HEAVY-FLAVOR PHYSICS

Helen Quinn
Stanford Linear Accelerator Center, Stanford University, Stanford, California

Abstract
These lectures are intended as an introduction to flavor physics for graduate students in experimental high-energy physics. In these four lectures the basic ideas of the subject and some current issues are presented, but no attempt is made to teach calculational techniques and methods.

1. WHAT IS FLAVOR PHYSICS? WHY STUDY IT?

The flavor sector is that part of the Standard Model which arises from the interplay of quark weak gauge couplings and quark-Higgs couplings. The Standard Model physics of the quark masses and matrix of quark weak couplings in the mass eigenstate basis that encodes these effects will be briefly reviewed. From a theorist's perspective, the aim of the game in flavor physics today is to search for those places where Standard Model predictions are clean enough that effects from physics from beyond the Standard Model could be recognized if indeed they occur. Along the way we also want to refine our knowledge of the Standard Model parameters of this sector.

The Standard Model encodes the physics of weak decays of quarks. Predictions at the quark level are clean and simple, since leading-order weak effects are all we need. However we observe hadrons, not quarks. Hence we must explore how and when the quark-level physics can be studied at the hadron level. In particular we search for those places where the (hard-to-calculate) hadronic physics effects can be separated from the calculable quark-level predictions. This will direct our attention to two classes of results: those on rare decays or effects suppressed by symmetries in the Standard Model; and those on CP violation in the decays of neutral, but flavor non-singlet mesons. In the first case there is a possibility that new physics effects are large compared to Standard Model contributions. In the second, Standard Model predictions for a broad range of effects are related because there is only one CP violating parameter in the quark sector of the Standard Model. These relationships provide many tests of the theory. I will review this latter topic in some detail, particularly for the case of B decays, because these are the topic of current and planned experimental programs. I will also discuss some examples of rare K decays, and the issue of D-mixing, as examples of effects that are suppressed in the Standard Model and hence provide other possible probes for new physics.

I will review the calculational tools that are at our disposal to control the uncertainties that arise from hadronic physics. These include the use of symmetries, perturbative QCD and the operator product expansion, heavy quark expansion, lattice QCD calculations and QCD sum rules. I will discuss the roles of these techniques, with some examples. This short lecture series will not teach you how to do any of these calculations, my emphasis is rather to help you as experimentalists to become informed consumers in the theory market.

There are some excellent textbooks available covering B physics and CP Violation in much more detail than these four lectures can. I refer students to these texts for further study [1]. Another useful reference is the BaBar Physics Book [2], which summarizes a year-long study effort planning the experimental program for that detector. As well as some introductory chapters and appendices that cover general issues in the theory, this report discusses in detail both the theoretical and the experimental issues for a number of interesting B decay modes.

2. LECTURE 1

2.1. Preliminaries

Gauge theories require universality in couplings. However, for the weak interactions, this universality is masked because of spontaneous symmetry breaking. The down-type quark and neutrino mass-eigenstates are admixtures of the states that couple in a simple universal fashion to the distinct up-type quark or charged lepton mass eigenstates respectively. This occurs because there is no universality constraint on the Higgs-fermion couplings. All the richness and variety of the flavor sector is the image in our world of the richness and variety of Higgs couplings, which give all but four of the parameters in the Standard Model.

I will restrict my attention to quark flavor physics. Alvaro's lectures on neutrino physics [3] cover lepton flavor physics. The flavor parameters in the quark sector are the quark masses, and the elements of the quark mixing matrix, the matrix that specifies quark-W couplings in the quark mass-eigenstate basis. In a world without strong interactions the measurement and definition of all these parameters would be a simple matter. However, in the real world of quark confinement many challenges arise, both in defining and in determining these parameters.

Putting those challenges aside for the moment, why is that task of such great import? The answer is that the Standard Model is predictive, and thus testable, because there are so few parameters. Said another way, there are multiple measurements that depend on the same few parameters in the Standard Model. The comparison of values of the parameters obtained by different measurements provides a test of the theory, or alternatively a probe for physics beyond the Standard Model. Interesting tests also arise in cases where the Standard Model predicts that a particular effect is either zero, or very small. Often these predictions can be dramatically altered in extensions of the Standard Model containing new particles. Searches for such effects thus provide another set of tests of the theory. The possibility that a test may fail is what makes it interesting!

Because of strong interactions, neither the quark masses nor the mixing matrix elements can be measured directly. In all cases theory must be used to connect a given measurement to the relevant Lagrangian parameters. Ambiguities due to convention dependence (such as subtraction method) can readily be taken care of, although of course they require attention and commonly accepted conventions. Much more troublesome ambiguities arise from the fact that strong interaction physics effects can obscure the relationship between the measured quantities and the quark-level parameters. In order to test the theory by comparing two measurements that should be related to the same Standard-Model parameter, we must be able to quantify the differences that could be accounted for by hadronic physics effects. If we cannot do this our test is lost. Hence, while the flavor parameters are weak interaction parameters, much of what one must think about in discussing measurements to determine these parameters is how to constrain or remove the impact of hadronic physics.

These lectures will likewise emphasize this issue. The challenge is to find cases where hadronic effects are small, or where combinations of multiple measurements plus predictions based on symmetries of the theory can be used to measure or constrain the impact of hadronic physics on the parameter of interest. As data improve, we theorists must continue to improve calculational methods and our ability to quantify the impacts of hadronic physics. Indeed that is a major focus of theoretical work in flavor physics today.

Of course testing the Standard Model is not a new area of investigation; there are already many tests of flavor-physics predictions of the Standard Model. The continued success of the theory in accommodating all results already enforces some strong constraints on extensions of the theory. Any additional particles that could affect new flavor physics measurements must also be consistent with all past results.

The Standard Model generation structure was indeed invented to explain one of the strong constraints, to give a gauge theory of weak interactions that includes a Z boson with no flavor changing neutral coupling. Such effects are strongly constrained by the small mass difference between the neutral-kaon mass eigenstates. Likewise any CP-violating and thus, via CPT, T-violating effects in extensions of the Standard Model are strongly constrained by the small upper bound on the neutron electric dipole moment. The match between theory and experiment for the quantity ($g - 2$) of both the electron and the muon also provides constraints. All these effects put lower bounds on the masses of additional particles, and hence, in general, on the size of amplitude contributions involving such particles. New tests, the focus of much of these lectures, will provide further constraints on model building, or, if we are lucky, a glimpse of new physics.

What are the patterns of flavor physics? Where can we make incisive tests of the Standard Model? Why are heavy quarks so interesting in this regard? To answer these questions one must first step back and look at the hierarchy of mass scales that enter the physics of flavor. Only after this can we understand what it means to say a quark is heavy or light–heavy or light with respect to what scale? Aside from the masses of the quarks themselves two other scales define the physics. The first of these is the mass of the W boson, the other the scale Λ_{QCD}, which is much smaller than the W mass.

What is the physical meaning of Λ_{QCD}? The formal definition as the scale where perturbation theory gives an infinite coupling constant for the strong gauge interactions is clearly not physical; no measurement can give infinity as its result, and perturbation theory clearly breaks down well before the coupling grows so large. A better definition is to say that Λ_{QCD} is the scale that defines the running of the strong coupling constant that should be seen in high-energy jet physics, and in the binding of massive "onium" type states. Indeed these are the measurements used to determine it. But this is still a rather esoteric definition. What actually occurs at this scale, what quantity in low-energy physics depends on it? One needs to understand this to see the role that Λ_{QCD} plays in flavor physics. The answer is that this scale sets the size of hadrons and thus also gives the scale of the kinetic energy of quarks confined within these hadrons. For hadrons built solely of light quarks, which we can now understand to mean quarks whose mass is small compared to the scale Λ_{QCD}, it thus also gives the scale of hadronic masses. The up and down quarks are light quarks, but the strange quark is a borderline case.

Conversely, quarks with masses large compared to Λ_{QCD} are heavy quarks. There are two consequences of being heavy: the first and most obvious one is that the quark mass dominates the mass of any hadron containing that quark, and thus such quarks are effectively static components of hadrons (mass large compared to kinetic energy); the second is that the strong interaction coupling at the scale of that quark mass is small. Thus there are two small parameters for heavy quark physics Λ_{QCD}/m_q and $\alpha_s(m_q)$. Expansions in both of these parameters are useful in calculating the impact of hadronic physics on weak decay processes. This means that we have better control over these effects for hadrons constraining heavy quarks that we do in the case of light hadrons.

Weak interactions of hadrons are rare compared to electromagnetic ones because the masses of the W and Z bosons are large compared to hadron masses. The exception comes when we get to the extremely heavy top quark, with a mass greater than the W. The weak decays of the top quark occur so rapidly that it decays before it ever has time to form a hadron. Further, its decays are very strongly dominated by the decay $t \to b + W$. All other decays are so suppressed that there is no chance we will detect them any time soon (especially since the only way we can tell a top quark was produced is via the signatures of the $t \to b + W$ decay)! Thus there is no interesting flavor physics that can be studied for the top quark, nor do we see hadrons containing that quark.

When we talk of the heavy-quark limit for hadronic physics, we take that limit while ignoring weak decays. One can ask how hadronic wave-functions and other hadronic properties scale as the quark mass goes to infinity, without considering the fact that no such hadron is ever formed. The rigorous scaling properties derivable in this limit can then be used to constrain models and to make predictions about the behavior of hadrons in the interesting heavy-quark mass range – namely around the mass of the b-quark, which is conveniently large compared to Λ_{QCD} while still small compared to M_W.

It would be very convenient for the study of flavor physics if there were more than one quark in this mass range. In fact there is a second quark that is almost so, the charm quark. The ratio Λ_{QCD}/m_c

is about 0.3, small, but not quite small enough. The QCD corrections to weak decay patterns, and the leading order Λ_{QCD}/m_c effects are both quite large for charm-quark states, which limits our ability to make clean predictions about charm decays. We can, and do, use heavy-quark theory to simplify the analysis of $b \to c$ decays, but must take care to allow for the leading order Λ_{QCD}/m_c corrections to these predictions.

At the opposite extreme, when quark masses are light compared to Λ_{QCD} we can use this fact to derive symmetries of the light hadrons. Since both m_u and m_d are small on this scale, both isospin symmetry, which is broken by their difference, and chiral symmetry, which is exact in the zero-mass limit, provide useful inputs for the study of hadronic physics. The additional symmetries that involve the strange quark as well – the full SU(3) flavor symmetry or its SU(2) subgroups U-spin (the symmetry of interchange of s and d quarks) and V-spin (s and u quarks) – have larger symmetry-breaking effects. These are scaled by m_s/Λ_{QCD}, which is again borderline as a small parameter. However, as we will see throughout these lectures, these symmetries provide useful constraints on hadronic effects. The question of how to quantify corrections to the symmetry limit dominates the discussion of theoretical uncertainties for the relationship between measurement and Standard-Model predictions in many cases. But that is always a gain over trying to quantify uncertainties in a case with no known good limit.

In addition to these flavor symmetries, a set of discrete symmetries are of interest here. Three discrete transformations can be defined for all fields. These are: C, charge conjugation, which interchanges particle and antiparticle; P, parity, which reverses all spatial co-ordinates; and T, time-reversal, which interchanges in-states and out-states. The product of these three operations, CPT, is an exact symmetry in any local Lagrangian field theory. This follows from the locality, Lorentz Invariance, and hermiticity of the Lagrangian. This means that any two rates which are related to one another by the operation of CPT must be equal in any field theory. Tests for CPT violation are thus testing for physics which lies outside the realm of local Lagrangian field theory.

The combination CP, and thus T, is also an automatic symmetry in pure gauge theories, and hence, in particular in QED and QCD. These theories also have separate C and P conservation, as long as all quarks are massless. In this limit, weak interactions maximally violate C and P but conserve CP. The pattern of automatic CP conservation applies for many theories beyond QED, and indeed was the only one known to theorists prior to the experimental discovery of CP violation [4], which explains why this discovery was such a shock to the physics community. However we now understand that CP violation, unlike CPT violation, is readily accommodated in field theories. CP violation can arise whenever there are complex coupling constants in a theory that cannot be removed by any set of phase redefinitions of the fields. (I will explain this point in more detail later in this lecture.)

In the Standard Model the quark–Higgs Yukawa couplings introduce a large number of additional parameters, and allow the possibility of CP violation. The quark masses, as well as the W and Z masses, arise due to the spontaneous breaking of the $U(1)$ symmetry of the Higgs field. Quark masses come from the Higgs field vacuum expectation value via the Yukawa couplings of the quarks to the Higgs field. These couplings thus define what combination of quark weak eigenstates (states paired to a given up quark in weak decays) form a definite mass down-type quark. Thus, in the quark mass-eigenstate basis, these couplings are the source of the quark mixing-matrix parameters, which define the strength of the various W-emission transitions.

In a two-generation Standard Model, all couplings can be made real by field redefinitions, starting from the most general complex, but hermitian, Lagrangian with the symmetries of this theory imposed. However, as one adds quark multiplets the number of independent quark–Higgs Yukawa couplings grows faster than the number of quark fields. A priori these couplings are allowed to take the most general complex form. Hermitian conjugate couplings must also be included, these reverse the types of quark and antiquark. Then all possible phase redefinitions of the fields can be used to make as many couplings as possible real.

For three generations of quarks in the Standard Model, after field redefinitions have removed as many phases as possible, and the constraints due to unitarity of the theory have been imposed, the quark mixing matrix, known as the Cabibbo–Kobayashi–Maskawa (CKM) matrix [5], contains four independent parameters, one of which is a complex phase that causes CP violation. These are the basic parameters of flavor physics, and the subject of these lectures. Our aim is to determine the CKM parameters as cleanly as possible by many independent sets of measurements and to search for effects that are not consistent with the predictions of this theory.

In K-decay physics, long-range hadronic effects are large, and thus there are few precise predictions. The exceptions are decays that are very rare or forbidden in the Standard Model. Searches for these decays provide interesting tests for new physics. CP violation in the decays of neutral K mesons is one such rare effect that has been studied in some detail. As we will show later it provides some interesting constraints on the CP-violating parameter in the CKM matrix.

Charm decays suffer a similar fate; corrections to leading-order predictions are significant and difficult to quantify. Thus there are few clean tests of the Standard Model to be found in studying these decays. Again rare effects are an exception. One particularly interesting rare effect in the charm sector is $D-\overline{D}$ mixing. This is very suppressed in the Standard Model, and so provides a search arena for new physics effects.

The cleanest arena for flavor physics is b decay physics, because the mass of the b-quark is large enough that we have some reliable small expansion parameters to help to control the impact of hadronic physics. This explains why this b-physics is such a central part of the program of particle physics experiments at present, and will continue to be so for the next ten years at least. Hence the major focus of these lectures will be b-decay physics, and within that the physics of the B_d and B_s meson systems. These neutral but flavored mesons, like K mesons, have interesting quantum mechanics. The mass eigenstates cannot be the flavor eigenstates, but would be CP eigenstates if that symmetry were exact. These systems thus provide sensitive probes of the CP violation structure of the Standard Model, with multiple predictions that depend on the same few Standard Model parameters.

2.2. The CKM matrix

The parameters of interest are those that appear in the CKM matrix [5]. This matrix defines the relative strengths of the quark flavor transitions that occur due to W emission. The matrix elements V_{ij} denote the transition between an up-type quark i and a down-type quark j. The magnitudes of these matrix elements are physically measurable quantities. We will discuss ways in which they can be determined in the next lecture. The phase of any of the matrix elements can be changed by phase redefinitions of one of the quark fields that couples via that term. Under the redefinitions $q_i \to e^{i\alpha_i} q_i'$ the term $V_{ij}\overline{q}_i \gamma_\mu W^\mu q_j$ becomes

$$e^{i(\alpha_j - \alpha_i)} V_{ij} \overline{q}_i' \gamma_\mu W^\mu q_j' = V_{ij}' \overline{q}_i' \gamma_\mu W^\mu q_j' \; . \tag{1}$$

This redefinition has no physical consequences. Clearly, however, the relative phases of two terms that involve the same quark field are not changed by a redefinition of that field. So phase differences are the physically meaningful and measurable quantities, whereas the phase of any one term is convention dependent and unphysical.

The weak interaction gauge symmetry requires that the CKM matrix is unitary, unless there are additional quark types beyond the three generations of the Standard Model. The relationships that arise from this requirement reduce the number of independent parameters in the matrix to four, including the one phase. One commonly used way to define these parameters was suggested by Wolfenstein [6], namely

$$V = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} = \begin{pmatrix} 1-\lambda^2/2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1-\lambda^2/2 & A\lambda^2 \\ A\lambda^3(1-\rho-i\eta) & -A\lambda^2 & 1 \end{pmatrix} + O(\lambda^4) \; . \tag{2}$$

The parameter $\lambda \equiv |V_{us}|$ is a small number, of order 0.2. The higher powers of this parameter that appear in the more off-diagonal matrix elements V_{cb} and V_{ub} have no theoretical basis; they are simply a way of denoting the empirical fact that these matrix elements are successively smaller. The powers of λ are chosen so that the parameters A, defined by $|V_{cb}|$, and $\rho^2 + \eta^2$, defined by $|V_{ub}|$, are of order 1. The matrix elements V_{cd}, V_{td} and V_{ts} are then fixed by unitarity to take the form given here, up to corrections of order λ^4. The Wolfenstein parametrization is also a choice of phase convention for this matrix. The parameter η is a CP-violating parameter; the couplings in which it appears are complex. (We will see later how complex couplings give rise to CP-violating effects.)

The unitarity constraints take the form

$$\sum_{(i=u,c,t)} V_{ij} V_{ik}^* = \delta_{jk} \tag{3}$$

and likewise

$$\sum_{(j=d,s,b)} V_{ij} V_{kj}^* = \delta_{ik} \; . \tag{4}$$

The off-diagonal relationships in Eq. (3) or (4) take the form of a sum of three complex numbers equal to zero. They can be represented as a closed triangle of vectors in the complex plane. These are called the unitarity triangles. Notice that the angles in any of these triangles, that is the relative phases of the terms in any one of the sum relationships, cannot be changed by any set of phase redefinitions of the quark fields; they are rephasing-invariant quantities. In fact, all these triangles are related; the true rephasing invariant statement, that there is only one independent CP-violating parameter in the matrix, is the condition that all these triangles have the same area, J/2, where J is called the Jarlskog invariant, for Cecilia Jarlskog who first proved this fact [7]. Obviously J is zero if all couplings are relatively real.

While all the triangles have the same area they come in three distinct types. Consider the case for Eq. (3) with $k = d$ and $j = s$. Then two of the terms are of order λ and the third is of order λ^5. The area of this triangle is thus of order λ^6; its small angle is of order λ^4. Asymmetries proportional to such a small parameter are extremely unlikely to be measured. For the case $k = s$ and $j = b$ one finds two sides of order λ^2 and one of order λ^4, again an area of order λ^6. Here the small angle is of order λ^2, difficult but perhaps not impossible to measure. Finally for the case $k = d$ and $j = b$ we find all three sides are of order λ^3 and thus all angles are of order unity.

This last is the interesting case for CP violation studies, which measure quantities directly proportional to the angles of the triangle. While the overall effect is of order λ^6, here we have CP asymmetries of order unity in rare processes, as compared to the first case where the CP asymmetries are of order λ^4 but could occur in leading weak decay rates. In B physics, when people talk of "the unitarity triangle" they mean this last triangle. Conventionally it is drawn with the sides rescaled by the quantity $V_{cd}V_{cb}^*$ so that the base is real and of unit length and the apex of the triangle is the point $\overline{\rho}, \overline{\eta}$ in the complex plane where $\overline{\rho} = \rho(1 - \lambda^2/2)$, and $\overline{\eta} = \eta(1 - \lambda^2/2)$.

The angles of this triangle have, unfortunately, two conventionally used sets of names, they are either α, β, γ or ϕ_2, ϕ_1, ϕ_3, where the first named is at the apex and the order is clockwise around the triangle. One can of course determine this triangle by measuring the lengths of its sides, all of which are CP-conserving quantities. The match between the angles determined in that way and those found by measuring CP-violating quantities is a test of the Standard Model. In the next lecture we will discuss how, and how well these various quantities are measured.

2.3. Consequences of unitarity

The unitarity relationships are the residual effect of the initial gauge symmetry. Thus they serve to protect the theory from certain divergences that could arise in a less tightly-constrained theory. The generational structure of the Standard Model was designed to cancel all anomalies in the theory and to give a theory which, at the leading order, has no strangeness-changing neutral current couplings. At one loop there are divergent diagrams of the form of Fig. 1 that introduce the process $Z \to s\overline{d}$ and similarly for other such flavor-changing currents.

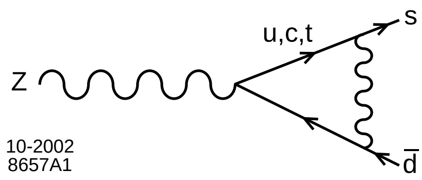

Fig. 1: Flavor changing Z-coupling induced at one loop in the Standard Model

The internal quarks in the loop in Fig. 1 can be any one of the three up-type quarks. The sum of all such diagrams is given by

$$A(Z \to s\overline{d}) = \Sigma_{i=u,c,t} V_{is} V^*_{id} f(m_i) \; . \tag{5}$$

Here the quantity $f(m_i)$ represents the Feynman integral over loop momenta for the diagram containing the i-th up-type quark in the internal lines of the loop. If all three up-type quarks had equal mass, then the unitarity condition (Eq. (3)) would say that this amplitude vanishes. Since the divergent term in the integral does not depend on the quark mass, the unitarity relationship guarantees that the divergences cancel. Thus one is left with a finite result. One way to see this explicitly is to use the unitarity relationship to eliminate one set of CKM coefficients, say $V_{cs}V^*_{cd}$. This gives

$$A(Z \to s\overline{d}) = V_{us}V^*_{ud}[f(m_u) - f(m_c)] + V_{ts}V^*_{td}[f(m_t) - f(m_c)]. \tag{6}$$

The divergence cancellation is now explicit in the differences of the $f(m_i)$. This shows that there is no need to introduce a bare flavor-changing neutral current coupling. Furthermore, the resulting finite effect is small. The first term in equation (6) shows the usual “GIM mechanism” suppression of the two-generation theory; it is proportional to $(m_c - m_u)/M_W$. The second term is suppressed by an additional four powers of the small parameter λ.

This same unitarity pattern is used over and over again in the Standard Model to combine terms from similar diagrams with different internal quarks. In addition to demonstrating divergence cancellations it is a useful way to group terms to display both the size of the resulting terms and any possible CKM phase structure. For example consider the loop diagrams that contribute to mixing between a D^0 and a $\overline{D}^0$ meson. The diagrams are shown in Fig. 2. Any one continuous quark line in these diagrams can have any of the three down-type quarks in the intermediate state. One can write the contribution of such a quark line as a sum of propagator terms with their respective CKM coefficients. One can then use unitarity to group the terms, this gives

$$\begin{aligned} Q(k, m_i) &= V_{cd}V^*_{ud}D(k, m_d) + V_{cs}V^*_{us}D(k, m_s) + V_{cb}V^*_{ub}D(k, m_b) \\ &= \lambda[D(k, m_d) - D(k, m_s)] + \text{terms of order } (\lambda^5) \; . \end{aligned} \tag{7}$$

Since there are two such quark lines in each diagram, this gives a finite loop integral proportional to $[\lambda(m_s - m_d)/M_W]^2$.

One ends up with a very small mixing effect, naively of order 10^{-5}. This means that the D mesons decay more rapidly than they mix in the Standard Model. Early papers on this topic [8] therefore suggested that any observable mixing effect would be a sign of new physics; that is of physics beyond the Standard Model. However the first diagram of Fig. 2 has real intermediate states when both the two inter-

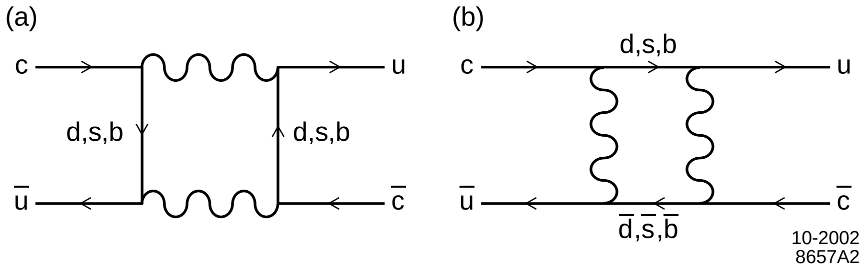

Fig. 2: Diagrams that give D^0-$\overline{D}^0$ mixing in the Standard Model

mediate quarks are either d or s quarks. Real intermediate states always bring in long-distance hadronic physics effects; these are not completely accounted for in evaluating the quark-level Feynman integral. These real intermediate states contribute to the width difference for the D mesons in the Standard Model. Furthermore, in the "operator-product expansion", there are higher order operators which contribute to the width difference without the same quark-mass suppression factors. Current estimates suggest that the quantity $y = \Delta\Gamma/2\Gamma$ could be as large as a few percent, while $x = \Delta M/\Gamma$ is probably less than 10^{-3} [9].

Experiments now set bounds at the few percent level on a linear combination of x and y (different combinations in different experiments) [10]. The challenge now is to untangle these two terms. Current experimental accuracy is insufficient to do so. If it turns out $y \gg x$ and of order a few percent one could not be sure that this is not Standard Model physics, however if $x \gg y$ with x at the few percent level that would be difficult to understand without invoking new physics effects. So there is still some discovery potential in this search. If the theorists could clean up the calculations further, there may be even more.

This example illustrates the general problem in interpreting experimental results in flavor physics. The goal is to search for effects that do not fit Standard Model calculations. Long range hadronic physics introduces contributions that cannot be calculated with great accuracy. The uncertainty in the size of these effects becomes the theoretical uncertainty of the Standard Model prediction. Our ability to see physics from beyond the Standard Model is obscured if this uncertainty cannot be well constrained.

This is generally the case in dealing with hadronic charm meson decays. The overall patterns of charm decays clearly show that long-distance physics effects are significant. Final-state interaction effects, including those that involve $q\overline{q}$ annihilation, must be invoked to fit the observations. It is difficult to make any tests of the Standard Model in this context, as the predictions are simply not precise enough. There are, however, a number of B-meson decays that can only be interpreted using precise knowledge of certain D-decay rates, so those measurements are important. I will not talk further about the interpretation of charm decays in these lectures; for a thorough discussion of this topic see Ref. [11].

Figure 3 shows what is meant by a tree diagram and a loop diagram at the level of quark weak decays. These two types of quark-level processes are the dominant contributions for most weak decay amplitudes. Two factors govern the general size of an amplitude contribution. One is the size of the CKM coefficient that appears in it, the more factors of λ there are the smaller the contribution. The second is whether it is a tree or a loop diagram. Loop diagrams have an additional factor of order $\alpha_S(m_q)/4\pi$, due to the gluon emission. Thus, for heavy-flavor decays they are suppressed relative to trees.

One way around the hadronic physics impasse is to look for decays that are forbidden at the tree level and have CKM-suppressed one loop contributions. Since any new physics process is, by definition, mediated by heavy particles, it is unlikely to compete with unsuppressed tree-level Standard Model processes. However such effects could be large compared to a rare or forbidden Standard Model contribution. Hadronic corrections introduce uncertainty in the size of the Standard Model contribution, but for such channels the discrepancy arising from new physics effects could be large even compared to this uncertainty.

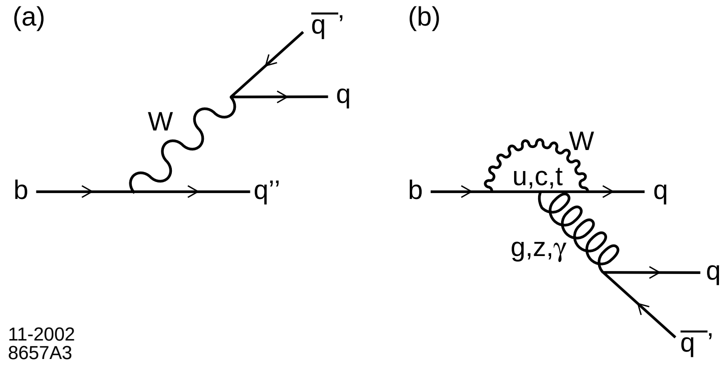

Fig. 3: Quark level tree (a) and penguin (b) graphs for weak decay.

In the terminology of weak decays the loop-diagrams are called penguin processes. At the quark-level they are diagrams of the type shown in Fig. 3(b), in which the W-boson is emitted and reabsorbed on a single quark line, and a gluon, photon or Z-boson is emitted from the loop either as a real particle (photon case only) or produces an additional quark–antiquark or lepton–antilepton pair. Examples in kaon physics are decays such as $K \to \pi\mu^+\mu^-$ and, even more rare, $K \to \pi\nu\overline{\nu}$; B-decay processes that occur only via loop diagrams are $B \to s\gamma$, and $b \to s\overline{s}s$ (such as $B \to \phi K_S$). In the latter case both the rate and the size of the time-dependent CP-violation are interesting places to search for new physics.

A short comment on what channels are "pure penguin" may be helpful. It illustrates the general problem of hadronic physics effects. At the quark level it appears that there are four pure penguin channels for b decays: $b \to s\overline{s}s$, $s\overline{s}d$, $d\overline{d}d$, and $d\overline{d}s$. In general the strong interactions readily mix $d\overline{d}$ or $s\overline{s}$ with $u\overline{u}$ – especially if these quark pairs make a neutral pseudoscalar meson. The tree contributions $b \to u\overline{u}d$ are thus a possible contributor to many channels accessible via the $s\overline{s}d$ or $d\overline{d}d$ quark-level process, so these cannot be considered "pure penguin" channels. In the case where the third quark is an s quark, the tree contribution $b \to u\overline{u}s$ is suppressed by the small CKM factor $V_{ub}V_{us}^*$ and so does not create a large effect. Finally when the $s\overline{s}$ pair makes a vector meson ϕ, the fact that this meson is almost pure $s\overline{s}$ further suppresses any impact of the $u\overline{u}$ tree graph.

In what follows I will make a number of statements about graphs and the sizes of certain processes. In the fourth lecture I discuss these estimates in a little more detail. Students unfamiliar with weak decay physics should to sit down this afternoon and draw all the types of graphs by which one quark can become three (two quarks and one antiquark) and label all the quark lines to see how the various quark-level final states are reached, that is how the three quarks from the b decay plus the second quark of the original B meson combine to give the two final state particles. Try this for at least three or four possible two-body hadronic decay channels for a B meson.

To go from quark-level final states to two body hadronic final states one usually requires that the quarks existing after the b quark decay become the valence quarks of the final hadrons. This is an assumption, it goes along with the idea that the final-state hadronic processes are suppressed. Strong interaction scattering (meson exchange) readily alters the quark composition, particularly that of the light quarks. Quark pair annihilation to a gluon and regeneration with different flavor is also an allowed strong interaction process. Thus one must take some care about interpreting the quark-level diagrammatics too literally. All the rescattering-type corrections to quark-level rules of thumb are comprehended in the general term hadronic effects. For each channel of interest theorists discuss just how well we can calculate or constrain these effects. Once rescattering effects are allowed the distinction between tree and penguin contributions becomes much more murky, as we shall presently see.

3. LECTURE 2: DETERMINING CKM MATRIX ELEMENTS

3.1. Magnitudes from B physics generalities

The Wolfenstein parametrization encodes some of what is known about the magnitudes of CKM matrix elements in terms of powers of the Cabibbo quantity λ, which is well measured. The remaining parameters A , ρ and η (or equivalently $\overline{\rho} = \rho(1 - \lambda^2/2)$, and $\overline{\eta} = 1 - \lambda^2/2)$ enter into b-physics. These are less well known. This lecture is about how, and how well, one can determine them from the magnitudes of the CKM matrix elements in which they enter. An excellent review of this subject is included in the particle data book [12].

In almost all cases the largest uncertainty comes from the theory. Theory uncertainties have two unfortunate features, first they are difficult to quantify, and second they are generally not statistical, so it is not clear that one treats them correctly by adding them in quadrature. (These same two statements apply also to many types of systematic errors in experiments.) This is why results are typically quoted nowadays with three terms in the error: first the experimental statistical uncertainty; second the systematic error that is intrinsic to the experimental measurement; and third the systematic error that arises when one tries to interpret the result as a measurement of some parameter in the theory. The difference between the second and third terms is not always well-defined. For example one measures a particular branching fraction but must impose some cuts to reduce backgrounds. The cuts introduce a theoretical uncertainty already at the level of a branching fraction result because one must use some theoretical input to determine the impact of the cut on the branching fraction. Often a Monte-Carlo model is used, but that model is built using some theory, as well as data inputs. This kind of uncertainty is usually called an experimental systematic uncertainty. When one makes the step of using a measured branching fraction to extract the value of a CKM parameter another set of theoretical inputs are needed. One must calculate the ratio of the branching fraction (or other measured quantity) to the desired parameter. It is the uncertainty in this last step that is usually called the theoretical uncertainty. (It too may depend on what cuts are used in the measurement.)

3.2. V_{cb}

The transition $b \to cl\nu$ can be used to extract the $b \to cW$ coupling and thus the mixing matrix element V_{cb}. There are two ways to approach this, inclusive measurements and exclusive decays to particular channels. As a first guess you might think it obvious that the inclusive hadronic semileptonic decay rate measures the quark level semileptonic decay rate; after all, if the quark decays then the hadron must. This is called quark-hadron duality, or to be more precise local quark hadron duality (as opposed to the more rigorously defendable "global quark hadron duality" that applies for example for the energy-averaged cross-section for $e^+e^- \to$ hadrons) [13].

The fact that this logic is a completely wrong in nuclear physics should make you pause. Try to measure the neutron half-life by studying the beta decays that occur for a lump of iron (Fe_{56}). You would conclude that the neutron is a stable particle. So it is clear that the environment in which a particle finds itself can affect its decay rate; in this extreme example the decay becomes forbidden, due to the exclusion principle, because all lower-energy allowed states for the decay product proton are occupied. The question is then how much does the environment affect the decay for a b-quark in a hadron. Here the naive guess – not much – seems a good one; one would not expect the energetics of the decay to be greatly changed by the confinement of an additional light quark around the heavy b-quark, and certainly no exclusion principle forbids the decay. This argument is compelling but not rigorous. Its biggest flaw is that it does not tell us how to calculate the size of the error we are making in carrying out an approximate treatment of the effect of the environment. How big is "not much"?

The formal treatment that gives some improvement over the naive statements goes by the name of the operator product expansion. The physical idea behind this game is that we can separate the hard or short-distance physics and its time scale from the soft or long distance effects. This is a weak form of the quark–hadron duality assumption; it assumes that the environment is properly accounted for in the matrix elements, and does not affect the short-distance or hard part of the physics. The hard physics gives us a set of local operators (products of fields and their derivatives) that are generated by the quark-level weak decay vertex and the hard QCD corrections to that vertex. At each order in an expansion in $\alpha_S(m_q)$ and Λ_{QCD}/m_q further operators can appear. The coefficients of these operators are calculated from the weak decay and hard QCD corrections to it; these are quark and gluon level diagrams. The soft physics gets lumped into the matrix elements of the operators. These are not perturbatively calculable. However, there are a finite number of such matrix elements that appear in inclusive B decays at each order in Λ_{QCD}/m_q. The hope is that these can be determined by combining several sets of measurements that, according to this theory, depend on the same set of matrix elements.

The same quantities that enter into the meson semileptonic decay rate, weighted integrals over quark distributions in the B meson, also govern the moments of the charmless hadronic spectra seen in the decay $B \to X_s\gamma$. Here X_s denotes any final states containing non-cancelled strangeness. So eventually, with enough data, and assuming the expansions in Λ_{QCD}/m_q and $\alpha_S(m_b)$ converge well enough, we should be able to fit for the leading set of matrix elements and V_{cb} simultaneously. The terms we have dropped via our heavy-quark and QCD expansions can be estimated on the basis of the sizes of the terms that we have kept. Indeed with sufficient data the parameters of the leading and next-leading terms are over constrained, so the basic assumptions of the method can be tested by checking for self-consistent results.

In defining the operator matrix elements and their coefficients we introduced an artificial scale, the division between what we call a hard gluon and what we call a soft one. Both the operator coefficients and the matrix elements depend on this scale. If we could treat both exactly there should be no dependence on it in the final result. The calculations also depend on a second scale, which is the scale at which we choose to renormalize the strong coupling constant. Again this scale should not enter into the final result if the calculation is done consistently, but can do so when approximations are made. Both scales appear in a similar form in the results; they are usually chosen to be equal. Scale dependence appears both because we truncate the perturbative calculation of operator coefficients at some low order, and because we do not have exact methods to determine the matrix elements of the operators.

Additional scale dependence appears because of the dependence of the rate on another unphysical parameter, the mass of the b-quark. Quark masses cannot be directly measured. There are a number of different prescriptions used to define them. These prescriptions can introduce some scale dependence if used in leading or next-to-leading order in perturbation theory. In addition, care is needed to ensure that a consistent definition is used for all parts of an analysis of data.

If you recall the calculation for muon decay, which I hope you have all seen at some point in your education, semileptonic decay rates scale as the mass of the decaying particle to the fifth power. The situation is not quite that bad, as three of the five powers of mass are actually phase-space factors, which scale as the mass difference $m_b - m_c$. This is more readily determined; it is given by the difference of meson masses up to corrections suppressed by Λ_{QCD}/m_c.

The true answer for any physical parameter should not depend on the artificially-introduced scales at all. However, all calculations yield results that do have some scale dependence because approximations are made that do not correctly treat these details. What scale should we choose to define the answer? What theoretical uncertainty arises because of this choice? The usual prescription in b-decay physics is to say the right scale is the mass of the b quark, since this sets the physical scale of energy release in the problem. The uncertainty due to scale-dependence is typically estimated from the amount the answer changes as the scale is varied from $m_b/2 \leq \mu \leq 2m_b$. Clearly this prescription is quite arbitrary! Fortunately, scale-dependence is much decreased when higher-order QCD effects are calculated. Then there is a range of choices for the scale around m_b over which the result is quite stable. It is generally assumed that this gives the correct scale-independent result with small uncertainty.

The upshot for all this is that the extraction of V_{cb} from inclusive data has an uncertainty of order 10% when only the total rate and the leading order perturbative calculation are used. This is the story as I told it in Greece. More recent work includes calculation of the higher order perturbative corrections and the use of moments of the hadronic spectral distribution to determine the non-perturbative parameters (the operator matrix elements that appear up to the first two powers of the Λ_{QCD}/m_q expansion). The parameters so determined include m_b. This approach shows promise of giving a very accurate value (less than 5% uncertainty) for V_{cb}. However as yet not all the details of this method are fully worked out and understood. At present there appear to be some inconsistencies between details of data and theory. More work is needed to clarify this situation [14].

The alternative determination of V_{cb} comes from an exclusive decay to $D^*l\nu$ or $D\, l\nu$. In this case the fact that both the b and c quarks are massive on the scale of Λ_{QCD} gives us a very nice situation. (For the moment let us put aside the concern that the charm quark is not really so very massive on this scale, and talk as if this is a good limit.) In this limit both the B and the D mesons can be pictured as a massive static quark around which the light quark is located in a distribution with a size (and thus a light-quark momentum) scaled by Λ_{QCD}. We don't know a lot about this distribution; so it is often referred to as "the brown muck". However we do know that QCD is flavor blind, so, up to terms of order Λ_{QCD}/m_q, the light-quark distribution is independent of which massive quark is at its core. Indeed, it is also independent of the spin orientation of the massive quark, so, in the heavy quark limit, it is the same for the B, B^*, D and D^* mesons. Now consider the weak $B \to D(orD^*)l\nu$ decay at the kinematic point where the D-type meson is at rest in the B rest frame. The rate is the quark decay times the matrix element of the operator between the meson wave functions. But the heavy quark limit tells us that the meson wave functions are identical at this particular kinematic point! So we know the matrix element. The operator simply switches the core quark type (and perhaps also its spin) and the wave function overlap is 1. Of course there are corrections to this statement, for finite mass quarks. It turns out that for the transition $B \to D^*l\nu$ the corrections begin at order $(\Lambda_{QCD}/m_q)^2$, while for $Dl\nu$ the first correction is of order Λ_{QCD}/m_q [15]. This makes the $D^*l\nu$ decay a particularly good way to fix the parameter V_{cb}, since even for the charm quark the second-order correction is small. In addition there are calculable perturbative QCD corrections.

There is a catch however. The situation in which the leptons carry off all the energy of the $b \to c$ transition is clearly a kinematic endpoint. The cross sections vanishes at this point, because of phase space factors going to zero! So, in actuality, one must measure at some distance from this end point and then extrapolate to it. This introduces some theoretical uncertainty in the relationship between the measurement and the parameter V_{cb} because we must postulate and fit how the wave-function overlap changes as we move away from the known end-point. The uncertainty from this fitting can be reduced if the measurement is made closer to the end-point, but of course the rate is smaller there. So there is an interplay here between theoretical uncertainty and statistical uncertainty. In such a case more data can shift the result to smaller uncertainty. Currently the accuracy obtainable by this method is also at the $5-10\%$ level, with the range depending on how one combines various non-statistical sources of error.

3.3. Determinations of V_{ub}

The situation for the parameter V_{ub} is in principle quite similar to that for V_{cb}; one can pursue either an inclusive or an exclusive semi-leptonic measurement to fix this quantity. Additional difficulties arise in both cases.

In the inclusive case the problem is to discriminate the $b \to u$ decays from the much more copious $b \to c$ decays (including the effects where the c-quark decays to a d-quark, so no strange particles flag its presence). This requires kinematic cuts to exclude the region reachable via a charm quark decay. Then one must determine what fraction of the $b \to u$ events is excluded by this cut. This determination depends on theoretical modelling of the spectrum. At the quark level the spectrum is readily calculated, but the hadron level spectrum has a different end-point and there are also noticeable effects from hadronic resonant states near the end-point. These do not appear in the quark-level calculation.

As in the case of decays to charm the assumption of quark–hadron duality can be more reasonably applied for a set of a few moments of the spectral function, than for the detailed spectrum itself. The calculation of V_{ub} can be given in terms of such moments. The moments of the quark distribution in the B meson obtained from $B \to X_s\gamma$ can also be used to fix some of the non-perturbative quantities that enter the extraction of V_{ub}. In addition, because of the unseen neutrino, there are several different choices for how to impose the cut to exclude charm decays. One can use the charged lepton momentum, or the hadronic invariant mass, or some combination of these two. Like the inclusive determination of V_{cb} this method is based on the assumption that quark–hadron duality correctly gives the total rate and leading moments of the spectrum, though not the all details of the spectrum. It is difficult to quantify the residual theoretical uncertainty that comes from that assumption. One test is to check whether the result is stable as the choice of cuts is varied. Again this is currently a work in progress; it holds promise for accurate results [16].

For exclusive decays such as $B \to \rho\ell\nu$ or $B \to \pi\ell\nu$ one cannot use the heavy-quark limit to constrain the transition matrix element. The heavy-quark theory suggests that one could use comparison with the corresponding D decays in matched kinematic regions for the transition matrix element, but the Λ_{QCD}/m_c corrections can be large and, at least to date, are not well-controlled. Furthermore data on both the B and D decays are quite limited at present. The alternative approach is to fix the transition matrix element by a lattice calculation. At present such calculations have only been done in the "quenched" approximation, which means that the effect of internal light-quark loops is set to zero. Furthermore, the quark mass used for the light quarks is generally large compared to the physical value, so an extrapolation in that parameter is also needed. Both effects are sources of theoretical uncertainties. Both these issues can be clarified with sufficient computing time available. Methods to treat the quark-loop effects, and the computing power to calculate with lighter quark masses are beginning to appear, and certainly will be developed over the next few years. Perhaps by the time there are sufficient data to give a statistical accuracy of order of 5% for these decays there will also be sufficiently good lattice determinations of the transition matrix elements to give an overall 5% level accuracy for V_{ub}. However that day is at least a few years in the future.

3.4. The third side of the unitarity triangle

Like V_{ub}, the quantity V_{td} is of order λ^3, so the prospect of measuring it directly in top-quark decays, where it must compete with the order 1 leading $t \to b$ decays is remote at best! Instead one must find loop effects that are dominated by top-quarks in the loop and use these to fix the magnitude of V_{td} and likewise V_{ts}. Fortunately these are common, the most readily measured being the mass (and width) differences in the neutral B meson systems, in other words the effects due to B–$\overline{B}$ mixing. These are mediated by diagrams like those of Fig. 3, but now with external b and d quarks for B_d and b and s quarks for B_s instead of the c and u quarks of the D meson case, and internal up-type quarks.

As in the D-mixing case we can look at any one quark line and write the contribution to the loop integrand for this line, summing over all three up-type quarks in the intermediate state. This gives

$$Q(k, m_t, m_c, m_u) = V_{tb}V_{ti}^*D(k, m_t) + V_{cb}V_{ci}^*D(k, m_c) + V_{ub}V_{ui}^*D(k, m_u)\,, \tag{8}$$

where the functions $D(k, m)$ are the quark propagators, and i can be either d or s. Again we use the unitarity relationship Eq. (3) to eliminate the term proportional to $V_{cb}V_{ci}^*$, giving

$$Q(k, m_t, m_c, m_u) = V_{tb}V_{ti}^*(m_t - m_c)D(k, m_t)D(k, m_c) + V_{ub}V_{ui}^*(m_u - m_c)D(k, m_u)D(k, m_c). \tag{9}$$

Since the top-quark mass is so much larger than the others and the typical loop momentum k is large compared to the charm quark mass, the first term dominates (in the case $i = s$ the second term is also CKM suppressed). There are two such quark lines in each diagram, so the dominant contribution to the mixing amplitude is proportional to V_{ti}^2. This quantity multiplies a known coefficient times the matrix element of a local four-quark operator between the B and $\overline{B}$ meson states.

For the B_d system the mass-difference between the two mass-eigenstates is well-measured, so the dominant uncertainty in the extraction of V_{td} comes from the uncertainty in the theoretical calculation of the operator matrix element. Lattice calculations for this quantity are steadily improving, but the resulting theoretical uncertainty is still quite large.

Since V_{ts} is strongly constrained by unitarity, a precise measurement of the B_s system mass difference can give V_{td} from the ratio of B_d to B_s mass differences. In this ratio much of the uncertainty in the matrix element cancels, since, up to the effect of the mass difference between an s and a d quark, the two matrix elements are the same. The correction is an SU(3)-breaking effect, where the SU(3) in question is the flavor symmetry of the three light quarks. Naively one might expect this effect to be of order $(m_s - m_d)/m_b$, but care must be taken to ensure that any $(m_s - m_d)/\Lambda_{QCD}$ corrections are properly included. (There are also calculable perturbative QCD corrections; these are well understood.) For the B_s only a lower limit on the mass difference has so far been established [17]. It is hoped that the next round of experiments at the TeVatron will yield an actual value for the B_s mass difference. Even the lower limit currently available significantly improves the constraints on V_{td}.

The three measurements V_{cb}, V_{ub} and V_{td}, complemented with V_{us}, are in principle sufficient to determine the unitarity triangle, but the uncertainties in their values at present are quite large. In particular, if these measurements were all we had, it would be difficult to say with certainty that the CP-violating parameter is non-zero in the Standard Model. Of course, we know from observation of CP violating effects in both K and B decays that this is the case. The K decay result gives a constraint on a combination of ρ and η. The constraint has a large theoretical uncertainty, but excludes $\eta = 0$ which would give a vanishing rate for $K_L \to \pi\pi$. The theoretical uncertainty arises from the matrix element for the K-mixing operator, which is calculated on the lattice. It would however provide a very nice test of the Standard Model to see the result $\eta \neq 0$ appear from measurement of the CP conserving quantities only, and consistent with the result given by combining the CP violating measurements in K and B decays. All that takes is an improvement in the theoretical uncertainties, reducing them in all cases to the few percent level. The current progress, both in lattice calculations and in applications of the moment method of operator-product calculations is certainly bringing that day closer.

3.5. Determining the unitarity triangle from K decays alone

Another way to test the Standard Model predictions would be to determine the unitarity triangle parameters in two separate sets of processes, the B decay processes discussed above and then, independently, from K decay processes. As I stated earlier, in general there are large long-distance hadronic physics effects in K decays, so the relationship between measurement and parameters is fraught with large theoretical uncertainties in many cases. Nevertheless one can search for cases where the relationship is somewhat cleaner, and use these measurements to fix the values of ρ and η and thus the shape of the unitarity triangle [18].

The CP violation in the decay $K_L \to \pi\pi$ is one such quantity. This decay gives the parameter ϵ_K which defines how the kaon mass eigenstates differ from CP eigenstates. This can be calculated as a function of CKM parameters times an operator matrix element as was described for D-mixing above. The matrix element can be evaluated using lattice calculation. The result is that the known value of ϵ_K defines a broad allowed band in the ρ–η plane and clearly excludes the value $\eta = 0$. This result is usually presented in combination with the B-decay results as discussed above.

Further constraints can, at least in principle, come from the measurements of certain rare K decays. The decay $K_L \to \pi^0\nu\overline{\nu}$ is a CP-violating effect that is directly proportional to η in the Standard Model. The problem is that the proportionality constant is so small that the rate has yet to be measured; the prediction is well below current limits. (Of course, if we are talking about testing the Standard Model, rather than defining the unitarity triangle, the search for this decay provides an interesting window for new physics effects.) The decays $K^+ \to \pi^+\nu\overline{\nu}$ and $K_L \to \pi^0\nu\overline{\nu}$, once both are well-measured, can be combined to give a determination of ρ and η albeit with significant residual theoretical uncertainties.

Eventually the match between the values obtained from B-decays and those from K decays can provide a test of the Standard Model, but it seems to me unlikely that the theoretical uncertainties that obscure this comparison can be reduced much below the 10% level.

4. LECTURE 3. CP VIOLATION

"... I would like to have the asymmetry between positive and negative electricity in the laws of nature (it does not satisfy me to shift the empirically established asymmetry to one of initial conditions)" Wolfgang Pauli, in a letter to Heisenberg, June 1933.

This remarkable quote from Pauli shows he felt that matter–antimatter asymmetry in the equations, the asymmetry we now know as CP violation, is preferable to an initial condition for understanding the matter–antimatter asymmetry of the Universe. Pauli aside, it seems that most physicists accepted a complete symmetry in the laws of physics between those for matter and those for antimatter as a natural condition of their theories until the empirical discovery that this could not be true with the observation of the two pion decay of the long-lived neutral kaon (the supposed odd-CP eigenstate).

Now, almost forty years later, not only do we have a theory that accommodates CP violation, namely the three-generation Standard Model, but also we have observations of new CP-violating effects in B meson decays, and the expectation that further such effects will soon be observed. This puts us in the exciting position of being able to test whether the observed patterns of CP-violation fit the tightly-constrained predictions of the Standard Model.

In this lecture I will first develop a general discussion of why and how CP-violation occurs in field theories, and then review what we now know, and what further tests are likely. I have already stated that CP violation arises when there is a phase difference between two couplings in the theory that cannot be removed by any set of phase redefinitions of the fields. But why do complex couplings give CP violation? The basic answer is that they set up a situation where two amplitudes which interfere with one another in contributing to the same process do so with opposite sign for a decay and the CP-conjugate decay – leading to CP-violating rate differences. We will see that such phase differences of couplings can also lead to a situation where there is no choice of phase conventions under which certain mass eigenstates can be defined as CP-eigenstates. Either of these two effects is an explicit violation of CP symmetry.

4.1. Types of CP violation

In generality, we can define three types of CP violation. The first is a CP-violating difference in the magnitude of the amplitude A for any process and the amplitude $\overline{A}$ for the CP-conjugate process (and thus a difference in the rates). This can occur for both charged and neutral particle decays. It requires that there are two non-zero phase differences, both in the phases of the Lagrangian couplings, and in the absorptive parts of the two amplitude contributions. The Lagrangian phases are generally called "weak phases" because, in the Standard Model, they arise in the weak interaction parameters, namely in the CKM matrix. The absorptive phases are called "strong phases" because they arise from rescattering effects that are dominated by the strong interactions. Strong phases occur because multiple channels, as labelled by particle content, are mixed by the strong interactions. As you should remember from whenever you learned scattering theory, such mixing can make the transition matrix element for a given channel complex.

The rate difference can be seen by the following algebra. Let the amplitude for the decay of interest be

$$A = A_1 e^{i\phi_1} e^{i\delta_1} + A_2 e^{i\phi_2} e^{i\delta_2} \, , \tag{10}$$

where the A_i are real amplitudes, the ϕ_i are the coupling constant phases (weak phases), and the δ_i are the phases from absorptive parts in the amplitude (strong phases). Now the CP conjugate amplitude is given by

$$\overline{A} = A_1 e^{-i\phi_1} e^{i\delta_1} + A_2 e^{-i\phi_2} e^{i\delta_2}. \tag{11}$$

The weak phases reverse sign between A and $\overline{A}$ because the CP-conjugate rate is governed by the complex conjugate couplings. The strong phases, however, stay the same as before, because whatever absorptive parts contribute them are matched by the CP-conjugate absorptive parts. One then readily sees that the difference in rates is given by

$$|A|^2 - |\overline{A}|^2 = 2A_1A_2 \sin(\phi_1 - \phi_2)\sin(\delta_1 - \delta_2)\ . \tag{12}$$

Clearly then such a CP-violating rate difference requires that both the weak and strong phases are different for the two terms in the amplitude. This is called direct CP violation, or CP violation in the decay amplitude.

Note also that since the terms that gave the two phases ϕ_1 and ϕ_2 contribute to the same rate they must correspond to the same overall set of quark fields for the external particles of the diagram. Thus any rephasing of the fields changes them both in the same way; the difference is rephasing invariant. (This must be true for any physically observable quantity.) Direct CP violation has been a topic of considerable theoretical and experimental attention, chiefly because it distinguishes between a class of theories called "superweak" that predict no such effect, and all others. However, because of its dependence on strong interaction phases, which are notoriously difficult to calculate in most situations, it is generally very difficult to use an observation of direct CP violation to pin down theoretical parameters or otherwise test the Standard Model theory. One exception is obvious, those cases where the Standard Model predicts no, or very small, direct CP violation effects. In such cases observation of significant direct CP violation would be a clean signature that some new physics effect is contributing to the amplitudes. Hence such channels are important to identify and to study.

The remaining two types of CP violating effect are peculiar to the systems of neutral but flavored pseudoscalar mesons, $K^0, \overline{K}^0$; $D^0, \overline{D}^0$; $B^0, \overline{B}^0$; and $B_s, \overline{B}_s$. In each case there are two distinguishable quark flavor eigenstates. Let us use the notation P^0 and $\overline{P}^0$ to denote any such pair of particles, with the phase convention chosen so that $(CP)P^0 = \overline{P}^0$. If CP were an exact symmetry then the two CP-eigenstates $(P^0 \pm \overline{P}^0)/\sqrt{2}$ would have to be the mass eigenstates. Note that if a mixing amplitude between P^0 and $\overline{P}^0$ would exist, as we showed for the case of the D mesons in Lecture 1, it can give both a mass and width difference between these two states. In general the mass eigenstates can be written as

$$P_{L(H)} = pP^0 \pm q\overline{P^0} \tag{13}$$

with the constraint $p^2 + q^2 = 1$. The subscripts H and L denote the two mass eigenstates – H for the heavier and L for the lighter, with mass difference ΔM. A second type of CP violation occurs if $|q/p| \neq 1$. In that case it is clear that the mass eigenstates cannot be the CP eigenstates. This is called CP violation in the mixing. It is seen in the kaon system where $q/p - 1$ is of order ϵ_K. A similar effect is expected also for the B_s system; in the B_d system it is very small.

The third type of CP violation can occur even when $|\overline{A}/A| = 1$ and $|q/p| = 1$. It occurs for decays of neutral pseudoscalar mesons to a CP eigenstate f. Such a state is accessible both from decay of P^0 (with amplitude A_f) and that of $\overline{P}^0$ (with amplitude $\overline{A}_f = \eta_f \overline{A}_{\overline{f}}$ where $CP|f >= |\overline{f} >= \eta_f|f >$). The CP quantum number $\eta_f = \pm 1$ depends on the particular state f under study. A particle that is (somehow) known to be P^0 at time $t = 0$ can decay to f either directly, or by first mixing to $\overline{P}^0$ and then decaying to the final state. These two paths can interfere to give a CP violating effect, a difference in the (time-dependent) rate for an initial P^0 and that for an initial $\overline{P}^0$ to produce the state f. Let us define

$$\lambda_f = \frac{q}{p}\frac{\overline{A}_f}{A_f}\ . \tag{14}$$

As will be shown later, one finds a contribution to the rate difference that is proportional to $|A_f|^2 \sin(\Delta M t)\text{Im}\lambda_f$. This third type of CP violation occurs whenever the weak phase of the decay amplitude is different from the weak phase of the mixing amplitude that is when $\text{Im}\lambda_f \neq 0$. (In addition there is a possible direct CP violation contribution to this rate difference, proportional to $|A_f|^2 \cos(\Delta M t)(1 - |\lambda_f|^2)$.) The quantity λ_f appears over and again in discussions of CP violation in neutral B decays, so it is worthwhile remembering its definition.

What final states are CP eigenstates? Obviously the quark content of the state must be CP self conjugate, every quark matched by its own antiquark type. However that is not sufficient. In general any such state can include a mixture of CP-odd and CP-even states. One clear exception is when the decay is to a two-body or quasi- two-body state of definite orbital angular momentum. (Here quasi-two-body simply means any state dominated by two particle contributions, even if one or both of the two particles is an unstable hadron). If at least one of the two particles has spin zero then the orbital angular momentum must match the spin of the other particle. Then the state is a CP eigenstate; some examples are $J/\psi K_S$, $\pi^+\pi^-$ and $\rho^0\pi^0$. For two higher-spin particles one can often separate out contributions of even and odd relative angular momentum (and thus those of even or odd CP) by angular analysis of the decays of one of the particles [19]. For multi-hadron final states it is more difficult to isolate contributions of definite CP. This is why much of the attention in searches for CP violation in B-decays is on the two-body or quasi-two-body final states.

This third type of CP violation is particulary interesting in the case $|\lambda_f| = 1$. Then the imaginary part of λ_f directly measures the phase difference between the mixing and the decay amplitudes, a quantity that is cleanly predicted in the Standard Model. In this case the magnitudes and strong phases of the decay amplitudes do not enter the asymmetry result (the difference of rates divided by the sum), so there are no hadronic physics uncertainties in extracting CKM phases from such a measurement.

The case ψK_S (where ψ stands for any $c\overline{c}$ resonance, including the η_c type) is an example of this type; in this case the asymmetry is proportional to sin2β where β (or ϕ_1) is the bottom right-hand corner of the b-decay unitarity triangle, the angle between $V_{cb}V_{cd}^*$ and $V_{tb}V_{td}^*$. This will be explained in more detail a little later.

All three types of CP violation have been observed. In K decays the quantity ϵ' is the first type – direct CP violation or $|\overline{A}/A| \neq 1$. The quantity $\text{Re}\,\epsilon$ measured in the decays $K_L \to \pi\pi$ is the second type, CP violation in the mixing or $|q/p| \neq 1$, and the asymmetry in B decays to ψK_S (and ψK_L) is of the third type, asymmetry due to interference between decay channels with and without mixing, or $\text{Im}\lambda_f \neq 0$. For an up-to-date review on current and prospective tests of the Standard Model via CP violation see [20].

4.2. Formalism for B decays

Let us now examine this general situation more specifically for the case of B decay, since this is the topic of much current and future experimental attention. I will develop the formalism in some detail for the case of the B factories, that is, for an asymmetric e^+e^- collider running at the Υ_{4s}.

To dispel any thought that this is the only interesting way to do B physics I first make a few comments on the case of hadron colliders. The two types of experiments have different advantages and disadvantages; both are needed to carry out the full program of B physics measurements. One can produce many more B-mesons per hour in a hadron collider, but along with them one also produces many other hadrons, and many more events that contain only other hadrons. The question that must be considered for each decay mode is whether one can devise a way to trigger on the events of interest, separate the particles produced in the B decay from other hadrons, and from backgrounds that fake a B event, and tag the initial flavor of the B meson. Each of these steps is somewhat more difficult, and less efficient, in the hadronic environment. How much more difficult depends on the mode in question. However since one is starting with a much higher production rate, lower efficiencies can be acceptable.

In addition, all types of b-hadrons are produced, so a hadronic collider can study processes that are inaccessible at an e^+e^- collider-based B factory, which makes only B_d type mesons when running at the Υ_{4s} resonance. Most importantly, the B_s mesons are not accessible to the current B-factories, and their decays are as interesting for testing the Standard Model as those of B_d mesons. Conversely, the mode $B \to \pi^0\pi^0$ is important and cannot be readily studied except at the e^+e^- B factories. [These are examples to show why both approaches are needed; they are not the only cases.]

4.3. *B* decay formalism

Let us define $M = (M_H + M_L)/2$, and $\Delta M = M_H - M_L$, and similarly for Γ and $\Delta\Gamma$, where the subscripts H and L denote the heavier and lighter mass eigenstates respectively. Another warning about conventions is in order here; there are, unfortunately, two of them floating around. With the convention defined above ΔM is obviously positive, however the sign of q/p is a physical quantity to be explored. The other convention labels the two states as 1 and 2 and defines q/p to be positive (where 1 is the state with +q, and 2 has $-q$ in the superposition Eq. (13)). In this alternate convention the sign of ΔM is a priori undefined.

Now, using the first convention, let us define the states $B^0(t)$ ($\overline{B}^0(t)$) as the time-dependent superposition of a B^0 and a $\overline{B}^0$, (or, equivalently, of a B_H and a B_L) which at time $t = 0$ is a pure B^0 (or $\overline{B}^0$ respectively)

$$\begin{aligned} B^0(t) &= g_+(t)\, B^0 + \frac{q}{p}\, g_-(t)\overline{B^0}\,, \\ \overline{B^0}(t) &= \frac{p}{q}\, g_-(t)\, B^0 + g_+(t)\overline{B^0}\,. \end{aligned} \tag{15}$$

The functions $g_\pm(t)$ can readily be found by writing the state $B^0(t = 0)$ as a superposition of B_H and B_L and then allowing that state to time evolve. A little algebra gives

$$\begin{aligned} g_+(t) &= \frac{1}{2}\, e^{-(\Gamma t/2)} e^{iMt} \\ &\quad \times \left\{ \cos\frac{\Delta M t}{2}\left(e^{-(\Delta\Gamma t/4)} + e^{+(\Delta\Gamma t/4)}\right) + i\sin\frac{\Delta M t}{2}\left(e^{-(\Delta\Gamma t/4)} - e^{+(\Delta\Gamma t/4)}\right)\right\} \\ &\to e^{-(\Gamma t/2)} e^{iMt}\cos\frac{\Delta M t}{2} \text{ for } \Delta\Gamma = 0 \end{aligned} \tag{16}$$

and

$$\begin{aligned} g_-(t) &= \frac{1}{2}\, e^{-(\Gamma t/2)} e^{iMt} \\ &\quad \times \left\{ i\sin\frac{\Delta M t}{2}\left(e^{-(\Delta\Gamma t/4)} + e^{+(\Delta\Gamma t/4)}\right) + \cos\frac{\Delta M t}{2}\left(e^{-(\Delta\Gamma t/4)} - e^{+(\Delta\Gamma |t|/4)}\right)\right\} \\ &\to ie^{-(\Gamma t/2)} e^{iMt}\sin\frac{\Delta M t}{2} \text{ for } \Delta\Gamma = 0\,. \end{aligned} \tag{17}$$

Note that the states $B(t)$ are perfectly well-defined for $t < 0$. What this means physically is that superposition of B^0 and $\overline{B}^0$ which, if it does not decay, would evolve to be pure B^0 (or pure $\overline{B}^0$) at time $t = 0$. (Note however that the state so defined has norm greater than 1 at $t < 0$ in order to arrive at time $t = 0$ with norm 1. This is a bit artificial, in any real case we normalize the state for any particle at production time and then evolve that state with a decaying exponential.)

In an e^+e^- B factory the initial system is produced in a coherent state which remains exactly $B^0\overline{B}^0$ until such time as one of the particles decays. (Better said, both particles oscillate, but they do so coherently, so that the probability of finding two B^0 particles or two $\overline{B}^0$ particles vanishes at all times, as long as both are present.) However once one particle decays the other continues to oscillate until such time as it decays. If one B decays to a flavor-tagging mode and the other decays to a CP-study mode we have an event that can be used to reconstruct the time dependence of the asymmetry. We find the rate for the production of such events is given by

$$\begin{aligned} R(t_{\rm tag}, t_f) &\propto e^{-\Gamma(t_{\rm tag}+t_f)/2} \left|\overline{A}_{\rm tag}\right|^2 |A_f|^2 \\ &\times \left\{ \frac{1+|\lambda_f|^2}{2} \mp \cos\,\Delta m(t_f - t_{\rm tag}) \left(\frac{1-|\lambda_f|^2)}{2}\right) \pm\, \sin\,\Delta m(t_f - t_{\rm tag})\,{\rm Im}\,\lambda_f \right\} . \end{aligned} \tag{18}$$

The CP asymmetry for a final state f is thus

$$a_f = \frac{R(B_{\rm tag}) - R(\overline{B}_{\rm tag})}{R(\overline{B}_{\rm tag}) + R(B_{\rm tag})} = \frac{-\cos(\Delta M t)(1-|\lambda_f|^2) + 2\sin(\Delta M t){\rm Im}\lambda_f}{1+|\lambda_f|^2} . \tag{19}$$

In this last equation we have set $t = t_f - t_{\rm tag}$. In an asymmetric B factory we can measure this time from the physical separation of the two B-decay vertices, since the pair is produced with known large momentum in the direction of the higher-energy beam. (Note that this time difference is negative when the tagging decay occurs later than the CP-eigenstate decay.) One CP-violating term survives in the case $|\lambda_f| = 1$. It is proportional to an odd function of time, and hence would vanish if one were to integrate over all times. This quantity is particularly interesting because it gives us a result that directly measures the difference of weak phases of the mixing and decay terms, and thus the relative phases of certain CKM matrix elements, with no uncertainties from hadronic physics effects.

An example of this type is the decay $f = \psi K_S$. At the quark level this is a $b \to c\bar{c}s$ decay. There are both tree graph and penguin graph contributions that give this quark content. The tree graph CKM coefficient is $V_{cb}V_{cs}^*$. We can use unitarity to write the penguin graph contributions as two terms, one with the same CKM coefficient as the tree graph and the other proportional to $V_{ub}V_{us}^*$. This term is small, it is both a penguin only term and furthermore suppressed by λ^2 compared to the other (tree+penguin) term. Thus for this channel $|\lambda_f| = 1$ at the level of few percent accuracy. For the two decay paths to interfere we need both a B mixing and a K mixing. Thus the phase measured here is

$$\begin{aligned} -\arg(q/p)_B \arg\frac{\overline{A}_{(\psi K_S)}}{A_{(\psi K_S)}}\ \arg(q/p)_K &= -2\arg V_{tb}^*V_{tb}\arg(V_{cb}V_{cs}^*)\arg(V_{cs}V_{cd}^*) \\ &= -2\arg V_{tb}^*V_{tb}V_{cb}V_{cd}^* = 2\beta . \end{aligned} \tag{20}$$

This decay is a particularly attractive one to study. It is clean both theoretically and experimentally, a rare situation! The channel is readily recognized, for example by the two-lepton decay of the ψ-type resonance and the two-charged-pion decay of the K_S.

The results from both BaBar and Belle now show a clear CP violation in this channel. The extracted value for sin(2β) is in good agreement with the value given by measuring the sides of the triangle. Figure 4 shows these results, and that from CP violation in K decays, as well as the allowed regions for the apex given by measuring the sides. The figure is taken from the Particle Data Group report on the CKM Triangle [12]. That report describes in detail what measurements have been used for each quantity in this figure. You can see that there is a common region for the apex that is consistent with all these measurements. The Standard Model has survived yet another test!

It will take some years more work on B decay physics to complete the next set of tests to a comparable level of accuracy. There are many interesting channels to study. Few of them have sufficient statistical accuracy as yet to give refined tests of the theory. Modes where a similar analysis predicts no CP violation because the decay weak phase cancels the mixing weak phase provide a good test of the theory. This is the case for example, up to few percent corrections, for the channel $B_s \to \psi\phi$. A large observed CP violation this channel would be a clear indication of physics beyond the Standard Model.

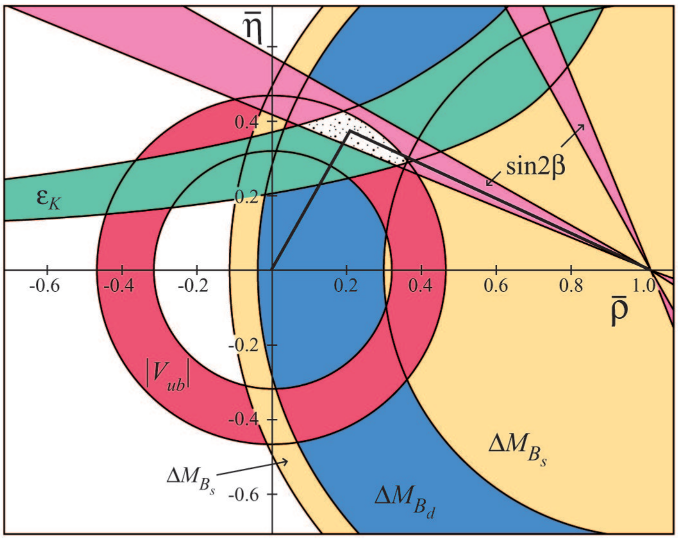

Fig. 4: Figure taken from Particle data group review, showing the consistency of all measurements with a single choice of CKM parameters.

However the existing B factories cannot study it, so this and other B_s results await a good B-physics detector at a hadron collider. Another interesting case is $B \to \phi K_S$, here the argument is similar to that for ψK_S, except that there is no tree diagram contribution. This makes the ratio of the dominant and sub-dominant terms smaller, but the subdominant term is still suppressed by λ^2 compared to the dominant one. Ignoring the sub-dominant terms, the asymmetry in this channel should be the same as that for ψK_S. As yet statistics is poor, the results seem to disagree with that prediction at about the two standard deviation level [21]; further statistics is needed to clarify this situation.

As an example of a case where $|\lambda_f| \neq 1$, consider the channel $B \to D^+D^-$. Here the b quark must decay to give $c\bar{c}d$. This can be done with either a tree diagram, which enters with CKM coefficient $V_{cb}V^*_{cd}$, or via a W-loop transition $b \to d$ where a gluon or Z-boson emitted from the loop creates the $c\bar{c}$ pair. Inside the loop the quark line can be either a top, charm or up quark, giving three terms. Once again we use the unitarity condition to rewrite the coefficient of one of these terms as the negative of the other two, thereby grouping the three terms into two. In this case none of the three terms is small, so it is quite arbitrary which one we eliminate. I choose to keep one that has the same form as the tree term, and one that has the same weak phase as the mixing diagram. Thus I eliminate $V_{ub}V^*_{ud}$ to get a contribution from the loop diagrams of the form

$$V_{cb}V^*_{cd}[f(m_c) - f(m_u)] + V_{tb}V^*_{td}[f(m_t) - f(m_u)]\,, \tag{21}$$

where the function $f(m_i)$ denotes the result of the Feynman loop-integral with an internal i-type quark (with $i = u$, c, or t). The final amplitude thus has two terms that have different weak phase factors because they have different CKM coefficients; these two terms also possibly have different strong phase factors. The operators that enter for the tree term and the penguin term are different. The combination of those operators that appears is different for the two terms with different CKM factors. Thus there is no reason why the strong phases of those two amplitude contributions must be the same. This means that we have a situation for which $|\overline{A}_f/A_f|$ is not necessarily unity, i.e. there can be direct CP violation in this channel. The argument of λ_f in such a case is not given directly by the phases of the CKM factors that enter. One can write the measured quantity proportional to sin($\Delta M t$) as

$$Im\lambda_{f=D^+D^-} = Im[(q/p)_B\overline{A}_f/A_f] = |\lambda_f|\sin(2\beta + \theta_f)\,. \tag{22}$$

The magnitude of λ_f can be measured from the coefficient of cos(ΔMt), but the unknown angle θ is not determined by this measurement. It depends on the relative magnitudes and strong phases of the two amplitude contributions. The uncertainty in the size of θ_f must be accounted for in assigning the theoretical uncertainty to the value of β extracted in this way. One can argue that the tree contribution dominates over the penguin one, in which case $|\lambda_f|$ should be close to unity and θ_f should be small. Even if the measurement agrees with the first of these statements this does not guarantee the second is correct. We need further inputs to give us a more reliable numerical constraint on θ_f.

There are many other channels where this pattern of tree plus penguin amplitudes leaves us with a shift between the measured phase quantity and the CKM phase differences that we want to evaluate. In some cases we can use further inputs, measured in other related channels, to determine or constrain these shifts. Tomorrow's lecture will focus on calculational methods. We will discuss some examples that show how symmetries of the strong interactions can be used to relate measurements for different channels and thereby constrain theoretical uncertainties such as the value of the shift θ_f in some cases.

5. LECTURE 4: THE THEORISTS' TOOLKIT

Up until now I have not discussed any details of what theorists calculate and how the calculations proceed. Clearly in this one remaining one lecture we can only brush the surface of that topic. My intention today is that you become familiar enough with the jargon of theory for B decays that you can tell when theorists are talking about solid calculations and when they are talking about estimates. Even when you think you know that much, you should always ask any theorist to try to quantify the uncertainty in his or her result; and then you should go ask several more theorists to give their assessment of the uncertainty in that particular calculation. Human nature comes into the game, because there is seldom any rigorous way to evaluate theoretical uncertainties once we are dealing with hadronic physics effects. It is human nature for people working on a problem for a long time to believe that the work they have done is definitive, and hence to tend to underestimate the uncertainties in it. However it is just as true that people who have thought little about the same problem will tend to be skeptical about the work they have yet to examine and understand and so perhaps overestimate the uncertainties in it. The only way we can get out of this bind is to apply the method in enough cases that we learn from empirical results how well it works in general. Then if we see a discrepancy much larger than expected we can begin to investigate it as a possible evidence for new physics. As you can see this is dangerous ground, logically speaking. What if the evidence we start with in testing how well the theory works has new physics effects in it, how could we tell? Unless the effect is clearly larger than our uncertainties, we cannot. Over and again we meet this issue. There is no way out except for theorists is to try to quantify uncertainties as reliably and as honestly as possible. As soon as any apparent discrepancy arises, everyone will re-examine the possible theoretical uncertainties. Only when we have solid methods to constrain the theoretical uncertainties can any discrepancy tell us about new physics. Frustrating as it is, that is the world in which we live.

5.1. Terminology for diagrams

Figure 5 shows a large collection of diagrams for two body decays of a meson. These diagrams are not really Feynman diagrams, they are a hybrid picture where the weak interaction and possibly one or two hard gluons are drawn as a Feynman diagram and the way valence quarks combine to make hadrons is indicated by a circle connecting a quark and antiquark line. Of course there are many soft gluons and quark-antiquark pairs that can contribute that are not indicated in these diagrams. In addition I have drawn dotted loops to indicate color index connectivity within each diagram. The same Feynman diagram at the quark level can become more than one diagram in Fig. 5.

Let us examine first diagrams (a) and (b). At the quark weak-interaction level these are both the same, the b-quark to become some up-type quark (u, or c) and the W then produces an additional quark (down type) and antiquark (u or c). (That is for decay of a $\overline{B}$ meson which contains, by the peculiarity of

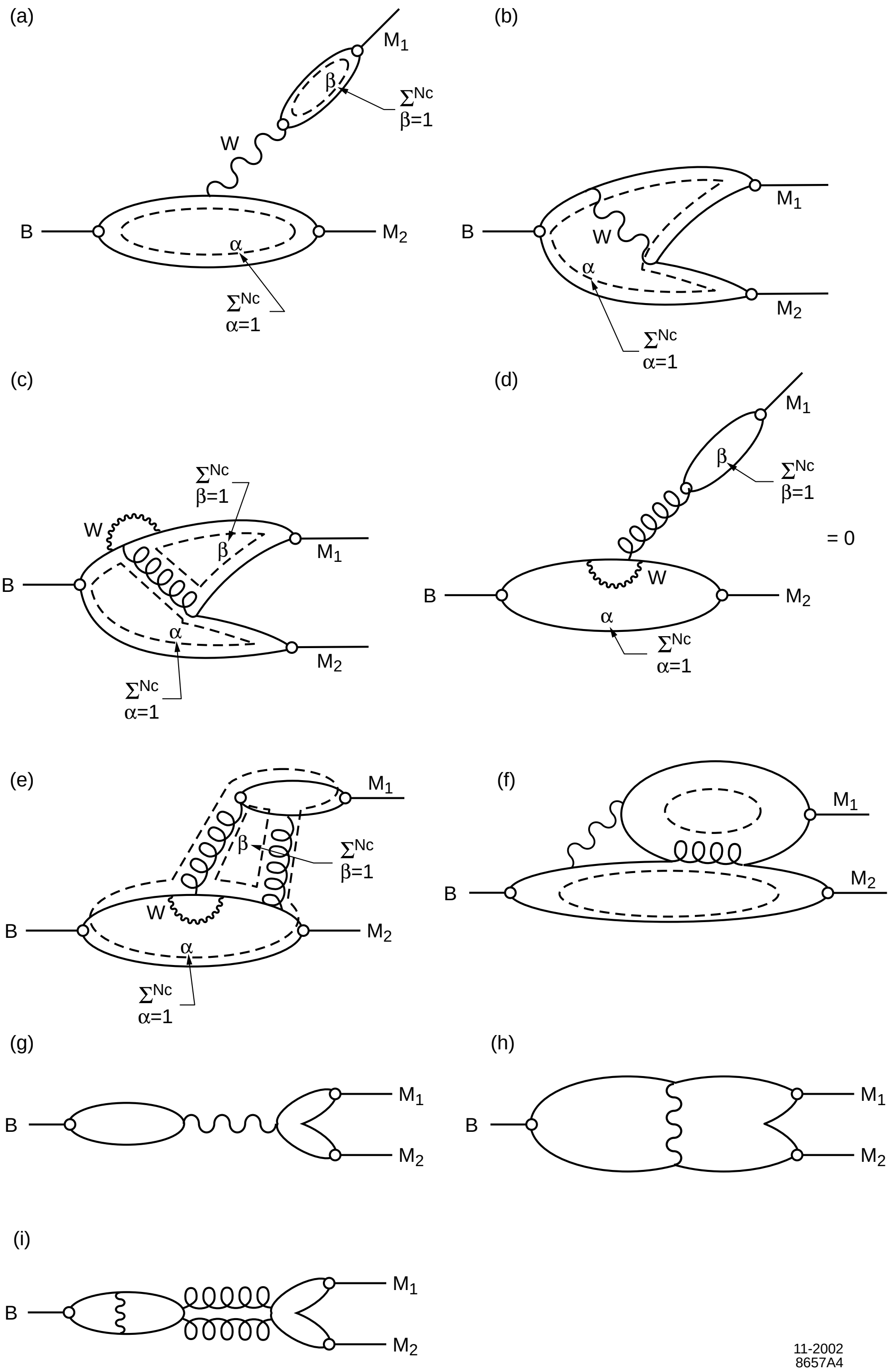

Fig. 5: Diagrams for two-body B decays.

our conventions, a b-type quark. For a B meson, you can follow the same pictures, but reverse the roles of quark and antiquark.) That topology for the weak decay makes both these diagrams tree contributions. However they differ in the topology of how the quarks combine in the final mesons. Depending on the particular final state one or the other or both of these diagrams might contribute to the amplitude.

In diagram (a) the quark and antiquark coming from the W end up in the same final meson. This is called a color-allowed tree contribution, because the color singlet nature of the W completely separates the color flow loops of the two parts of the diagram, so one sums over possible colors twice in this diagram, once for each color-loop. In diagram (b) the quarks from the W end up in two different final mesons. This has the consequence that there is only a single color-flow loop. Thus there is only a single sum of colors in this diagram. Since that gives diagram (a) a color advantage over diagram (b) of one factor of N_C (the number of terms in the color sum) one expects contributions of type (a) to be larger than those of type (b) by something like a factor of N_C. Thus the second diagram is called a "color suppressed" contribution. This is not a rigorous calculation of the ratio of the two contributions, they are differ by more than the color flow. If one writes the short-distance part of these diagrams as a sum of operators the relative coefficients of the operators are different in the two cases. Since operator matrix elements are not perturbatively calculable, one cannot give any perturbative prescription for the ratio of the two contributions. So the $1/N_C$ color suppression is a rule of thumb for making estimates of expected branching ratios, not a precision prediction.

Diagrams (c), (d), (e) and even (f) all have the common feature that the W-boson is emitted and reabsorbed from a single quark line. Any such diagram is called a "penguin diagram". What can we say about these various pictures? First note that diagram (f) is in fact no different (in Feynman diagram terms) from diagram (c), it is simply a distorted copy of the same diagram which emphasizes the fact that there is a possible time-ordering of this picture which looks like a tree process plus a strong interaction rescattering, and which, for u or c type quarks inside the loop, can have an real intermediate states with different quark content from the final states. Diagram (d) has no color flow loops shown, because there is no consistent way to draw them. A gluon cannot produce a single meson, because a gluon is an octet and a meson is a singlet under the color SU(3). Diagram (d) therefore vanishes for a gluonic penguin. There is a possible contribution that looks like this for an electroweak penguin – that is if I replace the gluon by a Z or a photon. But all I have to do to correct the color problem is exchange another gluon, as shown in diagram (e). Furthermore there is no argument that tells me a priori that this additional gluon is a hard gluon, so I cannot say that diagram (e) is suppressed (at least not without some significant further work.) Thus as far as which quarks end up in which meson, I have two types of penguins just as I have two types of tree diagrams. For penguins there not any simple rule of thumb about which contribution is larger. Further, as I have shown by drawing diagram (f) the conventional separation of tree and penguin is not even so clear when I include rescattering. Of course when the quark in the loop is the same as the final quarks produced by the gluon this part of a penguin diagram enters with the same CKM factors as a tree contribution to that process. If all we are doing is tracking weak phase factors we do not need to distinguish the tree and penguin parts of that term. However we certainly need to recognize that both are there if we want to try to estimate the size of any such contribution. In all these diagrams the light quark in the B meson does not participate in the hard interactions, so it is called the "spectator" quark.

Finally we have a set of possible diagrams where the second quark in the B meson participates in the weak interaction. This are called annihilation (g), exchange (h) and penguin annihilation (i). Diagram (g) is possible only for charged B mesons, diagrams (h) and (i) contribute only for neutral B decays. In all three cases the second quark-antiquark pair needed to make the final mesons is somehow created from the soup of gluons in the final state. This does not require a hard gluon, so I do not draw one, nor do I attempt to illustrate color flow for these diagrams. All three of these diagrams are expected to give very small contributions because they are suppressed by the need for the two quarks in the B meson to come together. (This is so even in the exchange case, since the W is far off shell and hence the weak interaction is effectively point-like.)

To turn these diagrams into rule-of-thumb estimates from amplitudes, the first factor to consider in the size of a contribution is the size of the CKM coefficient that enters. For tree diagrams one then counts whether the contribution is color allowed or color suppressed, with a factor of $1/N_C = 1/3$ for the latter case. Beyond that penguin diagrams are expected to be suppressed relative to any tree contribution because they have both a loop integral (giving a factor $1/16\pi^2$) and a hard gluon (giving two powers of a strong coupling constant, defined at a hard scale of order m_b). Thus such a diagram is suppressed by a factor of $\alpha_S(m_b)/4\pi$ compared to the tree diagram. Finally all the diagrams that directly involve the spectator quark in the hard interaction are suppressed by the probability for overlap of the spectator quark with the heavy quark in the hadron wave function. Any such suppression may of course be partially compensated by kinematic or dynamical factors which are not so readily estimated.

5.2. Operator-product calculations

All these diagrams are useful for two reasons, first because they allow us to keep track of CKM factors that contribute for a given process, and second because they do give us some rules of thumb about relative sizes of amplitudes for various channels. More rigorous calculations replace the weak interaction part of these diagrams as a sum of operators plus coefficients and the hadronization part by a set of operator matrix elements. One then must add the impacts of additional hard gluons which can both rescale the operator coefficients and add additional operators into the picture.

Then the question becomes how to determine the operator matrix elements. If the same matrix element enters in more than one process we may be able to fit for them along with the theoretically interesting parameters such as CKM elements. The ability to do this, even in an approximate fashion, is based on the fact that the decaying quark is massive, a fact that we can use in three ways. We have already mentioned that the massive quark gives us two useful expansions in Λ_{QCD}/m_q and in $\alpha_S(m_q)$, both of which are small parameters. The first of these expansions orders the operators by dimension, with higher dimension operators relatively suppressed by powers of this factor. The second expansion counts hard gluons; here the energy release in the heavy quark decay gives us a meaningful sense of what is meant by a hard gluon in the problem, and thus a way to pick the hard/soft separation scale in our operator-coefficient separation that makes physical sense. Furthermore it gives a suppression of high-dimension operators and hence limits the number of matrix elements that must be determined.

The large mass of the b quark does even more than that for two body decays. It gives us a sensible separation of time scales in the problem that matches the operator product formalism but also goes a bit beyond it. The hard/soft gluon separation can be thought of as a "factorization" of the time scale of short distance processes vs the time scale of long distance or hadronization processes. Intermediate between these scales is another time we could define for a two body process, that is the time that the two final state mesons are close enough (in the B rest frame) to interact strongly. Because they are produced with a large relative momentum in a heavy quark decay, this time scale is short compared to the hadronization time scale.

So what can we learn from that observation? Let us first think about a tree diagram process. Because the weak interaction is essentially a point-like four fermion vertex the two quarks that combine to form the meson that does not contain the spectator quark are produced very close to one another. Only in the configuration where these two quarks happen to move off almost parallel with similar momenta are they likely to combine to form a single meson. All other phase-space configurations for these two quarks will tend to give multiple final state mesons.

Notice, however, that in the case when the quark and antiquark are moving close to parallel they quickly fly far from the "brown muck" of the spectator quark, and the third quark of the b-decay flies rapidly in the opposite direction. As the pair of quark and antiquark travel through space they evolve from a local color-singlet to a color-singlet with the size of a physical hadron. However, because at the time they pass through the "brown muck" they are a *local* color singlet, small in transverse size compared to Λ_{QCD}, they escape with essentially no strong interaction with it [22]. This idea goes by the name of "color transparency" –it says that a local color-singlet system does not interact as strongly as one that is spread out.

In the context of B decays this argument says that final state interactions should be small for color-allowed tree-dominated processes. This then gives a way to determine matrix elements for these processes, because they can be "factorized". For example you might use the semileptonic decay $B \to \pi l \nu$ to fix the transition matrix element for $B \to \pi$ and then set the $W \to \pi$ transition to f_π thus evaluating diagram (a) for $B \to \pi^+\pi^-$. This is called the "factorization approximation" because it ignores any possible final state (hadronic) interactions between the two pions, and also because it factorizes the four quark operator matrix element into two distinct two quark matrix elements.

What about the penguin diagram, or a color-suppressed tree diagram? Again one can write the process in terms of a sum of local four-quark operators. One can always Fierz transform these operators into a form where the two quarks that finish up in one meson are paired and the other quark field is paired with the b quark. However one can only apply the factorization argument if this arrangement also pairs the color-index of the quark and antiquark in the same way. Otherwise the thing that is automatically a local quark-antiquark pair is not automatically a color singlet, in which case the color transparency argument does not apply. For color-suppressed tree diagrams and for penguin diagrams this is generally the situation. Final state interactions effects are sometimes ignored in estimating these terms anyway, but the approximation is less well-justified in these cases.

There are two sets of papers in the literature that go beyond the sort of hand-waving argument I have presented here and try to build these ideas into a predictive formalism for decays to two pseudoscalars, using both the heavy quark expansion, and the operator product expansion with QCD perturbation theory corrections [23]. The treatment gives the factorization result as a leading term, but finds that there are further operators that contribute at higher orders in the expansions. For example, if a hard gluon is exchanged to the spectator quark then this gives a local six-quark operator in addition to the local four quark operators of the leading terms. The coefficient of this operator is suppressed by a factor of Λ_{QCD}/m_b^3 as well as by the factor of $\alpha_S(m_b)$. However differences in the operator matrix elements may make up for these powers of Λ_{QCD}/m_b. For the leading operator the spectator quark is soft but must form a meson by combining with a fast-moving quark from the b-decay vertex. This gives a suppression of the matrix element. With the six-quark operator the two quarks that form each final-state meson can be produced moving together, giving an unsuppressed matrix element. One major difference between the two groups who attempt to apply this formalism is how much they assume about the suppression for the formation of a meson from one soft and one hard quark. This and other important but detailed differences in assumptions gives significant numerical differences in the results. For both groups certain Λ_{QCD}/m_b suppressed contributions turn out to be numerically significant, because they are "chirally enhanced" by factors of pseudoscalar meson mass over the sum of the relevant quark masses. This too reduces the predictive power of the formalism.

The more operators that enter, the more unknown operator matrix elements must be determined. This means that the predictive power of the method is best when it can be seen that the leading few terms dominate the result, and when several different pieces of information can be related. The pieces of information can be rates in channels related by symmetries, which I will discuss later in this lecture, or they can be moments of a spectrum, as described previously in the discussion of magnitudes of CKM parameters.

Unfortunately there are not always a sufficient number of related channels to fix all independent non-perturbative quantities. Furthermore the question of how well the heavy-quark expansion works, that is how dominant the leading terms are compared to formally (Λ_{QCD}/m_b)-suppressed terms, can have different answers for different channels. Thus the power of the formalism must be investigated on a case by case basis. This is still a work in progress. One thing is already clear, the application of this formalism requires close interaction of theorists and experimentalists, to deal with both the complexity of the formalism and experimental issues such as correlated errors in data when evaluating a fit to several moments of the same spectrum.

5.3. Symmetries of hadronic physics

The couplings of QCD are flavor-blind. Thus in hadronic physics the only thing that distinguishes quark flavors are the quark masses and charges. Up until now we have used heavy quark symmetry, which applies for masses large compared to Λ_{QCD}. But there are several "old fashioned" symmetries that apply for quark masses that are small compared to Λ_{QCD}. The best of these is isospin, which is the SU(2) symmetry of interchange of u and d quarks in hadrons. (It is called a spin because of the SU(2) structure was best known to physicists as the mathematics of spin, not because it has anything to do with angular momentum.) Isospin breaking effects come from electromagnetism as well as from the phase space differences and other impacts of the different quark masses. One can also apply a full SU(3) symmetry of interchange among the three lightest quark flavors. Since the strange quark mass is not very small on the scale of Λ_{QCD}, SU(3) breaking effects can be significant. In B physics many interesting results arise from the SU(2) subgroup of SU(3) known as U-spin, symmetry under interchange of s and d quarks. Here there are no electromagnetic breaking effects as the two quarks have equal charges, but the mass effects are still an issue. A third class of symmetries arise from the $m_q \to 0$ limit; these are chiral symmetries which constrain the couplings of soft pseudoscalar mesons. In two-body B decays there are not really any soft mesons, so this symmetry is useful only as a theoretical limit, for example for the process $B \to D\pi$ in the limit that the charm quark mass approaches the b-quark mass. Such limits often provide useful constraints on models for a given process, as any good physical model must incorporate the correct behavior in this limit, as well as in the heavy quark limit. (The chiral limit is also useful in relating multibody decay channels that differ by one pion.)

Isospin is particularly useful when it allows us to distinguish a pure tree amplitude contribution from those arising from penguin diagrams. A pure tree amplitude can have no direct CP violation. As we saw above, CP violating effects involving such amplitudes with no direct CP violation allow us to determine certain combinations of CKM phases. In the decays mediated by the quark process $b \to u\overline{u}d$ isospin can be used to distinguish a pure tree part. For the tree diagram the three quarks arising from the B decay can have total isospin of either $1/2$ or $3/2$. Adding the spectator quark gives final states with isospin 0, 1 or 2. However for a penguin diagram mediated by a gluon the gluon has isospin zero and thus so must the quark-antiquark pair produced by it. Thus this diagram gives only $I = 1/2$ for the three quarks from the b-decay, and only $I = 0$ or 1 when the spectator quark is added. Hence, if we can isolate the pure $I = 2$ amplitudes, we have essentially a pure tree amplitude (up to corrections from Z-penguins) and hence a single CKM coefficient.

This idea has been investigated in detail for the case of $B \to \pi\pi$ [24]. Without isospin analysis the situation is similar to that discussed for D^+D^- above; the measured CP-violating phase has an unknown contribution because there are two different hadronic amplitudes with different weak phases contributing. One finds

$$\mathrm{Im}\,\lambda_{\pi^+\pi^-} = |\lambda_{\pi^+\pi^-}|\sin(2\alpha + \theta)\ . \quad (23)$$

With an isospin analysis one can, in principle, cleanly extract from the data the CKM phase $\alpha = \pi - \beta - \gamma$, the angle that sits at the vertex of the unitarity triangle. The shift θ can be determined, up to a four-fold ambiguity, once one measures the rates for the charged $B^\pm \to \pi^\pm\pi^0$ and the neutral $B \to \pi^0\pi^0$ as well as the rates and CP asymmetries for the $B \to \pi^+\pi^-$ channel.

As always, there are a few caveats. The first is that the same argument does not apply for the contribution of a Z or a photon in the penguin diagram. Both these particles have isospin 1 parts as well as isospin zero. Their contributions are a priori small, except that the Z contribution is enhanced by a factor of $(M_t/M_Z)^2$, so one must check its importance as a correction to the analysis. A small theoretical uncertainty arises from this effect. The second problem is an experimental one, the rate for $B \to \pi^0\pi^0$ is small. One can however use the analysis to bound the unknown shift θ, and thus the theoretical uncertainty in α, even when only an upper limit on this branching ratio is available [25].

An important feature of this analysis is that the isospin of the two pion final state produced from a B decay can only be even, hence all $I_{\rm final} = 1$ contributions vanish. This follows from Bose statistics for two pions in final state of total angular momentum zero. (Note that in the context of isospin we treat two pions as identical particles, independent of their charges, as they differ only by interchanges of u and d type quarks.) The restriction from Bose statistics is crucial as it reduces the number of independent amplitudes to the point where there are enough measurements to determine them (including determining the strong phases between them).

Another piece of old fashioned physics, known as a Dalitz plot, is added to the isospin decomposition of amplitudes when the same three-quark decay of the b gives a three pion final state. If the three pion final state is dominated by resonant (quasi-two-body) contributions, such as $\rho\pi$, then we can use these channels to fix the CKM parameter α. Here we do not have the Bose statistics argument to eliminate the $I_{\rm final} = 1$ amplitudes, so there are more amplitudes and relative phases to determine, though again there are some, with $I_{\rm final} = 2$, that have no gluonic-penguin contribution. There are also more neutral B channels to study, as $\rho^+\pi^-$ is different from $\rho^-\pi^+$.

A Dalitz plot is a plot of the kinematically allowed region for the three pions from the decay of the B^0 (or $\overline{B}^0$). Each recorded event is a point on this plot. Quasi-two-body states show up as dense bands of events, because they are constrained to have two pions with an invariant mass close to that of the resonant state, in this case a ρ meson. Each charge of ρ can be formed, so there are three $\rho\pi$ bands in the plot. The crucial point for the analysis is that we assume that the various isospin amplitudes for $B \to \rho\pi$ are constant over any such band. However the decay of the ρ adds a characteristic Breit-Wigner factor, which gives an additional known complex factor that varies across the plot. The phase of this factor is thus a known (and large) strong phase effect. A further critical point is that there are regions in the corners of the plot where two different charges of ρ are both kinematically allowed; their bands overlap. This means that their amplitudes can interfere in this region. This is crucial to the analysis; there is information about the relative phases of the amplitudes in this interference behavior.

We use unitarity as usual to write the penguin terms as those that have the same weak phase of the tree, which we combine with the same-isospin tree amplitudes, and those that have the same as the weak phase as the mixing in the Standard Model, which we treat as an independent but isospin-restricted set of amplitudes. We do not use the diagrams to calculate anything beyond the CKM coefficient structure relevant for each isospin amplitude. There is (again in principle) enough information in the plot to fix the relative magnitudes and strong phases of all the independent isospin amplitudes and at the same time determine the parameter α, the difference of CKM phases at the vertex of the unitarity triangle. This angle is the difference of weak phases between the $B - \overline{B}$ mixing term and the tree graph for $b \to u\overline{u}d$.

I say "in principle", because such a multiparameter fit requires a lot of data. Further one must consider both backgrounds from non-resonant B decays and non-B backgrounds, and possibly also fit for other resonances, such as f_0, if these contribute significantly to the Dalitz plot. Additional resonances are not background, in the sense that their amplitudes too have the same CKM structure, and their CP properties are well-defined, but they do add more parameters to be fitted. The channel $\rho^0\pi^0$ is essential to this analysis, one must at least be able to detect it via its interference effects. As in the $\pi^0\pi^0$ case, the rate for this doubly neutral channel is expected to be small because it is color-suppressed. So in the real world this analysis will probably need about ten times the data currently available from the B factories to yield reliable results. No matter how good the data, there remain some small theoretical uncertainties introduced by the effects of electroweak penguins, particularly those mediated by a Z.

An example of application of SU(3) symmetry is its use in analyzing B decays to two pseudoscalars. Here the U-spin subgroup of SU(3) that comes into play, relating the penguin contributions to $B \to \pi\pi$ to those that dominate the decay $B \to K\pi$. Notice that the SU(3) argument is applied to relate the contributions of similar diagrams in different but SU(3)-related channels. It cannot be applied to the rate as a whole, because the relative CKM magnitude of tree and penguin contributions changes under the U-spin quark substitution (weak interactions do not respect the symmetry).

In the operator-product-based calculations mentioned earlier, the SU(3) relationships reduce the number of independent operators that enter for the set of related decays, and so make the technique more predictive. One then has the problem of estimating the impact of the corrections to SU(3) symmetry relationships, but this on the whole gives smaller theoretical uncertainties than any attempt to constrain the operator matrix elements without the use of SU(3).

Again and again in these lectures I have stressed the point that an essential issue for any theoretical calculation in heavy flavor physics is the estimation of theoretical uncertainties. The tools that theorists have at their disposal to constrain these effects are limited, and there is a lot of subjectivity in how various estimates of uncertainty are made. My general point here is that the uncertainties are best understood in cases where some systematic approximation is used. Heavy quark expansion, QCD perturbation theory and symmetries are all important tools in this respect. It is always better to be estimating the size of a non-leading or symmetry-breaking effect than it is to be estimating the entire effect.

Further constraints on non-perturbative quantities come from QCD sum rules, and also from lattice calculations. QCD sum rules provide another set of rigorous limits by which models can be constrained [26]. Lattice calculations can evaluate certain non-perturbative matrix elements with good precision. Since they are done in the Euclidean region they are not useful for evaluating physical strong interaction phases which appear in matrix elements for scattering or hadronic decay processes. However the method is very powerful for $1 \rightarrow 1$-body or $0 \rightarrow 1$-body transition matrix elements. For accuracy, one needs the so-called "unquenched" calculations that explicitly include the effect of light-quark loops. One also needs good control over the extrapolation needed to reach physical quark masses for light quarks. This requires that the calculations are done for a range of masses sufficiently close to physical values. Both these requirements can be met, it is chiefly a matter of paying the high price for them in computing time. Such calculations are beginning to appear for a few quantities, and more are promised in the next few years [27].

Eventually all these tools run out and we are forced to resort to models to calculate expected rates for some processes. Even when all the limits are correctly reproduced by a model there is no guarantee that its application gives accurate estimates. The usual way that the uncertainty in model estimates is obtained is to vary model parameters, or even to compare two or three different models, and see how the result varies. This can be instructive but it is hardly rigorous. Whenever a measurement disagrees with such an estimate most physicists will conclude that there is something wrong with the models used before they will conclude that they are seeing physics beyond the Standard Model theory – and rightly so. Thus while such calculations are a useful guide as to what to expect in first investigating an area they do not, to my mind, provide any basis for testing the Standard Model. To do that we must look for cases where the theoretical uncertainty can be well constrained, and then look for discrepancies that are large compared to those Standard Model uncertainties.

None of my emphasis on theoretical uncertainties should be construed as saying one cannot test the Standard Model in heavy flavor decays. One can do so, but one must to it carefully. Many so-called clean predictions made in early papers rely on simple rules of thumb to drop so-called suppressed contributions to an amplitude and thereby remove the complications of hadronic physics. As the field matures more careful work has been done, as the limitations of these early approximations have become relevant. Experimental information also helps. One finds empirically that some of the rules of thumb cannot be trusted. That gives the theorists motivation to re-examine their calculations and to make more reasoned estimates of uncertainties. So far no indication of non-Standard Model physics has been seen that has stood the test of better statistics and more careful theory work. But there are many channels with interesting predictions, and only a few have so far accumulated sufficient statistics to make the interesting tests. There is plenty still to do!

References

[1] G.C. Branco, L. Lavoura, J.P. Silva. *"CP Violation"*. Oxford Univ. Press, 1999, 511pp. (International Series of Monographs on Physics, Vol. 103); I.I. Bigi and A.I. Sanda. *"CP Violation"*, Cambridge Monogr. Part. Phys. Nucl. Phys. Cosmol. Vol. 9, 2000, 382pp.

[2] SLAC Report 504 (1998) *www.slac.stanford.edu/pubs/slacreports/slac-r-504.html*

[3] A. de Rujula, Lectures given in this same school.

[4] J.H. Christenson, J.W. Cronin, V.L. Fitch and R. Turlay, Phys. Rev. Lett. **13** (1964) 138.

[5] N. Cabibbo, Phys. Rev. Lett. **10** (1963) 531;
M. Kobayashi and T. Maskawa, Prog. Theor. Phys. **49** (1973) 652.

[6] L. Wolfenstein, Phys. Rev. Lett. **51** (1983) 1945.

[7] C. Jarlskog, Phys. Rev. Lett. **55** (1985) 1039.

[8] See, for example, H. Georgi, Phys. Lett. **B297** (1992) 353, [hep-ph/9209291].

[9] See for example A.F. Falk, Y. Grossman, Z. Ligeti and A.A. Petrov, Phys. Rev. **D65** (2002) 054034, [hep-ph/0110317]; Z. Ligeti, hep-ph/0205316.

[10] For current mass difference and width difference limits see *http://pdg.lbl.gov/2002/s032.pdf* ;
See also the discussion in S. Bergmann *et al.*, Phys. Lett. **B486** (2000) 418, [hep-ph/0005181].

[11] J.L. Rosner, Phys. Rev. **D60** (1999) 114026, [hep-ph/9905366].

[12] F.J. Gilman. K. Kleinknecht and B. Renk, *http://pdg.lbl.gov/2002/kmmixrpp.pdf*

[13] E.C. Poggio, H.R. Quinn and S. Weinberg, Phys. Rev. **D13** (1976) 1958.

[14] C.W. Bauer, Z. Ligeti, M. Luke and A.V. Manohar, hep-ph/0210027.

[15] M. Neubert, Phys. Lett. **B264** (1991) 455.

[16] C.W. Bauer, M. Luke and T. Mannel, Phys. Lett. **B 543** (2002) 261; [hep-ph/0205150].

[17] For recent numbers see the section on B_s mixing of *http://pdg.lbl.gov/2002/s086.pdf*

[18] L. Littenberg, AIP Conf. Proc. **618** (2002) 89, [hep-ex/0201026].

[19] I. Dunietz, H.R. Quinn, A. Snyder, W. Toki and H.J. Lipkin, Phys. Rev. **D43** (1991) 2193.

[20] Y. Nir, hep-ph/0208080.

[21] G. Hiller, hep-ph/0207356.

[22] J.D. Bjorken, Nucl. Phys. Proc. Suppl. **11** (1989) 325.

[23] M. Beneke, G. Buchalla, M. Neubert and C. T. Sachrajda, Phys. Rev. Lett. **83** (1999) 1914, [hep-ph/9905312]; M. Beneke, G. Buchalla, M. Neubert and C.T. Sachrajda, Nucl. Phys. **B606** (2001) 245, [hep-ph/0104110]; Y.Y. Keum, H. Li and A.I. Sanda, Phys. Rev. **D63** (2001) 054008, [hep-ph/0004173].

[24] M. Gronau and D. London, Phys. Rev. Lett. **65** (1990) 3381.

[25] Y. Grossman and H.R. Quinn, Phys. Rev. **D58** (1998) 017504, [hep-ph/9712306];
M. Gronau, D. London, N. Sinha and R. Sinha, Phys. Lett. **B514** (2001) 315, [hep-ph/0105308].

[26] See, for example, M. Shifman TASI Lectures 1995 in *"QCD and Beyond"* (World Scientific, 1995).

[27] For recent reviews see: N. Yamada, hep-lat/0210035; A.S. Kronfeld, hep-ph/0209231.