161}$$

One may make a rough estimate of $|P_{\pi\pi}/T_{\pi\pi}|$ from the decay $B \to K\pi$, which can be parameterised as follows:

$$A(B^0 \to K^+\pi^-) = T_{K\pi}V_{ub}^*V_{us} + P_{K\pi}V_{tb}V_{ts}^*. \tag{162}$$

In this case $|P_{K\pi}/T_{K\pi}| = \mathcal{O}(r_{PT}^{K\pi}/\lambda^2)$. If QCD enhances the penguin contribution to $B \to \pi\pi$ by a significant amount, that is $r_{PT} \gg \lambda^2$, then $B \to K\pi$ would be dominated by the penguin process. Let us provisionally make the following assumptions: (i) flavor SU(3)

symmetry in the QCD matrix elements; (ii) electroweak penguins and "color suppressed" processes are negligible; (iii) penguins dominate $B \to K\pi$, so $T_{K\pi}$ may be ignored in BR($B^0 \to K^+\pi^-$); (iv) penguins make a small enough contribution to $B \to \pi\pi$ that $P_{\pi\pi}$ may be ignored in BR($B^0 \to \pi^+\pi^-$). Then

$$\left|\frac{P_{\pi\pi}}{T_{\pi\pi}}\right| = \left|\frac{P_{\pi\pi}}{P_{K\pi}}\right|\left|\frac{P_{K\pi}}{T_{\pi\pi}}\right| = \left|\frac{V_{ub}V_{ud}}{V_{ts}V_{tb}}\right|\left(\frac{\text{BR}(B^0 \to K^+\pi^-)}{\text{BR}(B^0 \to \pi^+\pi^-)}\right)^{1/2}. \tag{163}$$

Recent measurements [100, 101, 102] give world averages BR($B^0 \to \pi^+\pi^-$) = $(4.4 \pm 0.9) \times 10^{-6}$ and BR($B^0 \to K^+\pi^-$) = $(17.3 \pm 1.5) \times 10^{-6}$. We thus find BR($B^0 \to K^+\pi^-$)/BR($B^0 \to \pi^+\pi^-$) ≈ 3.9 and obtain the rough estimate

$$|r_{PT}^{\pi\pi}| \sim 0.2 - 0.3. \tag{164}$$

It is clear that penguin effects are unlikely to be negligible in $B \to \pi\pi$.

Combining Equations (148) and (160), we find

$$\lambda(B \to \pi\pi) = \left(\frac{V_{tb}^*V_{td}}{V_{tb}V_{td}^*}\right)\left(\frac{V_{ub}V_{ud}^*}{V_{ub}^*V_{ud}}\right)\left[\frac{1 + r_{PT}^{\pi\pi}(V_{tb}V_{td}^*)/(V_{ub}V_{ud}^*)}{1 + r_{PT}^{\pi\pi}(V_{tb}^*V_{td})/(V_{ub}^*V_{ud})}\right]. \tag{165}$$

If the last factor could be approximated by unity, that is $r_{PT}^{\pi\pi} = 0$, we would obtain $|\lambda_{\pi\pi}| = 1$ and

$$a_{\pi\pi} = \sin 2\alpha. \tag{166}$$

This approximation is however unjustified. To get an idea of the effects of $P_{\pi\pi} \neq 0$, we give the leading corrections due to a small $|r_{PT}|$:

$$\begin{aligned}
|\lambda_{\pi\pi}| &= 1 - 2(R_t/R_u)\text{Im}(r_{PT}^{\pi\pi})\sin\alpha, \\
\frac{\text{Im}\lambda_{\pi\pi}}{|\lambda_{\pi\pi}|} &= \sin 2\alpha + 2(R_t/R_u)\text{Re}(r_{PT}^{\pi\pi})\cos 2\alpha \sin\alpha.
\end{aligned} \tag{167}$$

(For a more detailed discussion, see [98].) Note that if strong phases can be neglected, r_{PT} is real and $|\lambda_{\pi\pi}| = 1$ would be a good approximation. But it is not clear whether the strong phases are indeed small. In any case, one needs to know $r_{PT}^{\pi\pi}$ to extract α from $a_{\pi\pi}(t)$. This is the problem of penguin pollution.

A variety of solutions to this problem have been proposed, falling roughly into two classes. The first type of approach is to convert the estimate given above into an actual measurement of $|P_{K\pi}|$. (The list of papers on this subject is long. Early works include [103, 104, 105]. For a much more comprehensive list of references, see [98].) Once $|P_{K\pi}|$ is known, flavor $SU(3)$ is used to relate $|P_{K\pi}|$ to $|P_{\pi\pi}|$. One must then include a number of additional effects:

- Electroweak penguins. The effects are calculable [106].

- Color suppressed and re-scattering processes. These must be bounded or estimated using data and some further assumptions.

- $SU(3)$ corrections. Some, such as f_K/f_π, can be included, but $SU(3)$ corrections generally remain a source of irreducible uncertainty.

The second type of approach is to exploit the fact that the penguin contribution to $P_{\pi\pi}$ is pure $\Delta I = 1/2$, while the tree contribution to $T_{\pi\pi}$ contains a piece which is $\Delta I = 3/2$. (This is not true of the electroweak penguins [107], but these are expected to be small.) Isospin symmetry allows one to form a relation among the amplitudes $B^0 \to \pi^+\pi^-$, $B^0 \to \pi^0\pi^0$, and $B^+ \to \pi^+\pi^0$,

$$\frac{1}{\sqrt{2}} A(B^0 \to \pi^+\pi^-) + A(B^0 \to \pi^0\pi^0) = A(B^+ \to \pi^+\pi^0). \tag{168}$$

There is also a relation for the charge conjugate processes. A simple geometric construction then allows one to disentangle the unpolluted $\Delta I = \frac{3}{2}$ amplitudes, from which $\sin 2\alpha$ may be extracted cleanly [108].

The key experimental difficulty is that one must measure accurately the flavor-tagged rate for $B^0 \to \pi^0\pi^0$. Since the final state consists of only four photons, and the branching fraction is expected to be of $\mathcal{O}(10^{-6})$, this is very hard. It has been noted that an upper bound on this rate, if sufficiently strong, would also allow one to bound $P_{\pi\pi}$ usefully [109, 98, 110].

An alternative is to perform an isospin analysis of the process $B^0 \to \rho\pi \to \pi^+\pi^-\pi^0$ [111, 112, 113, 114]. Here one must study the time-dependent asymmetry over the entire Dalitz plot, probing variously the intermediate states $\rho^\pm\pi^\mp$ and $\rho^0\pi^0$. The advantage here is that the final states with two π^0's need not be considered. On the other hand, thousands of cleanly reconstructed events would be needed.

Finally, one can attempt to calculate the penguin matrix elements. Model-dependent analyses are not really adequate for this purpose, since the goal is the extraction of fundamental parameters. Precise calculations of such matrix elements from lattice QCD are far in the future, given the large energies of the π's and the need for an unquenched treatment. Recently, a new QCD-based analysis of the $B \to \pi\pi$ matrix elements has been proposed [115, 116, 117, 118]. For details, see [119].

7 CP violation in supersymmetry

7.1 CP violation as a probe of new physics

We have argued that the Standard Model picture of CP violation is rather unique and highly predictive. We have also stated that reasonable extensions of the Standard Model have a very different picture of CP violation. Experimental results are now starting to decide between the various possibilities. Our discussion of CP violation in the presence of new physics aims to demonstrate that, indeed, models of new physics can significantly modify the Standard Model predictions and that measurements in the near future will therefore have a strong impact on the theoretical understanding of CP violation.

To understand how the Standard Model predictions could be modified by new physics, we focus on CP violation in the interference between decays with and without mixing. As explained above, this type of CP violation may give, due to its theoretical cleanliness, unambiguous evidence for new physics most easily. We now list some of the questions can be answered when many such observables are measured.

- *Is there new physics in $B - \bar{B}$ mixing?*

  Consider $a_{\psi K_S}$, the CP asymmetry in $B \to \psi K_S$. This measurement will cleanly determine the relative phase between the $B - \bar{B}$ mixing amplitude and the $b \to c\bar{c}s$ decay amplitude ($\sin 2\beta$ in the SM). The $b \to c\bar{c}s$ decay has Standard Model tree contributions and therefore is very unlikely to be significantly affected by new physics. On the other hand, the mixing amplitude can be easily modified by new physics. We parameterise such a modification by a phase θ_d:

  $$2\theta_d = \arg(M_{12}/M_{12}^{\mathrm{SM}}) \tag{169}$$

  so that the the SM prediction for $a_{\psi K_S}$ becomes

  $$a_{\psi K_S} = \sin[2(\beta + \theta_d)]. \tag{170}$$

  $\theta_d \neq 0$ would be clear evidence of new physics (see *e.g.* [120]).

  It is interesting to note that already now the measured value of $a_{\psi K_S}$ (3), which is consistent with the SM range, excludes many models that require a modification of CP violation in $B - \bar{B}$ mixing due to new physics. Among these are various models of soft CP violation [121, 122] aimed at solving the strong CP problem, models of geometric CP violation due to extra dimensions [123], models of spontaneous CP violation in the left-right symmetric framework [124, 125], and several models that aim to solve the supersymmetric CP problems [126, 127, 128].

- *Is the new physics related to $\Delta B = 1$ or $\Delta B = 2$ processes, or both?*

  Consider $a_{\phi K_S}$, the CP asymmetry in $B \to \phi K_S$. This measurement will cleanly determine the relative phase between the $B - \bar{B}$ mixing amplitude and the $b \to s\bar{s}s$ decay amplitude ($\sin 2\beta$ in the SM). The $b \to s\bar{s}s$ decay has only Standard Model penguin contributions and therefore is sensitive to new physics. We parameterise the modification of the decay amplitude by a phase θ_A [129]:

  $$\theta_A = \arg(\bar{A}_{\phi K_S}/\bar{A}_{\phi K_S}^{\mathrm{SM}}). \tag{171}$$

  Then

  $$a_{\phi K_S} = \sin[2(\beta + \theta_d + \theta_A)]. \tag{172}$$

  Comparing $a_{\phi K_S}$ to $a_{\psi K_S}$, that is, examining whether $\theta_A \neq 0$, will tell us if the new physics is related to $\Delta B = 1$ or $\Delta B = 2$ processes.

- *Is the new physics related to the third generation, or all generations?*

  Consider $a_{\pi\nu\bar{\nu}}$, the CP violating ratio of $K \to \pi\nu\bar{\nu}$ decays, defined in (120). This measurement will cleanly determine the relative phase between the $K - \bar{K}$ mixing amplitude and the $s \to d\nu\bar{\nu}$ decay amplitude (of order $\sin^2 \beta$ in the SM). The experimentally measured small value of ε_K requires that the phase of the $K - \bar{K}$ mixing amplitude is not modified from the Standard Model prediction. (More precisely, it requires that the phase of the mixing amplitude is very close to twice the phase of the $s \to d\bar{u}u$ decay amplitude [130].) On the other hand, the decay, which in the SM is a loop process with small mixing angles, can be easily modified by new physics. Examining whether the SM correlation between $a_{\pi\nu\bar{\nu}}$ and $a_{\psi K_S}$ is fulfilled will give us information on the generations involved.

- *Is the new physics related to the down sector or the up sector or both?*

  Consider $a_{D \to K\pi}$, the CP violating quantity in $D \to K^{\pm}\pi^{\mp}$ decays defined in (134). It depends on ϕ_D, the relative phase between the $D - \bar{D}$ mixing amplitude and the $c \to d\bar{s}u$ and $c \to s\bar{d}u$ decay amplitudes. Within the Standard Model, the two decay channels are tree level. It is unlikely that they are affected by new physics. On the other hand, the mixing amplitude can be easily modified by new physics. Examining whether $a_{D \to K\pi} = 0$, that is, whether ϕ_D (and/or $\theta_d) \neq 0$, will provide information on the sectors involved.

- *Are the new sources of CP violation flavor changing, flavor diagonal or both?*

  Consider d_N, the electric dipole moment of the neutron. We did not discuss this quantity so far because, unlike CP violation in meson decays, flavor changing couplings are not necessary for d_N. In other words, the CP violation that induces d_N is *flavor diagonal*. It does in general get contributions from flavor changing physics, but it could be induced by sectors that are flavor blind. Within the SM (and ignoring θ_{QCD}), the contribution from δ_{KM} arises at the three loop level and is at least six orders of magnitude below the experimental bound (7). If the bound is further improved (or a signal observed), we may elucidate the flavor dependence.

It is no wonder then that with such rich information, flavor and CP violation provide an excellent probe of new physics. We will now demonstrate this situation more concretely by discussing CP violation in supersymmetry.

7.2 The supersymmetric framework

Supersymmetry solves the fine-tuning problem of the Standard Model and has many other virtues. But at the same time, it leads to new problems: baryon number violation, lepton number violation, large flavor changing neutral current processes and large CP violation. The first two problems can be solved by imposing R-parity on supersymmetric models. There is no such simple, symmetry-related solution to the problems of flavor and CP violation. Instead, suppression of the relevant couplings can be achieved by demanding very constrained structures of the soft supersymmetry breaking terms. There are two important questions here. First, can theories of dynamical supersymmetry breaking naturally induce such structures? (For an excellent review of dynamical supersymmetry breaking, see [131].) Second, can measurements of flavor changing and/or CP violating processes shed light on the structure of the soft supersymmetry breaking terms? Since the answer to both questions is in the affirmative, we conclude that flavor changing neutral current processes and, in particular, CP violating observables will provide clues to the crucial question of how supersymmetry breaks.

7.2.1 CP violating parameters

A generic supersymmetric extension of the Standard Model contains a host of new flavor and CP violating parameters. (For a review of CP violation in supersymmetry see [132, 133].) It is an amusing exercise to count the number of parameters [134]. The supersymmetric part of the Lagrangian depends, in addition to the three gauge couplings

of $G_{\rm SM}$, on the parameters of the superpotential W:

$$W = \sum_{i,j} \left(Y^u_{ij} H_u Q_{Li} \overline{U}_{Lj} + Y^d_{ij} H_d Q_{Li} \overline{D}_{Lj} + Y^\ell_{ij} H_d L_{Li} \overline{E}_{Lj} \right) + \mu H_u H_d. \qquad (173)$$

In addition, we have to add soft supersymmetry breaking terms:

$$\mathcal{L}_{\rm soft} = - \left(A^u_{ij} H_u \tilde{Q}_{Li} \tilde{\overline{U}}_{Lj} + A^d_{ij} H_d \tilde{Q}_{Li} \tilde{\overline{D}}_{Lj} + A^\ell_{ij} H_d \tilde{L}_{Li} \tilde{\overline{E}}_{Lj} + B H_u H_d + {\rm h.c.} \right)$$

$$- \sum_{\rm all\ scalars} (m_S^2)_{ij} A_i \bar{A}_j - \frac{1}{2} \sum_{(a)=1}^{3} \left(\tilde{m}_{(a)} (\lambda\lambda)_{(a)} + {\rm h.c.} \right). \qquad (174)$$

where $S = Q_L, \overline{D}_L, \overline{U}_L, L_L, \overline{E}_L$. The three Yukawa matrices Y^f depend on 27 real and 27 imaginary parameters. Similarly, the three A^f-matrices depend on 27 real and 27 imaginary parameters. The five m_S^2 hermitian 3×3 mass-squared matrices for sfermions have 30 real parameters and 15 phases. The gauge and Higgs sectors depend on

$$\theta_{\rm QCD}, \tilde{m}_{(1)}, \tilde{m}_{(2)}, \tilde{m}_{(3)}, g_1, g_2, g_3, \mu, B, m_{h_u}^2, m_{h_d}^2, \qquad (175)$$

that is 11 real and 5 imaginary parameters. Summing over all sectors, we get 95 real and 74 imaginary parameters. The various couplings (other than the gauge couplings) can be thought of as spurions that break a global symmetry,

$$U(3)^5 \times U(1)_{\rm PQ} \times U(1)_R \to U(1)_B \times U(1)_L. \qquad (176)$$

The $U(1)_{\rm PQ} \times U(1)_R$ charge assignments are:

$$\begin{array}{c c c c c c} & H_u & H_d & Q\overline{U} & Q\overline{D} & L\overline{E} \\ U(1)_{\rm PQ} & 1 & 1 & -1 & -1 & -1 \\ U(1)_{\rm R} & 1 & 1 & 1 & 1 & 1 \end{array}. \qquad (177)$$

Consequently, we can remove 15 real and 30 imaginary parameters, which leaves

$$124 = \begin{cases} 80 & \text{real physical parameters.} \\ 44 & \text{imaginary physical parameters.} \end{cases} \qquad (178)$$

In particular, there are 43 new CP violating phases! In addition to the single Kobayashi-Maskawa of the SM, we can put 3 phases in M_1, M_2, μ (we used the $U(1)_{\rm PQ}$ and $U(1)_R$ to remove the phases from μB^* and M_3, respectively) and the other 40 phases appear in the mixing matrices of the fermion-sfermion-gaugino couplings. (Of the 80 real parameters, there are 11 absolute values of the parameters in (175), 9 fermion masses, 21 sfermion masses, 3 CKM angles and 36 SCKM angles.) Supersymmetry provides a nice example to our statement that reasonable extensions of the Standard Model may have more than one source of CP violation.

The requirement of consistency with experimental data provides strong constraints on many of these parameters. For this reason, the physics of flavor and CP violation has had a profound impact on supersymmetric model building. A discussion of CP violation in this context can hardly avoid addressing the flavor problem itself. Indeed, many of the supersymmetric models that we analyze below were originally aimed at solving flavor problems. For details on the supersymmetric flavor problem, see [135].

For CP violation, one can distinguish two classes of experimental constraints. First, bounds on nuclear and atomic electric dipole moments determine what is usually called the *supersymmetric CP problem*. Second, the physics of neutral mesons and, most importantly, the small experimental value of ε_K pose the *supersymmetric ε_K problem*. In the next two subsections we describe the two problems.

7.2.2   The supersymmetric CP problem

One aspect of supersymmetric CP violation involves effects that are flavor preserving. Then, for simplicity, we describe this aspect in a supersymmetric model without additional flavor mixings, *i.e.* the minimal supersymmetric standard model (MSSM) with universal sfermion masses and with the trilinear SUSY-breaking scalar couplings proportional to the corresponding Yukawa couplings. (The generalisation to the case of non-universal soft terms is straightforward.) In such a constrained framework, there are four new phases beyond the two phases of the SM (δ_{KM} and θ_{QCD}). One arises in the bilinear μ-term of the superpotential (173), while the other three arise in the soft supersymmetry breaking parameters of (174): $\tilde{m}$ (the gaugino mass), A (the trilinear scalar coupling) and B (the bilinear scalar coupling). Only two combinations of the four phases are physical [136, 137]:

$$\phi_A = \arg(A^*\tilde{m}), \quad \phi_B = \arg(\tilde{m}\mu B^*). \tag{179}$$

In the more general case of non-universal soft terms there is one independent phase ϕ_A, for each quark and lepton flavor. Moreover, complex off-diagonal entries in the sfermion mass-squared matrices represent additional sources of CP violation.

The most significant effect of ϕ_A and ϕ_B is their contribution to electric dipole moments (EDMs). For example, the contribution from one-loop gluino diagrams to the down quark EDM is given by [138, 139]:

$$d_d = m_d \frac{e\alpha_3}{18\pi\tilde{m}^3} \left(|A|\sin\phi_A + \tan\beta|\mu|\sin\phi_B\right), \tag{180}$$

where we have taken $m_Q^2 \sim m_D^2 \sim m_{\tilde{g}}^2 \sim \tilde{m}^2$, for left- and right-handed squark and gluino masses. We define, as usual, $\tan\beta = \langle H_u \rangle / \langle H_d \rangle$. Similar one-loop diagrams give rise to chromoelectric dipole moments. The electric and chromoelectric dipole moments of the light quarks (u, d, s) are the main source of d_N (the EDM of the neutron), giving [140]

$$d_N \sim 2 \left(\frac{100\,GeV}{\tilde{m}}\right)^2 \sin\phi_{A,B} \times 10^{-23}\ \mathrm{e\,cm}, \tag{181}$$

where, as above, $\tilde{m}$ represents the overall SUSY scale. In a generic supersymmetric framework, we expect $\tilde{m} = \mathcal{O}(m_Z)$ and $\sin\phi_{A,B} = \mathcal{O}(1)$. Then the constraint (7) is generically violated by about two orders of magnitude. This is *the Supersymmetric CP Problem*.

Equation (181) shows two possible ways to solve the supersymmetric CP problem:

- Heavy squarks: $\tilde{m} \gtrsim 1\ TeV$.

- Approximate CP: $\sin\phi_{A,B} \ll 1$.

Recently, a third way has been investigated, that is cancellations between various contributions to the electric dipole moments. However, there seems to be no symmetry that can guarantee such a cancellation. This is in contrast to the other two mechanisms mentioned above that were shown to arise naturally in specific models. We therefore do not discuss any further this third mechanism.

The electric dipole moment of the electron is also a sensitive probe of flavor diagonal CP phases. The present experimental bound,

$$|d_e| \leq 4 \times 10^{-27} \, e \, cm[141],$$ (182)

is also violated by about two orders of magnitude for 'natural' values of supersymmetric parameters. A new experiment [142] has been proposed to search for the electric dipole moment of the muon at a level smaller by five orders of magnitude than present bounds; such improvement will make d_μ another sensitive probe of supersymmetry [143].

7.2.3   The supersymmetric ε_K problem

The supersymmetric contribution to the ε_K parameter is dominated by diagrams involving Q and $\bar{d}$ squarks in the same loop. For $\tilde{m} = m_{\tilde{g}} \simeq m_Q \simeq m_D$ (our results depend only weakly on this assumption) and focusing on the contribution from the first two squark families, one gets (see, for example, [144]):

$$\varepsilon_K = \frac{5\,\alpha_3^2}{162\sqrt{2}} \frac{f_K^2 m_K}{\tilde{m}^2 \Delta m_K} \left[\left(\frac{m_K}{m_s + m_d} \right)^2 + \frac{3}{25} \right] \mathrm{Im}\left[(\delta_{12}^d)_{LL} (\delta_{12}^d)_{RR} \right].$$ (183)

Here

$$
(\delta_{12}^d)_{LL} = \left(\frac{m_{\tilde{Q}_2}^2 - m_{\tilde{Q}_1}^2}{m_{\tilde{Q}}^2} \right) \left| K_{12}^{dL} \right|,
$$

$$
(\delta_{12}^d)_{RR} = \left(\frac{m_{\tilde{D}_2}^2 - m_{\tilde{D}_1}^2}{m_{\tilde{D}}^2} \right) \left| K_{12}^{dR} \right|,
$$ (184)

where K_{12}^{dL} (K_{12}^{dR}) are the mixing angles in the gluino couplings to left-handed (right-handed) down quarks and their scalar partners. Note that CP would be violated even if there were two families only [145]. Using the experimental value of ε_K, we get

$$\frac{(\Delta m_K \varepsilon_K)^{\mathrm{SUSY}}}{(\Delta m_K \varepsilon_K)^{\mathrm{EXP}}} \sim 10^7 \left(\frac{300\,GeV}{\tilde{m}} \right)^2 \left(\frac{m_{\tilde{Q}_2}^2 - m_{\tilde{Q}_1}^2}{m_{\tilde{Q}}^2} \right) \left(\frac{m_{\tilde{D}_2}^2 - m_{\tilde{D}_1}^2}{m_{\tilde{D}}^2} \right) |K_{12}^{dL} K_{12}^{dR}| \sin\phi,$$ (185)

where ϕ is the CP violating phase. In a generic supersymmetric framework, we expect $\tilde{m} = \mathcal{O}(m_Z)$, $\delta m_{Q,D}^2 / m_{Q,D}^2 = \mathcal{O}(1)$, $K_{ij}^{Q,D} = \mathcal{O}(1)$ and $\sin\phi = \mathcal{O}(1)$. Then the constraint (185) is generically violated by about seven orders of magnitude.

The Δm_K constraint on $\mathrm{Re}\left[(\delta_{12}^d)_{LL} (\delta_{12}^d)_{RR} \right]$ is about two orders of magnitude weaker. One can distinguish then three interesting regions for $(\delta_{12}^d) = \sqrt{(\delta_{12}^d)_{LL} (\delta_{12}^d)_{RR}}$:

$$
\begin{array}{lll}
0.003 & \ll (\delta_{12}^d) & \text{excluded,} \\
0.0002 & \ll (\delta_{12}^d) \lesssim 0.003 & \text{viable with small phases,} \\
& (\delta_{12}^d) \ll 0.0002 & \text{viable with } \mathcal{O}(1) \text{ phases.}
\end{array}
$$ (186)

The first bound comes from the Δm_K constraint (assuming that the relevant phase is not particularly close to $\pi/2$). The bounds here apply to squark masses of order 500 GeV and scale like $\tilde{m}$. There is also dependence on $m_{\tilde{g}}/\tilde{m}$, which is here taken to be one.

Equation (185) shows other possible ways of solving the supersymmetric ε_K problem:

- Heavy squarks: $\tilde{m} \gg 300 \, GeV$;

- Universality: $\delta m_{Q,D}^2 \ll m_{Q,D}^2$;

- Alignment: $|K_{12}^d| \ll 1$;

- Approximate CP: $\sin \phi \ll 1$.

7.2.4 A supersymmetric ε'/ε?

In this section we discuss the question of whether supersymmetric contributions to ε'/ε can be dominant. A typical supersymmetric contribution to ε'/ε is given by [146]

$$\left| \frac{\varepsilon'}{\varepsilon} \right| = 58 B_G \left[\frac{\alpha_s(m_{\tilde{g}})}{\alpha_s(500 \, \text{GeV})} \right]^{\frac{23}{21}} \left(\frac{158 \, \text{MeV}}{m_s + m_d} \right) \left(\frac{500 \, \text{GeV}}{m_{\tilde{g}}} \right) \left| \text{Im} \left[(\delta_{LR}^d)_{12} - (\delta_{LR}^d)_{21}^* \right] \right| , \quad (187)$$

where B_G parameterises the matrix element of the relevant four-quark operator. Consequently, the supersymmetric contribution saturates ε'/ε for

$$\text{Im} \left[(\delta_{LR}^d)_{12} - (\delta_{LR}^d)_{21}^* \right] \sim \lambda^7 \left(\frac{m_{\tilde{g}}}{500 \, \text{GeV}} \right) \quad (188)$$

where, motivated by flavor symmetries (see below), we parameterise the suppression by powers of $\lambda \sim 0.2$. Without proportionality, a naive guess would give

$$
\begin{aligned}
(\delta_{LR}^d)_{12} &\sim \frac{m_s |V_{us}|}{\tilde{m}} \sim \lambda^{5-6} \frac{m_t}{\tilde{m}}, \\
(\delta_{LR}^d)_{21} &\sim \frac{m_d}{|V_{us}| \tilde{m}} \sim \lambda^{5-6} \frac{m_t}{\tilde{m}} .
\end{aligned}
\quad (189)
$$

This is not far from the value required to account for ε'/ε [147]. Thus, it is certainly *possible* that supersymmetry accounts for, at least, a large part of ε'/ε (see, for example, the models of references [148]–[154]). Yet, it has been argued that such a situation is not *generic* [155]. The problem is that Equation (189) gives an overestimate of the supersymmetric contribution in most viable models of supersymmetry breaking that have appeared in the literature. We will encounter concrete examples of this statement when we survey the various supersymmetric flavor models.

7.3 Supersymmetry breaking and flavor models

Before turning to a detailed discussion, we define two scales that play an important role in supersymmetry: Λ_S, where the soft supersymmetry breaking terms are generated, and Λ_F, where flavor dynamics takes place. When $\Lambda_F \gg \Lambda_S$, it is possible that there are no genuinely new sources of flavor and CP violation. This leads to models with exact universality. When $\Lambda_F \lesssim \Lambda_S$, we do not expect, in general, that flavor and CP violation are limited to the Yukawa matrices. One way to suppress CP violation would be to assume that, similarly to the Standard Model, CP violating phases are large, but their effects are screened, possibly by the same physics that explains the various flavor puzzles, such as models with Abelian or non-Abelian horizontal symmetries. It is also possible that CP violating effects are suppressed because squarks are heavy. Another option, which is now excluded, was to assume that CP is an approximate symmetry of the full theory (namely, CP violating phases are all small).

7.3.1 Gauge mediated supersymmetry breaking

If at some high energy scale squarks are exactly degenerate and the A terms proportional to the Yukawa couplings, then the contribution to ε_K comes from Renormalisation Group Evolution (RGE) and is GIM suppressed, that is

$$\varepsilon_K \propto \mathrm{Im}\left[(V_{td}V_{ts}^*)^2\right] Y_t^4 \left[\frac{\log(\Lambda_S/m_W)}{16\pi^2}\right]^2 . \tag{190}$$

This contribution is negligibly small [136]. The contribution from genuinely supersymmetric phases (*i.e.* the phases in A_t and μ) is also negligible [156, 157]. (This does not necessarily mean that there is no supersymmetric effect on ε_K. In some small corner of parameter space the supersymmetric contribution from stop-chargino diagrams can give up to 20% of ε_K [158, 159].)

In Gauge Mediated Supersymmetry Breaking (GMSB) [160, 161], superpartner masses are generated by the SM gauge interactions. These masses are then exactly universal at the scale Λ_S at which they are generated (up to tiny high order effects associated with Yukawa couplings). Furthermore, A terms are suppressed by loop factors. The only contribution to ε_K is then from the running, and since Λ_S is low it is highly suppressed.

These models can also readily satisfy the EDM constraints. In most models, the A terms and gaugino masses arise from the same supersymmetry breaking auxiliary field, that is, they are generated by the same SUSY and $U(1)_R$ breaking source. They therefore carry the same phase (up to corrections from the Standard Model Yukawa couplings), and ϕ_A vanishes to a very good approximation:

$$\phi_A \propto Y_t^4 Y_c^2 Y_b^2 J_{\mathrm{CKM}} \left[\frac{\log(\Lambda_S/m_W)}{16\pi^2}\right]^4 . \tag{191}$$

The resulting EDM is $d_N \lesssim 10^{-31}$ e cm. This maximum can be reached only for very large $\tan\beta \sim 60$ while, for small $\tan\beta \sim 1$, d_N is about 5 orders of magnitude smaller. This range of values for d_N is much below the present ($\sim 10^{-25}$ e cm) and foreseen ($\sim 10^{-28}$ e cm) experimental sensitivities (see *e.g.* [162]).

The value of ϕ_B in general depends on the mechanism for generating the μ term. However, running effects can generate an adequate B term at low scales in these models even if $B(\Lambda_S) = 0$. One then finds [163]

$$\frac{B}{\mu} = A_t(\Lambda_S) + M_2(\Lambda_S)\left(-0.12 + 0.17|Y_t|^2\right) , \tag{192}$$

where M_2 is the $SU(2)$ gaugino mass. Since $\phi_A \simeq 0$, the resulting ϕ_B vanishes, again up to corrections involving the Standard Model Yukawa couplings [164].

There is therefore no CP problem in simple models of gauge mediation, even with phases of order one. The supersymmetric contribution to $D-\bar{D}$ mixing is similarly small and we expect no observable effects. For the B_d system, GMSB models predict then a large CP asymmetry in $B \to \psi K_S$, with small deviations (at most 20%) from the SM.

More generally, in any supersymmetric model where there are no new flavor violating sources beyond the Yukawa couplings, CP violation in meson decays is hardly modified from the SM predictions [165].

7.3.2 Gravity, anomaly and gaugino mediation

If different moduli of string theory obtain supersymmetry breaking F terms, they would typically induce flavor-dependent soft terms through their tree-level couplings to Standard Model fields. There are however various scenarios in which the leading contribution to the soft terms is flavor independent. The three most intensively studied frameworks are dilaton dominance, anomaly mediation and gaugino mediation.

Dilaton dominance assumes that the dilaton F term is the dominant one. Then, at tree level, the resulting soft masses are universal and the A terms proportional to the Yukawa couplings. Both universality and proportionality are, however, violated by string loop effects. These induce corrections to squark masses of order $\frac{\alpha_X}{\pi} m_{3/2}^2$, where $\alpha_X = [2\pi(S + S^*)]^{-1}$ is the string coupling. There is no reason why these corrections would be flavor blind. However, RGE effects enhance the universal part of the squark masses by roughly a factor of five, leaving the off-diagonal entries essentially unchanged. The flavor suppression factor is then [166]

$$\langle \delta_{12}^d \rangle \simeq \frac{m_{12}^{2 \text{ one-loop}}}{m_{\tilde{q}}^2} \simeq \frac{\alpha_X}{\pi} \frac{1}{25} \simeq 4 \times 10^{-4} . \tag{193}$$

Dilaton dominance relies on the assumption that loop corrections are small. This probably presents the most serious theoretical difficulty for this idea, because it is hard to see how non-perturbative effects, which are probably required to stabilise the dilaton, could do so in a region of weak coupling. In the strong coupling regime, these corrections could be much larger. However, this idea at least gives some plausible theoretical explanation for how universal masses might emerge in hidden sector models. Given that dilaton stabilisation might require that non-perturbative effects are important, the estimate of flavor suppression (193) might well turn out to be an underestimate.

We now turn to the flavor diagonal phases that enter in various EDMs. The phase ϕ_A vanishes at tree-level, so that [166, 167] $\phi_A = \mathcal{O}(\alpha_X/\pi)$. [The smallness of ϕ_A implies that there is a suppression of $\mathcal{O}(\alpha_X/\pi) \sim 10^{-2}$ compared to (189) and the supersymmetric contribution to ε'/ε is small.] However, ϕ_B is unsuppressed, even when μ, and through it B, are generated by Kahler potential effects through supersymmetry breaking, in which case $B = 2m_{3/2}^* \mu$ [168]. While the size of $m_{3/2}$ is determined from the requirement that the cosmological constant vanishes, its phase remains arbitrary, and in fact depends on the phase of the constant term that is added to the superpotential in order to cancel the cosmological constant.

We conclude that the supersymmetric ε_K problem is solved in these models but the EDM problem, in general, is not. For EDM contributions to be small in these models, the gravitino mass had better give a small physical phase.

Anomaly mediation (AMSB) provides another approach to solving the flavor problems of supersymmetric theories, as well as to obtaining a predictive spectrum. In the presence of some truly 'hidden' supersymmetry breaking sector, with no couplings to the SM fields (apart from indirect couplings through the supergravity multiplet) the conformal anomaly of the Standard Model gives rise to soft supersymmetry breaking terms for the Standard Model fields [169, 170]. These terms are generated purely by gravitational

effects and are given by

$$m_0^2(\mu) = -\frac{1}{4}\frac{\partial\gamma(\mu)}{\ln\mu}m_{3/2}^2, \quad m_{1/2}(\mu) = \frac{\beta(\mu)}{g(\mu)}m_{3/2}, \quad A(\mu) = -\frac{1}{2}\gamma(\mu)m_{3/2} , \qquad (194)$$

where β and γ are the appropriate beta function and anomalous dimension. Thus, apart from the Standard Model gauge and Yukawa couplings, the soft terms only involve the parameter $m_{3/2}$.

In general, naturalness considerations suggest that couplings of hidden and visible sectors should appear in the Kahler potential, leading to soft masses for scalars already at tree level, and certainly by one loop. As a result, one would expect the contributions (194) to be irrelevant. However, in "sequestered sector models" [169], in which the visible sector fields and supersymmetry breaking fields live on different branes separated by some distance, the anomaly mediated contribution (194) could be the dominant effect. This leads to a predictive picture for scalar masses. Since the soft terms (194) are generated by the Standard Model gauge and Yukawa couplings, they are universal, up to corrections involving the third generation Yukawa couplings. However, the resulting slepton masses-squared are negative, so this model requires some modification. We will not attempt a complete review of this subject here. Our principal concerns are the sources of CP violation, and the extent to which the AMSB formulae receive corrections, leading to non-degeneracy of the squark masses.

For Equation (194) to correctly give the leading order soft terms, it is necessary that all moduli obtain large masses before supersymmetry breaking, and that there be no Planck scale VEVs in the supersymmetry breaking sector [171]. A possible scenario for this to happen is if all moduli but the fifth dimensional radius, R, sit at an enhanced symmetry point, and that R obtains a large mass compared to the supersymmetry-breaking scale (say, by a racetrack mechanism). Even in this case, however, there is a difficulty. One might expect that some of the moduli have masses well below the fundamental scale. If there are light moduli in the bulk, there are typically one-loop contributions to scalar masses-squared from exchanges of bulk fields, proportional to $m_{3/2}^2/R^3$ times a loop factor [169]. If these contributions are non-universal, they may easily violate the Δm_K and ε_K constraints [133].

If there are no light moduli, and if the contributions described above are adequately suppressed, some modification of the visible sector is required in order to generate acceptable slepton masses. Different such solutions have been suggested. In some of these models, there are no new contributions to CP violation simply because there are few enough new parameters in the theory that they can all be chosen real by field redefinitions [172, 173, 174]. Furthermore, it is possible to generate the μ term in these models from AMSB, so that ϕ_B vanishes. These models are then similar to GMSB models from the point of view of CP violation.

We conclude then, that in generic sequestered sector models it is difficult to obtain strong degeneracy and a special phase structure is required. It is conceivable that there might be theories with a high degree of degeneracy, or with no new sources of CP violation. In such theories, the SM predictions for CP violation are approximately maintained.

Gaugino mediation [175, 176] is in many ways similar to anomaly mediation, and poses similar issues. These models also suppress dangerous tree level contact terms by invoking extra dimensions, with the Standard Model matter fields localised on one brane

and the supersymmetry breaking sector on another brane. In this case, however, the Standard Model gauge fields are in the bulk, so gauginos get masses at tree level, and as a result scalar masses are generated by running. Scalar masses are therefore universal. Furthermore, the soft terms typically involve only one new parameter, namely, the singlet F VEV that gives rise to gaugino masses. Therefore, they do not induce any new CP violation.

Again, however, if there are non-universal tree and one loop contributions to scalar masses, significant violations of degeneracy and proportionality can be expected, and a special structure of CP violating phases is required.

7.3.3 Supersymmetric flavor models

Various frameworks have been suggested in which flavor symmetries, designed to explain the hierarchy of the Yukawa couplings, impose at the same time a special flavor structure on the soft supersymmetry breaking terms that helps to alleviate the flavor and CP problems.

In the framework of **alignment**, one does not assume any squark degeneracy. Instead, flavor violation is suppressed because the squark mass matrices are approximately diagonal in the quark mass basis. This is the case in models of Abelian flavor symmetries, in which the off-diagonal entries in both the quark mass matrices and in the squark mass matrices are suppressed by some power of a small parameter, λ, that quantifies the breaking of some Abelian flavor symmetry. A natural choice for the value of λ is $\sin\theta_C$, so we will take $\lambda \sim 0.2$. One would naively expect the first two generation squark mixing to be of the order of λ. However, the Δm_K constraint is not satisfied with the 'naive alignment', $K_{12}^d \sim \lambda$, and one has to construct more complicated models to achieve the required suppression [177, 178]. One can solve the supersymmetric ε_K problem by flavor suppression, that is, models with $\langle \delta_{12}^d \rangle \sim \lambda^5$ [179]. These models are highly constrained and almost unique. It is simpler to construct models where $\langle \delta_{12}^d \rangle \sim \lambda^3$ but the CP violating phases are also suppressed [128]. Such models predict that $a_{\psi K_S} \ll 1$ and are therefore now excluded. (Models with $\langle \delta_{12}^d \rangle \sim \lambda^3$ could still be viable with phases of order one if the RGE contributions enhance squark degeneracy.)

For flavor diagonal phases, the question is more model dependent. There is however a way to suppress these phases without assuming approximate CP [179]. The mechanism requires that CP is spontaneously broken by the same fields that break the flavor symmetry ("flavons"). It is based on the observation that a Yukawa coupling and the corresponding A term carry the same horizontal charge and therefore their dependence on the flavon fields is similar. In particular, if a single flavon dominates a certain coupling, the CP phase is the same for the Yukawa coupling and for the corresponding A term and the resulting ϕ_A vanishes. Similarly, if the μ term and the B term depend on one (and the same) flavon, ϕ_B is suppressed.

For ε'/ε, the ε_K constraint requires that the relevant terms are suppressed by at least a factor of λ^2 compared to (189) [155] and the contribution is therefore small.

We conclude that one can construct models in which an Abelian horizontal symmetry solves both the ε_K and the d_N problems. These models are however not the generic ones in this framework.

Non-Abelian horizontal symmetries can induce approximate degeneracy between the first two squark generations, thus relaxing the flavor and CP problems [180]. A review of ε_K in this class of models can be found in [132]. Quite generically, the supersymmetric contributions to ε_K are too large and require small phases (see, for example, the models of reference [181]). There are however specific models where the ε_K problem is solved without the need for small phases [182, 183]. Furthermore, universal contributions from RGE running might further relax the problem.

For flavor diagonal phases, it is difficult (though not entirely impossible) to avoid $\phi_A \gtrsim \lambda^2 \sim 0.04$ [132]. This, however, might be just enough to satisfy the d_N constraint.

With a horizontal U(2) symmetry, the two contributions to ε'/ε in (187) cancel each other. (More generally, this happens for a symmetric A matrix with $A_{11} = 0$ [184].)

We conclude that, similar to models of Abelian flavor symmetries, one can construct models of non-Abelian symmetries in which the symmetry solves both the ε_K and the d_N problems. These models are however not the generic ones in this framework.

Finally, one can construct models of **heavy first two generation squarks**. Here, the basic mechanism to suppress flavor changing processes is actually flavor diagonal so that $m_{\tilde{q}_{1,2}} \sim 20$ TeV. Naturalness does not allow higher masses, but this mass scale is not enough to satisfy even the Δm_K constraint [185], and one has to invoke alignment, $K_{12}^d \sim \lambda$. This is still not enough to satisfy the ε_K constraint of Equation (185), and a somewhat small phase is required.

Three more comments are in order: First, in this framework, gauginos are significantly lighter than the first two generation squarks, and so RGE cannot induce degeneracy. Second, the large mass of the squarks is enough to solve the EDM related problems, and so it is only the ε_K constraint that motivates a special phase structure. Finally, the contribution to ε'/ε is negligibly small. Instead of (189), a more likely estimate is [155] $(\delta_{LR}^d)_{ij} \sim \dfrac{m_Z (M_d)_{ij}}{(10 \text{ TeV})^2}$, which suppresses the relevant matrix elements by a factor of order 10^4.

7.4   (S)Conclusions

We would like to emphasise the following points:

- For supersymmetry to be established, a direct observation of supersymmetric particles is necessary. Once it is discovered, then measurements of CP violating observables will be a very sensitive probe of its flavor structure and, consequently, of the mechanism of dynamical supersymmetry breaking.

- It seems possible to distinguish between models of exact universality and models with genuine supersymmetric flavor and CP violation. The former tend to give $d_N \lesssim 10^{-31}$ e cm while the latter usually predict $d_N \gtrsim 10^{-28}$ e cm.

- The proximity of $a_{\psi K_S}$ to the SM predictions is obviously consistent with models of exact universality. It disfavors models of heavy squarks such as that of reference [185]. Models of flavor symmetries allow deviations of order 20% (or smaller) from the SM predictions. To be convincingly signalled, an improvement in the

theoretical calculations that lead to the SM predictions for $a_{\psi K_S}$ will be required [186].

- Alternatively, the fact that $K \to \pi \nu \bar{\nu}$ decays are not affected by most supersymmetric flavor models [187, 188, 189] is an advantage here. The Standard Model correlation between $a_{\pi \nu \bar{\nu}}$ and $a_{\psi K_S}$ is a much cleaner test than a comparison of $a_{\psi K_S}$ to the CKM constraints.

- The neutral D system provides a stringent test of alignment. Observation of CP violation in the $D \to K \pi$ decays will make a convincing case for new physics.

Acknowledgments

References

[1] M. Kobayashi and T. Maskawa, Prog. Theor. Phys. **49**, 652 (1973).

[2] J. H. Christenson, J. W. Cronin, V. L. Fitch and R. Turlay, Phys. Rev. Lett. **13**, 138 (1964).

[3] H. Burkhardt *et al.* [NA31 Collaboration], Phys. Lett. B **206**, 169 (1988).

[4] G. D. Barr *et al.* [NA31 Collaboration], Phys. Lett. B **317**, 233 (1993).

[5] L. K. Gibbons *et al.*, Phys. Rev. Lett. **70**, 1203 (1993).

[6] J. Graham, FNAL seminar on '*A new measurement of ϵ'/ϵ from KTeV*,' June 8th, 2001.

[7] G. Unal, CERN seminar on '*A new Measurement of Direct CP Violation by NA48*,' May 10th, 2001.

[8] K. Ackerstaff *et al.* [OPAL collaboration], Eur. Phys. J. C **5**, 379 (1998) [hep-ex/9801022].

[9] T. Affolder *et al.* [CDF Collaboration], Phys. Rev. D **61**, 072005 (2000) [hep-ex/9909003].

[10] R. Barate *et al.* [ALEPH Collaboration], Phys. Lett. B **492**, 259 (2000) [hep-ex/0009058].

[11] B. Aubert *et al.* [BaBar Collaboration], Phys. Rev. Lett. **87**, 091801 (2001).

[12] K. Abe *et al.* [Belle Collaboration], Phys. Rev. Lett. **87**, 091802 (2001) [hep-ex/0107061].

[13] L. Wolfenstein, Phys. Rev. Lett. **13**, 562 (1964).

[14] A. D. Sakharov, Pisma Zh. Eksp. Teor. Fizx. **5**, 32 (1967) [JETP Lett. **5**, 24 (1967)].

[15] S. Burles, K. M. Nollett and M. S. Turner, Astrophys. J. **552**, L1 (2001) [astro-ph/0010171].

[16] G. R. Farrar and M. E. Shaposhnikov, Phys. Rev. D **50**, 774 (1994) [hep-ph/9305275].

[17] P. Huet and E. Sather, Phys. Rev. D **51**, 379 (1995) [hep-ph/9404302].

[18] M. B. Gavela, M. Lozano, J. Orloff and O. Pene, Nucl. Phys. B **430**, 345 (1994).

[19] A. Riotto and M. Trodden, Ann. Rev. Nucl. Part. Sci. **49**, 35 (1999) [hep-ph/9901362].

[20] W. Buchmuller and M. Plumacher, hep-ph/0007176.

[21] G. C. Branco, T. Morozumi, B. M. Nobre and M. N. Rebelo, hep-ph/0107164.

[22] A. G. Cohen, D. B. Kaplan and A. E. Nelson, Ann. Rev. Nucl. Part. Sci. **43**, 27 (1993) [hep-ph/9302210].

[23] R. J. Crewther, P. Di Vecchia, G. Veneziano and E. Witten, Phys. Lett. B **88**, 123 (1979) [Erratum-ibid. B **91**, 487 (1979)].

[24] L. J. Dixon, A. Langnau, Y. Nir and B. Warr, Phys. Lett. B **253**, 459 (1991).

[25] P. G. Harris *et al.*, Phys. Rev. Lett. **82**, 904 (1999).

[26] M. Dine, hep-ph/0011376.

[27] T. Banks, Y. Nir and N. Seiberg, hep-ph/9403203.

[28] R. D. Peccei and H. R. Quinn, Phys. Rev. Lett. **38**, 1440 (1977).

[29] R. D. Peccei and H. R. Quinn, Phys. Rev. D **16**, 1791 (1977).

[30] M. Dine, R. G. Leigh and D. A. MacIntire, Phys. Rev. Lett. **69**, 2030 (1992).

[31] K. Choi, D. B. Kaplan and A. E. Nelson, Nucl. Phys. B **391**, 515 (1993) [hep-ph/9205202].

[32] M. P. Worah, Phys. Rev. Lett. **79**, 3810 (1997) [hep-ph/9704389].

[33] M. P. Worah, Phys. Rev. D **56**, 2010 (1997) [hep-ph/9702423].

[34] C. Jarlskog, Phys. Rev. Lett. **55**, 1039 (1985).

[35] N. Cabibbo, Phys. Rev. Lett. **10**, 531 (1963).

[36] Z. Maki, M. Nakagawa and S. Sakata, Prog. Theor. Phys. **28**, 870 (1962).

[37] L. Chau and W. Keung, Phys. Rev. Lett. **53**, 1802 (1984).

[38] L. Wolfenstein, Phys. Rev. Lett. **51**, 1945 (1983).

[39] C. Dib, I. Dunietz, F. J. Gilman and Y. Nir, Phys. Rev. D **41**, 1522 (1990).

[40] J. L. Rosner, A. I. Sanda and M. P. Schmidt, EFI-88-12-CHICAGO *Presented at Workshop on High Sensitivity Beauty Physics, Batavia, IL, Nov 11-14, 1987.*

[41] A. B. Carter and A. I. Sanda, Phys. Rev. Lett. **45**, 952 (1980).

[42] A. B. Carter and A. I. Sanda, Phys. Rev. D **23**, 1567 (1981).

[43] I. I. Bigi and A. I. Sanda, Nucl. Phys. B **193**, 85 (1981).

[44] I. Dunietz and J. L. Rosner, Phys. Rev. D **34**, 1404 (1986).

[45] I. I. Bigi and A. I. Sanda, Nucl. Phys. B **281**, 41 (1987).

[46] H. Harari and Y. Nir, Phys. Lett. B **195**, 586 (1987).

[47] A. Hocker, H. Lacker, S. Laplace and F. Le Diberder, Eur. Phys. J. C **21**, 225 (2001) [hep-ph/0104062] (updates on http://www.slac.stanford.edu/~laplace/ckmfitter.html)

[48] G. Buchalla and A. J. Buras, Phys. Lett. B **333**, 221 (1994) [hep-ph/9405259].

[49] G. Buchalla and A. J. Buras, Phys. Rev. D **54**, 6782 (1996) [hep-ph/9607447].

[50] S. Bergmann and G. Perez, JHEP **0008**, 034 (2000) [hep-ph/0007170].

[51] G. C. Branco, L. Lavoura, J. P. Silva, "CP violation," *Oxford, UK: Clarendon (1999).*

[52] Y. Grossman, B. Kayser and Y. Nir, Phys. Lett. B **415**, 90 (1997) [hep-ph/9708398].

[53] A. J. Buras, hep-ph/0101336.

[54] S. L. Glashow, J. Iliopoulos and L. Maiani, Phys. Rev. D **2**, 1285 (1970).

[55] L. S. Littenberg, Phys. Rev. D **39**, 3322 (1989).

[56] G. Buchalla and G. Isidori, Phys. Lett. B **440**, 170 (1998) [hep-ph/9806501].

[57] G. Perez, JHEP **9909**, 019 (1999) [hep-ph/9907205].

[58] G. Perez, JHEP **0002**, 043 (2000) [hep-ph/0001037].

[59] S. Adler *et al.* [E787 Collaboration], Phys. Rev. Lett. **84**, 3768 (2000) [hep-ex/0002015].

[60] A. Alavi-Harati *et al.* [The E799-II/KTeV Collaboration], Phys. Rev. D **61**, 072006 (2000).

[61] Y. Grossman and Y. Nir, Phys. Lett. B **398**, 163 (1997) [hep-ph/9701313].

[62] S. Bergmann, Y. Grossman, Z. Ligeti, Y. Nir and A. A. Petrov, Phys. Lett. B **486**, 418 (2000) [hep-ph/0005181].

[63] S. Bergmann and Y. Nir, JHEP **9909**, 031 (1999) [hep-ph/9909391].

[64] G. D'Ambrosio and D. Gao, Phys. Lett. B **513**, 123 (2001) [hep-ph/0105078].

[65] Y. Nir, hep-ph/9911321.

[66] G. Blaylock, A. Seiden and Y. Nir, Phys. Lett. B **355**, 555 (1995) [hep-ph/9504306].

[67] L. Wolfenstein, Phys. Rev. Lett. **75**, 2460 (1995) [hep-ph/9505285].

[68] R. Godang *et al.* [CLEO Collaboration], Phys. Rev. Lett. **84**, 5038 (2000) [hep-ex/0001060].

[69] L. Chau and H. Cheng, Phys. Lett. B **333**, 514 (1994) [hep-ph/9404207].

[70] T. E. Browder and S. Pakvasa, Phys. Lett. B **383**, 475 (1996) [hep-ph/9508362].

[71] A. F. Falk, Y. Nir and A. A. Petrov, JHEP **9912**, 019 (1999) [hep-ph/9911369].

[72] E. Golowich and S. Pakvasa, Phys. Lett. B **505**, 94 (2001) [hep-ph/0102068].

[73] M. Gronau, Y. Grossman and J. L. Rosner, Phys. Lett. B **508**, 37 (2001) [hep-ph/0103110].

[74] J. M. Link *et al.* [FOCUS Collaboration], Phys. Lett. B **485**, 62 (2000) [hep-ex/0004034].

[75] E. M. Aitala *et al.* [E791 Collaboration], Phys. Rev. Lett. **83**, 32 (1999) [hep-ex/9903012].

[76] A. B. Smith [CLEO Collaboration], hep-ex/0104008.

[77] B. Yabsley [Belle Collaboration], talk at the International Europhysics Conference on High Energy Physics (Budapest, Hungary, July 2001).

[78] J. Bartelt *et al.* [CLEO Collaboration], Phys. Rev. D **52**, 4860 (1995).

[79] E. M. Aitala *et al.* [E791 Collaboration], Phys. Lett. B **421**, 405 (1998) [hep-ex/9711003].

[80] J. M. Link *et al.* [FOCUS Collaboration], Phys. Lett. B **491**, 232 (2000) [Erratum-ibid. B **495**, 443 (2000)] [hep-ex/0005037].

[81] G. Bonvicini *et al.* [CLEO Collaboration], Phys. Rev. D **63**, 071101 (2001) [hep-ex/0012054].

[82] D. E. Jaffe *et al.* [CLEO Collaboration], Phys. Rev. Lett. **86**, 5000 (2001) [hep-ex/0101006].

[83] G. Abbiendi *et al.* [OPAL Collaboration], Eur. Phys. J. C **12**, 609 (2000) [hep-ex/9901017].

[84] R. Barate *et al.* [ALEPH Collaboration], Eur. Phys. J. C **20**, 431 (2001).

[85] B. Aubert, hep-ex/0107059.

[86] I. I. Bigi, V. Khoze, N. G. Uraltsev and A. I. Sanda, in *CP Violation*, ed. C. Jarlskog (World Scientific, Singapore, 1992).

[87] R. Aleksan, A. Le Yaouanc, L. Oliver, O. Pene, J. C. Raynal, Phys. Lett. B **316**, 567 (1993).

[88] M. Beneke, G. Buchalla, C. Greub, A. Lenz and U. Nierste, Phys. Lett. B **459**, 631 (1999).

[89] L. Wolfenstein, Phys. Rev. D **57**, 5453 (1998).

[90] G. Buchalla, A. J. Buras and M. E. Lautenbacher, Rev. Mod. Phys. **68**, 1125 (1996).

[91] J. S. Hagelin, Nucl. Phys. B **193**, 123 (1981).

[92] A. J. Buras, W. Slominski and H. Steger, Nucl. Phys. B **245**, 369 (1984).

[93] M. Beneke, G. Buchalla and I. Dunietz, Phys. Rev. D **54**, 4419 (1996) [hep-ph/9605259].

[94] L. Randall and S. Su, Nucl. Phys. B **540**, 37 (1999) [hep-ph/9807377].

[95] R. N. Cahn and M. P. Worah, Phys. Rev. D **60**, 076006 (1999) [hep-ph/9904480].

[96] G. Barenboim, G. Eyal and Y. Nir, Phys. Rev. Lett. **83**, 4486 (1999) [hep-ph/9905397].

[97] G. Eyal and Y. Nir, JHEP **9909**, 013 (1999) [hep-ph/9908296].

[98] J. Charles, Phys. Rev. D **59**, 054007 (1999) [hep-ph/9806468].

[99] B. Aubert, hep-ex/0107074.

[100] D. Cronin-Hennessy *et al.* [CLEO Collaboration], hep-ex/0001010.

[101] K. Abe *et al.* [BELLE Collaboration], Phys. Rev. Lett. **87**, 101801 (2001) [hep-ex/0104030].

[102] B. Aubert *et al.* [BABAR Collaboration], hep-ex/0105061.

[103] Y. Nir and H. R. Quinn, Phys. Rev. Lett. **67**, 541 (1991).

[104] J. P. Silva and L. Wolfenstein, Phys. Rev. D **49**, 1151 (1994) [hep-ph/9309283].

[105] O. F. Hernandez, D. London, M. Gronau and J. L. Rosner, Phys. Lett. B **333**, 500 (1994).

[106] M. Neubert and J. L. Rosner, Phys. Lett. B **441**, 403 (1998) [hep-ph/9808493].

[107] R. Fleischer, Int. J. Mod. Phys. A **12**, 2459 (1997) [hep-ph/9612446].

[108] M. Gronau and D. London, Phys. Rev. Lett. **65**, 3381 (1990).

[109] Y. Grossman and H. R. Quinn, Phys. Rev. D **58**, 017504 (1998) [hep-ph/9712306].

[110] M. Gronau, D. London, N. Sinha and R. Sinha, Phys. Lett. B **514**, 315 (2001).

[111] H. J. Lipkin, Y. Nir, H. R. Quinn and A. Snyder, Phys. Rev. D **44**, 1454 (1991).

[112] M. Gronau, Phys. Lett. B **265**, 389 (1991).

[113] A. E. Snyder and H. R. Quinn, Phys. Rev. D **48**, 2139 (1993).

[114] H. R. Quinn and J. P. Silva, Phys. Rev. D **62**, 054002 (2000) [hep-ph/0001290].

[115] M. Beneke, G. Buchalla, M. Neubert, C. T. Sachrajda, Phys. Rev. Lett. **83**, 1914 (1999).

[116] M. Beneke, G. Buchalla, M. Neubert and C. T. Sachrajda, Nucl. Phys. B **591**, 313 (2000).

[117] M. Beneke, G. Buchalla, M. Neubert and C. T. Sachrajda, Nucl. Phys. B **606**, 245 (2001).

[118] Y. Keum, H. Li and A. I. Sanda, Phys. Lett. B **504**, 6 (2001) [hep-ph/0004004].

[119] G. Buchalla, these proceedings.

[120] Y. Grossman, Y. Nir and M. P. Worah, Phys. Lett. B **407**, 307 (1997) [hep-ph/9704287].

[121] H. Georgi and S. L. Glashow, Phys. Lett. B **451**, 372 (1999) [hep-ph/9807399].

[122] P. H. Frampton, S. L. Glashow and T. Yoshikawa, Phys. Rev. Lett. **87**, 011801 (2001).

[123] D. Chang, W. Keung and R. N. Mohapatra, Phys. Lett. B **515**, 431 (2001).

[124] P. Ball, J. M. Frere and J. Matias, Nucl. Phys. B **572**, 3 (2000) [hep-ph/9910211].

[125] S. Bergmann and G. Perez, hep-ph/0103299.

[126] A. Pomarol, Phys. Rev. D **47**, 273 (1993) [hep-ph/9208205].

[127] K. S. Babu and S. M. Barr, Phys. Rev. Lett. **72**, 2831 (1994) [hep-ph/9309249].

[128] G. Eyal and Y. Nir, Nucl. Phys. B **528**, 21 (1998) [hep-ph/9801411].

[129] Y. Grossman and M. P. Worah, Phys. Lett. B **395**, 241 (1997) [hep-ph/9612269].

[130] Y. Nir and D. J. Silverman, Nucl. Phys. B **345**, 301 (1990).

[131] Y. Shadmi and Y. Shirman, Rev. Mod. Phys. **72**, 25 (2000) [hep-th/9907225].

[132] Y. Grossman, Y. Nir and R. Rattazzi, hep-ph/9701231.

[133] M. Dine, E. Kramer, Y. Nir and Y. Shadmi, Phys. Rev. D **63**, 116005 (2001).

[134] H. E. Haber, Nucl. Phys. Proc. Suppl. **62**, 469 (1998) [hep-ph/9709450].

[135] S. Abel, these proceedings.

[136] M. Dugan, B. Grinstein and L. Hall, Nucl. Phys. B **255**, 413 (1985).

[137] S. Dimopoulos and S. Thomas, Nucl. Phys. B **465**, 23 (1996) [hep-ph/9510220].

[138] W. Buchmuller and D. Wyler, Phys. Lett. B **121**, 321 (1983).

[139] J. Polchinski and M. B. Wise, Phys. Lett. B **125**, 393 (1983).

[140] W. Fischler, S. Paban and S. Thomas, Phys. Lett. B **289**, 373 (1992) [hep-ph/9205233].

[141] E. D. Commins, S. B. Ross, D. DeMille and B. C. Regan, Phys. Rev. A **50**, 2960 (1994).

[142] Y. K. Semertzidis *et al.*, hep-ph/0012087.

[143] J. L. Feng, K. T. Matchev and Y. Shadmi, hep-ph/0107182.

[144] F. Gabbiani, E. Gabrielli, A. Masiero and L. Silvestrini, Nucl. Phys. B **477**, 321 (1996).

[145] Y. Nir, Nucl. Phys. B **273**, 567 (1986).

[146] A. J. Buras, G. Colangelo, G. Isidori, A. Romanino and L. Silvestrini, Nucl. Phys. B **566**, 3 (2000) [hep-ph/9908371].

[147] A. Masiero and H. Murayama, Phys. Rev. Lett. **83**, 907 (1999) [hep-ph/9903363].

[148] S. Baek, J. H. Jang, P. Ko and J. H. Park, Phys. Rev. D **62**, 117701 (2000).

[149] S. Baek, J. H. Jang, P. Ko and J. H. Park, Nucl. Phys. B **609**, 442 (2001) [hep-ph/0105028].

[150] S. Khalil, T. Kobayashi and A. Masiero, Phys. Rev. D **60**, 075003 (1999) [hep-ph/9903544].

[151] K. S. Babu, B. Dutta and R. N. Mohapatra, Phys. Rev. D **61**, 091701 (2000).

[152] S. Khalil and T. Kobayashi, Phys. Lett. B **460**, 341 (1999) [hep-ph/9906374].

[153] A. L. Kagan and M. Neubert, Phys. Rev. Lett. **83**, 4929 (1999) [hep-ph/9908404].

[154] S. Khalil, T. Kobayashi and O. Vives, Nucl. Phys. B **580**, 275 (2000) [hep-ph/0003086].

[155] G. Eyal, A. Masiero, Y. Nir and L. Silvestrini, JHEP **9911**, 032 (1999) [hep-ph/9908382].

[156] S. A. Abel and J. M. Frere, Phys. Rev. D **55**, 1623 (1997) [hep-ph/9608251].

[157] S. Baek and P. Ko, Phys. Lett. B **462**, 95 (1999) [hep-ph/9904283].

[158] G. C. Branco, G. C. Cho, Y. Kizukuri and N. Oshimo, Phys. Lett. B **337**, 316 (1994).

[159] T. Goto, T. Nihei and Y. Okada, Phys. Rev. D **53**, 5233 (1996) [Erratum-ibid. D **54**, 5904 (1996)] [hep-ph/9510286].

[160] M. Dine, A. E. Nelson and Y. Shirman, Phys. Rev. D **51**, 1362 (1995).

[161] M. Dine, A. E. Nelson, Y. Nir and Y. Shirman, Phys. Rev. D **53**, 2658 (1996) [hep-ph/9507378].

[162] A. Romanino and A. Strumia, Nucl. Phys. B **490**, 3 (1997) [hep-ph/9610485].

[163] K. S. Babu, C. Kolda and F. Wilczek, Phys. Rev. Lett. **77**, 3070 (1996) [hep-ph/9605408].

[164] M. Dine, Y. Nir and Y. Shirman, Phys. Rev. D **55**, 1501 (1997) [hep-ph/9607397].

[165] D. A. Demir, A. Masiero and O. Vives, Phys. Lett. B **479**, 230 (2000) [hep-ph/9911337].

[166] J. Louis and Y. Nir, Nucl. Phys. B **447**, 18 (1995) [hep-ph/9411429].

[167] A. Brignole, L. E. Ibanez and C. Munoz, Nucl. Phys. B **422**, 125 (1994) [Erratum-ibid. B **436**, 747 (1994)] [hep-ph/9308271].

[168] R. Barbieri, J. Louis and M. Moretti, Phys. Lett. B **312**, 451 (1993) [Erratum-ibid. B **316**, 632 (1993)] [hep-ph/9305262].

[169] L. Randall and R. Sundrum, Nucl. Phys. B **557**, 79 (1999) [hep-th/9810155].

[170] G. F. Giudice, M. A. Luty, H. Murayama and R. Rattazzi, JHEP **9812**, 027 (1998) [hep-ph/9810442].

[171] J. A. Bagger, T. Moroi and E. Poppitz, JHEP **0004**, 009 (2000) [hep-th/9911029].

[172] A. Pomarol and R. Rattazzi, JHEP **9905**, 013 (1999) [hep-ph/9903448].

[173] E. Katz, Y. Shadmi and Y. Shirman, JHEP **9908**, 015 (1999) [hep-ph/9906296].

[174] Z. Chacko, M. A. Luty, E. Ponton, Y. Shadmi and Y. Shirman, Phys. Rev. D **64**, 055009 (2001).

[175] D. E. Kaplan, G. D. Kribs and M. Schmaltz, Phys. Rev. D **62**, 035010 (2000).

[176] Z. Chacko, M. A. Luty, A. E. Nelson and E. Ponton, JHEP **0001**, 003 (2000).

[177] Y. Nir and N. Seiberg, Phys. Lett. B **309**, 337 (1993) [hep-ph/9304307].

[178] M. Leurer, Y. Nir and N. Seiberg, Nucl. Phys. B **420**, 468 (1994) [hep-ph/9310320].

[179] Y. Nir and R. Rattazzi, Phys. Lett. B **382**, 363 (1996) [hep-ph/9603233].

[180] M. Dine, R. Leigh and A. Kagan, Phys. Rev. D **48**, 4269 (1993) [hep-ph/9304299].

[181] R. Barbieri, G. Dvali and L. J. Hall, Phys. Lett. B **377**, 76 (1996) [hep-ph/9512388].

[182] C. D. Carone, L. J. Hall and H. Murayama, Phys. Rev. D **54**, 2328 (1996).

[183] R. Barbieri, L. J. Hall, S. Raby and A. Romanino, Nucl. Phys. B **493**, 3 (1997) [hep-ph/9610449].

[184] R. Barbieri, R. Contino and A. Strumia, Nucl. Phys. B **578**, 153 (2000) [hep-ph/9908255].

[185] A. G. Cohen, D. B. Kaplan, F. Lepeintre, A. E. Nelson, Phys. Rev. Lett. **78**, 2300 (1997).

[186] G. Eyal, Y. Nir and G. Perez, JHEP **0008**, 028 (2000) [hep-ph/0008009].

[187] Y. Nir and M. P. Worah, Phys. Lett. B **423**, 319 (1998) [hep-ph/9711215].

[188] A. J. Buras, A. Romanino and L. Silvestrini, Nucl. Phys. B **520**, 3 (1998) [hep-ph/9712398].

[189] G. Colangelo and G. Isidori, JHEP **9809**, 009 (1998) [hep-ph/9808487].

CP and flavour in supersymmetry

Steven Abel

Institute for Particle Physics Phenomenology
Durham University, U.K.

1   Introduction

The brief for these lectures is to introduce some of the ideas and challenges for super-symmetry coming from flavour and CP physics, aimed at an audience that includes 'non-experts'. It is usual for courses on this subject to jump in at the superpotential for the Minimal Supersymmetric Standard Model, and leave the reader to sort out for him or herself what the superpotential actually means. In this course I have instead decided to begin at square one, assuming no prior knowledge other than field theory of scalars and fermions. While it might seem a challenge to get from this point to the MSSM in a just page or two of working, this is in fact possible (and only slightly painful) if we do not spend too long worrying about things like 'superspace', which although quite fundamen-tal, do not really have much bearing on the question of phenomenology. Thus the first section will develop the idea of a superpotential by giving a quick and dirty demonstration of how to 'supersymmetrise' a fundamental scalar. By following the (very explicit) steps in this part, those readers who are unfamiliar with supersymmetry can get a grasp of the various concepts involved. In particular we will see at first hand the famous cancellation of divergences that makes supersymmetry such an important idea. This will lead natu-rally to the MSSM superpotential and from there we can go on to analyse the particular problems that flavour and CP throw up for supersymmetry.

The interests of a SUSSP audience are typically rather broad, so the course is organised as follows. The reader who requires a brief overview of the subject will be able to follow the main body of the text which is self contained. However, since some of the audience may want to develop a deeper knowledge of the subject, I have made copious use of appendices to include more detailed discussions of various aspects as we go along. From these it should be possible for the reader to follow some of the more formal texts in [1].

Why supersymmetry?

Although there are several responses to this question, the most important is undoubtedly cancellation of divergences. This is a remarkable feature of supersymmetric models that gives stability to any theory, specifically stability of mass scales. We shall shortly derive it in detail, but in brief it goes as follows.

Consider a theory involving scalars with self coupling λ_b, the Higgs particles, and also fermions which couple to the Higgs particles in Yukawa couplings with strength λ_f.
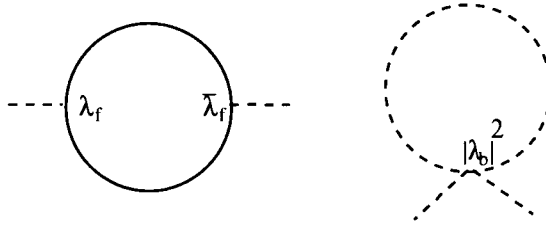
Figure 1. *Contributions to the Higgs potential.*

At one loop order, the Higgs potential will get a divergent contribution from the diagrams shown in Figure 1, which is of the form

$$\delta V_{\text{Higgs}} \sim (|\lambda_b|^2 - |\lambda_f|^2)H^2\Lambda^2. \tag{1}$$

The fermionic loop gets an extra minus sign. This type of contribution is a disaster for electroweak symmetry breaking, since we will have to tune our λ_b and λ_f couplings very precisely (to at least one part in 10^{14} if we take $\Lambda > M_{\text{GUT}} \sim 10^{16}\text{GeV}$) if the physical Higgs mass and hence weak gauge bosons are to end up around the 100GeV mark. Moreover, the λ couplings will themselves have logarithmic radiative corrections from one-loop diagrams, and in fact we need to re-tune at each order in perturbation theory. Finally the low energy theory will be sensitive to unknown high energy physics - for example an unknown GUT theory or string theory. The problem introduced by the existence of vastly different mass scales is referred to as the *gauge-hierarchy* problem. If you believe that the Higgs fields are the result of some non-perturbative process (a top quark condensate perhaps) happening just above the weak scale, or that there are no new scales above M_W, or more generally, that there are no fundamental scalars, then this will not be a problem for you.

Supersymmetry solves these technical problems by ensuring that $\lambda_f = \lambda_b$, and in fact all contributions to the effective potential then vanish. The only renormalisation of the bare couplings which takes place comes from field renormalisations. The price we have to pay for such a useful cancellation is high; we have to introduce a fermionic(bosonic) degree of freedom for every boson(fermion) (known collectively as *superpartners*). The general form of the couplings will be

$$\lambda(\mu_2) \sim \lambda(\mu_1) \log \frac{\mu_1^2}{\mu_2^2}. \tag{2}$$

Now couplings which start off small at the GUT scale remain small. (This holds to all orders in perturbation theory.)

This then is the main reason for taking supersymmetry seriously. Other reasons (in descending order of compulsion) are that the gauge couplings of the supersymmetric standard model (MSSM) appear to exactly unify at some (presumably GUT) scale, or that the only known theory of quantum gravity (*i.e.* string theory) strongly suggests it, or that the cosmological constant is zero. This last reason is probably invalid, since supersymmetry must be broken at some scale, reintroducing a cosmological constant (albeit one which is only out by 52 orders of magnitude instead of 120). One additional interesting hint for extra scalars comes from the fact that for relatively low mass Higgs particles (around 115GeV say) the Standard Model vacuum is unstable to decay, as shown in [2] and references therein, because the top mass is large. This instability can be removed if there are more scalars than just the Higgs fields.

2 Express supersymmetry

Our aim in this section is to derive (or at least explain the form of) supersymmetric interactions from their non-supersymmetric counterparts (SM). We begin in this subsection by supersymmetrising a simple Lagrangian describing a massless, free boson. We will then discuss how to write down interactions using the superpotential and examine the top quark Yukawa coupling in detail, in particular verifying the cancellation discussed above. Readers already familiar with supersymmetry can skip to Section 3. Readers who would like to see the more detailed conventional approach can consult the standard texts [1].

Let us begin by considering the Lagrangian for a free boson:

$$\mathcal{L} = -(\partial_\mu \phi^*)(\partial^\mu \phi). \tag{3}$$

Since we wish to describe both particles and anti-particles, ϕ is complex with two degrees of freedom. Also ϕ has spin 0 and dimension 1. This Lagrangian resembles that of the Higgs which is already present in the SM. Eventually, each chirality (left or right) of matter fields will aquire a scalar superpartner of this type, together forming *chiral supermultiplets*.

Now let us introduce our fermion superpartner which must also have two degrees of freedom for a sensible transformation. Before continuing the reader should familiarise themselves with the Weyl notation for fermions which has become standard for this subject and is outlined in Appendix A. The supersymmetry transformation should by definition turn the scalars into fermions so, noting that a two component spinor has the same number of degrees of freedom as the scalar, let us introduce an infinitesimal (linear) supersymmetry transformation δ_ξ on the complex boson above, involving such a spinor:

$$\delta_\xi \phi = a \xi^\alpha \chi_\alpha. \tag{4}$$

The fermionic parameter ξ^α describes the magnitude of the transformation, and a is some as yet undetermined constant. We will often drop the fermionic indices α when they are contracted. The spin and dimension on both sides must match, so that ξ^α must have spin 1/2 (it anticommutes with itself and χ) and dimension 1/2. Next we wish to have an equivalent transformation for the spinor itself. Matching spin and dimensions again, this can be of the form,

$$\delta_\xi \chi_\alpha = b \sigma^\mu_{\alpha\dot\alpha} \bar\xi^{\dot\alpha} \partial_\mu \phi + c \xi_\alpha F. \tag{5}$$

where F is some field of spin 0 and dimension 2 whose existence will be explained shortly. Then for F we can have;

$$\delta_\xi F = d\overline{\xi}_{\dot{\alpha}}\overline{\sigma}^{\mu\dot{\alpha}\alpha}\partial_\mu\chi_\alpha. \tag{6}$$

In order to determine the constants a, b, c, d we need to consider the commutation of two transformations. For example,

$$[\delta_\xi, \delta_\eta]\phi = ab(\eta\sigma^\mu\overline{\chi} - \chi\sigma^\mu\overline{\eta})\partial_\mu\phi. \tag{7}$$

This commutator involves only one field on each side, and is therefore related to the underlying symmetry (supersymmetry). It describes a relation between the δ_ξ and ∂_μ operators, and should not depend on which field (ϕ, χ or F) we choose to express it. (The equivalent statement in gauge theory, is that the Lie-algebra $[T^a, T^b] = if_{abc}T^c$ is formed by successive gauge transformations on any of the fields in the theory, irrespective of their representation.) We need to ensure that the algebra is the same for the field χ_α and this is why the field F is required; in fact

$$[\delta_\xi, \delta_\eta]\chi = cd(\eta\sigma^\mu\overline{\chi} - \chi\sigma^\mu\overline{\eta})\partial_\mu\chi, \tag{8}$$

giving $ab = cd$. The extra field F is known as an *auxiliary* field since it is there simply to make the algebra close. It does not propagate, and so it is easily removed from the Lagrangian using its equation of motion. The full chiral supermultiplet consists of the three fields $\Phi = \{\phi, \chi, F\}$ (and the anti-chiral supermultiplet, $\overline{\Phi} = \{\phi^*, \overline{\chi}, F^*\}$). Having determined the multiplet structure, we are able to write down the simplest invariant Lagrangian. Using the usual normalisations, our first guess is,

$$\mathcal{L}_{\text{KE}} = -i\chi\sigma^\mu\partial_\mu\overline{\chi} + \phi^*\partial^2\phi + F^*F \tag{9}$$

In order for this to be invariant (*i.e.* to transform into itself plus total derivatives under the operations (4)–(6)), we find that we require

$$b = ia, \qquad d = ic, \tag{10}$$

which leaves only one degree of freedom a which normalises the operators δ_ξ. Conventionally we take $a = \sqrt{2}$.

So far we have only the kinetic terms of our Lagrangian. In order to be able to include Yukawa couplings, for example, we need to consider additional interaction terms. As a starting point, notice that the auxiliary field F transforms into a total derivative under a supersymmetry transformation. Thus

$$\mathcal{L}_F = F \tag{11}$$

is a suitable invariant. By itself this is of course not much use, since we also wish to include mass terms and gauge symmetries. In order to do this we must first generalise our infinitesimal transformations to make them finite; a transformation through a finite *spinor* parameter θ_α is given by the operator $\exp \delta_\theta$. Arbitrary transformations through θ_α of the scalar field $\phi(x)$ will change its value in a one-to one manner. We can now construct a scalar quantity, $\Phi(x, \theta, \overline{\theta})$, which gives $\phi(x)$ for arbitrary values of θ_α. Taking $\phi(x)$ as the starting value when $\theta_\alpha = 0$,

$$\Phi(x, 0, 0) = \phi(x), \tag{12}$$

we can find the general form of Φ by transforming ϕ;

$$
\begin{aligned}
\Phi(x,\theta,\bar{\theta}) &= \left(1 + \delta_\theta + \frac{1}{2!}\delta_\theta{}^2 + \frac{1}{3!}\delta_\theta{}^3 + \frac{1}{4!}\delta_\theta{}^4 + \ldots\right)\phi(x) \\
&= \phi(y) + \sqrt{2}\theta\chi(y) + \theta\theta F(y)
\end{aligned} \tag{13}
$$

where $y^\mu = x^\mu + i\theta\sigma^\mu\bar{\theta}$. We have used the identities $\theta^2\theta_\alpha = 0$ (because θ_α has only two components which anticommute), and $\epsilon^{\alpha\beta}\theta\theta = -2\theta^\alpha\theta^\beta$. These scalar functions are referred to as chiral-*superfields*. The Taylor expansion in θs of any function terminates at the $\theta\theta$ term, and consequently any general function of y and θ is a chiral superfield, and its components must transform as the components of a chiral supermultiplet. Beginning with ϕ^* results instead in *anti*-chiral superfields (denoted $\bar{\Phi}$) which are functions of y^* and $\bar{\theta}$ only.

Chiral superfields enable us to build up more general Lagrangians, once we notice that (trivially) the product of two or more chiral superfields is also of this form and therefore is also a chiral superfield. So, for example, the F-term of the product of two chiral fields gives mass terms,

$$
\mathcal{L}_m = \Phi^2\Big|_F + \text{h.c.} = \phi F - \frac{1}{2}\psi^\alpha\psi_\alpha + \text{h.c.} \tag{14}
$$

It is easy to check that this is invariant under the infinitesimal supersymmetry transformations. For the Yukawa interactions we need the product of three superfields,

$$
\mathcal{L}_{\text{Yuk}} = \frac{1}{3}\lambda_{ijk}\Phi_i\Phi_j\Phi_k\Big|_F + \text{h.c.} = \lambda_{ijk}F_i\phi_j\phi_k - \lambda_{ijk}\psi_i^\alpha\psi_j\alpha\phi_k + \text{h.c.} \tag{15}
$$

The generic indices i,j,k stand for summation over gauge representation and/or field labels (e.g. Higgs, matter etc.). Terms of order higher than three make the theory non-renormalisable and will not be considered further. The most general interaction Lagrangian is therefore of the form

$$
\mathcal{L}_{\text{int}} = W(\Phi_i)|_F + \text{h.c.} \tag{16}
$$

where $W(\Phi_i)$ is a third order polynomial of the chiral superfields, known as the *superpotential*. In terms of the superpotential, the Lagrangian is

$$
\mathcal{L}_{\text{int}} = F_i\frac{\partial W}{\partial\phi_i} - \frac{1}{2}\psi_i\psi_j\frac{\partial^2 W}{\partial\phi_i\partial\phi_j} + \text{h.c.} \tag{17}
$$

where only the scalar part of the potential is taken, and summation over indices is implied. Finally, by using the equations of motion of F_i,

$$
F_i = -\frac{\partial W}{\partial\phi_i} \tag{18}
$$

we are able to eliminate it, to give

$$
\mathcal{L}_{\text{int}} = -\left(\frac{\partial W}{\partial\phi_i}\right)^2 - \frac{1}{2}\psi_i\psi_j\frac{\partial^2 W}{\partial\phi_i\partial\phi_j} + \text{h.c.} \tag{19}
$$

In particular we can see that the supersymmetry potential is positive definite, and has a minimum at $\partial W/\partial\phi_i = 0$ where it vanishes.

So we have derived a Lagrangian which is able to describe scalars and their fermionic superpartners. For this introduction we have been avoiding the notion of *superspace*; the interested reader can consult Appendix B which explains the connection. Now let us progress to the simplest non-trivial example, the top quark Yukawa, in order to see the promised cancellation of one loop divergences.

The top quark Yukawa

To write down a top-quark Yukawa we naturally define the superfields to have the exactly the same charges as they would in the Standard Model:

$H_u =$	h_u	$\tilde{h}_u$	F_{h_u}
$Q =$	$\tilde{q}$	q	F_Q
$t_c =$	$\tilde{t}_c$	t_c	F_{t_c}

Thus we have

$$W_{\text{top-Yukawa}} = \lambda_t Q H_u t_c \tag{20}$$

leading to a Lagrangian of the form

$$\mathcal{L}_{\text{top-Yukawa}} = -\lambda_t q h_u t_c - \lambda_t \tilde{q}(\tilde{h}_u t_c) - \lambda_t (q\tilde{h}_u)\tilde{t}_c + \lambda_t^2 |h_u \tilde{t}_c|^2 + \lambda_t^2 |h_u \tilde{q}|^2 + \lambda_t^2 |\tilde{q}\tilde{t}_c|^2. \tag{21}$$

It is customary to write the superpartner of any field as that field with a tilde and we shall follow that convention here. Now consider the one loop contributions to the h_u mass squareds. The internal fermions can only be the pair q and t_c, whereas the scalar 'bubble' diagram gets two contributions separately from $\tilde{q}$ and $\tilde{t}$. Including the symmetry factor of $1/2$ in the scalar diagram, and -1 for the fermion loop, we find exact cancellation as promised.

As yet we have not included the gauge fields (we have no spin one fields), however we can guess what their Lagrangian might look like from the above. Supersymmetry makes fermions and bosons indistinguishable, so not surprisingly the net effect on the Yukawa coupling was to introduce duplicate 'super'-Yukawa couplings with the fermions and bosons permuted (the first three terms in Equation 21). The same duplication happens in the gauge sector as well. Here the physical components of the vector supermultiplets are $\{\lambda_A, v_A^\mu, D_A\}$ (where A is a group index) which are respectively the gauginos, gauge bosons (of the Standard Model) and auxiliary fields. The matter-gauge-matter couplings become

$$\mathcal{L}_{\Phi_i - \Phi_j - V_A} = g\chi_i \sigma_\mu v_{ij}^\mu \overline{\chi}_j + i\sqrt{2}g\phi_i^* \lambda_{ij}\chi_j + h.c. \tag{22}$$

Where we use the usual notation that $v_{ij}^\mu = v_A^\mu T_{ij}^A$. As with the Yukawa, the introduction of supersymmetry has simply added supersymmetric counterpart couplings with the roles of boson and fermion reversed.

The inclusion of gauge symmetry and the supersymmetrisation of the Yang-Mills term involves so-called vector supermultiplets. It is possible (although much more arduous) to build these supermultiplets beginning with the lowest component, as we did above for the chiral supermultiplets. However, for the discussion of CP and flavour phenomenology we now have everything we need; the reader interested in these more technical aspects can turn to Appendix C where a discussion of the total gauge invariant Lagrangian is included for completeness.

3 The softly broken MSSM

We now have a general framework for building supersymmetric gauge invariant Lagrangians. Our aim is to build a model which includes this, and yet mimics the Standard Model at low enough energies. A consensus seems to have been established as to the most likely way this might happen in Nature. This is referred to as the softly broken Minimal Supersymmetric Standard Model (MSSM).

There are a number of influences guiding the form of the MSSM, one of the most overriding being 'naturalness'. A case in point is the scale of supersymmetry breaking itself (clearly a requirement). Most would agree that this should not be too high, otherwise fine-tuning is re-introduced. Because of this, a ceiling of a few TeV is usually placed on this parameter. One should bear in mind however that there are no other upper limits on the scale of supersymmetry breaking. This problem is compounded by the fact that there is still a large number of additional parameters left in the model (supersymmetry breaking scalar masses for example). Often these are reduced by considering a particular kind of supersymmetry breaking mediation – for example mediation by supergravity (local supersymmetry) or gauge mediation, but also they are reduced by considering experiment itself. (For example, the absence of large FCNCs is often cited as the reason for taking scalar masses to be degenerate at the GUT scale.) Because of this, getting genuine predictions out of supersymmetry is difficult, and usually the 'prediction' is merely that supersymmetry is *consistent* with a large $(g-2)_\mu$ or electric dipole moment of the neutron or $b \to s\gamma$ or whatever. In fact, usually supersymmetry is also consistent with values of $(g-2)_\mu$ *etc.* which totally violate current experimental bounds, thus lending an air of 'just around the corner'-ness to the whole business.

This difficulty really arises out of our ignorance about the nature of supersymmetry breaking and leads to what has become known as the supersymmetric flavour and CP problems. These general problems will be illustrated in this section, and expanded upon later. First however, let us present the superpotential of the MSSM.

3.1 The MSSM superpotential

Since we wish to stay as close to the SM as possible, we assume that the same gauge groups and representations will appear. Thus, the group will be $SU(3)_c \times SU(2)_L \times U(1)_Y$. The gauge bosons ($g$, $W_{1,2,3}$ and B) are accompanied by fermionic superpartners, the *gluinos*, *winos* and *bino*. The matter and Higgs fields become chiral supermultiplets. These are usually represented by capital letters, so that we have $Q = \{\tilde{q}_L, q_L\}$, $L = \{\tilde{l}_L, l_L\}$, $U^c = \{\tilde{u}_R^*, u_R^\dagger\}$ and so on for each of the quarks and leptons, and $H = \{h, \tilde{h}\}$ for the Higgs doublet. L, Q and H are $SU(2)$ doublets with

$$L = \begin{pmatrix} \nu_L \\ e_L^- \end{pmatrix} \qquad Q = \begin{pmatrix} u_L \\ d_L \end{pmatrix} \qquad H = \begin{pmatrix} h^0 \\ h^+ \end{pmatrix} \tag{23}$$

The superpartners are the *squarks*, *sleptons* and *Higgsinos* respectively.

We now encounter a minor problem. In the Standard Model we were able to make do with only one Higgs field, to generate two masses. The Yukawa couplings which did this

were of the form
$$\mathcal{L}_{\text{Yuk}} = \lambda_e \bar{l}_L h e_R + \lambda_d \bar{q}_L h d_R + \lambda_u \bar{q}_L \hat{h} u_R + \text{h.c.}. \tag{24}$$
The h scalar has hypercharge $Y = 1$, and the $\hat{h}$ scalar is given by $\hat{h} = i\tau_2 h^*$ with $Y = -1$. (The τ_i are the Dirac matrix generators of $SU(2)$, acting on isospin indices, not to be confused with the σ_i.) Clearly, if we turn these fields into chiral supermultiplets, the two quark mass terms cannot both be derived from a superpotential since if h corresponds to a chiral superfield then h^* corresponds to an *anti*chiral one. An additional problem comes from triangle anomalies. These are a potential problem in any theory with chiral fermions, and their disappearance requires that $\sum_i Y_i = \sum_i Y_i^3 = 0$. In the Standard Model, these cancel exactly, but now this is no longer the case since we have introduced an extra fermion, the Higgsino, which also couples to isospin. Happily, both these problems are solved by the introduction of a new Higgs superfield with hypercharge -1, which replaces $\hat{h}$. Our total particle content has the following assignments and transformations under the Standard Model gauge group,

$$
\begin{aligned}
L &= \begin{pmatrix} \nu_L \\ e_L^- \end{pmatrix} \quad (1,2,-1) & H_1 &= \begin{pmatrix} H_1^0 \\ H_1^- \end{pmatrix} \quad (1,2,-1) \\[2ex]
Q &= \begin{pmatrix} u_L \\ d_L \end{pmatrix} \quad (3,2,1/3) & H_2 &= \begin{pmatrix} H_2^+ \\ H_2^0 \end{pmatrix} \quad (1,2,1) \\[2ex]
E^c & \quad\quad\quad (1,1,2) \\
D^c & \quad\quad\quad (\bar{3},1,2/3) \\
U^c & \quad\quad\quad (\bar{3},1,-4/3)
\end{aligned}
\tag{25}
$$

where H_1 and H_2 are two entirely separate superfields. (There is, by definition, no right handed neutrino in the MSSM although clearly experimental data now suggests one. Neutrino physics requires a review in itself and we shall not attempt to cover it here.) The Yukawa piece of the superpotential (taking a little more care with our isospin indices for a moment) may now be taken to be
$$W_{\text{Yuk}} = \lambda_E E^c H_1^T \epsilon' L + \lambda_D D^c H_1^T \epsilon' Q + \lambda_U U^c H_2^T \epsilon' Q + \text{h.c.}, \tag{26}$$
where we have suppressed colour and generation indices. The $\epsilon' = i\tau_2$ is the same matrix as ϵ, the prime being there simply to remind us that its indices act on isospin. $SU(2)$ gauge invariance follows from the fact that $\epsilon' \tau^i = -(\tau^i)^T \epsilon$, so that $(e^{i\lambda \cdot \tau})^T \epsilon' e^{i\lambda \cdot \tau} = \epsilon'$.

The extra Higgs superfield has introduced another parameter; $\tan \beta = \langle H_2^0 \rangle / \langle H_1^0 \rangle$. The fact that this ratio can be large is seen as a possible explanation for the large value of the top quark mass.

In addition to the Yukawa terms, we will require some Higgs self-couplings in order to generate the correct electro-weak breaking;
$$W_{\text{Higgs}} = \mu H_1^T \epsilon' H_2, \tag{27}$$
where μ has dimensions of mass. In order to get the correct scale for breaking (*i.e.* the correct values of M_W and M_Z) it turns out that we need to have $\mu = \mathcal{O}(M_W)$. The fact that we have an additional dimensional parameter (besides the scale of SUSY breaking) that needs to be set by hand, is known as the 'μ-*problem*'.

3.2 Soft SUSY breaking

If supersymmetry is the correct extension of the Standard Model, then it is expected to be softly broken by dimensionful parameters (such as squark mass terms), in order for the broken supersymmetry to still remove quadratic divergences and provide a solution to the problem of maintaining a hierarchy between the electroweak scale and the Planck scale.

The physics of the breaking is not yet well understood and there are several mechanisms that can achieve it. To take account of our ignorance one can write a general, gauge invariant, Lorentz invariant, effective Lagrangian;

$$\mathcal{L} = \mathcal{L}_{\text{SUSY}} + \mathcal{L}_{\text{soft}} \tag{28}$$

where the first term preserves supersymmetry invariance and the second violates supersymmetry and is given by (dropping the ϵ')

$$
\begin{aligned}
-\mathcal{L}_{\text{soft}} = \ & \tfrac{1}{2}\left(M_3\tilde{g}\tilde{g} + M_2\widetilde{W}\widetilde{W} + M_1\tilde{B}\tilde{B} + \text{h.c.}\right) \\
& + M_{\tilde{Q}}^2\tilde{Q}^*\tilde{Q} + M_{\tilde{U}}^2\tilde{U}^{c^*}\tilde{U}^c + M_{\tilde{D}}^2\widetilde{D}^{c^*}\widetilde{D}^c + M_{\tilde{L}}^2\tilde{L}^*\tilde{L} + M_{\tilde{E}}^2\tilde{E}^{c^*}\tilde{E}^c \\
& + A_U\lambda_U\tilde{U}^c\tilde{Q}H_2 + A_D\lambda_D\widetilde{D}^c\tilde{Q}H_1 + A_E\lambda_E\tilde{E}^c\tilde{L}H_1 + \text{h.c.} \\
& + m_{H_2}^2 H_2^* H_2 + m_{H_1}^2 H_1^* H_1 + (BH_2H_1 + \text{h.c.})
\end{aligned} \tag{29}
$$

where all generation indices have been suppressed. The most general softly broken MSSM has a total of a 124 physical parameters. The count goes as follows; 6 of the parameters come from gaugino mass terms of the form $M_i = |M_i|\, e^{i\phi_i}$. The squark and slepton masses are in principle 3×3 Hermitian matrices with complex matrix elements, contributing $5\times6\times2 = 60$ parameters. Trilinear couplings between the sfermions and Higgs bosons are arbitrary 3×3 complex matrices which constitute $2\times9\times2 = 36$ parameters. Additional parameters arise from the μ term, resulting in 33 mass eigenstates, 43 phases and the CKM angle.

3.3 Flavour and CP problems: mass insertion approximation

Given the most general form of supersymmetry breaking in (29) we see immediately that there are likely to be phenomenological problems with both flavour changing and CP violating processes. For example, using the new supersymmetric contributions to the Lagrangian in equations 21 and 22, we can construct a diagram (Figure 2) with an internal wino and squark—one of many possible supersymmetric contributions to $b \to s\gamma$.
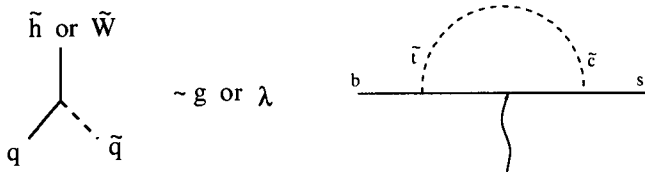
Figure 2. *A supersymmetric contribution to $b \to s\gamma$.*

The diagram includes a flavour changing mass insertion on the squark line and our lack of knowledge about supersymmetry breaking means that in principle this insertion could be as large as any of the other terms. Likewise, the supersymmetry breaking terms can have arbitrary CP violating phases and it is not difficult to construct diagrams that contribute to electric dipole moments (of which more later).

The appropriate way to treat these problem generally is to diagonalise all the mass matrices and calculate the diagrams with the mass-eigenstates. However there is a useful approximation, the mass-insertion approximation, that can be used when the supersymmetry breaking is approximately generation independent. This type of situation is usually motivated by particular kinds of supersymmetry breaking mediation, for example mediation by gravity resulting in mSUGRA models. In such models the canonical assumption is that the supersymmetry breaking terms are flavour universal at the GUT scale. Thus the most commonly considered scenario has

$$
\begin{aligned}
m^2_{\tilde{q}_{ij}} &= \delta_{ij} m^2_0 \\
M_A &= M_{1/2} \\
A_{U_{ij}} = A_{D_{ij}} = A_{E_{ij}} &= A \delta_{ij} .
\end{aligned}
\tag{30}
$$

Together with the parameters μ and B we then have 5 independent parameters. $|\mu|^2$ can be determined by electroweak symmetry breaking, and traditionally we trade μB for $\tan \beta$, the ratio of Higgs VEVs, leaving us with 4 independent parameters, $m_0, M_{1/2}, A, \tan \beta$. In addition there are 2 physical phases $\phi_A = \arg(A M^*_{1/2})$ and $\phi_\mu = \arg(\mu M^*_{1/2})$. All other phases can be rotated away. Typically these phases must be small in order to satisfy all electric dipole moment (EDM) requirements (including that of the mercury EDM [6]). This model, sometimes referred to as the Constrained MSSM (CMSSM), is an example of a Minimal Flavour Violating model (MFV model), defined as models where all flavour and CP violation is attributed to the Yukawa couplings.

Typically, running the renormalisation group equation from the GUT to the electroweak scales, will result in flavour non-universality thanks to the contribution from the Yukawa couplings, however they will generally be small. (Note that splittings between the squared masses of left and right-handed scalars, and also between squarks and sleptons will generally be large because of their different gauge couplings.) Thus, when considering a particular process, it is useful to consider a situation where the squark masses matrices are flavour universal and real, and to find the experimental limits on perturbations. This approach, pioneered by Gabbiani *et al* in [3], allows one to state the flavour and CP problems more quantitatively since it leads to model independent bounds on mass insertions. One begins by defining perturbations from universal scalar masses, $m^2_{\tilde{f}}$, as follows:

$$
\delta^q_{LR_{ij}} = \frac{(m^2_{\tilde{q}})_{LR_{ij}}}{m^2_{\tilde{f}}}, \qquad \delta^q_{LL_{ij}} = \frac{(m^2_{\tilde{q}})_{LL_{ij}}}{m^2_{\tilde{f}}}, \qquad \delta^q_{RR_{ij}} = \frac{(m^2_{\tilde{q}})_{RR_{ij}}}{m^2_{\tilde{f}}}.
\tag{31}
$$

It is then fairly straightforward to calculate the leading contributions to various processes by considering the diagrams with a single insertion. The approximation is represented pictorially in Figure 3. The fully diagonalised quark propagator would have diagonalisation matrices for the masses, V_{ia} where i labels flavour and a labels one the 6 squark mass eigenstates, on either end. However, this is equivalent to treating the flavour changing

$$V_{ia} \mathrel{-\!-\!-\!-\!-\!-\!-\!-\!-} V^+_{aj} \quad = \quad \mathrel{-\!-\!-\!-\!-\!-\!-\!-\!-} \quad + \quad \mathrel{-\!-\!} \overset{\times}{\mathrel{-\!-\!-}} \quad + \ldots$$
$$\tilde{q}_a \qquad\qquad\qquad\qquad \tilde{q}_i \qquad\qquad \tilde{q}_i \quad \tilde{q}_j$$

Figure 3. *The insertion approximation for the quark propagator.*

as an interaction vertex (denoted X). The experimental bounds can then be expressed as limits on the δ's, given in Table 1 and Table 2. For $b \to s\gamma$ decays with $m_{\tilde{f}}{=}500\mathrm{GeV}$ the main contributions are give in Figure 4 and the limits in Table 3. CP violating processes constrain imaginary insertions. e.g. for $M_3 = m_{\tilde{f}}{=}500\mathrm{GeV}$ we have the limits in Table 4.

Limits on-	$\sqrt{\|Re(\delta^q_{LL_{ij}})^2\|}$	$\sqrt{\|Re(\delta^q_{LR_{ij}})^2\|}$	$\sqrt{\|Re(\delta^q_{LL_{ij}})Re(\delta^q_{RR_{ij}})\|}$
$ij = sd$	0.04	0.004	0.003
$ij = bd$	0.1	0.03	0.02
$ij = cu$	0.1	0.03	0.02

Table 1. *Limits from $\Delta F = 2$ with $M_3(M_W) = m_{\tilde{f}}{=}500GeV$.*

Limits on-	$\|\delta^e_{LL_{ij}}\|$	$\|\delta^e_{LR_{ij}}\|$
$ij = \mu e$	0.008	2.10^{-6}
$ij = \tau e$	29	0.1
$ij = \tau \mu$	5	0.02

Table 2. *Limits from $l_i \to l_j\gamma$ decays with $M_1 = m_{\tilde{f}}{=}100GeV$.*
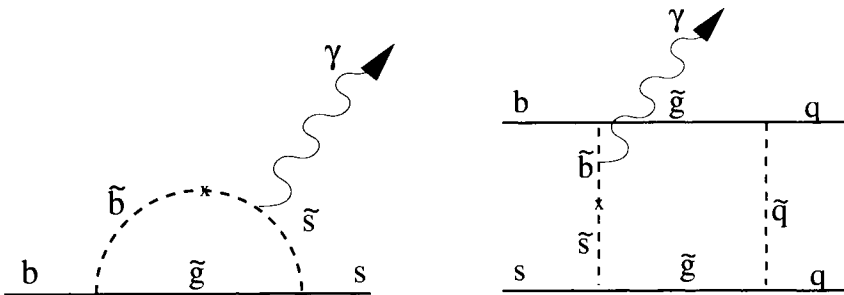
Figure 4. *Main contributions to $b \to s\gamma$ decays with $m_{\tilde{f}}{=}500GeV$.*

$x = M_3^2/m_{\tilde{f}}^2$	$|\delta^d_{LL_{23}}|$	$|\delta^d_{LR_{23}}|$
0.3	4.4	0.013
1	8	0.016
4	26	0.03

Table 3. *Limits from $b \to s\gamma$ decays with $m_{\tilde{f}}$=500GeV.*

ε	$\sqrt{|Im(\delta^d_{LL_{12}})^2|} < 0.002$	$\sqrt{|Im(\delta^d_{LR_{12}})^2} < 0.0006$
ε'/ε	$|Im(\delta^d_{LL_{12}})| < 0.5$	$|Im(\delta^d_{LR_{12}})| < 2.10^{-5}$
$d_n < 6.10^{-26}$	$|Im(\delta^d_{LR_{11}})| < 1.10^{-6}$	$|Im(\delta^u_{LR_{11}})| < 3.10^{-6}$

Table 4. *CP violating processes constrain imaginary insertions. Limits are for $M_3 = m_{\tilde{f}}$=500GeV.*

Satisfying experimental bounds/results on flavour changing and CP violating processes obviously prohibits significant perturbations from universality for many of these parameters (although other reasons besides universality have been considered in the literature, and will be mentioned later). This apparent fine tuning is the quantitative form of the flavour and CP problems.

Incidentally, in the case of the mSUGRA, all of these constraints are satisfied. Note that, had this not been the case (i.e. if running the RGE's in mSUGRA had resulted in supersymmetry breaking terms that violated these bounds) there would have been very little hope of ever explaining electroweak phenomena (e.g. symmetry breaking and masses) in terms of Planck or GUT scale physics.

4   The mass-spectrum and generic SUSY breaking

The mass insertion approximation outlined in the previous section is useful for addressing the question: how far can we deviate from universality and still satisfy experimental bounds? However, there may be (and often are) other wildly different patterns of soft-supersymmetry breaking which can have a significant degree of non-universality and yet have cancellation of contributions to various processes. (One example is models with Hermitian flavour structure, in which the EDMs are suppressed).

In order to calculate in more general scenarios, it is necessary to diagonalise all mass matrices and find the physical mass-eigenstates. The particle content and diagonalisations proceed in the following manner. First we collect the components with the same e.m. charges together into squarks, charginos, neutralinos:

$$
\begin{aligned}
\tilde{q}_{L_{i=1..3}} \oplus \tilde{q}_{R_{i=1..3}} &\to \tilde{q}_{a=1..6} \\
\tilde{h}_2^+ \oplus \widetilde{W}^+ &\to \tilde{\chi}^+ \\
\tilde{h}_1^- \oplus \widetilde{W}^- &\to \tilde{\chi}^- \\
\tilde{h}_u^0 \oplus \tilde{h}_d^0 \oplus \widetilde{W}^0 \oplus \tilde{B} &\to \tilde{\chi}_{a=1..4}
\end{aligned}
\tag{32}
$$

The sfermion mass-squared terms are given by

$$L_{\tilde{f}-mass-squared} = (\tilde{f}_L^*, \tilde{f}_R^*)_i \, M_{\tilde{f}_{ij}}^2 \begin{pmatrix} \tilde{f}_L \\ \tilde{f}_R \end{pmatrix}_j \tag{33}$$

where the $M_{\tilde{f}_{ij}}^2$ are 6×6 matrices

$$M_{\tilde{f}}^2 = \begin{pmatrix} M_L^2 + m_f^2 + M_z^2(\frac{1}{2} - Q_f \sin^2 \theta_W) \cos 2\beta & m_f(A_f - \mu^* R_f) \\ \text{h.c.} & M_R^2 + m_f^2 + M_z^2 Q_f \sin^2 \theta_W \cos 2\beta \end{pmatrix} \tag{34}$$

with $R_f = \cot \beta$ (or $\tan \beta$) for $I_3 = 1/2$ (or $-1/2$) and $\tan \beta = \langle h_2^0 \rangle / \langle h_1^0 \rangle$. We diagonalise this mass matrix by making a unitary transformation on the squarks

$$M_{\tilde{d}_a}^2 \delta_{ab} = \left(V_{\tilde{d}}^\dagger M_{\tilde{d}}^2 V_{\tilde{d}} \right)_{ab} \tag{35}$$

The charginos have a mass term

$$L_{charginos} = \left(\widetilde{W}^+, \tilde{h}_d^+ \right)_i M_{C_{ij}} \begin{pmatrix} \widetilde{W}^- \\ \tilde{h}_u^- \end{pmatrix}_j \tag{36}$$

where

$$M_C = \begin{pmatrix} M_2 & \sqrt{2} M_W \sin \beta \\ \sqrt{2} M_W \cos \beta & \mu \end{pmatrix}, \tag{37}$$

and we require a bi-unitary transformation to diagonalise the mass matrix:

$$M_{\tilde{\chi}^\pm} = U_C^* M_C V_C^{-1} \tag{38}$$

For the neutralinos we have:

$$L_{neutralinos} = \left(\tilde{B}^0, \widetilde{W}^0, \tilde{h}_d^0, \tilde{h}_u^0 \right)_i M_{N_{ij}} \begin{pmatrix} \tilde{B}^0 \\ \widetilde{W}^0 \\ \tilde{h}_d^0 \\ \tilde{h}_u^0 \end{pmatrix}_j \tag{39}$$

where M_N is symmetric:

$$M_N = \begin{pmatrix} M_1 & 0 & -M_Z \sin \theta_W \cos \beta & M_Z \sin \theta_W \sin \beta \\ T & M_2 & M_Z \cos \theta_W \cos \beta & -M_Z \cos \theta_W \sin \beta \\ T & T & 0 & -\mu \\ T & T & T & 0 \end{pmatrix} \tag{40}$$

with

$$M_{\tilde{\chi}^0} = V_N^T M_N V_N \tag{41}$$

The quantities which then enter calculations of FCNC and CP violating processes are then of course the diagonalisation matrices, $V_{\tilde{q}}, V_C, U_C, V_N$ which appear on the vertices.

4.1 An example calculation and the SuperGIM mechanism

All calculations follow more or less the same format, so let us examine just one in detail, the supersymmetric contribution to the EDMs. EDMs are quite complicated beasts, and get contributions from electromagnetic, chromomagnetic as well as purely gluonic contributions shown in Figure 5. See references [4, 5, 6].
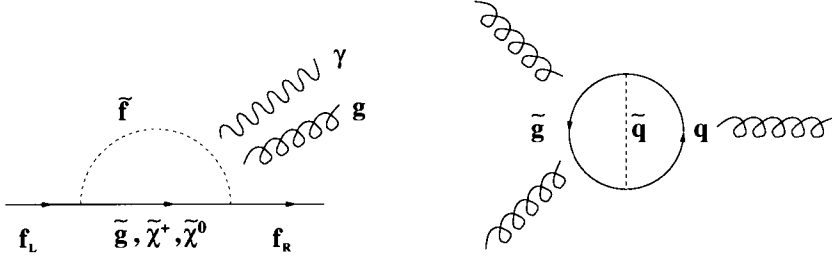
Figure 5. *Leading SUSY contributions to the EDMs. The photon and gluon lines are to be attached to the loop in all possible ways.*

For simplicity I will consider only the electromagnetic contributions to the fermion EDMs. These are given by [5]

$$
\frac{d_q^{\tilde{g}\,(E)}}{e} = \frac{-2\alpha_s}{3\pi} \sum_{k=1}^{6} \mathrm{Im}(\Gamma_q^{1k}) \frac{m_{\tilde{g}}}{M_{q_k}^2} Q_{\tilde{q}}\, \mathrm{B}\!\left(\frac{m_{\tilde{g}}^2}{M_{q_k}^2}\right),
$$

$$
\frac{d_u^{\chi^+\,(E)}}{e} = \frac{-\alpha_{\mathrm{em}}}{4\pi \sin^2\theta_W} \sum_{k=1}^{6}\sum_{i=1}^{2} \mathrm{Im}(\Gamma_{uik}) \frac{m_{\chi_i^+}}{M_{d_k}^2}\left[Q_{\tilde{d}}\, \mathrm{B}\!\left(\frac{m_{\chi_i^+}^2}{M_{d_k}^2}\right) + (Q_u - Q_{\tilde{d}})\, \mathrm{A}\!\left(\frac{m_{\chi_i^+}^2}{M_{d_k}^2}\right)\right],
$$

$$
\frac{d_d^{\chi^+\,(E)}}{e} = \frac{-\alpha_{\mathrm{em}}}{4\pi \sin^2\theta_W} \sum_{k=1}^{6}\sum_{i=1}^{2} \mathrm{Im}(\Gamma_{dik}) \frac{m_{\chi_i^+}}{M_{u_k}^2}\left[Q_{\tilde{u}}\, \mathrm{B}\!\left(\frac{m_{\chi_i^+}^2}{M_{u_k}^2}\right) + (Q_d - Q_{\tilde{u}})\, \mathrm{A}\!\left(\frac{m_{\chi_i^+}^2}{M_{u_k}^2}\right)\right],
$$

$$
\frac{d_e^{\chi^+\,(E)}}{e} = \frac{\alpha_{\mathrm{em}}}{4\pi \sin^2\theta_W} \sum_{i=1}^{2} \frac{m_{\chi_i^+}}{m_{\tilde{\nu}}^2} \mathrm{Im}(\Gamma_{ei})\, \mathrm{A}\!\left(\frac{m_{\chi_i^+}^2}{m_{\tilde{\nu}}^2}\right),
$$

$$
\frac{d_f^{\chi^0\,(E)}}{e} = \frac{\alpha_{\mathrm{em}}}{4\pi \sin^2\theta_W} \sum_{k=1}^{6}\sum_{i=1}^{4} \mathrm{Im}(\eta_{fik}) \frac{m_{\chi_i^0}}{M_{\tilde{f}_k}^2} Q_{\tilde{f}}\, \mathrm{B}\!\left(\frac{m_{\chi_i^0}^2}{M_{\tilde{f}_k}^2}\right), \tag{42}
$$

where

$$
\Gamma_q^{1k} = e^{-i\phi_3} V_{\tilde{q}4k} V_{\tilde{q}1k}^*, \tag{43}
$$

with ϕ_3 being the gluino phase. The chargino factor Γ_{fik} comes from products of vertices with attached charged Higgsino or wino:

$$
\Gamma_{uik} = \lambda_u V_{Ci2}^* V_{\tilde{d}1k}^* (U_{Ci1}^* V_{\tilde{d}1k}^* - \lambda_d U_{Ci2}^* V_{\tilde{d}4k}^*),
$$

$$
\Gamma_{dik} = \lambda_d U_{Ci2}^* V_{\tilde{u}1k}^* (V_{Ci1}^* V_{\tilde{u}1k}^* - \lambda_u V_{Ci2}^* V_{\tilde{u}4k}^*) \tag{44}
$$

and analogously for the electron. The λ_f are the Yukawa couplings

$$
\lambda_u = \frac{m_u}{\sqrt{2}m_W \sin\beta}, \quad \lambda_{d,e} = \frac{m_{d,e}}{\sqrt{2}m_W \cos\beta}. \tag{45}
$$

The neutralino vertex η_{fik} is given by

$$
\eta_{fik} = \left[-\sqrt{2} \{ \tan\theta_W (Q_f - I_{3_f}) V_{N1i} + I_{3_f} V_{N2i} \} V^*_{\tilde{f}1k} - \lambda_f V_{Nbi} V^*_{\tilde{f}4k} \right]
$$
$$
\times \left[\sqrt{2} \tan\theta_W Q_f V_{N1i} V_{\tilde{f}4k} - \lambda_f V_{Nbi} V_{\tilde{f}1k} \right] , \tag{46}
$$

where I_3 is the third component of the isospin, $b = 3$ (or 4) for $I_3 = -1/2$ (or 1/2). The loop functions $A(r), B(r)$, and $C(r)$ are defined by

$$
A(r) = \frac{1}{2(1-r)^2} \left(3 - r + \frac{2\ln r}{1-r} \right) ,
$$
$$
B(r) = \frac{1}{2(r-1)^2} \left(1 + r + \frac{2r\ln r}{1-r} \right) ,
$$
$$
C(r) = \frac{1}{6(r-1)^2} \left(10r - 26 + \frac{2r\ln r}{1-r} - \frac{18\ln r}{1-r} \right) . \tag{47}
$$

Note that the indices '1' and '4' appear in the above because for the EDMs of the down, up and electron we pick out the left and right chirality fermions of the first generation.

This completes the electromagnetic contribution to the quark and electron EDMs. We can now see in a more general way why flavour universality gives cancellation of contributions. Consider for example the gluino contribution to EDMs. In the event that the squark masses are universal and equal to $M_{\tilde{q}}^2$ this can be written as

$$
\frac{d_q^{\tilde{g}\,(E)}}{e} = \frac{-2\alpha_s}{3\pi} \frac{m_{\tilde{g}}}{M_{\tilde{q}}^2} Q_{\tilde{q}} \, \mathrm{B}\left(\frac{m_{\tilde{g}}^2}{M_{\tilde{q}}^2} \right) \sum_{k=1}^{6} \mathrm{Im}(\Gamma_q^{1k}) . \tag{48}
$$

The last factor in this expression is

$$
\mathrm{Im}(\Gamma_q^{1k}) = e^{-i\phi_3} \left(V_{\tilde{q}} V_{\tilde{q}}^{\dagger} \right)_{41} = 0 , \tag{49}
$$

and consequently the gluino contribution exactly vanishes because of the unitarity of the diagonalisation matrices, $V_{\tilde{q}}$. With exact universality the leading contribution is from the chargino diagram, however these are proportional to $\mathrm{Im} \sum_{k=1}^{6} (\Gamma_{dik})$. Again unitarity means that the main contribution here is from the phase on the μ term (unfortunately the least well understood of all the parameters in the MSSM). Note that even in the CMSSM, unitarity is broken between left and right chiralities at the weak scale because of the different gauge coupling contributions to the mass-squared RGEs. Because of this there is also a small dependence on the phase of the A-term (ϕ_A) even in the CMSSM.

A similar sort of cancellation occurs for flavour changing processes, where it is referred to as the *superGIM* mechanism. For example consider the chargino contributions to ΔM_K shown in Figure 6.

The $K_S - K_L$ mass difference is given by

$$
\Delta M_K = 2\mathrm{Re}(M_{12}) \tag{50}
$$

where

$$
M_{12} = \frac{\alpha_2^2}{24} B_K \eta_K f_K^2 \sum_{kk'}^{6} \sum_{ij}^{2} \Lambda_{2ik} \Lambda^*_{1ik'} \Lambda_{2jk'} \Lambda^*_{1jk} \frac{D(M_{\tilde{u}_k}, M_{\tilde{u}_{k'}}, M_{\tilde{\chi}_i}^2, M_{\tilde{\chi}_j}^2)}{M_{\tilde{\chi}_i}^2} , \tag{51}
$$
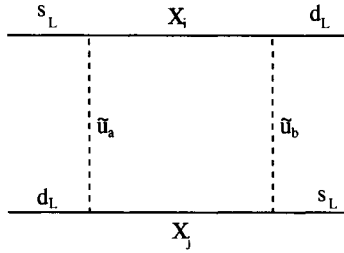
Figure 6. *Chargino contribution to ΔM_K.*

and

$$\Lambda_{fik} = V_{Ci1}^* V_{\tilde{u}fk}^* - \sum_{f'=1}^{3} \lambda_{U_{ff'}} V_{Ci2} V_{\tilde{u}(f'+3)k}^* \tag{52}$$

and where f, f' label flavours, and D is a dimensionless one-loop integral which is of order one if the chargino and squark masses are similar. Here the first term in Λ comes from the $\tilde{u}d\widetilde{W}$ term and the second from the $\tilde{u}d\tilde{h}$ term in the Lagrangian.

Again we see that degenerate squark masses leads to suppression because of the universality of the diagonalisation matrices. Indeed because only left handed external quarks are involved in this particular diagram, the contribution to ΔM_K from the first term in the Λs is proportional to the mass splitting in the first two generations of left handed up-squarks which in the CMSSM is driven by the renormalisation group effects of the Yukawas:

$$\Delta M_K \propto \frac{m_c^2}{m_u^2}. \tag{53}$$

In addition to this contribution there are contributions involving cross-terms in the Λs, but because we need two cross terms (there are 4 Λs) they will be of the same order of magnitude. This overall suppression is the super-GIM mechanism.

5   Aspects of CMSSM phenomenology

The universality of supersymmetry breaking is found not to hold in most string derived models and in general supergravity models, and we expect at least some non-universality in the flavour structure of the A terms. However because of the super-GIM mechanism, the CMSSM is the minimal model from the point of view of avoiding nasty FCNC and CP violating contributions. It is important therefore to estimate what the new supersymmetric contributions to these processes are likely to be in the CMSSM, in order to provide a kind of lower bound on the expectations for new physics. These effects come from the running of the renormalisation group equations, whence the flavour dependence of the Yukawas enters the supersymmetry breaking. These effects are small and so the mass insertion approximation is valid. However, there is no general way to estimate the resulting size of the δs induced by the renormalisation group, because the RGEs are non-linear and in any case some of the parameters depend on the minimisation of the one-loop electroweak potential. Thus a useful approach is to concentrate on the dominant contributions in the CMSSM (and indeed Minimal Flavour models). As we have

seen, contributions from diagrams involving gluinos are suppressed in minimal flavour models (this is a general result) and the neutralino contributions are also suppressed in the CMSSM. Hence the dominant contributions are those involving charginos and charged Higgs particles. This greatly reduces the number of parameters and means that one can survey the parameter space rather generally without having to perform the minimisation of the one-loop potential.

A study along these lines was carried out in [7] and recently improved and updated in [8] where many more rare decays were also considered. (This reference provides a useful compendium of expressions for calculating various rare processes in SUSY, and summarises the new operators that are introduced into the basis for H_{eff}. Many of the expressions for the supersymmetric contributions are also found in [9].) Here the discussion will concentrate on the supersymmetric contributions to the Wilson coefficients calculated at $\sim M_W$, so the reader is referred to that reference for more information on how to turn these into the measured parameters.) The later study includes NLO-QCD and QED corrections as well as constraints from electroweak precision studies, exclusion bounds on M_h, and the $B \to X_s \gamma$ decay rate. The input parameters required for such a study are only those that effect the chargino and Higgs masses and mixings:

$$m_{\tilde{t}_1}, \; m_{\tilde{t}_2}, \; \theta_{\tilde{t}}, \; m_{\tilde{\chi}_1}, \; \tan\beta, \; m_{H^\pm}, \; \mu, \; m_{\tilde{l}} \tag{54}$$

where $m_{\tilde{t}_{1,2}}$ are the heavy and light stop masses, $\theta_{\tilde{t}}$ is the stop mixing angle, $m_{\tilde{\chi}_1}$ the lightest chargino mass, and $m_{\tilde{l}}$ the common slepton mass (for the rare decays). Note that there are more parameters here than in the CMSSM (including the breaking of electroweak symmetry), but in performing a study using these parameters one is addressing a wider class of models, including all MSSM models that have Minimal Flavour Violation (for example those in which the scalar masses are flavour independent but not necessarily degenerate). One then scans over reasonable values of these parameters. Again what constitutes 'reasonable' is open to interpretation. However, imposing masses less than 1TeV seems appropriate in order to avoid reintroducing fine tuning. The parameter $|\mu|$ can typically be larger than the range $|\mu| < 500\mathrm{GeV}$ chosen in this study, although larger values reduce the mixing in the chargino and therefore its contributions. In addition values of large $\tan\beta > 6$ are excluded in this study. Although the results tend to asymptote for medium $\tan\beta$ there is a region of very high $\tan\beta$ which is interesting from a theoretical point of view; it allows Yukawa unification at the GUT scale. Unfortunately this implies large b-quark Yukawas and hence the top-quark dominance approximation made for the low-medium $\tan\beta$ studies is no longer valid.

Before summarising the results, let us briefly discuss the main effects of supersymmetry. For the kaons, the most interesting effect is that one tends to find a depletion of ε'/ε, unfortunately contrasting with what we would perhaps like given the fact that the measured value is slightly larger than one might expect from the Standard Model (modulo uncertainties in hadronic matrix elements). The dominant contributions are from Z^0 and photo-penguin diagrams involving charginos and charged Higgs. The two contributions come with opposite sign (the chargino being positive), so that there are indeed regions where an enhancement over the SM value can occur, although this is only of the order of 10%. On the other hand the depletion can be up to 40%. A summary of the most interesting processes is given in Table 5. Again, the result for ε'/ε is plagued by hadronic uncertainties, whereas the rare decays are theoretically much cleaner places to search for

Process \Ratio w.r.t SM	max	min
ε'/ε	0.53	1.07
$K^+ \to \pi^+\nu\bar{\nu}$	0.65	1.02
$K_L \to \pi^0\nu\bar{\nu}$	0.41	1.03
$K_L \to \pi^0 e^+ e^-$	0.73	1.10
$B \to X_s\mu\bar{\mu}$	0.73	1.34
$B \to \mu\bar{\mu}$	0.68	1.53

Table 5. *Ratios of CMSSM to SM results for various processes.*

signals of supersymmetry.

One additional interesting feature of Minimal Flavour models is that they predict an absolute lower bound of $\sin 2\beta > 0.42$ [9, 8]. Such a lower bound is possible because one can generalise the expression for the ε hyperbola to include supersymmetric contributions.

$$\bar{\eta}\left[(1 - \bar{\rho})A^2\eta_2 F_{tt} + 0.31\right]A^2\widehat{B}_K = 0.266 \tag{55}$$

where 0.31 is the charm contribution. From

$$R_t^2 = 1.6\frac{R_0^2}{A^2 F_{tt}} \tag{56}$$

we get

$$\sin 2\beta = \frac{2\bar{\eta}(1 - \bar{\rho})}{R_t^2} = \frac{1.6}{R_0^2\eta_2}\left[\frac{0.226}{A^2\widehat{B}_K} - 0.31\bar{\eta}(F_{tt})\right] \tag{57}$$

This function may now be plotted as a function of the parameter F_{tt} which is a collection of all SM and SUSY contributions and is typically of order 1 in the CMSSM. (Note that it is not monotonic because the variables are not independent of F_{tt}.) A value of $\sin 2\beta$ less than this value would have been the first indication of flavour structure in soft supersymmetry breaking and arguably an indirect detection of supersymmetry. Unfortunately, at the time of writing, the value of $\sin 2\beta$ coming from $B \to K_S\psi$ measurements at Belle and BaBar indicate higher values.

6   Beyond the CMSSM and MFV

Up to this point we have been looking at the minimal case where all flavour and CP violation in the MSSM is attributed to the Yukawa sector. However, an interesting question to ask is whether there might be other, non-universal scenarios, in which some of these processes are consistent with experimental limits. This is an important question because string models do not generally predict the favourable CMSSM scenario very naturally.

There are a number of ideas for suppressing FCNC and CP processes apart from universality. They are: cancellations of SUSY contributions amongst themselves, flavour alignment, particular flavour structures (e.g. Hermitian or off-diagonal phases), decoupling and dynamical (i.e. derived from minimising a string derived potential). In the

next subsection I will discuss cancellations with respect to EDMs and make some general comments. Then I will mention decoupling and in the following section discuss some recent developments in the last of these, FCNC and CP suppression in dynamical string derived patterns of supersymmetry breaking.

6.1 Non-universality and cancellations

When the flavour structure is non-universal at the GUT scale there can be cancellations to various processes. Intuitively it seems unlikely that they will occur simultaneously for different processes, and this indeed turns out to be the case when numerical studies are done, as we shall now see.

Non-universality has been studied as a general idea in [10] but more recently there has been a resurgence of interest thanks to 'new' string inspired scenarios, motivated by explicit constructions of SM-like models in for example type I models. These models can lead to flavour non-universality which is relatively 'well-controlled', so naturally there has been some speculation as to whether a particular type of flavour non-universality might lead to cancellations 'across the board' as effectively as universality [11, 12]. The answer to this, at least for those cases that have been studied, unfortunately appears to be negative as we shall now see.

The measurables that first received attention in this context were again the EDM's of the neutron and electron, both of which have rather strict experimental bounds. In [11] it was noted that for particular patterns of supersymmetry breaking there can be a simultaneous cancellation of these parameters, even if the CP violating phases on the μ term is large, much larger than would be allowed in the naive CMSSM with phases. This idea has since attracted some attention and has become known as the 'cancellation mechanism'.

The analysis was based on supersymmetry breaking pattern inspired by so-called D-brane models in which the $SU(2)$ gauge group was associated with a different brane from the $SU(3)$ and $U(1)_Y$ groups of the SM. The result is that the supersymmetry breaking gauginos masses can have different values at the string scale (assumed to be near the GUT scale). On the other hand the A-terms can have a good deal of universality. The overall pattern of supersymmetry breaking can therefore be assumed to be identical to that of the CMSSM with the exception of the gaugino masses which are of the form:

$$
\begin{aligned}
M_1 = M_3 &= \sqrt{3} m_{3/2} (\cos\theta) \Theta_1 e^{i\gamma_1} \\
M_2 &= \sqrt{3} m_{3/2} (\cos\theta) \Theta_2 e^{i\gamma_2} \,,
\end{aligned}
\tag{58}
$$

where $m_{3/2} \sim M_W$ is the gravitino mass and the angles here are a convenient parameterisation of the distribution of supersymmetry breaking amongst the different fields. If we in addition allow an arbitrary phase, ϕ_μ, for the μ term, then one finds large values of phases where the neutron and electron EDMs cancel simultaneously as shown in Figure 7(a).

Unfortunately, when additional experiments are considered, the cancellation region begins to be excluded. In this case the experiment in question is the bound on the mercury EDM [6]. This parameter also has a cancellation region, but unfortunately it tends to cut across the previous cancellation region, and all three EDMs are only consistent with bounds in the trivial small phase region [6].
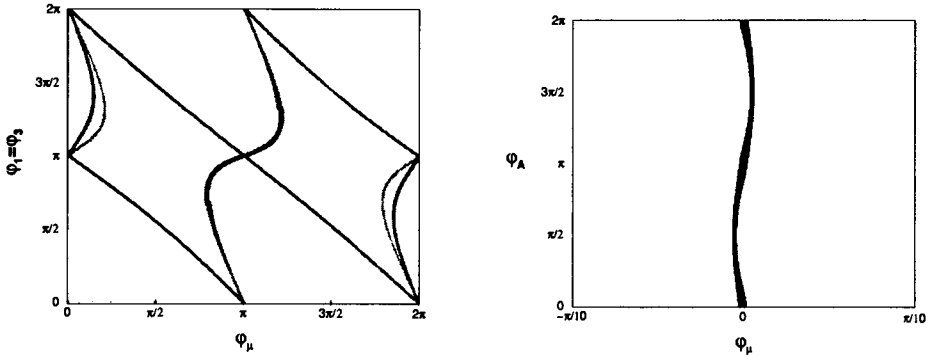
Figure 7. *The cancellation regions for EDMS of the electron (dark vertical curves), neutron (lighter vertical curves overlapping those of the electron) and mercury atom (diagonal lines). The parameters are* $\tan\beta = 2$, *of simultaneous n,e,Hg EDM cancellation in mSUGRA for the same* $\tan\beta$ *and* $m_{3/2}$.

Note that in contrast, mSUGRA (CMSSM with phases) has a region of cancellation for all three EDMs as shown in Figure 7b which allows any value of ϕ_A but restricts $\phi_\mu < \pi/100$. (The weak constraint on ϕ_A is a result of the small contribution of ϕ_A to the EDMs, as explained earlier). Adding a small phase to any of the gauginos separates the regions of small n, e and Hg EDMs and destroys this cancellation. In both cases the cancellation regions decrease in size at more realistic, larger, $\tan\beta$. Thus for this example the cancellations are not 'across the board'. Indeed, given the fact that the cancellation regions shift considerably depending on the adopted model of the neutron (valence-quark or parton), finding a (non-symmetry based) stringy underlying reason for cancellation would in any case be difficult. Although here I have concentrated on EDMs I see no reason why the conclusion would be different with any other CP or FCNC process unless there was some underlying flavour symmetry.

6.2   Decoupling

This possibility is based on the fact that as the supersymmetric particles become heavier their contribution to FCNC and CP processes begins to decouple. For example, even if one allows $\mathcal{O}(1)$ CP violating phases, the contribution to the EDMs will be negligible if the SUSY spectrum is sufficiently heavy [13]. Generally, SUSY fermions are required to be lighter than the SUSY scalars by, for example, cosmological considerations. So the decoupling scenario can be implemented with heavy sfermions only. Note that for FCNCs for example, the contribution is still proportional to

$$\frac{m_{\tilde{i}}^2 - m_{\tilde{j}}^2}{m_{\tilde{i}}^2 + m_{\tilde{j}}^2}, \tag{59}$$

where i and j label different flavours. Thus if, for example, we put down the suppression of $\Delta M_{K_{\text{SUSY}}}$ to large first and second generation masses, we still require a proportionate universality in their values. So this scenario is in a sense 'universality where we need it'.

Here the SUSY contributions to the EDMs are suppressed even with maximal SUSY phases because the squarks in the loop are heavy and the mixing angles are small. In Figure 8 we display the EDMs as functions of the universal scalar mass parameter m_0 for the mSUGRA model with maximal CP-phases $\phi_\mu = \phi_A = \pi/2$ and $m_{1/2} = A = 200 \text{GeV}$. We observe that all EDM constraints except for that of the electron require m_0 to be around 5TeV or more. The mercury constraint is the strongest one and requires

$$(m_0)_{\text{decoupl.}} \simeq 10 \text{ TeV} . \qquad (60)$$

This leads to a serious fine-tuning problem. Recall that one of the primary motivations for supersymmetry was a solution to the naturalness problem. Certainly this motivation will be entirely lost if a SUSY model reintroduces the same problem in a different sector, i.e. for example a large hierarchy between the scalar mass and the electroweak scale.
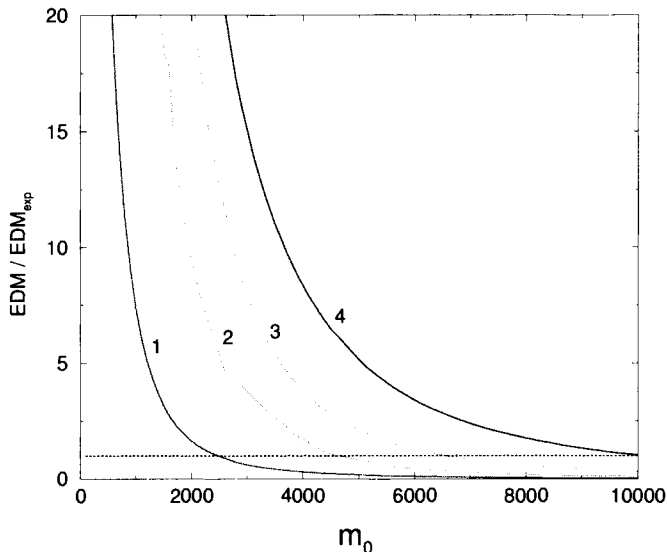
Figure 8. *EDMs as a function of the universal mass parameter m_0. 1: electron. 2: neutron (chiral model). 3: neutron (parton model). 4: mercury. The experimental limit is given by the horizontal line. Here $\tan \beta = 3$, $m_{1/2} = A = 200 GeV$ and $\phi_\mu = \phi_A = \pi/2$.*

The degree of fine-tuning can be quantified as follows. The Z boson mass is determined at tree level by

$$\frac{1}{2} m_Z^2 = \frac{m_{H_1}^2 - m_{H_2}^2 \tan^2 \beta}{\tan^2 \beta - 1} - \mu^2 . \qquad (61)$$

One can define the sensitivity coefficients [14],[15]

$$c_i \equiv \left| \frac{\partial \ln m_Z^2}{\partial \ln a_i^2} \right| , \qquad (62)$$

where a_i are the high energy scale input parameters such as $m_{1/2}$, m_0, etc. Note that μ is treated here as an independent input parameter. A value of c_i much greater than 1 would

indicate a large degree of fine-tuning. The Higgs mass parameters are quite sensitive to m_0 so, for $m_0=$10TeV, we find

$$c_{m_0} \sim 5000 \qquad (63)$$

for the parameters of Figure 8. For the universal scalar mass of 5 (or 3) TeV the sensitivity coefficient reduces to 1300 (or 500). This clearly indicates an unacceptable degree of fine-tuning.

This problem can be mitigated in models with focus point supersymmetry, that is when m_{H_2} is insensitive to m_0 [15]. However, this mechanism works for m_0 no greater than 2–3TeV which is not sufficient to suppress the EDMs. Another interesting possibility is presented by models with a radiatively driven inverted mass hierarchy, i.e. the models in which a large hierarchy between the Higgs and the first two generations scalar masses is created radiatively [16]. However, a successful implementation of this idea is far from trivial [17]. One can also break the scalar mass universality at the high energy scale [18]. In this case, either a mass hierarchy appears already in the soft breaking terms or certain relations among the soft parameters must be imposed (for a review see [19]). These significant complications disfavour the decoupling scenario as a way to solve the SUSY CP problem, yet it remains a possibility.

Concerning the other phenomenological consequences, we remark that the SUSY contributions to the CP-observables involving the first two generations (such as $\varepsilon, \varepsilon'$) are negligible, although those involving the third generation may be considerable. The corresponding CP-phases are constrained through the Weinberg operator contribution to the neutron EDM, which typically prohibits the maximal phase $\phi_{A_{t,b}}\big|_{\text{GUT}} = \pi/2$ while still allowing for smaller $\mathcal{O}(1)$ phases.

6.3   New contributions to FCNC processes

As well as the question of how SUSY contribution can be generally suppressed with non-universality, some attention has been paid to the question of whether non-universality might be able to explain excesses of, for example, ε'/ε. One of the reasons for the emphasis on non-universality is that there are rather general arguments that show that, in the absence of a CKM phase, there are no new contributions to ε, ε'/ε or hadronic B^0 asymmetries even with arbitrary soft phases. Thus supersymmetric

CP violation is intimately connected with flavour non-universality. CP in the presence of new flavour structures was reviewed recently in [20], mainly concentrating on enhancements to ε'/ε. Contributions come from the chargino diagrams in Figure 9 as well as the usual gluino diagrams, and these were recently analysed in [21]. Thus flavour and CP constrain low energy supersymmetry but at the same time provide interesting new areas to search for supersymmetry indirectly. Generally we do expect new CP violating phases in the SUSY sector; however the conclusion of these studies is that they will not produce sizable effects in ε and ε'/ε unless there is also some flavour structure in the SUSY breaking.
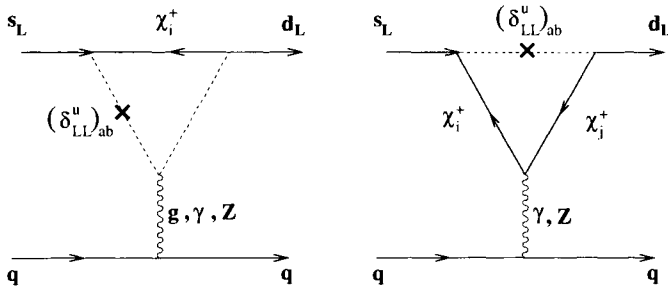
Figure 9. *Chargino contributions to ε'/ε.*

7 The future

Supersymmetry breaking is clearly the biggest factor in flavour and CP phenomenology and yet remains the biggest unknown. Ideally one would like to have a picture of CP and flavour from first principles but at the time of writing this seems some way off. Nevertheless there are two hints to do with CP violation that some progress might be possible. The first hint is that in the SM the CP violation is insufficient to account for the observed baryon asymmetry. Secondly, general arguments show that the four dimensional CP is a discrete gauge symmetry in string theory. Thus any viable string scenario must break CP spontaneously. (See [23] for the original observation, and some more recent references on the general connection between supersymmetry breaking and CP.)

If the CP violation responsible for baryogenesis is associated with the supersymmetry breaking sector then at some point in the evolution of the Universe these phases must have been quite large. On the other hand the SUSY CP problem seems to indicate that today they are small. An interesting idea that connects with the stringy picture of CP violation is therefore that the spontaneous breaking of CP happens at the same time as supersymmetry breaking, and that at the minimum of the potential the CP phases are zero. Thus at the time of supersymmetry breaking there was sufficient time dependent CP violation to account for baryogenesis, but today these phases have been set dynamically to zero. Some progress in this direction has been made in recent investigations into the nature of CP in string theory and the precise way in which it can be violated, either by string theory moduli (describing the compactification) or by the Froggatt–Nielsen fields which are responsible for flavour structure in effective string models [24, 25]. There it was found quite generally that it is quite difficult to generate a flavour and CP structure in the Yukawa couplings (by whatever means) without at the same time generating large EDMs [25]. This problem, which is at once more specific and more serious than the supersymmetric CP problem seems to indicate that there is more to be learnt by adopting more of a 'top-down' approach.

Eventually such approaches might shed some light on the origin of CP violation in string theory and even provide a direct connection between string theory and experiment. Certainly CP and flavour remain two of the most important aspects of supersymmetry phenomenology, and as we have seen may allow detection of supersymmetry indirectly before the LHC. Beyond that, should supersymmetry be found, experiments directed at flavour and CP violating phenomenology will help us to pin down the parameters of the

supersymmetry breaking sector and allow us to piece together the type of mechanism responsible for breaking and mediating supersymmetry, providing a window onto string theory.

In these lectures I have given a purely personal overview of the subject and have omitted many topics, my main aim being to equip the reader with the techniques rather than to be comprehensive. To finish, I will try to make some amends by summarising some of these developments and directing the reader to relevant works. Muon dipole moments have been discussed in [26] where it has been shown that a muon EDM may in the future be at the 10^{-22}e.cm level and hence measurable. The relations between dipole moment operators have been examined in [27] where it was shown that the muon anomalous magnetic moment is correlated with the electron EDMs. The question of $(g-2)_\mu$ in mSUGRA with and without phases and universality has been analysed in citearn1. In mSUGRA there are correlations between this quantity and $B_s \to \mu^- \mu^+$ (see Dedes *et al* in the same reference). Alternative approaches to spontaneous CP violation that concentrate instead on CP breaking in extensions of the MSSM with gauge singlet have been considered in [29]. Other possibilities for CP violation are reviewed in [22]. The neutrino sector (to which I have paid scandalously little attention) deserves a review of its own. As well as the direct interest in the flavour structure of the neutrino sector, various patterns of supersymmetry breaking in the neutrino sector may lead to observable effects in rare decays such as $\mu \to e\gamma$ and $\mu \to eee$, as well as interesting possibilities for CP and leptogenesis. The reader is referred to [30] and references therein. I have also been unable to cover R-parity violating models which obviously have interesting implication for CP and flavour. For a review see [33]. The masses of the Higgs sector are strongly dependent on the CP violating phase in the μ term (something that was somewhat skated over in the text). This is discussed in [32]. Finally there are a host of other interesting ideas and I cannot even begin to summarise them. One of many intriguing examples is the idea that all fermion masses are actually the result of radiative corrections and therefore soft. This is discussed in [34].

Acknowledgments

Appendix A: Conventions

We here briefly recall the essential features of the Weyl notation for the discussion of Section 1. In the Weyl representation, a four component, Dirac fermion looks like

$$\psi = \begin{pmatrix} \chi_\alpha \\ \overline{\chi}_c^\beta \end{pmatrix}.$$

where χ and χ_c are two-component spinors (α, $\beta = 1, 2$). The gamma matrices are

$$\gamma^\mu = i \begin{pmatrix} 0 & \sigma^\mu \\ \overline{\sigma}^\mu & 0 \end{pmatrix}, \tag{64}$$

where $\sigma^\mu = (1, \sigma^i)$, $\overline{\sigma}^\mu = (1, -\sigma^i)$, and σ^i are the Dirac matrices. The γ_5 matrix looks like

$$\gamma_5 = i \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{65}$$

so that the chiral projections (given by $R/L \equiv (1 \pm i\gamma_5)/2$) on the fermions are,

$$\psi_L = \begin{pmatrix} \chi \\ 0 \end{pmatrix} \qquad \psi_R = \begin{pmatrix} 0 \\ \overline{\chi}_c \end{pmatrix} \tag{66}$$

The suffix 'c' refers to charge conjugate, since

$$\psi_c = -\gamma^2 \psi^* = \begin{pmatrix} -i\sigma^2 \overline{\chi}_c^* \\ i\sigma^2 \overline{\chi} \end{pmatrix}, \tag{67}$$

so that $\chi_{c\alpha} = -i\sigma_{\alpha\beta}^2 \overline{\chi}_c^{*\beta}$. The matrix

$$\epsilon_{\alpha\beta} = -i\sigma_{\alpha\beta}^2 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \tag{68}$$

acts like a metric tensor for the unbarred spinors. For example the invariant mass terms are

$$\begin{aligned} \overline{\psi}\psi &= \overline{\chi}_c^{*\beta} \chi_\beta + \text{h.c.} \\ &= \epsilon^{\beta\alpha} \chi_{c\alpha} \chi_\beta + \text{h.c.} \\ &= (\psi_R^\dagger)^\alpha (\psi_{L\alpha}) + \text{h.c.} \end{aligned} \tag{69}$$

Because of this we identify the complex conjugate of a field, $\overline{\chi}^{*\beta}$ with χ^β. (The last line we included to make the point that we may identify the two component spinors in terms of the original Dirac ones, so that $(\psi_R^\dagger)^\alpha \equiv \psi_{Lc}^\alpha$. The dagger we shall reserve to indicate an operation on the Dirac spinors.) This means that $(\eta\chi)^\dagger = \overline{\chi}\overline{\eta}$. Similarly, the complex conjugates form their own representation of the Lorentz algebra, and have their

own Lorentz invariant quantities. To avoid confusion the indices of the latter are written with a dot, so that

$$\epsilon_{\dot\alpha\dot\beta} = -i\sigma^2_{\dot\alpha\dot\beta}, \tag{70}$$

and, in general, Lorentz invariant quantities are written as

$$\chi^\alpha \eta_\alpha, \quad \overline{\chi}_{\dot\alpha}\overline{\eta}^{\dot\alpha}, \quad \chi^\alpha \sigma^\mu_{\alpha\dot\alpha}\overline{\eta}^{\dot\alpha}\partial_\mu, \quad \overline{\chi}_{\dot\alpha}\sigma^{\mu\dot\alpha\alpha}\eta_\alpha\partial_\mu, \quad \dots \tag{71}$$

and so on.

Appendix B: Superspace

In the text we saw that the transformations of the $\{\phi, \chi, F\}$ supermultiplet, could be expressed as the behaviour of a single scalar quantity, the superfield. The eight 'coordinates' $(x, \theta, \overline{\theta})$ are known collectively as *superspace*. Supersymmetry transformations are equivalent to transformations in θ. Transformations in Minkowski space are generated by the momentum P_μ, which is realised over space-time coordinates as a differential operator $P_\mu \equiv -i\partial_\mu$, so that if $x \to x + a$, then $\exp(ia{\cdot}P)$ induces the transformation $\phi(x) \to \phi(x) + a^\mu\partial_\mu\phi(x)$. In fact P_μ is the conserved charge associated with the transformation $x \to x + a$, which, by Noether's theorem is

$$P_\mu = \int \mathrm{d}x^3 \frac{\partial\mathcal{L}}{\partial\dot{f}}\partial_\mu f - \delta^0_\mu\mathcal{L} \tag{72}$$

where f represents summation over all fields (fermions and bosons) in the Lagrangian. In a similar way, transformations in θ can be written in terms of differential operators acting in the superspace;

$$\delta_\xi = \xi\,Q + \overline{\xi}\,\overline{Q} \tag{73}$$

where

$$
\begin{aligned}
Q_\alpha &= \frac{\partial}{\partial\theta^\alpha} - i\sigma^\mu_{\alpha\dot\alpha}\overline{\theta}^{\dot\alpha}\partial_\mu \\
Q^{\dot\alpha} &= \frac{\partial}{\partial\theta_{\dot\alpha}} - i\theta^\alpha\sigma^{\mu\dot\alpha}_\alpha\partial_\mu.
\end{aligned} \tag{74}
$$

Acting (infinitesimally) with these operators on the coordinates gives the following transformations;

$$
\begin{aligned}
\delta_\xi x &= i\theta\sigma\overline{\xi} - i\xi\sigma\overline{\theta} \\
\delta_\xi \theta &= \xi \\
\delta_\xi \overline{\theta} &= \overline{\xi} \\
\delta_\xi y &= 2i\theta\sigma\overline{\xi}.
\end{aligned} \tag{75}
$$

It is easy to check that, firstly,

$$
\begin{aligned}
\Phi(y,\theta) + \delta_\xi\Phi(y,\theta) &= \Phi(y + 2i\theta\sigma\overline{\xi}, \theta + \xi) \\
&= \Phi(y,\theta) + \delta_\xi\phi + \sqrt{2}\theta\delta_\xi\chi + \theta\theta\delta_\xi F \tag{76}
\end{aligned}
$$

so that the differential operators above do indeed produce the infinitesimal transformations, and secondly, that the momentum is related to the supersymmetry generators as,

$$\{Q_\alpha, \overline{Q}_{\dot\alpha}\} = 2i\sigma^\mu_{\alpha\dot\alpha}\partial_\mu \tag{77}$$

all other anti-commutators vanishing. This is one piece of the supersymmetry algebra and leads to Equation 7. The remaining pieces are provided by the Poincaré algebra of normal special relativity (*i.e.* the algebra of commutations of the momentum P_μ and six $M_{\mu\nu}$ generators of the Lorentz group – three boosts M_{i0} and three angular momentum M_{ij} – which represent currents conserved under a Poincaré transformation $x^\mu \to \Lambda^\mu_\nu x^\nu + a^\mu$). The generator Q_α has a current associated with it too, the *super*-current. This is simply the No-ether current for the infinitesimal supersymmetry transformations given above;

$$S_{\alpha\mu} = -\sqrt{2}\left(2\chi_{i\alpha}\partial_\mu\phi_i^* - i\frac{\partial W}{\partial\phi_i}(\sigma_\mu\overline{\chi}_i)_\alpha\right) \tag{78}$$

with $Q_\alpha = \int dx^3 S_{\alpha 0}$.

The algebra that we have described is called $N = 1$ supersymmetry, because there is one generator Q_α. Models with N generators are referred to as having *N-extended* supersymmetry, and their multiplets become increasingly complicated. $N > 8$ is not normally considered because it requires helicities of $5/2$ and higher. Besides, we also have more than one graviton for these cases. $N = 4$ is the largest value of N that we see coming from heterotic strings. However, $N = 1$ is favoured because it allows chiral generations to exist (in the sense that we may have different representations of left and right handed fields).

Appendix C: Gauge invariant supersymmetry

Using the technology presented in the main text, we are already able to construct gauge invariant interactions such as Yukawa couplings using the superpotential. However gauge invariance also requires some modification of the kinetic terms, as well as a Yang-Mills term. This appendix is devoted to the task of determining gauge invariant expressions for these and is included for completeness. I should point out that the calculations required to produce the results of this appendix become more tedious as we progress. Because of this it is not possible (and would not be very useful) to give their derivation in the same detail that was possible for the superpotential, and the main aim will be to present the flow of the argument..

Gauge invariant kinetic terms

Our first task is to introduce *vector*-superfields. Vector superfields are the other convenient way of deriving supersymmetry invariant Lagrangians. Like the chiral-superfields, they are dimensionless, but obey

$$V(x, \theta, \overline{\theta}) = \overline{V}(x, \theta, \overline{\theta}). \tag{79}$$

In fact (in a roundabout way) we already know how to construct a supersymmetry invariant Lagrangian from such a field; the kinetic energy terms may be derived from the

vector superfield

$$K(x,\theta,\overline{\theta}) = \overline{\Phi}\Phi. \tag{80}$$

Clearly this obeys (79), and Taylor expanded in θs, it becomes

$$
\begin{aligned}
K(x,\theta,\overline{\theta}) = {}& \phi^*\phi + \sqrt{2}(\theta X + \overline{\theta}\overline{X}) + \theta\theta\phi^* F + \overline{\theta\theta}\phi F^* \\
& - \theta\sigma^\mu\overline{\theta}\left(X\sigma\overline{X} + i\phi\partial_\mu\phi^* - i\phi^*\partial_\mu\phi\right) \\
& + \theta\theta\overline{\theta}_{\dot\alpha}\left(i/\sqrt{2}\partial_\mu(\overline{\sigma}^{\mu\dot\alpha\alpha}X_\alpha\phi^*)\right) - \overline{\theta\theta}\theta^\alpha\left(i/\sqrt{2}\partial_\mu(\sigma^\mu_{\alpha\dot\alpha}\overline{X}^{\dot\alpha}\phi)\right) \\
& + \theta\theta\overline{\theta\theta}\left(\frac{i}{2}(\partial\overline{X}\sigma X + \partial X\sigma\overline{X}) - \frac{1}{2}|\partial\phi|^2 + \frac{1}{4}\phi\partial^2\phi^* + \frac{1}{4}\phi^*\partial^2\phi + FF^*\right). \tag{81}
\end{aligned}
$$

Notice that the series terminates at the $\theta\theta\overline{\theta\theta}$ term, which is the highest possible term. Also notice that, apart from some total derivatives, the final term is the kinetic energy piece of the Lagrangian. Traditionally, the $\theta\theta\overline{\theta\theta}$ term of a vector field is referred to as the *D-term*, so that we may now write the kinetic energy more succinctly as

$$\mathcal{L}_{\mathrm{KE}} = \overline{\Phi}\Phi\Big|_D. \tag{82}$$

As we have seen this term transforms into a total derivative under a supersymmetry transformation. In fact this is a general property of D-terms, which is perhaps not so surprising when we realise that D-terms, like F-terms, are auxiliary fields which do not propagate.

In order to show this we need to find the supersymmetry transformations for the components of an arbitrary vector superfield, which we can write in its Taylor expanded form as,

$$
\begin{aligned}
V(x,\theta,\overline{\theta}) = {}& C(x) + \theta\zeta(x) + \overline{\theta}\overline{\zeta}(x) \\
& + \theta\theta G(x) + \overline{\theta\theta}G^*(x) - \theta\sigma^\mu\overline{\theta}v_\mu(x) \\
& + i\theta\theta\overline{\theta}_{\dot\alpha}\left(\overline{\lambda}(x)^{\dot\alpha} + \overline{\sigma}^{\mu\dot\alpha\alpha}\partial_\mu\zeta_\alpha\right) - i\overline{\theta\theta}\theta^\alpha\left(\lambda_\alpha(x) - \sigma^\mu_{\alpha\dot\alpha}\partial_\mu\overline{\zeta}^{\dot\alpha}\right) \\
& + \theta\theta\overline{\theta\theta}\left(\frac{1}{2}D(x) + \frac{1}{4}\partial^2 C(x)\right). \tag{83}
\end{aligned}
$$

The spin one component v^μ lends its name to the entire vector superfield. In gauge superfields, the components λ, $\overline{\lambda}$ and v^μ will be the gaugino, anti-gaugino and gauge boson respectively. There are a number of ways to determine the supersymmetry transformations of the components in this multiplet. We could perform the same operation as in the text for the chiral supermultliplet (*i.e.* begin with $\delta_\xi C(x) = \xi\zeta + \overline{\xi}\overline{\zeta}$, and continue up through the multiplet, checking at each stage that the supersymmetry algebra operates correctly on each component), or we could deduce it by comparison with $K(x,\theta,\overline{\theta})$, or most simply, we could extract the components of

$$V(x+i\theta\sigma\overline{\xi}-i\xi\sigma\overline{\theta},\,\theta+\xi,\,\overline{\theta}+\overline{\xi}) = V(x,\theta,\overline{\theta}) + \delta_\xi V = V(x,\theta,\overline{\theta}) + (\xi Q + \overline{\xi}\overline{Q})V \tag{84}$$

where $Q,\overline{Q}$ are the differential operators defined in the Introduction. The result for the last three components is

$$
\begin{aligned}
\delta_\xi\lambda^A_\alpha(x) &= i\xi_\alpha D^A - \frac{1}{2}(\sigma^\mu\overline{\sigma}^\nu)^\beta_\alpha\xi_\beta F^A_{\mu\nu} \\
\delta_\xi v^A_\mu(x) &= -i\left(\xi\sigma^\mu\overline{\lambda}^A + \lambda^A\sigma^\mu\overline{\xi i}\right) - \partial_\mu\left(\xi\sigma^\mu\zeta^A + \overline{\xi}\overline{\zeta}^A\right) \\
\delta_\xi D^A(x) &= \left(\xi\sigma^\mu\overline{\partial}_\mu\lambda^A + \partial_\mu\lambda^A\sigma^\mu\overline{\xi}\right). \tag{85}
\end{aligned}
$$

The D-term transforms into a total derivative as required. Having determined this fact, let us now try to introduce gauge invariance into our kinetic term. Under a gauge transformation, we wish our supermultiplets to retain all their properties, so that the natural extension of the transformation

$$\phi_i \to \phi_i' = \left(e^{i\lambda}\right)_{ij} \phi_j \tag{86}$$

(different λ, with $\lambda_{ij} = \lambda^A T_{ij}^A$, and A, i, j as group indices) is

$$\Phi_i \to \Phi_i' = \left(e^{i\Lambda}\right)_{ij} \Phi_j \tag{87}$$

where now Λ is a chiral superfield (remember that the product of two chiral superfields is another chiral superfield). From now on we shall drop the explicit gauge indices. The first thing to notice is that our kinetic term

$$K'(x, \theta, \bar{\theta}) = \overline{\Phi} e^{-i\overline{\Lambda}} e^{i\Lambda} \Phi \tag{88}$$

is not gauge invariant, since the two exponentials do not cancel. It can be made gauge invariant by sandwiching another superfield with the correct transformation properties, between the chiral superfields. This field is the *gauge*-superfield, and since the kinetic term must transform as a vector superfield, it is also a vector superfield. Conventionally the gauge superfield is defined as $gV_{ij} = g_A V_A T_{ij}^A$, such that the kinetic term is

$$K(x, \theta, \bar{\theta}) = \overline{\Phi} e^{2gV} \Phi. \tag{89}$$

The constant g will turn out to be the conventional coupling constant. For gauge invariance, under a gauge transformation V must transform so that it cancels the troublesome phases in Equation 88. That is, we must have

$$e^{2gV} \to e^{i\overline{\Lambda}} e^{2gV} e^{-i\Lambda}. \tag{90}$$

We now have nearly all the ingredients we need for our Lagrangian. The kinetic energy term is found, as before, by extracting the D-term which transforms as a total derivative, but is now gauge invariant too,

$$\mathcal{L}_{\text{KE}} = \overline{\Phi} e^{2gV} \Phi \Big|_D. \tag{91}$$

In order to extract the D-term, we need to expand the exponential. To do so, we first make a gauge transformation such that $C, \zeta, \bar{\zeta}, G = 0$. As we shall see below, by examining infinitesimal gauge transformations, one finds that this gauge, the Wess-Zumino gauge, is always possible. In the Wess-Zumino gauge, things are particularly simple since,

$$
\begin{aligned}
V_A(x, \theta, \bar{\theta}) &= -\theta \sigma^\mu \bar{\theta} v_\mu(x) + i\theta\theta\bar{\theta}\bar{\lambda}(x) - i\bar{\theta}\bar{\theta}\theta\lambda + \frac{1}{2}\theta\theta\bar{\theta}\bar{\theta}D(x), \\
V_A V_B &= -\frac{1}{2}\theta\theta\bar{\theta}\bar{\theta} v_A^\mu v_{A\mu}, \\
V_A V_B V_C &= 0
\end{aligned}
\tag{92}
$$

so that $e^{2gV} = 1 + V + V^2/2$. After some lengthy manipulations and Fierz-rearranging, the kinetic terms in the Lagrangian are found to be

$$
\begin{aligned}
\mathcal{L}_{\text{KE}} &= -i\chi\sigma^\mu \overline{D}_\mu \overline{\chi} - (D_\mu \phi)^\dagger D^\mu \phi + F^* F \\
&\quad + i\sqrt{2}g\left(\phi^* \lambda\chi - \overline{\chi}\overline{\lambda}_A\phi\right) + g\phi D\phi^*.
\end{aligned}
\tag{93}
$$

Note that there are suppressed indices floating around here. For example we have $g\phi D\phi^* \equiv \sum_{Aij} g_A\phi_i D_A T^A_{ij}\phi^*_j$ and $g\phi^*\lambda\chi \equiv \sum_{Aij} g_A\phi^*_i T^A_{ij}\lambda^\alpha_A\chi_{i\alpha}$ where g_A allows for group factors and $U(1)$ charges. The derivatives are the usual covariant ones;

$$\begin{aligned}
\mathcal{D}_{\mu ij} &= \delta_{ij}\partial_\mu + ig_A T^A_{ij}v_{A\mu} \\
\mathcal{D}_{\mu AB} &= \delta_{AB}\partial_\mu + g_A f^A_{BC}v^{C\mu}.
\end{aligned} \tag{94}$$

In order to maintain the Wess-Zumino gauge, supersymmetry transformations must be accompanied by the required gauge transformation. That is, one performs an infinitesimal supersymmetry transformation on the gauge superfield. This generates new $C, \zeta, \bar{\zeta}, G$ terms which are found to cancel for a certain additional infinitesimal gauge transformation on V. Thus the purpose of Wess-Zumino gauge is to make explicit all the physical degrees of freedom. (Explicitly, the supersymmetry transformations give $\delta_\xi C = 0$, $\delta_\xi\zeta = -i\sigma^\mu\bar{\xi}v_\mu$, $\delta_\xi G = 2\bar{\xi}\bar{\lambda}$, and the gauge transformation which cancels it is given by $\Lambda = \{0, \frac{i}{\sqrt{2}}\sigma^\mu\bar{\xi}v_\mu, -\bar{\xi}\bar{\lambda}\}$.) The combined transformation we shall refer to as a *supergauge* transformation δ_ξ^{SG}. Roughly speaking, this changes the derivatives in the supersymmetry transformations into covariant ones. That is, for the chiral superfields we have

$$\begin{aligned}
\delta_\xi^{SG}\phi &= \sqrt{2}\xi\chi \\
\delta_\xi^{SG}\chi &= i\sqrt{2}\sigma^\mu\bar{\xi}\mathcal{D}_\mu\phi + \sqrt{2}\xi F \\
\delta_\xi^{SG}F &= i\sqrt{2}\bar{\xi}\bar{\sigma}^\mu\mathcal{D}_\mu\chi + 2g\phi\bar{\lambda}\bar{\xi},
\end{aligned} \tag{95}$$

and for the gauge boson and gauginos and D-field we have

$$\begin{aligned}
\delta_\xi^{SG}\lambda^A_\alpha(x) &= i\xi_\alpha D^A - \frac{1}{2}(\sigma^\mu\bar{\sigma}^\nu)^\beta_\alpha\xi_\beta F^A_{\mu\nu} \\
\delta_\xi^{SG}v^A_\mu(x) &= -i\left(\xi\sigma^\mu\bar{\lambda}^A - \lambda^A\sigma^\mu\bar{x}\bar{i}\right) \\
\delta_\xi^{SG}D^A(x) &= \left(\xi\sigma^\mu\overline{\mathcal{D}}^A_B\mu\lambda^B + \mathcal{D}^A_B\lambda^B\sigma^\mu\bar{\xi}\right).
\end{aligned} \tag{96}$$

The Yang Mills terms

This completes the kinetic terms for the matter fields. The final terms that we need are the Yang-Mills terms. These can be found as in the text by forming an additional chiral superfield $\mathcal{W}_\alpha(y,\theta)$ out of the three physical fields, λ^A_α, $F^A_{\mu\nu}$, and D^A. Since the gaugino is the component with lowest dimension, we begin by defining;

$$\mathcal{W}_\alpha(x,0,0) = -i\lambda_\alpha(x). \tag{97}$$

This chiral superfield is a function with dimension $3/2$ and helicity $1/2$, and is the *field-strength superfield*. Its general form may be found by repeating what we did in the main text but this time for finite supergauge transformations. Making an arbitrary supergauge transformation, and using the transformation rules for the components above we find,

$$\begin{aligned}
\mathcal{W}^A_\alpha(y,\theta) &= \delta_\theta^{SG}\mathcal{W}^A_\alpha(x,0) \\
&= -i\lambda^A_\alpha(y) + \theta_\beta\left(\delta^\beta_\alpha D^A + \frac{i}{2}(\sigma^\mu\bar{\sigma}^\nu)^\beta_\alpha F^A_{\mu\nu}\right) + \theta\theta\sigma^\mu_{\alpha\dot{\alpha}}\mathcal{D}^{AC}_\mu\bar{\lambda}^{\dot{\alpha}}_C
\end{aligned} \tag{98}$$

where

$$F^A_{\mu\nu} = \partial_\mu v^A_\nu - \partial_\nu v^A_\mu - g_A f^A_{BC}v^A_\mu v^B_\nu.$$

The required, supergauge invariant, Lagrangian can be formed in the usual way by taking the F-component of a combination of these superfields; it is given by

$$
\begin{aligned}
\mathcal{L}_{\mathrm{YM}} &= \frac{1}{4} \mathrm{Tr} \left(\mathcal{W}_A^\alpha \mathcal{W}_\alpha^A \big|_F + \overline{\mathcal{W}}_{\dot\alpha}^A \overline{\mathcal{W}}_A^{\dot\alpha} \big|_{F*} \right) \\
&= -\frac{1}{4} F_{\mu\nu}^A F_A^{\mu\nu} - i\lambda^A \sigma^\mu \mathcal{D}_{\mu A}^C \bar\lambda_C + \frac{1}{2} D^A D_A
\end{aligned}
\tag{99}
$$

in the Wess-Zumino gauge. The full $N = 1$ supersymmetric, gauge invariant Lagrangian is given by the sum of equations 91,16,99:

$$
\mathcal{L} = \mathcal{L}_{\mathrm{KE}} + \mathcal{L}_{\mathrm{int}} + \mathcal{L}_{\mathrm{YM}}
\tag{100}
$$

This is the main result of this Appendix. Those who would rather avoid any discussion of superfields at all, could simply verify that this Lagrangian is indeed invariant under the general supergauge transformations given above. Notice again that the auxiliary D field may be easily removed by using its equations of motion,

$$
D^A = g_A \phi_i^* T_{ij}^A \phi_j.
\tag{101}
$$

The remainder of this Appendix expands a little on the role played by the chiral $\mathcal{W}_\alpha$ field; it may be skipped at the reader's discretion.

In standard gauge theory, the gauge field is the connection required for covariant derivatives. These are quantities which, under a gauge transformation

$$
\phi_i \rightarrow \phi_i' = \left(e^{i\lambda} \right)_{ij} \phi_j,
\tag{102}
$$

transform covariantly:

$$
\mathcal{D}_\mu \phi_i \rightarrow (\mathcal{D}_\mu \phi_i)' = \left(e^{i\lambda} \right)_{ij} \mathcal{D}_\mu \phi_j.
\tag{103}
$$

The connection is everything other than the derivative in $\mathcal{D}_\mu = \partial_\mu - ig v_\mu$, and the transformation of v_μ is therefore defined so that

$$
\mathcal{D}_\mu' = \left(e^{i\lambda} \right)_{ij} \mathcal{D}_\mu \left(e^{-i\lambda} \right)_{ij}.
\tag{104}
$$

Quite generally, there is only one non-trivial tensor that can be made from the the connection, by taking its covariant derivative. This is the curvature, which is given by

$$
F_{\mu\nu} = \mathcal{D}_\mu A_\nu - \mathcal{D}_\nu A_\mu.
\tag{105}
$$

So much for standard Yang-Mills theory, but what about its supersymmetric counterpart? The field strength superfield $\mathcal{W}_\alpha$, may also be related to the connection in a super/gauge covariant derivative (just as the usual field strength $F_{\mu\nu}$ is related to the gauge bosons). We begin with the derivatives which anti-commute with Q_α and $\overline{Q}^{\dot\alpha}$ which we shall attempt to make gauge covariant; these are

$$
\begin{aligned}
D_\alpha &= \frac{\partial}{\partial\theta^\alpha} + i\sigma_{\alpha\dot\alpha}^\mu \bar\theta^{\dot\alpha} \partial_\mu \\
D^{\dot\alpha} &= -\frac{\partial}{\partial\theta_{\dot\alpha}} - i\theta^\alpha \sigma_\alpha^{\mu\dot\alpha} \partial_\mu.
\end{aligned}
\tag{106}
$$

We can easily verify that $\{D_\alpha, Q_\alpha\} = \{D_\alpha, \overline{Q}_{\dot{\alpha}}\} = \{\overline{D}_{\dot{\alpha}}, \overline{Q}_{\dot{\alpha}}\} = \{\overline{D}_{\dot{\alpha}}, Q_\alpha\} = 0$. In addition $\overline{D}_{\dot{\alpha}} y = 0$. This means that an alternative (and in fact the usual) definition of a chiral superfield is $\overline{D}_{\dot{\alpha}} \Phi(y, \theta) = 0$. The fact that $\overline{D}_{\dot{\alpha}} \overline{D}^2 = 0$ gives us a useful *chiral projection operator*,

$$P = -\frac{1}{4}\overline{D}^2, \tag{107}$$

which, when applied any function of $x, \theta, \overline{\theta}$, turns it into a chiral superfield (that is a function of y and θ only). Now consider the derivative

$$D_\alpha \Phi. \tag{108}$$

This derivative is not gauge covariant. A covariant derivative is one which transforms as follows under gauge transformations:

$$\nabla'_\alpha \Phi' \;=\; \nabla'_\alpha e^{i\Lambda}\Phi \;=\; e^{i\Lambda}(\nabla_\alpha \Phi). \tag{109}$$

Thus

$$\nabla'_\alpha = e^{i\Lambda}\nabla_\alpha e^{-i\Lambda}. \tag{110}$$

By examining the gauge transformations of V in Equation 90, we see that

$$\nabla_\alpha = e^{-2gV} D_\alpha e^{2gV} \tag{111}$$

transforms in the correct way, where D_α acts on everything to its right. The connection superfield is defined as the difference between this term and D_α. *i.e.*

$$\Gamma_\alpha = e^{-2gV}(D_\alpha e^{2gV}). \tag{112}$$

Finally the chiral field $\mathcal{W}_\alpha$ which we derived earlier is the chiral projection of this:

$$\mathcal{W}_\alpha = P\Gamma_\alpha. \tag{113}$$

The gauge invariance of the Yang-Mills term is now manifest.

8   Appendix D: *R*-symmetry and *R*-parity

As well as the terms in the superpotential of the MSSM there are many other Yukawa couplings that we could have included but didn't. For example,

$$W_{\text{problem}} = \mu' L^T \epsilon' H_2 + \lambda' EL^T \epsilon' L + \lambda'' DL^T \epsilon' Q + \lambda''' \epsilon_{abc} U^a D^b D^c, \tag{114}$$

where a, b, c are colour indices, are all invariant under $SU(3)_c \times SU(2)_L \times U(1)_Y$. Including *all* of the these terms would lead to disastrously fast proton decay (suppressed by only the inverse squark mass squared). Thus, counting over generation indices, we have to explain the absence of 30 possible couplings in the MSSM superpotential. This can be done by the introduction of a global $U(1)_R$ symmetry known as *R–symmetry*. An *R*–transformation may be defined as a transformation of the spinor coordinates,

$$\begin{aligned}
\theta &\rightarrow e^{i\alpha}\theta \\
\overline{\theta} &\rightarrow e^{-i\alpha}\overline{\theta}.
\end{aligned} \tag{115}$$

Under this transformation we may assign a charge, the *R–charge*, to each of the chiral superfields:

$$\phi'_i(x, \theta', \overline{\theta}') = e^{i\alpha R_i} \phi'_i(x, \theta, \overline{\theta}).$$ (116)

For the elements of the supermultiplet this means

$$\{\phi_i, \chi_i, F_i\} \rightarrow e^{i\alpha R_i}\{\phi_i, e^{-i\alpha}\chi_i, e^{-2i\alpha}F_i\}$$
$$\{\phi_i^*, \overline{\chi}_i, F_i^*\} \rightarrow e^{-i\alpha R_i}\{\phi_i^*, e^{-i\alpha}\overline{\chi}_i, e^{2i\alpha}F_i^*\}.$$ (117)

Since the elements transform differently this also means that R transformations do not commute with supersymmetry transformations (*i.e.* in order to be consistent the supersymmetry generators in Appendix A must obey $Q \rightarrow e^{-i\alpha}Q$ which may easily be checked). Now let us examine which of the terms in the Lagrangian is invariant under R-transformations. The ϕ kinetic term is invariant only if we assign $R_v = 0$ to the gauge vector, since $R_\phi = -R_{\overline{\phi}}$. Examining the kinetic terms in, for example, Equation 22, we find $\lambda \rightarrow e^{i\alpha}\lambda$. The supersymmetric Yang-Mills Lagrangian (shown in Appendix C) is also invariant under this transformation. This leaves the interaction term which looks like

$$\begin{aligned}\mathcal{L}_{\text{int}} &= W(\Phi)|_F + \text{h.c.} \\ &= \lambda_{ijk}\Phi_i\Phi_j\Phi_k|_F + \text{h.c.}\end{aligned}$$ (118)

By looking at the transformations in (117), the interaction Lagrangian transforms as

$$\lambda_{ijk}\Phi_i\Phi_j\Phi_k|_F \rightarrow e^{i\alpha(\sum_i R_i - 2)}\lambda_{ijk}\Phi_i\Phi_j\Phi_k|_F$$ (119)

and therefore only the terms with

$$\sum_i R_i = 2$$ (120)

survive in the R–invariant superpotential. A choice of charges which results in the MSSM is

$$R_i = 3B_i + L_i$$ (121)

where B_i and L_i are the baryon number and lepton number of Φ_i. Clearly Equation 120 is not satisfied by any of the 30 couplings in Equation 114. When supersymmetry is broken (for example by a Majorana mass for the gauginos), a residual *discrete* symmetry remains in the Lagrangian, given by

$$f_i \rightarrow (-1)^{2S_i + 3B_i + L_i} f_i$$ (122)

where f_i is an arbitrary field of spin S_i. This particular choice of discrete symmetry is called *R–parity* or *matter–parity*. Other (equally valid) choices are said to be *R*-parity breaking. For example, one particularly interesting choice, called *baryon–parity* is [35],

$$f_i \rightarrow (-1)^{2S_i + 3B_i} f_i.$$ (123)

In this case only the first two terms in Equation 114 survive, but nucleon decay is absent. This model is interesting because this discrete symmetry can arise in certain string models (for more details see [35, 36]).

References

[1] J. Wess and J. Bagger, 'Supersymmetry and Supergravity', (Princeton Series in Physics 1982)
P. Srivastava, 'Supersymmetry, Superfields and Supergravity', (Adam Hilger 1986)
R. N. Mohapatra, 'Unification and Supersymmetry', (Springer Verlag 1986)
A.B. Lahanas, D.V. Nanopoulos, "The road to no scale supergravity", Phys.Rept.145:1,1987

[2] J. A. Casas, V. Di Clemente and M. Quiros, Nucl. Phys. B **581** (2000) 61 [arXiv:hep-ph/0002205].

[3] E. Gabrielli, A. Masiero and L. Silvestrini, Phys. Lett. B **374** (1996) 80 [arXiv:hep-ph/9509379]; F. Gabbiani, E. Gabrielli, A. Masiero and L. Silvestrini, Nucl. Phys. B **477** (1996) 321 [arXiv:hep-ph/9604387].

[4] T. Falk and K. A. Olive, Phys. Lett. **B375**, 196 (1996) [hep-ph/9602299]; Phys. Lett. B **439**, 71 (1998) [hep-ph/9806236]; C. Hamzaoui, M. Pospelov and R. Roiban, Phys. Rev. D **56**, 4295 (1997) [hep-ph/9702292]; A. Bartl, T. Gajdosik, W. Porod, P. Stockinger and H. Stremnitzer, Phys. Rev. D **60**, 073003 (1999) [hep-ph/9903402].

[5] T. Ibrahim and P. Nath, Phys. Lett. **B418** (1998) 98; T. Ibrahim and P. Nath, Phys. Rev. **D57** (1998) 478. T. Ibrahim and P. Nath, Phys. Rev. D **58**, 111301 (1998) [hep-ph/9807501]; S. Pokorski, J. Rosiek and C. A. Savoy, Nucl. Phys. B **570**, 81 (2000) [hep-ph/9906206].

[6] T. Falk, K. A. Olive, M. Pospelov and R. Roiban, Nucl. Phys. B **560** (1999) 3 [arXiv:hep-ph/9904393]; S. Abel, S. Khalil and O. Lebedev, Phys. Rev. Lett. **86** (2001) 5850 [arXiv:hep-ph/0103031]; S. Abel, S. Khalil and O. Lebedev, Nucl. Phys. B **606** (2001) 151 [arXiv:hep-ph/0103320].

[7] E. Gabrielli and G. F. Giudice, Nucl. Phys. B **433** (1995) 3 [Erratum-ibid. B **507** (1995) 549] [arXiv:hep-lat/9407029].

[8] A. J. Buras, P. Gambino, M. Gorbahn, S. Jager and L. Silvestrini, Nucl. Phys. B **592** (2001) 55 [arXiv:hep-ph/0007313].

[9] A. Ali and D. London, Phys. Rept. **320** (1999) 79 [arXiv:hep-ph/9907243].

[10] S. A. Abel and J. M. Frere, Phys. Rev. D **55** (1997) 1623 [arXiv:hep-ph/9608251]; S. Khalil, T. Kobayashi and A. Masiero, Phys. Rev. D **60** (1999) 075003 [arXiv:hep-ph/9903544]; A. Masiero and O. Vives, arXiv:hep-ph/0001298; A. Masiero and O. Vives, Phys. Rev. Lett. **86** (2001) 26 [arXiv:hep-ph/0007320]; T. Kobayashi and O. Vives, Phys. Lett. B **506** (2001) 323 [arXiv:hep-ph/0011200]. S. Abel, D. Bailin, S. Khalil and O. Lebedev, Phys. Lett. B **504** (2001) 241 [arXiv:hep-ph/0012145]; D. F. Carvalho, M. E. Gomez and S. Khalil, JHEP **0107** (2001) 001 [arXiv:hep-ph/0101250]; S. w. Baek, J. H. Jang, P. Ko and J. H. Park, Nucl. Phys. B **609** (2001) 442 [arXiv:hep-ph/0105028].

[11] M. Brhlik, L. Everett, G.L. Kane and J. Lykken, Phys. Rev. Lett. **83**, 2124 (1999); M. Brhlik, L. Everett, G. L. Kane and J. Lykken, Phys. Rev. D **62**, 035005 (2000).

[12] E. Accomando, R. Arnowitt and B. Dutta, Phys. Rev. D **61**, 075010 (2000); T. Ibrahim and P. Nath, Phys. Rev. D **61**, 093004 (2000).

[13] P. Nath, Phys. Rev. Lett. **66** (1991), 2565; Y. Kizukuri and N. Oshimo, Phys. Rev. **D46** (1992) 3025.

[14] R. Barbieri and G.F. Giudice, Nucl. Phys. **B306** (1988), 63; J. Ellis, K. Enqvist, D.V. Nanopoulos, F. Zwirner, Mod. Phys. Lett. **A1** (1986), 57; G.G. Ross and R.G. Roberts, Nucl. Phys. **B377** (1992), 571.

[15] J.L. Feng, K.T. Matchev, and T. Moroi, Phys. Rev. **D61** (2000), 075005; Phys. Rev. Lett. **84** (2000), 2322.

[16] J. L. Feng, C. Kolda and N. Polonsky, Nucl. Phys. B **546**, 3 (1999).

[17] H. Baer, C. Balazs, M. Brhlik, P. Mercadante, X. Tata, Y. Wang, hep-ph/0102156.

[18] S. Dimopoulos and G. F. Giudice, Phys. Lett. **B357** (1995), 573.

[19] H. Baer, M. A. Diaz, P. Quintana and X. Tata, JHEP**0004**, 016 (2000).

[20] A. Masiero and O. Vives, Nucl. Phys. Proc. Suppl. **99B** (2001) 228 [arXiv:hep-ph/0107276].

[21] S. Khalil and O. Lebedev, Phys. Lett. B **515** (2001) 387 [arXiv:hep-ph/0106023].

[22] Y. Nir, arXiv:hep-ph/9911321.

[23] M. Dine, R. G. Leigh and D. A. MacIntire, Phys. Rev. Lett. **69** (1992) 2030 [arXiv:hep-th/9205011]; M. Dine, E. Kramer, Y. Nir and Y. Shadmi, Phys. Rev. D **63** (2001) 116005 [arXiv:hep-ph/0101092];

[24] B. Acharya, D. Bailin, A. Love, W. A. Sabra and S. Thomas, Phys. Lett. B **357** (1995) 387 [arXiv:hep-th/9506143]; D. Bailin, G. V. Kraniotis and A. Love, Nucl. Phys. B **518** (1998) 92 [arXiv:hep-th/9707105]. D. Bailin, G. V. Kraniotis and A. Love, Phys. Lett. B **432** (1998) 343 [arXiv:hep-th/9804135]; D. Bailin, G. V. Kraniotis and A. Love, Phys. Lett. B **435** (1998) 323 [arXiv:hep-th/9805111]; T. Dent, Phys. Rev. D **64** (2001) 056005 [arXiv:hep-ph/0105285]; O. Lebedev, Phys. Lett. B **521** (2001) 71 [arXiv:hep-th/0108218]; T. Dent, arXiv:hep-th/0110110.

[25] S. A. Abel and G. Servant, Nucl. Phys. B **597** (2001) 3 [arXiv:hep-th/0009089]; S. A. Abel and G. Servant, Nucl. Phys. B **611** (2001) 43 [arXiv:hep-ph/0105262]; S. Khalil, O. Lebedev and S. Morris, arXiv:hep-th/0110063.

[26] J. L. Feng, K. T. Matchev and Y. Shadmi, hep-ph/0107182; A. Romanino and A. Strumia, arXiv:hep-ph/0108275.

[27] M. Graesser and S. Thomas, arXiv:hep-ph/0104254.

[28] J. Feng and K. Matchev, *Phys. Rev. Lett.* **86**, (2001), 3480; U. Chattopadhyay and P. Nath, *Phys. Rev. Lett.* **86**, (2001), 5854; S. Komine, T. Moroi and M. Yamaguchi, *Phys. Lett.* B **506**, (2001), 93; *Phys. Lett.* B **507**, (2001), 224; T. Ibrahim, U. Chattopadhyay and P. Nath, *Phys. Rev.* D **64**, (2001), 016010; J. Ellis, D.V. Nanopoulos and K. A. Olive, *Phys. Lett.* B **508**, (2001), 65; S. Martin and J. Wells, *Phys. Rev.* D **64**, (2001), 035003; H. Baer, C. Balazs, J. Ferrandis and X. Tata, *Phys. Rev.* D **64**, (2001), 035004; F. Richard, hep-ph/0104106; D. Carvalho, J. Ellis, M. Gomez and S. Lola, hep-ph/0103256; S. Baek, T. Goto, Y. Okada and K. Okumura, hep-ph/0104146; Y. Kim and M. Nojiri, hep-ph/0104258; K. Choi, K. Hwang, S.K. Kang, K.Y. Lee and W.Y. Song, *Phys. Rev.* D **64**, (2001), 055001; W. de Boer, M. Huber, C. Sander, D.I. Kazakov, hep-ph/0106311; L. Roszkowski, R.R. de Austri and T. Nihei, hep-ph/0106334; A. Dedes, H.K. Dreiner and U. Nierste, hep-ph/0108037; R. Arnowitt, B. Dutta, B. Hu and Y. Santoso, *Phys. Lett.* B **505**, (2001), 177; D. G. Cerdeno, E. Gabrielli, S. Khalil, C. Munoz and E. Torrente-Lujan, Phys. Rev. D **64** (2001) 093012 [arXiv:hep-ph/0104242].

[29] A. Pomarol, *Phys. Lett.* B **287**, (1992), 331; K.S. Babu and S.M. Barr, *Phys. Rev.* D **49**, (1994), R2156; A.T. Davies, C.D. Froggatt and A. Usai, *Proc. of the International Europhysics Conference on High Energy Physics*, Jerusalem (1997), Eds. D. Lellouch, G. Mikenberg and E. Rabinovici , Springer-Verlag (1998),p891; A.T.Davies, C.D. Froggatt and A. Usai, *Proc. of Conference on Strong and Electroweak Matter, Copenhagen* (1998), Eds. J. Ambjorn, P.H. Damgaard, K. Kainulainen, K. Rummukainen, World Scientific (1999), p227; G.C. Branco, F. Krüger, J.C. Romão, A. Teixeira, hep-ph/0012318; A. T. Davies, C. D. Froggatt and A. Usai, Phys. Lett. B **517** (2001) 375 [arXiv:hep-ph/0105266].

[30] J. R. Ellis, J. Hisano, S. Lola and M. Raidal, arXiv:hep-ph/0109125.

[31] T. Nihei, Prog. Theor. Phys. **98**, 1157 (1997) [hep-ph/9707336]; A. Ali, H. Asatrian and C. Greub, Phys. Lett. **B429**, 87 (1998) [hep-ph/9803314]; A. L. Kagan and M. Neubert, Phys. Rev. D **58**, 094012 (1998) [hep-ph/9803368]; S. Baek and P. Ko, Phys. Rev. Lett. **83**, 488 (1999) [hep-ph/9812229]; D. A. Demir, A. Masiero and O. Vives, Phys. Rev. Lett. **82**, 2447 (1999) [hep-ph/9812337]; D. Bailin and S. Khalil, Phys. Rev. Lett. **86** (2001) 4227 [arXiv:hep-ph/0010058]; A. Ali and E. Lunghi, hep-ph/0105200; D. A. Demir and K. A. Olive, arXiv:hep-ph/0107329.

[32] A. Pilaftsis, Phys. Lett. B **435**, 88 (1998) [hep-ph/9805373]; D. A. Demir, Phys. Rev. D **60**, 055006 (1999) [hep-ph/9901389]; A. Pilaftsis and C. E. Wagner, Nucl. Phys. B **553**, 3 (1999) [hep-ph/9902371]; T. Ibrahim and P. Nath, Phys. Rev. D **63**, 035009 (2001) [hep-ph/0008237]; M. Carena, J. Ellis, A. Pilaftsis and C. E. Wagner, Nucl. Phys. B **586**, 92 (2000) [hep-ph/0003180]; Phys. Lett. B **495**, 155 (2000) [hep-ph/0009212].

[33] B. Allanach *et al.*, arXiv:hep-ph/9906224.

[34] F. Borzumati, G. R. Farrar, N. Polonsky and S. Thomas, Nucl. Phys. B **555** (1999) 53 [arXiv:hep-ph/9902443].

[35] L. E. Ibanez and G. G. Ross, Nucl. Phys. B **368** (1992) 3.

[36] H. Dreiner and G. G. Ross, Nucl. Phys. B **410** (1993) 188 [arXiv:hep-ph/9207221].

B phenomenology

Sheldon Stone

Syracuse University, USA

1 The Standard Model and B decays

Studies of b and c physics are focused on two main goals. The first is to look for new phenomena beyond the Standard Model and the second is to measure Standard Model parameters including CKM elements and decay constants. These lectures concern "B Phenomenology," a topic so broad that it can include almost anything concerning b quark decay or production. I will cover an eclectic ensemble of topics that I find interesting and hope will be educational.

1.1 Theoretical basis

The physical states of the "Standard Model" are comprised of left-handed doublets containing leptons and quarks and right handed singlets (Rosner 2001)

$$\begin{pmatrix} u \\ d \end{pmatrix}_L \quad \begin{pmatrix} c \\ s \end{pmatrix}_L \quad \begin{pmatrix} t \\ b \end{pmatrix}_L, \quad u_R, \quad d_R, \quad c_R, \quad s_R, \quad t_R, \quad b_R \tag{1}$$

$$\begin{pmatrix} e^- \\ \nu_e \end{pmatrix}_L \quad \begin{pmatrix} \mu^- \\ \nu_\mu \end{pmatrix}_L \quad \begin{pmatrix} \tau^- \\ \nu_\tau \end{pmatrix}_L, \quad e_R^-, \quad \mu_R^-, \quad \tau_R^-, \quad \nu_{eR}, \quad \nu_{\mu R}, \quad \nu_{\tau R}. \tag{2}$$

The gauge bosons, $W^\pm$, γ and Z^0 couple to mixtures of the physical d, s and b states. This mixing is described by the Cabibbo-Kobayashi-Maskawa (CKM) matrix (Kobayashi 1973).

The Lagrangian for charged current weak decays is

$$L_{cc} = -\frac{g}{\sqrt{2}} J_{cc}^\mu W_\mu^\dagger + h.c., \tag{3}$$

where

$$J_{cc}^\mu = (\bar{\nu}_e, \ \bar{\nu}_\mu, \ \bar{\nu}_\tau) \, \gamma^\mu V_{MNS} \begin{pmatrix} e_L \\ \mu_L \\ \tau_L \end{pmatrix} + (\bar{u}_L, \ \bar{c}_L, \ \bar{t}_L) \, \gamma^\mu V_{CKM} \begin{pmatrix} d_L \\ s_L \\ b_L \end{pmatrix} \tag{4}$$

and

$$V_{CKM} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}. \tag{5}$$

Multiplying the mass eigenstates (d_L, s_L, b_L) by the CKM matrix leads to the weak eigenstates. V_{MNS} is the analogous matrix required for massive neutrinos (we will not discuss this matrix any further). There are nine complex CKM elements. These 18 numbers can be reduced to four independent quantities by applying unitarity constraints and the fact that the phases of the quark wave functions are arbitrary. These four remaining numbers are fundamental constants of nature that need to be determined from experiment, like any other fundamental constant such as α or G. In the Wolfenstein approximation the matrix is written to order λ^3 for the real part and λ^4 for the imaginary part as (Wolfenstein 1983)

$$V_{CKM} = \begin{pmatrix} 1 - \lambda^2/2 & \lambda & A\lambda^3(\rho - i\eta)(1 - \lambda^2/2) \\ -\lambda & 1 - \lambda^2/2 - i\eta A^2\lambda^4 & A\lambda^2(1 + i\eta\lambda^2) \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix}. \tag{6}$$

The constants λ and A are determined from charged-current weak decays. The measured values are $\lambda = 0.2205 \pm 0.0018$ and A=0.784±0.043. There are constraints on ρ and η from other measurements that we will discuss. Usually the matrix is viewed only up to order λ^3. To explain CP violation in the K^0 system the term of order λ^4 in V_{cs} is necessary.
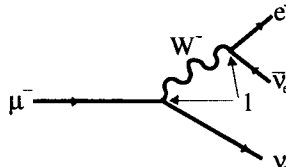
Figure 1. *Diagram for muon decay.*

1.1.1   Determination of G_F

Muons, being lighter than the lightest hadrons, must decay purely into leptons. The process is $\mu^- \to e^- \bar{\nu}_e \nu_\mu$ as shown in Figure 1. The total width for this decay process is given by

$$\Gamma_\mu = \frac{G_F^2}{192\pi^3} m_\mu^5 \times (\text{phase space}) \times (\text{radiative corrections}). \tag{7}$$

Since $\Gamma_\mu \cdot \tau_\mu = \hbar$, measuring the muon lifetime gives a direct measure of G_F.

1.1.2   Determination of $|V_{us}|$

A charged current decay diagram for strange quark decay is shown in Figure 2. Here the CKM element V_{us} is present. The decay rate is given by a formula similar to Equation (7), with the muon mass replaced by the s-quark mass and an additional factor
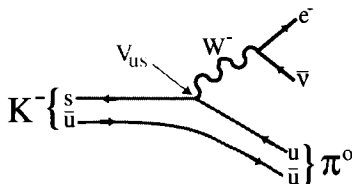
Figure 2. *Semileptonic K^- decay diagram.*

of $|V_{us}|^2$. Two complications arise since we are now measuring a decay process involving hadrons, $K^- \to \pi^0 e^- \bar{\nu}$ rather than elementary constituents. One is that the s-quark mass is not well defined and the other is that we must make corrections for the probability that the $\bar{u}$-spectator-quark indeed forms a π^0 with the u-quark from the s-quark decay. These considerations will be discussed in greater detail in the semileptonic B decays section. Fortunately there are theoretical calculations that allow for a relatively precise measurement because they deal with hadron rather than quark masses and have good constraints on the form-factors. Using these models we have $\lambda = V_{us} = 0.2205 \pm 0.0018$.

1.1.3 Using semileptonic B decays to determine V_{cb} and V_{ub}

Semileptonic b decays arise from a similar diagram to Figure 2, where the b quark replaces the s quark. In this case the b quark can decay into either the c quark or the the u quark, so we can use these decays to determine V_{cb} and V_{ub} providing we have three ingredients:

- B lifetimes,
- relevant B branching fractions,
- a theory or model to take care of hadronic physics.

1.2 Lifetime measurements

The "b-lifetime" was first measured at the 30GeV e^+e^- colliders PEP and PETRA, where b-quarks were produced via e^+e^- annihilation into $b\bar{b}$ pairs. Experiments at these colliders measured the average lifetime of all b-hadron species using the distribution of the "impact parameter", which is the minimum distance of approach of a track to the primary production vertex. This distance is related to the lifetime (Atwood 1994).

A more direct measurement would be to measure the actual decay distance L and the momentum of the b hadron. Since $L = \gamma\beta ct$, where t is the decay time of the individual particle (also called the proper time), and $\beta = p/E$ and $\gamma = E/m_b$, the distribution of decay times t can be determined. Events will be distributed exponentially in t as $e^{-t/\tau}$, with τ being the lifetime. Uncertainties come from errors on L, the b hadron momentum, and contributions from non b-quark backgrounds.

Precision lifetimes of individual b-flavored hadrons have been measured at LEP where the production process is $e^+e^- \to Z^0 \to b\bar{b}$ and at CDF in 1.8TeV $p\bar{p}$ collisions. Large samples of semileptonic decays have been used to determined the B^0 and B^- lifetimes. Note that the CPT theorem guarantees that the lifetime of the anti-particles is the same
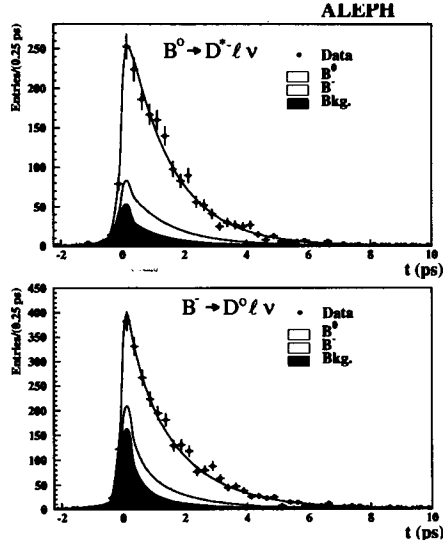
Figure 3. *Proper time distributions of exclusive semi-leptonic decays from ALEPH.*

as the particles. The decay distributions for two semileptonic B decay channels are shown in Figure 3. The $B^0 \to D^{*-}\ell^+\nu$ channel has mostly signal with some background from B^- decays and other backgrounds as indicated in the figure. It takes a great deal of careful work to accurately estimate these background contributions. The clear exponential lifetime shapes can be seen in these plots. Some data has also been obtained using purely hadronic final states (Sharma 1994).
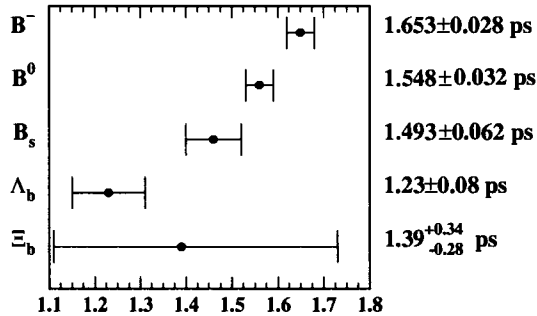
Figure 4. *Measured lifetimes of different B species.*

A summary of the lifetimes of specific b-flavored hadrons is given in Figure 4 (Groom 2001). Note that the ratio of B^+ to B^0 lifetimes is 1.074 ± 0.028, a 2.6σ difference from unity, which is quite significant. Also, the Λ_b lifetime is much shorter than the B^0 lifetime. According to proponents of the Heavy Quark Expansion model, there should be at most a 10% difference between them (Bigi 1997). To understand lifetime differences we must first analyze hadronic decays.

1.3   *B* decay mechanisms

Figure 5 shows sample diagrams for B decays. Semileptonic decays are depicted in Figure 5(a), when the virtual W^- materialises as a lepton-antineutrino pair. The name "semileptonic" is given, since there are both hadrons and leptons in the final state. The leptons arise from the virtual W^-, while the hadrons come from the coupling of the spectator anti-quark with either the c or u quark from the b quark decay. Note that the B is massive enough that all three lepton species can be produced. The simple spectator diagram for hadronic decays (Figure 5(a)) occurs when the virtual W^- materialises as a quark-antiquark pair, rather than a lepton pair. The terminology *simple spectator* comes
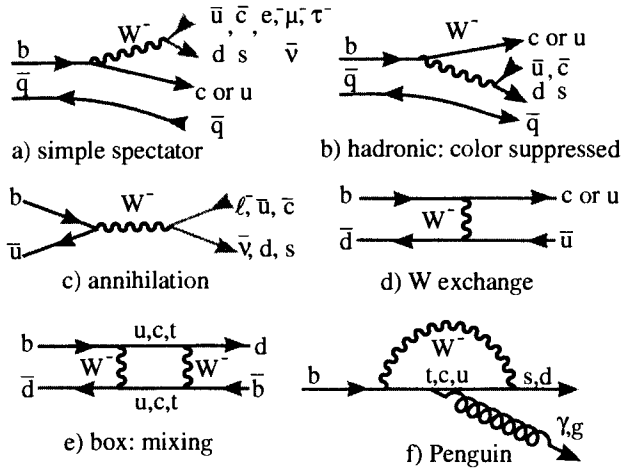
Figure 5. *Various mechanisms for B meson decay.*

from viewing the decay of the b quark, while ignoring the presence of the *spectator* antiquark. If the colors of the quarks from the virtual W^- are the same as the initial b quark, then the color suppressed diagram, Figure 5(c), can occur. While the amount of color suppression is not well understood, a good first order guess is that these modes are suppressed in amplitude by the color factor 1/3 and thus in rate by 1/9, with respect to the non-color suppressed spectator diagram.

The annihilation diagram shown in Figure 5(c) occurs when the b quark and spectator anti-quark find themselves in the same space-time region and annihilate by coupling to a virtual W^-. The probability of such a wave function overlap between the b and $\bar{u}$-quarks is proportional to a numerical factor called f_B. The decay amplitude is also proportional to the coupling V_{ub}. The mixing and penguin diagrams will be discussed later.

Each diagram contributes differently to the decay width of the individual species, although the simple spectator diagram is always expected to be the dominant one. Note that for b baryons there are even more diagrams than those shown in Figure 5. Currently there is no direct evidence for diagrams (c) and (d), although (c) is expected to be responsible for the purely leptonic decay $B^- \to \tau^- \bar{\nu}$.

The semileptonic decay width, Γ_{sl}, is defined as the decay rate in units of inverse seconds into a hadron (or hadrons) plus a lepton-antineutrino pair. Decay rates can also

be expressed in units of MeV by multiplying by $\hbar$. The width, Γ_{sl}, is related to the semileptonic branching ratio $\mathcal{B}_{sl}$ and the lifetime τ by:

$$\Gamma_{sl} = \mathcal{B}_{sl} \cdot \Gamma_{total} = \mathcal{B}_{sl}/\tau. \tag{8}$$

The semileptonic width should be equal for all b species. This is true for D^0 and D^+ mesons, even though their lifetimes differ by more than a factor of two. Thus, it is differences in the hadronic widths among the different b hadrons that drive the lifetime differences.

Let us now consider the case of $\overline{B}^0$ and B^- lifetime differences. There is some indication that the $\overline{B}^0$ has a shorter lifetime, which would imply that there are more decay channels available. Figure 6 shows the color allowed and color suppressed decay diagrams for *two-body* decays into a ground-state charmed meson and a π^-. The color suppressed diagram only exists for the B^-. The relative rate

$$\frac{\Gamma(B^- \to D^0\pi^-)}{\Gamma(\overline{B}^0 \to D^+\pi^-)} = 1.8 \pm 0.3, \tag{9}$$

and the same is true for all other similar two-body channels, such as $D^*\rho^-$. Thus we would expect, if most B decays are given by these diagrams that the B^- would have a shorter lifetime than the $\overline{B}^0$, opposite to what the data suggests. An explanation is that this ratio reverses for higher multiplicity decays, but this is an interesting discrepancy that needs to be kept in mind.
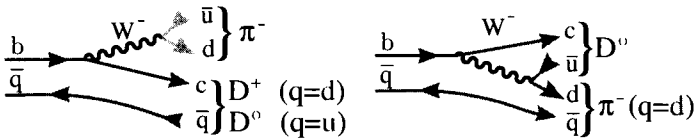
Figure 6. *(left) Spectator diagram for $\overline{B}^0 \to D^+\pi^-$ and $B^- \to D^0\pi^-$. (right) Color suppressed spectator diagram for $B^- \to D^0\pi^-$ only.*

2   Semileptonic B decays

2.1   Formalism of exclusive semileptonic B decays

The same type of semileptonic charged current decays used to find V_{us} are used to find V_{cb} and V_{ub}. The basic diagram is shown in Figure 5(a). We can use either inclusive decays, where we look only at the lepton and ignore the hadronic system at the lower vertex, or exclusive decays where we focus on a particular single hadron. Theory currently can predict either the inclusive decay rate, or the exclusive decay rate when there is only a single hadron in the final state.

Now let us briefly go through the mathematical formalism of semileptonic decays. We start with pseudoscalar B to pseudoscalar m transitions. The decay amplitude is given by Grinstein (1986) and Gilman (1990):

$$A(\bar{B} \to me^-\bar{\nu}) = \frac{G_F}{\sqrt{2}}V_{ij}L^\mu H_\mu, \tag{10}$$

where

$$L^\mu = \bar{u}_e \gamma^\mu \left(1 - \gamma_5\right) v_\nu, \tag{11}$$

and

$$H_\mu = \langle m | J^\mu_{\text{had}}(0) | \overline{B} \rangle = f_+(q^2)(P + p)_\mu + f_-(q^2)(P - p)_\mu, \tag{12}$$

where q^2 is the four-momentum transfer squared between the B and the m, and P, p are four-vectors of the B, m respectively. H_μ is the most general form of the hadronic matrix element. It is written in terms of the unknown f functions that are called "form-factors." It turns out that the term multiplying the $f_-(q^2)$ form-factor is the mass of lepton squared. Thus for electrons and muons (but not τ's), the decay width is given by

$$\frac{d\Gamma_{\text{sl}}}{dq^2} = \frac{G_F^2 |V_{ij}|^2 K^3}{24\pi^2} |f_+(q^2)|^2, \tag{13}$$

where

$$K = \frac{1}{2M_B} \left[\left(M_B^2 + m^2 - q^2\right)^2 - 4m^2 M_B^2 \right]^{1/2} \tag{14}$$

is the momentum of the particle m (with mass m) in the B rest frame. In principle, $d\Gamma_{\text{sl}}/dq^2$ can be measured over all q^2. Thus the shape of $f_+(q^2)$ can be determined experimentally. However, the normalisation, $f_+(0)$ must be obtained from theory, for V_{ij} to be measured. In other words,

$$\Gamma_{\text{SL}} \propto |V_{ij}|^2 |f_+(0)|^2 \frac{1}{\tau_B} \int K^3 g(q^2) dq^2, \tag{15}$$

where $g(q^2) = f_+(q^2)/f_+(0)$. Measurements of semileptonic B decays give the integral term, while the lifetimes are measured separately, allowing the product $|V_{ij}|^2 |f_+(0)|^2$ to be experimentally determined.

For pseudoscalar to vector transitions there are three independent form-factors whose shapes and normalisations must be determined (Richman 1995).

2.2 Measurement of $|V_{cb}|$

2.2.1 Heavy quark effective theory and $\bar{B} \rightarrow D^* \ell^- \bar{\nu}$

We can use measurements of exclusive B decays, coupled with "Heavy Quark Effective Theory" (HQET) to find V_{cb} (Isgur 1994). We start with a quick introduction to HQET. It is difficult to solve QCD at long distances, but it is possible at short distances. Asymptotic freedom, the fact that the strong coupling constant α_s becomes weak in processes with large q^2, allows perturbative calculations. Large distances are of the order $\sim 1/\Lambda_{\text{QCD}} \sim 1$ fm, since Λ_{QCD} is about 0.2GeV. Short distances, on the other hand, are of the order of the quark Compton wavelength; $\lambda_Q \sim 1/m_Q$ equals 0.04 fm for the b quark and 0.13fm for the c quark.

For hadrons, on the order of 1fm, the light quarks are sensitive only to the heavy quark's color electric field, not the flavor or spin direction. Thus, as $m_Q \rightarrow \infty$, hadronic systems which differ only in flavor or heavy quark spin have the same configuration of

their light degrees of freedom. The following two predictions follow immediately (the actual experimental values are shown below):

$$m_{B_s} - m_{B_d} \;=\; m_{D_s} - m_{D^+} \tag{16}$$
$$(90.2 \pm 2.5) \text{ MeV} \qquad (99.2 \pm 0.5) \text{ MeV} \;,$$

$$m_{B^*}^2 - m_B^2 \;=\; m_{D^*}^2 - m_D^2. \tag{17}$$
$$0.49 \text{ GeV}^2 \qquad 0.55 \text{ GeV}^2.$$

The agreement is quite good but not exceptional. Since the charmed quark is not that heavy, there is some heavy quark symmetry breaking. This must be accounted for in quantitative predictions, and can probably explain the discrepancies above. The basic idea is that if you replace a b quark with a c quark moving at the same **velocity**, there should only be small and calculable changes.

In lowest order HQET there is only one form-factor function F which is a function of the Lorentz invariant four-velocity transfer ω, where

$$\omega = \frac{M_B^2 + M_{D^*}^2 - q^2}{2 M_B M_{D^*}}. \tag{18}$$

The point ω equals one corresponds to the situation where the B decays to a D^* which is at rest in the B frame. Here the "universal" form-factor function $F(\omega)$ has the value, $F(1) = 1$, in lowest order. This is the point in phase space where the b quark changes to a c quark with zero velocity transfer. The idea is to measure the decay rate at this point, since we know the value of the form-factor, namely unity, and then apply the hopefully small and hopefully well understood corrections.

In HQET, the form factor F can also be applied to the analysis of $\bar{B} \to D \ell^- \bar{\nu}$, but the overall decay rate for the pseudoscalar final state is only 40% of $D^* \ell^- \bar{\nu}$,and the decay rate vanishes at $\omega = 1$ much faster, making the measurement more difficult.

2.2.2 Detection of $B \to D^* \ell \nu$

Since this is a semileptonic final state containing a missing neutrino, the decay cannot be identified or reconstructed by merely measuring the 4-vectors of the final state particles. One technique used in the past relies on evaluating the missing mass (MM) where

$$\begin{aligned} MM^2 &= (E_B - E_{D^*} - E_\ell)^2 - (\mathbf{p}_B - \mathbf{p}_{D^*} - \mathbf{p}_\ell)^2 \\ &= M_B^2 + M_{D^*}^2 + M_\ell^2 - 2E_B \cdot (E_{D^*} + E_\ell) + 2E_{D^*} E_\ell \\ &\quad -2\mathbf{p}_{D^*}\cdot\mathbf{p}_\ell + 2\mathbf{p}_B\cdot(\mathbf{p}_{D^*} + \mathbf{p}_\ell). \end{aligned} \tag{19}$$

For experiments using $e^+e^- \to \Upsilon(4S) \to B\bar{B}$, the B energy, E_B, is set equal to the beam energy, E_{beam}, and all quantities are known except the angle between the B direction and the sum of the D^* and lepton 3-vectors (the last term). A reasonable estimate of MM^2 is obtained by setting this term to zero. The signal for the $D^* \ell \nu$ final state should appear at the neutrino mass, namely at $MM^2 = 0$. In an alternative technique MM^2 is set to zero and the angle between the B momentum and the sum of the D^* and lepton 3-vectors is evaluated as

$$\cos\left(\Theta_{B \cdot D^* \ell}\right) = \frac{2E_B(E_{D^*} + E_\ell) - M_B^2 - M_{D^* \ell}^2}{2 \, |\mathbf{p}_B| \, \big|(\mathbf{p}_{D^*} + \mathbf{p}_\ell)\big|}, \tag{20}$$

where $M_{D^*\ell}$ indicates the invariant mass of the D^*-lepton combination.

A Monte-Carlo simulation of $\cos\left(\Theta_{B.D^*\ell}\right)$ is given for the final state of interest and for the main background reaction in Figure 7. For the correct final state only a few events are outside the "legal" region of ± 1, while when there are extra pions in the final state the shape changes and many events are below -1.
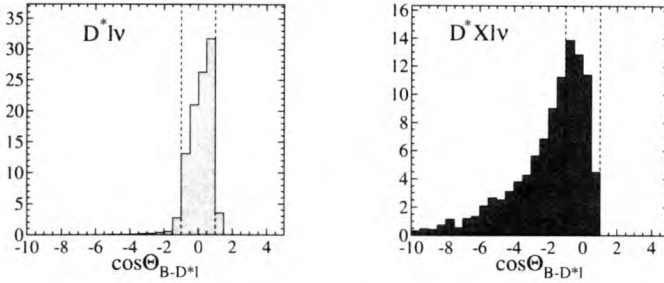
Figure 7. *The cosine of the angle between the B momentum vector and the sum of D^* and lepton momentum vectors for (left) the final state $D^*\ell\nu$ and (right) $D^*X\ell\nu$, where X refers to an additional pion.*

Recent CLEO data has been analyzed with such a technique. The data in two specific ω bins is shown in Figure 8. The final result for all ω bins is shown in Figure 9. The result is characterised by both a value for $F(1)|V_{cb}|$ and a shape parameter ρ^2.
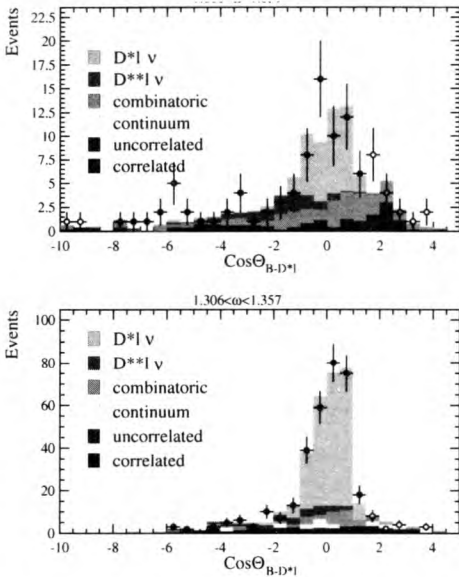
Figure 8. *The signal and background contributions in two different ω bins for the final state $D^*\ell\nu$.*
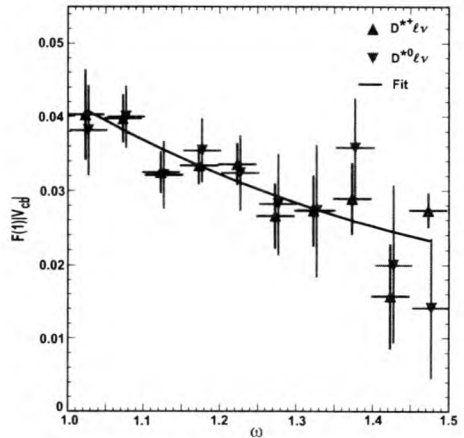
Figure 9. *The CLEO results for $D^*\ell\nu$ for both $\overline{B}^0$ and B^-. The curve is a fit to a shape suggested in (Caprini 1998).*
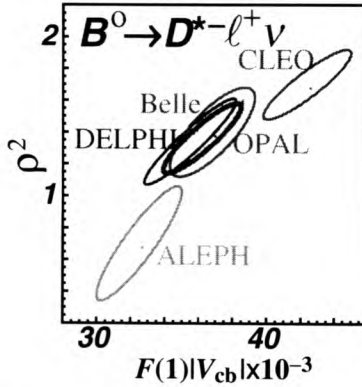
Figure 10. *The correlation between the slope parameter ρ^2 and $F(1)|V_{cb}|$. The contours are for a change in the fit χ^2 of one unit.*

2.2.3   Evaluation of V_{cb} using $B \to D^*\ell\nu$

Figure 10 and Table 1 give recent experimental results on exclusive $B \to D^*\ell\nu$ decays. The CLEO results are not in particularly good agreement with the rest of the world including the BELLE results.

Experiment	$F(1)|V_{cb}|$ $(\times 10^{-3})$	ρ^2
ALEPH (Buskulic 1997)	$33.0 \pm 2.1 \pm 1.6$	$0.74 \pm 0.3 \pm 0.4$
BELLE (Tajima 2001)	$35.4 \pm 1.9 \pm 1.9$	$1.35 \pm 0.17 \pm 0.18$
CLEO (Heltsley 2001)	$42.2 \pm 1.3 \pm 1.8$	$1.61 \pm 0.09 \pm 0.21$
DELPHI (Abreu 2001)	$34.5 \pm 1.4 \pm 2.5$	$1.2 \pm 0.1 \pm 0.4$
OPAL($\pi\ell$) (Abbiendi 2000)	$37.9 \pm 1.3 \pm 2.4$	$1.2 \pm 0.2 \pm 0.4$
OPAL (Abbiendi 2000)	$37.5 \pm 1.7 \pm 1.8$	$1.4 \pm 0.2 \pm 0.2$
Average	37.8 ± 1.4	1.37 ± 0.13

Table 1. *Modern determinations of $F(1)|V_{cb}|$ using $B \to D^*\ell^-\overline{\nu}$ decays.*

To extract the value of $|V_{cb}|$ we have to determine the corrections to $F(1)$ that lower its value from unity. The corrections are of two types: quark mass, characterised as some coefficient times Λ_{QCD}/m_Q, and hard gluon, characterised as η_A. The value of the form-factor can then be expressed as (Neubert 1996)

$$F(1) = \eta_A \left(1 + 0 \cdot \Lambda_{QCD}/m_Q + c_2 \cdot (\Lambda_{QCD}/m_Q)^2 + ...\right) = \eta_A(1 + \delta). \qquad (21)$$

The zero coefficient in front of the $1/m_Q$ term reflects the fact that the first order correction in quark mass vanishes at ω equals one. This is called Luke's Theorem (Luke 1990). Recent estimates for η_A and δ are 0.967 ± 0.007 and -0.55 ± 0.025, respectively. The value predicted for $F(1)$ is then 0.91 ± 0.05. This is the conclusion of the PDG review done by

Artuso and Barberio (Artuso 2001). There has been much controversy surrounding the theoretical prediction of this number. In the future Lattice-Gauge Theory calculations will presumably become more accurate when unquenched calculations become possible. Current lattice calculations are done in the quenched approximation, without including light quark loops. They give $F(1) = 0.913^{+0.024}_{-0.017} \pm 0.016^{+0.003+0.000+0.006}_{-0.014-0.016-0.014}$, where the uncertainties come respectively, from statistics and fitting, matching lattice gauge theory to QCD, lattice spacing dependence, light quark mass effects and the quenching approximation. (Hashimoto 2001).

Using the Artuso-Barberio value for $F(1)$ we have

$$|V_{cb}| = (41.5 \pm 1.5 \pm 2.3) \times 10^{-3}, \tag{22}$$

where the first error is experimental and the second the error on the calculation of $F(1)$.

2.2.4 $|V_{cb}|$ from inclusive semileptonic decays

The inclusive semileptonic branching ratio $b \to X\ell\nu$ has been measured by both CLEO and LEP to reasonable accuracy. CLEO finds $(10.49 \pm 0.17 \pm 0.43)\%$, while LEP has $(10.56 \pm 0.11 \pm 0.18)\%$. These are not quite the same quantities as the CLEO number is an average over B^- and B^0 only, while the LEP number is a weighted average over all b hadron species produced in Z^0 decays. Thus using the LEP number one should use the average "b quark" lifetime of 1.560 ± 0.014ps.

Using the Heavy Quark Expansion (HQE) model (Bigi 1997) we can relate the total semileptonic decay rate at the quark level to $|V_{cb}|$:

$$
\begin{aligned}
|V_{cb}|^2 &= h(\lambda_1, \lambda_2, \overline{\Lambda}) \times \Gamma(b \to c\ell\nu) \\
&= h(\lambda_1, \lambda_2, \overline{\Lambda}) \times \mathcal{B}(b \to c\ell\nu) \times \frac{1}{\tau_b},
\end{aligned}
\tag{23}
$$

where $\mathcal{B}(b \to c\ell\nu)$ is the inclusive semileptonic branching ratio minus a small $b \to u\ell\nu$ component. To be precise, it is the decay of a B meson to a lepton-antineutrino pair plus any charmed hadron. τ_b is the lifetime of that particular meson or average lifetime of the combination of the b-flavored hadrons used in the analysis, suitably weighted. In the HQE model the semileptonic rate is described to order $(\Lambda_{\text{QCD}}/m_b)^2$ by three parameters:

- λ_1 is the kinetic energy of the residual motion of the b quark in the hadron:

$$\lambda_1 = \frac{M_B}{2} \left\langle B(v) \left| \overline{h_v}(iD)^2 h_v \right| B(v) \right\rangle, \tag{24}$$

- λ_2 is the chromo-magnetic coupling of the b quark spin to the gluon field:

$$\lambda_2 = -\frac{M_B}{2} \left\langle B(v) \left| \overline{h_v} \frac{g}{2} \cdot \sigma^{\mu\nu} G_{\mu\nu} h_v \right| B(v) \right\rangle, \tag{25}$$

- $\overline{\Lambda}$ is the strong interaction coupling, given in terms of $\overline{M}_B = (M_B + 3M_{B^*})/4$, the spin averaged B meson mass, as

$$\overline{\Lambda} = \overline{M}_B - m_b + \frac{\lambda_1}{2M_B}. \tag{26}$$

These parameters are further related by

$$M_B = m_b + \overline{\Lambda} - \frac{(\lambda_1 + 3\lambda_2)}{2m_b}$$

$$M_{B^*} = m_b + \overline{\Lambda} - \frac{(\lambda_1 - \lambda_2)}{2m_b}. \tag{27}$$

These relations allow us to determine λ_2 from the $M_{B^*} - M_B$ mass splitting as $0.12\,\text{GeV}^2$. The function $h(\lambda_1, \lambda_2, \overline{\Lambda})$ can be calculated from the Heavy Quark Expansion (HQE) model. This involves both perturbative and non-perturbative pieces.

Although we will go through this derivation of V_{cb}, there is a disturbing assumption of quark-hadron duality; the idea of duality is that if you integrate over enough exclusive charm bound states and enough phase space, the inclusive hadronic result will match the quark level calculation. However, we do not know what size is the uncertainty associated with the duality *assumption*. In fact, Isgur said "I identify a source of Λ_{QCD}/m_Q corrections to the assumption of quark-hadron duality in the application of heavy quark methods to inclusive heavy quark decays. These corrections could substantially affect the accuracy of such methods in practical applications and in particular compromise their utility for the extraction of the CKM matrix element V_{cb}" (Isgur 1999).

Let us move to the details of the calculation. In one implementation λ_1 and $\overline{\Lambda}$ are derived; the relationship between the inclusive $b \to c\ell\nu$ branching fraction (about 99% of $b \to X\ell\nu$) is given as

$$|V_{cb}| = 0.0411\sqrt{\frac{\mathcal{B}(b \to X_c\ell\nu)}{0.105}}\sqrt{\frac{1.55\ \text{ps}}{\tau_b}}\left(1 - 0.012\frac{\mu_\pi^2 - 0.5\ \text{GeV}^2}{0.1\ \text{GeV}^2}\right) \tag{28}$$

$$\times \left(1 \pm 0.015_{\text{pert.}} \pm 0.010_{m_b} \pm 0.012_{1/m_Q^3}\right), \tag{29}$$

where μ_π^2 is the negative of λ_1 modulo QCD corrections and is taken as $(0.5 \pm 0.1)\ \text{GeV}^2$ (Bigi 1997). This leads to a value of

$$|V_{cb}| = (40.7 \pm 0.5 \pm 2.4) \times 10^{-3} \tag{30}$$

from the LEP data alone, with a similar value from CLEO. The first error contains the statistical and systematic error from the experiments while the second error contains an estimate of the theory error from sources other than duality.

In another implementation the parameters λ_1 and $\overline{\Lambda}$ are obtained from data. Here we use the HQE formula

$$\Gamma_{\text{sl}} = \frac{G_F^2 |V_{cb}|^2 M_B^5}{192\pi^3} \times 0.369 \times \tag{31}$$

$$\left[1 - 1.54\frac{\alpha_s}{\pi} - 1.65\frac{\overline{\Lambda}}{M_B}\left(1 - 1.087\frac{\alpha_s}{\pi}\right) - 0.95\frac{\overline{\Lambda}^2}{M_B^2} - 3.18\frac{\lambda_1}{M_B^2} + 0.02\frac{\lambda_2}{M_B^2}\right].$$

Determining λ_1 and $\overline{\Lambda}$ can be accomplished by measuring, for example, the "moments" of the hadronic mass produced in $b \to c\ell\nu$ decays. The first moment is defined as the deviation from the D mass (M_D) as $\langle M_X^2 - \overline{M}_D^2 \rangle$ and the second moment as $\langle \left(M_X^2 - \overline{M}_D^2\right)^2 \rangle$.

It is also possible to use the first and second moments of the lepton energy distribution in these decays, or moments of the photon energy in the process $b \to s\gamma$. In fact any two distributions can be used; in practice it will be critical to use all of them to try and ascertain if any violations of quark-hadron duality appear and to check that terms of order $(\Lambda_{\rm QCD}/m_b)^3$ are not important.

CLEO has used the first and second moments of the hadron mass in $b \to c\ell\nu$ decays. They find the M_X distributions by using missing energy and momentum in the event to define the ν four-vector. Then detecting only the lepton and requiring it to have a momentum above 1.5GeV/c, they calculate:

$$
\begin{aligned}
M_X^2 &= (E_B - E_\ell - E_\nu)^2 - (\mathbf{p}_B - \mathbf{p}_\ell - \mathbf{p}_\nu)^2 \qquad (32) \\
&= M_B^2 + M_{\ell\nu}^2 - 2E_B(E_\ell + E_\nu) + 2\mathbf{p}_B \cdot (\mathbf{p}_\ell - \mathbf{p}_\nu) \\
&\approx M_B^2 + M_{\ell\nu}^2 - 2E_B E_{\ell\nu}, \qquad (33)
\end{aligned}
$$

where $M_{\ell\nu}$ is the invariant mass of the lepton-neutrino pair. The measured M_X^2 distribution is shown in Figure 11. We do not see distinct peaks at the mass of the D and D^* mesons because ignoring the last term in Equation (32) causes poor resolution. This term must be ignored, because using this technique, we do not know the direction of the B meson.
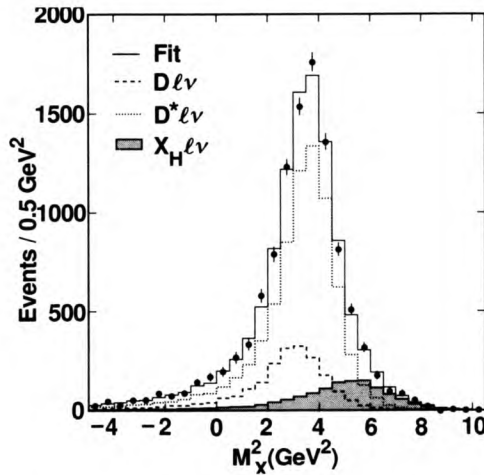
Figure 11. *The mass distribution for $b \to c\ell\nu$ events from CLEO. X_H indicates an additional pion plus a D or D^*.*

CLEO finds values of the first and second moments of $(0.287\pm0.023\pm0.061)$ GeV^2 and $(0.712\pm0.056\pm0.176)$GeV^4, respectively. These lead to the determination of λ_1, $\bar{\Lambda}$ and V_{cb} shown in Figure 12. Later we will see a different determination using $b \to s\gamma$ (Section 5.3.1).

In summary the exclusive measurements of V_{cb} are to be trusted while the inclusive determination, though consistent, has an unknown source of systematic error and should not be used now.
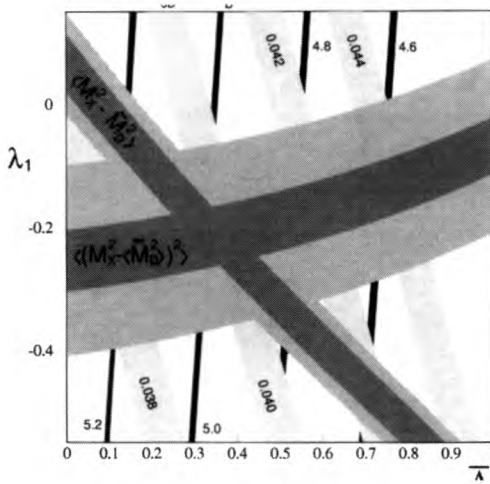
Figure 12. *Constraints in λ_1 versus $\overline{\Lambda}$ from CLEO measurements of first and second hadronic moments in semileptonic decay. The darker band gives experimental uncertainties alone, while the lighter outer band also includes uncertainties from unknown 3rd order theoretical parameters. Bands of constant V_{cb} and b quark mass, m_b, are also shown, where the band widths represents theoretical uncertainties due to unknown 3rd order parameters.*

2.3   Measurement of $|V_{ub}|$

This is a heavy to light quark transition where HQET cannot be used directly as in finding V_{cb}. Unfortunately the theoretical models that can be used to extract a value from the data do not currently give precise predictions.

Three techniques have been used. The first measurement of V_{ub} done by CLEO (Fulton 1990) and subsequently confirmed by ARGUS (Albrecht 1990), used only leptons which were more energetic than those that could come from $b \to c\ell^-\bar{\nu}$ decays. These "endpoint leptons" can occur $b \to c$ background free at the $\Upsilon(4S)$, because the B's are almost at rest. The CLEO data are shown in Figure 13. Since the lepton momentum for $B \to D\ell\nu$ decays is cut off by phase space, this data provides incontrovertible evidence for $b \to u\ell\nu$ decays.

Unfortunately, there is only a small fraction of the $b \to u\ell^-\bar{\nu}$ lepton spectrum that can be seen this way, leading to model dependent errors. The models used are either inclusive predictions, sums of exclusive channels, or both: see Isgur (1995), Bauer (1989), Körner (1989), Melikhov (1996), Altarelli (1982) and Ramirez (1990). The average among the models is $|V_{ub}/V_{cb}| = 0.079 \pm 0.006$, without a model dependent error. These models differ by at most 11%, making it tempting to assign a $\pm 6\%$ error. However, there is no quantitative way of estimating the error.

ALEPH (Barate 1999), L3 (Acciarri 1998) and DELPHI (Abreu 2000) isolate a class of events where the hadron system associated with the lepton is enriched in $b \to u$ and thus depleted in $b \to c$. They define a likelihood that hadron tracks come from b decay by using a large number of variables including, vertex information, transverse momentum, not being a kaon etc.. Then they require the hadronic mass to be less than 1.6GeV, which greatly reduces $b \to c$, since a completely reconstructed $b \to c$ decay has a mass greater
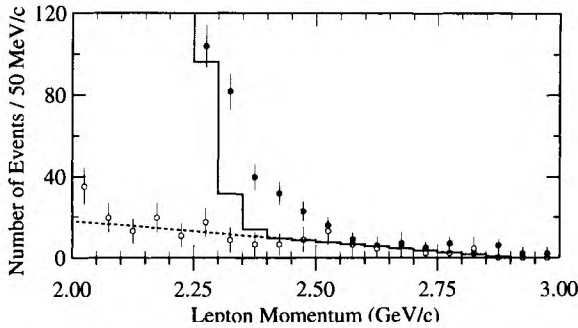
Figure 13. *Sum of inclusive electron and muon distributions from CLEO. The solid points are data taken on the peak of the $\Upsilon(4S)$ while the open circles are data taken on the continuum 30MeV below the resonance(suitably normalised). The dashed line is a fit to the continuum data and the solid line is the predicted curve from $b \to c\ell\nu$ dominated by $B \to D\ell\nu$ near the end of the allowed lepton spectrum.*
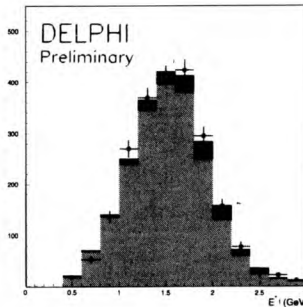
Figure 14. *The lepton energy distribution in the B rest frame from DELPHI. The data have been enriched in $b \to u$ events, and the mass of the recoiling hadronic system is required to be below 1.6GeV. The points indicate data, the light shaded region, the fitted background and the dark shaded region, the fitted $b \to u\ell\nu$ signal.*

than that of the D (1.83GeV). They then examine the lepton energy distribution, shown in Figure 14 for DELPHI.

The average of all three results is $|V_{ub}| = (4.13^{+0.42+0.43+0.24}_{-0.47-0.48-0.25}\pm0.20)\times10^{-3}$, resulting in a value for $|V_{ub}/V_{cb}| = 0.102\pm0.018$, using $|V_{cb}| = 0.0405\pm0.0025$. I have several misgivings about this result. First of all the experiments have to understand the systematic errors very well. To understand semileptonic b and c decays and thus find their $b \to u\ell\nu$ efficiency, they employ different models and Monte Carlo manifestations of these models. To find the error they take half the spread that different models give. This alone may be a serious underestimate. Secondly they use one model, the HQE model, to translate their measured rate to a value for $|V_{ub}|$. This model assumes duality, and there are no successful experimental checks: The model fails on the Λ_b lifetime prediction. Furthermore, the quoted theoretical error, even in the context of the model, has been estimated by Neubert

to be much larger at 10% (Neubert 2000). Others have questioned the effect of the hadron mass cut and estimate 10–20% errors due to this alone (Bauer 2001).

It may be possible to use the spectrum of photons in $b \to s\gamma$ to reduce the theoretical error in the endpoint lepton method or to make judicious cuts in q^2 instead of hadronic mass to help reduce the theoretical errors: see Wise (2001).

The third method uses exclusive decays. CLEO has measured the decay rates for the exclusive final states $\pi\ell\nu$ and $\rho\ell\nu$ (Alexander 1996). The data are shown in Figure 15. The model of Körner and Schuler (KS) is ruled out by the measured ratio of ρ/π. Other models include those of Isgur (1995, 1989), Wirbel (1985), Bauer (1989), Korner (1988), Melikhov (1996), Altarelli (1982) and Ramierz (1990). CLEO has presented an updated analysis for $\rho\ell\nu$ where they have used several different models to evaluate their efficiencies and extract V_{ub}. These theoretical approaches include quark models, light cone sum rules (LCRS), and lattice QCD. The CLEO values are shown in Table 2.
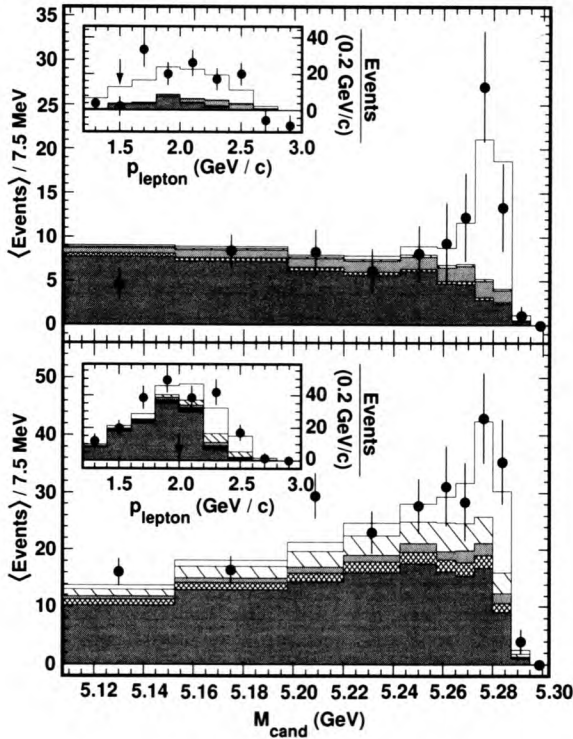
Figure 15. *The B candidate mass distributions and the signal bin lepton momentum spectra (insets) for the pion modes (top) and the sum of ρ and ω (vector) modes (bottom). The points are the data after continuum and fake background subtractions; the dark-shaded, cross-hatched and unshaded histograms are $b \to cX$, $b \to u\ell\nu$ feed-down and signal, respectively. For the π (vector) modes, the light-shaded and hatched histograms are $\pi \to \pi$ (vector$\to$ vector) and vector $\to \pi$ ($\pi \to$ vector) cross-feed, respectively (charge final states can feed neutral and vice-versa). The histogram normalisations are from the nominal fit. The arrows indicate the momentum cuts.*

Model	V_{ub} $(\times 10^{-3})$
ISGW2 (Isgur 1989)	$3.23 \pm 0.14^{+0.22}_{-0.29}$
Beyer/Melikhov (Beyer 1998)	$3.32 \pm 0.15^{+0.21}_{-0.30}$
Ligeti/Wise (Ligeti 1996)	$2.92 \pm 0.13^{+0.19}_{-0.26}$
LCSR (Ball 1998)	$3.45 \pm 0.15^{+0.22}_{-0.31}$
UKQCD (Debbio 1998)	$3.32 \pm 0.14^{+0.21}_{-0.30}$

Table 2. *Values of $|V_{ub}|$ using $B \to \rho \ell^- \overline{\nu}$ and some theoretical models.*

The uncertainties in the quark model calculations (first three in the table) are guessed to be 25–50% in the rate. The Ligeti/Wise model uses charm data and SU(3) symmetry to reduce the model dependent errors. The other models estimate their errors at about 30–50% in the rate, leading conservatively to a 25% error in $|V_{ub}|$. Note that the models differ by 18%, but it would be incorrect to assume that this spread allows us to take a smaller error. It may be that the models share common assumptions, e.g. the shape of the form-factors. At this time it is prudent to assign a 25% model dependent error realising that the errors in the models cannot be averaged. The fact that the models do not differ much allows us to comfortably assign a central value $|V_{ub}| = (3.25 \pm 0.14^{+0.22}_{-0.29} \pm 0.80) \times 10^{-3}$, and a derived value $|V_{ub}/V_{cb}| = 0.08 \pm 0.02$.

Lattice QCD has predicted form-factors and resulting rates for the exclusive semileptonic final states $\pi \ell \nu$ and $\rho \ell \nu$ (Sachrajda 1999) in the quenched approximation. These calculations require the momentum of the final-state light meson to be small in order to avoid discretisation errors. This means we only obtain results at large values of the invariant four-momentum transfer squared, q^2. Figure 16 shows the predictions of the $B \to \rho \ell \nu$ width as a function of q^2. Note that the horizontal scale is highly zero suppressed. The region marked "phase space only" is not calculated but estimated using a phase space extrapolation from the last lattice point.
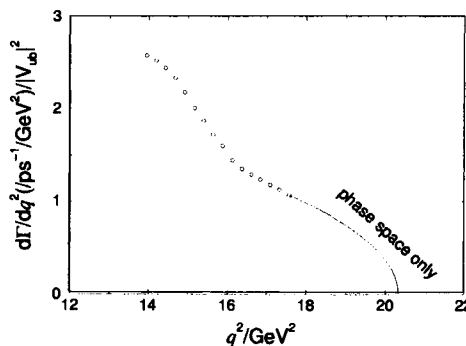
Figure 16. *The UKQCD lattice calculation for $B \to \rho \ell \nu$ shown as circles. The line is an estimate.*

The integral over the region for $q^2 > 14 \text{GeV}^2$ gives a rate of $\Delta\Gamma = 8.3 |V_{ub}|^2 \text{ ps}^{-1}\text{-GeV}^2$. CLEO measurements in the same interval give $(7.1 \pm 2.4) \times 10^{-5} \text{ps}^{-1}\text{-GeV}^2$, yielding a value for $V_{ub} = (2.9 \pm 0.5) \times 10^{-3}$ (Sachrajda 1999). Ultimately unquenched lattice calculations when coupled with more precise data will yield a much better value for V_{ub}.

3   Facilities for *b* studies

3.1   *b* production mechanisms

Although most of what is known about *b* physics has been obtained from e^+e^- colliders
operating either at the $\Upsilon(4S)$ or at the Z^0 pole, interesting information is now appearing
from the hadron collider experiments, CDF and D0, which were designed to look for
considerably higher energy phenomena. The appeal of hadron colliders arises mainly
from the large measured *b* production cross-sections. At the $\Upsilon(4S)$ the total *b* production
cross-section is only 1.05 nb, whereas at the FNAL Tevatron collider, with 1.8TeV in the
$p\bar{p}$ center-of-mass, the *b* cross-section has been measured as $\sim 100\mu b$. It is expected to be
about five times higher at the LHC (Artuso 1994).

The different production mechanisms of *b* quarks at various accelerators leads to dif-
ferent methods of measurement. Figure 17 shows the production of B^- and B^0 mesons at
the $\Upsilon(4S)$, while Figure 18 shows the production mechanism of the different *b* species at a
higher energy e^+e^- collider such as LEP. Figure 19 shows the production mechanisms for
a *b* or *c* quark at a hadron collider. The third order diagrams appear to be as important
as the second order diagrams and the overall theoretical calculation gives about 1/2 of
the measured value.
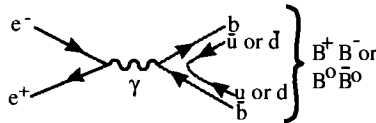
Figure 17. *B production at the* $\Upsilon(4S)$.
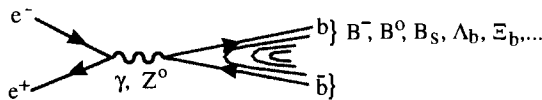
Figure 18. *b production in the continuum at* e^+e^- *colliders.*
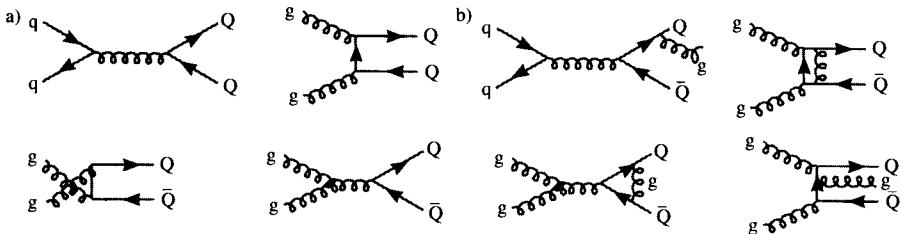
Figure 19. *Heavy quark, Q, production at hadron colliders. Left four diagrams are
second order in* α_s, *while right four are third order.*

3.2 Accelerators for *b* physics

Experiments on *b* decays started with CLEO and ARGUS using e^+e^- colliders operating at the $\Upsilon(4S)$. They were quickly joined by the PEP and PETRA machines operating around 30GeV. Table 3 lists some of the machines used to study *b* quarks in the last century (Artuso 1994).

Machine	Beams	Energy (GeV)	$\sigma(b)$	*b* fraction	$\mathcal{L}$ cm^{-2}s^{-1}	Total # *b*-pairs
CESR	e^+e^-	10.8	1.05 nb	0.25	1.3×10^{33}	9.8×10^6
DORIS	e^+e^-	10.8	1.05 nb	0.23	$\sim10^{31}$	0.4×10^6
PEP	e^+e^-	29	0.4 nb	0.09	3.2×10^{31}	
PETRA	e^+e^-	35	0.3 nb	0.09	1.7×10^{31}	
LEP	e^+e^-	91	9.2 nb	0.22	2.4×10^{31}	1.8×10^6
SLC	e^+e^-	91	9.2 nb	0.22	3.0×10^{30}	8.8×10^4
TEVATRON	$\bar{p}p$	1800	100 μb	0.002	3×10^{31}	

Table 3. *Machines used for b physics in the 20th century. The total number of $\bar{b}b$ pairs accumulated per experiment is also listed when known.*

In the year 2000 the PEP II and KEK-B storage rings began operation. These machines have separate e^- and e^+ magnet rings so they can operate at asymmetric energies; PEP II has beam energies of 9.0GeV and 3.1GeV, while KEK-B has energies of 8.0GeV and 3.5GeV. This allows the *B* mesons produced at the $\Upsilon(4S)$ to move with a velocity $\beta \sim 0.6$, which turns out to be very important in measurements of CP violation, since time integrated CP violation via mixing must be exactly zero due to the C odd nature of the $\Upsilon(4S)$. These machines are operating at very high luminosities. Current and future machines for *B* physics are listed in Table 4. The CDF and D0 experiments will continue at the Tevatron with higher luminosities. CDF has already made significant contributions including studies of *b* production, lifetimes and the discovery of the B_c meson (Abe 1998).

Machine	Exp.	Beam	Energy (GeV)	$\sigma(b)$	*b* fraction	$\mathcal{L}$(Design) cm^{-2}s^{-1}	Interactions per crossing
PEP II	BABAR	e^+e^-	10.8	1.05 nb	1/4	3×10^{33}(†)	$\ll 1$
KEK-B	BELLE	e^+e^-	10.8	1.05 nb	1/4	10^{34}	$\ll 1$
HERA	HERA-b	pN	800	10 nb	$2\cdot10^{-6}$		4
Tevatron	BTeV	$\bar{p}p$	2000	100 μb	1/500	2×10^{32}	2
LHC	LHCb	$\bar{p}p$	14000	500 μb	1/160	2×10^{32}	0.6

† Machine has already exceeded design luminosity.

Table 4. *Machines in use or approved for dedicated b physics experiments.*

BTeV and LHCb will go into operation around 2007 with much larger event rates. The CMS and ATLAS experiments at the LHC will also contribute to b physics especially in the early stages when the luminosity will be relatively low; at design luminosity these experiments have an average of 23 interactions per crossing making the finding of detached vertices difficult.

3.3 e^+e^- detectors

Most experiments at e^+e^- storage rings look quite similar. CLEO II, shown in Figure 20, was the first detector to have both an excellent tracking system and an excellent electromagnetic calorimeter. Starting from the inside there is a thin beryllium beam pipe
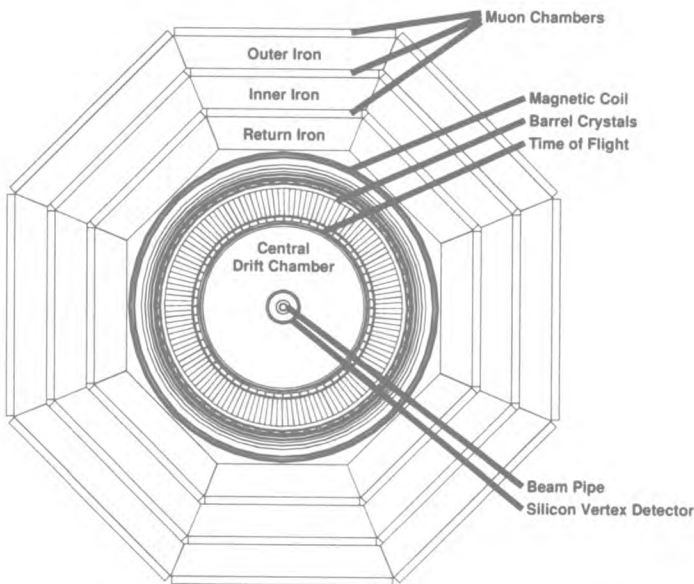
Figure 20. *Electrons view of the CLEO II detector.*

surrounded by a silicon vertex detector; this detector measures positions with an accuracy of $\approx 10\mu$m. Then there is a wire drift chamber whose main function is to measure the curving trajectories of particles in the 1.5T solenoidal magnetic field. The next device radially outward is time-of-flight system to distinguish pions, kaons and protons. This system only works for lower momenta. The next device is an electromagnetic calorimeter that uses Thallium doped CsI crystals; indeed this was the most important new technical implementation done in CLEO II and has also been adopted by BABAR and BELLE. Afterwards there is segmented iron that serves as both a magnetic flux return and a filter for muon identification.

The most important advance in the new CLEO III, BELLE and BABAR detectors is much better charged hadron identification. Each experiment uses different techniques based on Cherenkov radiation to extend π/K separation up to the limit from B decays. Figure 21 shows a view of the BELLE detector parallel to the beam.
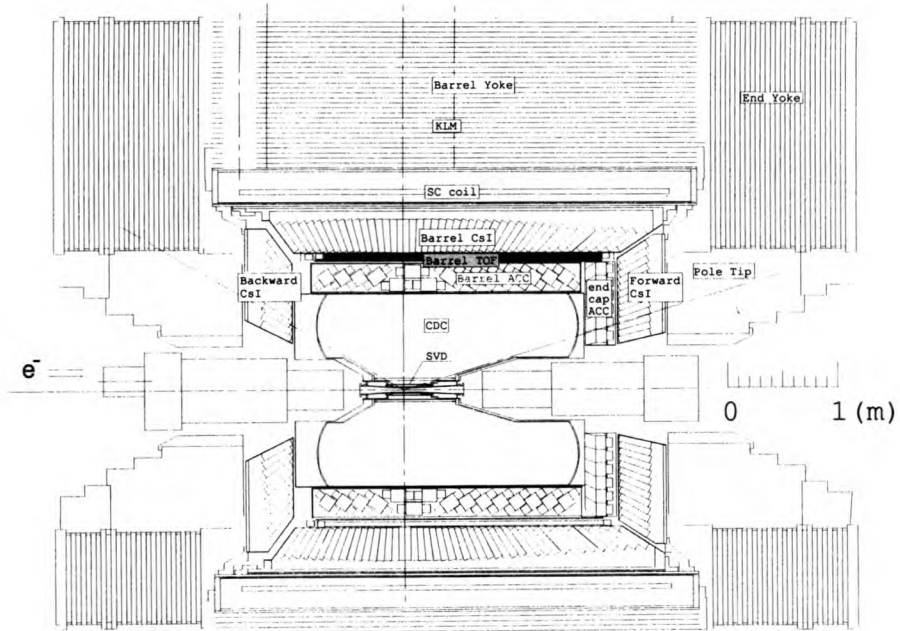
Figure 21. *The BELLE detector.*

3.4   *b* production characteristics at hadron colliders

To make precision measurements, large samples of *b*'s are necessary. Fortunately, these are available. With the Fermilab Main Injector, the Tevatron collider will produce $\approx 4 \times 10^{11}$ *b* hadrons/10^7 s at a luminosity of $2 \times 10^{32} \text{cm}^{-2}\text{s}^{-1}$. These rates compare very favorably to e^+e^- machines operating at the $\Upsilon(4S)$. At a luminosity of $10^{34}\text{cm}^{-2}\text{s}^{-1}$ they would produce 2×10^8 B's/10^7 s. Furthermore B_s, Λ_b and other *b*-flavored hadrons are accessible for study at hadron colliders. The LHC has about a five times larger *b* production cross-section than the Tevatron. Also important are the large charm rates, $\sim$10 times larger than the *b* rate.

   In order to understand the detector design it is useful to examine the characteristics of *b* quark production at $p\bar{p}$ collider. It is often customary to characterise heavy quark production in hadron collisions with the two variables p_t and η, where $\eta = -\ln(\tan(\theta/2))$, and θ is the angle of the particle with respect to the beam direction. According to QCD based calculations of *b* quark production, the *B*'s are produced "uniformly" in η and have a truncated transverse momentum spectrum, p_t, characterised by a mean value approximately equal to the *B* mass (Artuso 1994). The distribution in η is shown in Figure 22(a).

   The flat η distribution hides an important correlation of $b\bar{b}$ production at hadronic colliders. In Figure 22(c) the production angles of the hadron containing the *b* quark is plotted versus the production angle of the hadron containing the $\bar{b}$ quark according to the Pythia generator. Many important measurements require the reconstruction of a *b* decay and the determination of the flavor of the other $\bar{b}$, thus requiring both *b*'s to be
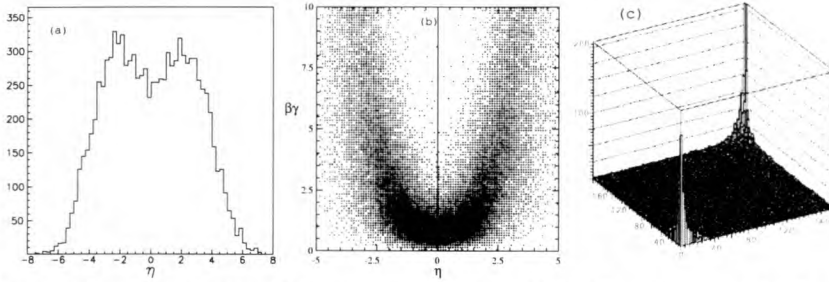
Figure 22. *(a) The B yield versus η. (b) $\beta\gamma$ of the B versus η. (c) The production angle (in degrees) for the hadron containing a b quark plotted versus the production angle for a hadron containing $\bar{b}$ quark.*

observed in the detector. There is a very strong correlation in the forward (and backward) direction: when the B is forward the $\bar{B}$ is also forward. This correlation is not present in the central region (near 90°). By instrumenting a relative small region of angular phase space, a large number of $b\bar{b}$ pairs can be detected. Furthermore the B's populating the forward and backward regions have large values of $\beta\gamma$ Figure 22(b).

BTeV is a dedicated heavy flavor experiment approved to run at the Fermilab Tevatron collider, uses two forward spectrometers (along both the p and $\bar{p}$ directions) that utilise the boost of the B's at large rapidities. This is of crucial importance because the main way to distinguish b decays is by the separation of decay vertices from the main interaction. LHCb, approved for operation at the LHC, needs a larger detector to analyze the higher momentum decay products, and thus has only one arm.

3.4.1   The BTeV detector description

I will describe BTeV here though LHCb shares many of the same features. There are difficulties that heavy quark experiments at hadron colliders must overcome. First of all, the huge b rate is accompanied by an even larger rate of uninteresting interactions. At the Tevatron the b-fraction is only 1/500. In addition all b experiments have problems searching for rare decay processes, at the level of parts per million, where the background from other b decays is dominant. To observe such decays the large data rate of b's must be handled. For example, BTeV, has 1 kHz of b's into the detector, and these events must be selected and written out. The electromagnetic calorimeter must be robust enough to deal with the particles from the underlying event and still have useful efficiency. Furthermore, radiation damage can destroy detector elements.

The BTeV Detector is shown in Figure 23 (Skwarnicki 2001) and the LHCb detector in Figure 24 (Muheim 2001). The central part of the BTeV detector has a silicon pixel detector inside a 1.5T dipole magnet. The LHCb experiment uses silicon strips. The BTeV pixel detector provides precision space points for use in both the offline analysis and the trigger. The pixel geometry is sketched in Figure 25(a). Pulse heights are measured on each pixel. Prototype detectors were tested in a beam at Fermilab, and excellent resolutions were obtained (Figure 25(b)), especially when reading out pulse heights (Appel 2001). The final design uses a 3-bit ADC for this.
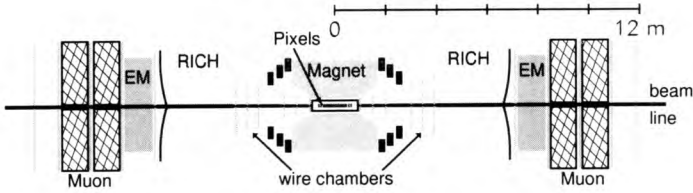
Figure 23. *Schematic of the BTeV detector.*
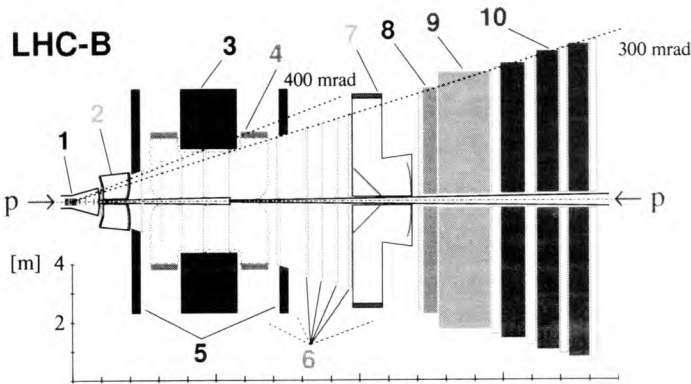
Figure 24. *A schematic diagram of the LHCb detector . Key: 1, vertex detectors; 2,aerogel and gas RICH's; 3, magnet yoke; 4, coils; 5, magnetic shielding; 6, tracking chambers; 7, gas RICH; 8, EM-calorimeter; 9, hadron calorimeter; 10, muon system.*
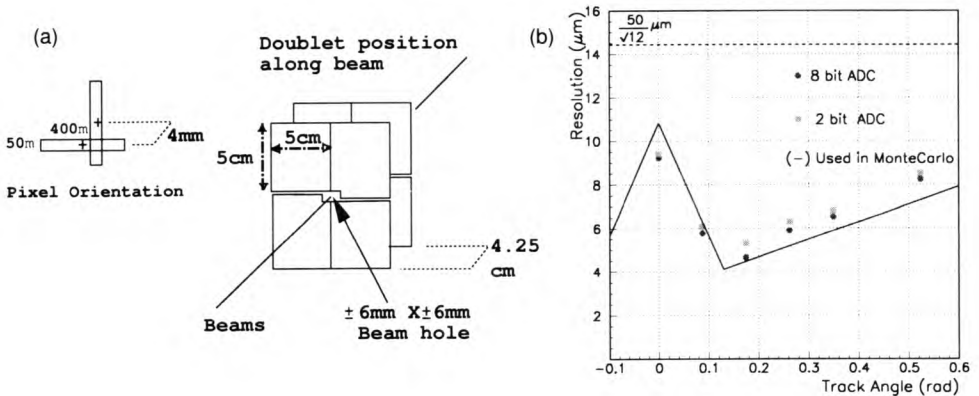
Figure 25. *(a) Pixel detector geometry in BTeV. The detector is inside the beam pipe. (b) The spatial resolution as a function of the incident track angle for both 2-bit and 8-bit ADC's as measured in an 800GeV/c pion beam. The straight lines are piecewise fits to the data used in the Monte Carlo simulation. The dashed line near the top indicates the resolution obtainable without using pulse height information.*

The pixel tracker provides excellent vertex resolution which is good enough to trigger on events with detached vertices characteristic of b or c decays. BTeV shows a rejection of 100:1 for minimum bias events in the first trigger level while accepting about 50% of the usable b decays. A good explanation of the trigger algorithm can be found at the website www-btev.fnal.gov/public_documents/animations/Animated_Trigger/. Further trigger levels reduce the background by about a factor of twenty while decreasing the b sample by only 10%. The trigger system stores data in a pipeline that is long enough to ensure no deadtime. The data acquisition system has sufficient throughput to accommodate an output of 1kHz of b's, 1kHz of c's and 2kHz of junk. Tracking is accomplished using straw tube wire chambers with silicon strip chambers in the high track density region near the beam.

Charged particle identification is done using a Ring Imaging CHerenkov detector. A gaseous C_4F_{10} radiator is used with a large mirror that focuses light onto a plane of photon detectors; these are currently Hybrid Photo-Diodes. They have a photocathode and a 20KV potential difference between the photocathode and a silicon diode that is segmented into 163 individual pads. The photoelectron is accelerated and focused onto the diode yielding position information for the initial photon. The system will provide four standard-deviation kaon/pion separation between 3–70GeV/c, electron/pion separation up to 22GeV/c and pion/muon separation up to 15GeV/c. Because protons don't radiate until 9GeV/c they can't be distinguished from kaons below this momenta. BTeV is considering an additional liquid C_6F_{14} radiator, 1cm thick, in front of the gas along with a proximity focused phototube array adjacent to the sides of the gas volume, to resolve this ambiguity.

BTeV also has an excellent Electromagnetic calorimeter made from $PbWO_4$ crystals, based on the design of CMS. Finally, the Muon system is used to both identify muons and provide an independent trigger on dimuons (BTeV 2000).

4 $B^0 - \overline{B}^0$ mixing

4.1   Introduction

Neutral B mesons can transform to their anti-particles before they decay. The diagrams for this process are shown in Figure 26 for the B_d. There is a similar diagram for the B_s. Although u, c and t quark exchanges are all shown, the t quark plays a dominant role mainly due to its mass, as the amplitude of this process is proportional to the mass of the exchanged fermion.
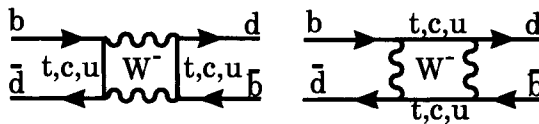
Figure 26. *The two diagrams for B_d mixing.*

Under the weak interactions the eigenstates of flavor, which are degenerate in pure QCD, can mix. Let the quantum mechanical basis vectors be $\{|1\rangle, |2\rangle\} \equiv \{|B^0\rangle, |\overline{B}^0\rangle\}$.

Then the Hamiltonian is

$$\mathcal{H} = M - \frac{i}{2}\Gamma = \begin{pmatrix} M & M_{12} \\ M_{12}^* & M \end{pmatrix} - \frac{i}{2} \begin{pmatrix} \Gamma & \Gamma_{12} \\ \Gamma_{12}^* & \Gamma \end{pmatrix}. \tag{34}$$

Diagonalising we have

$$\Delta m = m_{B_H} - m_{B_L} = 2\,|M_{12}|\,. \tag{35}$$

Here H refers to the heavier and L to the lighter of the two weak eigenstates.

B_d mixing was first discovered by the ARGUS experiment (Albrecht 1983), although there was a previous measurement by UA1 indicating mixing for a mixture of B_d^0 and B_s^0 (Albajar 1987). At the time it was quite a surprise, since m_t was thought to be in the 30GeV range. It is usual to define R as probability for a B^0 to materialise as a $\overline{B}^0$ divided by the probability it decays as a B^0. The OPAL data for R (Akers 1995) are shown in Figure 27.
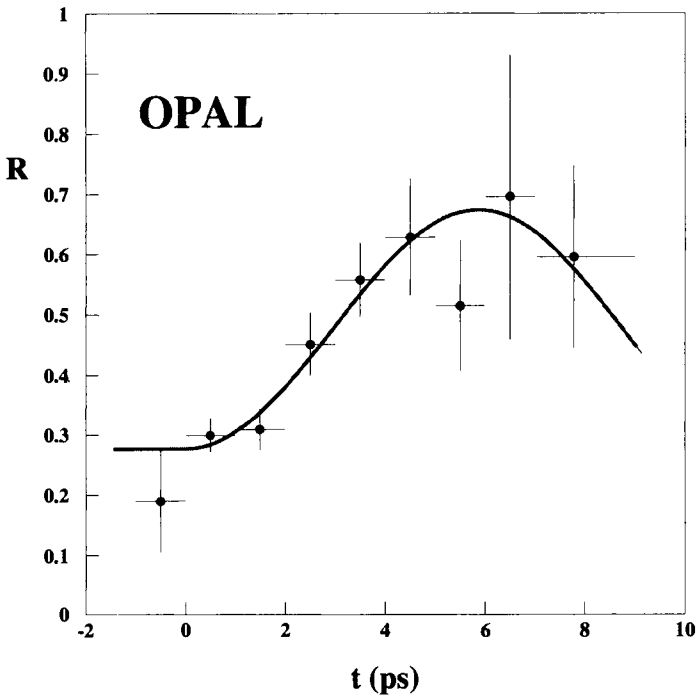
Figure 27. *The ratio, R, of like-sign to total events as a function of proper decay time, for selected $B \to D^{*+}X\ell^-\bar{\nu}$ events. The jet charge in the opposite hemisphere is used to determine the sign correlation. The curve is the result of a fit to the mixing parameter.*

Data from many experiments has been combined by a LEP Working Group to obtain an average value of $\Delta m_d = (0.489 \pm 0.0008) \times 10^{12}\,\hbar\mathrm{s}^{-1}$. Values from individual experiments are listed in Figure 28.
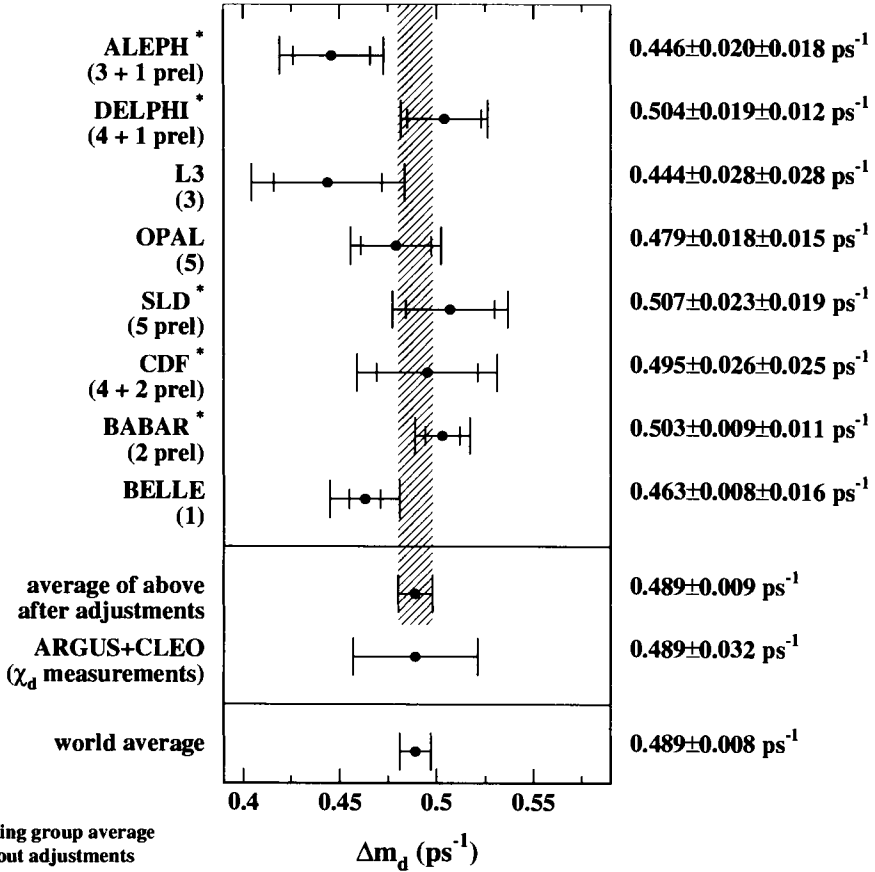
Figure 28. *Values of the B_d mixing parameter Δm_d for each experiment.*

The probability of mixing is given by (Gaillard 1974) (Bigi 2000)

$$x \equiv \frac{\Delta m}{\Gamma} = \frac{G_F^2}{6\pi^2} B_B f_B^2 m_B \tau_B |V_{tb}^* V_{td}|^2 m_t^2 \, F\left(\frac{m_t^2}{M_W^2}\right) \eta_{\text{QCD}}, \qquad (36)$$

where B_B is a parameter related to the probability of the d and $\bar{b}$ quarks forming a hadron which must be estimated theoretically, F is a known function which increases approximately as m_t^2, and η_{QCD} is a QCD correction, with a value of about 0.8. By far the largest uncertainty arises from the unknown decay constant, f_B. In principle f_B can be measured, since the decay rate of the annihilation process $B^- \to \ell^- \bar{\nu}$ is proportional to the product of $f_B^2 |V_{ub}|^2$. In practice, even if V_{ub} were well known, this is a very difficult process to measure. Our current best hope of determining f_B is to rely on unquenched lattice QCD which can use the measurements of the analogous $D^+ \to \mu^+ \nu$ decay as check. Measurements of these D decays requires the construction of a "τ-charm factory."

Since the B_d mixing measurement determines:

$$|V_{tb}^* V_{td}|^2 \propto |(1 - \rho - i\eta)|^2 = (\rho - 1)^2 + \eta^2, \qquad (37)$$

it can be related to the CKM parameters; giving a circle centered at (1,0) in the $\rho-\eta$ plane. This could in principle be a very powerful constraint. Unfortunately, the parameter B_B is not experimentally accessible, and f_B, although in principle measurable, has not been and may not be for a very long time, so it too must be calculated. The errors on these calculations are still quite large.

4.2 B_s mixing in the Standard Model

B_s^0 mesons can mix in a similar fashion to B_d^0 mesons. The diagrams in Figure 26 are modified by substituting s quarks for d quarks, thereby changing the relevant CKM matrix element from V_{td} to V_{ts}. The time dependent mixing fraction is

$$x_s \equiv \frac{\Delta m_s}{\Gamma_s} = \frac{G_F^2}{6\pi^2} B_{B_s} f_{B_s}^2 m_{B_s} \tau_{B_s} |V_{tb}^* V_{ts}|^2 m_t^2 F\left(\frac{m_t^2}{M_W^2}\right) \eta_{\rm QCD}, \tag{38}$$

which differs from Equation (36) by parameters relevant for the B_s rather than the B_d.

Measuring x_s allows us to use ratio of x_d/x_s to provide constraints on the CKM parameters ρ and η. We still obtain a circle in the (ρ, η) plane centered at (1,0):

$$|V_{td}|^2 = A^2 \lambda^4 \left[(1 - \rho)^2 + \eta^2\right] \tag{39}$$

$$\frac{|V_{td}|^2}{|V_{ts}|^2} = (1 - \rho)^2 + \eta^2.$$

However we must calculate only the SU(3) broken ratios B_{B_d}/B_{B_s} and f_{B_d}/f_{B_s}.

B_s^0 mixing has been searched for at LEP, the Tevatron, and the SLC. The probability, $\mathcal{P}(t)$ for a B_s to oscillate into a $\overline{B}_s$ is given by:

$$\mathcal{P}(t)\left(B_s \to \overline{B}_s\right) = \frac{1}{2}\Gamma_s e^{-\Gamma_s t}\left[1 + \cos\left(\Delta m_s t\right)\right], \tag{40}$$

where t is the proper time.

To combine different experiments a framework has been established where each experiment finds amplitudes A for a spectrum of test oscillation frequencies, ω, defined as

$$\mathcal{P}(t) = \frac{1}{2}\Gamma_s e^{-\Gamma_s t}\left[1 + A\cos\left(\omega t\right)\right]. \tag{41}$$

Figure 29 shows the world average measured amplitude A as a function of the test frequency $\omega = \Delta m_s$ (Leroy 2001). For each frequency the expected result is either zero for no mixing or one for mixing. No other value is physical, although allowing for measurement errors other values are possible. The data do indeed cross one at a Δm_s of 16ps^{-1}, however here the error on A is about 0.6, precluding a statistically significant discovery. The quoted upper limit at 95% confidence level is 14.6 ps^{-1}. This is the point where the value of A plus 1.645 times the error on A reaches one. Also indicated on the figure is the point where the error bar is small enough that a 4σ discovery would be possible. This is at $\Delta m_s = 11$ps^{-1}. One should also be aware that all the points are strongly correlated.

The upper limit on Δm_s translates into an upper limit $x_s < 21.6$, also at 95% confidence level. CDF plans to probe higher sensitivity and eventually LHCb and BTeV can reach values of $\sim$80.
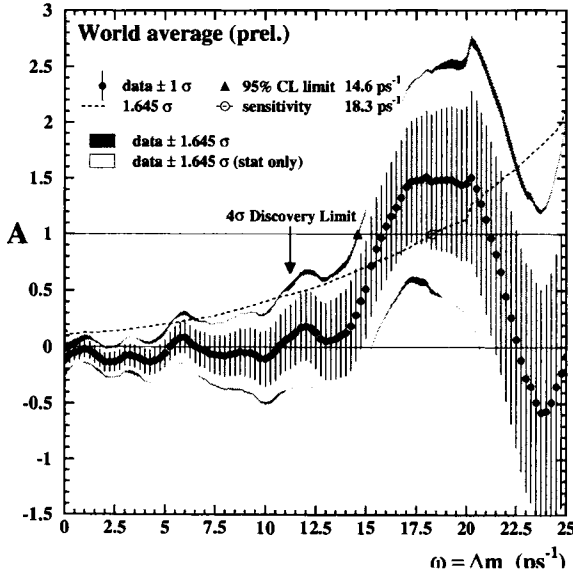
Figure 29. *Combined experimental values of the amplitude A versus the test frequency* $\omega = \Delta m_d$ *as defined in Equation 41. The inner (outer) envelopes give the 95% confidence levels using statistical (statistical and systematic) errors. The "sensitivity" shown at 18.3ps^{-1} is the likely place a 95% c.l. upper limit could be set. Also indicated is the maximum value, 11ps^{-1}, where a 4σ discovery would be possible.*

5   Rare b decays

5.1   Introduction

These processes proceed through higher order weak interactions involving loops, which are often called "penguin" processes, for unscientific reasons (Lingel 1998). A Feynman loop diagram is shown in Figure 30 that describes the transition of a b quark into a charged $-1/3$ s or d quark, which is effectively a neutral current transition. The dominant charged current decays change the b quark into a charged $+2/3$ quark, either c or u.
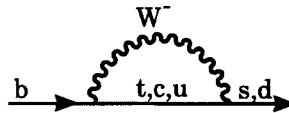
Figure 30. *Loop or "penguin" diagram for a $b \rightarrow s$ or $b \rightarrow d$ transition.*

The intermediate quark inside the loop can be any charge $+2/3$ quark. The relative size of the different contributions arises from different quark masses and CKM elements. In terms of the Cabibbo angle (λ=0.22), we have for the $b \rightarrow s$ case $t{:}c{:}u \approx \lambda^2{:}\lambda^2{:}\lambda^4$. The mass dependence favours the t loop, but the amplitude for c processes can be quite large $\approx 30\%$. Moreover, as pointed out by Bander, Silverman and Soni (1979), interference can

occur between t, c and u diagrams and lead to CP violation. In the Standard Model it is not expected to occur for $b \to s$, due to the lack of a CKM phase difference, but could occur for $b \to d$. It is always worth looking for such CP violation effects; all that needs to be done, for example, is to compare the number of $B^- \to K^{*-}\gamma$ events with the number of $B^+ \to K^{*+}\gamma$ events.

There are other ways for physics beyond the Standard Model to appear. For example, the W^- in the loop can be replaced by some other charged object such as a Higgs; it is also possible for a new object to replace the t.

5.2 Standard Model theory

In the Standard Model the effective Hamiltonian for the intermediate t quark is given by (Deshpande 1994)

$$H_{\text{eff}} = -\frac{4G_F}{\sqrt{2}} V_{tb} V_{ts}^* \sum_{i=1}^{10} C_i(\mu) O_i(\mu). \tag{42}$$

Two of the more important operators are

$$O_1 = \bar{s}_L^i \gamma_\mu b_L^j \bar{c}_L^j \gamma^\mu c_L^i, \qquad O_7 = \frac{e}{16\pi^2} m_b \bar{s}_L^i \sigma_{\mu\nu} b_R^j F^{\mu\nu}. \tag{43}$$

The matrix elements are evaluated at the scale $\mu = M_W$ and then evolved to the b mass scale using renormalisation group equations, which mix the operators:

$$C_i(\mu) = \sum_j U_{ij}(\mu, M_W) C_j(M_W). \tag{44}$$

5.3 $b \to s\gamma$

This process occurs when any of the charged particles in Figure 30 emits a photon. The only operator which enters into the calculation is $C_7(\mu)$. We have for the inclusive decay

$$H_{\text{eff}} = \frac{4G_F}{\sqrt{2}} (V_{tb}V_{ts})^2 C_7(m_b) O_7 \tag{45}$$

$$O_7 = \frac{e}{16\pi^2} m_b \bar{s}_L \sigma_{\mu\nu} b_R F^{\mu\nu} \tag{46}$$

$$\Gamma(b \to s\gamma) = \frac{G_F^2 \alpha m_b^5}{32\pi^4} |C_7|^2 \left| V_{tb} V_{ts}^* \right|^2. \tag{47}$$

It is far more difficult to calculate the exclusive radiative decay rates, but they are much easier to measure. Note that the reaction $B \to K\gamma$ would violate angular momentum conservation, so the simplest exclusive final state is $B \to K^*\gamma$.

Different techniques are used for reconstructing exclusive and inclusive decays and unique methods are invoked for exclusive decays on the $\Upsilon(4S)$. At other machines the decay products, i, from an exclusive B decays are used to reconstruct an "invariant mass" via $M^2 = \sum_i E_i^2 - \sum_i \mathbf{p}_i^2$. At the $\Upsilon(4S)$ the decay products are first tested to see if the sum of their energies is close to the beam energy, E_{beam}. If this is true then the "beam constrained" invariant mass is calculated as $M^2 = E_{\text{beam}}^2 - \sum_i \mathbf{p}_i^2$. These methods can be used for all exclusive B decays. Figure 31 shows the BELLE data for the reaction
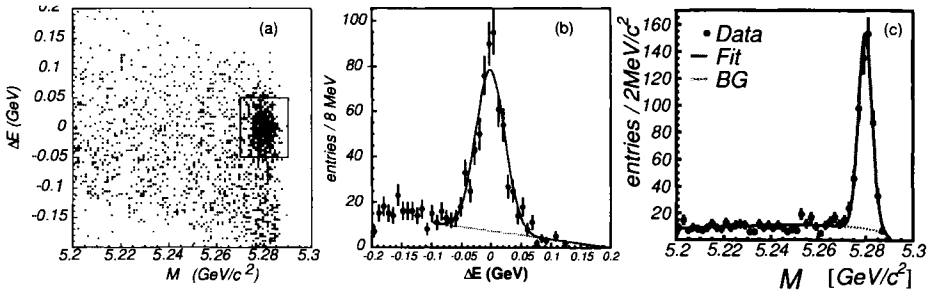
Figure 31. *BELLE data for the reaction $\overline{B}^0 \to D^{*+}\pi^-$. (a) The correlation between ΔE and M. The box shows the signal region. (b) The projection in ΔE for events in the M signal region; the line shows a fit to the background. (c) The M distribution for ΔE in the signal region; the line shows a fit to the background.*

$\overline{B}^0 \to D^{*+}\pi^-$, where the $D^{*+} \to \pi^+ D^0$. BELLE first selects events with candidate D^0's. Then they require an additional π^+ where the measured mass difference between the $D^0\pi^+$ minus the D^0 candidate is consistent with the known mass difference. Selecting the D^{*+} candidates they combine them with candidate π^-. Figure 31(a) shows the the correlation between the difference in measured energy $\Delta E = \sum_i E_i - E_{\text{beam}}$ and the beam-constrained invariant mass. In Figure 31(b) ΔE is shown after selecting the signal region in M, and in Figure 31(c) M is shown after selecting on ΔE. These plots show how clean signals can be selected.

CLEO first measured the exclusive rate into $K^*(892)\gamma$ (Ammar 1993) shown in Figure 32. Here several different decay modes of the $K^*(892)$ are used. The current world average value for $\mathcal{B}(B \to K^*\gamma) = (4.2 \pm 0.8 \pm 0.6) \times 10^{-5}$.
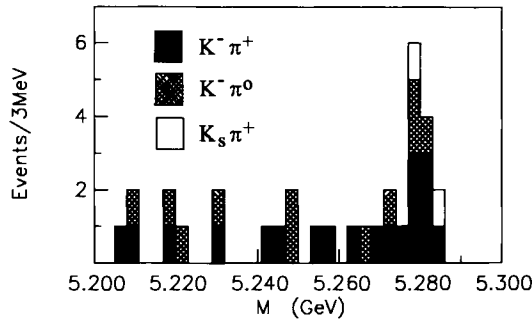
Figure 32. *First published CLEO data for the reaction $B \to K^*\gamma$ showing the M distribution for ΔE in the signal region.*

To find inclusive decays two techniques are used. The first one, which provides the cleanest signals, is to sum the exclusive decays for the final states $Kn\pi\gamma$, where $n \leq 4$ and only one of the pions is a π^0. These requirements are necessary or the backgrounds become extremely large. Both charged and neutral kaons are used. Of course, imposing these restrictions leads to a model dependence of the result that must be carefully evaluated. This is why having an independent technique is useful. That is provided by detecting only
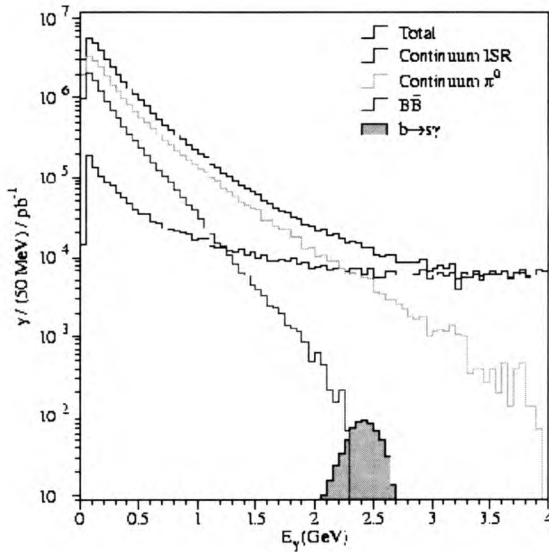
Figure 33. *Levels of inclusive photons from various background processes at the* $\Upsilon(4S)$ *labeled largest to smallest at 2.5GeV/c. Also shown is the expected signal from* $b \rightarrow s\gamma$.

the high energy photon. The technique used is to form a neutral network to discriminate between continuum and $\Upsilon(4S)$ data using shape variables.

The momentum spectrum of the γ peaks close to its maximum value at half the B mass. If we had data with only B mesons, it would be easy to pick out $b \rightarrow s\gamma$. We have, however, a large background from other processes. At the $\Upsilon(4S)$, the γ spectrum from the different background processes is shown. The largest is π^0 production from continuum e^+e^- collisions, but another large source is initial state radiation (ISR), where one of the beam electrons radiates a hard photon before annihilation. The backgrounds and the expected signal are illustrated in Figure 33. Similar backgrounds exist at LEP.
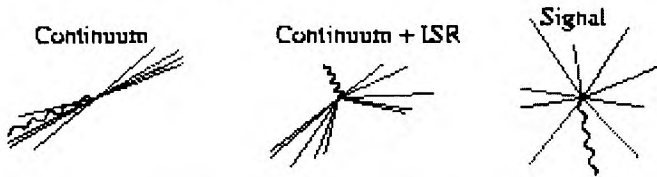
Figure 34. *Examples of idealised event shapes. The straight lines indicate hadrons and the wavy lines photons.*

CLEO also made the first measurement of the inclusive rate for $b \rightarrow s\gamma$ (Alam 1995). To remove background CLEO used two techniques, one based on "event shapes" and the other on summing exclusively reconstructed B samples. Examples of idealised events are shown in Figure 34. CLEO uses eight different shape variables and defines a variable r using a neural network to distinguish signal from background. The idea of the B reconstruction analysis is to find the inclusive branching ratio by summing over exclusive

modes. The allowed hadronic system is comprised of either a $K_s \to \pi^+\pi^-$ candidate or a $K^\mp$ combined with 1-4 pions, only one of which can be neutral. The restriction on the number and kind of pions maximises efficiency while minimising background. It does however lead to a model dependent error. For all combinations CLEO evaluates

$$\chi_B^2 = \left(\frac{M_B - 5.279}{\sigma_M}\right)^2 + \left(\frac{E_B - E_{\text{beam}}}{\sigma_E}\right)^2, \tag{48}$$

where M_B is the measured B mass for that hypothesis and E_B is its energy. χ_B^2 is required to be < 20. If any particular event has more than one hypothesis, the solution which minimises χ_B^2 is chosen. For events with a reconstructed B candidate CLEO also considers the angle between the direction of the B and the thrust axis of event with the B candidate removed, $\cos(\theta_t)$. This is highly effective in removing continuum background.

A neural network is used to combine r, χ_B^2, $\cos(\theta_t)$ into a new variable r_c and events are then weighted according to their value of r_c. This method maximises the statistical potential of the data. Figure 35 shows the photon energy spectrum of the inclusive signal from CLEO combining both reconstruction techniques. The signal is compared to a theoretical prediction based on the model of Ali and Greub (Ali 1991). A fit to the model over the photon energy range from 2.0 to 2.7GeV/c gives the branching ratio result shown in Table 5, where the first error is statistical, the second systematic and the third shows the dependence on the theoretical model (Chen 2001).
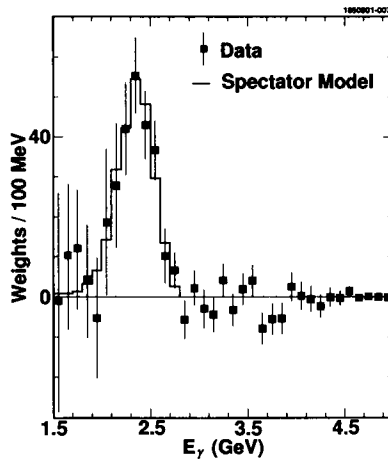
Figure 35. *The background subtracted photon energy spectrum from CLEO. The spectrum is not corrected for resolution or efficiency. The solid lines show the spectrum from a simulation of the Ali-Greub spectator model with the b quark mass set to 4.690GeV and the Fermi momentum set to 410MeV/c.*

ALEPH reduces the backgrounds by weighting candidate decay tracks in a $b \to s\gamma$ event by a combination of their momentum, impact parameter with respect to the main vertex and rapidity with respect to the b-hadron direction (Barate 1998).

Current results are shown in Table 5. The data are in agreement with the Standard Model theoretical prediction to next to leading order, including quark mass effects of

Experiment	$\mathcal{B} \times 10^{-4}$
CLEO	$3.21 \pm 0.43 \pm 0.27^{+0.18}_{-0.10}$
ALEPH	$3.11 \pm 0.80 \pm 0.72$
BELLE	$3.36 \pm 0.53 \pm 0.44^{+0.50}_{-0.54}$
Average	3.23 ± 0.42

Table 5. $\mathcal{B}(b \to s\gamma)$.

$(3.73 \pm 0.30) \times 10^{-4}$ (Hurth 2001). A deviation here would show physics beyond the Standard Model. More precise data and better theory are needed to further limit the parameter space of new physics models, or show an effect.

5.3.1 $|V_{cb}|$ using moments of the photon energy spectrum

In Section(2.2.4) we found a value of V_{cb} using the first and second hadronic mass moments. Here we use the first moment of the photon energy distribution in $b \to s\gamma$. The values found for the moments and $\overline{\Lambda}$ which is directly proportional to $\langle E_\gamma \rangle$ are (Chen 2001)

$$\langle E_\gamma \rangle = 2.346 \pm 0.032 \pm 0.011 \text{ GeV} \tag{49}$$

$$\langle E_\gamma^2 \rangle - \langle E_\gamma \rangle^2 = 0.0226 \pm 0.0066 \pm 0.0020 \text{ GeV}^2 \tag{50}$$

$$\overline{\Lambda} = 0.35 \pm 0.08 \pm 0.10 \text{ GeV.} \tag{51}$$

In Figure 36 we show the combination of first moments from photon energy in $b \to s\gamma$ and hadron moments in $b \to c\ell\nu$. This implies a value of V_{cb} around 0.0406.
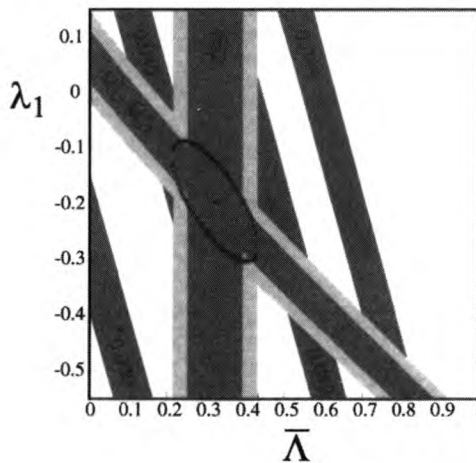
Figure 36. *Correlation between λ_1, $\overline{\Lambda}$ and V_{cb} derived from $\langle E_\gamma \rangle$ and $\langle M_X^2 - \overline{M}_D^2 \rangle$. The lighter bounds include both experimental and theoretical errors.*

5.4   Rare hadronic decays

The decays $\overline{B}^0 \to \pi^+\pi^-$ and $\overline{B}^0 \to K^-\pi^+$ do not contain any charm quarks in the final state, so they must proceed either via the tree level V_{ub} process shown in Figure 37(left), or via the penguin process shown in Figure 37(right).
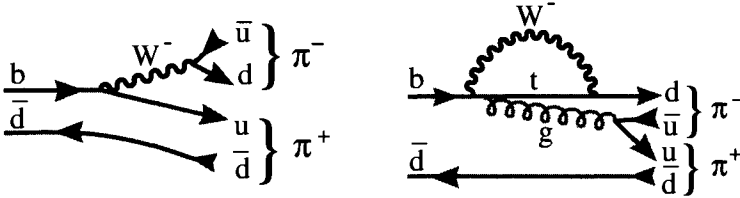
Figure 37. *Decay diagrams for $\overline{B}^0 \to \pi^+\pi^-$. (left) Via tree level V_{ub} process. (right) Via a penguin process.*

These diagrams can interfere with each other, and they can also interfere through B^0 mixing, thus complicating any weak phase extraction. The same diagrams are applicable for $\overline{B}^0 \to K^-\pi^+$ by replacing $W^- \to \overline{u}d$ in the tree level diagram by $W^- \to \overline{u}s$ and replacing the td coupling in the penguin by a ts coupling.
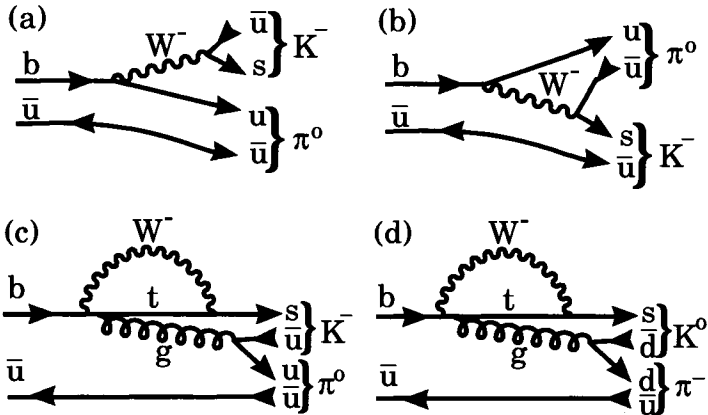
Figure 38. *Diagrams for $B^- \to K^-\pi^0$, (a) and (b) are tree level diagrams where (b) is color suppressed; (c) is a penguin diagram. (d) shows $B^- \to K^0\pi^-$, which cannot be produced via a tree diagram.*

Other diagrams for producing $K\pi$ final states are shown in Figure 38. In Section 7.1 it will be shown that CP violation can result from the interference between two distinct decay amplitudes leading to the same final state. Consider the possibility of observing
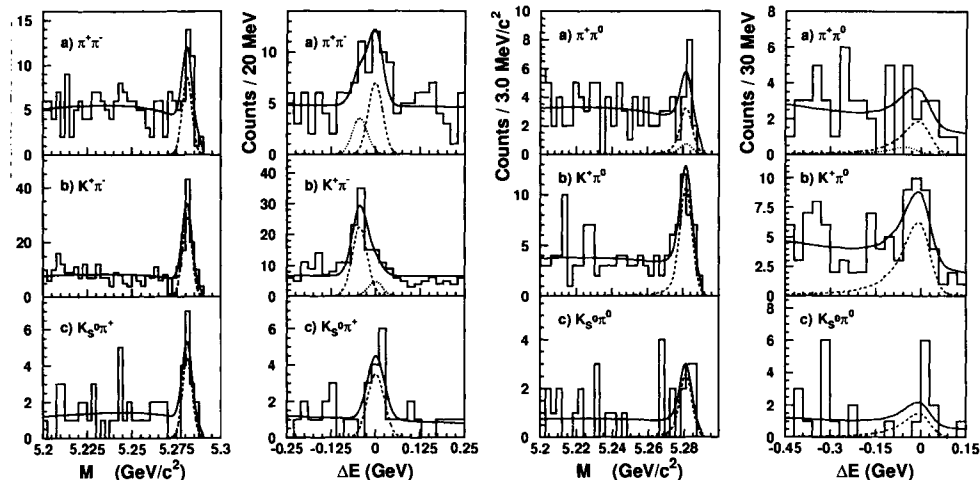
Figure 39. *Signals in M and ΔE for two-body decay modes from BELLE. The data result from projections of a Likelihood fit that takes into account event shape and particle identification information. The dashed lines are the signal projections. The dotted lines in the ΔE distributions are projections of the background component from the $\pi \Leftrightarrow K$ substitution.*

CP violation by measuring a rate difference between $B^- \to K^-\pi^0$ and $B^+ \to K^+\pi^0$. The $K^-\pi^0$ final state can be reached either by tree or penguin diagrams. The tree diagram has an imaginary part coming from the V_{ub} coupling, while the penguin term does not, thus ensuring a weak phase difference. This type of CP violation is called "direct." Note also that the process $B^- \to K^0\pi^-$ can only be produced by the penguin diagram in Figure 38(d). Therefore, we do not expect a rate difference between $B^- \to K^0\pi^-$ and $B^+ \to K^0\pi^+$.

Measurements of these rates have been made by several groups. Recent data from BELLE are shown in Figure 39 (Abe 2001a). Table 6 lists the currently measured branching ratios.

Mode	CLEO	BABAR	BELLE	Average
$\pi^+\pi^-$	$4.7^{+1.8}_{-1.5} \pm 0.6$	$4.1 \pm 1.0 \pm 0.7$	$5.6^{+2.3}_{-2.0} \pm 0.4$	$4.5^{+0.9}_{-0.8}$
$\pi^+\pi^0$	<12	<9.6	<13.4	
$K^\pm\pi^\mp$	$18.8^{+2.8}_{-2.6} \pm 1.3$	$16.7 \pm 1.6 \pm 1.3$	$19.3^{+3.4+1.5}_{-3.2-.06}$	$17.7^{+1.6}_{-1.5}$
$K^+\pi^0$	$12.1^{+3.0+2.1}_{-2.8-1.4}$	$10.8^{+2.1}_{-1.9} \pm 1.6$	$16.3^{+3.5+1.6}_{-3.3-1.8}$	$12.1^{+1.7}_{-1.6}$
$K^0\pi^-$	$18.2^{+4.6}_{-4.0} \pm 1.6$	$18.2^{+3.3}_{-3.0} \pm 2.0$	$13.7^{+5.7+1.9}_{-4.8-1.8}$	$17.3^{+2.7}_{-2.4}$
$K^0\pi^0$	$14.8^{+5.9+2.4}_{-5.1-3.3}$	$8.2^{+3.1}_{-2.7} \pm 1.2$	$16.0^{+7.2+2.5}_{-5.9-2.7}$	$10.4^{+2.7}_{-2.5}$

Table 6. *Branching Ratios for $B \to K\pi$ and $B \to \pi\pi$ in units of 10^{-6}.*

6   Hadronic decays

6.1   Introduction

Mark Wise in his talk at the 2001 Lepton Photon conference (Wise 2001) gave some advice to theorists: "If you drink the nonleptonic your physics career will be ruined and you will end up face down and in the gutter." Presumably Mark's statement was inspired by the difficulty in predicting hadronic decays. Here we have lots of gluon exchange with low energy gluons, while perturbation theory works well when the energies are large compared with $\Lambda_{QCD} \sim 200$MeV. Furthermore, multibody decays are currently impossible to predict, so we will consider only two-body decays.

6.2   Two-body decays into a charm or charmonium

We start by considering two-body decays into a charmed hadron (Neubert 1997). Figure 6 shows the processes for both B^- and $\overline{B}^0$ decays into a D and a π^-. There is only one possible process for the $\overline{B}^0$, the simple spectator process (left), while the color suppressed spectator (right) is also allowed for the B^-. We call decays with only the simple spectator diagram allowed "class I" and decays where both the simple and color suppressed diagrams are allowed "class III". Note, that because the colors of all the outgoing quarks must be the same in the color suppressed case, naively the amplitude is only $\sim 1/3$ that of the simple spectator case where the W^- can transform into quarks of all three colors. "Class II" decays are processes than can only be reached by the color suppressed spectator diagram, for example the $B^0 \to J/\psi K_s$ decay shown in Figure 40.
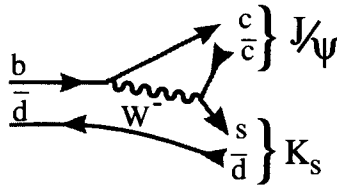
Figure 40. *Color suppressed spectator decay diagram for $B^0 \to J/\psi K_s$.*

The effective Hamiltonian consists of local 4-quark operators renormalised at the scale μ and the Wilson coefficients from the Operator Product Expansion, $c_i(\mu)$. We have

$$H_{\text{eff}} = \frac{G_F}{\sqrt{2}} \left\{ V_{cb} \left[c_1(\mu) Q_1^{cb} + c_2\mu) Q_2^{cb} \right] + V_{ub...} \right\} \tag{52}$$

$$Q_1^{cb} = \left[(\overline{d}u)_{V-A} + (\overline{s}c)_{V-A} \right] (\overline{c}b)_{V-A}$$

$$Q_2^{cb} = (\overline{c}u)_{V-A}(\overline{d}b)_{V-A} + (\overline{c}c)_{V-A}(\overline{s}b)_{V-A},$$

where the notation $(\overline{q}_1 q_2)_{V-A} \equiv \overline{q}_1 \gamma_\mu (1 - \gamma_5) q_2$. Without QCD corrections $c_1(\mu) = 1$ and $c_2(\mu) = 0$. From the non-leading order correction using the renormalisation group equations, we have $c_1(\mu) = 1.132$ and $c_2(\mu) = -0.249$.

We can factorise the amplitude by assuming that the current producing the π^- is independent of the one producing the charmed hadron. Let us consider a class I case, $\overline{B}^0 \to D^+ \pi^-$. The amplitude can be written as

$$A_{\text{fact}} = -\frac{G_F}{\sqrt{2}} V_{cb} V_{ud}^* a_1 \left\langle \pi^- \left| (\overline{d}u)_A \right| 0 \right\rangle \left\langle D^+ \left| (\overline{c}b)_V \right| \overline{B}^0 \right\rangle \tag{53}$$

The part of the amplitude dealing with the π^- is known from pion decay. We have $\langle \pi^- | \overline{d} \gamma_\mu \gamma_5 u | 0 \rangle = i f_\pi p_\mu$, where the axial vector structure is made explicit, p_μ is the pion four-vector and f_π is given by measuring the decay width for $\pi^- \to \mu^- \nu$. The term a_1 is defined as

$$a_1 = c_1(\mu_f) + \xi c_2(\mu_f), \tag{54}$$

where ξ is equal to the number of colors and the scale μ_f is on the order of the b quark mass. Then

$$A_{\text{fact}} = -\frac{G_F}{\sqrt{2}} V_{cb} V_{ud}^* a_1 f_\pi (m_B^2 - m_D^2) F_0^{B \to D}(m_\pi^2). \tag{55}$$

The F_0 form factor can either be calculated or measured in semileptonic decays.

For a class II process $B \to J/\psi K$:

$$A_{\text{fact}} = -\frac{G_F}{\sqrt{2}} V_{cb} V_{cs}^* a_2 \langle J/\psi | (\overline{c}c)_V | 0 \rangle \langle K | (\overline{s}b)_V | B \rangle, \tag{56}$$

where $a_2 = c_2(\mu_f) + \xi c_1(\mu_f)$.

A class III decay has a term $a_1 + x a_2$ in the amplitude, where x equals one from flavor symmetry. The actual values of a_1, x, and a_2 are not well predicted from theory, but we can obtain them from the data.

One method is to use the class I decays to obtain $a_1 = 1.08 \pm 0.04$. It is possible to calculate a_2/a_1 as shown in Figure 41 (Neubert 1997). Using these values the measured branching ratios are compared with the predicted ones in Table 7. Here $x = +1$ is used, taken from the data. This is opposite to the interference in D decays but is expected from the calculation shown in Figure 41.
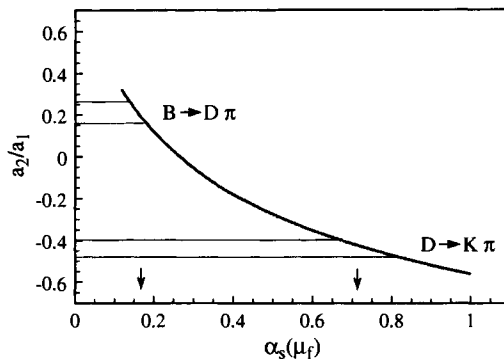
Figure 41. *Calculated values of a_2/a_1 versus α_s. The arrows indicate the ratios chosen for B and D decays.*

	Class I			Class II			Class III	
Mode	Model	Data	Mode	Model	Data	Mode	Model	Data
$D^+\pi^-$	30	30±4	$J/\psi K^-$	11	10±1	$D^0\pi^-$	48	53±5
$D^{*+}\pi^-$	30	28±2	$J/\psi K^*$	17	15±2	$D^{*o}\pi^-$	49	46±4
$D^+\rho^-$	70	79±14	$D^0\pi^0$	0.7	2.8±0.4	$D^0\rho^-$	110	134±18
$D^{*+}\rho^-$	85	73±15	$D^{*o}\pi^0$	1.0	2.1±0.9	$D^{*o}\rho^-$	119	155±31

Table 7. *Predicted Branching Ratios (10^{-4}) (Neubert 1997) Compared to Measurement For Two-Body Hadronic Decays.*

The agreement is rather good except for the newly measured $D^{o(*)}\pi^0$ modes where it is rather miserable. CLEO and BELLE both have rates for $D^0\pi^0$ of $(2.7 \pm 0.3 \pm 0.5) \times 10^{-4}$ and $(2.9^{+0.4}_{-0.3} \pm 0.6) \times 10^{-4}$, respectively, while CLEO alone has measured $D^{*o}\pi^0$ as $(2.1 \pm 0.5 \pm 0.8) \times 10^{-4}$.

6.2.1   Isospin analysis of the $B \to D\pi$ system

All the decay rates for $B \to D\pi$ have now been measured. The four-quark operator $(\bar{d}u)(\bar{c}b)$ has isospin I=1 and I_3=+1. It transforms the $\overline{B}^0$ into final states $D^+\pi^-$ and $D^0\pi^0$ with I=1/2 or I=3/2. The B^- decays into $D^0\pi^-$ with I=3/2 only. It is thought that the isospin amplitudes cannot be modified by final state interactions, so we can look for evidence of final state phase shifts by doing an "isospin analysis." The decay amplitudes form a triangle as shown in Figure 42.
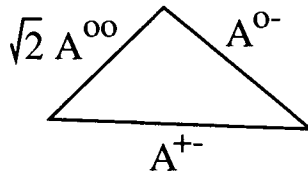
Figure 42. *The isospin amplitude triangle for $B \to D\pi^-$ decays.*

The relationship among B decay amplitudes and isospin amplitudes is given by

$$A(\overline{B}^0 \to D^+\pi^-) = \sqrt{\frac{1}{3}}A_{3/2} + \sqrt{\frac{2}{3}}A_{1/2}$$

$$A(\overline{B}^0 \to D^0\pi^0) = \sqrt{\frac{2}{3}}A_{3/2} - \sqrt{\frac{1}{3}}A_{1/2} \qquad (57)$$

$$A(B^- \to D^0\pi^-) = \sqrt{3}A_{3/2}.$$

These equations may be solved for the isospin amplitudes and the relative phase shift between the two amplitudes. The solution is

$$\left|A_{1/2}\right|^2 = \left|A(\overline{B}^0 \to D^+\pi^-)\right|^2 + \left|A(\overline{B}^0 \to D^0\pi^0)\right|^2 - \frac{1}{3}\left|A(B^- \to D^0\pi^-)\right|^2 \qquad (58)$$

$$\left|A_{1/2}\right|^2 = \frac{1}{3}\left|A(B^- \to D^0\pi^-)\right|^2$$

$$\cos\delta = \cos\left(\delta_{3/2} - \delta_{1/2}\right) = \frac{3\left|A(B^- \to D^0\pi^-)\right|^2 - 2\left|A_{1/2}\right|^2 - \left|A_{3/2}\right|^2}{2\sqrt{2}\left|A_{1/2}\right|\left|A_{3/2}\right|}$$

Solving these equations for the $D\pi$ final states gives $\cos\delta = 0.88\pm0.05$ which indicates a phase shift of about 28 ± 8 degrees, which is not statistically significant enough to rule out zero.

6.2.2 Factorisation tests using semileptonic decays

The factorised amplitude for $D\pi^-$ decays in Equation 56 is the product of two hadronic currents, one for $W^- \to \pi^-$ and the other for $B \to D$. In semileptonic decay (Figure 5) we have the product of the known lepton current and the pion current. At $q^2 = m_\pi^2$ the $B \to D$ should be the same in both decays, at least for class I. The comparison for the general case of any hadron h^- is

$$\Gamma\left(\overline{B} \to D^{(*)}h\right) = 6\pi^2 a_1^2 f_h^2 \left|V_{ud}\right|^2 \frac{d\Gamma}{dq^2}\left(\overline{B} \to D^{(*)}\ell^-\overline{\nu}\right)\Bigg|_{q^2=m_h^2}. \tag{59}$$

Tests of this equation for D^{*+} and a π^-, ρ^- or a_1^- are satisfied at about 15% accuracy (Bortoletto 1990) (Browder 1996).

Another test compares the polarisation of the D^* in both hadronic and semileptonic cases:

$$\frac{\Gamma_L}{\Gamma}\left(\overline{B} \to D^*h\right) = \frac{\Gamma_L}{\Gamma}\left(\overline{B} \to D^*h\right)\Bigg|_{q^2=m_h^2}, \tag{60}$$

where Γ_L denotes the longitudinally polarised fraction of the decay width. Comparisons with data will be shown in the next section.

There are more modern approaches to factorisation (Beneke 2001, and Bauer 2001). However these approaches predict

$$\frac{\Gamma(B^- \to D^0\pi^-)}{\Gamma(\overline{B}^0 \to D^+\pi^-)} = 1 + \mathcal{O}(\Lambda_{\text{QCD}}/m_b) \tag{61}$$

which seems to contradict current observations.

6.3 Observation of the ρ' in B decays

CLEO made the first statistically significant observations of six hadronic B decays shown in Table 8 that result from studying the reactions $B \to D^{(*)}\pi^+\pi^-\pi^-\pi^0$(Alexander 2001). The signal in one of these final states $\overline{B}^0 \to D^{(*+)}\pi^+\pi^-\pi^-\pi^0$, where $D^{*+} \to \pi^+D^0$ and $D^0 \to K^-\pi^+$ is shown in Figure 43.

In examining the substructure of the four-pions, a clear ω signal was observed in the $\pi^+\pi^-\pi^0$ mass as can be seen in Figure 44, leading to a significant amount of $D^{(*)}\omega\pi^-$. Furthermore, there is a low-mass resonant substructure in the $\omega\pi^-$ mass (Figure 45). The spin and parity of the $\omega\pi^-$ resonance (denoted by A temporarily) is determined by considering the decay sequence $B \to A\, D;\quad A \to \omega\pi\quad$ and $\quad\omega \to \pi^+\pi^-\pi^0$. The angular

Mode	$\mathcal{B}$ (%)
$\overline{B}^0 \to D^{*+}\pi^+\pi^-\pi^-\pi^0$	$1.72\pm0.14\pm0.24$
$\overline{B}^0 \to D^{*+}\omega\pi^-$	$0.29\pm0.03\pm0.04$
$\overline{B}^0 \to D^+\omega\pi^-$	$0.28\pm0.05\pm0.03$
$B^- \to D^{*o}\pi^+\pi^-\pi^-\pi^0$	$1.80\pm0.24\pm0.25$
$B^- \to D^{*o}\omega\pi^-$	$0.45\pm0.10\pm0.07$
$B^- \to D^o\omega\pi^-$	$0.41\pm0.07\pm0.04$

Table 8. *Measured Branching Ratios*
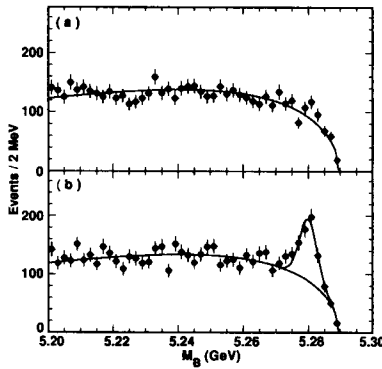
Figure 43. *The $\overline{B}$ candidate mass spectra for the final state $D^{*+}\pi^+\pi^-\pi^-\pi^0$, with $D^0 \to K^-\pi^+$ (a) for ΔE sidebands and (b) for ΔE consistent with zero. The curve in (a) is a fit to the background distribution described in the text, while in (b) the shape from (a) is used with the normalisation allowed to float and a signal Gaussian of width 2.7MeV is added.*
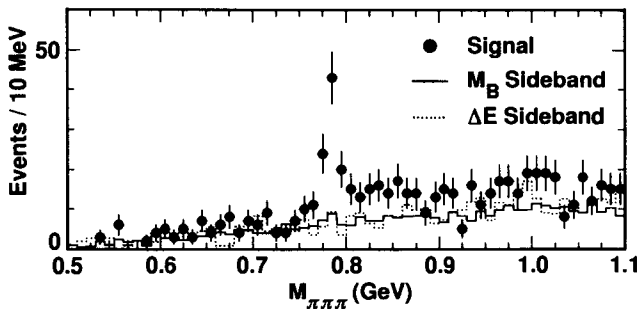
Figure 44. *The invariant mass spectra of $\pi^+\pi^-\pi^0$ for the final state $D^{*+}\pi^+\pi^-\pi^-\pi^0$ for all three D^0 decay modes $K^-\pi^+$, $K^-\pi^+\pi^+\pi^-$ and $K^-\pi^+\pi^0$. The solid histogram is the background estimate from the M_B lower sideband and the dashed histogram is from the ΔE sidebands; both are normalised to the fitted number of background events. There is an additional cut selecting the centre of the Dalitz plot.*
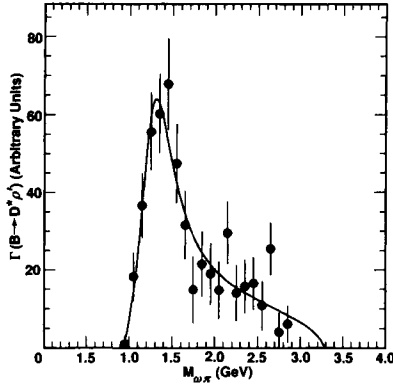
Figure 45. *(left) The background subtracted efficiency corrected $\omega\pi^-$ mass spectrum from $\overline{B}^0 \to (D^{*+} + D^0 + D^+)\omega\pi^-$ decays fit to a Breit-Wigner shape.*

distributions are shown in Figure 46. Here θ_A is the angle between the ω direction in the A rest frame and the A direction in the B rest frame; θ_ω is the orientation of the ω decay plane in the ω rest frame; and χ is the angle between the A and ω decay planes.
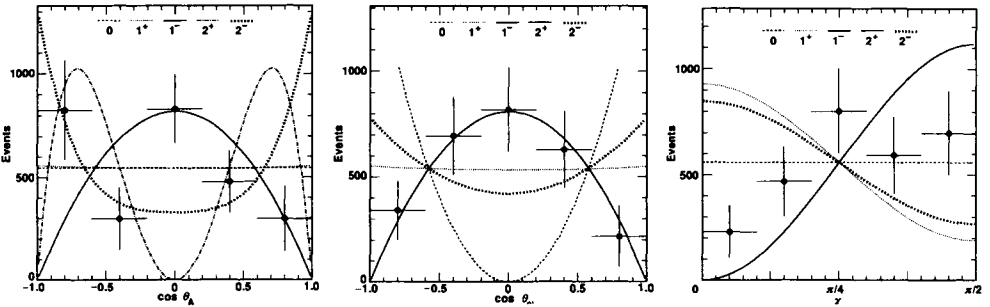
Figure 46. *The angular distribution of θ_A (top-left),θ_ω (top-right) and χ (bottom). The curves show the best fits to the data for different J^P assignments. The 0^- and 1^+ are almost indistinguishable in $\cos\theta_A$, while the 1^- and 2^+ are indistinguishable in $\cos\theta_\omega$ and χ. The vertical axis gives efficiency corrected events (104 data events are used).*

The data are fit to the expectations for the various J^P assignments. The ω polarisation is very clearly transverse ($\sin^2\theta_\omega$) and that infers a 1^- or 2^+ assignment. The $\cos\theta_A$ distribution prefers 1^-, as does the fit to all three projections.

This structure is identified with the ρ' because it has the correct J^P and is at approximately the right mass. To determine the mass and width parameters, which are not well known, we write the decay width as a function of $\omega\pi^-$ mass as

$$d\Gamma(B \to D\omega\pi^-) = \frac{1}{2M_B}\left|A(B \to D\rho') \cdot \mathcal{BW}(m_{\omega\pi}) \cdot A(\rho' \to \omega\pi^-)\right|^2 \quad (62)$$

$$\times \; dP(B \to D\rho') \cdot dP(\rho' \to \omega\pi^-)\frac{dm_{\omega\pi}^2}{2\pi},$$

where dP indicates phase space and the Breit-Wigner amplitude is given by

$$\mathcal{BW}(m_{\omega\pi}) = \frac{1}{(m_{\omega\pi}^2 - m_{\rho'}^2) - im_{\omega\pi}\Gamma_{\text{tot}}(m_{\omega\pi})}. \tag{63}$$

The Breit-Wigner fit assuming a single resonance and no background gives a mass of $1349\pm25^{+10}_{-5}$MeV with an intrinsic width of $547\pm86^{+46}_{-45}$MeV. The fit shows that the $\omega\pi^-$ mass spectrum is consistent with being entirely one resonance. This state is likely to be the elusive ρ' resonance (Clegg 1994). These are by far the most accurate and least model dependent measurements of the ρ' parameters. Since the ρ' dominates the final state, the branching ratios for $D^{(*)}\omega\pi^-$ apply also to $D^{(*)}\rho'^-$.

Heavy quark symmetry predicts equal partial widths for $D^*\rho'$ and $D\rho'$. CLEO measures the relative rates to be $\Gamma(B \to D^*\rho'^-)/\Gamma(B \to D\rho'^-) = 1.06\pm0.17\pm0.04$, consistent within the relatively large errors.

Factorisation predicts that the fraction of longitudinal polarisation of the D^{*+} is the same as in the related semileptonic decay $B \to D^*\ell^-\bar{\nu}$ at four-momentum transfer q^2 equal to the mass-squared of the ρ'

$$\frac{\Gamma_L(B \to D^{*+}\rho'^-)}{\Gamma(B \to D^{*+}\rho'^-)} = \left.\frac{\Gamma_L(B \to D^*\ell^-\bar{\nu})}{\Gamma(B \to D^*\ell^-\bar{\nu})}\right|_{q^2=m_{\rho'}^2} \tag{64}$$

CLEO's measurement of the D^{*+} polarisation is $(63\pm9)\%$. The model predictions in semileptonic decays for a q^2 of 2GeV^2, are between 66.9 and 72.6% (Isgur 1990, Wirbel 1985, Neubert 1991). Figure 47 shows the measured polarisations for the $D^{*+}\rho'^-$, the $D^{*+}\rho^-$, (Artuso 1999) and the $D^{*+}D_s^{*-}$ final states (Stone 2000). The latter is based on a new measurement using partial reconstruction of the D^{*+} (Ahmed 2000). Thus this prediction of factorisation is satisfied.
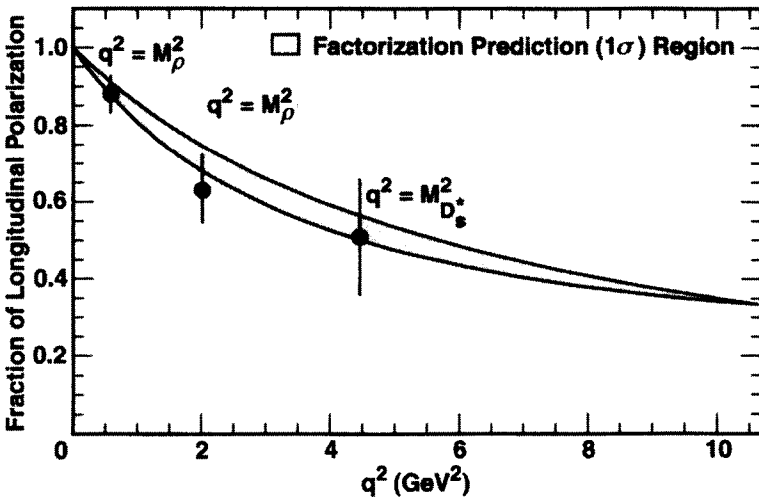
Figure 47. *Measured D^{*+} polarisation versus semileptonic model predictions.*

7 CP violation

7.1 Introduction

Recall that the operation of Charge Conjugation (C) takes particle to anti-particle and Parity (P) takes a vector $\mathbf{r}$ to $-\mathbf{r}$. CP violation can occur because of the imaginary term in the CKM matrix, proportional to η in the Wolfenstein representation (Bigi 2000).

Consider the case of a process $B \to f$ that goes via two amplitudes A and B each of which has a strong phase, s_A, and a weak phase, w_A. Then we have

$$\Gamma(B \to f) = \left(|\mathcal{A}| e^{i(s_A + w_A)} + |\mathcal{B}| e^{i(s_B + w_B)}\right)^2 \tag{65}$$

$$\Gamma(\overline{B} \to \overline{f}) = \left(|\mathcal{A}| e^{i(s_A - w_A)} + |\mathcal{B}| e^{i(s_B - w_B)}\right)^2 \tag{66}$$

$$\Gamma(B \to f) - \Gamma(\overline{B} \to \overline{f}) = 2|\mathcal{AB}| \sin(s_A - s_B) \sin(w_A - w_B). \tag{67}$$

Any two amplitudes will do, though it is better that they be of approximately equal size. Thus charged B decays can exhibit CP violation as well as neutral B decays. In some cases, we will see that it is possible to guarantee that $|\sin(s_A - s_B)|$ is unity, so we can get information on the weak phases. In the case of neutral B decays, mixing can be the second amplitude.

7.2 Unitarity triangles

The unitarity of the CKM matrix, namely

$$\begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} V_{ud}^* & V_{cd}^* & V_{td}^* \\ V_{us}^* & V_{cs}^* & V_{ts}^* \\ V_{ub}^* & V_{cb}^* & V_{tb}^* \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

allows us to construct six relationships of which the most useful turns out to be

$$V_{ud}V_{td}^* + V_{us}V_{ts}^* + V_{ub}V_{tb}^* = 0. \tag{68}$$

To a good approximation

$$V_{ud} \approx V_{tb}^* \approx 1, \qquad V_{ts}^* \approx -V_{cb}, \qquad V_{us} = \lambda, \tag{69}$$

so that Equation 68 becomes

$$\frac{1}{\lambda} \frac{V_{td}^*}{V_{cb}} - 1 + \frac{1}{\lambda} \frac{V_{ub}}{V_{cb}} = 0. \tag{70}$$

Now

$$\frac{1}{\lambda} \left| \frac{V_{td}}{V_{ts}} \right| = \frac{1}{\lambda} \sqrt{(\rho - 1)^2 + \eta^2} \tag{72}$$

$$\frac{1}{\lambda} \left| \frac{V_{ub}}{V_{cb}} \right| = \frac{1}{\lambda} \sqrt{\rho^2 + \eta^2}, \tag{73}$$
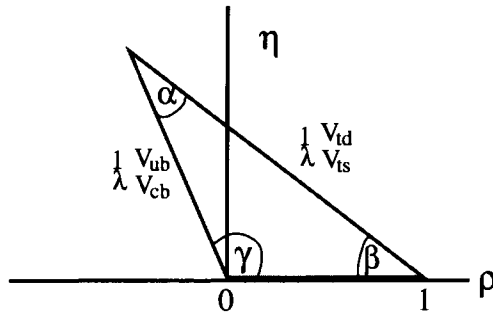
Figure 48. *The CKM unitarity triangle in the ρ–η plane. The left side is determined by measurements of $b \to u/b \to c$ and the right side can be determined using mixing measurements in the B_s and B_d systems. The angles can be found by making measurements of CP violating asymmetries in hadronic B decays.*

so that Equation 70 defines a triangle in the ρ–η plane. This is the CKM triangle depicted in Figure 48. We know the length of two sides already: the base is defined as unity and the left side is determined by the measurements of $|V_{ub}/V_{cb}|$, but the error is still quite substantial. The right side can be determined using mixing measurements in the neutral B systems. Figure 48 also shows the angles labelled α, β, and γ. These angles can be determined by measuring CP violation in the B system.

Another constraint on ρ and η is given by the measurement of ϵ in CP violation in K_L^0 decays (Buras 1995):

$$\eta \left[(1 - \rho) A^2 (1.4 \pm 0.2) + 0.35 \right] A^2 \frac{B_K}{0.75} = (0.30 \pm 0.06), \qquad (74)$$

where B_K is parameter that cannot be measured and thus must be calculated. A reasonable range is $0.6 < B_K < 0.9$, given by an assortment of theoretical calculations (Buras 1995). This number is one of the largest sources of

uncertainty on the ϵ constraint. Other constraints come from current measurements on V_{ub}/V_{cb}, B_d mixing and B_s mixing. The current status of these constraints on ρ and η is shown in Figure 49 (Hocker 2001). The width of these bands are also dominated by theoretical errors. This shows that the data are consistent with the standard model.

It is crucial to check if measurements of the sides and angles are consistent, i.e., whether or not they actually form a triangle. If they do not the standard model is incomplete. The standard model has many parameters including the four CKM numbers, six quark masses, gauge boson masses and coupling constants. Perhaps measurements of the angles and sides of the unitarity triangle will bring us beyond the standard model and show us how these parameters are related, or what is missing.

New physics can also be observed by measuring branching ratios which violate standard model predictions. Especially important are rare decay processes such as $B \to K\mu^+\mu^-$ or $D \to \pi\mu^+\mu^-$. These processes occur only through loops, and are an important class of penguin decays.
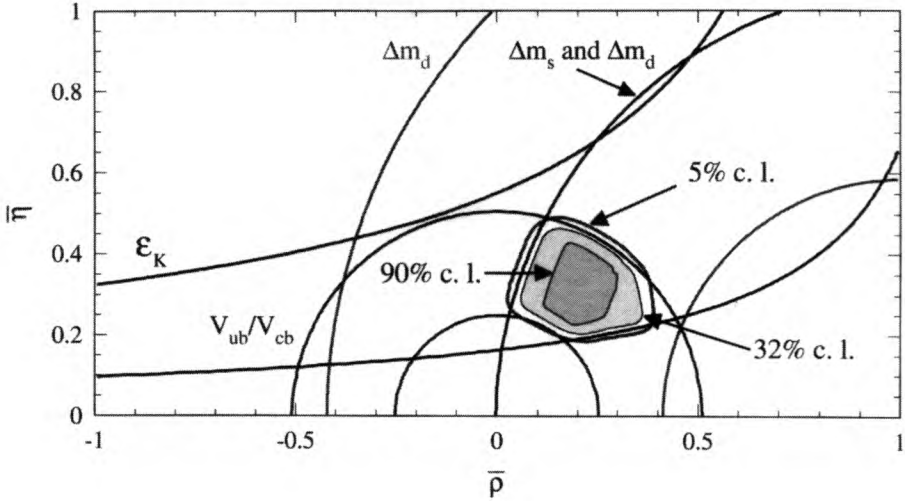
Figure 49. *The regions in $\bar{\rho}-\bar{\eta}$ space (shaded), where $\bar{\rho} = \rho(1-\lambda^2/2)$ and $\bar{\eta} = \eta(1-\lambda^2/2)$, consistent with measurements of CP violation in K_L^0 decay (ϵ), V_{ub}/V_{cb} in semileptonic B decay, B_d^0 mixing, and the excluded region from limits on B_s^0 mixing. The allowed region is defined by a fit from (Hocker 2001). The large width of the B_d mixing band is dominated by the uncertainty in $B_B f_B^2$. The lines that are not specified are at 5% confidence level.*

7.3   Formalism of CP violation in neutral B decays

Consider the operations of Charge Conjugation, C, and Parity, P:

$$C|B(\mathbf{p})\rangle \ = \ |\overline{B}(\mathbf{p})\rangle, \qquad\qquad C|\overline{B}(\mathbf{p})\rangle \ = \ |B(\mathbf{p})\rangle \tag{75}$$

$$P|B(\mathbf{p})\rangle \ = -|B(-\mathbf{p})\rangle, \qquad\quad P|\overline{B}(\mathbf{p})\rangle \ = -|\overline{B}(-\mathbf{p})\rangle \tag{76}$$

$$CP|B(\mathbf{p})\rangle \ = -|\overline{B}(-\mathbf{p})\rangle, \qquad CP|\overline{B}(\mathbf{p})\rangle \ = -|B(-\mathbf{p})\rangle. \tag{77}$$

For neutral mesons we can construct the CP eigenstates

$$|B_1^0\rangle \ = \ \frac{1}{\sqrt{2}}\left(|B^0\rangle - |\overline{B}^0\rangle\right), \tag{78}$$

$$|B_2^0\rangle \ = \ \frac{1}{\sqrt{2}}\left(|B^0\rangle + |\overline{B}^0\rangle\right), \tag{79}$$

where

$$CP|B_1^0\rangle \ = \ |B_1^0\rangle, \tag{80}$$

$$CP|B_2^0\rangle \ = \ -|B_2^0\rangle. \tag{81}$$

Since B^0 and $\overline{B}^0$ can mix, the mass eigenstates are a superposition of $a|B^0\rangle + b|\overline{B}^0\rangle$ which obey the Schrodinger equation

$$i\frac{d}{dt}\begin{pmatrix} a \\ b \end{pmatrix} = \mathcal{H}\begin{pmatrix} a \\ b \end{pmatrix} = \left(M - \frac{i}{2}\Gamma\right)\begin{pmatrix} a \\ b \end{pmatrix}. \tag{82}$$

If CP is not conserved then the eigenvectors, the mass eigenstates $|B_L\rangle$ and $|B_H\rangle$, are not the CP eigenstates but are

$$|B_L\rangle = p|B^0\rangle + q|\overline{B}^0\rangle, \qquad |B_H\rangle = p|B^0\rangle - q|\overline{B}^0\rangle, \tag{83}$$

where

$$p = \frac{1}{\sqrt{2}} \frac{1 + \epsilon_B}{\sqrt{1 + |\epsilon_B|^2}}, \qquad q = \frac{1}{\sqrt{2}} \frac{1 - \epsilon_B}{\sqrt{1 + |\epsilon_B|^2}}. \tag{84}$$

CP is violated if $\epsilon_B \neq 0$, which occurs if $|q/p| \neq 1$.

The time dependence of the mass eigenstates is

$$|B_L(t)\rangle = e^{-\Gamma_L t/2} e^{im_L t/2} |B_L(0)\rangle \tag{85}$$
$$|B_H(t)\rangle = e^{-\Gamma_H t/2} e^{im_H t/2} |B_H(0)\rangle, \tag{86}$$

leading to the time evolution of the flavor eigenstates as

$$|B^0(t)\rangle = e^{-(im+\Gamma/2)t} \left(\cos \frac{\Delta m t}{2} |B^0(0)\rangle + i\frac{q}{p} \sin \frac{\Delta m t}{2} |\overline{B}^0(0)\rangle \right) \tag{87}$$

$$|\overline{B}^0(t)\rangle = e^{-(im+\Gamma/2)t} \left(i\frac{p}{q} \sin \frac{\Delta m t}{2} |B^0(0)\rangle + \cos \frac{\Delta m t}{2} |\overline{B}^0(0)\rangle \right), \tag{88}$$

where $m = (m_L + m_H)/2$, $\Delta m = m_H - m_L$, $\Gamma = \Gamma_L \approx \Gamma_H$, and t is the decay time in the B^0 rest frame, the so called proper time. Note that the probability of a B^0 decay as a function of t is given by $\langle B^0(t)|B^0(t)\rangle^*$, and is a pure exponential, $e^{-\Gamma t/2}$, in the absence of CP violation.

7.3.1 CP violation for B via interference of mixing and decays

Here we choose a final state f which is accessible to both B^0 and $\overline{B}^0$ decays (Bigi 2000). The second amplitude necessary for interference is provided by mixing. Figure 50 shows the decay into f either directly or indirectly via mixing. It is necessary only that f be
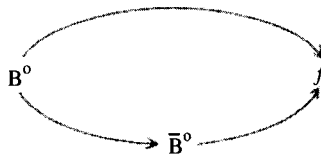
Figure 50. *Two interfering ways for a B^0 to decay into a final state f.*

accessible directly from either state; however if f is a CP eigenstate the situation is far simpler. For CP eigenstates

$$CP|f_{CP}\rangle = \pm|f_{CP}\rangle. \tag{89}$$

It is useful to define the amplitudes

$$A = \langle f_{CP}|\mathcal{H}|B^0\rangle, \qquad \overline{A} = \langle f_{CP}|\mathcal{H}|\overline{B}^0\rangle. \tag{90}$$

If $\left|\overline{A}/A\right| \neq 1$, then we have "direct" CP violation in the decay amplitude, which we will discuss in detail later. In general CP can be violated by having

$$\lambda = \frac{q}{p} \cdot \frac{\overline{A}}{A} \neq 1, \tag{91}$$

which is also the case if λ acquires a non-zero phase, i.e $|\lambda|$ could be unity and CP violation can occur.

A comment on neutral B production at e^+e^- colliders is in order. At the $\Upsilon(4S)$ resonance there is coherent production of $B^0\overline{B}^0$ pairs. This puts the B's in a $C = -1$ state. In hadron colliders, or at e^+e^- machines operating at the Z^0, the B's are produced incoherently. For the rest of this article I will assume incoherent production except where explicitly noted.

The asymmetry, in this case, is defined as

$$a_{f_{\mathrm{CP}}} = \frac{\Gamma\left(B^0(t) \to f_{\mathrm{CP}}\right) - \Gamma\left(\overline{B}^0(t) \to f_{\mathrm{CP}}\right)}{\Gamma\left(B^0(t) \to f_{\mathrm{CP}}\right) + \Gamma\left(\overline{B}^0(t) \to f_{\mathrm{CP}}\right)}, \tag{92}$$

which for $|q/p| = 1$ gives

$$a_{f_{\mathrm{CP}}} = \frac{(1 - |\lambda|^2)\cos(\Delta mt) - 2\mathrm{Im}\lambda\sin(\Delta mt)}{1 + |\lambda|^2}. \tag{93}$$

For the cases where there is only one decay amplitude A, $|\lambda|$ equals 1, and we have

$$a_{f_{\mathrm{CP}}} = -\mathrm{Im}\,\lambda\,\sin(\Delta mt). \tag{94}$$

Only the amplitude, $-\mathrm{Im}\lambda$, contains information about the level of CP violation, the sine term is determined only by B_d mixing. In fact, the time integrated asymmetry is given by

$$a_{f_{\mathrm{CP}}} = -\frac{x}{1 + x^2}\mathrm{Im}\,\lambda = -0.48\mathrm{Im}\,\lambda. \tag{95}$$

This is quite lucky as the maximum size of the coefficient for any x is -0.5.

Let us now find out how $\mathrm{Im}\,\lambda$ relates to the CKM parameters. Recall $\lambda = (q/p)(\overline{A}/A)$. The first term is the part that comes from mixing:

$$\frac{q}{p} = \frac{\left(V_{tb}^* V_{td}\right)^2}{|V_{tb}V_{td}|^2} = \frac{(1 - \rho - i\eta)^2}{(1 - \rho + i\eta)(1 - \rho - i\eta)} = e^{-2i\beta} \tag{96}$$

and

$$\mathrm{Im}\,\frac{q}{p} = -\frac{2(1 - \rho)\eta}{(1 - \rho)^2 + \eta^2} = \sin(2\beta). \tag{97}$$

To evaluate the decay part we need to consider specific final states. For example, consider $f \equiv \pi^+\pi^-$. The simple spectator decay diagram is shown in Figure 37 (left). For the moment we will assume that this is the only diagram which contributes, although we know this is not true. For this $b \to u\bar{u}d$ process we have

$$\frac{\overline{A}}{A} = \frac{\left(V_{ud}^* V_{ub}\right)^2}{|V_{ud}V_{ub}|^2} = \frac{(\rho - i\eta)^2}{(\rho - i\eta)(\rho + i\eta)} = e^{-2i\gamma}, \tag{98}$$

and

$$\text{Im}\,(\lambda) = \text{Im}\,(e^{-2i\beta}e^{-2i\gamma}) = \text{Im}\,(e^{2i\alpha}) = -\sin(2\alpha). \qquad (99)$$

For our next example let us consider the final state $J/\psi K_s$. The decay diagram is shown in Figure 40. In this case we do not get a phase from the decay part because

$$\frac{\overline{A}}{A} = \frac{\left(V_{cb}V_{cs}^*\right)^2}{|V_{cb}V_{cs}|^2} \qquad (100)$$

is real to order $1/\lambda^4$.

In this case the final state is a state of negative CP, i.e. $CP|J/\psi K_s\rangle = -|J/\psi K_s\rangle$. This introduces an additional minus sign in the result for $\text{Im}\,\lambda$. Before finishing discussion of this final state we need to consider in more detail the presence of the K_s in the final state. Since neutral kaons can mix, we pick up another mixing phase (similar diagrams as for B^0, see Figure 26). This term creates a phase given by

$$\left(\frac{q}{p}\right)_K = \frac{\left(V_{cd}^*V_{cs}\right)^2}{|V_{cd}V_{cs}|^2}, \qquad (101)$$

which is real to order λ^4. It necessary to include this term, however, since there are other formulations of the CKM matrix than Wolfenstein, which have the phase in a different location. It is important that the physics predictions not depend on the CKM convention. (Here we don't include CP violation in the neutral kaon since it is much smaller than what is expected in the B decay. The term of order λ^4 in V_{cs} is necessary to explain K^0 CP violation.)

In summary, for the case of $f = J/\psi K_s$, $\text{Im}\,\lambda = -\sin(2\beta)$.

7.3.2   Comment on the penguin amplitude

In principle all processes can have penguin components. One such diagram is shown in Figure 37(right). The $\pi^+\pi^-$ final state is expected to have a rather large penguin amplitude $\sim 10\%$ of the tree amplitude. Then $|\lambda| \neq 1$ and $a_{\pi\pi}(t)$ develops a $\cos(\Delta mt)$ term. It turns out that $\sin(2\alpha)$ can be extracted using isospin considerations and measurements of the branching ratios for $B^+ \to \pi^+\pi^0$ and $B^0 \to \pi^0\pi^0$, or other methods, the easiest of which appears to be the study of $B^0 \to \rho\pi$.

In the $J/\psi K_s$ case, the penguin amplitude is expected to be small since a $c\bar{c}$ pair must be "popped" from the vacuum. Even if the penguin decay amplitude were of significant size, the decay phase is the same as the tree level process, namely zero.

7.4   Results on $\sin 2\beta$

For years observation of large CP violation in the B system was considered to be one of the corner stone predictions of the Standard Model. Yet it took a very long time to come up with definitive evidence. The first statistically significant measurements of CP violation in the B system were made recently by BABAR and BELLE. This enormous achievement was accomplished using an asymmetric e^+e^- collider on the $\Upsilon(4S)$ which

Experiment	$\sin 2\beta$
BABAR	$0.59 \pm 0.14 \pm 0.05$
BELLE	$0.99 \pm 0.14 \pm 0.06$
Average	0.79 ± 0.11
CDF	$0.79^{+.41}_{-0.44}$
ALEPH	$0.84^{+0.82}_{-1.04} \pm 0.16$
OPAL	$3.2^{+1.8}_{-2.0} \pm 0.5$

Table 9. *Statistically significant measurements of* $\sin 2\beta$ *compared with previous results.*

was first suggested by Pierre Oddone. The measurements are listed in Table 9, along with other previous indications (Groom 2001).

The average value of 0.79 ± 0.11 is taken from BABAR and BELLE only. These two measurements do differ by a sizeable amount, but the confidence level that they correctly represent the same value is 6%. The value of 0.79 is consistent with what is expected from the other known constraints on ρ and η. We have

$$\bar{\eta} = (1 - \bar{\rho}) \left\{ \frac{1 \pm \sqrt{1 - \sin^2 2\beta}}{\sin 2\beta} \right\}. \tag{102}$$

There is a four fold ambiguity in the translation between $\sin 2\beta$ and the linear constraints in the $\rho - \eta$ plane. These occur at β, $\pi/2 - \beta$, $\pi + \beta$ and $3\pi/2 - \beta$. Two of these constraints are shown in Figure 51. The other two can be viewed by extending these to negative $\bar{\eta}$. We think $\eta > 0$ based only on the measurement of ϵ in the neutral kaon system.
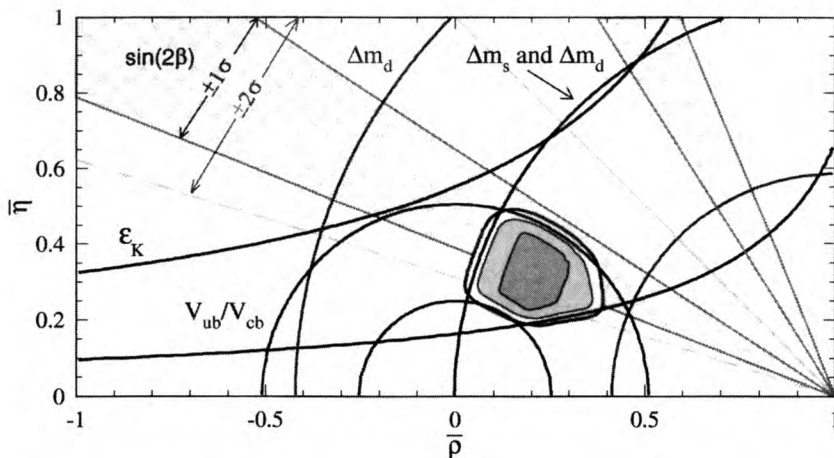
Figure 51. *Constraints from* $\sin 2\beta$ *measurement overlaid with other constraints (Hocker 2001). The inner band is at* 1σ *while the outer band, shown on one band only, is at* 2σ.

7.5   Remarks on global fits to CKM parameters

The fits shown in this paper (Hocker 2001) for ρ and η have been done by others with a somewhat different statistical framework (Ciuchini 2001, Mele 1999). The latter group uses "Bayesian" statistics which means that they use a priori knowledge of the probability distribution functions. The former are termed "frequentists" (Groom 2001), almost for the lack of a better term. The frequentists are more conservative than the Bayesians.

The crux of the issue is how to treat theoretically predicted parameters that are used to translate measurements into quantities such as V_{ub} or ϵ_K that form constraints in the $\rho - \eta$ plane. This is of course a problem because it is difficult to estimate the uncertainties in the theoretical predictions. Both groups treat the experimental measurements as Gaussian distributions with the 1σ width derived from combined statistical and systematic errors. Note, that experimental systematic errors are also difficult to quantify and are not necessarily Gaussian, but they are judged to be sufficiently well known to not cause a problem.

Hocker et al. (2001) have decided to use a method which restricts the theoretical quantities to a 95% confidence interval where the parameter in question is just as likely to lie anywhere in the interval. They call this the "Rfit" method. Of course assigning the 95% confidence interval is a matter of judgment which they fully admit. Ciuchini et al. (2001) treat the theoretical errors in the same manner as the experimental errors. They call theirs "the standard method" with just a bit of hubris. They argue that QCD is mature enough to trust its predictions, that they know the sign and rough magnitude of corrections and they can assign reasonable errors, so it would be wrong to throw away information.

Hocker et al. point out an interesting but generally unknown danger with the Bayesian approach, which is that in multi-dimension problems the Bayesian treatment unfairly predicts a narrowing of possible results. The following discussion will demonstrate this.

Let x_i denote N theoretical parameters over the identical ranges $[-\Delta, +\Delta]$; then the theoretical prediction is

$$T_P^{(N)} = \prod_i^N x_i. \tag{103}$$

In the 95% scan scheme $[T_P^{(N)}] = [-\Delta^N, +\Delta^N]$, while in the Bayesian approach the convoluted Probability Density Function (PDF) is

$$\rho(T) = \int_{-\infty}^{+\infty} ... \int_{-\infty}^{+\infty} \prod_i^N dx_i G(x_i)\delta(T - T_P^{(N)}), \tag{104}$$

where the $G(x_i)$ are the PDF's for the individual variables which taken to be equal here. This function has a singularity in $\rho(T)$ that goes as $(-\ln T)^{N-1}$, so it increases as N increases.

Now suppose $G(x_i)$ is flat, then for $N = 1$ both approaches are the same, but for $N \geq 2$, the Bayesian approach gets a $\rho(T)$ that peaks at zero. In effect, when the number of theoretical predictions entering the computation increases, and hence our knowledge of the corresponding observable decrease the Bayesian approach claims the converse.

Lets look at a specific example where $N = 3$ and $\Delta = \sqrt{3}$. Consider both the sum $T_S = x_1 + x_2 + x_3$ and product $T_P = x_1 x_2 x_3$ distributions. For Rfit the allowed ranges are
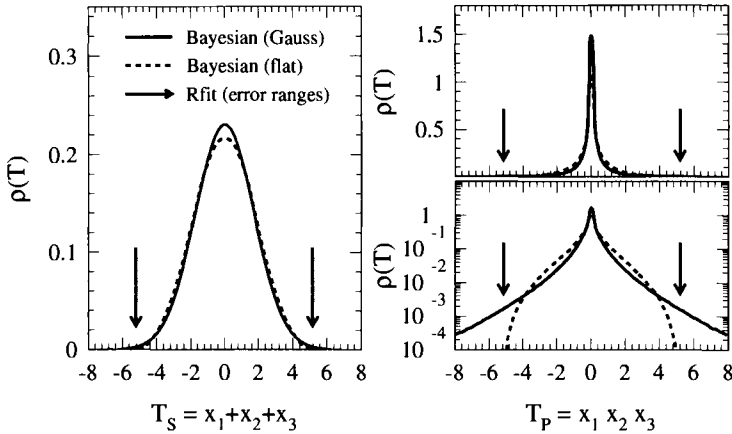
Figure 52. *Convolution of the sum $T_S = x_1 + x_2 + x_3$ for both Rfit and Bayesian methods (left) and the product $T_P = x_1 x_2 x_3$ (right) of 3 parameters for the Bayesian method only. Plotted is the PDF $\rho(T)$ obtained using for $G(x)$ a uniform (solid lines, $\Delta = \sqrt{3}$) or a Gaussian (dashed lines, $\sigma = 1$) distribution. Both PDF's of T_P have a singularity at the origin which is not shown. The Rfit ranges of T_S and T_P are indicated by the arrows located in both instances at $\pm 3\sqrt{3}$. From (Hocker 2001).*

identical being $[-3\Delta, 3\Delta]$. The left side of Figure 52 shows the probability density $\rho(T)$ for T_S, while the right side shows $\rho(T)$ for T_P with $G(x)$ in the Bayesian case being either a Gaussian with $\sigma = 1$ (solid lines) or a uniform distribution over the range $[-\sqrt{3}, +\sqrt{3}]$ (dashed lines). The later distribution is closest to the Rfit method where the allowed range for either T is $[-3\sqrt{3}, +3\sqrt{3}]$ indicated by the arrows.

In the Rfit scheme the two predictions for T_P and T_S are identical, while in the Bayesian scheme there is a large difference in the PDF's and it really doesn't matter if a Gaussian or uniform distribution is chosen. There is a clear distinction between the Rfit and Bayesian predictions for T_P in this case, and the Bayesian one is unreasonable because it produces a very narrow PDF peaked very close to zero.

An example of how this can effect the results is shown on Figure 53 where predictions of $\sin 2\beta$ from the indirect measurements are shown for the Rfit technique and either uniform or Gaussian Bayesian PDF's. The predictions are quite different.
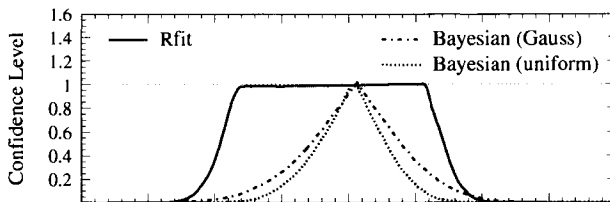
Figure 53. *Comparison between Rfit (broad solid curve) and Bayesian fits for the indirect CKM constraints on $\sin 2\beta$. For the Bayesian fits: Gaussian (uniform) theoretical PDF's are depicted as dashed-dotted (dotted) curves. (This example does not include the newer data.)*

8   New Physics studies

8.1   Introduction

There are many reasons why we believe that the Standard Model is incomplete and there must be physics beyond. One is the plethora of "fundamental parameters," for example quark masses and mixing angles. The Standard Model cannot explain the smallness of the weak scale compared to the GUT or Planck scales; this is often called "the hierarchy problem." In the Standard Model it is believed that the CKM source of CP violation extensively discussed here is not large enough to explain the baryon asymmetry of the Universe (Gavela 1993). Finally, gravity is not incorporated. John Ellis said "My personal interest in CP violation is driven by the search for physics beyond the Standard Model" (Ellis 2000). We can also take the view that we will discover additional large unexpected effects in b and/or c decays.

We must realise that *all* our current measurements are a combination of Standard Model and New Physics; any proposed models must satisfy current constraints. Since the Standard Model tree level diagrams are probably large, let us consider them as a background to New Physics. Therefore loop diagrams and CP violation are the best places to see New Physics. The most important current constraints on New Physics models are

- The neutron electric dipole moment, $d_N < 6.3 \times 10^{-26}$e-cm.

- $\mathcal{B}(b \to s\gamma) = (3.23 \pm 0.42) \times 10^{-4}$ and $\mathcal{B}(b \to s\ell^+\ell^-) < 4.2 \times 10^{-5}$.

- CP violation in K_L decay, $\epsilon_K = (2.271 \pm 0.017) \times 10^{-3}$.

- B^0 mixing parameter $\Delta m_d = (0.487 \pm 0.014)$ ps^{-1}.

8.2   Generic tests for new physics

We can look for New Physics either in the context of specific models or more generically, looking for any deviations from the Standard Model expectation. One example is to examine the rare decays $B \to K\ell^+\ell^-$ and $B \to K^*\ell^+\ell^-$ for branching ratios and polarisations. According to Greub et al., "Especially the decay into K^* yields a wealth of new information on the form of the new interactions since the Dalitz plot is sensitive to subtle interference effects" (Greub 1995). Another important tactic is to test for inconsistencies in Standard Model predictions independent of specific non-standard models.

The unitarity of the CKM matrix allows us to construct six relationships. These may be thought of as triangles in the complex plane shown in Figure 54. (The **bd** triangle is the one depicted in Figure 48.)

All six of these triangles can be constructed knowing four and only four independent angles (Silva 1997) (Aleksan 1994). These are defined as:

$$\beta = \arg\left(-\frac{V_{tb}V_{td}^*}{V_{cb}V_{cd}^*}\right), \qquad \gamma = \arg\left(-\frac{V_{ub}^*V_{ud}}{V_{cb}^*V_{cd}}\right), \tag{105}$$

$$\chi = \arg\left(-\frac{V_{cs}^*V_{cb}}{V_{ts}^*V_{tb}}\right), \qquad \chi' = \arg\left(-\frac{V_{ud}^*V_{us}}{V_{cd}^*V_{cs}}\right). \tag{106}$$
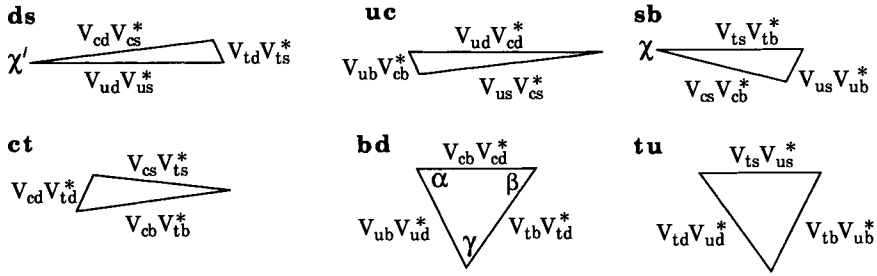
Figure 54. *The six CKM triangles. The bold labels, e.g.* **ds** *refer to the rows or columns used in the unitarity relationship. The angles defined in equation (105) are also shown.*

(α can be used instead of γ or β.) Two of the phases β and γ are probably large while χ is estimated to be small ≈ 0.02, but measurable, while χ' is likely to be much smaller.

It has been pointed out by Silva and Wolfenstein (Silva 1997) that measuring only angles may not be sufficient to detect new physics. For example, suppose there is new physics that arises in $B^0 - \overline{B}^0$ mixing. Let us assign a phase θ to this new physics. If we then measure CP violation in $B^0 \to J/\psi K_s$ and eliminate any penguin pollution problems in using $B^0 \to \pi^+\pi^-$, then we actually measure $2\beta' = 2\beta + \theta$ and $2\alpha' = 2\alpha - \theta$. So while there is new physics, we miss it, because $2\beta' + 2\alpha' = 2\alpha + 2\beta$ and $\alpha' + \beta' + \gamma = 180°$.

8.2.1 A critical check using χ

The angle χ, defined in Equation 105, can be extracted by measuring the time dependent CP violating asymmetry in the reaction $B_s \to J/\psi \eta^{(\prime)}$, or if one's detector is incapable of high quality photon detection, the $J/\psi\phi$ final state can be used. However, in this case there are two vector particles in the final state, making this a state of mixed CP, which requires a large statistics sample to do a time-dependent angular analysis to extract χ.

Measurements of the magnitudes of CKM matrix elements all come with theoretical errors. Some of these are hard to estimate. The best measured magnitude is that of $\lambda = |V_{us}/V_{ud}| = 0.2205 \pm 0.0018$.

Silva and Wolfenstein (Silva 1997) (Aleksan 1994) show that the Standard Model can be checked in a profound manner by seeing if:

$$\sin \chi = \left| \frac{V_{us}}{V_{ud}} \right|^2 \frac{\sin \beta \; \sin \gamma}{\sin(\beta + \gamma)}. \tag{107}$$

Here the precision of the check will be limited initially by the measurement of $\sin \chi$, not of λ. This check can reveal new physics, even if other measurements have not shown any anomalies.

Other relationships to check include:

$$\sin \chi = \left| \frac{V_{ub}}{V_{cb}} \right|^2 \frac{\sin \gamma \; \sin(\beta + \gamma)}{\sin \beta}, \tag{108}$$

$$\sin \chi = \left| \frac{V_{td}}{V_{ts}} \right|^2 \frac{\sin \beta \; \sin(\beta + \gamma)}{\sin \gamma}. \tag{109}$$

The astute reader will have noticed that these two equations lead to the non-trivial relationship:

$$\sin^2 \beta \left| \frac{V_{td}}{V_{ts}} \right|^2 = \sin^2 \gamma \left| \frac{V_{ub}}{V_{cb}} \right|^2 . \tag{110}$$

This constrains these two magnitudes in terms of two of the angles. Note, that it is in principle possible to determine the magnitudes of $|V_{ub}/V_{cb}|$ and $|V_{td}/V_{ts}|$ without model dependent errors by measuring β, γ and χ accurately. Alternatively, β, γ and λ can be used to give a much more precise value than is possible at present with direct methods. For example, once β and γ are known

$$\left| \frac{V_{ub}}{V_{cb}} \right|^2 = \lambda^2 \frac{\sin^2 \beta}{\sin^2(\beta + \gamma)} . \tag{111}$$

Table 10 lists the most important physics quantities and the decay modes that can be used to measure them. The necessary detector capabilities include the ability to collect purely hadronic final states labeled here as "Hadron Trigger," the ability to identify charged hadrons labeled as "$K\pi$ sep," the ability to detect photons with good efficiency and resolution and excellent time resolution required to analyze rapid B_s oscillations. Measurements of $\cos(2\phi)$ can eliminate 2 of the 4 ambiguities in ϕ that are present when $\sin(2\phi)$ is measured.

Physics Quantity	Decay Mode	Hadron Trigger	$K\pi$ sep	γ det	Decay time σ
$\sin(2\alpha)$	$B^0 \to \rho\pi \to \pi^+\pi^-\pi^0$	√	√	√	
$\cos(2\alpha)$	$B^0 \to \rho\pi \to \pi^+\pi^-\pi^0$	√	√	√	
$\mathrm{sign}(\sin(2\alpha))$	$B^0 \to \rho\pi$ & $B^0 \to \pi^+\pi^-$	√	√	√	
$\sin(\gamma)$	$B_s \to D_s^\pm K^\mp$	√	√		√
$\sin(\gamma)$	$B^- \to \overline{D}^0 K^-$	√	√		
$\sin(\gamma)$	$B^0 \to \pi^+\pi^-$ & $B_s \to K^+K^-$	√	√		√
$\sin(2\chi)$	$B_s \to J/\psi\eta', J/\psi\eta$			√	√
$\sin(2\beta)$	$B^0 \to J/\psi K_s$				
$\cos(2\beta)$	$B^0 \to J/\psi K^0, K^0 \to \pi\ell\nu$		√		
$\cos(2\beta)$	$B^0 \to J/\psi K^{*o}$ & $B_s \to J/\psi\phi$				
x_s	$B_s \to D_s^+\pi^-$	√			√
$\Delta\Gamma$ for B_s	$B_s \to J/\psi\eta', D_s^+\pi^-, K^+K^-$	√	√	√	√

Table 10. *Required CKM Measurements for b's*

8.2.2   Finding inconsistencies

Another interesting way of viewing the physics was given by Peskin (2000). Non-Standard Model physics would show up as discrepancies among the values of (ρ, η) derived from independent determinations using CKM magnitudes ($|V_{ub}/V_{cb}|$ and $|V_{td}/V_{ts}|$), or B_d^0 mixing (β and α), or B_s mixing (χ and γ).

8.2.3 Required measurements involving β

Besides a more precise measurement of $\sin 2\beta$ we need to resolve the ambiguities. There are two suggestions on how this may be accomplished. Kayser (1997) showed that time dependent measurements of the final state $J/\psi K^0$, where $K^0 \to \pi \ell \nu$, give a direct measurement of $\cos(2\beta)$ and can also be used for CPT tests. Another suggestion is to use the final state $J/\psi K^{*o}$, $K^{*o} \to K_s \pi^0$, and to compare with $B_s \to J/\psi \phi$ to extract the sign of the strong interaction phase shift assuming SU(3) symmetry, and thus determine $\cos(2\beta)$ (Dighe 1998).

8.2.4 Required measurements involving α and γ

It is well known that $\sin(2\beta)$ can be measured without problems caused by penguin processes using the reaction $B^0 \to J/\psi K_s$. The simplest reaction that can be used to measure $\sin(2\alpha)$ is $B^0 \to \pi^+ \pi^-$. This reaction can proceed via both the tree and penguin diagrams shown in Figure 37.

Current measurements (Table 6) show a large penguin component. The ratio of the penguin *amplitude* to the tree *amplitude* in the $\pi^+ \pi^-$ channel is about 15% in magnitude. Thus the effect of the penguin must be determined in order to extract α. The only model independent way of doing this was suggested by Gronau and London, but requires the measurement of $B^\mp \to \pi^\mp \pi^0$ and $B^0 \to \pi^0 \pi^0$, the latter being rather daunting.

There is however, a theoretically clean method to determine α. The interference between tree and penguin diagrams is exploited by measuring the time dependent CP violating effects in the decays $B^0 \to \rho\pi \to \pi^+ \pi^- \pi^0$ as shown by Snyder and Quinn (1993).

The $\rho\pi$ final state has many advantages. First of all, it has been seen with a relatively large rate. The branching ratio for the $\rho^0 \pi^+$ final state as measured by CLEO is $(1.5\pm0.5\pm 0.4) \times 10^{-5}$, and the rate for the neutral B final state $\rho^\pm \pi^\mp$ is $(3.5^{+1.1}_{-1.0} \pm 0.5) \times 10^{-5}$, while the $\rho^0 \pi^0$ final state is limited at 90% confidence level to $< 5.1 \times 10^{-6}$ (Gao 1999). BABAR (Bona 2001) measures $\mathcal{B}(B^0 \to \rho^\pm \pi^\mp)$ as $(28.9 \pm 5.4 \pm 4.3) \times 10^{-6}$. These measurements are consistent with some theoretical expectations (Ali 1999). The associated vector-pseudoscalar $K^* \pi$ decay modes have conquerable or smaller branching ratios. Since the ρ is spin-1, the π spin-0 and the initial B is spinless, the ρ is fully polarised in the (1,0) configuration, so it decays as $\cos^2 \theta$, where θ is the angle of one of the ρ decay products with respect to the other π from the B, determined in the ρ rest frame. This causes the periphery of the Dalitz plot to be heavily populated, especially the corners. A sample Dalitz plot is shown in Figure 55. This kind of distribution is good for maximising the interferences, which helps minimise the error. Furthermore, little information is lost by excluding the Dalitz plot interior, a good way to reduce backgrounds.

To estimate the required number of events Snyder and Quinn preformed an idealised analysis that showed that a background-free, flavor-tagged sample of 1000 to 2000 events was sufficient. The 1000 event sample usually yields good results for α, but sometimes does not resolve the ambiguity. With the 2000 event sample, however, they always succeeded.

This technique not only finds $\sin(2\alpha)$, it also determines $\cos(2\alpha)$, thereby removing two of the remaining ambiguities. The final ambiguity can be removed using the CP asymmetry in $B^0 \to \pi^+ \pi^-$ and a theoretical assumption (Grossman 1997).
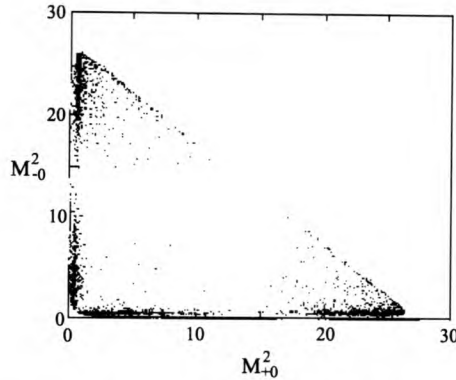
Figure 55. *The Dalitz plot for $B^0 \to \rho\pi \to \pi^+\pi^-\pi^0$ from Snyder and Quinn (1993).*

Several model dependent methods using the light two-body pseudoscalar decay rates have been suggested for measuring γ. The basic idea in all these methods can be sum-marised as follows: the tree diagram for $B^0 \to \pi^+\pi^-$ has the weak decay phase γ. In order to reproduce the observed suppression of the decay rate for $\pi^+\pi^-$ relative to $K^\pm\pi^\mp$ we require a large negative interference between the tree and penguin amplitudes. This puts γ in the range of 90°. A great deal of theoretical work is required to understand the effects of re-scattering, form-factors etc. We are left with several ways of obtaining model dependent limits, due to Fleischer and Mannel (1998), Neubert and Rosner (1998), Fleis-cher and Buras (2000), and Beneke *et al.* (2001). The latter use a sophisticated model of QCD factorisation and apply corrections. Figure 56 shows values of γ that can be found
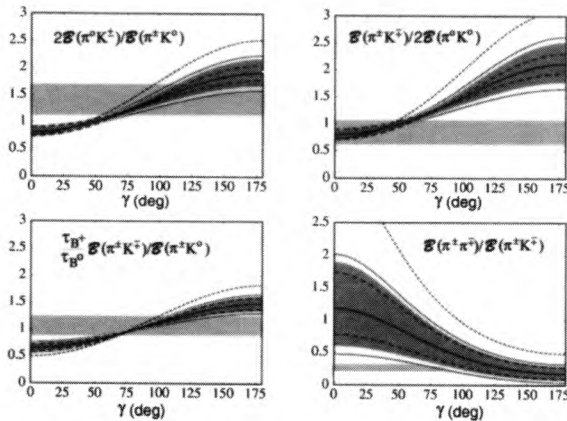
Figure 56. *Model predictions from (Beneke 2001) as a function of the indicated rate ratios. The dotted curve shows the predictions from naive factorisation. The curved bands show the total model uncertainties where the inner band comes from theoretical input uncertainties, while the outer band allows for errors to corrections on the theory. The specific sensitivity to $|V_{ub}|$ is showed as the dashed curves. The gray bands show the current data with a 1σ error while the lighter bands are at 2σ.*
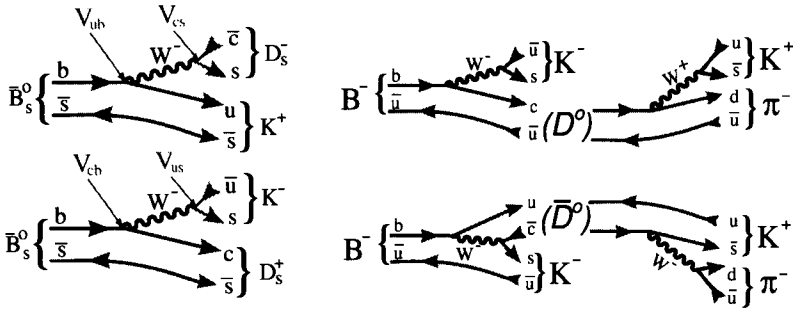
Figure 57. *(left) The two diagram diagrams for $B_s \to D_s^{\pm} K^{\mp}$ that interfere via B_s mixing. (right) The two interfering decay diagrams for $B^- \to \overline{D}^0 K^-$ where one is a $b \to u$ transition and the other a doubly-Cabibbo suppressed decay.*

in their framework, once better data are obtainable.

In fact, it may be easier to measure γ than α in a model independent manner. There have been two methods suggested.

(1) Time dependent flavor tagged analysis of $B_s \to D_s^{\pm} K^{\mp}$. This is a direct model independent measurement (Du 1986) (Aleksan 1992) (Aleksan 1995). Here the Cabibbo suppressed V_{ub} decay interferes with a somewhat less suppressed V_{cb} decay via B_s mixing as illustrated in Figure 57 (left). Even though we are not dealing with CP eigenstates here there are no hadronic uncertainties, though there are ambiguities.

(2) Measure the rate differences between $B^- \to \overline{D}^0 K^-$ and $B^+ \to D^0 K^+$ in two different D^0 decay modes such as $K^- \pi^+$ and $K^+ K^-$. This method makes use of the interference between the tree and doubly-Cabibbo suppressed decays of the D^0, and does not depend on any theoretical modeling (Atwood 1997, Gronau 1991). See Figure 57 (right).

8.3 New Physics tests in specific models

8.3.1 Supersymmetry

Supersymmetry is a kind of super-model. The basic idea is that for every fundamental fermion there is a companion boson and for every boson there is a companion fermion. There are many different implementations of couplings in this framework (Masiero 2000). In the most general case we pick up 80 new constants and 43 new phases. This is clearly too many to handle so we can try to see things in terms of simpler implementations. In the minimum model (MSSM) we have only two new fundamental phases. One θ_D would arise in B^0 mixing and the other, θ_A, would appear in B^0 decay. A combination would generate CP violation in D^0 mixing, call it $\phi_{K\pi}$ when the decay is $D^0 \to K^- \pi^+$ (Nir 1999). Table 11 shows the CP asymmetry in three different processes in the Standard Model and the MSSM.

Two direct effects of New Physics are clear here. First of all, the difference in CP asymmetries between $B^0 \to J/\psi K_s$ and $B^0 \to \phi K_s$ would show the phase ϕ_A. Secondly,

Process	Standard Model	New Physics
$B^0 \to J/\psi K_s$	$\sin 2\beta$	$\sin 2(\beta + \theta_D)$
$B^0 \to \phi K_s$	$\sin 2\beta$	$\sin 2(\beta + \theta_D + \theta_A)$
$D^0 \to K^-\pi^+$	0	$\sim \sin \phi_{K\pi}$

Table 11. *CP Violating Asymmetries in the Standard Model and the MSSM.*

there would be finite CP violation in $D^0 \to K^-\pi^+$ where none is expected in the Standard Model.

Manifestations of specific SUSY models lead to different patterns. Table 12 shows the expectations for some of these models in terms of these variables and the neutron electric dipole moment d_N; see (Nir 1999) for details. Note, that "Approximate CP" has already

Model	$d_N \times 10^{-25}$	θ_D	θ_A	$\sin \phi_{K\pi}$
Standard Model	$\leq 10^{-6}$	0	0	0
Approx. Universality	$\geq 10^{-2}$	$\mathcal{O}(0.2)$	$\mathcal{O}(1)$	0
Alignment	$\geq 10^{-3}$	$\mathcal{O}(0.2)$	$\mathcal{O}(1)$	$\mathcal{O}(1)$
Heavy squarks	$\sim 10^{-1}$	$\mathcal{O}(1)$	$\mathcal{O}(1)$	$\mathcal{O}(10^{-2})$
Approx. CP	$\sim 10^{-1}$	$-\beta$	0	$\mathcal{O}(10^{-3})$

Table 12. *Some SUSY Predictions.*

been ruled out by the measurements of $\sin 2\beta$.

In the context of the MSSM there will be significant contributions to B_s mixing, and the CP asymmetry in the charged decay $B^\mp \to \phi K^\mp$. The contribution to B_s mixing significantly enhances the CP violating asymmetry in modes such as $B_s \to J/\psi\eta$. (Recall the CP asymmetry in this mode is proportional to $\sin 2X$ in the Standard Model.) The Standard Model and MSSM diagrams are shown in Figure 58. The expected CP asymmetry in the MSSM is $\approx \sin \phi_\mu \cos \phi_A \sin(\Delta m_s t)$, which is approximately 10 times the expected value in the Standard Model (Hinchliff 2001a).
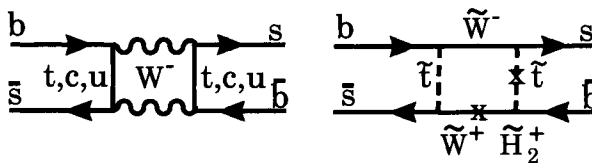
Figure 58. *The Standard Model (left) and MSSM (right) contributions to B_s^0 mixing.*

We observed that a difference between CP asymmetries in $B^0 \to J/\psi K_s$ and ϕK_s arises in the MSSM due to a CP asymmetry in the decay phase. It is possible to observe this directly by looking for a CP asymmetry in $B^\mp \to \phi K^\mp$. The Standard Model and MSSM
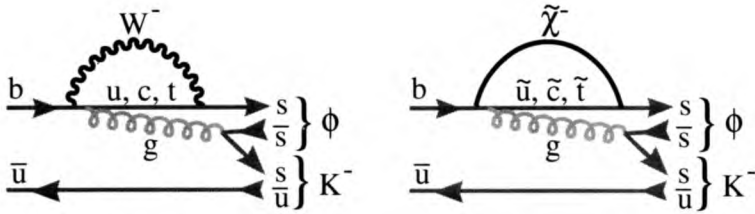
Figure 59. *The Standard Model (left) and MSSM (right) contributions to $B^- \to \phi K^-$.*

diagrams are shown in Figure 59. Here the interference of the two diagrams provides the CP asymmetry. The predicted asymmetry is equal to $(M_W/m_{\text{squark}})^2 \sin \phi_\mu$ in the MSSM, where m_{squark} is the relevant squark mass (Hinchliff 2001a).

The ϕK and ϕK^* final states have been observed, first by CLEO (Briere 2001) and subsequently by BABAR (Aubert 2001). The BABAR data is shown in Figure 60. The average branching ratio is $\mathcal{B}(B^- \to \phi K^-) = (6.8 \pm 1.3) \times 10^{-6}$ showing that in principle large samples can be acquired especially at hadronic machines.
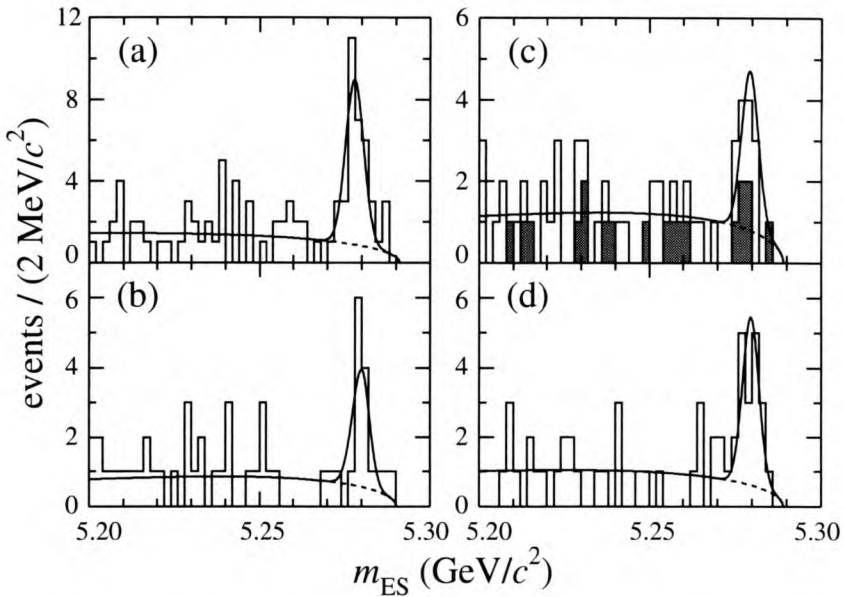
Figure 60. *Projections of the maximum likelihood fit onto the B candidate mass m_{ES} from BABAR. (a) $B^+ \to \phi K^+$; (b) $B^0 \to \phi K^0$; (c) $B^+ \to \phi K^{*+}$; (d) $B^0 \to \phi K^{*o}$. In (c) the histogram is the sum of the two contributing K^* channels while the shaded area is $K^0\pi^+$ alone (the other channel is $K^+\pi^0$). The solid line shows the projection of the signal plus background fit while the dashed line shows the projection of the background only.*

8.3.2   Other new physics models

There are many other specific models that predict New Physics in b decays. I list here a few of these with a woefully incomplete list of references, to give a flavor of what these models predict.

- *Two Higgs and Multi-Higgs Doublet Models* predict large effects in ϵ_K and CP violation in $D^0 \to K^- \pi^+$ with only a few percent effect in B^0 (Nir 1999). Expect to see 1–10% CP violating effects in $b \to s\gamma$ (Wolfenstein 1994).

- *Left-Right Symmetric Model* have contributions that compete with or even dominate over Standard Model contributions to B_d and B_s mixing. This means that CP asymmetries into CP eigenstates could be substantially different from the Standard Model prediction (Nir 1999).

- *Extra Down Singlet Quarks* are likely to give dramatic deviations from Standard Model predictions for CP asymmetries in b decays (Nir 1999).

- *FCNC Couplings of the Z boson* in the decay $B \to K^* \ell^+ \ell^-$ mean that both the sign and magnitude of the decay leptons carry sensitive information on new physics. Potential effects are on the of 10% compared to an entirely negligible Standard Model asymmetry of $\sim 10^{-3}$ (Buchalla 2000). These models also predict a factor of 20 enhancement of $b \to d\ell^+ \ell^-$ and could explain a low value of $\sin 2\beta$ (Barenboim 2001a).

- *Non-commutative Geometry.* If the geometry of space time is non-commutative, i.e. $[x_\mu, x_\nu] = i\theta_{\mu\nu}$, then CP violating effects may be manifest a low energy. For a scale $<2\text{TeV}$ there are comparable effects to the Standard Model (Hinchliffe 2001b).

- *MSSM without new flavor structure* can lead to CP violation in $b \to s\gamma$ of up to 5% (Bartl 2001). Ali and London propose (Ali 1999) that the Standard Model formulas are modified by Supersymmetry as

$$\Delta m_d = \Delta m_d(\text{SM}) \left[1 + f\left(m_{\chi_2^\pm}, m_{\tilde{i}_R}, m_{H^\pm}, \tan\beta \right) \right] \tag{112}$$

$$\Delta m_s = \Delta m_s(\text{SM}) \left[1 + f\left(m_{\chi_2^\pm}, m_{\tilde{i}_R}, m_{H^\pm}, \tan\beta \right) \right] \tag{113}$$

$$|\epsilon_K| = \frac{G_F^2 f_K^2 M_K M_W^2}{6\sqrt{2}\pi^2 \Delta M_K} B_K(A^2\lambda^6 \bar\eta) \left\{ y_c \left(\eta_{ct} f_3(y_c, y_t) - \eta_{cc} \right) \right.$$
$$\left. + \eta_{tt} y_t f_s(y_t) \left[1 + f\left(m_{\chi_2^\pm}, m_{\tilde{i}_R}, m_{H^\pm}, \tan\beta \right) \right] A^2\lambda^4(1 - \bar\rho) \right\}, \tag{114}$$

where $\Delta m(SM)$ refers to the Standard Model formula and the expression for $|\epsilon_K|$ would be the Standard Model expression if f were set equal to zero. Ali and London show that it is reasonable to expect that $0.8 > f > 0.2$. Since the CP violating angles will not change from the Standard Model, determining the value of (ρ, η) using the magnitudes $\Delta m_s / \Delta m_d$ and $|\epsilon_K|$ will show an inconsistency with values obtained using other magnitudes and angles.

- *Extra Dimensions* are beginning appear in papers predicting b decay phenomena. See Agashe (2001), Barenboim (2001b), Branco (2001), Chang (2001) and Papavassiliou (2000).

I close this section with a quote from Masiero and Vives (2001): "The relevance of SUSY searches in rare processes is not confined to the usually quoted possibility that indirect searches can arrive 'first' in signaling the presence of SUSY. Even after the possible direct observation of SUSY particles, the importance of FCNC and CP violation in testing SUSY remains of utmost relevance. They are and will be complementary to the Tevatron and LHC establishing low energy supersymmetry as the response to the electroweak breaking puzzle."

I agree, except that I would replace "SUSY" with "New Physics."

8.4 Expected data samples

It is clear that precision studies of b decays can bring a wealth of information to bear on new physics, that probably will be crucial in sorting out anything seen at the LHC. This is possible because we do expect to have data samples large enough to test these ideas from existing and approved experiments. In Table 13 I show the expected rates in BTeV for one year of running (10^7 s) and an e^+e^- B-factory operating at the $\Upsilon(4S)$ with a total accumulated sample of 500 fb-1, about what is expected around 2006. (LHCb numbers are the same order of magnitude as the BTeV numbers for many of the modes.)

Mode	BTeV $(10^7)s$			B-factory (500 fb^{-1})		
	Yield	Tagged†	S/B	Yield	Tagged†	S/B
$B_s \to J/\psi\eta^{(\prime)}$	22000	2200	>15			
$B^- \to \phi K^-$	11000	11000	>10	700	700	4
$B^0 \to \phi K_s$	2000	200	5.2	250	75	4
$B^0 \to K^{*o}\mu^+\mu^-$	4400	4400	11	$\sim$50	$\sim$50	?
$D^{*+} \to \pi^+ D^0; D^0 \to K^-\pi^+$	$\sim 10^8$	$\sim 10^8$	large	8×10^5	8×10^5	large

$\dagger$ Tagged here means that the initial flavor of the B is determined.

Table 13. *Comparisons of BTeV and B-factory Yields on Different Time Scales.*

9 Conclusions

I have attempted to cover the length and breadth of b physics, and have only scratched the surface. There is much more to be said and much more to learn. Why are there three families? What is the connection with neutrinos and that mixing matrix? How do we explain the mystery of flavor? These and many more unanswered questions I leave to the students.

Acknowledgments

References

Abe F, et al., 1998, *Phys. Rev. D* **58**, p.5513, [hep-ex/9804012].

Abe K, et al., 2001a, Measurement of Branching Fractions for $B \to \pi\pi$, $K\pi$ and KK Decays, *Phys. Rev. Lett.* **87**, p.101801, [hep-ex/0104030].

Abbiendi G, et al., 2000, *Phys. Lett. B* **482**, p.15 [hep-ex/0003013].

Abreu P, et al., 2000, Determination of $|V_{ub}|/|V_{cb}|$ with DELPHI at LEP, *Phys. Lett. B* **478**, p.14 [hep-ex/0105054].

Abreu P, et al., 2001, *Phys. Lett. B* **510**, p.55 [hep-ex/0104026].

Acciarri M, et al., 1998, *Phys. Lett. B* **436**, p.174.

Agashe K, Deshpande N G and Wu G H, 2001, Universal Extra Dimensions and $b \to s\gamma$, *Phys. Lett. B* **514**, p.309 [hep-ph/0105084].

Ahmed S, et al., 2000, *Phys. Rev. D*,**62**, p.112003 [hep-ex/0008015].

Akers R, et al., 1995, *Z. Phys. C* **66**, p.555.

Alam M S, et al., 1995, *Phys. Rev. Lett.*, **74**, p.2285.

Albajar, et al., 1987, *Phys. Lett. B* **186**, p.247, erratum-ibid B**197** p.565.

Albrecht H, et al., 1983, *Phys. Lett. B* **192**, p.245.

Albrecht H, et al., 1990, *Phys. Lett. B* **234**, p.16.

Aleksan R, Dunietz I and Kayser B, 1992, *Z. Phys. C* **54**, p.653.

Aleksan R, Kayser B and London D, 1994, *Phys. Rev. Lett.* **73**, p.18 [hep-ph/9403341].

Aleksan R, et al., 1995, *Z. Phys. C* **67**, p.251 [hep-ph/9407406].

Alexander J, et al., 1996, *Phys. Rev. Lett.* **77**, p.5000.

Alexander J, et al., 2001, *Phys. Rev. D* **64**, p.092001 [hep-ex/0103021].

Ali A and Greub C, 1991 , *Phys. Lett. B* **259**, p.182.

Ali A, Kramer G and Lu C D, 1999, *Phys. Rev. D* **59** p.014005 [hep-ph/9805403].

Altarelli G, 1982, *Nucl. Phys. B*, **208**, p.365.

Ammar R, et al., 1993, *Phys. Rev. Lett.*, **71**, p.674.

Appel J A, et al., 2001, Performance of Prototype BTeV Silicon Pixel Detectors in a High Energy Pion Beam, [hep-ex/0108014].

Artuso M, 1994, Experimental Facilities for b-quark Physics, in *B Decays 2nd Edition* ed. S. Stone, World Scientific, Singapore, p.80.

Artuso M and Barberio E, 2001, www-cdf.lbl.gov/~weiming/vcb_summary_ckm_workshop.ps .

Atwood D, Dunietz I and Soni A, 1997, *Phys. Rev. Lett.* **78**, p.3257.

Atwood M and Jaros R A, 1994, Lifetimes, *B Decays 2nd Edition* ed. S. Stone, World Scientific, Singapore, p. 364.

Aubert B, et al., 2001, *Phys. Rev. Lett.*, **87**, p.151801, [hep-ex/0105001].

Ball P and Braun V M, 1998, *Phys. Rev. D* **58**, p.094016.

Bander M, Silverman D and Soni A, 1979, *Phys. Rev. Lett.*, **43**, p.242.

Barate R, et al., 1998, *Phys. Lett. B* **429**, p.169.

Barate R, et al., 1999, *Eur. Phys. J.* **C6**, p.555.

Barenboim G, Botella F J and Vives O, 2001a, *Phys. Rev. D* **64**, p.015007 [hep-ph/0012197].

Barenboim G, Botella F J and Vives O, 2001b, Constraining Models With Vector-Like Fermions from FCNC in K and B Physics, *Nucl. Phys. B* **613**, p.285 [hep-ph/01050306].

Bartl A, et al., 2001, *Phys. Rev. D* **64**, p.076009 [hep-ph/0103324].

Bauer C W, Pirjol D and Steward I W, 2001, A Proof of Factorization for $B \to D\pi$, *UCSD-01-12* [hep-ph/0107002].

Bauer M and Wirbel M, 1989, *Z. Phys. C* **42**, p.671.

Bauer C W, Ligeti Z and Luke M, 2001, Precision Determination of $|V_{ub}|$, from Inclusive Decays, *UTPT-01-08* [hep-ph/0107074].

Beneke M, et al., 2000, *Nucl Phys. B* **591**, p.313 [hep-ph/0006124].

Beneke M, et al., 2001, *Nucl Phys. B* **606**, p.245 [hep-ph/0104110].

Beyer M and Melikhov D, 1998, *Phys. Lett. B* **436**, p.344 [hep-ph/9807223].

Bigi II, et al., 1997, *Annu. Rev. Nucl. Part. Sci.* **47**, p.591.

Bigi II and Sanda A I, 2000, *CP Violation*, Cambridge University Press, Cambridge, UK.

Bona M, 2001, *BABAR-CONF-01/71, SLAC-PUB-9045* [hep-ex/0111017].

Bortoletto D and Stone S, 1990, *Phys. Rev. Lett.* **65**, p.2951.

Branco G C, de Gouvea A and Rebelo M N, 2001, Split Fermions in Extra Dimensions and CP Violation, *Phys. Lett. B* **506**, p.115 [hep-ph/0012289].

Briere R A, et al., 2001, *Phys. Rev. Lett.* **86**, p.3718, [hep-ex/0101032].

Browder T E, Honscheid K and Pedrini D, 1996, Nonleptonic Decays and Lifetimes of b-quark and c-quark Hadrons, *Ann. Rev. Nucl. Part. Sci.* **46**, p.395 [hep-ph/9606354].

BTeV C, 2000, BTeV An Experiment to Measure Mixing, CP Violation and Rare Decays of Beauty and Charm at the Fermilab Collider, [hep-ex/0006037].

Buchalla G, Hiller G and Isidori G, 2000, *Phys. Rev. D* **63**, p.014015 [hep-ph/0006136].

Buras A J, 1995, *Nucl. Instrm. and Meth. A*, **368**, p.1.

Buskulic D, et al., 1997, *Phys. Lett. B* **395**, p.373.

Caprini I, Lellouch L and Neubert M. *Nucl. Phys. B* **530**, p.153 [hep-ph/9712417].

Chang D, Keung W Y and Mohapatra R N, 2001, Models for Geometric CP Violation With Extra Dimensions, *Phys. Lett. B* **515**, p.431 [hep-ph/0105177].

Chen S, 2001, Branching Fraction and Photon Energy Spectrum for $b \to s\gamma$, [hep-ex/0108032].

Ciuchini M, 2001, *JHEP* **0107**, p.23 [hep-ph/0012308].

Clegg A B and Donnachie A, 1994 *Z. Phys. C* **62**, p.455.

Debbio L D et al., 1998, *Phys. Lett. B* **416**, p.392.

Deshpande N G, 1994, Theory of Penguins in *B* Decays, in *B Decays 2nd Edition* ed. S. Stone, World Scientific, Singapore, p. 80 and references therein.

Dighe A, Dunietz I and Fleischer R, 1998, *Phys. Lett. B* **433**, p.147 [hep-ph/9804254].

Du D-S, Dunietz I and Wu D-D, 1986, *Phys. Rev. D* **34**, p.3414.

Ellis J, 2000, Highlights of CP 2000, *Nucl. Phys. Proc. Suppl.* **99A**, p.331 [hep-ph/0011396].

Fleischer R and Mannel T, 1998, *Phys. Rev. D* **57**, p.2752 [hep-ph/9704423].

Fleischer R and Buras A J, 2000, Constraints on γ and Strong Phases from $B \to \pi K$ Decays," presented at ICHEP 2000, Osaka, Japan, July 2000. To appear in the Proceedings [hep-ph/0008298].

Fulton R, et al., 1990, *Phys. Rev. Lett.*, **64**, p.16.

Gaillard M and Lee B, 1974, *Phys. Rev. D* **10**, p.897.

Gao Y and Würthwein F, 1999, [hep-ex/9904008].

Gavela M B, Hernández P, Orloff J and Pène O, 1993, Standard Model CP Violation and Baryon Asymmetry, *Mod. Phys. Lett. A* **9**, p.795 [hep-ph/9312215].

Gilman F J and Singleton R, 1990, *Phys. Rev. D* **41**, p.142.

Greub F J, Ioannissian A and Wyler D, 1995, *Phys. Lett. B* **346**, p.149 [hep-ph/9408382].

Grinstein B, Isgur N and Wise M B, 1986 *Phys. Rev. Lett.* **56**, p.258.

Gronau M and Wyler D, 1991, *Phys. Lett. B* **265**, p.172.

Groom D E, et al., 2001, The Particle Data Group, *The European Physics Journal C* **15**, p.1.

Grossman Y and Quinn H R, 1997, *Phys. Rev. D* **56** p.7259 [hep-ph/9705356].

Hashimoto S, et al., 2001, "Lattice Calculation of the Zero-recoil Form Factor of $\overline{B} \to D^* \ell^- \bar{\nu}$: Toward à Model Independent Determination of $|V_{cb}|$," [hep-ph/0110253].

Heltsley B K, 2001, CKM Status and Prospects, *To appear in Proceedings of XXI Physics in Collision, Seoul, Korea, June, 2001*, [hep-ph/0110260]/

Hinchliff I and Kersting N, 2001a, Constraining CP Violating Phases of the MSSM, *Phys. Rev. D* **63**, p.015003 [hep-ph/0003090].

Hinchliff I and Kersting N, 2001b, CP Violation from Noncommutative Geometry, *LBNL-47750* [hep-ph/0104137].

Hocker A, Lacker H, Laplace S, Le Diberder F, 2001, A New Approach to a Global Fit of the CKM Matrik, *Eur. Phys. J. C* **21**, p.25 [hep-ph/0104062].

Hurth T, 2001, "Inclusive Rare B Decays, *CERN-TH/2001-146*, [hep-ph/0106050].

Isgur N and Scora D, Grinstein B and Wise M B, 1989, *Phys. Rev. D* **39**, p.799.

Isgur N and Wise M B, 1990, *Phys. Lett. B* **237**, p.527.

Isgur N and Wise M B, 1994, "Heavy Quark Symmetry," in *B Decays 2nd Edition* ed. S. Stone, World Scientific, Singapore, p. 231 and references therein.

Isgur N and Scora D, 1995, *Phys. Rev. D* **52**, p.2783.

Isgur N, 1999, Duality-Violating $1/m_Q$ Effects in Heavy Quark Decays, *Phys. Lett. B*, **448**, p.111 [hep-ph/98113777].

Kayser B, 1997, Cascade Mixing and the CP-Violating Angle Beta, [hep-ph/9709382],

Kobayashi M and Maskawa K, 1973, *Prog. Theor. Phys.* **49**, p.652.

Körner J G and Schuler G A, 1988, *Z. Phys. C* **38**, p.511.

Leroy O, 2001, *http://lepbosc.web.cern.ch/LEPBOSC/people.html* and references therein.

Ligeti Z and Wise M, 1996, *Phys. Rev. D* **53**, p.4937.

Lingel K, Skwarnicki T and Smith J G, 1998, Penguin Decays of B Mesons, *Ann. Rev. of Nucl. & Part. Science* **48**, p.253 [hep-ex/9804015].

Luke M, 1990, *Phys. Lett. B* **252** p.447.

Masiero A and Vives O, 2000, New Physics Behind the Standard Model's Door?, *Int. School on Subnuclear Physics, Erice, Italy, 1999* [hep-ph/0003133].

Masiero A and Vives O, 2001, New Physics in CP Violation Experiments, *Ann. Rev. of Nucl. & Part. Science* **51**, [hep-ph/0104027].

Mele S, 1999, *Phys. Rev. D* **59**, p.113011.

Melikhov D, 1996, *Phys. Rev. D* **53**, p.2460.

Muheim F, 2001, Status of the LHCb Experiment, *Nucl. Instrum and Meth. A* **462**, p. 233 [hep-ex/0012059].

Neubert M, 1991, *Phys. Lett. B* **264**, p.455.

Neubert M, 1996, *Int. J. Mod. Phys. A* **11**, p.4173 [hep-ph/9604412].

Neubert M and Stech B, 1997, "Non-Leptonic Weak Decays of B Mesons, in *Heavy Flavours, 2nd Edition*, ed. Buras A J and Lindner M, World, Scientific, Singapore, p.294 [hep-ph/9705292].

Neubert M and Rosner J L, 1998, *Phys. Rev. Lett.* **81**, p.5076 [hep-ph/9809311].

Neubert M, 2000, *JHEP* **7**, p.22 [hep-ph/0006068].

Nir Y, 1999, CP Violation In and Beyond the Standard Model, *IASSNS-HEP-99-96* [hep-ph/9911321].

Papavassiliou J and Santamaria A, 2001, Extra Dimensions at the One Loop Level: $Z \to b\bar{b}$ and $B - \overline{B}$ Mixing, *Phys. Rev. D* **63**, p.016002 [hep-ph/016002].

Peskin M E, 2000, Theoretical Summary Lecture for EPS HEP99, to appear in proceedings [hep-ph/0002041].

Ramirez C, Donoghue J F, and Burdman G, 1990, *Phys. Rev. D* **41**, p.1496.

Richman J D and Burchat P R, 1995, *Rev. Mod. Phys.* **67**, p.893 and references therein.

Rosner J, 2001, The Standard Model in 2001, *In these proceedings,* and references therein [hep-ph/0108195].

Sachrajda C T, 1999, Lattice B-physics, *Nucl. Instru. and Meth. A* **462**, p.23 [hep-lat/9911016].

Sharma V A and Weber F V, 1994, Recent Measurements of the Lifetimes of b Hadrons, *B Decays 2nd Edition* ed. S. Stone, World Scientific, Singapore, p. 395.

Silva J P and Wolfenstein L, 1997, *Phys. Rev. D* **55**, p.5331 [hep-ph/9610208].

Skwarnicki T, 2001, Overview of the BTeV Experiment, *Nucl. Instrum and Meth. A* **462**, p. 227.

Snyder A E and Quinn H R, 1993, *Phys. Rev. D* **48**, p.2139.

Stone S, 2000, Hadronic *B* Decays to Charm from CLEO, in *Proc. of XXXth Int. Conf. on High Energy Physics, Aug. 2000, Osaka, Japan* p.842 [hep-ex/0008070].

Tajima H, 2001, "Belle B Physics Results," *To appear in Proceedings of the XX Int. Symp on Lepton and Photon Interactions at High Energies, July, 2001, Rome, Italy,* [hep-ex/0111037].

Wirbel M, et al., 1985, *Z. Phys. C* **25**, p.627.

Wirbel M, Stech B and Bauer M, 1989, *Z. Phys. C* **29**, p.637.

Wise M, 2001, "Recent Progress in Heavy Quark Physics," *To appear in Proceedings of the XX Int. Symp on Lepton and Photon Interactions at High Energies, July, 2001, Rome, Italy,* [hep-ph/0111167].

Wolfenstein L, 1983, *Phys. Rev. Lett.* **51**, p.1945.

Wolfenstein L and Wu Y L, 1994, *Phys. Rev. Lett.* **74**, p.2809 [hep-ph/9410253].

Heavy quarks at HERA-B and the Tevatron

Peter Križan

University of Ljubljana, Slovenia

1 Introduction

The two lectures which are written up in this paper, try to answer the questions why hadron machines look very promising for B physics studies, what are the problems in b hadron detection, and how these are solved in three experiments, CDF, an example of a general purpose collider experiment, HERA-B, a fixed target experiment, and BTeV, a specially designed collider experiment in the forward direction. Some of the advantages of hadron machines are evident from Table 1. The $b\bar{b}$ production cross section is typically much higher in a hadron machine, as compared to an e^+e^- machine. At HERA-B, where the cross-section is comparable to e^+e^- production, the $b\bar{b}$ production rate can

Production Mechanism	$e^+e^- \to \Upsilon(4s)$ $\to B\bar{B}$	$e^+e^- \to Z^0$ $\to b\bar{b}$	$pA \to b\bar{b}X$	$p\bar{p} \to b\bar{b}X$	$p\bar{p}(p) \to b\bar{b}X$ forward
Accelerators	CESR, DORIS	LEP, SLD	HERA (P)	Tevatron	Tevatron
	PEPII, KEKB				
Detectors	CLEO, ARGUS	ALEPH, DELPHI	HERA-B	CDF, D0	BTeV
	BABAR, BELLE	L3, OPAL, SLD			
$\sigma(b\bar{b})$	$\approx$ 1nb	$\approx$ 6nb	$\approx$ 12nb	$\approx 50\mu$b	$\approx 100\mu$b
$\sigma(b\bar{b}){:}\sigma(had)$	0.26	0.22	10^{-6}	10^{-3}	$2 \cdot 10^{-3}$
B^0, B^+	yes	yes	yes	yes	yes
$B_s^0, B_c^+, \Lambda_b^0$	no	yes	yes	yes	yes
boost $\langle\beta\gamma\rangle$	0.06 (0.5)	6	≈ 20	$\approx 2{-}4$	$\approx 4{-}20$
multiple events	no	no	yes, 4	yes	yes, 2
main trigger	inclusive	inclusive	lepton pairs	leptons	displaced vertex

Table 1. *Comparison of $b\bar{b}$ production at e^+e^- and hadron machines*

be improved by increasing the luminosity. Another advantage of b hadrons produced in hadronic collisions, compared to the B-factories at the $\Upsilon(4S)$, is the large boost which results in an average decay length, $\langle L \rangle = 480$ μm. Finally, we note that the $\Upsilon(4S)$ machines only produce B^0 and $B^{\pm}$ mesons, whereas the hadron machines also produce other b hadrons, including B_s, B_c and Λ_b.

When designing a detector for b hadrons at a hadron machine, we have to consider several issues. High interaction rates are needed because of the rather low ratio of the signal cross section to the total cross section. In most cases this means we are dealing with multiple interactions per event, so the detector has to be able to tolerate high track densities (see Figure 1). The detector components that are used have to be radiation hard, and have a sufficiently fine granularity to separate out the track hits.
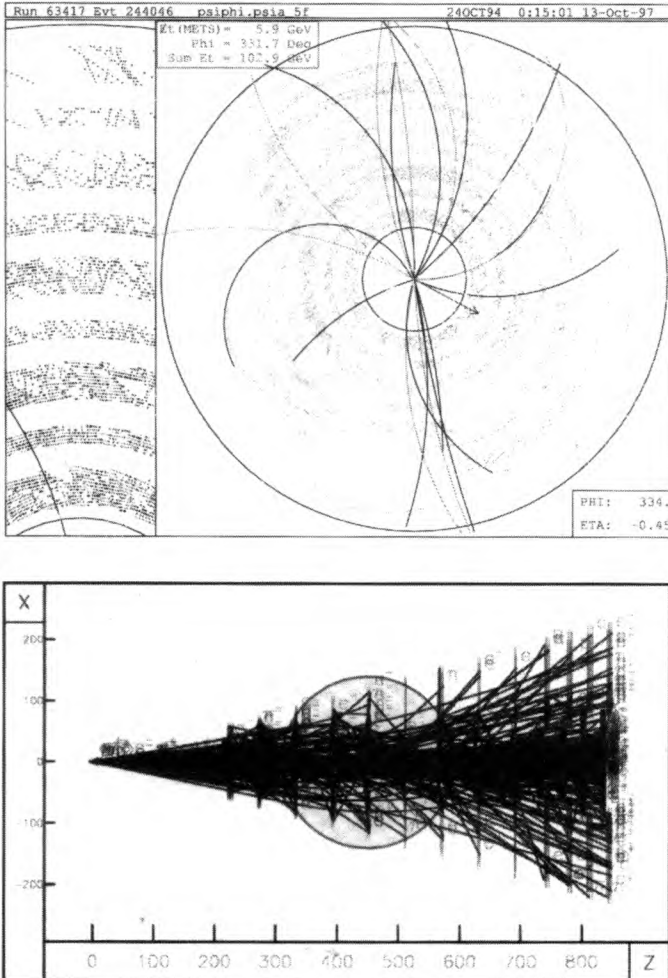
Figure 1. *Comparison of $b\bar{b}$ events: CDF(top), HERA-B(bottom)*

The low ratio of signal to total cross section also requires an early selection of interesting events. Two approaches are used in developing selective triggers with good background suppression. The first approach relies on the fact that the decay products of a B meson have rather high transverse momenta in few-body final states, such as $B \to \mu\nu X$, $B \to J/\psi K_s \to \mu\mu\pi\pi$, $B \to \pi^+\pi^-$. The final states with a single lepton or lepton pairs are particularly attractive because lepton identification can be done rather easily. A second approach takes advantage of the displacement of the decay vertices of b hadrons from the interaction point. In the case of HERA-B this distance is typically 1cm. Both approaches are used in the experiments presented in this paper. The detection of high p_T tracks only takes a few hundred nanoseconds, while the reconstruction of the vertex takes more time, and is a part of the higher trigger levels, although the BTeV experiment plans to reverse this order.

The next section reviews the excellent results from the CDF detector. These serve as a proof of principle for doing B physics at a hadron machine. Then follows a brief discussion of how to measure B_s mixing. The remaining sections describe the HERA-B and BTeV experiments.

2 B physics with CDF

2.1 CDF spectrometer

The Tevatron collider provides $p\bar{p}$ collisions at a center-of-mass energy of 2TeV in two interaction regions, for the two experiments, CDF and D0. The CDF detector (Abe 1988) is a general purpose magnetic spectrometer with a silicon microstrip vertex detector, a vertex TPC, a main tracking chamber, an electromagnetic and a hadronic calorimeter, and a system of muon chambers (Figure 2).
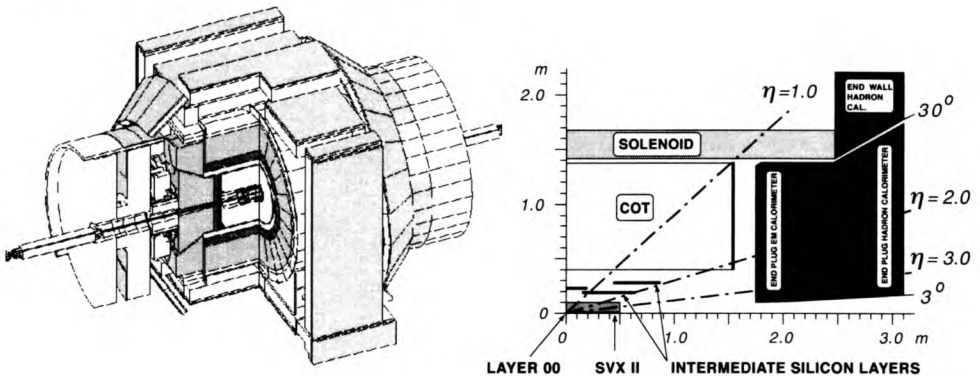
Figure 2. *CDF spectrometer: isometric view (left) with a vertex detector system, a central tracking chamber, electromagnetic and hadronic calorimeters, a superconducting solenoid, and muon chambers; side view after the recent upgrade (right) with a new silicon vertex detector including the additional layer (L00), silicon tracker, central tracking chamber (COT), and additional calorimeters.*
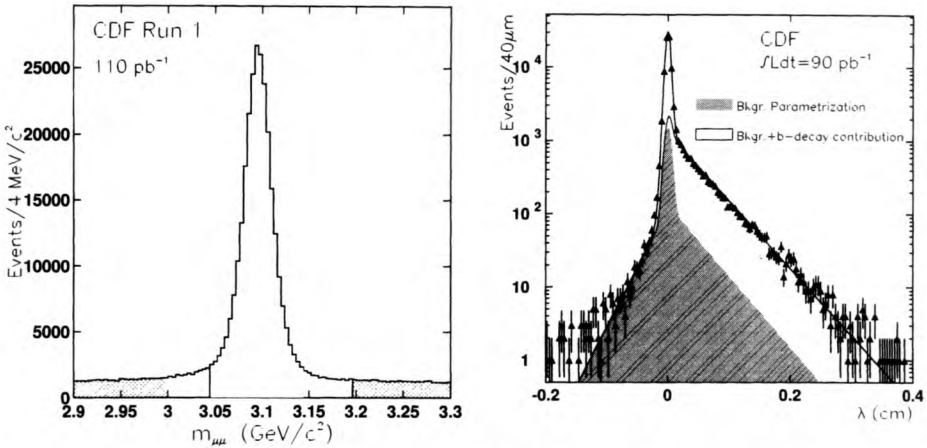
Figure 3. *Examples of CDF performance: distribution of di-muon invariant mass (left), decay length distribution for di-muon events from the J/ψ mass band (right).*

For b physics studies there are two interesting figures of merit illustrated in Figure 3. The di-muon invariant mass resolution in the J/ψ region is 16MeV/c^2, and the resolution on the position of the $J\psi$ vertex is 40–50 μm. The performance of the vertex detector is essential for improving the signal to background ratio, as can be seen for the $B_d \to J/\psi K^+$ decay channel in Figure 4.
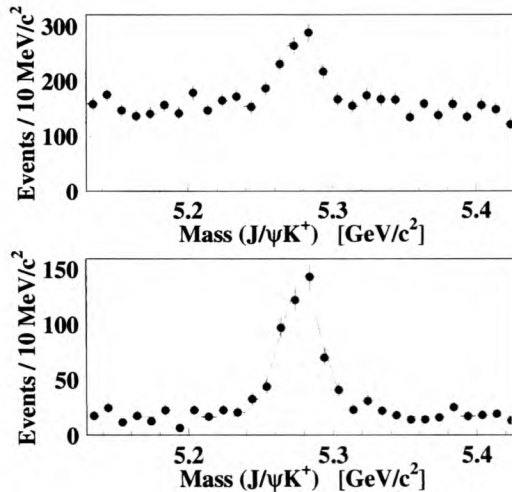
Figure 4. *Cleaning up the $J/\psi K^+$ signal: a cut on decay length ($>100\mu m$) significantly improves the signal to background ratio (bottom).*

2.2 CDF results

This is a brief summary of the results from CDF. The interested reader should consult overview papers (Paulini 1999, Pitts 2000) for more details.

The excellent performance of the vertex detector allowed measurements of the time evolution, and hence of b lifetimes and $B\bar{B}$ mixing. The results of the lifetime measurements of b hadrons are summarized in Figure 5. The precision of these results, and of the B_d mixing measurements (Gay 2000), match the precision of the LEP experiments (Figure 5). A search for a B_s mixing signal resulted in a lower limit $\Delta m_s > 5.8\text{ps}^{-1}$ at 95% confidence level.

Among the contributions to the spectroscopy of b hadrons the most important were the first observation of the B_c meson and the determination of the masses of the B_s, Λ_b and B_c hadrons. CDF also measured the production cross section $\sigma(p\bar{p} \to bX)$ which turns out to be larger than theoretical predictions by about a factor of 2 (Bedeschi 2001).

CDF studied time-dependent CP violation in $B_d \to J/\psi K_s$, obtaining the result $\sin 2\beta = 0.79^{+0.41}_{-0.44}$ (Affolder 2000). For the next generation of precision b physics experiments a measurement of the polarization in the $B_s \to J/\psi\phi$ decay is important as an input for the measurement of $\Delta\Gamma/\Gamma$ (Affolder 2000a). CDF has also shown that rare decays, such as $B_d \to K^*\mu\mu$, can be studied at hadron machines with sensitivities comparable to e^+e^- machines (Affolder 1999).
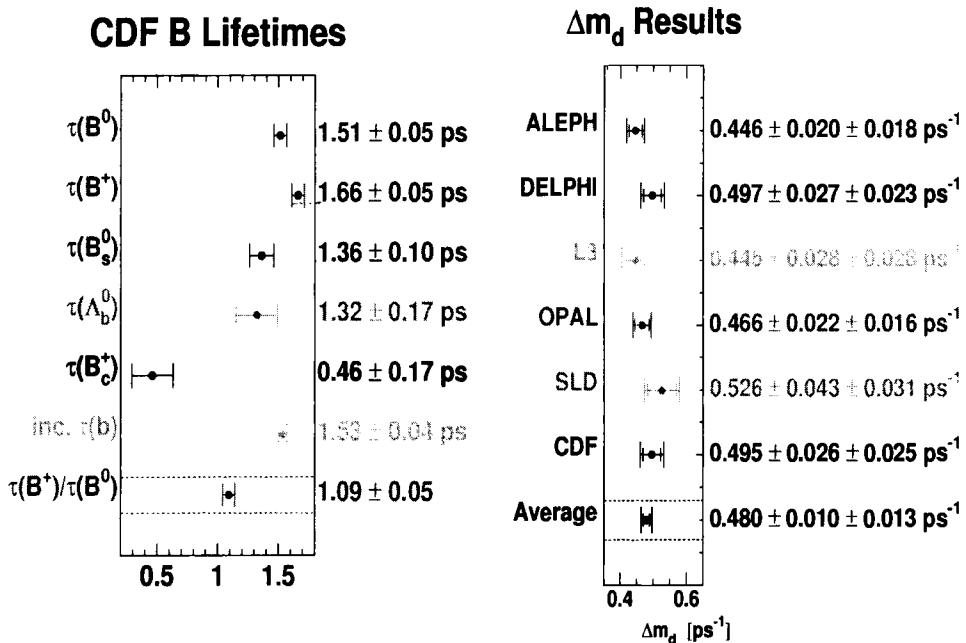
Figure 5. *CDF results: lifetimes of b hadrons (left), B_d mixing parameter Δm_d compared to LEP and SLC results(right).*

2.3   CDF plans for future

For Run II the Tevatron collider has been upgraded with a new injector and antiproton source. The bunch spacing has been changed from $3.5\mu s$ to 396ns, and will be further reduced to 132ns in the second stage. It is expected that 2fb^{-1} of integrated luminosity will be accumulated in the years 2001–2004, and as much as 15fb^{-1} by 2007. Both numbers should be compared to the 0.1fb^{-1} of data collected in Run I.

Major upgrades of the CDF and D0 detectors have been carried out to take advantage of the machine improvements. The upgrades to the CDF detector, shown in Figure 2, include increased muon system coverage, a larger vertex detector, and a new central tracker consisting of a drift chamber and additional silicon layers between the vertex detector and the drift chamber. A time-of-flight counter has been added for π/K separation up to 1.6GeV/c. An additional vertexing layer(L00) close to the beam significantly improves the reach of the experiment for B_s mixing (see next section). The trigger system now includes a fast track search in the first level, and a silicon vertex tracking algorithm in the second level. This will lower the p_t threshold for muons, as well as providing a two track trigger for hadronic decays

With an integrated luminosity of 2fb^{-1} in the first period of Run II, it is expected that the CP violation parameter $\sin 2\beta$ will be measured with an error of 0.043 (Tanaka 2001). A B_s mixing observation will be possible up to $x_s = 60$ at 5σ significance. $\Delta\Gamma/\Gamma$ for B_s mesons will be measured with an error of 0.05 using the decay channel $B_s \to J/\psi\phi$. It is also expected that the forward-backward asymmetry A_{FB} in the rare decay $B_d \to K^*\mu\mu$ will be measured. The $b\bar{b}$ production cross section will be studied further, as well as exclusive decays of the B_c meson, $B_c \to J/\psi\ell\nu$ and $J/\psi\pi$.

3   How to measure B_s mixing

To observe mixing in the B_s system requires a measurement of the probability that a B_s at time $t = 0$ turns into $\bar{B}_s$ when it decays at a later time t. The $\bar{B}_s$ has to decay to a flavour specific final state with a precisely measured decay vertex and $\bar{B}_s$ momentum. Several decay channels have been considered, including the semileptonic channel $B_s \to \mu D_s\nu$, where $D_s \to \phi\pi^+, K^+K_S, K^+K^*$, the hadronic channel $B_s \to J/\psi K^*$, where $J/\psi \to \mu\mu$, $K^* \to K\pi$, as well as more conventional hadronic decays such as $B_s \to D_s^{(*)}\pi^+$ or $D_s^{(*)}\pi^+\pi^+\pi^-$. A flavour tag is needed to determine the initial B_s flavour at $t = 0$. This is best provided by the charge of the kaon or lepton from the associated decay of the other b quark.

In what follows, a simplified picture is given of the maximal mixing parameter x_s an experiment can measure, as a function of the resolution and available statistics. The interested reader should consult Moser & Roussarie 1997 for a full discussion. For a perfectly tagged B_s at t=0, the probability that this B_s meson decays as a $\bar{B}_s$ depends on the proper time t according to $\{1 - \cos(x_s t/\tau)\}\exp(-t/\tau)$. Imperfect tagging and backgrounds reduce the amplitude of the cosine term by a "dilution" factor (D):

$$r(B_s \to \bar{B}_s) = e^{-t/\tau}\{1 - D\cos(x_s t/\tau)\}.$$

For lepton tagged events the value of D is about 0.19 (Figure 6).
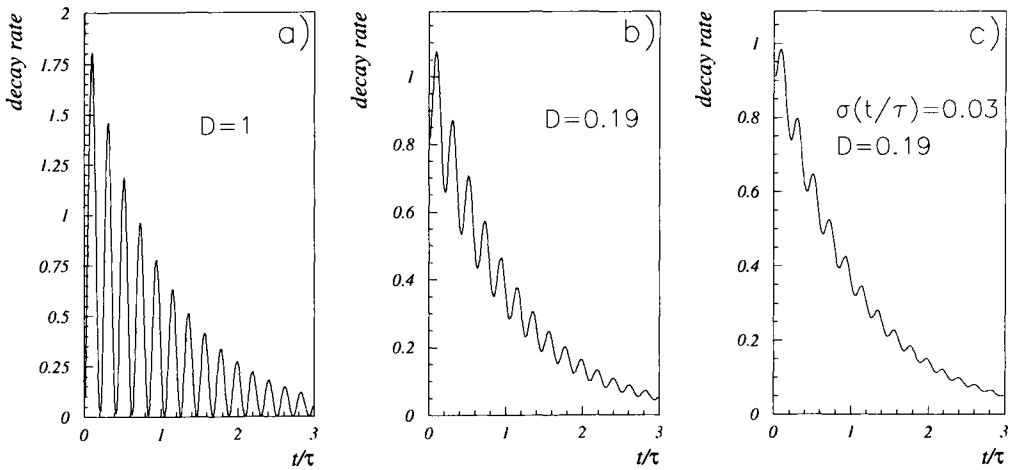
Figure 6. *Example of B_s mixing with $x_s = 30$ ($\Delta m_s = 22ps^{-1}$): (a) perfect resolution and no dilution, (b) with a dilution factor $D = 0.19$, (c) with dilution factor $D = 0.19$ and a finite resolution $\sigma_t = 45fs$*

The amplitude of the mixing signal is further reduced by the finite resolution in proper time t. In the limit of very good resolution ($\sigma_t \ll x_s\tau$), a simple quantitative estimate of the effect can be made by observing that the amplitude is reduced by events moving from the "up" side of the wave to the "down" side, the effect being most important in an interval of order σ_t around the node of the oscillating term (Figure 7). In this region the function is approximately linear, and the fraction of events in this interval can be roughly estimated to be $\sigma_t^2 x_s^2/2\tau^2$. A full derivation involves a convolution of the time evolution curve and the detector response. The amplitude turns out to be reduced by a factor $\exp(-\sigma_t^2 x_s^2/2\tau^2)$, in agreement with the rough estimate to first order in σ_t^2/τ^2.
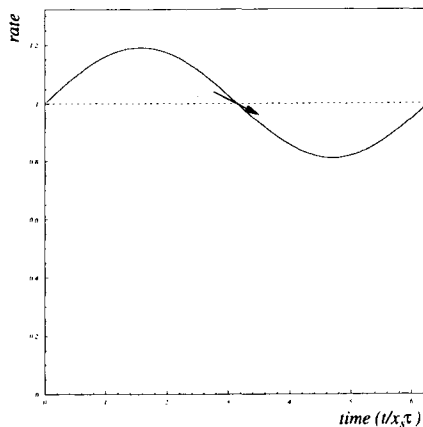
Figure 7. *Dilution of the oscillatory term due to finite resolution.*

The dilution D is replaced by $D' = D \exp(-\sigma_t^2 x_s^2 / 2\tau^2)$, giving the distribution shown in Figure 6(c).

A simple estimate of the statistics required for a significant measurement can be made by fixing x_s, and dividing the events into two classes, those from the "up" part of the wave, and those from the "down" part (Figure 7). The measured oscillation amplitude for a given x_s differs from zero if the two classes are found to be differently populated. The distribution over the two classes is binomial, with probability for the "up" part equal to $p = 0.5 + cD'$, where c is a constant of order 1. The error on p, $\sigma(p) = \sqrt{p(1-p)/N}$, for a measurement with N reconstructed and tagged events. This expression can be simplified for $p \approx 0.5$ to $\sigma(p) \approx 1/(2\sqrt{N})$. The resulting error on the amplitude D' is $\sigma_{D'} \approx 1/(2c\sqrt{N})$, and the significance of the measurement is:

$$D'/\sigma_{D'} \approx 2c\sqrt{N}D \exp\left(-\frac{\sigma_t^2 x_s^2}{2\tau^2}\right).$$

For a given significance, the number of events needed, and thus the duration of the measurement, is proportional to $\exp(\sigma_t^2 x_s^2 / 2\tau^2)$. This is a very steep function of the proper time resolution σ_t and the mixing parameter x_s, particularly in the region where the quadratic term in the exponent is greater than 1. In the upgraded CDF detector, where σ_t=45fs is expected, this value is reached at $x_s = 30$.

4   HERA-B

HERA-B is a fixed target experiment at the storage ring HERA at DESY (Lohse et al. 1994). The experiment was originally designed to measure CP violation in the decays of neutral B mesons to $J/\psi K_S^0$ (Figure 8). Bottom hadrons at HERA-B are produced with a beam of 920GeV protons from the storage ring which hits an internal wire target. The average momentum of the B mesons is 120GeV, so they travel about 1cm before they decay. The J/ψ from the decay $B^0 \to J/\psi K_S^0$ can be triggered and reconstructed in its
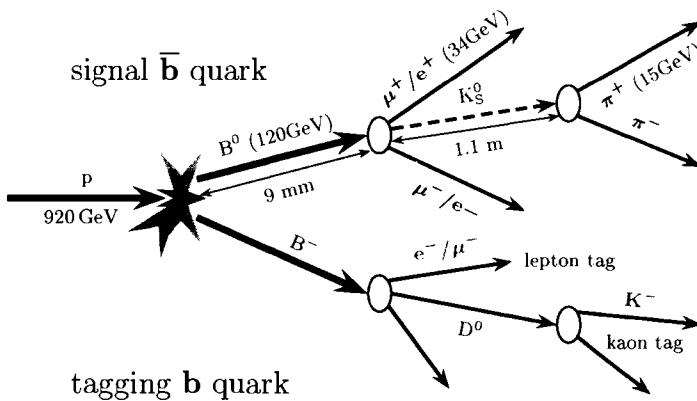
Figure 8. *A $B \to J/\psi K_S^0$ event showing the typical momenta and decay distances expected in the HERA-B experiment.*

dileptonic decay mode. From two existing measurements of $\sigma(b\bar{b})$ in the relevant kinematic region (Jansen et al. 1995, Alexopoulos et al. 1999), the fraction of events containing b quarks is of order 10^{-6}. Since the branching ratio for the decay $B^0 \rightarrow J/\psi K_S^0$ is of order 10^{-3}, and $J/\psi \rightarrow l^+l^-$ is 6%, the ratio of the signal channel to the total inelastic cross section is of order 10^{-11}. The HERA-B design requires a very high interaction rate of 40MHz, with an average of four proton–nucleon interactions occurring for each bunch crossing (every 96ns). This interaction rate leads to fluxes of charged particles through the detector components which are comparable to the fluxes expected at the LHC.
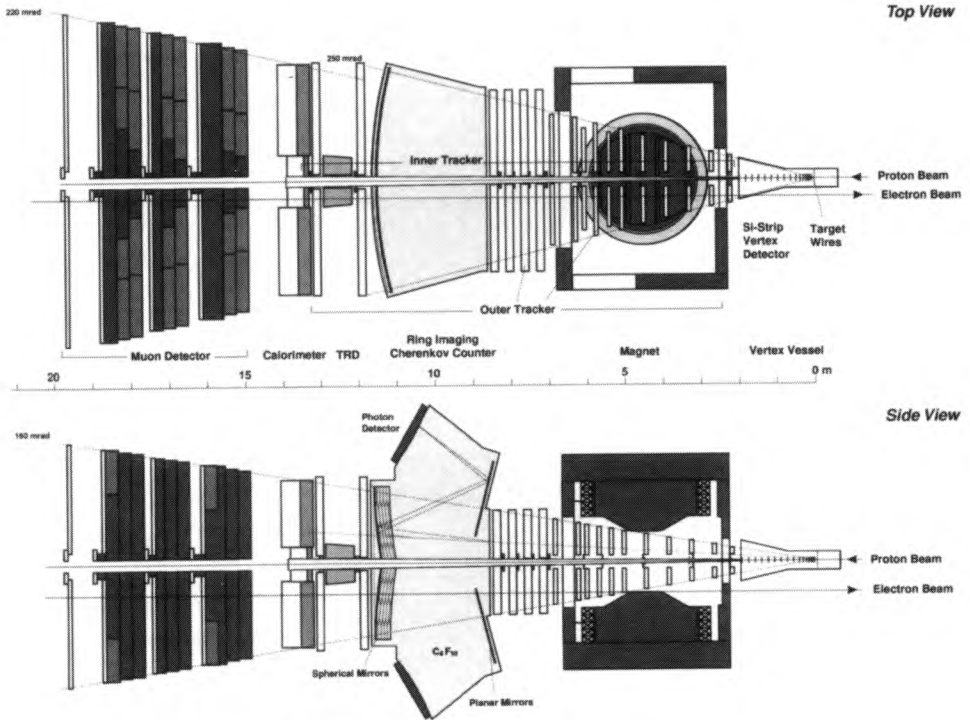
Figure 9. *The HERA-B spectrometer.*

4.1 The HERA-B spectrometer

HERA-B is a large magnetic forward spectrometer (Figure 9) with outer dimension of 20m×9m×7m. It covers a polar angle between 10mrad and 200mrad with respect to the beam direction corresponding to a 90% coverage in the center of mass system. The main detector components are wire targets, a vertex detector system, a tracking system, a ring imaging Čerenkov counter, an electromagnetic calorimeter, muon chambers and a trigger system.

An important requirement for HERA-B was to not interfere with HERA operation for *ep* collider experiments. The solution was to put several target wires in the halo of the proton beam (Ehret 2000). Protons which leave the core of the beam, and are therefore
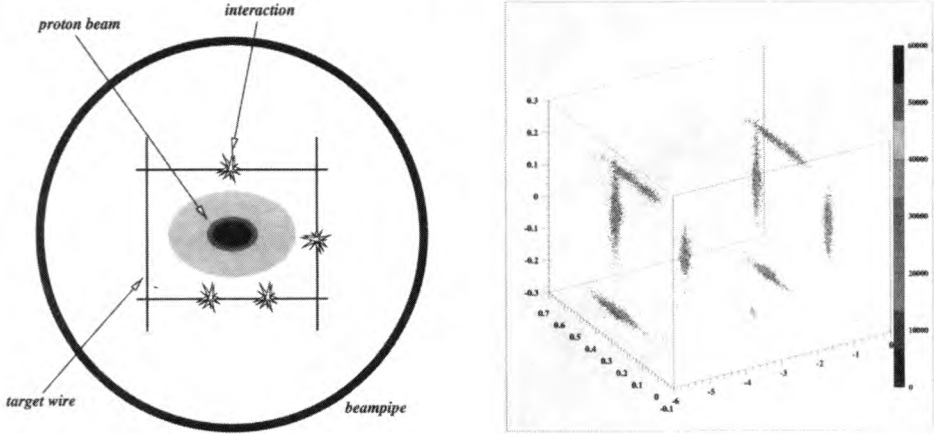
Figure 10. *HERA-B wire target concept (left), reconstructed vertex positions (right).*

lost for collisions, interact in 8 wires which are arranged in 2 stations (Figure 10). This arrangement allows for an efficient separation of primary vertices for events with multiple interactions. The interaction rate can be changed by moving the wire targets within the beam halo, and the number of wires in operation can be varied by retracting some of them. Operation of the wire target is very stable, and meets the criteria for small interference with collider operation.
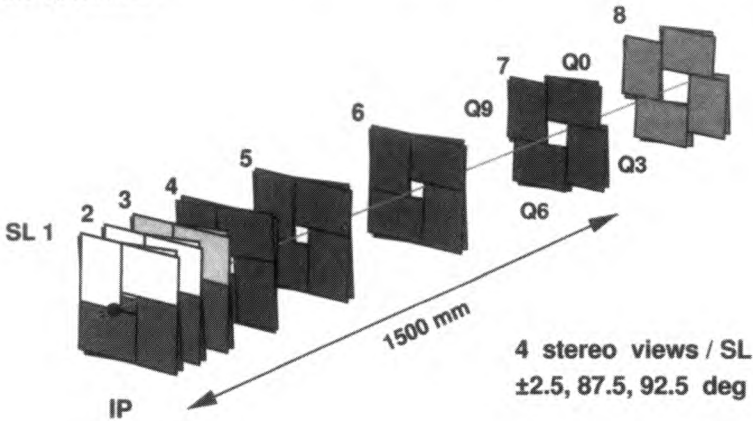
Figure 11. *The HERA-B vertex detector system.*

The vertex detector system (VDS) has to provide track co-ordinates for reconstructing decay vertices and the impact parameters of decay products. The VDS (Masciocchi 2001) consists of 64 double–sided silicon microstrip detectors assembled in eight super–layers of four quadrants around the proton beam axis (Figure 11). To minimize multiple scattering, the first seven super–layers are placed in a Roman pot configuration. The VDS has successfully operated for several years with interaction rates of up to 40MHz. It has a hit efficiency better than 97%, and a hit resolution of 10 μm.
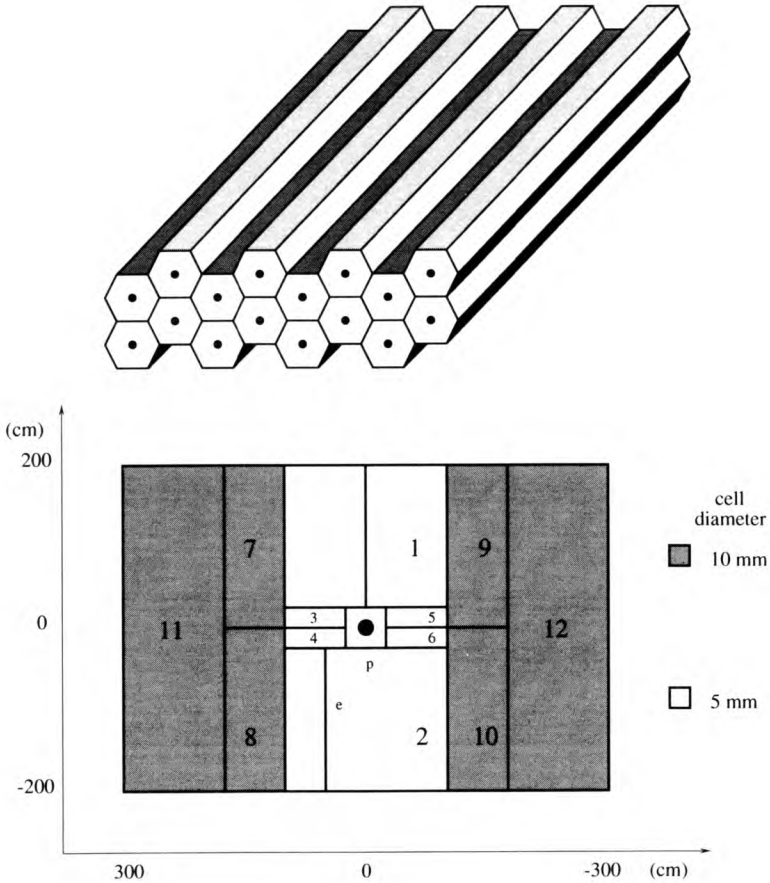
Figure 12. *The HERA-B Outer Tracking system: honeycomb drift chambers (top), modular structure (bottom).*

The tracking system extends over a total length of 11m. The first group of super-layers are situated inside the field of a dipole magnet, while the others are in the field–free region. The particle flux increases towards small polar angles, so the system is divided into inner and outer parts consisting of detector elements with different granularity. The Outer Tracker (OTR) measures charged tracks over most of the geometrical acceptance. It consists of honeycomb drift chambers of two different granularities (Figure 12), covering a total area of about 1000m^2. The system has to operate in a high radiation environment, which led to some serious aging problems (Hohlmann 2001). During the year 2000 run, the overall hit efficiency was also somewhat worse than expected, due to HV failures of individual wires. We note that the HERA-B outer tracking system is the first large area detector working in a LHC–like environment.

The inner tracker (ITR) measures tracks at small distances from the beam, and provides input for the first level trigger (Gradl 2001). The ITR elements cover the region from 6cm to 25cm around the beam pipe, where the particle flux can reach 10^6 per square centimeter per second. The detectors are microstrip gas chambers with a gas electron
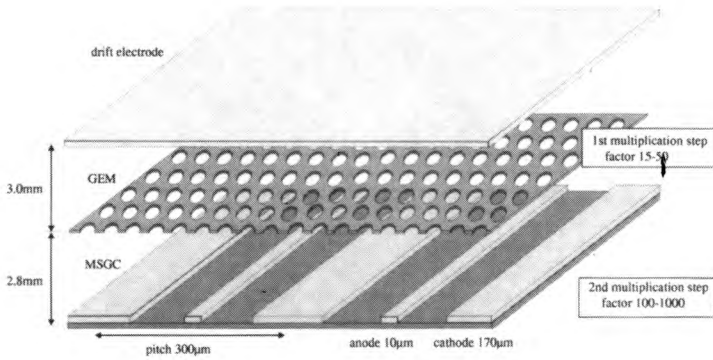
Figure 13. *MSGC-GEM detector structure.*

multiplier stage (GEM), as shown in Figure 13. After an extensive conditioning phase, the operation of the ITR was stable in the year 2000 run, with performance close to the design value. Performances of the complete tracking system for K_S^0 and Λ candidates are shown in Figure 14.
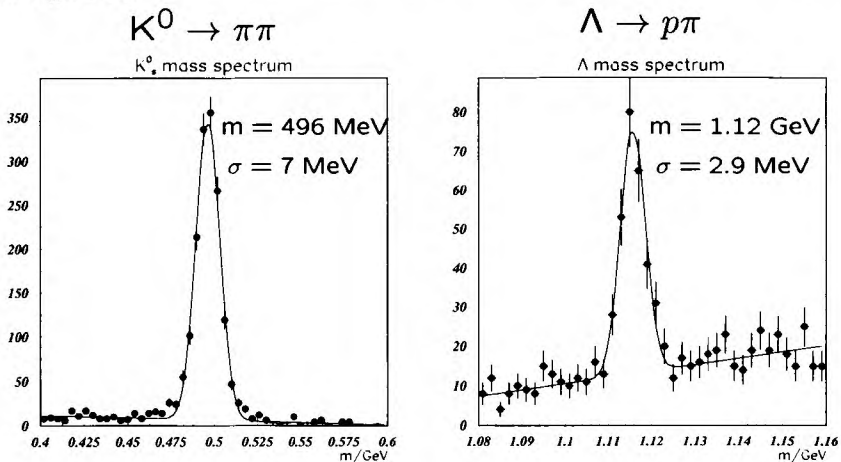
Figure 14. *Performance of the HERA-B tracking system: K_S^0 and Λ signals*

A ring imaging Čerenkov counter (RICH) identifies charged hadrons from the photons emitted in a gaseous C_4F_{10} radiator (Ariño et al. 2000). The photons are reflected by two mirrors and collected as rings on multianode photomultipliers (Broemmelsiek 1998), using a lens system (Figure 15). To match the variation in occupancy and reduce the cost, 4 and 16 channel versions of the Hamamatsu R5900 multianode PMT are used. The system has typical rates of 1MHz per channel in the hottest parts. It has been in operation since summer 1998, with very few noisy or dead channels (Figure 16).

For ultrarelativistic particles, the average number of photons in a ring is 32, and the single photon resolution is $\sigma_\theta^{RICH} = 0.7$mrad for the finer, and 1mrad for the coarser granularities, in excellent agreement with the design values. Figure 17 shows the probability that
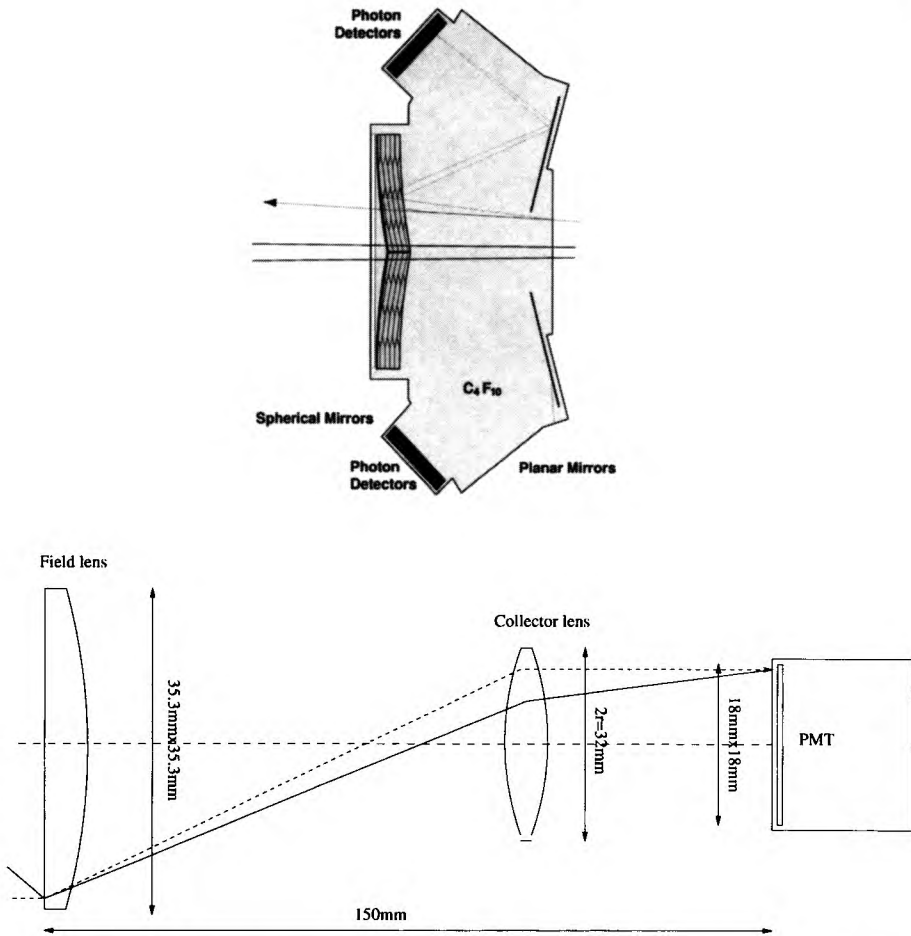
Figure 15. *HERA-B RICH: radiator vessel (left), light collection system (right)*
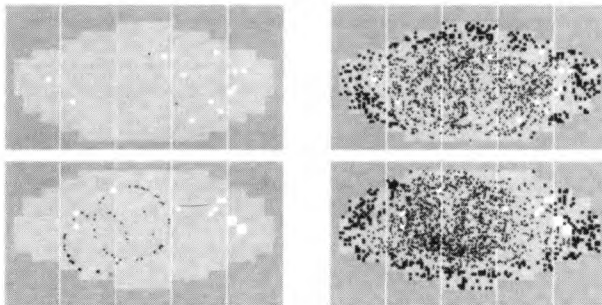
Figure 16. *HERA-B RICH: a low occupancy background event with two rings on the photon detector (left), a typical event (right).*
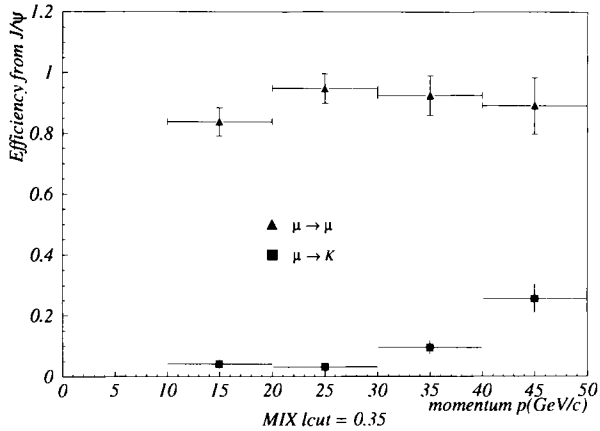
Figure 17. *HERA-B RICH: particle identification efficiency from $J/\psi \to \mu\mu$.*

a muon from a J/ψ decay is recognized as a muon by the RICH, as a function of muon momentum. The performance of RICH is expected to improve with better alignment of the tracking system.

The electromagnetic calorimeter (Avoni et al. 2001) measures photon energies, identifies electrons and provides an electron pretrigger. The calorimeter is of the "shashlik"-type with three different granularities in order to match occupancies. The measured energy resolution, $\sigma(E)/E = 1.7\% \oplus 22.5\%/\sqrt{E}$, is close to the design value.

The muon detector (Eiges et al. 2001) separate muons from hadrons and gives a muon pretrigger. It consists of four layers of detector elements embedded in iron absorbers. Three types of detector elements (tube, pad and pixel chambers) are used in order to match the required granularity. In the 2000 run, the muon pretrigger provided a good J/ψ signal using a fourfold coincidence of signals in the pad chambers. The performance was limited by inefficiencies in certain parts of the detector.

The huge background from inelastic proton-nucleon scattering requires a highly selective and efficient trigger. The main idea of the HERA-B trigger (Gerndt 2000, Kreuzer 2001) is to search for a pair of high-p_t leptons, with invariant mass consistent with a J/ψ. At the pretrigger level, data from the calorimeter, muon and tracking systems are used to define regions of interest which correspond to particles with high transverse momenta. Pattern recognition and track reconstruction are performed with dedicated processors at the tracking system stations (Figure 18), starting from the track candidates defined by the pretrigger system. A cut is made on the invariant mass of candidate track pairs The first trigger level aims to reduce the rate by a factor 200. Second and third level triggers refine track candidates and search for vertices. They are designed to give suppression factors of 100 and 10 respectively. A fourth level performs the complete event reconstruction and writes data to storage. The best results in the 2000 run were achieved with a first level trigger requirement of a single high-pt lepton, and a search for the second high-pt lepton in the second level trigger. We note that in this period the first level trigger efficiency was 20% for single muons and 54% for electrons, which is well below the design value of 90%. This is mainly due to inefficiencies in the tracking and muon chambers.
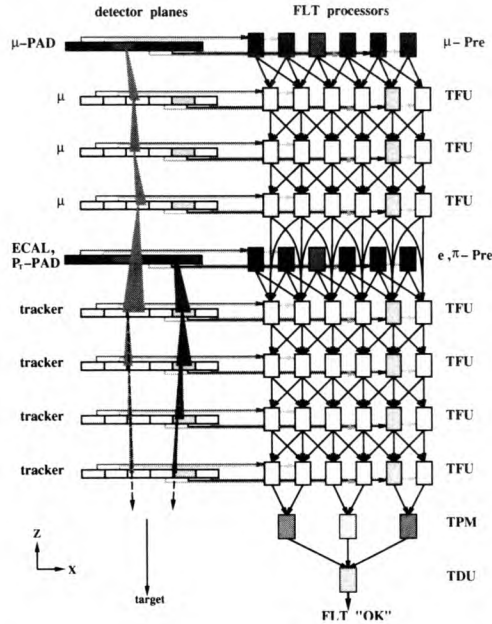
Figure 18. *HERA-B first level trigger system: pretrigger candidates are followed through track finding units (TFU), combined into pairs in track parameter units (TPM); before a final decision by the trigger decision units (TDU).*

4.2   HERA-B status and plans

The experiment was approved in 1995, but the design and construction of detector elements that can cope with the extremely high fluxes proved to be a technological challenge. Problems with the tracking system caused a substantial delay in the schedule, and installation of the full detector was only completed in February 2000. The first phase of commissioning lasted 6 months (HERA-B Collaboration 2000). The performance of the detector for $J\psi$ events is shown in Figure 19.

After the end of the current HERA shutdown, a continuation of the operation of HERA-B is expected in the second half of 2001. Due to the late start and the trigger inefficiencies, HERA-B is no longer competitive for a measurement of CP violation. In the 2001/2002 running period the experiment will concentrate on the measurement of events which contain a J/ψ meson, for which the detector is optimized. Based on the results of the 2000 run, and detector improvements, it is expected that 1500 reconstructed J/ψ events will be collected per hour, resulting in 2.1 million events in the full running period. With these data HERA-B can measure the nuclear dependence of charmonium production, and the $b\bar{b}$ production cross-section.

A suppression of the yield per nucleon of J/ψ in proton-nucleus, pion-nucleus, and heavy ion collisions is observed in heavy nuclei compared to light nuclei (Leitch et al. 2000, Alde et al. 1991, Leitch et al. 1995). The effect depends strongly on the x_F and p_T of the J/ψ. Various mechanisms can contribute to this suppression, but their relative importance is not clear. HERA-B can measure the production of J/ψ mesons down
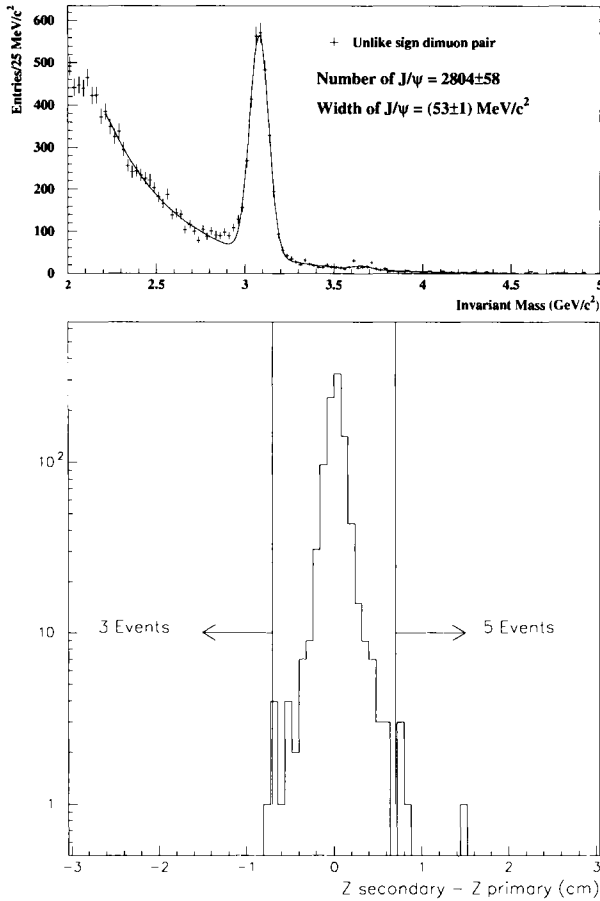
Figure 19. *HERA-B performance: di–muon invariant mass spectrum (top), decay length distributions in J/ψ mass region (bottom).*

to $x_F = -0.4$, as well as ψ' and χ_c, using the electromagnetic calorimeter to reconstruct $\chi_c \to J/\psi\gamma$. Running simultaneously with target wires made of different materials reduces the systematic errors. Last year HERA-B took data using Ti and C wires.

A measurement of of the production cross-section $\sigma_{b\bar{b}}$ is needed, because there are two inconsistent results (Jansen et al. 1995, Alexopoulos et al. 1999). For the related processes, $\pi A \to b\bar{b}X$, the theoretical predictions underestimate the cross section. The measurement of $\sigma_{b\bar{b}}$ at HERA-B will be based on inclusive $b \to J/\psi X$ compared to direct J/ψ production. The J/ψ candidates with decay vertices detached from the target wire will be attributed to decays of b hadrons. It is expected that 340 events of this type will survive all cuts with a small background. The result for $\sigma_{b\bar{b}}$ is expected to have a statistical accuracy of 5–10%.

HERA-B is the first experiment working in a LHC-like environment, As such it has accumulated valuable experience during construction and commissioning which is vital for the success of the next generation of hadron experiments. After further improvements during the 2001 shutdown, final commissioning of the detector will be completed.

5 BTeV - a second generation B factory

The motivation for the BTeV experiment is the expectation that in the next few years the e^+e^- B-factories will firmly establish the evidence for CP violation in B decays, but that several questions will remain open due to lack of statistics, and the restricted set of b hadron decays that are available at the e^+e^- machines. The goal of BTeV is to perform high precision measurements of the CP parameters γ, α, χ, in the decays of B^0, $B^\pm$ and B_s mesons, and to search for new physics in rare and flavour changing neutral current decays (Stone 2001).
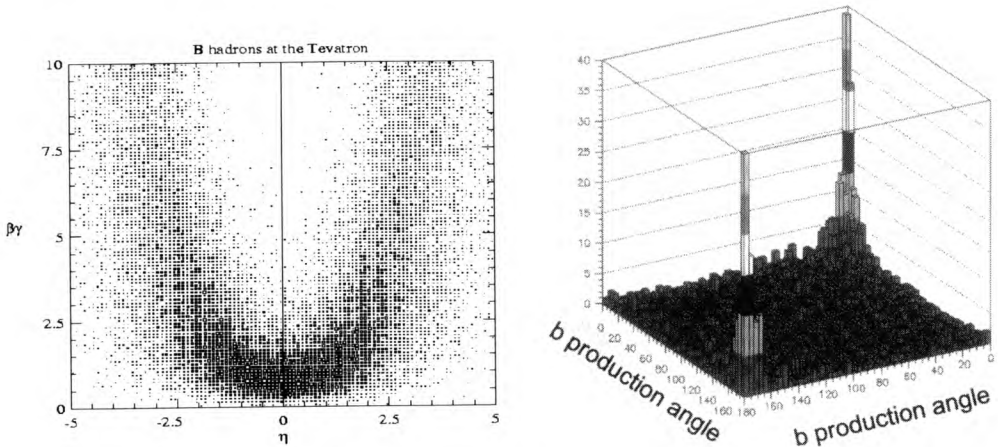
Figure 20. *Production of $b\bar{b}$ pairs at BTeV: dependence of boost on the pseudorapidity (left), correlation of production angles of the $b\bar{b}$ pair (right).*

The BTeV experiment (Santoro 1997, Skwarnicki 2001) will exploit the very large production cross section for $b\bar{b}$ pairs in the forward and backward direction in $p\bar{p}$ collision at the Tevatron (Stone 1997). The boost factor $\beta\gamma$ becomes very large in this region, compared to the central region covered by CDF and D0. Figure 20 shows that b quark pairs are either both in the forward or both in the backward direction, so a single arm spectrometer similar to HERA-B would in principle be sufficient for flavour-tagged CP measurements.

5.1 BTeV spectrometer

To accomplish its goals, the BTeV spectrometer (Sheldon 2001), shown in Figure 21, has to meet the following requirements. It has to efficiently trigger on hadronic B decays, and it needs a high resolution vertex detector to measure B_s oscillations. An excellent particle identification system is required for π/K separation, and a high resolution electromagnetic calorimeter for π^0 reconstruction. All the detector components have to survive high interaction rates. Finally, it needs a high capacity data acquisition system that can handle a large number of read-out channels.

The BTeV tracking system has three major parts, a vertex pixel detector, a system
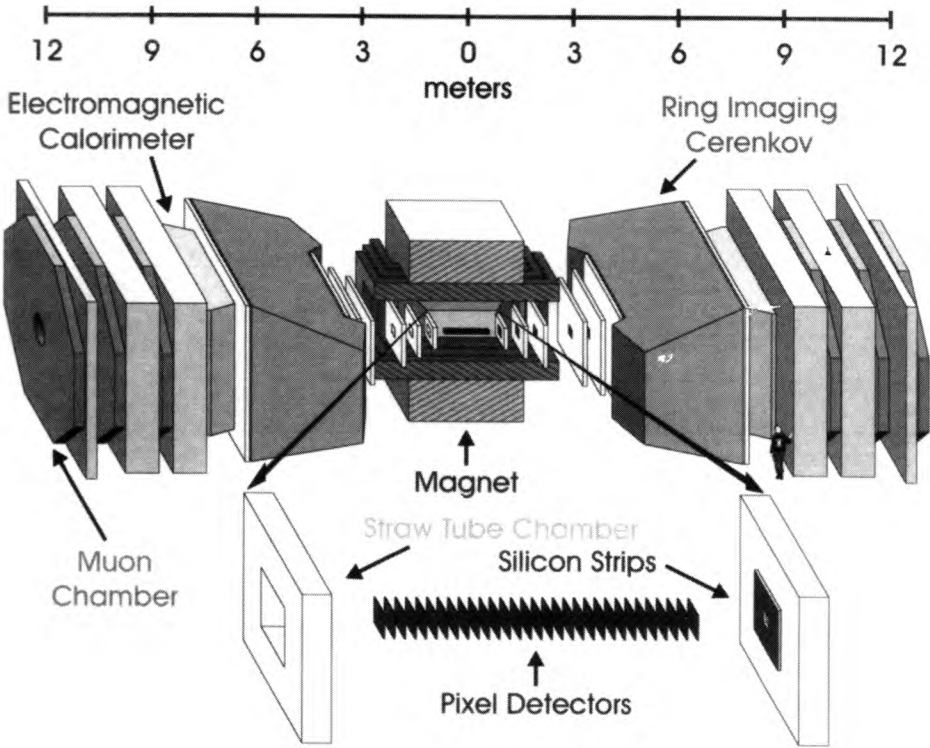
Figure 21. *BTeV spectrometer with an expanded view of the interaction region.*

of straw-tube chambers as the main tracking device, and silicon strip detectors in the vicinity of the beam pipe. The pixel detectors are essential to eliminate ambiguities and provide a detached vertex trigger in the high track density region (Artuso 2001). The detectors are expected to have a low noise. The read-out electronics is subject to the same radiation load as the detector, but it has been possible to develop radiation hard electronics using the deep sub-micron process (Christian 2001).

The identification of hadrons is essential for the tagging of b-flavour in CP violation and mixing measurements. It is also crucial for the identification of hadronic final states such as $B \rightarrow \pi^+\pi^-$. The importance of the latter is illustrated in Figure 22, where the $B \rightarrow \pi^+\pi^-$ events are obscured by a large background from other, kinematically similar, two-body decays. The kaon and pion momentum range that the identification device has to cover, can be deduced from Figure 23. To achieve the necessary π/K separation in the range $3 < p < 70 \text{GeV/c}$, a RICH counter will be used (Figure 24), with the same radiator gas, C_4F_{10}, as in HERA-B. To supplement the performance in the low momentum region and to improve p/K separation, plans are being discussed to add an aerogel radiator. A novel type of a position sensitive photon counter will be used (Figure 25, the hybrid photon detector (HPD) (Alemi 2000). A high electric field is used to accelerate photo-electrons onto a silicon pixel detector. The HPD is still under development, and is also
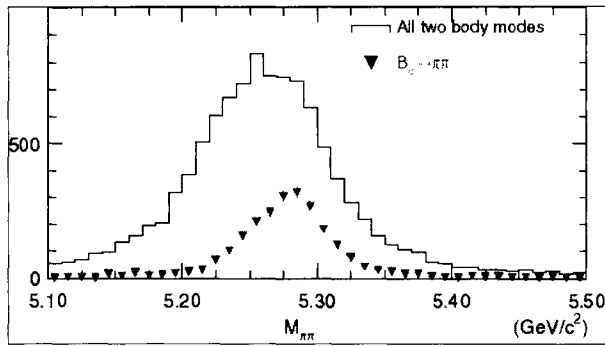
Figure 22. *Invariant mass distributions as determined by MC simulation for the B →*
π⁺π⁻ decay together with the two body background sources, B → Kπ, Bₛ → KK and
Bₛ → Kπ, without particle identification criteria.
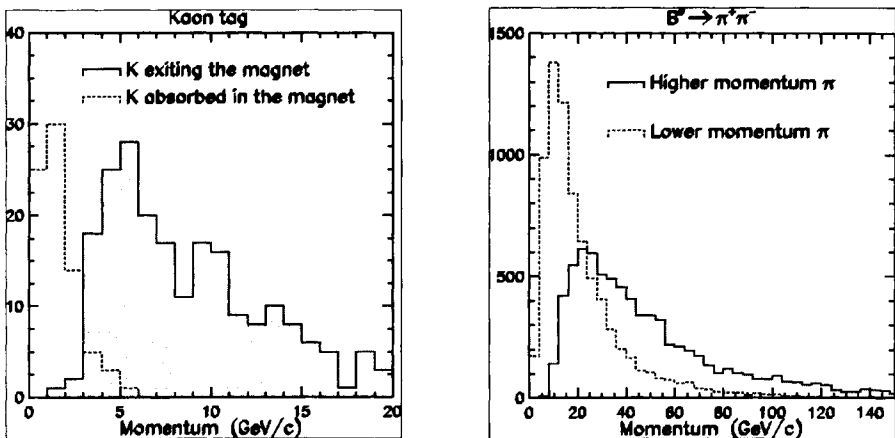
Figure 23. *Momentum spectra of tagging kaons (left) and pions from B → π⁺π⁻ (right).*

foreseen for use in the LHCb experiment (Wotton 2000).

The calorimeter needs to have high resolution and good radiation tolerance. For this BTeV will use 22000 Lead Tungstate crystals (PbWO₄), with size 27×27×2220mm³. This is the same technology as for the CMS experiment. The crystals are radiation hard, fast scintillators with 99% of the light emitted in less than 100ns. They will be read out with photomultiplier tubes. The energy resolution predicted by a full GEANT simulation is $\sigma(E)/E = 0.55\% \oplus 1.6\%/\sqrt{E}(\text{GeV})$, and the position resolution is given by $\sigma_x = 3500\mu\text{m}/E(\text{GeV}) \oplus 200\mu\text{m}$.

The muon detector will be used both to identify muons, and to provide an auxiliary di-muon trigger for checks of the detached vertex trigger, particularly in the commissioning phase of the experiment. It consists of two toroids with three detector sets.
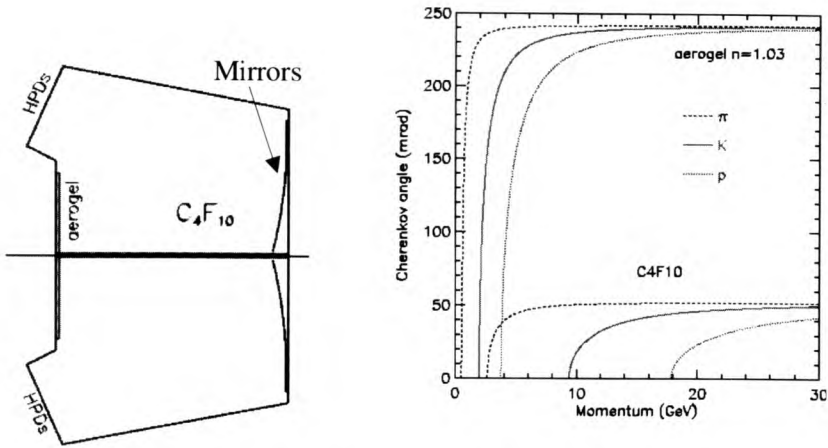
Figure 24. *BTEV RICH: detector layout with aerogel and C_4F_{10} radiators, spherical mirrors and HPD photon detectors (left), momentum dependence of the Čerenkov angle in the two radiators (right).*
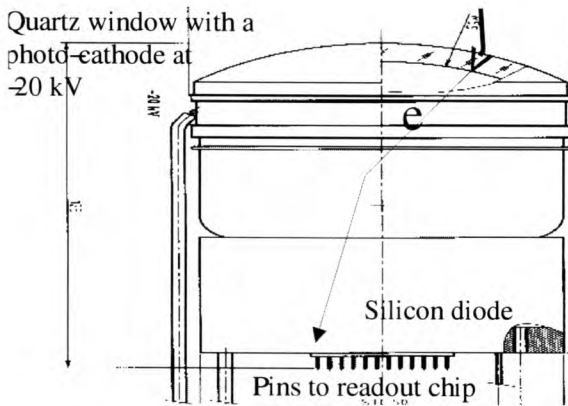
Figure 25. *The HPD design for the BTeV RICH.*

One of the main components of the BTeV spectrometer is the trigger system (Artuso 2001). The first level trigger is meant to find segments in the entrance and exit of the pixel detector as shown in Figure 26. It will then match segments, get crude momentum estimates, and reject low momentum tracks from further use for vertex determination. In the next step, the primary vertex will be found, and then for each of the tracks which miss the primary vertex, the significance of the detachment parameter, $b/\sigma(b)$, will be calculated (Figure 27).

The goal of the first level trigger is a factor of 100 rejection of bunch crossings which do not produce a $b\bar{b}$ pair, and an efficiency exceeding 50% for interesting events. As an example, a requirement of two tracks with $b/\sigma(b) > 6$ leaves 1% of beam crossings, and has 63% efficiency for $B^0 \rightarrow \pi^+\pi^-$, 70% for $B^- \rightarrow D^0 K^-$, and 50% for $B^0 \rightarrow J/\psi K_S^0$. The dependence of the signal efficiency and the background rejection power on the detachment parameter cut for different numbers of tracks is shown in Figure 28.
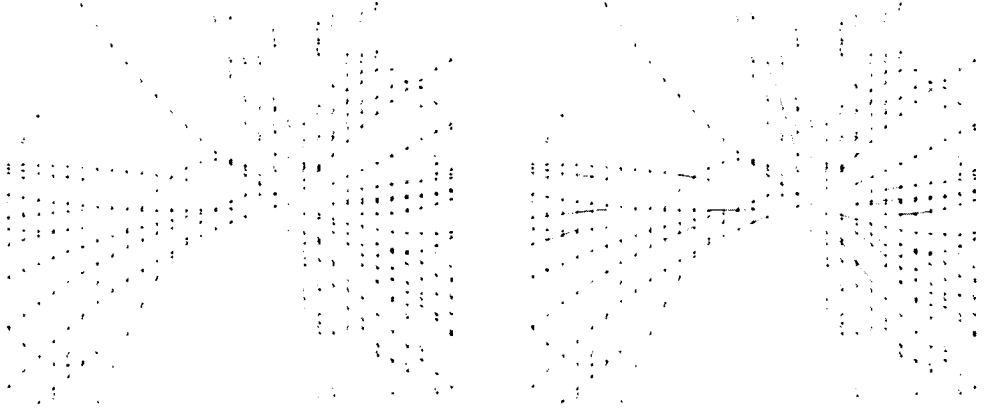
Figure 26. *Event in the pixel detector, with the horizontal coordinate compressed by factor 10. Left: raw event, right: segments found at the entrance and exit of the detector.*
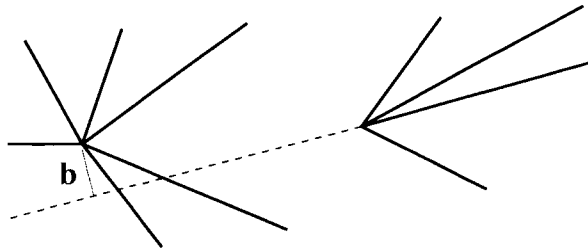
Figure 27. *Detachment of a track from the primary vertex.*
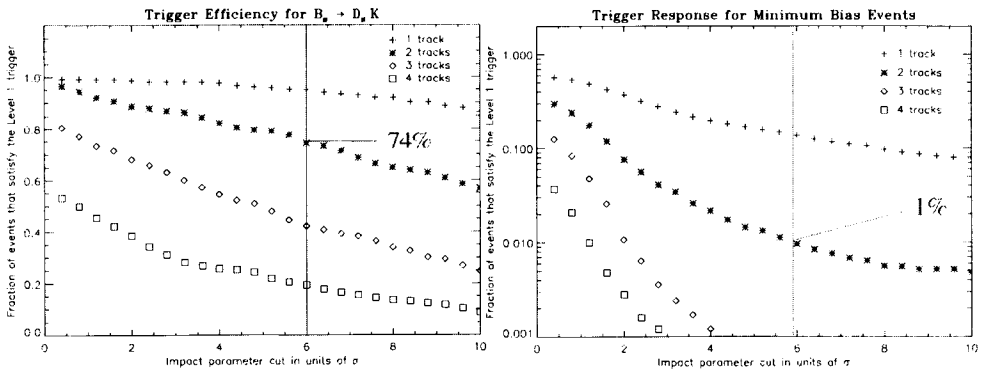
Figure 28. *Level 1 trigger efficiency for $B_s \to D_s K$ (left) compared to the efficiency for minimum bias events (right).*

The second trigger level will rely on re-fitting the pixel segments, including more hits, and linking these tracks to the downstream tracking systems. This will give an accurate momentum measurement, and a better determination of the detachment parameter of the track. The aim is to achieve a further rejection of a factor of 10 or more at 90% efficiency. Finally a third level will involve more processing, with the goal of reducing the event output rate of 4kHz, with an event size of 40kbyte.

5.2   BTeV physics reach

The comparisons shown in Table 2, are meant to illustrate in which respects the BTeV experiment is a second generation B-factory. In the first column we list the expected numbers of events collected by the e^+e^- B-factories with the integrated luminosities of 500fb^{-1} which are expected by 2005. In the second column we present the amount of data the BTeV experiment is expected to reconstruct in one effective year (10^7s). The interested reader is referred to a review article (Stone 1997) for further discussions, as well as a comparison between the BTeV and LHCb experiments.

Mode	e^+e^- B-factories with 500fb^{-1}	BTeV yield in 10^7 s
$B_d \to \pi^+\pi^-$	850	24000
$B_d \to \phi K^-$	700	11000
$B_d \to \phi K_S$	250 (75 tagged)	2000 (200 tagged)
$B_d \to K^*\mu\mu$	50	2200
$B_s \to J/\psi\eta^{(')}$	-	15000

Table 2. *Comparison of expected BTeV yield of reconstructed events in some interesting decay modes with the expectations from the e^+e^- B-factories by the year 2005.*

6   Summary

We summarize this review by observing that although the hadron production of B mesons has a large cross section, the background levels are high, so elaborate triggering strategies are needed to extract the signal events. In addition an efficient operation of the detector in a high rate environment is required over extended run periods. The results of B physics studies by the CDF experiment prove that B physics can be done at hadron machines. The HERA-B spectrometer components are ready after extensive R& D to develop radiation hardness and high rate tolerance. The trigger of the HERA-B spectrometer is being commissioned, and a year of data taking is expected in 2002. This will provide valuable experience for the next generation of experiments at hadron machines. Finally, the second generation of B factory experiments, BTeV and LHCb, are being developed, and are expected to open a new era of B physics after 2005.

Acknowledgments

References

Abe F *et al.* [CDF Collaboration], 1988, *Nucl. Instrum. Meth.* **271** 387.
Affolder T *et al.* [CDF Collaboration], 1999, *Phys. Rev. Lett.* **83** 3378.
Affolder T *et al.* [CDF Collaboration], 2000, *Phys. Rev.* **D61** 072005.
Affolder T *et al.* [CDF Collaboration], 2000a, *Phys. Rev. Lett.* **85** 4668.
Alde D M *et al.*, *Phys. Rev. Lett.* **66** 133.
Alemi M *et al.* [TA2 Collaboration], 2000, *Nucl. Instrum. Meth. A* **449** 48.
Alexopoulos T *et al.*, 1999, *Phys. Rev. Lett.* **82** 41.
Ariño I *et al.*, 2001, *Nucl. Instrum. Meth. A* **453** 289.
Artuso M, [BTeV Collaboration], 2001, *Nucl. Instrum. Meth. A* **462** 249.
Avoni G *et al.*, 2001, *Nucl. Instrum. Meth. A* **461** 332.
Bedeschi F, 2001, "Heavy flavor production in hadronic collisions," *Fermilab-Conf-01-006-E*
Christian D C *et al.*, 2001, "FPIX2: A radiation-hard pixel readout chip for BTeV," *Nucl. Instrum. Meth. A* **473** 152.
Ehret K, 2000, *Nucl. Instrum. Meth. A* **446** 198.
Eiges V *et al.*, 2001, *Nucl. Instrum. Meth. A* **461** 104.
Gay C, 2000, "B mixing," *Ann. Rev. Nucl. Part. Sci.* **50** 577, *hep-ex*/0103016.
Gradl W, 2001, *Nucl. Instrum. Meth. A* **461** 80 2001.
Hall B K *et al.*, 2001, *FERMILAB-CONF-01-335.*
HERA-B Collaboration, 2000, *DESY-PRC* 00/04.
Hohlmann M, 2001, *Nucl. Instrum. Meth. A* **461** 21.
Jansen D M *et al.*, 1995, *Phys. Rev. Lett.* **74** 3118.
Kreuzer P, 2001, *Nucl. Instrum. Meth. A* **462** 212.
Leitch M J *et al.*, *Phys. Rev. D* **84** 3526.
Leitch M J *et al.* , *Phys. Rev. D* **52** 4251.
Lohse T *et al.*, 1994, *DESY-PRC* 94/02.
Masciocchi S, 2001, *Nucl. Instrum. Meth. A* **461** 220.
Moser H G and Roussarie A, 1997, "Mathematical methods for B0 anti-B0 oscillation analyses," *Nucl. Instrum. Meth. A* **384** 491.
Paulini M, [CDF Collaboration], 1999, "B lifetimes, mixing and CP violation at CDF," *Int. J. Mod. Phys. A* **14** 2791, *hep-ex*/9903002.
Pitts K T [CDF Collaboration], 2000, "Electroweak and B physics results from the Fermilab Tevatron collider," *hep-ex*/0102010.
Santoro A *et al.*, 1997, "BTeV: An Expression of interest for a heavy quark program at C0," *FERMILAB-FN-660.*
Sheldon P D, [BTeV Collaboration], 2001, *Nucl. Instrum. Meth. A* **462** 291.
Skwarnicki T, [BTeV Collaboration], 2001, *Nucl. Instrum. Meth. A* **462** 227.
Stone S, 2001, these proceedings.

Stone S, 1997, "The goals and techniques of BTeV and LHC-B," *hep-ph*/9709500.

Stone S, 2001, these proceedings.

Tanaka M, [BTeV Collaboration], 2001, *Nucl. Instrum. Meth. A* **462** 165.

Wotton S A, [LHCb RICH Group Collaboration], 2000, "The LHCb RICH detectors," *Nucl. Instrum. Meth. A* **453** 296.

Observation of direct CP Violation in kaon decays

Konrad Kleinknecht

Johannes Gutenberg-Universität, Mainz, Germany

1 Introduction

A small matter-antimatter asymmetry of the weak force has been experimentally established. This CP violation may be related to the small excess of matter from the big bang. The nature of CP violation in the K^0 system has been clarified after 37 years of experimentation as being due to a non-trivial phase in the weak quark mixing matrix which generates "direct CP violation" in the weak Hamiltonian. The experiments demonstrating this direct CP violation are discussed.

2 The big bang and the expanding universe

In our present universe, we observe with optical telescopes about 100 billion (10^{11}) galaxies each of them with 100 billion stars. These stars for most of their life burn hydrogen to Helium, through nuclear fusion. They emit visible light from their surface and neutrinos from their interior. Both of these radiations have been observed from the sun.

The analysis of the visible wavelength spectrum reveals black absorption lines due to the absorption of light in the atoms of the outer layers of the sun or other stars. A large step in the understanding of the evolution of the universe was made when Edwin Hubble observed in 1929 that these absorption lines for a given element are shifted towards the red end of the wavelength scale for stars which are very far outside our galaxy, say 100 million light years away. This "redshift" is interpreted as being due to these stars receding from
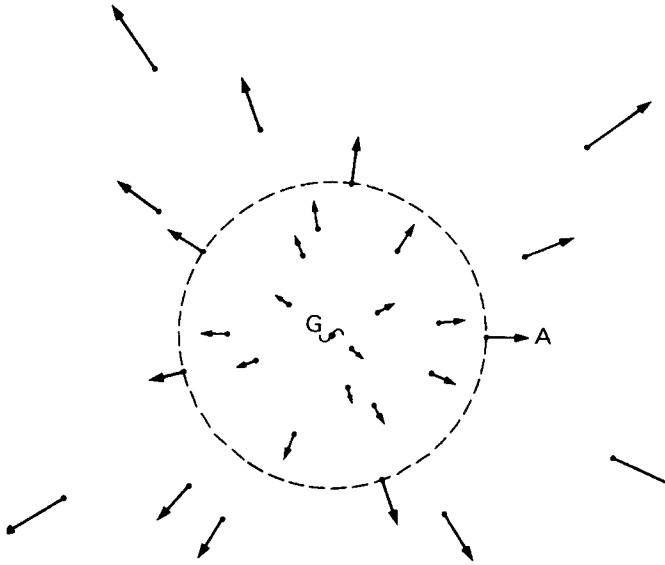
Figure 1. *Galaxies receding from us with velocities indicated by the length of the arrows*

us. Hubble observed that the size of the redshift, and the velocity of these receding stars is proportional to their distance from us. As an example, there is a galaxy in the Ursa Major which moves away with 1/7 of the light velocity, i.e. 42000km/sec. Figure 1 visualizes the flight velocities of such galaxies, with arrows indicating the size of the recession velocities.

Of course, our point of reference, the earth, is by no means a special one. According to relativity theory, each reference system is equivalent in our finite, but unlimited universe. Each galaxy recedes from every other with a velocity proportional to their distance, like two points on a expanding rubber balloon. Turning the time arrow backwards, one concludes that these objects must have been very close to each other, about 15 billion years ago. At that time, an incredibly hot fireball of matter and radiation started to expand in the "big bang".

During the big bang, at high temperatures above 10^{30} degrees Kelvin, matter does not consist only of the stable building blocks which we observe in our old and cold universe, the protons and neutrons of the nucleus, and the electrons and neutrinos. The fundamental constituents of protons and neutrons, the up (u) and down (d) quarks, play their role, but there are also two more massive families of fundamental particles which have been produced and discovered in modern accelerator laboratories over the last few decades. Thus we have altogether three families:

    1:     up (u) and down (d) quarks, electrons (e) and their neutrinos (ν)

    2:     strange (s) and charm (c) quarks, muons (μ) and their neutrinos (ν_μ)

    3:     bottom (b) and top (t) quarks, tau leptons (τ) and their neutrinos (ν_τ)

In addition to these matter particles, it was possible to create their antiparticles at
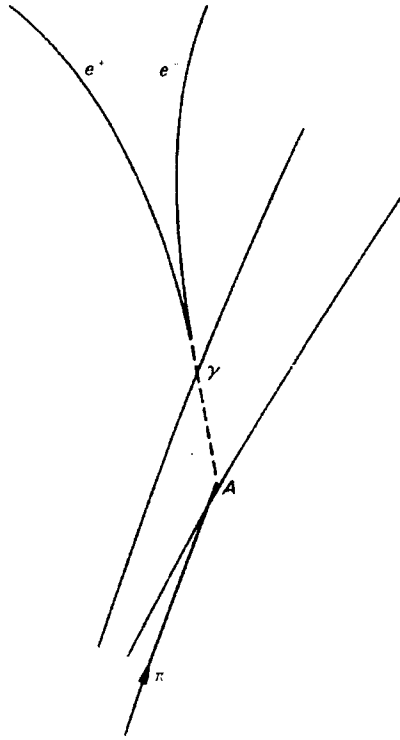
Figure 2. *Production of an electron-positron pair in a hydrogen bubble chamber*

the time of the big bang . In the universe today there is no indication of antimatter left over from the big bang. If antimatter exists, it must show up through the γ radiation which is produced when antimatter annihilates with matter. In the big bang, matter and antimatter were made at extremely high temperatures and in equal quantities because the forces which are responsible for their production are completely symmetric with respect to matter and antimatter. As an example Figure 2 shows the production of an electron and a positron from a photon hitting liquid hydrogen in a bubble chamber. The curvature of the two trajectories is due to a magnetic field which allows to distinguish between positive and negative electric charges. In a similar way, the other elementary constituents of matter were produced in pairs during the initial phases of the big bang. In this hot fireball, creation and annihilation of particles and antiparticles led to an equilibrium of approximately equal numbers of particles, antiparticles and photons. The temperature (measured in degrees Kelvin) corresponds to the average energy of the particles (measured in electron-volts, eV). The expanding fireball cools down, and below a certain temperature the creation of particle-antiparticle pairs stops while annihilation goes on. If all forces were symmetric with respect to matter and antimatter, there would be no matter left over in the cold phase except for photons. As a result of the expansion of the universe these photons would be shifted to lower energies. This cosmic microwave background was discovered in 1965 by Penzias and Wilson. The wave-length spectrum perfectly matches the Planck radiation of a black body at a temperature of 2.73 degrees Kelvin (Figure 3).
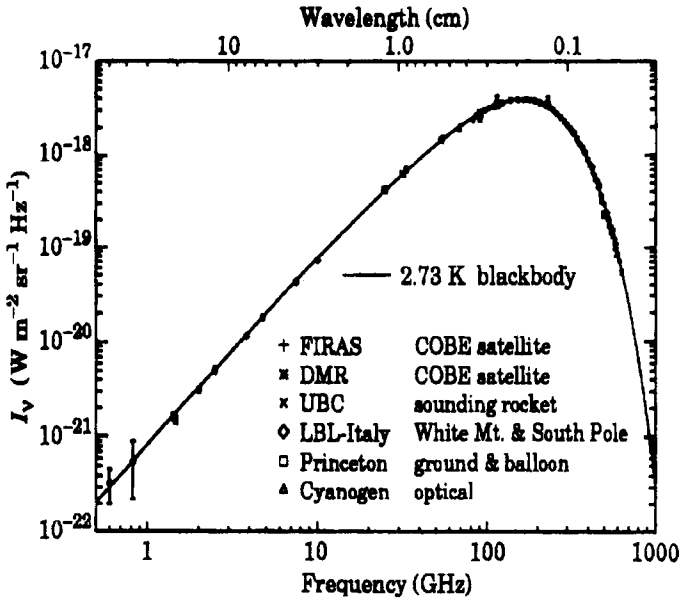
Figure 3. *Frequency distribution of the cosmic microwave background variation, as measured by the COBE satellite*

This is the echo of the hot photons from the big bang. Today the density of this radiation is 500000 photons per litre.

Not only is there no primordial antimatter in our universe. The ratio of nucleons over photons is now only about 10^{-10}, whereas it was of order one in the early phases of the universe. A possible explanation was given by Sacharov and Kuzmin. They stated that this small surplus of matter is only possible if

- one force violates matter-antimatter symmetry,

- baryon number is violated and

- the expansion goes through phases when there is no thermodynamic equilibrium.

3 Symmetries

Symmetries and conservation laws have long played an important role in physics. The simplest examples of macroscopic relevance are the conservation of energy and momentum, which are due to the invariance of forces under translation in time and space, respectively. This was demonstrated by Emmy Noether.

Sacharov postulated that one of the known forces or a new force violated the symmetry between matter and antimatter, thus producing a small surplus of matter. All the remaining matter annihilated with the corresponding amount of antimatter, and at

the end we are left with a surplus of 10^{-10} of matter compared to the photons. Such a symmetry violation goes against principles which were cherished for centuries.

In the domain of quantum phenomena, there are conservation laws corresponding to discrete transformations. One of these is reflection in space by the parity operator P. Invariance of the laws of nature under P means that the mirror image of an experiment yields the same result in its reflected frame of reference as the original experiment in the original frame of reference. This means that "left" and "right" cannot be defined in an absolute sense.

Similarly, the particle-antiparticle conjugation operator C, transforms each particle into its antiparticle, whereby all additive quantum numbers change their sign. C invariance means that experiments in a world consisting of antiparticles will give identical results to experiments in our world. Here again it will be not possible to define in an absolute way whether an object consists of antimatter or matter. An anti-atom composed of antinucleons and positrons emits the same spectral lines as the corresponding atom.

A third transformation of this kind is the time reversal operator T, which reverses momenta and angular momenta. According to the CPT theorem (Lüders 1954, Pauli 1955), there is a connection between these three transformations such that under rather weak assumptions in a local field theory all processes are invariant under the combined operation CPT.

For a long time it was assumed that all elementary processes are also invariant under the application of each of the three operations C, P, and T separately. This assumption was questioned by (Lee & Yang 1956), and the subsequent experiments demonstrated the violation of P and C invariance in weak decays of nuclei and of pions and muons. This violation can be visualized by the longitudinal polarization of neutrinos emerging from a weak vertex: they are left-handed when they are particles and right-handed when antiparticles. Application of P or C to a neutrino leads to an unphysical state (Figure 4).
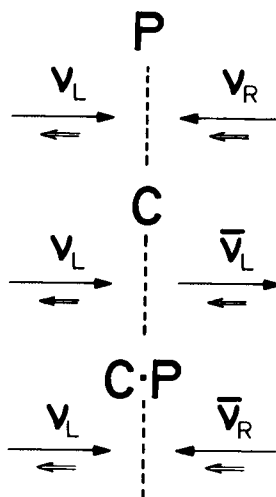
Figure 4. *The mirror image of a left-handed neutrino under P, C, and CP*

The combined operation CP, transforms a left-handed neutrino into a right-handed antineutrino, thus connecting two physical states. CP invariance was therefore considered (Landau 1957) to be a valid replacement for the separate P and C invariance which was no longer true in weak interactions.

A unique testing ground for CP invariance are elementary particles called neutral K mesons. They have the property that the particle K^0 and its antiparticle $\overline{K^0}$ differ in one quantum number, called strangeness, but can mix through a transition mediated by the weak interaction. A consequence of CP invariance for the neutral K mesons was predicted by (Gell-Mann & Pais 1955). It was then known that there was a V^0 or K_1^0 particle that had a short lifetime (10^{-10} sec). It was predicted that there should be a long-lived partner, K_2^0, and that these two physical particles are mixtures of the two strangeness eigenstates, K^0 (S = + 1) and $\overline{K^0}$(S = - 1) that are produced in strong interactions. Weak interactions do not conserve strangeness, but the physical particles should be eigenstates of CP if the weak interactions are CP invariant. With $\overline{K^0} = $ CP K^0, the eigenstates are:

$$\text{CP } K_1 = \text{CP}\,[(K^0 + \overline{K^0})/\sqrt{2}] = (\overline{K^0} + K^0)\sqrt{2} = K_1$$

$$\text{CP } K_2 = \text{CP}\,[(K^0 - \overline{K^0})/\sqrt{2}] = (\overline{K^0} - K^0)/\sqrt{2} = -K_2$$

Because CP $(\pi^+\pi^-) = (\pi^+\pi^-)$ for two π mesons in a state with zero angular momentum, the decay into $\pi^+\pi^-$ is allowed for the K_1, but forbidden for the K_2. Hence the K_2 is predicted to have a longer lifetime, because it must decay to three pions. This was indeed confirmed when the long-lived K_2 was discovered.

It was discovered by (Christenson, Cronin, Fitch & Turlay 1964), that the long-lived neutral K meson also decays to $\pi^+\pi^-$, but with a small branching ratio of $\sim 2 \times 10^{-3}$. From then on the long-lived state was called K_L because it was no longer identical to the CP eigenstate K_2. Similarly, the short-lived state was called K_S.

4   Phenomenology and models of CP violation

The phenomenon of CP violation in decays of neutral K^0 mesons has been with us for 37 years. The first ten years of intense experimentation after the discovery of the decay $K_L \to \pi^+\pi^-$ were devoted to the observation of other manifestations of the phenomenon, like the decay $K_L \to \pi^0\pi^0$ (Gaillard 1967, Cronin 1967), and the charge asymmetry in the decays $K_L \to \pi^\pm e^\mp \nu$ and $K_L \to \pi^\pm \mu^\mp \nu$ (Bennett 1967, Dorfan 1967), and to precision experiments on the moduli and phases of the CP violating amplitudes (Kleinknecht 1976). These experimental results excluded a large number of theoretical models proposed to explain CP violation, such that at the time of the London conference (Kleinknecht 1974) only two classes of models survived. The superweak model postulated a new, very weak, CP violating interaction (Wolfenstein 1964) with $\Delta S = 2$, whereas milliweak models invoked a small (10^{-3}) part of the normal $\Delta S = 1$ weak interaction as the source of CP violation. In the milliweak model there is also a direct decay of a Kaon state with CP quantum number -1 into a two-pion state with CP $+1$ through the weak Hamiltonian. This is called "direct" CP violation, as opposed to "indirect" CP violation by $K^0/\overline{K}^0$ mixing, which is a feature of both models. The key question is can one devise experiments to distinguish between these models?

In this context it was very important that a specific milliweak model within the Standard Model was proposed (Kobayashi & Maskawa 1973). At the time of the discovery of CP violation, only 3 quarks were known, and there was no possibility of explaining CP violation as a genuine phenomenon of weak interactions. This situation remained unchanged with the discovery of the fourth quark, because the 2×2 weak quark mixing matrix has only one free parameter, the Cabibbo angle, and no non-trivial complex phase. Kobayashi and Maskawa remarked that this picture would change if six quarks are present. Then the 3×3 mixing matrix naturally contains a complex phase, as well as three mixing angles. It is then possible to construct CP violating weak amplitudes from "box diagrams" of the form shown in Figure 5.
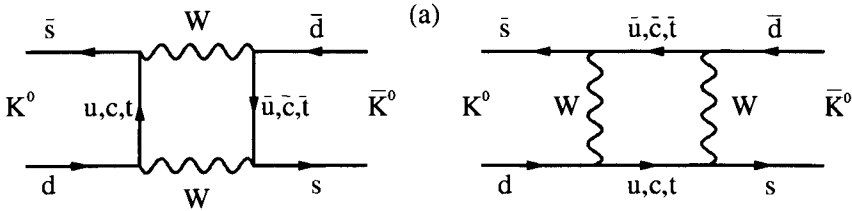
Figure 5. *Box diagram for $K^0-\bar{K}^0$ mixing connected to CP violating parameter ϵ.*

A consequence of this model of CP violation is that the relative decay rates for $K_L \to \pi^+\pi^-$ and $K_L \to \pi^0\pi^0$ are not necessarily equal. This direct CP violation is due to "Penguin diagrams" of the form given in Figure 6.
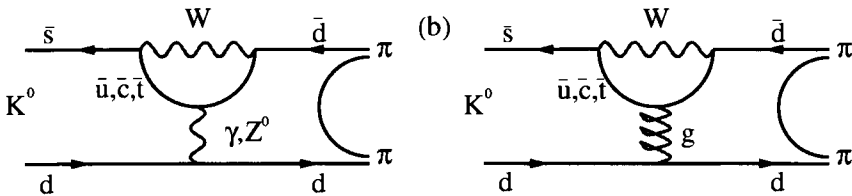
Figure 6. *Penguin diagrams for $K^0 \to 2\pi$ decay with direct CP violation.*

For a quantitative discussion, we use the conventional notation. The eigenstates of CP are:

$$K_1 = (K^0 + \bar{K}^0)/\sqrt{2} = +CPK_1$$
$$K_2 = (K^0 - \bar{K}^0)/\sqrt{2} = -CPK_2$$

The physical long-lived (K_L) and short-lived (K_S) states are defined as:

$$K_S = (K_1 + \epsilon_S K_2)/(1 + |\epsilon_S|^2)^{1/2}$$
$$K_L = (K_2 + \epsilon_L K_1)/(1 + |\epsilon_L|^2)^{1/2}$$

With CPT invariance, $\epsilon_S = \epsilon_L = \epsilon$, where:

$$\epsilon = \frac{\mathrm{Im}\Gamma_{12}/2 + i\mathrm{Im}M_{12}}{i(\Gamma_S - \Gamma_L)/2 - (m_S - m_L)}$$

is given in terms of the elements of the $K^0 - \bar{K}^0$ mass matrix M and decay matrix Γ. The phase of ϵ, is

$$\arg(\epsilon) = \arctan(2\Delta m/\Gamma_S) = 43.7° \pm 0.2°.$$

The experimentally observable quantities are

$$|\eta_{+-}|e^{i\phi_{+-}} = \eta_{+-} = \frac{< \pi^+\pi^-|T|K_L >}{< \pi^+\pi^-|T|K_S >},$$

$$|\eta_{00}|e^{i\phi_{00}} = \eta_{00} = \frac{< \pi^0\pi^0|T|K_L >}{< \pi^0\pi^0|T|K_S >}.$$

It can be shown that these amplitude ratios consist of a contribution from CP violation in the box diagrams describing $K^0 - \bar{K}^0$ mixing, called ϵ above, and another one from CP violation in the penguin diagrams associated with weak $K \rightarrow 2\pi$ amplitudes, called ϵ':

$$\eta_{+-} = \epsilon + \epsilon' \qquad \eta_{00} = \epsilon - 2\epsilon'$$

The quantities η_{+-}, η_{00} and $3\epsilon'$ form the Wu-Yang triangle in the complex plane. The CP violating decay amplitude ϵ' is due to the interference of a $\Delta I = 1/2$ amplitude, A_0, and a $\Delta I = 3/2$ amplitudes, A_2:

$$\epsilon' = \frac{i \mathrm{Im} A_2}{2 A_0} \exp[i(\delta_2 - \delta_0)]$$

and its phase is given by the $\pi\pi$ phase shifts in the $I = 0$ and $I = 2$ states, δ_0 and δ_2:

$$\arg(\epsilon') = (\delta_2 - \delta_0) + \pi/2$$

which is measured experimentally to be $(61 \pm 9)°$ (Biswas 1981) or $(45 \pm 9)°$ (Ochs 1981). The two models discussed above differ significantly in their predictions. In the superweak model there is no direct CP violation in weak decays, $\epsilon' = 0$, and therefore $\eta_{+-} = \eta_{00} = \epsilon$. In milliweak models it is likely that $\epsilon' \neq 0$.

5   Theoretical estimates for the parameter ϵ'

The parameter ϵ' can be estimated within the weak quark-mixing model of Kobayashi and Maskawa if one infers the magnitude of the quark mixing angles from other experiments, and calculates the hadronic matrix elements for box graphs and penguin graphs. Typical values for $|\epsilon'/\epsilon|$ are in the range $+0.2 - 3.0 \times 10^{-3}$ for three generations of quarks. A measurement of this quantity to this level of precision therefore becomes crucial for our understanding of CP violation. Since the phase of ϵ' is close to that of ϵ, and since $|\epsilon'/\epsilon| \ll 1$, we get:

$$\frac{\epsilon'}{\epsilon} = \mathrm{Re}\left(\frac{\epsilon'}{\epsilon}\right) = \frac{1}{6}\left(1 - |\eta_{00}/\eta_{+-}|^2\right)$$

Various methods are used to calculate the value of $\mathrm{Re}(\epsilon'/\epsilon)$. Due to the difficulties in calculating hadronic matrix elements, which involve long distance effects, the task turns out to be very difficult. The following results have been obtained recently:

1. The Dortmund group uses the $1/N_C$ expansion and Chiral Perturbation Theory (χPT). They quote a range of $1.5 \times 10^{-4} < \epsilon'/\epsilon < 31.6 \times 10^{-4}$ (Hambye 1999) from scanning the complete range of input parameters.

2. The Munich group uses a phenomenological approach in which as many parameters as possible are taken from experiment. Their result (Buras 1999) is $1.5 \times 10^{-4} < \epsilon'/\epsilon < 28.8 \times 10^{-4}$ from a scanning of the input parameters and $\epsilon'/\epsilon = (7.7^{+6.0}_{-3.5}) \times 10^{-4}$ using a Monte Carlo method to determine the error.

3. The Rome group uses lattice calculation results for the input parameters. Their result is $\epsilon'/\epsilon = (4.7^{+6.7}_{-5.9}) \times 10^{-4}$ (Martinelli 1999).

4. The Trieste group uses a chiral quark model to calculate ϵ'/ϵ. Their result is $7 \times 10^{-4} < \epsilon'/\epsilon < 31 \times 10^{-4}$ (Bertolini 1998).

It is hoped that reliable hadronic matrix elements will be obtained in the near future by lattice gauge theory calculations.

A measurement of the double ratio

$$R = \frac{|\eta_{00}|^2}{|\eta_{+-}|^2} = \frac{\Gamma(K_L \to 2\pi^0)/\Gamma(K_L \to \pi^+\pi^-)}{\Gamma(K_S \to 2\pi^0)/\Gamma(K_S \to \pi^+\pi^-)}$$

to a precision of better than 0.3% is required to distinguish between the milliweak and superweak models, and test these theoretical predictions.

6 Experiments on direct CP violation

6.1 The first observation of direct CP violation

The first observation of direct CP violation was made in 1988 by a collaboration of physicists at CERN. Their experiment, called NA31, was based on the concurrent detection of $\pi^0\pi^0$ and $\pi^+\pi^-$ decays from alternating collinear beams of K_S and K_L. Kaons with energies around 100GeV were produced by a 450GeV proton beam from the SPS accelerator at CERN. The energies of the decay products were measured in a combination of a high-resolution Liquid-Argon electromagnetic calorimeter and an iron-scintillator hadronic calorimeter. In the first runs, about 100000 decays of the type $K_L \to \pi^0\pi^0$ and 295000 decays of $K_L \to \pi^+\pi^-$ were observed, and the result for the CP parameter was $\mathrm{Re}(\epsilon'/\epsilon) = (33 \pm 11) \times 10^{-4}$ (Burkhardt 1998). In further data-taking, the number of observed $K_L \to \pi^0\pi^0$ decays was increased to 4×10^5.

A similar sensitivity was achieved by an experiment at Fermilab near Chicago, called E731. In 1992/93 the experiments NA31 at CERN and E731 at FNAL presented final results. The CERN result (Barr 1993) of $(23.0 \pm 6.5) \times 10^{-4}$ shows evidence for direct CP violation with a significance of more than 3σ, whereas the Fermilab result (Gibbons 1993) of $(7.4 \pm 5.9) \times 10^{-4}$ is consistent with zero. As a consequence of this disagreement, two new experiments, NA48 at CERN and kTeV at Fermilab, were constructed to resolve the issue.
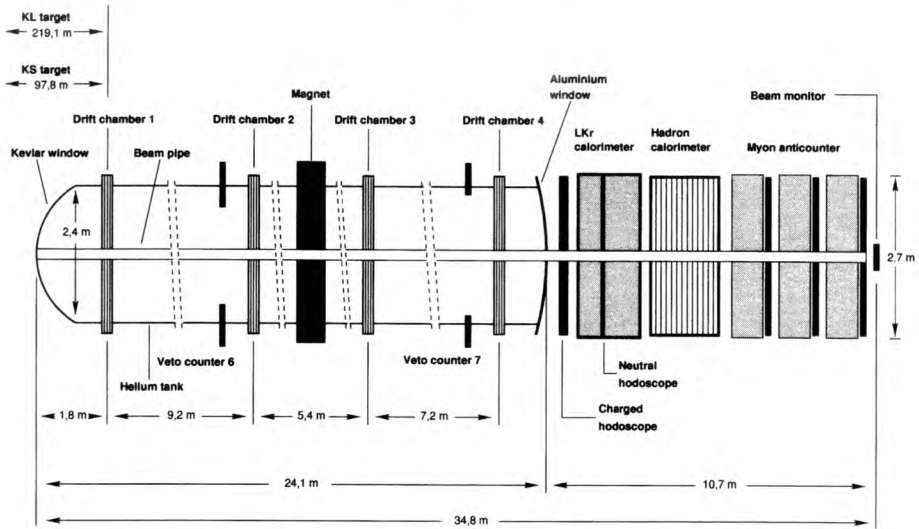
Figure 7. *Layout of the main detector components of the NA48 experiment.*

6.2   The NA48 detector

The new CERN experiment NA48 (Figure 7) was designed

- to measure all four decay modes concurrently,

- to register data at a rate 10 times higher than NA31,

- to achieve an improved energy resolution for photons (liquid Krypton calorimeter) and for charged pions (magnetic spectrometer).

In the design of the NA48 detector the cancellation of systematic uncertainties in the double ratio is exploited as much as possible. Important properties of the experiment are:

1. Two almost collinear beams which lead to an almost identical illumination of the detector.

2. The lifetime weighting of the events defined as K_L events, to remove systematic errors due to detector acceptance.

The K_L target is located 126m upstream of the beginning of the decay region. As the decay lengths at the average kaon momentum of 110GeV/c are $\lambda_S = 5.9$m and $\lambda_L = 3\ 400$m respectively, the neutral beam derived from this target gives predominantly K_L decays in the decay region. A separate K_S target is located only 6m upstream of the decay region, so this beam is a source of mainly K_S decays. The two beams are almost collinear, with the K_S target situated 7.2cm above the center of the K_L beam. The relative angle of the beams is 0.6mrad so that they converge at the position of the electromagnetic calorimeter.

The beginning of the K_S decay region is defined by an anti-counter (AKS). which is used to veto Kaon decays occurring upstream of the decay region. The position of the

AKS also defines the global Kaon energy scale which is directly correlated to the distance scale. The decay region itself is contained in a 90m long evacuated tank.

The identification of decays from the K_S beam is done by a tagger consisting of an array of scintillators situated in the proton beam which is directed onto the K_S target. If a proton signal is detected within a time window of ± 2ns with respect to a decay, the event is defined as K_S event. The absence of a proton tag defines a K_L event.

A magnetic spectrometer is used to reconstruct $K_{S,L} \to \pi^+\pi^-$ decays. The spectrometer consists of a dipole magnet with a transverse momentum kick of 265MeV/c, and four drift chambers which have a spatial resolution of $\sim 90\mu$m. This leads to a mass resolution of 2.5MeV/c^2. A hodoscope consisting of two planes of plastic scintillator provides the time of a charged event with a resolution of about 200ps.

A liquid krypton electromagnetic calorimeter (LKR) is used to identify the four photons from a $\pi^0\pi^0$ event. Liquid krypton has a radiation length of $X_0 = 4.7$cm which allows one to build a compact calorimeter with high energy resolution $\Delta E/E = 1.35\%$, as measured with electrons coming from a $K_L \to \pi e\nu$. It also has a very good time resolution of less than 300ps and very good linearity. It consists of 13212 cells (2×2cm^2) pointing to the average K_S decay position. The transverse spatial resolution is better than 1.3mm.

The electromagnetic calorimeter is complemented by an iron-scintillator sandwich calorimeter with a depth of 6.8 nuclear interaction lengths which measures the remaining energy of hadrons for use in the trigger for charged events.

A muon veto detector, consisting of three planes of scintillator shielded by 80cm of iron, is used to identify muons and veto $K_L \to \pi\mu\nu$ events.

6.3 Analysis of NA48 data

To identify events coming from the K_S target a coincidence window of ± 2ns between the proton signal in the tagger and the event time is chosen (see Figure 8a,c). Due to inefficiencies in the tagger and in the proton reconstruction a fraction α_{SL} of true K_S events are misidentified as K_L events. On the other hand there is a constant background of protons in the tagger which have not led to a good K_S event. If those protons accidentally coincide with a true K_L event, this event is misidentified as a K_S decay. This fraction α_{LS} depends only on the proton rate in the tagger and the width of the coincidence window.

Both effects, α_{SL}^{+-} and α_{LS}^{+-}, can be measured in the charged mode as K_S and K_L can be distinguished by the vertical position of the decay vertex (see Figure 8b). The results are $\alpha_{SL}^{+-} = (1.63 \pm 0.03)\times 10^{-4}$ and $\alpha_{LS}^{+-} = (10.649 \pm 0.008)\%$. This means that about 11% of true K_L events are misidentified as K_S events, however, this quantity is precisely measured to the 10^{-4} level. What is important for the measurement of the double ratio R is the difference between the charged and the neutral decay modes $\Delta\alpha_{LS} = \alpha_{LS}^{00} - \alpha_{LS}^{+-}$. Proton rates for both decays modes are measured in the sidebands of the tagging window to determine $\Delta\alpha_{LS}$. The result is $\Delta\alpha_{LS} = (4.3 \pm 1.8)\times 10^{-4}$. Several methods have been used to measure $\Delta\alpha_{SL}$, leading to the conclusion that there is no measurable difference between these mistaggings with an uncertainty of $\pm 0.5\times 10^{-4}$.

Another important correction is the background subtraction. Decays $K_L \to \pi e\nu$ and $K_L \to \pi\mu\nu$ can be misidentified as $K \to \pi^+\pi^-$ decays, as the ν is undetectable. However, since the ν carries away momentum and energy, these events can be identified by their
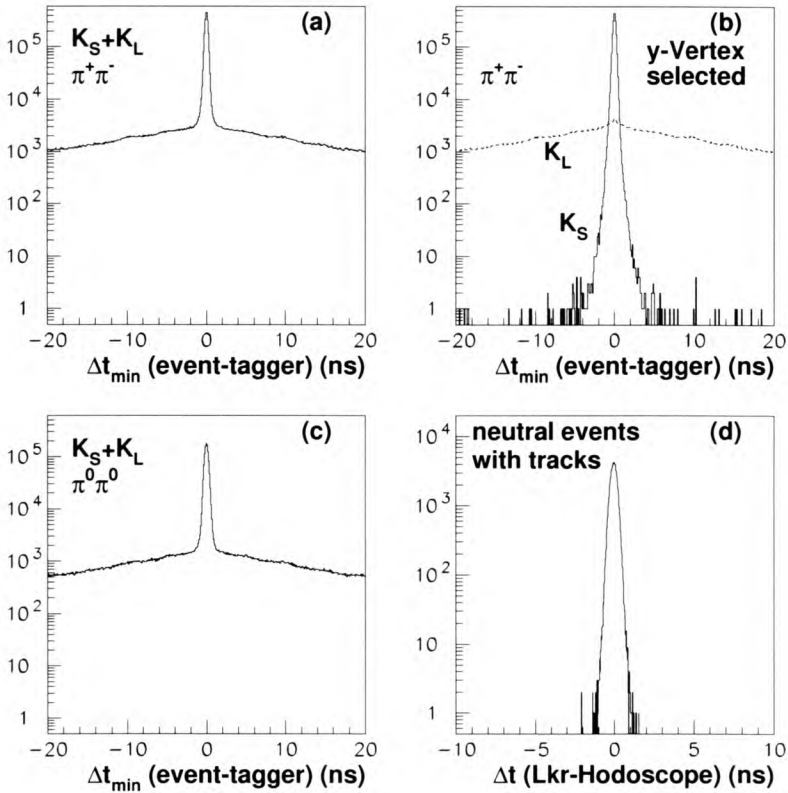
Figure 8. (a) and (c): minimal difference between tagger time and event time (Δt_{min}). (b) Δt_{min} for charged K_L and K_S events. (d) Comparison between charged and neutral event time. For this measurement decays, selected by the neutral trigger, with tracks are used (γ conversion and Dalitz decays $K_S \rightarrow \pi^0 \pi_D^0 \rightarrow \gamma\gamma\gamma e^+ e^-$).

high transverse momentum p'_t and their reconstructed invariant mass. The remaining background can be measured by extrapolating the shape of the background in the $m - p'^2_t$ plane into the signal region. In this way the charged background fraction leads to an overall correction on R of $(16.9 \pm 3.0) \times 10^{-4}$.

A similar extrapolation can be done in the neutral decay mode. Here the background comes from $K_L \rightarrow 3\pi^0$ decays, where two γ are not detected. This leads to a misreconstruction of the invariant π^0 masses. In this case the background leads to a correction of R by $(-5.9 \pm 2.0) \times 10^{-4}$.

The number of signal events after these corrections are summarised in Table 1. The efficiencies of the triggers used to record neutral and charged events have been determined. Independent triggers are used which accept a down-scaled fraction of events. In the neutral decay mode the efficiency is measured to be 0.99920 ± 0.00009 without measurable difference between K_S and K_L decays. The $\pi^+\pi^-$ trigger efficiency is measured to be

Event statistics ($\times\ 10^6$)					
	NA48	KTeV		NA48	KTeV
$K_S \to \pi^+\pi^-$	22.221	4.52	$K_L \to \pi^+\pi^-$	14.453	2.61
$K_S \to \pi^0\pi^0$	5.209	1.42	$K_L \to \pi^0\pi^0$	3.290	0.86

Table 1. *Event numbers after tagging correction (NA48) and background subtraction.*

(98.319 ± 0.038)% for K_L and (98.353 ± 0.022%) for K_S decays. Here a small difference between the trigger efficiency in K_S and K_L decays is found. This leads to a correction to the double ratio of $(-4.5 \pm 4.7) \times 10^{-4}$. The error on this measurement is dominated by the total number of events registered with the independent trigger. This error is one of the main contributions to the systematic error on the measurement of R.

The distance D from the LKR to the decay vertex is reconstructed using the position of the four γ clusters. From the kinematics of the decay one obtains

$$D = \frac{1}{M_K} \sqrt{\sum_{i,j} E_i\, E_j\, r_{ij}^2},$$

where E_i is the energy of cluster i and r_{ij} the distance between cluster i and cluster j. This formula directly relates the distance scale to the energy scale. It is therefore possible to fix the global energy scale with the measurement of the known AKS position in the neutral decay mode. More checks on the energy scale and the linearity of the energy measurement can be performed using the invariant π^0 mass and the known position of an added thin CH_2 target in which a π^- beam produces $\pi^0 \to 2\gamma$. The comparison of all methods gives an uncertainty of $\pm 3 \times 10^{-4}$ in the global energy scale.

Another systematic problem is the minimization of acceptance corrections, which is done by weighting the K_L events. The difference in the lifetime between K_S and K_L events produces a different illumination of the detector. There are more K_L events decaying close to the detector, whose decays products are measured at smaller radii closer to the beampipe. NA48 weights the K_L events according to the measured lifetime such that the distribution of the z-position of the decay vertices of K_S and K_L events and the detector acceptances become equal. Using this method the influence of detector inhomogeneities is minimised and the analysis becomes nearly independent of acceptance calculations by Monte Carlo methods. The price to pay for this reduction in a systematic error is a loss in statistical accuracy.

Although the acceptance is almost equal there are nevertheless small differences in the beam geometry and detector illumination between decays coming from the K_S and the K_L target. These remaining differences are corrected using Monte Carlo methods. to calculate that the deviation of the double ratio R from an input value of 1 is $(26.7 \pm 5.7) \times 10^{-4}$.

Summing up the systematic uncertainties of all different sources in quadrature, gives a total of 12.4×10^{-4} in R, and 2.1×10^{-4} in ϵ'/ϵ.

The result of NA48 using the data samples from 1997, 1998 and 1999 is(Fanti 1999, Lai 2001)

$$\epsilon'/\epsilon = (\ 15.3\ \pm\ 2.6\ )\ \times 10^{-4} \qquad (1)$$

6.4   The KTeV detector
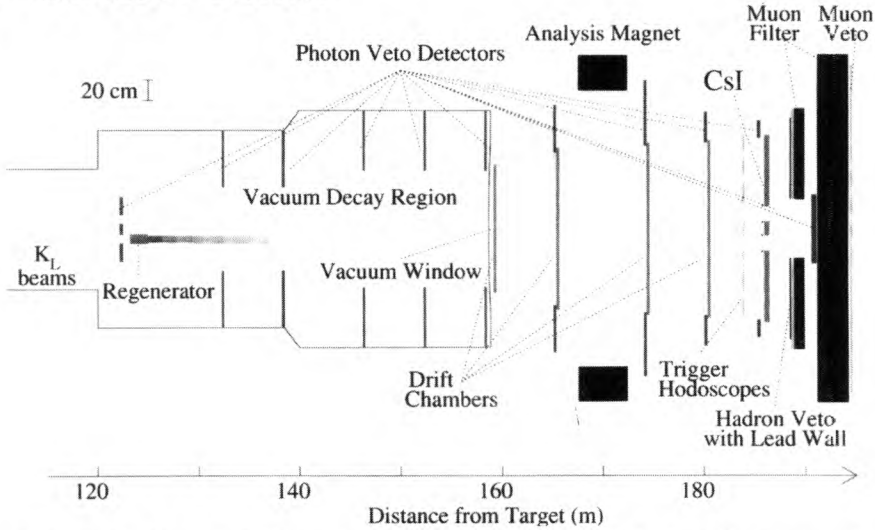
Figure 9. *Layout of the main detector components of the KTeV experiment*

The main elements of the KTeV detector (Figure 9) are similar to those in NA48, since both experiments work in a similar environment. The main difference is the way in which the K_S mesons are produced. KTeV uses two parallel well separated kaon beams derived from a single target. In one beam, the K_L mesons from the target pass through a collimator and then decay in an evacuated decay region of 30m length. In the other beam the K_L mesons traverse a slab of matter, known as a "regenerator". In the regenerator, the K^0 and $\overline{K^0}$ components of the K_L are affected differently by interactions with matter. Therefore the beam emerging behind the regenerator is a slightly different superposition of these two components. In the forward direction, it contains a K_S component which is coherent with the outgoing K_L component. This is the regenerated K_S beam used in the experiment. The regenerator alternates between both beams every minute in order to keep the detector illumination identical for the K_S and the K_L components.

The decay region of the regenerator beam is defined by a lead-scintillator counter at the downstream end of the regenerator. The decay region of the vacuum beam starts at the first photon veto detector.

The KTeV magnetic spectrometer consists of four drift chambers, with the magnet providing a momentum kick of $p_t = 412\text{MeV/c}$, leading to a mass resolution given by $\sigma(m_{\pi^+\pi^-}) = 1.6\text{MeV/c}^2$. For triggering of charged events a scintillator hodoscope is used.

The electromagnetic calorimeter consists of 3100 pure Cesium-Iodide (CsI) crystals with a radiation length of $X_0 = 1.85\text{cm}$. In the inner region the crystal size is $2.5 \times 2.5\,\text{cm}^2$ and in the outer region $5.0 \times 5.0\,\text{cm}^2$. Two beam holes of $15 \times 15\,\text{cm}^2$ allow the kaon beams to pass to the beam dump. The energy resolution at large photon energies above 10GeV is 0.75% as measured with K_{e3} decays.

Ten lead-scintillator "photon veto" counters are used to detect particles escaping the decay volume. The background is further reduced by a muon veto counter consisting of 4 m of steel and a hodoscope.
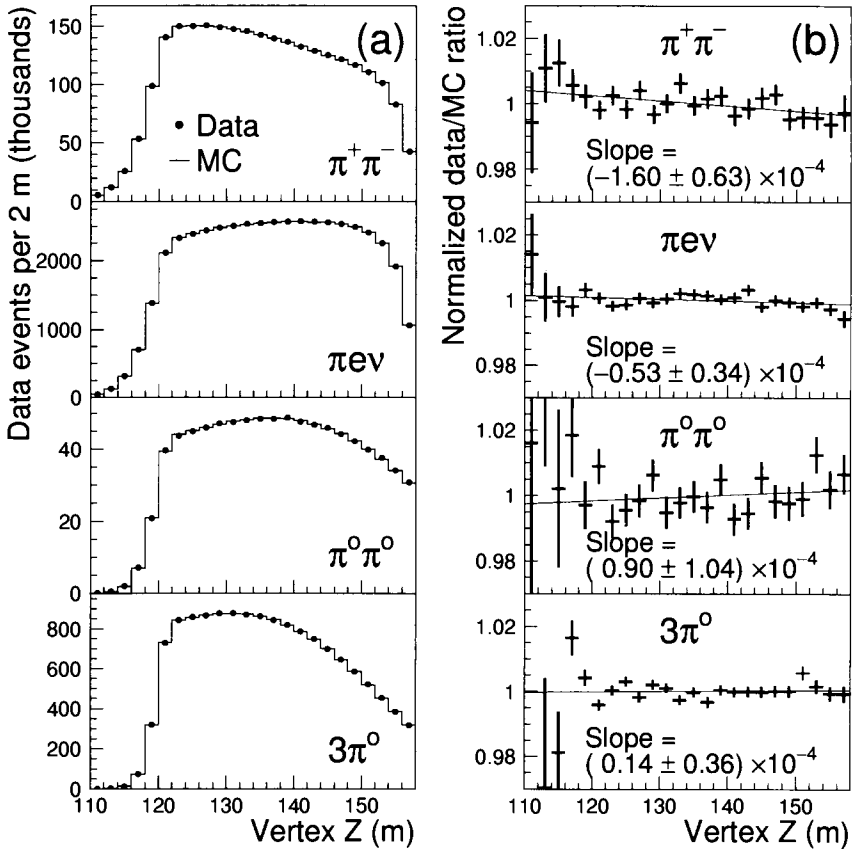
Figure 10. *(a) Data versus Monte Carlo comparisons of z-vertex distributions. (b) Linear fits to the data/MC ratio of (a).*

6.5 Analysis of KTeV data

KTeV has similar physical backgrounds to NA48, except that two-pion decays from K_S mesons produced by incoherent regeneration in the regenerator have to be subtracted. Typical numbers for the background fractions are 6.9×10^{-4} charged background from K_{e3} and $K_{\mu3}$ decays and 27×10^{-4} neutral background from $3\pi^0$ decays. The background levels in the regenerated beam are 107×10^{-4} in the neutral mode, which gives rise to a large systematic error, and 7.2×10^{-4} in the charged mode. The event numbers after background subtraction are given in Table 1.

The main difference in the analysis techniques of the two experiments is the treatment of the acceptance correction. KTeV does not use event weighting but uses Monte Carlo studies to correct for detection differences between the K_S and K_L decays. K_{e3} and $3\pi^0$ decays are used to model the detector and the agreement between data and Monte Carlo is good (see Figure 10). The acceptance correction to R is calculated by Monte Carlo

simulation to be $(231 \pm 13) \times 10^{-4}$. This can be compared to the size of the total effect of ϵ'/ϵ on R, 168×10^{-4}, as measured by KTeV. The main source of systematic uncertainty is a slight disagreement between data and Monte Carlo comparison in the $\pi^+\pi^-$ decay mode in the vacuum beam. A variation of $(-1.60 \pm 0.63) \times 10^{-4} \mathrm{m}^{-1}$ has been found which is applied as a systematic error.

The energy scale is determined using the known position of the regenerator edge. The comparison of the measured position of the vacuum window with the real position gives an uncertainty of the global energy scale which contributes an error of 4.2×10^{-4} on R.

The result is obtained by a fit of $\mathcal{R}e(\epsilon'/\epsilon)$, the regeneration amplitude and phase to the event numbers per energy bin. The result published in 1999 was (Alavi-Harati 1999)

$$\epsilon'/\epsilon = (\ 28.0\ \pm\ 3.0\ \text{(stat)}\ \pm\ 2.8\ \text{(sys)}\ )\times 10^{-4} \tag{2}$$

but this was subsequently corrected due to an error in background subtraction:

$$\epsilon'/\epsilon = (\ 23.2\ \pm\ 3.0\ \text{(stat)}\ \pm\ 3.2\ \text{(sys)}\ \pm\ 0.7\ \text{(MC)}\ )\times 10^{-4} \tag{3}$$
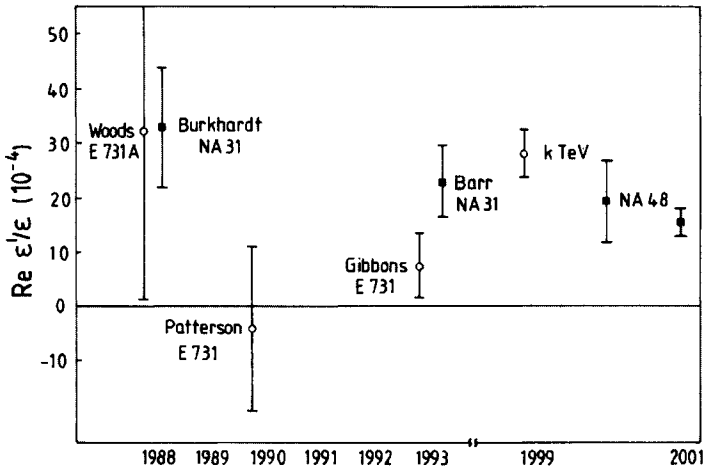
Figure 11. *Measurements of the parameter ϵ'/ϵ*

7   Conclusion

The two experiments KTeV and NA48 have definitively confirmed the original observation of NA31 that direct CP violation exists. The results of all published results on ϵ'/ϵ are shown in Figure 11. Therefore, the CP violation observed in the K meson system is a part of the weak interaction, due to weak quark mixing. Exotic new interactions such as the superweak interaction are not needed. With more data, both experiments will reach a precision of $\mathcal{O}(2 \times 10^{-4})$ on ϵ'/ϵ. The theoretical calculations of ϵ'/ϵ within the Standard Model are still not very precise, but this does not change the main conclusion of

the experiments, that ϵ' is different from zero and positive, and hence direct CP violation exists.

Unfortunately, this milliweak CP violation in the Kobayashi-Maskawa quark mixing matrix is probably not large enough to explain the observed matter-antimatter asymmetry from the big bang. An additional, stronger CP violation is needed for this. There are speculations that this might be due to CP violation in the lepton sector in the early universe. Very heavy Majorana neutrinos could play a role in the formation of an antimatter-matter asymmetry at this time (Buchmüller 2001).

References

Alavi-Harati A. et al. (KTeV Collaboration), 1999, *Phys. Rev. Lett.*,**83**, 22

Barr G.D. et al. (NA31 Collaboration), 1993, *Phys. Lett.*,**B 317**, 233

Bennett S. et al., 1967, *Phys. Rev. Lett.*,**19**, 993

Bertolini S., Fabbrichesi M., Eeg J.O., 1998, hep-ph/9802405

Biswas N.N. et al., 1981, *Phys. Rev. Lett.*,**47**, 1378

Buchmüller W., 2001, *Ann. Phys.*, **10**, 95

Buras A.J., Gorbahn M., Jäger S., Jamin M, Lautenbacher M.E., Silvestrini L., hep-ph/9904408, hep-ph/9908395

Burkhardt H. et al. (NA31 Collaboration), 1988, *Phys. Lett.*,**B 206**, 169

Christenson J.H., Cronin J.W., Fitch V.L., Turlay R., 1964, *Phys. Rev. Lett.*,**13**, 138

Cronin J.W. et al., 1967, *Phys. Rev. Lett.*,**18**, 25

Dorfan D. et al., 1967, *Phys. Rev. Lett.*, **19**, 987

Fanti V. et al. (NA48 Collaboration), 1999, *Phys. Lett.*, **B 465**, 335

Gaillard J.M. et al., 1967, *Phys. Rev. Lett.***18**, 20

Gell-Mann M., Pais A., 1955, *Phys. Rev.*,**97**, 1387

Gibbons L.K. et al. (E731 Collaboration), 1993, *Phys. Rev. Lett.*,**70**, 1203

Hambye T., Köhler G.O., Paschos E.A., Soldan P.H., 1999, hep-ph/9906434

Kleinknecht K., 1974, Proc. 17th Int. Conf. High Energy Phys., London ed. J.R. Smith, p.III-23, (Chilton, Didcot, UK)

Kleinknecht K., 1976, *Ann. Rev. Nucl. Sci.*,**26**, 1

Kobayashi M., Maskawa T., 1973, *Prog. Theor. Phys.*,**49**, 652

Cabibbo N., 1963, *Phys. Rev. Lett.*,**10**, 531

Lai A. et al., 2001, "A precise measurement of the direct CP violation parameter $Re(\epsilon'/\epsilon)$", *Europ. Phys. Journal*

Landau L. D., 1957, *Nucl. Phys.*, **3**, 127

Lee T. D., Yang C. N., 1956, *Phys. Rev.*,**104**, 254

Lüders, G, 1954, Kgl. Danske Videnskab, Selskab, *Matfys. Medd.*, **28(5)**, 1

Martinelli G., Presentation at Kaon99, Chicago
http://hep.uchicago.edu/kaon99/talks/martinelli/

Ochs W., 1981, private communication. from analysis of CERN-Munich data

Pauli W, 1955, *in "Niels Bohr and the Development of Physics"*, ed. W. Pauli, p. 30, (Oxford: Pergamon)

Wolfenstein L., 1964, *Phys. Rev. Lett.*,**13**, 562

Results from the B-meson factories

Klaus R Schubert

Technische Universität Dresden, Germany

1   Introduction and short history

Two experiments, BABAR at SLAC and Belle at KEK, succeeded this year in presenting new evidence for CP violation, so that 37 years after its discovery in decays of neutral K mesons (Christenson 1964), the effect is now established for the first time in a second system, the decays of neutral B mesons (Aubert 2001a, Abe 2001). Both experiments use double storage rings with asymmetric energies to produce B mesons in e^+e^- collisions on the $\Upsilon(4S)$ resonance. These lectures describe these so-called "B-Meson Factories", the experimental detectors, the special features of experiments on the $\Upsilon(4S)$, and the description of CP violation in quantum mechanics and in the Standard Model. The most recent experimental results are discussed.

Let me start with a short history of CP violation in B meson decays. The b quark was discovered in 1977 in high-energy proton-tungsten collisions at FNAL (Herb 1977), in the bound $b\bar{b}$ state $\Upsilon(1S) = 1^3S_1$, with a mass of 9.46GeV. Soon afterwards, the $\Upsilon(1S)$ (Berger 1978, Darden 1978a), and $\Upsilon(2S)$ (Bienlein 1978, Darden 1978b), were also produced in e^+e^- annihilation with the DORIS storage ring at DESY. Both these states, as well as the $\Upsilon(3S)$, are not heavy enough to decay into the lightest B mesons, $B^0(\bar{b}d)$ and $B^+(\bar{b}u)$, and their antiparticles $\overline{B}^0$ and B^-. The next $b\bar{b}$ resonance, the $\Upsilon(4S)$ with a mass of 10.58GeV, was first produced in Cornell's e^+e^- storage ring CESR (Andrews 1980, Finocchiaro 1980). Its decays into B^+B^- and $B^0\overline{B}^0$ pairs were first seen as an enhancement in lepton production from semileptonic B decays (Bebek 1981, Spencer 1981).

Soon after the discovery of the $\Upsilon(1S)$, it was already clear that B mesons could show large CP-violating effects (Ellis 1977). Plans to look for these effects experimentally led to proposals for B-meson factories with e^+e^- luminosities much higher than those achieved by the storage rings CESR and DORIS. The first proposal was worked out at the Swiss National Laboratory PSI (Eichler 1986, Wacker 1988), but was not funded. During this period, the ARGUS experiment at DORIS discovered $B^0\overline{B}^0$ oscillations (Albrecht 1987), leading to an enhanced interest in B-meson factories, and in the search for CP violation in B-meson decays. The subsequent proposals at SLAC (Hitlin 1989) and KEK (KEK 1990) were approved in 1993, and construction of the two machines PEP-II and KEK-B started more or less immediately.

The designs of the detectors BABAR and Belle were finalized and approved in 1995, and the first e^+e^- collisions in PEP-II took place in July 1998, with the first events recorded by BABAR in May 1999. In July 2000, the first BABAR and Belle results were reported at the International Conference for High Energy Physics at Osaka. In October 2000, PEP-II reaches its design luminosity of 3×10^{33}/cm^2/s, and in July 2001 BABAR and Belle discover CP violation in B-meson decays.

2   PEP-II and KEK-B

Figure 1 shows a schematic view of the B-meson factory PEP-II at SLAC. It consists of a double storage ring for electrons of 9GeV and positrons of 3.1GeV located in the 2.2km circumference tunnel of the dismantled PEP storage ring. The new electron ring uses bending magnets from PEP, but all the other components of PEP-II have been built in 1994–1998. The electrons are extracted from the SLAC linac at their nominal energy and then transferred into the HER (high energy ring) of PEP-II. The positrons are produced by high-energy electrons near the end of the linac, returned to the linac start point, damped by synchrotron radiation in two small rings, extracted from the linac at their nominal energy, and then transferred into the LER (low energy ring). The two rings intersect at only one point where the BABAR detector is located.
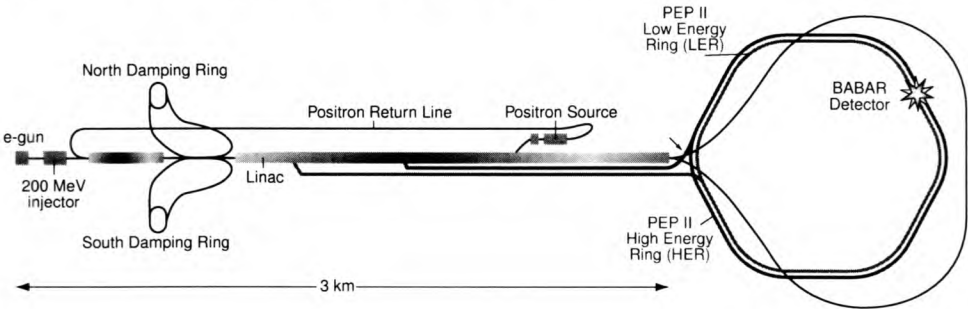
Figure 1. *Schematic view of the B-meson factory PEP-II at SLAC.*

By August 2001, a peak luminosity of $\mathcal{L} = 3.3\times10^{33}$/cm^2/s has been reached, (exceeding the design value of 3×10^{33}); this corresponds to 3.6 $e^+e^- \rightarrow \Upsilon(4S) \rightarrow B\overline{B}$ events per second, is obtained with an electron current of 0.88 A, and a positron current of 1.55A. Averaged over one day the best luminosity has been 220/pb/day, compared to a design value of 135/pb/day. Figure 2 shows the luminosity accumulated up to the end of August 2001. PEP-II has delivered 45 events/fb, and BABAR has recorded 43 events/fb. Out of these, 4/fb were recorded at a center-of-mass energy 40MeV below the $\Upsilon(4S)$ resonance for background studies, and 39/fb are on the resonance corresponding to 42×10^6 $B\overline{B}$ events. The asymmetric energies of the two beams cause the $B\overline{B}$ pairs to be produced with a significant velocity in the direction of the high energy beam, unlike in the symmetric storage rings CESR and DORIS. The Lorentz boost of the Centre of Mass system is:

$$\beta\gamma = \frac{|\mathbf{p}(e^+) + \mathbf{p}(e^-)|}{m_{\Upsilon(4S)}} = 0.56 . \tag{1}$$
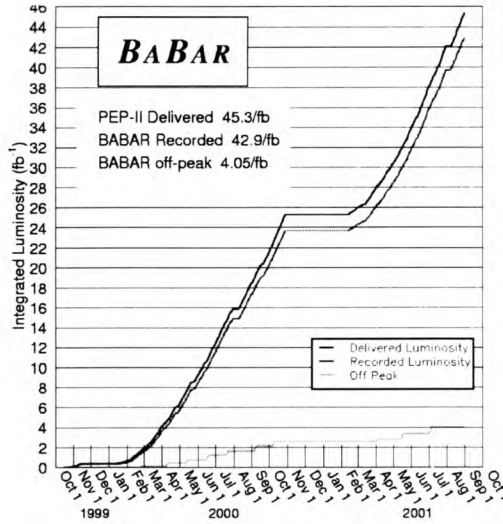
Figure 2. *Integrated luminosity of PEP-II up to August 2001.*

Figure 3 shows the B-meson factory, at the Japanese National Laboratory KEK. It is a double storage ring for electrons of 8GeV and positrons of 3.5GeV leading to a boost of $\beta\gamma = 0.45$, which is slightly smaller than in PEP-II. The second difference is a beam crossing angle of 22mrad in the interaction region, whereas PEP-II collides e^+ and e^- beams with zero crossing angle. The non-zero angle could make it possible to reach higher luminosities. Up to July 2001, KEKB has reached a peak luminosity of $\mathcal{L} = 4.5 \times 10^{33}/\text{cm}^2/\text{s}$, but the integrated luminosity of 32.8/fb is slightly lower than PEP-II. There is only one interaction region in KEK-B, which houses the Belle detector.
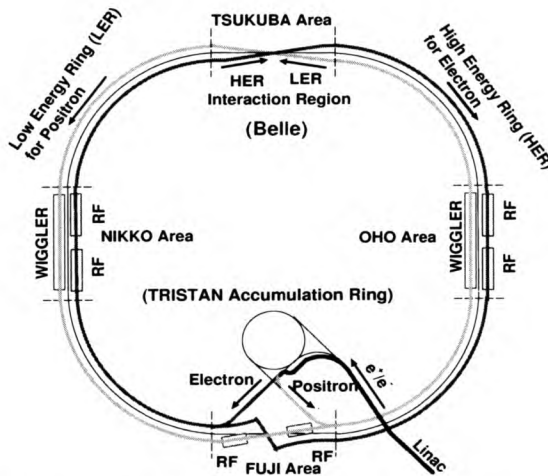
Figure 3. *Schematic view of the B-meson factory KEK-B at KEK.*

3   BABAR and Belle

Figure 4 shows the layout of the BABAR detector (Aubert 2001f). The BABAR Collaboration consists of about 520 physicists from 9 countries: Canada, China, France, Germany, Italy, Norway, Russia, UK, and USA. The detector geometry reflects the boosted event topology, with the e^+e^- interaction region displaced from the centre of the detector. There are six main detector components, a silicon vertex detector (SVX), a drift chamber (DCH), a Čerenkov detector (DIRC), an electromagnetic calorimeter (EMC), a magnet coil, and an instrumented flux return yoke (IFR).
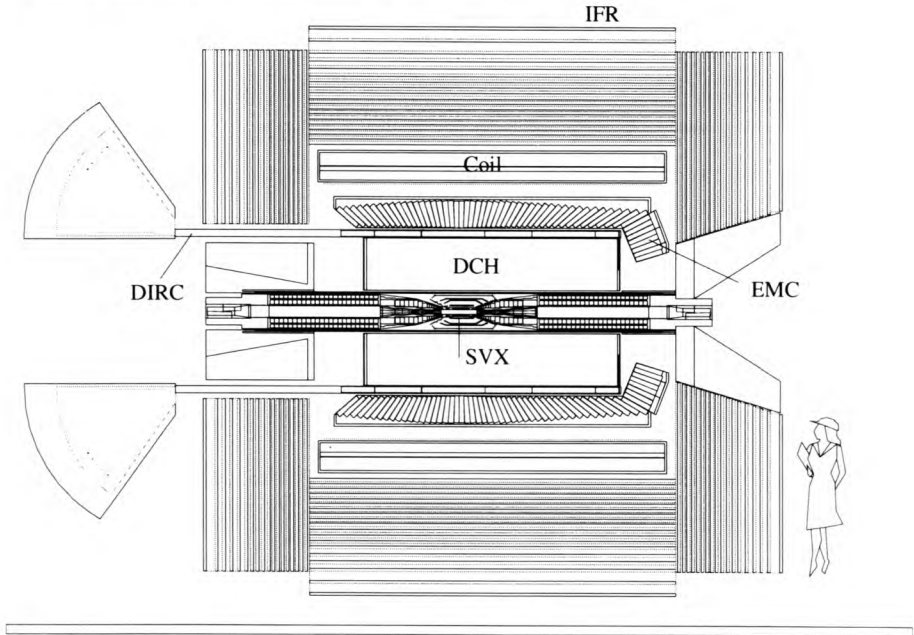
Figure 4. *Scheme of the BABAR detector. Abbreviations are explained in the text.*

The SVX reconstructs the decay vertices of the two B mesons. With the boost of $\beta\gamma = 0.56$, the average distance between the production and decay of a B meson is:

$$\beta\gamma c\tau = 260\mu m. \tag{2}$$

The production point $e^+e^- \to \Upsilon(4S) \to B\overline{B}$, cannot be reconstructed, and lies somewhere in the bunch-crossing region which has a length $\sigma_z \approx 1$cm. The decay point of a fully reconstructed B meson is found with a resolution $\sigma_z \approx 70\mu m$, which depends slightly on reconstructed final state. Partially reconstructed B mesons, which are used for flavour tagging, have a vertex resolution $\sigma_z \approx 180\mu m$. As explained in Section 4, all time-dependent B-meson decay information can be obtained from the distance between the two decay points, without any knowledge of the $B\overline{B}$ pair production point. This distance has a typical resolution of $\sqrt{70^2 + 180^2} = 190\mu m$, dominated by the tagging vertex. This is below the typical decay length of $260\mu m$ in Equation 2, and the $B^0\overline{B}^0$ oscillation length:

$$\frac{2\pi\beta\gamma c}{\Delta m_d} = 2.0\,\text{mm}. \tag{3}$$
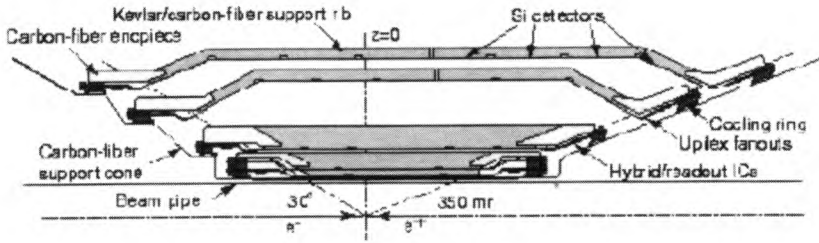
Figure 5. *The Silicon Vertex Detector of BABAR.*

The vertex resolution is achieved by tracking charged decay products with five layers of double-sided-readout silicon strip detectors in the geometry shown in Figure 5.

The drift chamber (DCH) tracks charged decay products in three dimensions through the magnetic field of 1.5T produced by the solenoid coil. It measures the track momenta, p_x, p_y, p_z, and their specific ionisation dE/dx. The DIRC determines their velocity by measuring the emission angle of Čerenkov light in 2.5cm thick quartz bars. Combining momentum and velocity information identifies the mass of the charged particle. The EMC measures the directions and energies of photons, identifies electrons, and also determines the direction of K_L^0 mesons that interact in the calorimeter. The IFR identifies muons, and measures the direction of K_L^0 mesons that do not interact in the EMC.
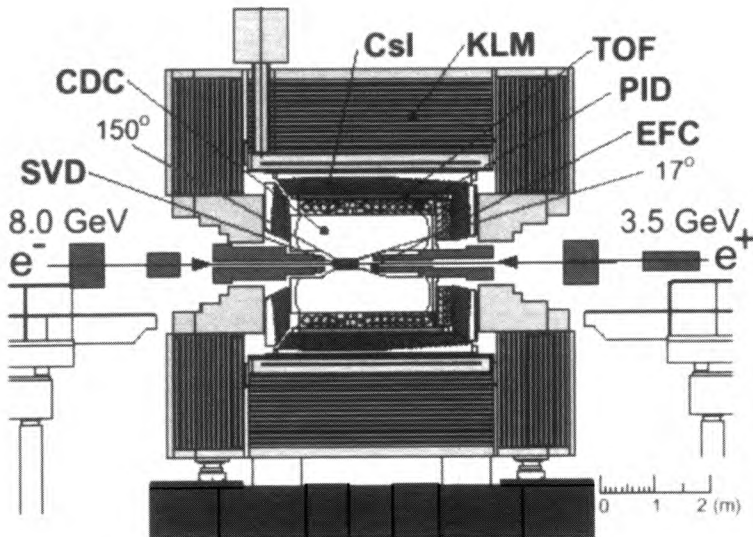
Figure 6. *Scheme of the Belle detector. Abbreviations are explained in the text.*

The Belle detector (Abashian 2001) is shown in Figure 6. The Belle Collaboration has about 300 physicists from 14 countries: Australia, Austria, China, Germany, India, Korea, Japan, Philippines, Poland, Russia, Slowenia, Switzerland, Taiwan, and USA. The
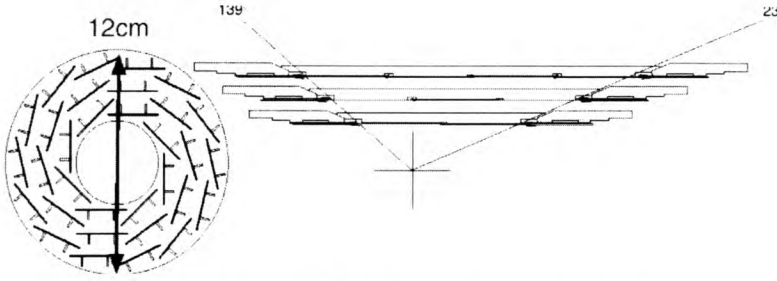
Figure 7. *Scheme of Belle's Silicon Vertex Detector.*

detector consists of a silicon vertex detector (SVD), a drift chamber (CDC), a particle identification system (PID), an electromagnetic calorimeter (CsI), and an instrumented flux return for K_L and muon identification (KLM). It is quite similar to BABAR, with the main difference being in the particle identification. Instead of a DIRC, where the Čerenkov angle is measured, Belle uses aerogel Čerenkov counters (EFC), which only determine velocity thresholds, and time-of-flight scintillators (TOF), which determine slow velocities. Belle's silicon vertex detector has only three layers, with the geometry shown in Figure 7. An upgrade to five layers is in progress.
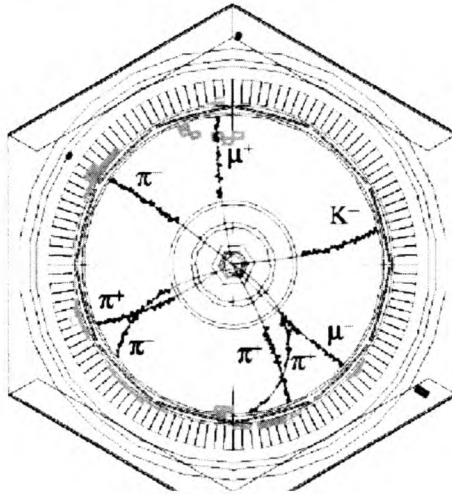
Figure 8. *A fully reconstructed event $B^0\overline{B}^0 \rightarrow (\psi(2S)K_S^0)(D^{*+}\pi^-)$ in BABAR.*

Figure 8 shows a $B^0\overline{B}^0$ event with both B-meson decays fully reconstructed in BABAR. One B meson decays into $D^{*+}\pi^-$ with $D^{*+} \rightarrow D^0\pi^+$, $D^0 \rightarrow K^-\pi^+$. This final state is a flavour eigenstate, i.e. the neutral B was a $\overline{B}^0$ at its time of decay. The other B decays into the final state $\psi(2S)K_S^0$ with $\psi(2S) \rightarrow \mu^+\mu^-$, $K_S^0 \rightarrow \pi^+\pi^-$, which is a CP eigenstate that can come from either flavour of neutral B. Such events are very rare, since a single B is fully reconstructed, with all its decay products detected and identified in BABAR, with a probability of only 10^{-3}.

4 Special features of $B\overline{B}$ pair decays

The $\Upsilon(4S)$ is a $b\bar{b}$ meson with mass $m = (10580.0 \pm 3.5)$MeV and width $\Gamma = (14 \pm 5)$MeV (PDG 2000). The cross section for its formation in e^+e^- annihilation is 1.1nb, and it decays by a strong interaction into $B\overline{B}$ pairs, 50% B^+B^- and 50% $B^0\overline{B}^0$. These final states consists of two and only two particles which decay by weak interactions on a time scale of 10^{-12}s.

The experimental study of these weak decays is influenced by the fact that B mesons from the $\Upsilon(4S)$ come in pairs. The mass of a B meson is 5279MeV (PDG2000), so in the center-of-mass system of the $\Upsilon(4S)$, each B has a kinetic energy of only $10580/2 - 5279 = 11$MeV, corresponding to a velocity $\beta^* = 0.064$, and a momentum $p^* = 340$MeV/c. Let us look first at the determination of the total decay rate of neutral and charged B mesons, through measurements of their lifetimes.
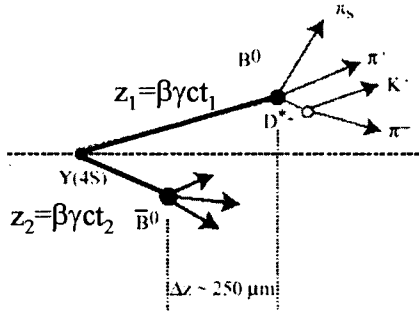
Figure 9. *Decay topology of $B\overline{B}$ pairs and definition of Δz. In the example, a B^0 decays into $D^{*-}\pi^+$ and the other neutral B into three non-identified tracks.*

Figure 9 sketches the decay topology of a $B\overline{B}$ pair. The two decay points are measured with the help of the silicon vertex detectors, but the production point is not measurable. Due to the small value of β^* in the $\Upsilon(4S)$ frame, compared to the boost factor $\beta\gamma = 0.56$ in BABAR (0.45 in Belle), we can approximate

$$\Delta z = z_1 - z_2 = \beta\gamma c(t_1 - t_2) = \beta\gamma c\Delta t . \tag{4}$$

Therefore, a good measurement of Δz provides a measurement of Δt. The lifetime, $\tau(B^+)$, gives the decay distribution of a B^+B^- pair:

$$d^2N = \frac{N}{\tau^2} e^{-t_1/\tau} e^{-t_2/\tau} dt_1 \, dt_2$$

$$= \frac{N}{\tau^2} e^{-\Delta t/\tau} e^{-2t_2/\tau} d\Delta t \, dt_2 ,$$

$$\frac{dN}{d\Delta t} = \begin{cases} \frac{N}{\tau} e^{-\Delta t/\tau} \int_0^\infty e^{-2t_2/\tau} dt_2/\tau & \text{for } t_1 > t_2, \\ \frac{N}{\tau} e^{-\Delta t/\tau} \int_{-\Delta t}^\infty e^{-2t_2/\tau} dt_2/\tau & \text{for } t_2 > t_1, \end{cases}$$

$$dN = \frac{N}{2\tau} e^{-|\Delta t|/\tau} d\Delta t = \frac{N}{2\tau} e^{-|t_1-t_2|/\tau} d(t_1 - t_2). \tag{5}$$
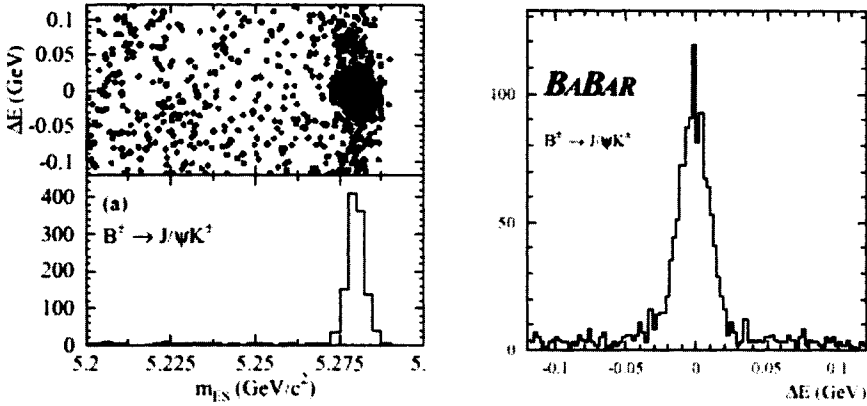
Figure 10. *Plots of* m_{ES} *and* ΔE^* *for* $B^{\pm} \to J/\psi K^{\pm}$ *decays in BABAR.*

Thus there are decay pairs with positive and negative values of $\Delta t = t_1 - t_2$, and the decay law is symmetric, $dN(\Delta t) = dN(-\Delta t)$, and depends only on the absolute decay time difference, $|\Delta t|$. The first BABAR measurement of τ uses pairs where one B is fully reconstructed with its decay vertex at z_1, and the second vertex at z_2 is obtained from the intersection of at least two reconstructed tracks from the other B.

B^+ mesons are fully reconstructed in the decay modes $D^0\pi^+$, $D^{*0}\pi^+$, $J/\psi K^+$, and $\psi(2S)K^+$. B^- mesons are reconstructed in the corresponding charge-conjugated modes. J/ψ mesons are reconstructed in their e^+e^- and $\mu^+\mu^-$ decay modes. Figure 10 shows how combinations of J/ψ candidates and K^+ tracks are identified as $B^+ \to J/\psi K^+$ decays, using the kinematic quantities, "beam-energy-substituted mass":

$$m_{\mathrm{ES}} = \sqrt{(m_{\Upsilon(4S)}/2)^2 - (\mathbf{p}^*_{J/\psi} + \mathbf{p}^*_K)^2} \,, \tag{6}$$

which has to be equal to $m(B^+)$ within the experimental resolution of $\pm 2.5 \mathrm{MeV}$, and the centre-of-mass "energy difference":

$$\Delta E^* = E^*_{J/\psi} + E^*_K - m_{\Upsilon(4S)}/2 \,, \tag{7}$$

which must be zero within the experimental resolution of $\pm 15 \mathrm{MeV}$. Figure 11 shows the m_{ES} distribution of all the ≈ 7300 $B^{\pm}$ candidates. Within the signal region, there are 93% signal events and 7% background. The Δt distribution of all events in the signal region is shown in Figure 12. The lifetime $\tau(B^+)$ is obtained by fitting the expectation in Equation 5 to the data, folding in the experimental resolution for Δt. This is described in detail in (Stark 2001). The result (Aubert 2001b) is given in Table 1. The lifetime of the B^0 is obtained in a similar way, but in this case the presence of $B^0\overline{B}^0$ oscillations has to be taken into account. Let us have a quick look at oscillation phenomenology before returning to the $\tau(B^0)$ measurement.

A B^0 meson produced at $t = 0$ decays at time t either as a B^0 or as a $\overline{B}^0$. Owing to second order weak interactions, illustrated in Figure 13, a neutral B meson is a linear superposition

$$|B\rangle = \psi_1|B^0\rangle + \psi_2|\overline{B}^0\rangle \tag{8}$$
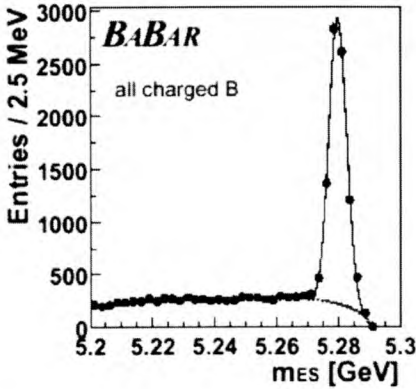
Figure 11. m_{ES} distribution of all fully reconstructed $B^{\pm}$ decay candidates in the BABAR lifetime measurement. The dotted curve is the background.
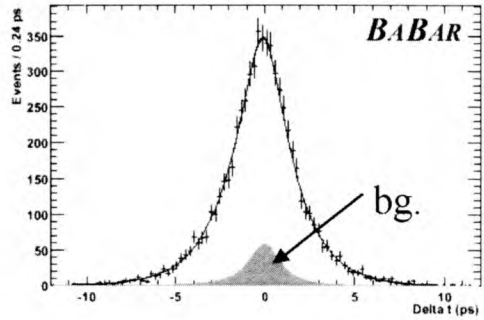
Figure 12. Δt distribution of all fully reconstructed $B^{\pm}$ decay candidates in the BABAR lifetime measurement.

	PDG 2000	BABAR	New Average
$\tau(B^+)$ in ps	1.653 ± 0.028	$1.673 \pm 0.032 \pm 0.023$	1.662 ± 0.023
$\tau(B^0)$ in ps	1.548 ± 0.032	$1.546 \pm 0.032 \pm 0.022$	1.542 ± 0.021
$\tau(B^+)/\tau(B^0)$	1.062 ± 0.029	$1.082 \pm 0.029 \pm 0.012$	1.080 ± 0.021

Table 1. *B Meson Lifetime Results (Stark 2001).*
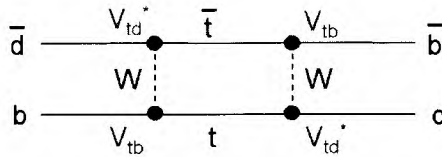
Figure 13. *Diagram describing $B^0\overline{B}^0$ oscillations.*

with a time dependence of ψ_1 and ψ_2 determined by initial conditions and the coupled differential equations,

$$ i\,\frac{\partial \psi_{ij}}{\partial t} = \mu_{ij}\,\psi_j \quad \text{where} \quad \mu_{ij} = \left[\begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} - \frac{i}{2} \begin{pmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{pmatrix} \right]. \tag{9} $$

The matrices m_{ij} and Γ_{ij} are Hermitean, with $m_{ji} = m_{ij}^*$ and $\Gamma_{ji} = \Gamma_{ij}^*$. CPT symmetry requires that $m_{11} = m_{22} = m$, $\Gamma_{11} = \Gamma_{22} = \Gamma$. There is also an unobservable phase, so the matrix μ_{ij} has only five physical parameters, m, Γ, $|m_{12}|$, Re (Γ_{12}/m_{12}), and Im (Γ_{12}/m_{12}). There are two and only two solutions of equation (9) with an exponential decay law:

$$ \begin{aligned} B_H^0(t) &= (pB^0 + q\overline{B}^0)e^{-\gamma_H t} & \gamma_H &= im_H + \Gamma_H/2, \\ B_L^0(t) &= (pB^0 - q\overline{B}^0)e^{-\gamma_L t} & \gamma_L &= im_L + \Gamma_L/2. \end{aligned} \tag{10} $$

The normalisation requirement is $|p|^2 + |q|^2 = 1$. Again the phase of q/p is not an observable, and the five observable quantities in equation (10) are m_H, m_L, Γ_H, Γ_L, and $|q/p|$. Explicit formulae relating these five observables to the five parameters of μ_{ij} can be found in the literature (Branco 1999). By convention (H=heavy, L=light), we have

$$\Delta m_d = m_H - m_L > 0. \tag{11}$$

The non-exponential decay laws of the two initial states:

$$\psi_B(0) = B^0, \qquad \psi_{\overline{B}}(0) = \overline{B}^0, \tag{12}$$

are given by:

$$\psi_B(t) = \frac{e^{-\gamma_H t} + e^{-\gamma_L t}}{2} B^0 + \frac{q}{p} \frac{e^{-\gamma_H t} - e^{-\gamma_L t}}{2} \overline{B}^0, \tag{13}$$

$$\psi_{\overline{B}}(t) = \frac{e^{-\gamma_H t} + e^{-\gamma_L t}}{2} \overline{B}^0 + \frac{p}{q} \frac{e^{-\gamma_H t} - e^{-\gamma_L t}}{2} B^0. \tag{14}$$

These decay laws include the probabilities P that a B^0 changes flavour, becoming a $\overline{B}^0$, and vice versa:

$$P(B^0 \to \overline{B}^0) = \frac{1}{4} \left| \frac{q}{p} \right|^2 \left[e^{-\Gamma_H t} + e^{-\Gamma_L t} - 2e^{-(\Gamma_H + \Gamma_L)t/2} \cos(\Delta m_d \cdot t) \right],$$

$$P(\overline{B}^0 \to B^0) = \frac{1}{4} \left| \frac{p}{q} \right|^2 \left[e^{-\Gamma_H t} + e^{-\Gamma_L t} - 2e^{-(\Gamma_H + \Gamma_L)t/2} \cos(\Delta m_d \cdot t) \right]. \tag{15}$$

In the Standard Model, CP and T violation in $B^0 \overline{B}^0$ oscillations is expected to be very small, and these probabilities are almost equal:

$$P(B^0 \to \overline{B}^0) = P(\overline{B}^0 \to B^0) + O(10^{-3}) \;\Leftrightarrow\; |q/p| = 1 + O(10^{-3}). \tag{16}$$

The lifetime difference should also be less than a percent:

$$\Gamma_H = \Gamma_L = \Gamma. \tag{17}$$

This leads to the good approximation:

$$P(B^0 \to \overline{B}^0) = P(\overline{B}^0 \to B^0) = e^{-\Gamma t}[1 - \cos(\Delta m_d \cdot t)]/2, \tag{18}$$

and a corresponding derivation gives:

$$P(B^0 \to B^0) = P(\overline{B}^0 \to \overline{B}^0) = e^{-\Gamma t}[1 + \cos(\Delta m_d \cdot t)]/2. \tag{19}$$

$B^0 \overline{B}^0$ pairs from $\Upsilon(4S)$ decays have to be parity eigenstates because of parity conservation in strong interactions. A two-particle state $B^0(\theta)\overline{B}^0(\pi - \theta)$, as sketched in Figure 14, is not a parity eigenstate owing to

$$\mathcal{P} \; B^0(\theta)\overline{B}^0(\pi - \theta) = B^0(\pi - \theta)\overline{B}^0(\theta). \tag{20}$$

Since the parity of the $\Upsilon(4S)$ is -1, the allowed state is

$$\psi = B^0(\theta) \; \overline{B}^0(\pi - \theta) - B^0(\pi - \theta) \; \overline{B}^0(\theta). \tag{21}$$
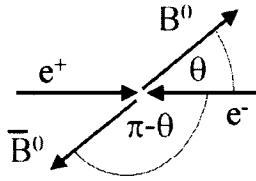
Figure 14. $B^0(\theta)\overline{B}^0(\pi - \theta)$ *is not a parity eigenstate.*

It remains in this coherent state until the first of the two B mesons decays. The state of the second B meson is defined at that time t_1. If the first decay is a flavour-specific $\overline{B}^0$, we have:

$$B_1(t_1) = \overline{B}^0, \qquad B_2(t_1) = B^0, \qquad B_2(t_2) = \alpha(t_2 - t_1)B^0 + \beta(t_2 - t_1)\overline{B}^0, \qquad (22)$$

with $\alpha(t)$ and $\beta(t)$ from equation (13), where $t = t_2 - t_1$. Equation 22 describes $B^0\overline{B}^0$ pair decays also for the case when only the second B decay at time t_2 is flavour-specific. $\alpha(t)$ and $\beta(t)$ then describe the time-changing flavour of the first B which decays at $t < 0$. If both decays are flavour-specific, we have measurements of B^0B^0, $B^0\overline{B}^0$, or $\overline{B}^0\overline{B}^0$ pairs as a function of t.

In a recent BABAR measurement of Δm_d (Aubert 2001c), neutral B mesons are fully reconstructed in flavour-specific decay modes, in the same manner as B^+ and B^- in the lifetime measurement discussed above. The second neutral B is partially reconstructed to obtain its flavour. This flavour measurement, which will be described in more detail in the CP violation Section 5.4, has a certain fraction of wrong flavour determinations:

$$w = \frac{N_{\text{wrong}}}{N_{\text{right}} + N_{\text{wrong}}} \approx 0.20. \qquad (23)$$

From measurements of the decay vertices, z_1, z_2, and equation (4), we obtain Δt for each reconstructed pair. Including the wrong flavour assignments, and the oscillation formalism with $|q/p| = 1$, and $\Gamma_L = \Gamma_H = \Gamma$, we have the number of unmixed and mixed pairs, and their time-dependent asymmetry $a(\Delta t)$:

$$N(B^0\overline{B}^0) = ne^{-\Gamma|\Delta t|}\left[(1-w)[1 + \cos(\Delta m_d\Delta t)] + w[1 - \cos(\Delta m_d\Delta t)]\right],$$

$$= ne^{-\Gamma|\Delta t|}\left[1 + (1 - 2w)\cos(\Delta m_d\Delta t)\right], \qquad (24)$$

$$N(B^0 B^0) = N(\overline{B}^0\overline{B}^0) = ne^{-\Gamma|\Delta t|}\left[1 - (1 - 2w)\cos(\Delta m_d\Delta t)\right], \qquad (25)$$

$$a(\Delta t) = \frac{N(B^0 B^0) - N(B^0\overline{B}^0)}{N(B^0 B^0) + N(B^0\overline{B}^0)} = (1 - 2w)\cos(\Delta m_d\Delta t). \qquad (26)$$

The time dependences of Equations 24 and 25 are shown in Figure 15a with perfect flavour-tagging, $w = 0$. In Figure 15b the effects of $w > 0$ and the experimental resolution in Δt are included. The BABAR results are shown in Figure 16. A fit gives:

$$\Delta m_d = (0.516 \pm 0.016 \pm 0.010)/\text{ps}. \qquad (27)$$
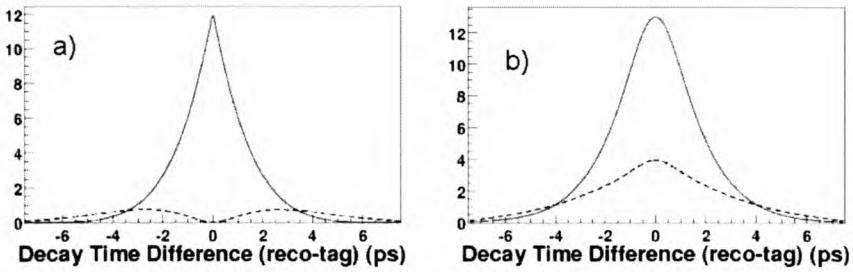
Figure 15. *Time dependence of unmixed and mixed pairs of neutral B mesons from the* $\Upsilon(4S)$, *(a) without and (b) with experimental resolutions, solid curves = unmixed, dotted curves = mixed.*
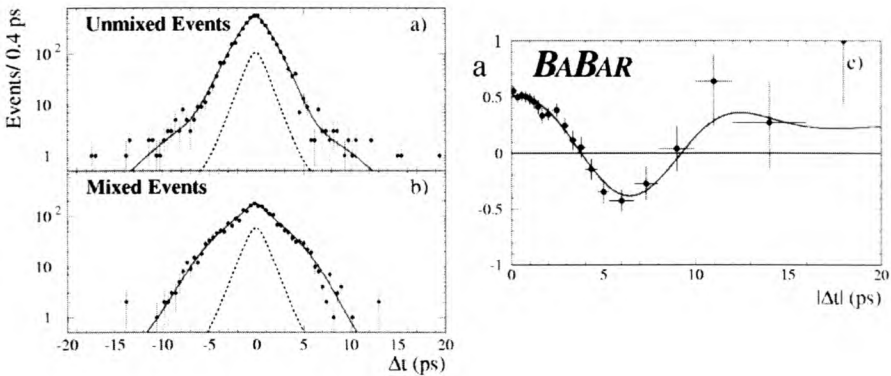
Figure 16. *BABAR result on* $B^0\overline{B}^0$ *mixing with fully reconstructed B mesons. The dotted curves indicate the backgrounds.*

BABAR and Belle have also measured Δm_d using other methods. A clean sample of "dilepton" events can be obtained, $\Upsilon(4S) \to \ell\ell X$, where ℓ is a positive or negative high momentum electron or muon. Positive leptons tag the decay of a B^+ or a B^0, and negative leptons a B^- or a $\overline{B}^0$, with a low mistag fraction w if the momentum cut is high enough. In 23 million $B\overline{B}$ events, BABAR finds 99000 dilepton events (Aubert 2001d), with a resolution on the vertex separation, $\sigma(\Delta z) \approx 250\mu m$. The asymmetry

$$a(\Delta t) = \frac{N(\ell^+\ell^-) - N(\ell^\pm\ell^\pm)}{N(\ell^+\ell^-) + N(\ell^\pm\ell^\pm)}, \tag{28}$$

is shown in Figure 17. Its expectation is different from that in Equation 26 and Figure 16 because of the presence of $\ell^+\ell^-$ pairs from B^+B^- events. Taking this into account, the fit gives the preliminary result

$$\Delta m_d = (0.499 \pm 0.010 \pm 0.012)/\text{ps}. \tag{29}$$

A compilation of Δm_d measurements is given in Figure 18. The "world average" of
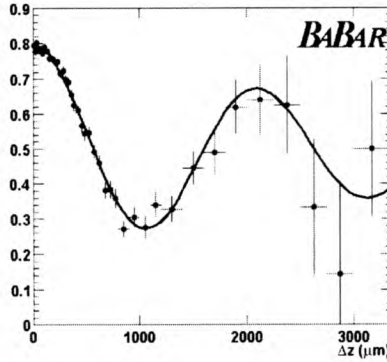
Figure 17. *BABAR result on $B^0\overline{B}^0$ mixing with lepton pairs.*
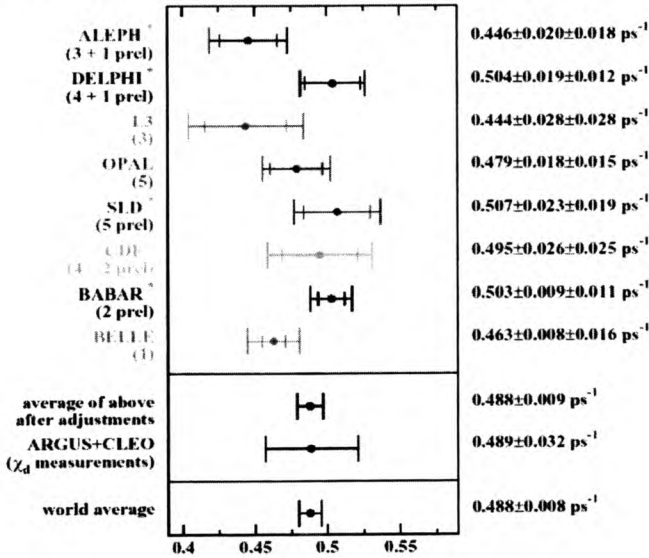
Figure 18. *Compilation of Δm_d results.*

these measurements is

$$\Delta m_d = (0.488 \pm 0.009)/\text{ps}. \tag{30}$$

When ARGUS discovered $B^0\overline{B}^0$ oscillations (Albrecht 1987), they used a time-integrated measurement of dilepton events, which gave:

$$\chi = \frac{(\Delta m_d)^2}{2\Gamma^2 + 2(\Delta m_D)^2} = 0.17 \pm 0.05, \qquad \Delta m_d = (0.47 \pm 0.10)/\text{ps}, \tag{31}$$

where the value of Δm_d has been adjusted for the present value of $\Gamma = 1/\tau(B^0)$. The Standard Model expectation is:

$$\Delta m_d = \frac{G_F^2}{6\pi^2}\eta_B m(B^0)\, f(B^0)^2\, B(B^0)\, m(W)^2\, S\left(\frac{m(t)}{m(W)}\right) |V_{tb}|^2 |V_{td}|^2 \tag{32}$$

which relates Δm_d to the CKM matrix element $|V_{td}|$ with the Fermi constant G_F, a parameter η_B from perturbative QCD, and a parameter $f\sqrt{B}$ from non-perturbative QCD. The Inami-Lim function S and $\eta_B = 0.55 \pm 0.01$ are known with small errors. The dominant theoretical uncertainty in the relationship between Δm_d and $|V_{td}|$ comes from low-energy QCD. For this, we will use the lattice QCD result:

$$f(B^0)\sqrt{B(B^0)} = (230 \pm 25 \pm 20) \text{ MeV}. \qquad (33)$$

Let me conclude this section by returning to the B^0 lifetime measurement. Adding Equations 24 and 25 leads to the result

$$dN(\Delta t) = N(B^0\overline{B}^0) + N(B^0 B^0) = \frac{N \cdot \Gamma}{2} e^{-\Gamma|\Delta t|} d\Delta t, \qquad \Gamma = \frac{1}{\tau(B^0)}, \qquad (34)$$

which is as simple as Equation (5) for the determination of $\tau(B^+)$ from $B^+ B^-$ pairs. In the same analysis (Stark 2001, Aubert 2001b), BABAR fully reconstructs 7000 neutral B mesons in the modes $D^{(*)\pm}\pi^\mp$, $D^{(*)\pm}\rho^\mp$, $D^{(*)\pm}a_1^\mp$, and $J/\psi K^{*0}$, where $D^{(*)\pm}$ is either a D or a D^*. The sample purity of 90% is similar to charged B reconstruction. The other neutral B is partially reconstructed to measure the decay vertex, but without requiring a flavour determination. The fit to the Δt distribution gives the results in Table 1 (on page 353). The lifetime ratio $\tau(B^+)/\tau(B^0)$ is significantly different from unity. This starts to be well understood, see (Buchalla 2001) in this School.

5 CP Violation in B-Meson decays

5.1 CP violation in the Standard Model

Understanding CP violation is the main goal of the two experiments BABAR and Belle:

- Is there CP violation in B decays and how large is it?

- Can this violation be explained by the Standard Model weak interaction?

- Does the same explanation hold for the known CP violation in K meson decays?

The answers to the second and third questions also require precision measurements of B meson decays without CP violation. Note that the discovery of CP violation in K meson decays does not mean that CP is violated in the Standard Model, since its origin could be in a non-Standard new interaction.

In the Standard Model, CP violation is a simple property of the Higgs couplings to quarks if there are more than two families of quarks. Since the three families of leptons and quarks are well established and the Higgs mechanism has been treated in the lecture of (Rosner 2001) at this School, I will not go into details here. The essence is that the Higgs mechanism mixes quarks in a different way from antiquarks. The Glashow doublet partners of the quark mass eigenstates u, c, and t, are linear combinations d', s', and

b' of the mass eigenstates d, s, and b. These combinations are different for quarks and antiquarks:

$$
\begin{aligned}
d' &= V_{ud} \cdot d + V_{us} \cdot s + V_{ub} \cdot b\,, \\
s' &= V_{cd} \cdot d + V_{cs} \cdot s + V_{cb} \cdot b\,, \\
b' &= V_{td} \cdot d + V_{ts} \cdot s + V_{tb} \cdot b\,,
\end{aligned}
\tag{35}
$$

$$
\begin{aligned}
\overline{d}' &= V_{ud}^* \cdot \overline{d} + V_{us}^* \cdot \overline{s} + V_{ub}^* \cdot \overline{b}\,, \\
\overline{s}' &= V_{cd}^* \cdot \overline{d} + V_{cs}^* \cdot \overline{s} + V_{cb}^* \cdot \overline{b}\,, \\
\overline{b}' &= V_{td}^* \cdot \overline{d} + V_{ts}^* \cdot \overline{s} + V_{tb}^* \cdot \overline{b}\,.
\end{aligned}
\tag{36}
$$

Here V_{ij} is a unitary quark mixing matrix , named CKM after (Cabibbo 1963) and (Kobayashi Maskawa 1973). CP violation arises only if $V_{ij} \neq V_{ij}^*$. Since the gauge invariance of the Higgs coupling term in the Standard Model Lagrangian does not explicitly forbid complex coupling matrices, it is expected to lead to a complex matrix V_{ij}. This complexity is a necessary but not sufficient condition for CP violation. Because of unobservable phases, we must have:

$$
\mathrm{Im}\,(V_{ik}V_{il}^* V_{jl}V_{jk}^*) = J \neq 0.
\tag{37}
$$

The quantity J is the same for all $i \neq j$ and $k \neq l$ (Jarlskog 1985).

The CKM matrix has four observable parameters. There are various parametrisations, with experimentalists prefering that of (Wolfenstein 1984) ,

$$
V = \begin{pmatrix}
1 - \lambda^2/2 & \lambda & A\lambda^3(\rho - i\eta) \\
-\lambda & 1 - \lambda^2/2 & A\lambda^2 \\
A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1
\end{pmatrix} + O(\lambda^4)\,, \qquad J \approx A^2\lambda^6\eta.
\tag{38}
$$

The four parameters of the matrix, A, λ, ρ, and η, are 4 of the 18 free parameters of the Standard Model. Six unitarity conditions $VV^\dagger = 1$ can be drawn as triangles (Bjorken 1987) in the complex plane. All six have the same area $J/2$. The most popular one is shown in Figure 19:

$$
V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0,
\tag{39}
$$

$$
V_{ub}^*/(A\lambda^3) - 1 + V_{td}/(A\lambda^3) = 0.
\tag{40}
$$

Measurements of K and B semileptonic decays, which are not discussed here, give

$$
\lambda = 0.220 \pm 1\%, \qquad A = 0.83 \pm 5\%.
\tag{41}
$$

Information on ρ and η comes from several observations, as summarized in Figure 20. The figure shows $\overline{\rho} = \rho\sqrt{1 - \lambda^2}$ and $\overline{\eta} = \eta\sqrt{1 - \lambda^2}$, which are different from ρ and η by about 2.5%. As can be seen from Equation 39, $V_{td} = A\lambda^3(1 - \overline{\rho} - i\overline{\eta})$, together with $V_{ub} = A\lambda^3(\rho - i\eta)$, fulfill unitarity more accurately than Equation 40. Consequently the use of $\overline{\rho}$ and $\overline{\eta}$ will become important when the experiments have reached a precision of 3% or better.
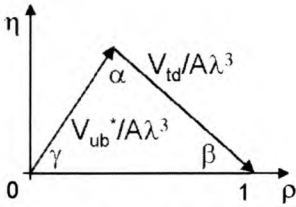
Figure 19. *The Unitarity Triangle.*
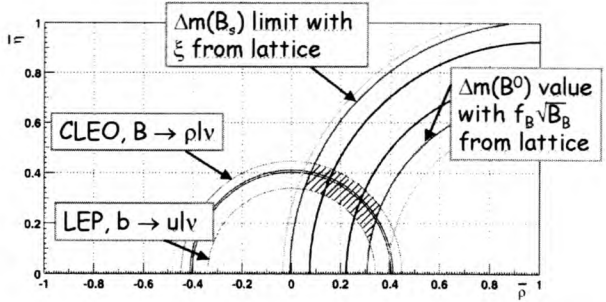
Figure 20. *Circular bands for ρ and η from measurements of $|V_{ub}|$ and $|V_{td}|$. The inner circles give the regions of maximum likelihood, the outer ones give one and two standard deviations.*

The measurements of $\Gamma(b \to u\ell\nu)$ (LEP 1999), and $\Gamma(B \to \rho\ell\nu)$ (CLEO 1999), are sensitive to $|V_{ub}|$. The measurement of Δm_d, discussed in the previous section, and the upper limit on B_s mixing, Δm_s (LEP 2000), are both sensitive to $|V_{td}|$. All these results are consistent, and we conclude that, with a significance of $\geq 1.7\sigma$, $\eta \neq 0$, and $|J| \approx 3 \times 10^{-5} \neq 0$. These CP conserving measurements already show that CP symmetry is violated in the Standard Model. The significance will increase when the theoretical uncertainties in the circular bands of Figure 20 are reduced.

The B-Meson Factories have so far contributed to the measurement of Δm_d, but they do not intend to measure Δm_s which would require running the storage rings on the $\Upsilon(5S)$ resonance. They will contribute new results on $|V_{ub}|$ starting next year.

Two immediate questions arise from Figure 20:

- Does $|J| = 3 \times 10^{-5}$ explain the known value of ϵ_K for CP violation in K decays?

- What are the predictions for CP violation in B meson decays?

The first answer is yes, as shown in Figure 21. CP violation in K decays has been discussed here by (Kleinknecht 2001). The band in Figure 21 follows from $|\eta_{+-}| \approx |\eta_{00}| \approx 2.3 \times 10^{-3}$. The experimental width of the band is less than 1%, so the width is dominated by the theoretical uncertainty on B_K. Adding in the ϵ_K constraint fixes the signs of η and J to be positive. Figure 22 shows the result of a maximum-likelihood fit to the experimental results for $|V_{ub}|$, $|V_{td}|$, and $\epsilon(K)$ including theoretical uncertainties in the relationship of these measurements to ρ and η.

As discussed here by (Nir 2001), there are three classes of CP violation which can be observed in the B-meson system; in $B^0\overline{B}^0$ oscillations, in decays ("direct CP violation"), and in the interference between oscillations and decay. In the Standard Model with ρ and η as fitted in Figure 22, the largest effects are expected in the third class, where the decay mode is a CP eigenstate. There are some modes where the prediction for the CP asymmetry is very insensitive to theoretical uncertainties, especially

$$a_f = \frac{dN(\overline{B}^0 \to J/\psi K_S^0) - dN(B^0 \to J/\psi K_S^0)}{dN(\overline{B}^0 \to J/\psi K_S^0) + dN(B^0 \to J/\psi K_S^0)} = [\sin 2\beta \pm O(1\%)]\sin(\Delta m_d \Delta t), \quad (42)$$
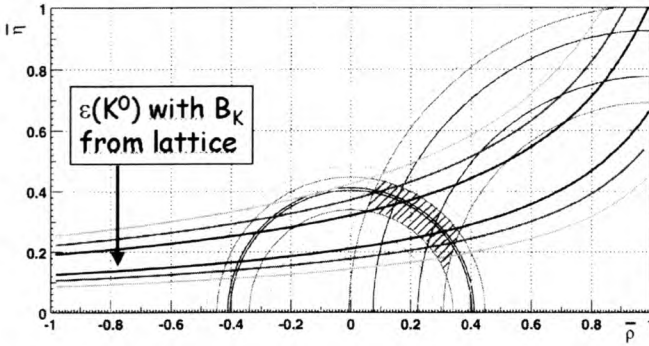
Figure 21. *The hyperbolic band shows the relation of $\overline{\rho}$ and $\overline{\eta}$ to the measured value of ϵ_K.*
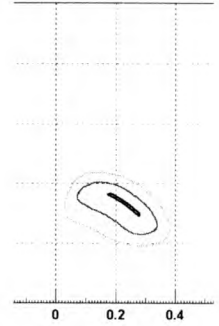
Figure 22. *Fit result for $\overline{\rho}$ and $\overline{\eta}$.*

where β has been defined in Figure 19:

$$\sin 2\beta = \frac{2\eta(1-\rho)}{\eta^2 + (1-\rho)^2}, \tag{43}$$

and Δt is the difference between the decay time into $J/\psi K$, and the decay time of other B in the $B^0\overline{B}^0$ pair. Before coming to the BABAR and Belle results for a_f, I would like to discuss briefly the other two classes.

5.2 CP violation in $B^0\overline{B}^0$ oscillations

The two mass and lifetime eigenstates in the neutral B system have been defined in equation (10). CP conservation in $B^0\overline{B}^0$ oscillations is equivalent to $|q/p| = 1$. If CP is violated, we may use the Kaon language for p and q,

$$\epsilon_B = \frac{p-q}{p+q}, \qquad \frac{p}{q} = \frac{1+\epsilon_B}{1-\epsilon_B}, \qquad \left|\frac{p}{q}\right| = 1 + \frac{2\mathrm{Re}\,(\epsilon_B)}{1+|\epsilon_B|^2}. \tag{44}$$

Because there are several ϵ parameters in Kaon physics, we choose the notation ϵ_B for ϵ in mixing. The imaginary part of ϵ_B is unobservable. Since the Standard Model expects $\mathrm{Re}\,(\epsilon_B)$ to be small, as ϵ_K is in $K^0\overline{K}^0$ oscillations, a useful convention is to take $\mathrm{Im}\,(\epsilon_B) = 0$, giving $|p/q| = 1 + 2\mathrm{Re}\,(\epsilon_B)$.

CP violation in oscillations shows up as a different number of B^0B^0 and $\overline{B}^0\overline{B}^0$ decay pairs. A straightforward calculation gives

$$a_m(\Delta t) = \frac{dN(B^0B^0)(\Delta t) - dN(\overline{B}^0\overline{B}^0)(\Delta t)}{dN(B^0B^0)(\Delta t) + dN(\overline{B}^0\overline{B}^0)(\Delta t)} = \frac{4 \cdot \mathrm{Re}\,(\epsilon_B)}{1+|\epsilon_B|^2}. \tag{45}$$

This asymmetry is independent of Δt. Table 2 gives results for $\mathrm{Re}\,(\epsilon_B)$, including a recent BABAR result (Brandt 2001) which uses 21 million $\Upsilon(4S)$ events. No CP violation is seen. This is not surprising since the Standard Model expectation is below the sensitivities reached by current experiments.

Experiment	Reference	Method	Sample	$\epsilon_B/(1+|\epsilon_B|^2)$
OPAL	(Ackerstaff 1998)	$\ell^+\ell^+/\ell^-\ell^-$	3M Z^0	$0.002 \pm 0.007 \pm 0.003$
OPAL	(Abbiendi 1999)	inclusive	4M Z^0	$0.001 \pm 0.014 \pm 0.003$
ALEPH	(Barate 2001)	combined	4M Z^0	-0.003 ± 0.007
CLEO	(Jaffe 2001)	$\ell^+\ell^+/\ell^-\ell^-$	10M $\Upsilon(4S)$	$0.004 \pm 0.010 \pm 0.002$
BABAR	(Brandt 2001)	$\ell^+\ell^+/\ell^-\ell^-$	21M $\Upsilon(4S)$	$0.0012 \pm 0.0029 \pm 0.0036$

Table 2. *Searches for CP violation in $B^0\overline{B}^0$ oscillations.*

5.3   CP violation in decays

BABAR, Belle, and CLEO have searched for direct CP asymmetries in various decay modes. As explained in (Nir 2001), the decay amplitude $A(B \to f)$ with $B = B^+$ or B^0 has to be described by at least two diagrams, with different weak phases and with different strong phases, in order to show a direct CP asymmetry:

$$a_{\mathrm{CP}}(f) = \frac{\Gamma(\overline{B} \to \overline{f}) - \Gamma(B \to f)}{\Gamma(\overline{B} \to \overline{f}) + \Gamma(B \to f)}. \tag{46}$$

Table 3 shows the limits on asymmetry values for five decay modes. For details and references, see (Hamel de Monchenault 2001). Again, no indication for CP violation is seen, but this is not in contradiction with Standard Model expectations.

Mode	BABAR	Belle	CLEO
$K^+\pi^-$	-0.19 ± 0.10	$+0.04 + 0.19 - 0.17$	-0.04 ± 0.16
$K^+\pi^0$	0.00 ± 0.18	$-0.06 + 0.22 - 0.20$	-0.29 ± 0.23
$K_S^0\pi^+$	-0.21 ± 0.18	$+0.10 + 0.43 - 0.34$	-0.18 ± 0.24
$J/\psi K^+$	$-0.009 \pm 0.027 \pm 0.005$		$+0.018 \pm 0.043 \pm 0.004$
$K^*\gamma$	$-0.035 \pm 0.076 \pm 0.012$		$+0.08 \pm 0.13 \pm 0.03$

Table 3. *Search for direct CP violation in hadronic and radiative B decay modes. The $K^*\gamma$ result is an average of K^{*+} and K^{*0}.*

5.4   CP violation in interference

This third class of CP violation requires decays of neutral B mesons into CP eigenstates,

$$|\overline{f}\rangle = \eta_f|f\rangle \quad \text{with } \eta_f = +1 \text{ or } -1. \tag{47}$$

Examples are $f = \pi^+\pi^-$, where $\eta_f = +1$, since the two pions are in a state with $\ell = 0$, $P = +1$, and $f = J/\psi K_S^0$, where $\eta_f = -1$. The notation $J/\psi K_S^0$ is a short-hand: in reality

K_S^0 and $J/\psi K_S^0$ are not CP eigenstates, and what are observed are the decay cascades:

$$
\begin{aligned}
B^0 &\to J/\psi K^0 \quad \text{with} \quad K^0 \to \pi^+\pi^- , \\
\overline{B}^0 &\to J/\psi \overline{K}^0 \quad \text{with} \quad \overline{K}^0 \to \pi^+\pi^- ,
\end{aligned}
\tag{48}
$$

where the reconstructed state $J/\psi(\pi^+\pi^-)$ with $m_{\pi\pi} = m_K$ is a strict CP eigenstate. The CP eigenvalue is -1 because $CP(J/\psi) = +1$, $CP(\pi\pi) = +1$, and $\ell = 1$ between the J/ψ and the $\pi\pi$ state is required by angular momentum conservation with $J(B) = 0$, $J(J/\psi) = 1$, and $J(\pi\pi) = J(K) = 0$.

We now introduce the decay amplitudes

$$
\langle f|T|B^0\rangle = A_f , \qquad \langle f|T|\overline{B}^0\rangle = \overline{A}_f .
\tag{49}
$$

Using the notations of equations (13) and (14), the time-dependent decay rates dN and $d\overline{N}$ of initial flavour states ψ_B and $\psi_{\overline{B}}$ are given by

$$
dN(t) = \left| \frac{e^{-\gamma_H t} + e^{-\gamma_L t}}{2} A_f + \frac{q}{p} \frac{e^{-\gamma_H t} - e^{-\gamma_L t}}{2} \overline{A}_f \right|^2 ,
\tag{50}
$$

$$
d\overline{N}(t) = \left| \frac{e^{-\gamma_H t} + e^{-\gamma_L t}}{2} \overline{A}_f + \frac{p}{q} \frac{e^{-\gamma_H t} - e^{-\gamma_L t}}{2} A_f \right|^2 .
\tag{51}
$$

Using the abbreviated notation:

$$
r_f = \frac{q\overline{A}_f}{pA_f} ,
\tag{52}
$$

and making the approximations $\Gamma_H = \Gamma_L = \Gamma$ and $|q/p| = 1$, a straightforward calculation gives:

$$
\begin{aligned}
dN(\Delta t) &= \frac{|A_f|^2}{2} e^{-\Gamma|\Delta t|} \left[1 + \frac{1 - |r_f|^2}{1 + |r_f|^2} \cos(\Delta m_d \Delta t) - \frac{2\mathrm{Im}\,(r_f)}{1 + |r_f|^2} \sin(\Delta m_d \Delta t) \right] , \\
d\overline{N}(\Delta t) &= \frac{|A_f|^2}{2} e^{-\Gamma|\Delta t|} \left[1 - \frac{1 - |r_f|^2}{1 + |r_f|^2} \cos(\Delta m_d \Delta t) + \frac{2\mathrm{Im}\,(r_f)}{1 + |r_f|^2} \sin(\Delta m_d \Delta t) \right] .
\end{aligned}
\tag{53}
$$

For the case $f = J/\psi K_S^0$, this gives, with a precision of better than 1% (Nir 2001):

$$
r_f = e^{2i\beta},
\tag{54}
$$

and we obtain, with the proper normalisation of the decay amplitude A_f:

$$
\begin{aligned}
dN(\Delta t) &= \frac{\Gamma e^{-\Gamma|\Delta t|}}{2} \left[1 - \sin(2\beta)\sin(\Delta m_d \Delta t) \right] , \\
d\overline{N}(\Delta t) &= \frac{\Gamma e^{-\Gamma|\Delta t|}}{2} \left[1 + \sin(2\beta)\sin(\Delta m_d \Delta t) \right] , \\
a_f(\Delta t) &= \frac{d\overline{N} - dN}{d\overline{N} + dN} = \sin(2\beta)\sin(\Delta m_d \Delta t) .
\end{aligned}
\tag{55}
$$

The integral of a_f over all Δt is zero. Therefore, a time-dependent measurement of the two rates is necessary in order to obtain $\sin 2\beta$. The experimental determination of $\sin 2\beta$ requires eight steps:

1. Reconstruct J/ψ mesons in their e^+e^- and $\mu^+\mu^-$ decays.

2. Reconstruct neutral K mesons in their $\pi^+\pi^-$ decays.

3. Reconstruct neutral B mesons from the J/ψ and K candidates.

4. Partially reconstruct the second B meson of the pair in a flavour-specific way. This "tags" the flavour of the first B at the time of decay of the second B.

5. Reconstruct the z coordinates of the two decay vertices to obtain $\Delta\tilde{t}$ for each event.

6. Determine the fraction w of wrong flavour tags.

7. Determine the vertex resolution function $R(\Delta\tilde{t} - \Delta t)$ between reconstructed and true Δt values

8. Determine $\sin 2\beta$ from a fit to the two time distributions $dN_w(\Delta\tilde{t})$ and $d\overline{N}_w(\Delta\tilde{t})$. (Here $dN_w = (1 - w)\, dN + w\, d\overline{N}$ and similarly for $d\overline{N}_w$).

In the BABAR experiment (Aubert 2001a) with a sample of about 32 Million $B\overline{B}$ pairs, there are 380 reconstructed and tagged $J/\psi K_S^0$ events, including about 15% with a $K \to \pi^0\pi^0$ decay. Steps 1, 2, and 3 have already been discussed in the lifetime section. In addition to $J/\psi K_S^0$, BABAR has used other charmonium CP eigenstates in their $\sin 2\beta$ analysis, $\psi(2S)K_S^0$ and $\chi_{c1}K_S^0$ with $\eta_f = -1$, $J/\psi K_L^0$ with $\eta_f = +1$, as well as $J/\psi K^{0*}$ with $K^{0*} \to K_S^0\pi^0$ which is a mixture of CP eigenstates with $\eta_f = -1$ and $+1$, but with a well-determined effective value of $\eta_f = +0.65 \pm 0.07$ (Aubert 2001e).
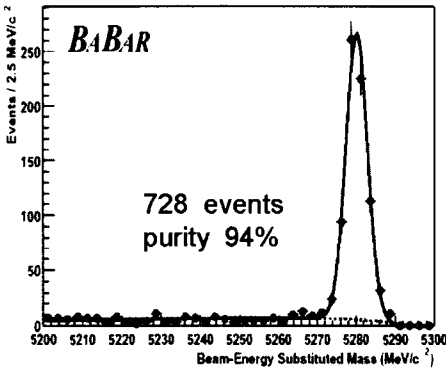
Figure 23. *Mass distribution for all B decays into CP eigenstates with a K_S^0 in the BABAR $\sin 2\beta$ measurement. The dotted curve shows the background.*
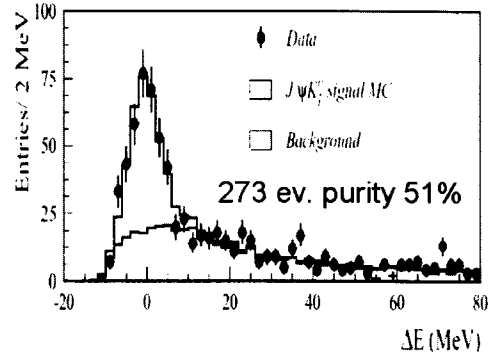
Figure 24. *Distribution of ΔE^*, see text, for the $B \to J/\psi K_L^0$ candidates in the BABAR $\sin 2\beta$ measurement.*

The modes with $\eta_f = +1$ have $r_f = -e^{2i\beta}$, and an asymmetry a_f with the opposite sign. The m_{ES} distribution of all events in modes with a K_S^0 is shown in Figure 23. The signal of the $J/\psi K_L^0$ mode is shown in Figure 24. The EMC and IFR measure the directions of K_L^0 mesons, but not their energies. Using the $m(B^0)$ requirement to constrain the K_L^0 energy, leaves only one kinematic constraint, $\Delta E^* = E^*(J/\psi) + E^*(K_L^0) - m(\Upsilon 4S)/2$,

which is shown in Figure 24. The background in the K_L sample is much higher than in the K_S sample. Background levels are often expressed as purities:

$$P = \frac{N_{\text{signal}}}{N_{\text{signal}} + N_{\text{background}}} \tag{56}$$

The event numbers and purities in Figures 23 and 24 are those before flavour-tagging.

Flavour-tagging is achieved by requiring a second B decay vertex with high momentum electrons, high momentum muons, charged Kaons, or other flavour-specific combinations determined using a neural net algorithm. In order to maximize the tagging rate, tags with only one track are accepted, and their decay coordinate z_2 is given by the intersection of this track with an assumed B flight line. The neural net tags use poorly identified leptons, slow pions, and fast hadrons. They are grouped into two classes NT1 and NT2. Table 4 gives the tagging efficiency ϵ_i, the mistag fraction w_i, and the quality factor Q_i

Tag category	ϵ_i (%)	w_i (%)	Q_i (%)
Leptons	10.9 ± 0.3	8.9 ± 1.3	7.4 ± 0.5
Kaons	35.8 ± 0.5	17.6 ± 1.0	15.0 ± 0.9
NT1	7.8 ± 0.3	22.0 ± 2.1	2.5 ± 0.4
NT2	13.8 ± 0.3	35.1 ± 1.9	1.2 ± 0.3
Sum	68.4 ± 0.7		26.1 ± 1.2

Table 4. *Flavour-tagging in the BABAR experiment.*

for each tag category i. The quality factors are defined as

$$Q_i = \epsilon_i \left(1 - 2w_i\right)^2, \qquad Q = \sum_i Q_i. \tag{57}$$

They determine the statistical error of the fit result:

$$\sigma(\sin 2\beta) \approx \frac{1.9}{\sqrt{N_{\text{sig}} P Q}} \tag{58}$$

where N_{sig} is the number of signal events before tagging. Table 5 gives the number of tagged events and their purity for the BABAR experiment (Aubert 2001a).

Step 5, the determination of Δt with equation (4), is the same as in the lifetime and oscillation measurements. Step 6, the determination of the mistag fractions, uses the same sample of fully reconstructed neutral B mesons with identified flavour as in the oscillation measurement. The 9400 events of this "flavour sample" are tagged with the same four tagging categories as the "CP sample". Since their flavour is known, the tags are easily classified as right or wrong. The results are given in Table 4.

Step 7, the determination of the time resolution, is achieved by simultaneously fitting parametrized resolution functions $R_i(\Delta \tilde{t} - \Delta t)$, which are different for each tagging category i and each decay mode, to both the flavour sample and the CP sample. This is

Mode	N	P (%)
$J/\psi K_S^0(\pi^+\pi^-)$	316	98
$J/\psi K_S^0(\pi^0\pi^0)$	64	94
$\psi(2S)K_S^0$	67	98
$\chi_{c1}K_S^0$	33	97
$J/\psi K^{*0}$	50	74
$J/\psi K_L^0$	273	51
Sum	803	80

Table 5. *Number and purity of tagged events in the BABAR experiment.*
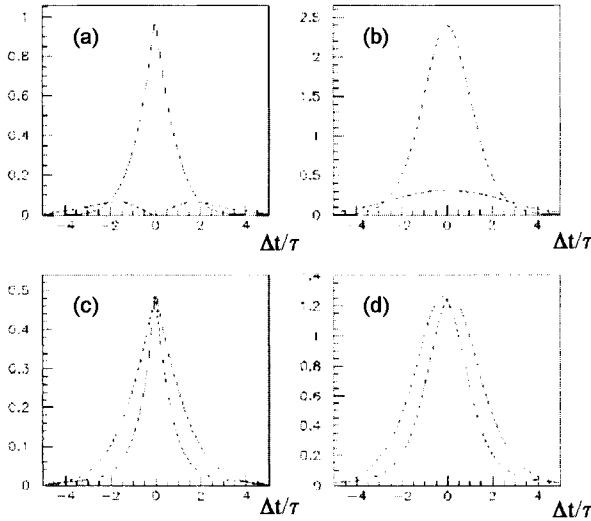
Figure 25. *(a) and (b) show the time dependences for the flavour sample with unmixed (solid curves) and mixed events (dashed curves). (c) and (d) show these curves for the CP sample with B^0 tags (solid) and $\overline{B}^0$ tags (dashed). (a) and (c) are without Δt smearing and without mistags, (b) and (d) with smearing and mistags.*

illustrated in Figure 25. For each i, the observed $\Delta\tilde{t}$ distributions can be described by

$$dN_1(\Delta\tilde{t}) = \int e^{-\Gamma|\Delta t|}[1 + D_i\cos(\Delta m_d\Delta t)]R_i(\Delta\tilde{t} - \Delta t)\,d\Delta t,$$

$$dN_2(\Delta\tilde{t}) = \int e^{-\Gamma|\Delta t|}[1 - D_i\cos(\Delta m_d\Delta t)]R_i(\Delta\tilde{t} - \Delta t)\,d\Delta t, \tag{59}$$

for the "unmixed" flavour sample events dN_1 with the opposite-flavour tag, and the "mixed" events dN_2 with the same-flavour tag, and by

$$dN_3(\Delta\tilde{t}) = \int e^{-\Gamma|\Delta t|}[1 + D_i\sin 2\beta\sin(\Delta m_d\Delta t)]R_i(\Delta\tilde{t} - \Delta t)\,d\Delta t,$$

$$dN_4(\Delta\tilde{t}) = \int e^{-\Gamma|\Delta t|}[1 - D_i\sin 2\beta\sin(\Delta m_d\Delta t)]R_i(\Delta\tilde{t} - \Delta t)\,d\Delta t, \tag{60}$$
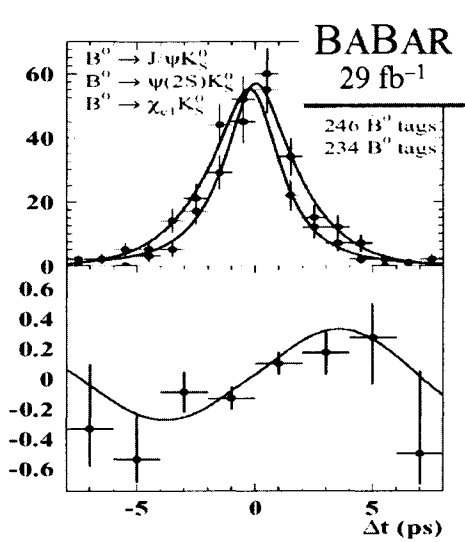
Figure 26. *Rates and asymmetry for the K_S^0 modes in the BABAR measurement*

for the CP sample events with a B^0 and a $\overline{B}^0$ tag, dN_3 and dN_4, respectively. The "dilution" factors are $D_i = 1 - 2w_i$. Since the resolution functions R_i are the same for all four distributions, the fits to the high statistics samples 1 and 2, fix their shapes in the low statistics samples 3 and 4.

This combined fit in step 7 is in practice performed together with steps 6 and 8, which determine the mistag fractions w_i, and the value of $\sin 2\beta$. The results are

$$
\begin{aligned}
\sin 2\beta &= 0.56 \pm 0.15 && \text{for the } K_S \text{ modes,} \\
&= 0.70 \pm 0.34 && \text{for } J/\psi K_L, \\
&= 0.82 \pm 1.00 && \text{for } J/\psi K^{0*},
\end{aligned}
$$

where the errors are statistical only. A combination of these results leads to

$$\sin 2\beta = 0.59 \pm 0.14 \pm 0.05 . \tag{61}$$

The systematic error is dominated by the knowledge of the resolution function for K_S modes, and by the knowledge of the CP properties of the background for the K_L mode. The rates and the asymmetry a_f for the K_S modes are shown in Figure 26. If the condition $|r_f| = 1$ is not imposed in the fit, the data can be fitted to the expressions in equations (53). This fit, using all modes, gives:

$$|r_f| = 0.93 \pm 0.09 \pm 0.03 . \tag{62}$$

Hence there is no indication for direct CP violation in the decay modes used for the $\sin 2\beta$ measurement.

In the Belle experiment (Abe 2001), 31M $B\overline{B}$ events are analysed in a very similar way. The same CP eigenmodes are reconstructed as in BABAR, plus an additional mode, $\eta_c K_S$,
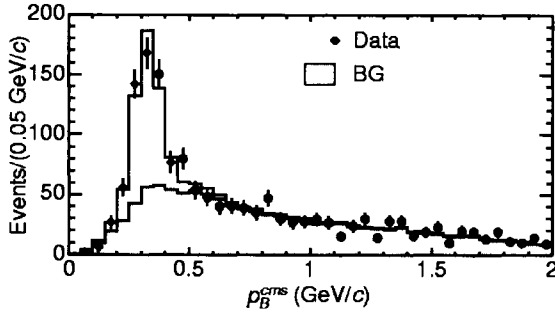
Figure 27. *Signal and background in the $J/\psi K_L^0$ sample of Belle's $\sin 2\beta$ measurement.*
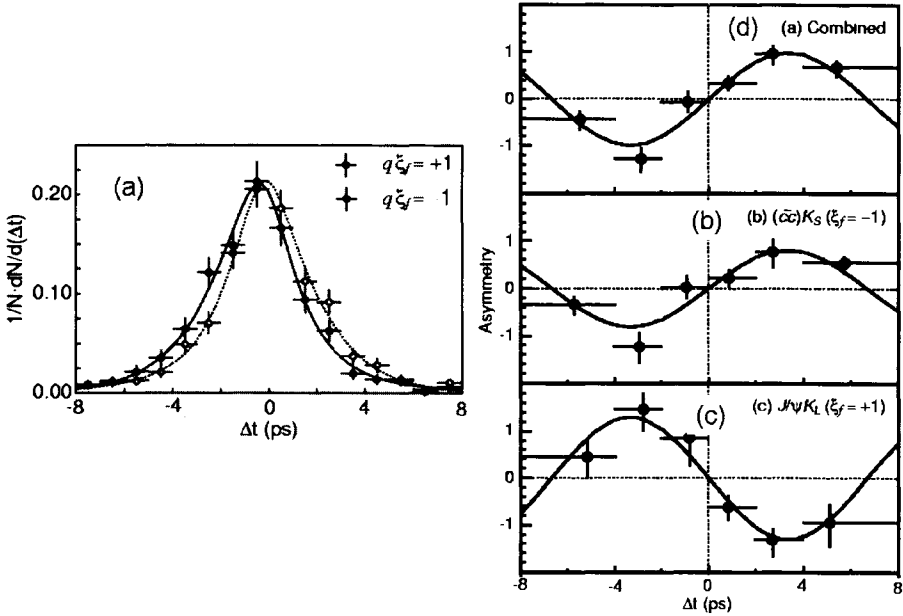
Figure 28. *The $\sin 2\beta$ result of Belle. (a) Rates for all tagged events; B^0 tags for $\eta_f = -1$ modes are combined with $\overline{B}^0$ tags for $\eta_f = +1$ and vice versa. (b) Asymmetry a_f for modes with $\eta_f = -1$, (c) with $\eta_f = +1$, (d) combined asymmetry.*

which has $\eta_f = -1$. Figure 27 shows Belle's signal for the mode $J/\psi K_L^0$, whose purity of 61% is slightly better than BABAR. Another difference is in the tagging strategy, where Belle tries to tag each event, and uses all 1137 reconstructed events of their CP sample in the final fit. Their "tagging efficiency" is $\epsilon = 1$ compared to $\epsilon = 0.68$ in BABAR, but their overall tagging quality is $Q = 0.270 \pm 0.012$, which is very similar to $Q = 0.261 \pm 0.012$ in BABAR. Belles's final fit, combined all modes, is shown in Figure 28. Their result is

$$\sin 2\phi_1 = 0.99 \pm 0.14 \pm 0.06 , \qquad (63)$$

where ϕ_1 is the Japanese name for β.
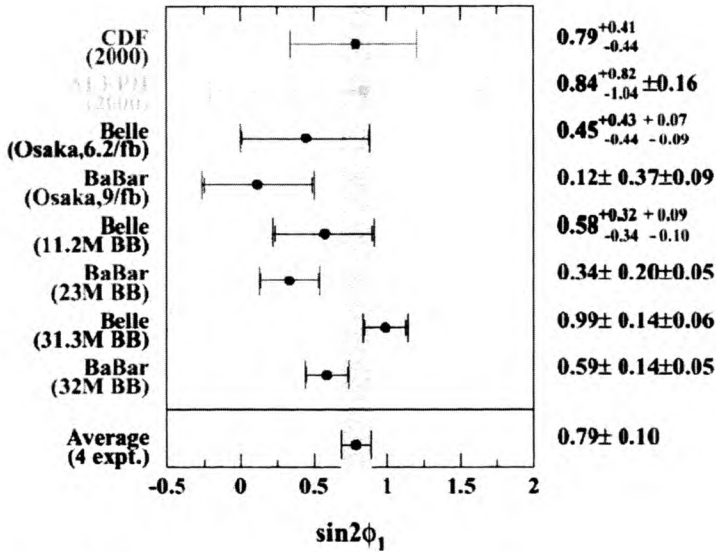
Figure 29. *Compilation and average of* sin 2β *measurements.*
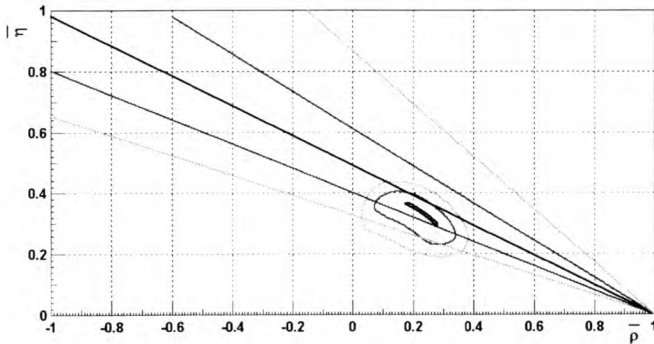
Figure 30. *Fit result from Figure 22 and lines of constant likelihood (max, ±1σ, ±2σ) for* sin 2β.

Figure 31. *Final fit to all observations including* sin 2β.

Figure 29 compiles all results for sin 2β. The world average is

$$\sin 2\beta = 0.79 \pm 0.10 \,, \tag{64}$$

which is in very good agreement with the ρ/η region favoured by the measurements of $|V_{ub}|$, $|V_{td}|$, and $\epsilon(K^0)$, as shown in Figure 30. A common fit to $|V_{ub}|$, $|V_{td}|$, $\epsilon(K^0)$, and sin 2β gives the region in Figure 31. With one-standard-deviation errors, this region has

$$
\begin{aligned}
\bar{\rho} &= 0.22 \pm 0.12 \,, \\
\bar{\eta} &= 0.34 \pm 0.06 \,, \\
J &= (2.8 \pm 0.6) \times 10^{-5} \,.
\end{aligned}
\tag{65}
$$
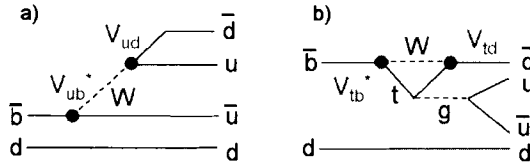
Figure 32. *Tree (a) and penguin (b) diagram for π^+ and π^- decays of B mesons.*

BABAR has also performed a first measurement (Aubert 2001g) of the angle α in the Unitarity Triangle (see Figure 19 for its definition). $\sin 2\alpha$ is obtained from the time dependence of tagged $\pi^+\pi^-$ decays. If only the tree diagram of Figure 32a contributes to these decays, we would have in complete analogy to Equations 54 and 55

$$
\begin{aligned}
r_{\pi^+\pi^-} &= e^{2i\alpha}, \\
|r_{\pi^+\pi^-}| &= 1, \\
a_{\pi^+\pi^-} &= \sin 2\alpha \sin(\Delta m_d \Delta t),
\end{aligned}
\tag{66}
$$

where $a_{\pi^+\pi^-}$ is the asymmetry between $d\overline{N}$, the number of $\pi^+\pi^-$ decays with B^0 tags, and dN, the number with $\overline{B}^0$ tags. However, there is a contribution from the penguin diagram in Figure 32b, which has a different weak and a different strong phase. This leads to direct CP violation, $|r_{\pi^+\pi^-}| \neq 1$, in addition to CP violation in the interference between oscillations and decays. This means Im $(r_{\pi^+\pi^-}) \neq \sin 2\alpha$, and instead of equations (55), we have to use equations (53) for the expected time dependence, leading to the asymmetry:

$$
\begin{aligned}
a_{\pi^+\pi^-}(\Delta t) &= -C_{\pi^+\pi^-}\cos(\Delta m_d \Delta t) + S_{\pi^+\pi^-}\sin(\Delta m_d \Delta t), \\
C_{\pi^+\pi^-} &= \frac{1 - |r_{\pi^+\pi^-}|^2}{1 + |r_{\pi^+\pi^-}|^2}, \\
S_{\pi^+\pi^-} &= \frac{2\text{Im}\,(r_{\pi^+\pi^-})}{1 + |r_{\pi^+\pi^-}|^2} = \sin(2\alpha_{\text{eff}}).
\end{aligned}
\tag{67}
$$

Using the same 32 M $B\overline{B}$ events as in their $\sin 2\beta$ analysis, BABAR finds the preliminary result

$$
\begin{aligned}
C_{\pi^+\pi^-} &= -0.25^{+0.45}_{-0.47} \pm 0.14, \\
S_{\pi^+\pi^-} &= 0.03^{+0.53}_{-0.56} \pm 0.11.
\end{aligned}
\tag{68}
$$

There have been theoretical attempts (Beneke 2001, Luo 2001) to relate $\sin(2\alpha_{\text{eff}})$ to $\sin(2\alpha)$, using additional measurements of $K\pi$ and $\pi\pi$ decay rates of neutral B mesons. The central value of the BABAR result, $\sin(2\alpha_{\text{eff}}) = 0.03$, gives $\alpha_{\text{eff}} = 1°$ or $89°$, which they translate into α close to zero or equal to $75°$. The first value does not agree with the preferred ρ/η region from all our previous discussions, but the second one agrees very well. Using the one-standard-deviation errors of the BABAR result in Equation 68 and the α_{eff} to α translation curve in (Luo 2001):

$$
\alpha = (75 \pm 18)°.
\tag{69}
$$

This result is drawn with a central and two one-standard-deviation circles in Figure 33. There are no 2σ contours because the BABAR result still allows any value of α within two standard deviations.
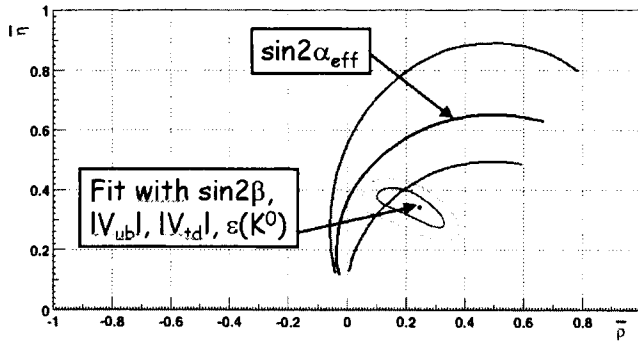
Figure 33. *Contours of constant likelihood (max, $\pm 1\sigma$), from the BABAR* $\sin 2\alpha_{\text{eff}}$ *measurement.*

6 Conclusion

CP violation in the B meson sector has been firmly established in 2001 by two experiments, BABAR at SLAC and Belle at KEK. Their observation of $a_f \neq 0$ in charmonium final states, has a combined significance of more than seven standard deviations. It is compatible with the Standard Model explanation of CP violation in the Higgs mechanism. It is also compatible with the CKM matrix parameter values of ρ and η as determined from measurements of $|V_{ub}|$, $|V_{td}|$, and $\epsilon(K)$.

With the combined efforts of theorists and experimentalists, the next years will show how precisely the measurements will determine ρ and η, and if the five independent constraints on ρ and η from $|V_{ub}|$, $|V_{td}|$, $\epsilon(K)$, β, and α will establish a consistent description, or will falsify the assumption that all weak processes including CP violation are well described by the Standard Model.

Acknowledgments

References

Abashian A et al (Belle), 2001, "The Belle Detector", submitted to *Nucl Instr Meth A*, KEK Progress Report 2000-4
Abbiendi G et al (OPAL), 1999, *Eur Phys J C*, **12** 609
Abe K et al (Belle), 2001, *Phys Rev Lett* **87** 091802
Ackerstaff K et al (OPAL), 1998, *Eur Phys J C*, **5** 379
Albrecht H et al (ARGUS), 1987, *Phys Lett B* **192** 245
Andrews D et al (CLEO), 1980, *Phys Rev Lett* **45** 219
Aubert B et al (BABAR), 2001a, *Phys Rev Lett* **87** 091801
Aubert B et al (BABAR), 2001b, *Phys Rev Lett* **87** 201803

Aubert B et al (BABAR), 2001c, submitted to *Phys Rev Lett*, hep-ex/0112044

Aubert B et al (BABAR), 2001d, submitted to *Phys Rev Lett*, hep-ex/0112045

Aubert B et al (BABAR), 2001e, *Phys Rev Lett* **87** 241801

Aubert B et al (BABAR), 2001f, submitted to *Nucl Instr Meth*, hep-ex/0105044

Aubert B et al (BABAR), 2001g, submitted to *Phys Rev Lett*, hep-ex/0110062

Barate R et al (ALEPH), 2001, *Eur Phys J C* **20** 431

Bebek C et al (CLEO), 1981, *Phys Rev Lett* **46** 84

Beneke M, Buchalla G, Neubert M and Sachrajda C T, 2001, *Nucl Phys B* **606** 245

Berger Ch et al (PLUTO), 1978, *Phys Lett* **76** 243

Bienlein J K et al (DHHM), 1978, *Phys Lett* **78** 360

Bjorken J D, 1987, presented at the Workshop on Experiments for the Supercollider, Berkeley

Branco G C, Lavoura L and Silva J P, 1999, *"CP Violation"*, Oxford University Press

Brandt T (BABAR), 2001, presented at 9th Int Symp on Heavy Flavour Physics, Pasadena

Buchalla G, 2001, *these Proceedings*

Cabibbo N, 1963, *Phys Rev Lett* **10** 531

Christenson J H, Cronin J W, Fitch V L and Turlay R, 1964, *Phys Rev Lett* **13** 138

Darden C W et al (DASP-2), 1978a, *Phys Lett B* **76** 246

Darden C W et al (DASP-2), 1978b, *Phys Lett B* **78** 364

Eichler R, Nakada T, Schubert K R, Weseler S and Wille K, 1986, *SIN-PR-86-13*

Ellis J R, Gaillard M K, Nanopoulos D V and Rudaz S, 1977, *Nucl Phys B* **131** 285

Finocchiaro G et al (CUSB), 1980, *Phys Rev Lett* **45** 222

Hamel de Montchanault G , 2001, *Proc Europhys Conf High Energy Phys, Budapest*

Herb S W et al, 1977, *Phys Rev Lett* **39** 252

Hitlin D et al (SLAC, LBL, Caltech), 1989, *SLAC-352*

Jarlskog C, 1985, *Phys Rev Lett* **55** 1039

Hitlin D et al (SLAC, LBL, Caltech), 1989, *SLAC-352*

Jaffe D E et al (CLEO), 2001, *Phys Rev Lett* **86** 5000

KEK Laboratory, 1990, Task Force Reports *KEK-90-23* and *KEK-90-24*

Kleinknecht K, 2001, *these Proceedings*

Kobayashi M and Maskawa T, 1973, *Progr Theor Phys* **49** 652

Luo Z and Rosner J L, submitted to *Phys Rev D*, hep-ph/0108024

Nir J, 2001, *these Proceedings*

PDG (Particle Data Group), 2000, *Eur Phys J C* **15** 1

Rosner J L, 2001, *these Proceedings*

Spencer L J et al (CUSB), 1981, *Phys Rev Lett* **47** 771

Stark J (BABAR), 2001, *Ph D Thesis*, Université Paris VI, in French

Wacker K et al (PSI), 1988, *PSI-PR-88-09*

Wolfenstein L, 1985, *Phys Rev Lett* **51** 1945

B physics at the LHC

Tatsuya Nakada

CERN and University of Lausanne, Switzerland

1 Introduction

Since its discovery, the b quark has brought us two big surprises. The first was the unexpectedly large lifetime. The second was the mass difference between the two mass eigenstates of the B_d meson system which is about 100 times larger than the mass difference in the neutral K meson system. From the first observation, we learned that the mixing between the second and third families is much smaller than between the first and second families. (The mixing between the first and third families is even more suppressed.) The second observation taught us that the mass of the top quark is much larger than had been anticipated at that time.

The main goal of B physics is to study the structure of quark mixing and its role in CP violation. In the Standard Model CP violation can be very naturally accommodated through the complex quark mixing matrix (Kobayashi and Maskawa 1972) defined by four parameters (Nir 2001). All current observations of CP violating phenomena in particle physics are in full agreement with the Standard Model calculations. However, there are still some reasons to speculate about CP violation generated by physics beyond the Standard Model. Firstly, the Standard Model alone cannot account for the large asymmetry between matter and antimatter observed in our universe (Shaposhnikov 1986). Secondly, various extensions to the Standard Model introduce new sources of CP violation. Since CP violation is expected in many B meson decay modes, and the Standard Model can make precise predictions for some of these decay modes, the B meson system appears to be a very attractive place to look for evidence for physics beyond the Standard Model.

In the presence of new physics, some assumptions made to extract the Cabibbo-Kobayashi-Maskawa (CKM) parameters are no longer valid. Indeed, the consistency of the CKM picture with current observations could be accidentally due to numerical cancellation between various effects from new physics. Therefore, it is essential to develop a strategy which allows the new physics and the Standard Model contributions to be clearly disentangled. Then the CKM parameters can be determined in a model-independent way. This is important since we hope that physics at higher energy scale will one day be able to explain the family structure, by deriving the CKM parameters.

In this article, we discuss how B meson decays can be explored at the LHC in order to obtain a better understanding of the CKM picture, and to look for physics beyond the Standard Model.

2   The CKM picture

2.1   The CKM matrix

In the Standard Model, CP violation is naturally introduced by a 3×3 complex quark mixing matrix, V_{CKM}, expressed as

$$V_{\text{CKM}} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}.$$

The charged current of the weak interaction is then proportional to

$$\overline{U}_L^i \left(1 - \gamma_5\right) \gamma_\mu V_{ij} D_L^j W^\mu$$

where U_L and D_L are the left-handed quark operators for the charge 2/3 up-type and the charge $-1/3$ down-type quarks respectively:

$$U_L = \begin{pmatrix} u_L \\ c_L \\ t_L \end{pmatrix}, \qquad D_L = \begin{pmatrix} d_L \\ s_L \\ b_L \end{pmatrix}.$$

The matrix V_{CKM} is unitary, $V^\dagger V = 1$. One of the unitarity relations is:

$$V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0. \tag{1}$$

This is illustrated in the complex plane in Figure 1(a). The unitarity condition can be illustrated more easily by translating $V_{cd}V_{cb}^*$ and $V_{td}V_{tb}^*$ to form a closed triangle, as shown in Figure 1(b). The three angles of the triangle, α, β and γ (also known as ϕ_2, ϕ_1 and ϕ_3 respectively) can be defined as:

$$\alpha = \tan^{-1}\left(-\frac{V_{td}V_{tb}^*}{V_{ud}V_{ub}^*}\right), \qquad \beta = \pi - \tan^{-1}\left(\frac{V_{td}V_{tb}^*}{V_{cd}V_{cb}^*}\right), \qquad \gamma = \tan^{-1}\left(-\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*}\right). \tag{2}$$
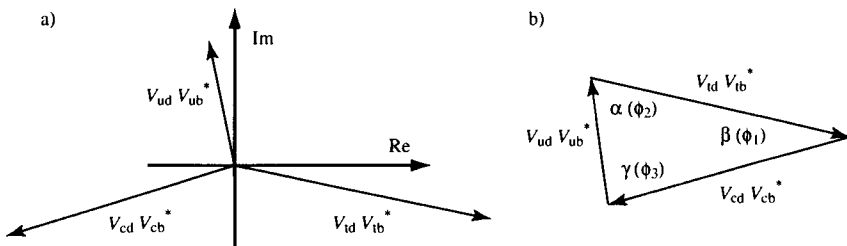
Figure 1.   (a) One CKM matrix unitarity condition: $V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0$. (b) Elements translated to form a triangle drawn in the complex plane.

It should be noted that a redefinition of the quark phases results in a rotation of the triangle in which the three angles, α, β and γ remain invariant.

Violation of the unitarity condition given by Equation (1) is expressed in a graphical form in Figure 2(a) and (b), where the three sides given by $V_{ud}V_{ub}^*$, $V_{cd}V_{cb}^*$ and $V_{td}V_{tb}^*$ do not form a closed triangle. Note that α, β and γ defined by Equation (2) still satisfy:

$$\alpha + \beta + \gamma = \pi.$$

If one forms the closed triangle from the length of the three sides, $|V_{ud}V_{ub}^*|$, $|V_{cd}V_{cb}^*|$ and $|V_{td}V_{tb}^*|$, the three angles of this triangle, α', β' and γ' defined in Figure 2(c), are not identical to α, β and γ. Therefore, a test of unitarity can be made by comparing the angles defined by the length of the three sides, and the angles measured by CP violating effects (as explained later).
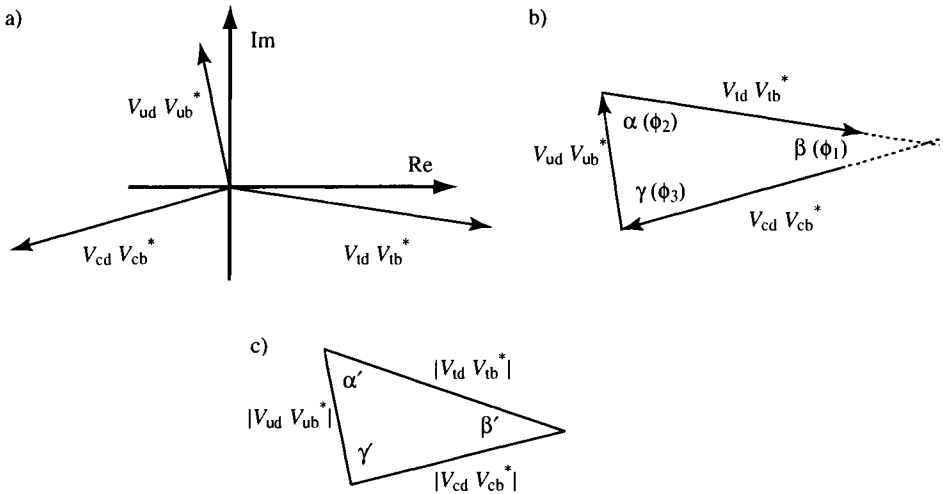
Figure 2. *(a) and (b), are similar to Figure 1, but unitarity is violated. In (c) a closed triangle is formed using the three sides.*

A unitary 3×3 matrix can be parameterized by four parameters. One possible choice (Particle Data Group 2000) is to use three angles, θ_{12}, θ_{23}, θ_{13} and one complex phase, δ. The standard parameterization for the CKM matrix is then given by:

$$V_{\text{CKM}} = R_{23} \times R_{13} \times R_{12}$$

where

$$R_{12} = \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad R_{23} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \quad R_{13} = \begin{pmatrix} c_{13} & 0 & s_{13}e^{-i\delta} \\ 0 & 1 & 0 \\ -s_{13}e^{i\delta} & 0 & c_{13} \end{pmatrix}$$

with $s_{ij} = \sin\theta_{ij}$ and $c_{ij} = \cos\theta_{ij}$.

A parameterization reflecting the observed pattern of the CKM matrix was proposed by Wolfenstein (Wolfenstein 1983). This introduces the parameters:

$$\lambda = s_{12}, \qquad A = \frac{s_{23}}{s_{12}^2}, \qquad \rho = \frac{s_{13}\cos\delta}{s_{12}s_{23}}, \qquad \eta = \frac{s_{13}\sin\delta}{s_{12}s_{23}},$$

and expands the matrix elements in powers of λ. Neglecting terms proportional to λ^n where $n > 5$, this gives:

$$V_{\text{CKM}} \approx \begin{pmatrix} 1 - \lambda^2/2 & \lambda & A\lambda^3\,(\rho - i\,\eta) \\ -\lambda - i\,A^2\lambda^5\eta & 1 - \lambda^2/2 & A\lambda^2 \\ A\lambda^3\,(1 - \tilde\rho - i\,\tilde\eta) & -A\lambda^2 - i\,A\lambda^4\eta & 1 \end{pmatrix} \tag{3}$$

where $\tilde\rho$ and $\tilde\eta$ are given by $\tilde\rho = \rho(1 - \lambda^2/2)$ and $\tilde\eta = \eta(1 - \lambda^2/2)$. The parameter λ is known from light hadron decays to be 0.221 ± 0.002 (Particle Data Group, 2000). As seen from Equation 3, the first 2×2 sub-matrix is almost unitary:

$$V_{ud}V_{cd}^* + V_{us}V_{cs}^* = iA^2\lambda^5\eta \approx 0 \qquad V_{ud}V_{us}^* + V_{cd}V_{cs}^* = -iA^2\lambda^5\eta \approx 0$$

and

$$|V_{ud}|^2 + |V_{us}|^2 = 1 - \lambda^4/4 \approx 1 \qquad |V_{cd}|^2 + |V_{cs}|^2 = 1 - \lambda^4/4 \approx 1$$

With the parameterization given in Equation 3, the imaginary part of V_{cd} becomes negligible in the unitarity relation given by Equation (1), and the phases of the matrix elements are:

$$\arg V_{td} = -\beta, \qquad \arg V_{ub} = -\gamma,$$

$$\arg V_{ud} = \arg V_{cb} = \arg V_{tb} = 0, \qquad \arg V_{cd} = \pi.$$

V_{us} and V_{cs} are also real, but the imaginary part of V_{ts} cannot be completely ignored:

$$\arg V_{ts} = \delta\gamma + \pi$$

The angles β, γ and $\delta\gamma$ are functions of ρ, η and λ:

$$\beta = \tan^{-1}\frac{\eta}{1 - \rho}, \qquad \gamma = \tan^{-1}\frac{\eta}{\rho}, \qquad \delta\gamma = \tan^{-1}\lambda^2\eta. \tag{4}$$

2.2   Extraction of the CKM parameters

From a measurement of $|V_{cb}|$, the Wolfenstein parameter A can be determined. The matrix element $|V_{cb}|$ is extracted from semileptonic (and hadronic) decays of B mesons into charmed meson final states (Stone 2001). In exclusive B meson decays, a description based on the quark tree level process $b \to c + W^-$ is obscured by soft hadronic interactions. This has to be taken into account in order to extract $|V_{cb}|$ from the data. Significant improvements in understanding these hadronic effects have been made both in theory and experiment. Further progress will come using the high statistics data samples from the BABAR and BELLE experiments (Schubert 2001).

The semileptonic decays of B mesons to final states containing only the light u and d quarks are used for the extraction of $|V_{ub}|$ (Stone 2001). They are generated by the

tree level process $b \to u + W^-$. Unlike the determination of $|V_{cb}|$, hadronic decays cannot be used since sizeable contributions from the penguin processes, $b \to s$ and $b \to d$, are present in the decay amplitudes. It is much more difficult to evaluate the effects of strong interactions for $|V_{ub}|$ than for $|V_{cb}|$, and the error on the current value of $|V_{ub}|$ is totally dominated by the theoretical uncertainties. This will remain the case for some time.

While $|V_{cb}|$ and $|V_{ub}|$ are determined from decays with tree processes, $|V_{td}|$ can be accessed indirectly through loop processes such as B^0–$\overline{B}^0$ oscillations. The oscillation in the Standard Model is described by the well known box diagrams shown in Figure 3. Due to the large top quark mass and the structure of the CKM matrix, only the top quark
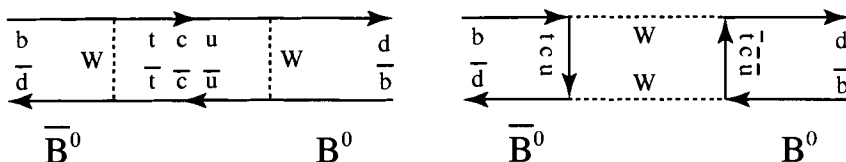
Figure 3. *Box diagrams describing B^0–$\overline{B}^0$ oscillation.*

contribution needs to be considered in the loop. Neglecting the absorptive part of the box diagrams, the oscillation amplitude for $B^0 \to \overline{B}^0$ is calculated to be (Hagelin 1981):

$$H_{21} = -\frac{G_F^2 f_{B_d}^2 B_{B_d} m_{B_d} m_W^2}{12\pi^2} \eta_{B_d} S(x_t)(V_{td}V_{tb}^*)^2 \tag{5}$$

where G_F, m_W and m_{B_d} are the Fermi coupling constant, W boson mass and B_d mass. The function $S(x_t)$ is determined from the mass ratio, $x_t = (m_t/m_W)^2$, where m_t is the top quark mass, and the QCD correction factor, η_{B_d}, can be calculated reliably with perturbation methods. The B meson decay constant, f_{B_d}, and the bag parameter, B_{B_d}, have never been measured directly. The decay constant f_{B_d} is given by the transition matrix element between B^0 and the hadronic vacuum state. Thus it could be experimentally obtained from the branching fraction for the leptonic decay, $B^\pm \to \tau^\pm \nu_\tau$ shown in Figure 4, assuming that $|V_{ub}|$ is known. The decay constants for $B^\pm$ and B^0 are expected to be very
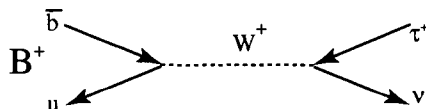
Figure 4. *Diagram for the $B^+ \to \tau^+ \nu_\tau$ decay.*

similar. The bag parameter takes into account the difference between the hadronic vacuum and the actual hadronic states virtually present in the B^0–$\overline{B}^0$ oscillation processes. Theoretical estimates exist with large uncertainties due to the difficulties in evaluating the effects of the non-perturbative soft hadronic interactions. The most promising theoretical approach to obtain f_{B_d} and B_{B_d} is the QCD lattice calculation discussed in these proceedings (Davies 2001).

The absorptive part in the B^0–$\overline{B}^0$ oscillation is due to c and u quarks. It is calculated to be very small compared to the dispersive part and can indeed be ignored. With this

approximation, the mass difference between the two mass eigenstates, B_d^H and B_d^L, can be derived,

$$\Delta m_d \equiv m_{B_d}^H - m_{B_d}^L = 2|H_{21}|,$$

and the decay width difference between them is negligible:

$$\Delta\Gamma_d \equiv \Gamma_{B_d}^L - \Gamma_{B_d}^H \approx 0.$$

There is no CP violation in the B^0-$\overline{B}^0$ oscillations in this approximation. Note that the phase of the oscillation amplitude is given by the phase $V_{td}^2 = -2\beta$.

The mass difference Δm_d is experimentally measured as the frequency of B^0-$\overline{B}^0$ oscillations. From the known structure of the CKM matrix, we have $|V_{tb}| = 1$ with very high accuracy. Therefore, $|V_{td}|$ can be obtained from Δm_d. The error on $|V_{td}|$ is totally dominated by the theoretical uncertainties on $f_B\sqrt{B_d}$.

From the measurements of V_{ub}, V_{cb} and V_{td}, the quantities:

$$\sqrt{\tilde{\rho}^2 + \tilde{\eta}^2} = \frac{1}{\lambda}\frac{|V_{ub}|}{|V_{cb}|}\left(1 - \frac{\lambda^2}{2}\right) \tag{6}$$

and

$$\sqrt{(1-\tilde{\rho})^2 + \tilde{\eta}^2} = \frac{1}{\lambda}\frac{|V_{td}|}{|V_{cb}|} \tag{7}$$

can be drawn as circles around $(0, 0)$ and $(1, 0)$ in the $\tilde{\rho}$-$\tilde{\eta}$ plane respectively. There are two possible solutions where these circles intersect, but the one with $\tilde{\eta} > 0$ is favoured from the analysis of CP violation in the K^0-$\overline{K}^0$ oscillations.

The mass difference in the B_s system, Δm_s, can be evaluated by calculating the B_s^0-$\overline{B}_s^0$ oscillation amplitude using box diagrams in a similar way to the B_d system. A measurement of Δm_s determines $|V_{ts}V_{tb}^*|^2$, which allows:

$$\sqrt{(1-\tilde{\rho})^2 + \tilde{\eta}^2} = \frac{1}{\lambda}\frac{|V_{td}|}{|V_{ts}|} \tag{8}$$

to be used instead of Equation (7). In the ratio:

$$\frac{1}{\lambda}\frac{|V_{td}|}{|V_{ts}|} = \frac{1}{\lambda}\frac{\sqrt{\Delta m_d}}{\sqrt{\Delta m_d}} \times \frac{\sqrt{m_{B_s}\eta_{B_s}}}{\sqrt{m_{B_d}\eta_{B_d}}} \times \frac{\sqrt{B_{B_s}}f_{B_s}}{\sqrt{B_{B_d}}f_{B_d}},$$

the assumption $|V_{tb}| = 1$ is no longer necessary. While the theoretical uncertainty in $f_B\sqrt{B}$ is quite considerable, the ratio between these quantities for the B_s and B_d mesons is theoretically much better understood. Therefore, Equation (8) will have a significantly smaller error than Equation (7). Unfortunately, only an experimental lower limit is available for Δm_s at the moment.

We now consider CP violation for decay final states which can be produced by both B^0 and $\overline{B}^0$. These can be CP eigenstates such as $J/\psi K_S$ or non-CP eigenstates such as $D^{*+}\pi^-$ and its CP-conjugate state $D^{*-}\pi^+$. If CP violation is present neither in the oscillation nor the decay amplitudes, it is well known that the only signature of CP violation that can appear is through phases:

$$\sin(\phi_0 + \phi_f)\sin(\Delta m_d t)$$

in the time-dependent decay rates of initial B^0 and $\overline{B}^0$, where ϕ_0 is the phase of the B^0–$\overline{B}^0$ oscillation amplitude, and ϕ_f is the phase of the ratio of the instantaneous decay amplitudes of B^0 and $\overline{B}^0$ into the particular final state.

For the $J/\psi K_S$ final state, the $b \to c + W^-$ tree process dominates, but there can also be a $b \to s$ penguin process (Figure 5). The phase of the tree process is given by V_{cb}, and is 0 in the Wolfenstein approximation. The penguin process is described by a virtual top quark loop with CKM matrix elements, $V_{tb}V_{ts}^*$, which has only a small phase $-\delta\gamma$. Therefore, $\phi_f = 0$, with a theoretical uncertainty less than a few percent.
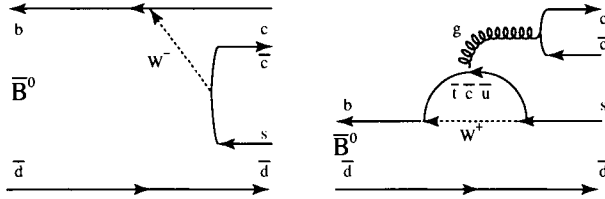
Figure 5. *Quark diagrams for the $\overline{B}^0 \to J/\psi K_S$ decay.*

The final state $D^{*+}\pi^-$ can be generated by the tree process $b \to c + W^-$, followed by $W^- \to \overline{u}d$ for the $\overline{B}^0$ decay or the doubly Cabibbo-suppressed tree process $\overline{b} \to \overline{u} + W^+$ followed by $W^+ \to c\overline{d}$ for the B^0 decay, as seen from Figure 6. From the relevant CKM matrix elements, it follows that $\phi_f = \gamma$. Note that only the tree processes contribute.
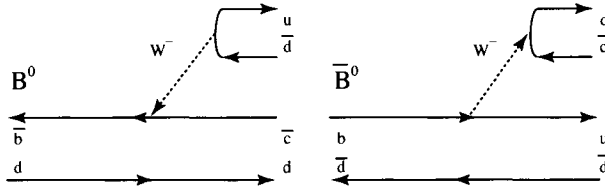
Figure 6. *Quark diagrams for the B^0 and $\overline{B}^0 \to D^{*+}\pi^-$ decays.*

The phase of the oscillation amplitude is given by $\phi_0 = 2\beta$ as discussed before. Therefore the CP violation signatures in B^0 and $\overline{B}^0$ decays are

$$\sin(2\beta)\sin(\Delta m_d t) \qquad \text{for} \quad J/\psi K_S$$

and

$$\sin(2\beta + \gamma)\sin(\Delta m_d t) \qquad \text{for} \quad D^{*+}\pi^- \text{ and } D^{*-}\pi^+.$$

For the $\sin(2\beta)$ measurement from $J/\psi K_S$, the average of the recent results from BABAR and BELLE gives (Schubert 2001):

$$\sin 2\beta = 0.79 \pm 0.17$$

where the error is scaled following the recipe of the Particle Data Group. It is unlikely that any definitive measurement of $\sin(2\beta + \gamma)$, using $B \to D^{(*\pm)}\pi^{(\mp)}$, will be done by the current generation of experiments.

Analogous decays for the B_s^0 meson system are $J/\psi\,\phi$ and $J/\psi\,\eta$, which have $\phi_f = 0$, and $D_s^+ K^-$ which has $\phi_f = \gamma$. Since the B_s^0–$\overline{B}_s^0$ oscillation phase is $-2\delta\gamma$ as discussed before, the CP violation signatures would be:

$$\sin(-2\delta\gamma)\sin(\Delta m_s t) \quad \text{for} \quad J/\psi\,\phi \ (\text{or}\ J/\psi\,\eta)$$

and

$$\sin(-2\delta\gamma + \gamma)\sin(\Delta m_s t) \quad \text{for} \quad D_s^+ K^- \text{ and } D_s^- K^+.$$

No definitive measurements on these quantities will be available from the current generation of experiments.

In Figure 7, we summarise the region of $\tilde{\rho}$ and $\tilde{\eta}$ allowed by the current data on $|V_{ub}|/(\lambda|V_{cb}|)$, Δm_d, and $\sin(2\beta)$, as well as the one-sided constraint on $\Delta m_d/\Delta m_s$ from the lower limit on Δm_s. All the data are interpreted within the framework of the Standard Model, so the constraints may be modified by new physics as discussed earlier. There is a common overlapping region indicating that the data are consistent with the CKM picture, and no sign of new physics is visible within the errors. Note that the $\sin(2\beta)$ measurement gives two allowed regions in the $\tilde{\rho}$–$\tilde{\eta}$ plane defined in Figure 7, and only the one consistent with the other constraints is drawn.
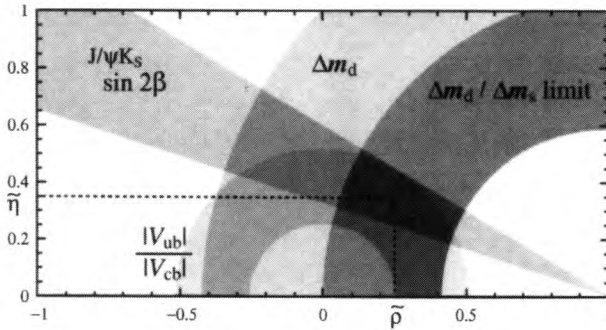
Figure 7. *Allowed region in the $\tilde{\rho}$-$\tilde{\eta}$ plane from $|V_{ub}|/|V_{cb}|$, Δm_d, and $\sin(2\beta)$ measurements, and from the upper limit on Δm_s.*

3   Possible CKM picture in 2006

Let us now speculate about the CKM landscape in 2006. Due to progress in theoretical understanding of the hadronic effect, strongly helped by the large statistics and high quality data from BABAR and BELLE, $|V_{ub}/V_{cb}|$ might become known with a relative error of $\pm 10\%$. Once Δm_s is measured by CDF, $|V_{td}|/(\lambda|V_{ts}|)$ will become known with a relative error of $\pm 7\%$. From CP violation in B^0 and $\overline{B}^0$ decays to $J/\psi\,K_S$, the combined results from BABAR, BELLE, CDF and D0 might yield a result of $\sin(2\beta)$ with an error as small as 0.03, which is still statistically limited. Estimating the progress in other CP violation channels is more difficult since they have various theoretical uncertainties, and future improvements are difficult to evaluate. Assuming that the currently preferred values of ρ and η remain unchanged, the situation expected in 2006 is illustrated in Figure 8.
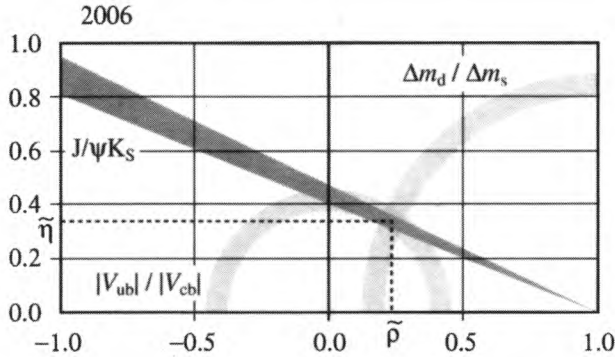
Figure 8. *Possible situation of the allowed region in the $\tilde{\rho}$-$\tilde{\eta}$ plane in 2006*

4   Presence of New Physics

Once we allow the presence of physics beyond the Standard Model, the determination of the CKM matrix elements becomes much more complicated. Most of the extensions of the Standard Model introduce new heavy particles. The contributions of these particles to the tree processes should be negligible, but for B^0-$\overline{B}^0$ and B_s^0-$\overline{B}_s^0$ oscillations, their contribution to the box processes could be sizeable. New heavy particles could also appear in the penguin processes, but the effects should be smaller than in the box processes, due to the different dependence on the masses of the particles appearing in the loops.

In the presence of new physics, the phases for the oscillation amplitudes are modified:

$$\phi_0^d = 2\beta + \phi_{\text{NP}}^d$$

and

$$\phi_0^s = -2\delta\gamma + \phi_{\text{NP}}^s$$

where ϕ_{NP}^d and ϕ_{NP}^s are new phases.

The new physics has very little effect on the decay amplitude if this is generated by tree processes. Therefore, we assume that the phases of the amplitudes for decay modes such as $B^0 \to J/\psi\, K_S$, $B^0 \to D^{*+}\pi^-$, $B_s^0 \to J/\psi\,\phi$ and $B_s^0 \to D_s^+ K^-$ remain unchanged. For the CP violation signatures in B^0 and $\overline{B}^0$ decays, we would then have

$$\sin(2\beta + \phi_{\text{NP}}^d)\sin\Delta m_d t \qquad \text{for } J/\psi\,K_S \tag{9}$$

and

$$\sin(2\beta + \phi_{\text{NP}}^d + \gamma)\sin\Delta m_d t \qquad \text{for } D^{*+}\pi^- \text{ and } D^{*-}\pi^+.$$

The value of β is modified, but it is possible to use both measurements to extract the angle, γ, in a theoretically clean way, even if new physics is present.

Similarly for the B_s^0 system:

$$\sin(-2\delta\gamma + \phi_{\text{NP}}^s)\sin(\Delta m_s t) \qquad \text{for } J/\psi\,\phi \text{ (or } J/\psi\,\eta) \tag{10}$$

and

$$\sin(-2\delta\gamma + \phi_{\text{NP}}^s + \gamma)\sin(\Delta m_s t) \qquad \text{for } D_s^+ K^- \text{ and } D_s^- K^+.$$

the value of $\delta\gamma$ is modified, but γ can be cleanly extracted by combining the two CP violation measurements.

Once γ is determined, ρ and η are given by:

$$\rho = \frac{1}{\lambda}\frac{|V_{ub}|}{|V_{cb}|}\cos\gamma, \qquad \eta = \frac{1}{\lambda}\frac{|V_{ub}|}{|V_{cb}|}\sin\gamma \,,$$

and the unmodified values of β and $\delta\gamma$ can be obtained from Equation (4). Hence all the parameters of the CKM matrix can be completely determined even if new physics exists. The contributions from new physics are then determined from Equations (9) and (10).

The allowed region in $\tilde{\rho}$-$\tilde{\eta}$ plane from the CP asymmetry measurement with B^0 and $\overline{B}^0 \rightarrow J/\psi\, K_S$ decays is no longer valid. What is measured from the asymmetry is not β, but $\beta + \phi_{NP}^d/2$. Similarly, the region given by the $\Delta m_d/\Delta m_s$ measurements is no longer valid since there could be sizeable contributions to both the Δm from new physics. The only valid region in Figure 8 is that from the $|V_{ub}|/(\lambda|V_{cb}|)$ measurement, and the apparent consistency seen in Figure 8 could be accidental. Once γ is measured in the theoretically clean way explained above, the situation will become clear. Figure 9 illustrates a possible situation where γ is measured with an accuracy of $\pm 5°$, and is not consistent with the mixing and CP asymmetry constraints. In this case, the measurements of γ reveal the presence of new physics contributions.
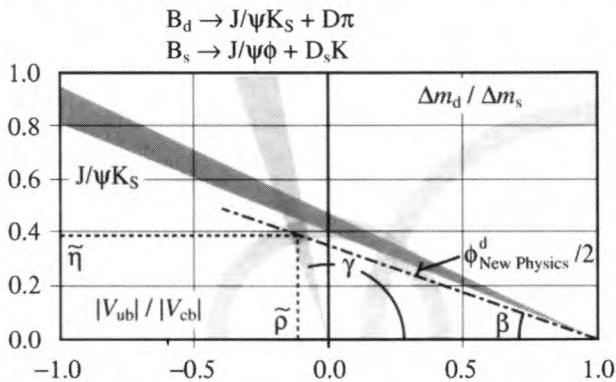
Figure 9. *Possible situation in the $\tilde{\rho}$-$\tilde{\eta}$ plane in 2006 in the presence of new physics contributions*

In the discussion above, we have made an assumption that new physics does not contribute to the decay amplitude. This may not be strictly true for the $B^0 \rightarrow J/\psi\, K_S$ and $B_s^0 \rightarrow J/\psi\,\phi$ decays, since these decays receive a small contribution from the $b \rightarrow s$ penguin process. Whether new physics can appear in the penguin process can be examined by studying CP violation in B^0 and $\overline{B}^0$ decays into ϕK_S. In the Standard Model the decay amplitude is dominated by a loop with the virtual top quark, so the phase is given by V_{ts}, and the CP violation signature is:

$$\sin(2\beta + \phi_{NP}^d - 2\delta\gamma + \phi_{NP}^P)\sin(\Delta m_d t) \,, \tag{11}$$

where ϕ_{NP}^P is the additional phase introduced in the $b \rightarrow s$ penguin process by new physics. By comparing Equations (9) and (11), it is possible to determine whether new physics contributes significantly to penguin processes.

New physics contributions to the $b \to d$ penguin processes can be studied from CP violation in B^0 and $\overline{B}^0$ decaying into $K^{*0}\overline{K}^{*0}$ or $\phi\pi^0$ where only the $b \to d$ penguin process contributes in the Standard Model. Independent studies of $b \to s$ and $b \to d$ penguins can be made with the B_s system.

In conclusion, a theoretically clean and model-independent determination of the parameters of the CKM matrix is possible, even if there are new physics contributions to CP violation observables. Once the CKM parameters are determined in a model independent way, the possible existence of new physics can be cleanly established.

5 *B* experiments at the LHC

5.1 General considerations

The goal of B physics in the LHC era (Ahmadov et al, 2000) is to determine the CKM parameters in a model-independent way and to isolate the effect of new physics so that its characteristics can be identified. This calls for an high statistics experiment capable of studying CP violation with both B_d^0 and B_s^0 systems decaying into various hadronic final states.

The production cross section for $b\bar{b}$ quark pairs at the LHC energy is estimated to be approximately $500\mu b$, which is far larger than at existing machines. The fraction of events with b quarks, $\sigma_{b\bar{b}}/\sigma_{\text{inelastic}}$, is about 6×10^{-3}, which is similar to the fraction of charm events in present fixed-target charm experiments. Thus, the LHC appears to be a very promising place to perform high precision CP violation measurements in B meson decays. At the LHC, B_s, $\overline{B}_s$, $B_c^{\pm}$ and b-baryons are abundantly produced, in addition to $B^{\pm}$, B^0 and $\overline{B}^0$.

In order to exploit the full potential of the LHC, the following capabilities are needed.

1. A trigger sensitive to both leptonic and hadronic final states.

2. A particle identification system capable of identifying p, K, π, μ and e within the required momentum range.

3. A vertex detector able to reconstruct primary and B vertices very precisely.

4. A tracking system with good momentum resolution.

5. An electromagnetic calorimeter capable of reconstructing π^0 decays.

5.2 The ATLAS and CMS experiments

ATLAS (Armstrong 1994) and CMS (Bayatian 1994) are two general purpose collider detectors designed to perform high p_T physics at the LHC, including studies of the top quark, detection of the Higgs boson, and searches for supersymmetric particles. They detect $b\bar{b}$ quark pairs in the central region of the pp interactions. As already demonstrated by CDF, such general purpose collider experiments can cleanly reconstruct B meson final states containing lepton pairs, such as $J/\psi K_S$, $J/\psi\phi$ and $\ell^+\ell^-$. Both ATLAS and CMS

have excellent muon and electron detection capabilities, and a good vertex resolution, allowing them to collect a high statistics sample of such decays. However, their first level trigger on a high p_T lepton is not sensitive to purely hadronic final states. Hadronic decay modes can only be triggered by a semileptonic decay of the other b quark, which has a low efficiency due to the relatively small branching fraction for the semileptonic decay. The two experiments also have no $p/K/\pi$ separation capability in the relevant momentum range, although the energy loss, dE/dx, can be used for particle identification at low momenta. In particular they cannot separate kaons from pions in the high momentum region relevant for two-body decay modes of B mesons such as $\pi^+\pi^-$ and D_sK.

For these reasons, ATLAS and CMS will not be able to study CP violation with all the final states necessary to perform model-independent analysis.

5.3   The LHCb experiment

The LHCb experiment is a single arm forward spectrometer, which covers the region of pp collisions which is not looked at by the two general purpose detectors. The LHCb detector (Amato 1998) is designed to have a trigger that is equally efficient for leptons and hadrons. It will be able to exploit the full b physics potential of the LHC at a much lower luminosity $(2 \times 10^{32} \text{cm}^{-2}\text{s}^{-1})$ than the nominal luminosity $(10^{34} \text{cm}^{-2}\text{s}^{-1})$, and will therefore be able to perform its full physics programme from the beginning of LHC operation.
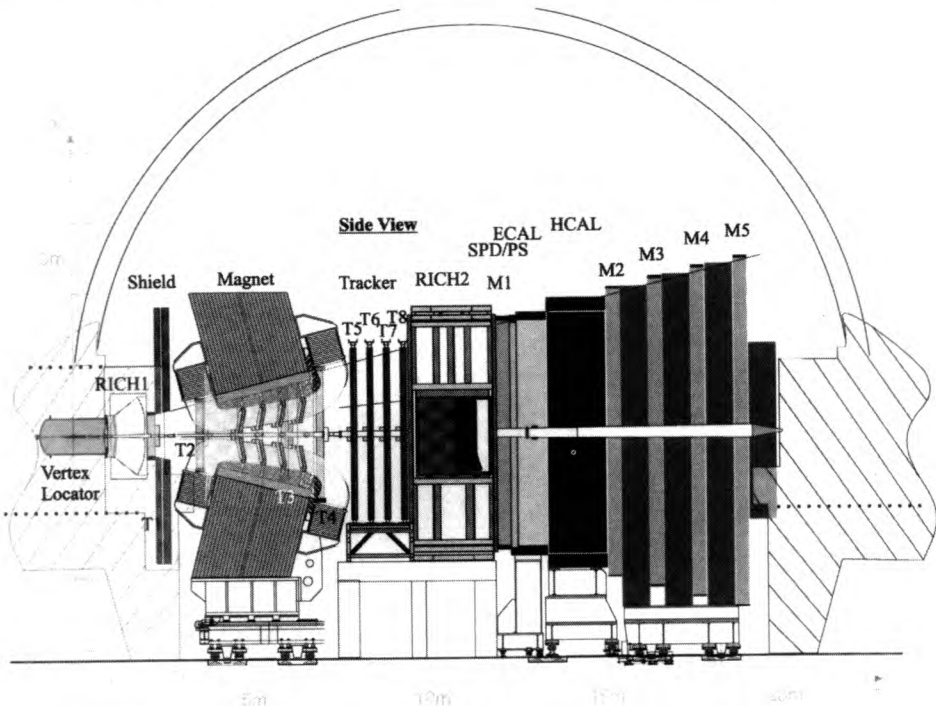
Figure 10. *The LHCb spectrometer placed in the LHC Intersection-8 experimental area.*
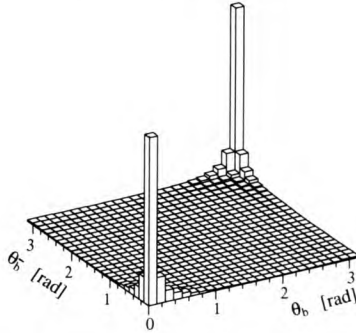
Figure 11. *Polar angles of the b- and b̄-hadrons*

The detector layout is shown in Figure 10. It resembles a typical fixed target spectrometer. A vertex detector is placed in "Roman pots" around the interaction region. It is followed by a tracking system, RICH counters with aerogel and gas radiators, a large-gap dipole magnet, a calorimeter system, and a muon system. An existing LEP experimental area will be reused to install the detector. The interaction point is shifted by 11m from the centre of the experimental hall, in order to accommodate the detector elements without extra excavation.

The choice of a single arm detector geometry is based on the fact that both the b- and b̄-hadrons are predominantly produced in the same forward (or backward) cone at high energies. This is demonstrated in Figure 11 where the polar angles of the b- and b̄-hadrons are shown in events generated with the PYTHIA simulation programme. The polar angle is defined with respect to the beam axis in the pp center-of-mass system. Detecting both b- and b̄-hadron at the same time is essential for the flavour tag.

Further advantages of the forward geometry are:

- The b-hadrons produced are faster than those in the central region. Their average momentum is about 80GeV/c, corresponding to a mean decay length of ∼ 7mm. Therefore, a good decay time resolution can be obtained for reconstructed B-mesons.

- The spectrometer can be built in an open geometry with an interaction region which is not surrounded by all the detector elements. This allows the vertex detector system to be built with sensors which can be extracted away from the beam during the injection. During data taking, the sensors are positioned closer to the beam in order to achieve a good vertex resolution.

- In the forward region, momenta are mainly carried by the longitudinal components. Therefore, the threshold value for the p_t trigger can be set low for electrons, muons and hadrons without being constrained by the detector requirements. This makes the p_t trigger more efficient than in the central region.

- The momentum range required for particle identification is well matched to the Ring Imaging Cherenkov Counters, and the required size for the counters remains affordable.

- The open geometry allows easy installation, maintenance and possible upgrades.

5.4   The LHCb detector

Beam pipe

A large vacuum tank with a length of 1.7m and a diameter of 1m is placed around the interaction point to accommodate the vertex detector system with its retraction mechanics. It has a 2mm Al forward exit window over the full detector acceptance. This part is followed by two conical beam pipe sections; the first part is 1.4m long with a 25mrad opening angle, and the second part is 16m long with a 10mrad opening angle. Except for bellows, flanges and the last 6.3m of the 10mrad cone, the beam pipe is made from Al-Be alloy in order to reduce the radiation length. This is essential to minimize the occupancies of the tracking and RICH systems, and to improve the detection efficiency for photons and electrons.

Magnet

A dipole magnet with an Al conductor provides a field integral of 4 Tm. The polarity of the field can be changed to reduce systematic errors in the CP-violation measurements that could result from a left-right asymmetry of the detector. The two pole faces form a wedge shape following the spectrometer acceptance, in order to reduce the power consumption.

Vertex locator (VELO)

A total of 27 stations of silicon microstrip detectors are placed perpendicular to the beam of which 25 stations are used as a vertex detector system. The remaining two stations are used for detecting bunch crossings with more than one pp interaction as a part of the Level-0 trigger (pile-up veto counters). All the stations are split into two halves, covering left and right 180° sections. Each vertex detector station consists of two sensor planes with different strip layouts, one for r and the other for ϕ measurements. The pile-up veto counters have only r measurement sensors.

The closest distance between the active silicon area and the beam is 8mm. In order to cope with the high radiation dose expected so close to the beam, it is planned to use n-on-n silicon sensors. The silicon detectors are placed in Roman pots surrounded by 250μm thick aluminium foil, which acts as a shield against RF pickup from the circulating beam bunches. To avoid mechanical collapse, a secondary vacuum is maintained inside the Roman pots. During injection and acceleration, the Roman pot system will be moved away from the beam to avoid interference with the machine operation and accidental irradiation of the detectors.

For each bunch crossing, i.e. every 25ns, the signals are read-out and stored in analogue pipeline buffers.

Tracking system

Because of the high particle density close to the beam pipe, the LHCb tracking detector is split into inner and outer systems. The boundary between the two is chosen so that the occupancy of the outer tracker does not exceed 15%.

The outer tracking system uses drift chambers based on a straw cell structure. The straws are made by winding 5mm diameter carbon-loaded Kapton foil around a central anode wire. The drift-time is sampled over 50ns, i.e.every two bunch crossings.

The inner tracking system is made from single sided p-on-n Si strip detectors with a

strip pitch of $\sim 200\mu$m. Since the sensitive regions of the Si sensor are several centimetres away from the beam the problem of radiation damage is less severe than for the VELO detectors.

The RICH detectors

The LHCb detector has two Ring Imaging Čerenkov (RICH) systems, with three different radiators, in order to cover the required momentum range, 1-100GeV/c. The first RICH uses aerogel and C_4F_{10} gas as radiators. The second RICH, uses CF_4 gas as a radiator. It is placed after the magnet, and is responsible for identifying high momentum particles. In both RICHs the Cherenkov light is reflected by mirrors and detected with planes of Hybrid Photon Detectors (HPD's) placed outside the spectrometer acceptance.

Calorimeters

The calorimeter system consists of a preshower detector followed by electromagnetic and hadronic calorimeters. It also serves as the initial part of the muon filter system. The cells of the preshower detector are made from two scintillator plates sandwiching 14mm-thick lead plates. The cell size of the preshower detector is matched to the module size of the electromagnetic calorimeter. For the electromagnetic part a Shashlik calorimeter is used since a modest energy resolution is required. The hadron calorimeter is based on a scintillating tile design similar to that used in the ATLAS experiment.

Muon detectors

For the muon stations Resistive Plate Chambers are used in the region where the charged particle rate is below 1kHz/cm^2. For the region with a charged particle rate from 1kHz/cm^2 to 100kHz/cm^2, Multi Wire Proportional Chambers (MWPC's) are used. For the small region of the first muon station closed to the beam pipe, where the charged particle rate exceeds 100kHz/cm^2, triple-GEM chambers or MWPC's with asymmetric gas gaps are being considered.

Trigger

The LHCb trigger is divided into four decision levels. The Level-0 decision is based on high-p_T hadrons, electrons or photons found in the calorimeter system, or muons found in the muon system. Information on these candidates are sent to the Level-0 Decision Unit. The number of primary vertex candidates is determined using the pile-up veto counters in the Vertex Locator and also sent to the Level-0 Decision Unit. Based on all the information, the Level-0 Decision Unit makes an overall Level-0 decision. Events with multiple pp interactions are discarded, since rejection of background is more difficult in those events. The Level-0 trigger provides a modest reduction of minimum bias events by a factor of ~ 10.

In the Level-1 trigger, data from the vertex detector is used to select events with multiple vertices. In addition, the high-p_T candidates used in Level-0 can be combined with the tracks having large impact parameter in the Vertex Locator. This provides a further reduction of ~ 25 for minimum bias events. After a positive decision of the Level-1 trigger, all the data is read out into an event buffer.

At Level-2, a further enhancement of events with b-hadrons can be achieved by combining different detector components, e.g. by adding momentum information from the main tracking system to the impact parameter calculation with the Vertex Locator. At

Level-3, a further decision is made after fully reconstructing the b decays.

The LHCb trigger system is designed to cope with the rather small $\sigma_{b\bar{b}}/\sigma_{\text{inelastic}}$ of $\sim 6 \times 10^{-3}$ at the LHC energies, while still maintaining a high efficiency for events with b hadrons. The strategy is to spread the suppression factors evenly and not to rely on any single trigger selection, in particular at early levels where available information from the detector is limited. This is reflected in the modest suppression factors of 10 and 25 for the ordinary pp interaction events at Level-0 and Level-1, respectively. Simulations of the trigger performance can be relied upon for such a modest suppression. By not heavily relaying on a particular selection criterion, the trigger system is flexible and can be readjusted to the actual running conditions of the experiment.

5.5   Performance of the LHCb detector

The benefit of having particle identification and good invariant mass resolution can be demonstrated by reconstructing $B_s \to D_s^+ K^-$ decays. The main background to this decay mode comes from the $B_s \to D_s^+ \pi^-$ decays which are used to study B_s–$\bar{B}_s$ oscillations. Compared to this decay mode, the branching fraction of $B_s \to D_s^+ K^-$ is suppressed by a factor of $1/\lambda^2 \approx 20$. The two decay modes have an identical decay topology, and can only be distinguished by invariant mass and particle identification. It should be noted that the momenta of the K^- and π^- from these decay modes are large since they are two-body decays.
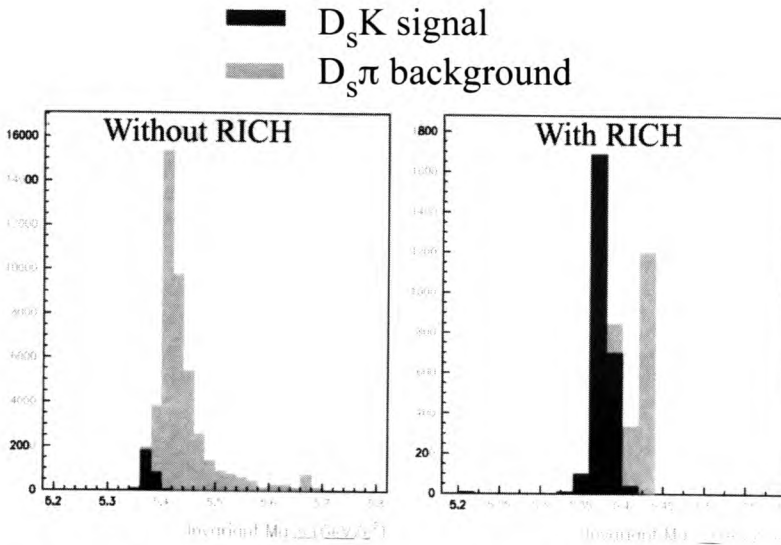
Figure 12. *The LHCb simulation results for the reconstructed invariant mass distributions for $D_s^+ K^-$ combinations without and with particle identification using RICH.*

Figure 12 shows the reconstructed invariant mass distributions expected with the LHCb detector for $D_s^+ K^-$ combinations without and with particle identification using RICH. The combination of the good mass resolution, $\sigma = 11 \text{MeV}/c^2$, and RICH par-

ticle identification, mean that the background from $B_s \to D_s^+\pi^-$ decays can be almost completely removed in the reconstructed $D_s^+K^-$ sample. It should be noted that no CP violation effect is expected in the background decay mode. The decay time resolution for these decays is found to be 40fs.

With the particle identification capability, an efficient flavour tag can be obtained using the charged kaons from the b hadron decays. No lepton is required in the analysis for flavour tagging. Therefore, the Level-0 high p_T hadron trigger increases significantly the statistics of the flavour-tagged sample of pure hadronic B decay final states, compared to using the high p_T lepton trigger alone. Table 1 summarises the Level-0 trigger efficiencies for various decay modes. Efficiencies are calculated for those events where the initial

Decay Mode	Level-0 high-p_T			Level-0
	muon	electron	hadron	all combined
$B^0 \to J/\psi(e^+e^-)K_S$	0.17	0.63	0.17	0.72
$B^0 \to J/\psi(\mu^+\mu^-)K_S$	0.87	0.06	0.16	0.88
$B_s \to D_s^+K^-$	0.15	0.09	0.45	0.54
$B^0 \to \pi^+\pi^-$	0.14	0.08	0.70	0.76

Table 1. *Level-0 trigger efficiencies for reconstructed and flavour tagged final states.*

flavour is identified and the final state is fully reconstructed with all the cuts applied to remove background. While $J/\psi K_S$ final states are mainly triggered by the muon and electron high-p_T triggers, the hadron high-p_T trigger is essential for the other final states.

The table also indicates that the Level-0 trigger efficiencies are very high for the events that are useful in the analysis. As a result, the LHCb experiment will run with a luminosity of 2×10^{32}cm^{-2}s^{-1} and still collect 2.4k reconstructed and flavour-tagged B_s and $\overline{B}_s$ decays into $D_s^+K^-$ and $D_s^-K^+$ in one year. With these statistics, γ can be measured with an accuracy of $\sim10°$. By combining the result from the decays of B^0 and $\overline{B}^0$ into $D^{*+}\pi^-$ and $D^{*-}\pi^+$, γ will be measured with a precision of better than $7°$ with one year of data taking.

At a low luminosity, the bunch interactions are dominated by events with only one pp collision. The running luminosity will be locally tuned at the LHCb intersection such that the experiment can run with this optimal luminosity while the other LHC experiments run at the design luminosity. It must be noted that running at lower luminosities has the additional benefit that the radiation damage to the detector is reduced.

6   Conclusions

CP violation outside the neutral kaon system has been observed for the first time by BABAR and BELLE experiments, in the decays of B^0 and $\overline{B}^0$ into the $J/\psi K_S$ final state. Current analysis shows that all the measurements related to b hadron decays are consistent with the CKM picture, including CP violation. Although BABAR, BELLE, CDF and D0 will further enhance these studies in the near future, a new generation of experiments at

the LHC is needed in order to examine the CKM picture in a model-independent way. It should be noted that a dedicated experiment at the Tevatron (Kulyavtsev 2000) would have a similar performance. LHCb is a dedicated detector for B physics at the LHC with particle identification capability and a trigger sensitive to both leptonic and hadronic final states, which should be able to study CP violation in both B^0 and B_s^0 meson systems in many decay modes. Some of those decay modes are essential to determine the CKM parameter in a theoretically clean and model-independent way. This will allow the effect of possible new physics to be identified unambiguously.

Acknowledgments

References

Ahmadov, A. et al (2000) *Proceedings of the Workshop on Standard Model Physics (and more) at the LHC*, CERN 2000-04 .

Amato, S. (1998), *LHCb Technical Proposal*, CERN/LHCC/98-4.

Armstrong, W. W. (1994), *ATLAS Technical Proposal*, CERN/LHCC/94-43.

Bayatian, G. L. (1994), *CMS Technical Proposal*, CERN/LHCC/94-38.

Davies, C. (2001) *This volume.*

Groom, D.E. et al. [Particle Data Group] (2000), *The European Physical Journal* **C15**, 1

Hagelin, J. S. (1981), *Nucl. Phys.* **B193**, 123. For more details, see Buras, A. J, (1999), *Proceedings of 14th Lake Louise Winter Institute* (Singapore: World Scientific).

Kobayashi, M. and Maskawa, K. (1972), *Prog. Theor. Phys.* **49**, 282.

Kulyavtsev, A. (2000), *BTeV Proposal*, BTeV-doc-66-v1 .

Nir, Y. (2001) *This volume.*

Shaposhnikov, M.E. (1986), *JETP Lett.* **44**, 364.

Schubert, K.R. (2001), *This volume.*

Stone, S. (2001), *This volume.*

Wolfenstein, L. (1983), *Phys. Rev. Lett.* **51**, 1945.

Index