SLAC-R-719

# The Proceedings of the 27th SLAC Summer Institute on CP Violation: In and Beyond the Standard Model (SSI 1999)

27th SLAC Summer Institute On Particle Physics: CP Violation In And Beyond
The Standard Model (SSI 99),
7-16 Jul 1999, Stanford, California

*Stanford Linear Accelerator Center, Stanford University, Stanford, CA 94309*

Work supported by Department of Energy contract DE–AC03–76SF00515.

# PROCEEDINGS

# OF THE

# XXVII SLAC Summer Institute on Particle Physics


**July 7-16, 1999**
**Stanford Linear Accelerator Center**
**Stanford, California, U.S.A.**

Sponsored by Stanford University and Stanford Linear Accelerator Center under contract with the U.S. Department of Energy, Contract DE-AC03-76SF00515.

Printed in the United States of America. Available from National Technical Information Service, U.S. Department of Commerce, 5285 Port Royal Road, Springfield, VA 22161.

# Table of Contents

# CP Violation In and Beyond the Standard Model

Yosef Nir

*School of Natural Sciences, Institute for Advanced Study, Princeton NJ 08540, USA*[1]

and

*Department of Particle Physics, Weizmann Institute of Science, Rehovot 76100, Israel*

The special features of CP violation in the Standard Model are presented. The significance of measuring CP violation in $B$, $K$ and $D$ decays is explained. The predictions of the Standard Model for CP asymmetries in $B$ decays are analyzed in detail. Then, four frameworks of new physics are reviewed: (i) Supersymmetry provides an excellent demonstration of the power of CP violation as a probe of new physics. (ii) Left-right symmetric models are discussed as an example of an extension of the gauge sector. CP violation suggests that the scale of LRS breaking is low. (iii) The variety of extensions of the scalar sector are presented and their unique CP violating signatures are emphasized. (iv) Vector-like down quarks are presented as an example of an extension of the fermion sector. Their implications for CP asymmetries in $B$ decays are highly interesting.

---

[1] Address for the academic year 1999-2000; nir@sns.ias.edu.

## 1. Introduction

CP violation arises naturally in the three generation Standard Model. The CP violation that has been measured in neutral $K$-meson decays ($\varepsilon_K$ and $\varepsilon'_K$) is accommodated in the Standard Model in a simple way [1]. Yet, CP violation is one of the least tested aspects of the Standard Model. The value of the $\varepsilon_K$ parameter [2] as well as bounds on other CP violating parameters (most noticeably, the electric dipole moments of the neutron, $d_N$, and of the electron, $d_e$) can be accounted for in models where CP violation has features that are very different from the Standard Model ones.

It is unlikely that the Standard Model provides the complete description of CP violation in nature. First, it is quite clear that there exists New Physics beyond the Standard Model. Almost any extension of the Standard Model has additional sources of CP violating effects (or effects that change the relationship of the measurable quantitites to the CP violating parameters of the Standard Model). In addition there is a great puzzle in cosmology that relates to CP violation, and that is the baryon asymmetry of the universe [3]. Theories that explain the observed asymmetry must include new sources of CP violation [4]: the Standard Model cannot generate a large enough matter-antimatter imbalance to produce the baryon number to entropy ratio observed in the universe today [5-7].

In the near future, significant new information on CP violation will be provided by various experiments. The main source of information will be measurements of CP violation in various $B$ decays, particularly neutral $B$ decays into final CP eigenstates [8-10]. First attempts have already been reported [11-13] and interpreted in the framework of various models of new physics [14-15]. Another piece of valuable information might come from a measurement of the $K_L \to \pi^0 \nu\bar{\nu}$ decay [16-19]. For the first time, the pattern of CP violation that is predicted by the Standard Model will be tested. Basic questions such as whether CP is an approximate symmetry in nature will be answered.

It could be that the scale where new CP violating sources appear is too high above the Standard Model scale (e.g. the GUT scale) to give any observable deviations from the Standard Model predictions. In such a case, the outcome of the experiments will be a (frustratingly) successful test of the Standard Model and a significant improvement in our knowledge of the CKM matrix.

A much more interesting situation will arise if the new sources of CP violation appear at a scale that is not too high above the electroweak scale. Then they might be discovered in the forthcoming experiments. Once enough independent observations of CP violating effects are made, we will find that there is no single choice of CKM parameters that is consistent with all measurements. There may even be enough information in the pattern of the inconsistencies to tell us something about the nature of the new physics contributions [20-22].

The aim of these lectures is to explain the theoretical tools with which we will analyze new information about CP violation. The first part, chapters 2-6, deals with the Standard Model while the second, chapters 7-11, discusses physics beyond the Standard Model. In chapter 2, we present the Standard Model picture of CP violation. We emphasize the features that are unique to the Standard Model. In chapter 3, we give a brief, model-independent discussion of CP violating observables in $B$ meson decays. In chapter 4, we discuss CP violation in the $K$ system (particularly, the $\varepsilon_K$ and $\varepsilon'_K$ parameters) in a model independent way and in the framework of the Standard Model. We also describe CP violation in $K \to \pi\nu\bar{\nu}$. In chapter 5, we discuss CP violation in $D \to K\pi$ decays. In chapter 6, we present in detail the Standard Model predictions for CP asymmetries in $B$ decays. In chapter 7, the power of CP violation as a probe of new physics is explained. Then, we discuss specific frameworks of new physics: Supersymmetry (chapter 8), Left-Right symmetry as an example of extensions of the gauge sector (chapter 9), extensions of the scalar sector (chapter 10), and extra down singlet-quarks as an example of extensions of the fermion sector (chapter 11). Finally, we summarize our main points in chapter 12.

## 2. Theory of CP Violation in the Standard Model

### *2.1. Yukawa Interactions Are the Source of CP Violation*

A model of the basic interactions between elementary particles is defined by the following three ingredients:

(i) The symmetries of the Lagrangian;

(ii) The representations of fermions and scalars;

(iii) The pattern of spontaneous symmetry breaking.

The Standard Model (SM) is defined as follows:

(i) The gauge symmetry is

$$G_{\rm SM} = SU(3)_{\rm C} \times SU(2)_{\rm L} \times U(1)_{\rm Y}. \tag{2.1}$$

(ii) There are three fermion generations, each consisting of five representations:

$$Q^I_{Li}(3,2)_{+1/6}, \quad u^I_{Ri}(3,1)_{+2/3}, \quad d^I_{Ri}(3,1)_{-1/3}, \quad L^I_{Li}(1,2)_{-1/2}, \quad \ell^I_{Ri}(1,1)_{-1}. \tag{2.2}$$

Our notations mean that, for example, the left-handed quarks, $Q^I_L$, are in a triplet (3) of the $SU(3)_{\rm C}$ group, a doublet (2) of $SU(2)_{\rm L}$ and carry hypercharge $Y = Q_{\rm EM} - T_3 = +1/6$. The index $I$ denotes interaction eigenstates. The index $i = 1,2,3$ is the flavor (or generation) index. There is a single scalar multiplet:

$$\phi(1,2)_{+1/2}. \tag{2.3}$$

(iii) The $\phi$ scalar assumes a VEV,

$$\langle\phi\rangle = \begin{pmatrix} 0 \\ \frac{v}{\sqrt{2}} \end{pmatrix}, \tag{2.4}$$

so that the gauge group is spontaneously broken:

$$G_{\rm SM} \to SU(3)_{\rm C} \times U(1)_{\rm EM}. \tag{2.5}$$

The Standard Model Lagrangian, $\mathcal{L}_{\rm SM}$, is the most general renormalizable Lagrangian that is consistent with the gauge symmetry $G_{\rm SM}$ of eq. (2.1). It can be divided to three parts:

$$\mathcal{L}_{\rm SM} = \mathcal{L}_{\rm kinetic} + \mathcal{L}_{\rm Higgs} + \mathcal{L}_{\rm Yukawa}. \tag{2.6}$$

As concerns the kinetic terms, to maintain gauge invariance, one has to replace the derivative with a covariant derivative:

$$D^\mu = \partial^\mu + ig_s G^\mu_a L_a + igW^\mu_b T_b + ig' B^\mu Y. \tag{2.7}$$

Here $G_a^\mu$ are the eight gluon fields, $W_b^\mu$ the three weak interaction bosons and $B^\mu$ the single hypercharge boson. The $L_a$'s are $SU(3)_{\rm C}$ generators (the $3\times 3$ Gell-Mann matrices $\frac{1}{2}\lambda_a$ for triplets, 0 for singlets), the $T_b$'s are $SU(2)_{\rm L}$ generators (the $2\times 2$ Pauli matrices $\frac{1}{2}\tau_b$ for doublets, 0 for singlets), and $Y$ are the $U(1)_{\rm Y}$ charges. For example, for the left-handed quarks $Q_L^I$, we have

$$\begin{aligned} \mathcal{L}_{\rm kinetic}(Q_L) &= i\overline{Q^I_{Li}}\gamma^\mu D_\mu Q^I_{Li}, \\ D^\mu Q^I_{Li} &= \left(\partial^\mu + \frac{i}{2}g_s G_a^\mu\lambda_a + \frac{i}{2}g W_b^\mu\tau_b + \frac{i}{6}g' B^\mu\right) Q^I_{Li}. \end{aligned} \tag{2.8}$$

This part of the interaction Lagrangian is always CP conserving.

The Higgs potential, which describes the scalar self interactions, is given by:

$$\mathcal{L}_{\rm Higgs} = \mu^2\phi^\dagger\phi - \lambda(\phi^\dagger\phi)^2. \tag{2.9}$$

For the Standard Model scalar sector, where there is a single doublet, this part of the Lagrangian is also CP conserving. For extended scalar sector, such as that of a two Higgs doublet model, $\mathcal{L}_{\rm Higgs}$ can be CP violating. Even in case that it is CP symmetric, it may lead to spontaneous CP violation.

The Yukawa interactions are given by

$$-\mathcal{L}_{\rm Yukawa} = Y^d_{ij}\overline{Q^I_{Li}}\phi d^I_{Rj} + Y^u_{ij}\overline{Q^I_{Li}}\tilde{\phi} u^I_{Rj} + Y^\ell_{ij}\overline{L^I_{Li}}\phi \ell^I_{Rj} + {\rm h.c.}. \tag{2.10}$$

This part of the Lagrangian is, in general, CP violating. More precisely, CP is violated if and only if [23]

$$\mathcal{I}m\left\{\det[Y^d Y^{d\dagger}, Y^u Y^{u\dagger}]\right\} \neq 0. \tag{2.11}$$

An intuitive explanation of why CP violation is related to *complex* Yukawa couplings goes as follows. The hermiticity of the Lagrangian implies that $\mathcal{L}_{\rm Yukawa}$ has its terms in pairs of the form

$$Y_{ij}\overline{\psi_{Li}}\phi\psi_{Rj} + Y^*_{ij}\overline{\psi_{Rj}}\phi^\dagger\psi_{Li}. \tag{2.12}$$

A CP transformation exchanges the operators

$$\overline{\psi_{Li}}\phi\psi_{Rj} \leftrightarrow \overline{\psi_{Rj}}\phi^\dagger\psi_{Li}, \tag{2.13}$$

but leaves their coefficients, $Y_{ij}$ and $Y_{ij}^*$, unchanged. This means that CP is a symmetry of $\mathcal{L}_{\rm Yukawa}$ if $Y_{ij} = Y_{ij}^*$.

How many independent CP violating parameters are there in $\mathcal{L}_{\rm Yukawa}$? Each of the three Yukawa matrices $Y^f$ is $3 \times 3$ and complex. Consequently, there are 27 real and 27 imaginary parameters in these matrices. Not all of them are, however, physical. If we switch off the Yukawa matrices, there is a global symmetry added to the Standard Model,

$$G^{\rm SM}_{\rm global}(Y^f = 0) = U(3)_Q \times U(3)_{\bar d} \times U(3)_{\bar u} \times U(3)_L \times U(3)_{\bar \ell}. \tag{2.14}$$

A unitary rotation of the three generations for each of the five representations in (2.2) would leave the Standard Model Lagrangian invariant. This means that the physics described by a given set of Yukawa matrices $(Y^d, Y^u, Y^\ell)$, and the physics described by another set,

$$\tilde Y^d = V_Q^\dagger Y^d V_{\bar d}, \quad \tilde Y^u = V_Q^\dagger Y^u V_{\bar u}, \quad \tilde Y^\ell = V_L^\dagger Y^\ell V_{\bar \ell}, \tag{2.15}$$

where $V$ are all unitary matrices, is the same. One can use this freedom to remove, at most, 15 real and 30 imaginary parameters (the number of parameters in five $3 \times 3$ unitary matrices). However, the fact that the Standard Model with the Yukawa matrices switched on has still a global symmetry of

$$G^{\rm SM}_{\rm global} = U(1)_B \times U(1)_e \times U(1)_\mu \times U(1)_\tau \tag{2.16}$$

means that only 26 imaginary parameters can be removed. We conclude that there are 13 flavor parameters: 12 real ones and a single phase. This single phase is the source of CP violation.

### *2.2. Quark Mixing is the (Only!) Source of CP Violation*

Upon the replacement $\mathcal{R}e(\phi^0) \to (v + H^0)/\sqrt{2}$ (see eq. (2.4)), the Yukawa interactions (2.10) give rise to mass terms:

$$-\mathcal{L}_M = (M_d)_{ij} \overline{d^I_{Li}} d^I_{Rj} + (M_u)_{ij} \overline{u^I_{Li}} u^I_{Rj} + (M_\ell)_{ij} \overline{\ell^I_{Li}} \ell^I_{Rj} + {\rm h.c.}, \tag{2.17}$$

where

$$M_f = \frac{v}{\sqrt{2}} Y^f, \tag{2.18}$$

and we decomposed the $SU(2)_\mathrm{L}$ doublets into their components:

$$Q^I_{Li} = \begin{pmatrix} u^I_{Li} \\ d^I_{Li} \end{pmatrix}, \quad L^I_{Li} = \begin{pmatrix} \nu^I_{Li} \\ \ell^I_{Li} \end{pmatrix}. \tag{2.19}$$

Since the Standard Model neutrinos have no Yukawa interactions, they are predicted to be massless.

The mass basis corresponds, by definition, to diagonal mass matrices. We can always find unitary matrices $V_{fL}$ and $V_{fR}$ such that

$$V_{fL} M_f V^\dagger_{fR} = M_f^{\mathrm{diag}}, \tag{2.20}$$

with $M_f^{\mathrm{diag}}$ diagonal and real. The mass eigenstates are then identified as

$$\begin{aligned} d_{Li} &= (V_{dL})_{ij} d^I_{Lj}, \quad d_{Ri} = (V_{dR})_{ij} d^I_{Rj}, \\ u_{Li} &= (V_{uL})_{ij} u^I_{Lj}, \quad u_{Ri} = (V_{uR})_{ij} u^I_{Rj}, \\ \ell_{Li} &= (V_{\ell L})_{ij} \ell^I_{Lj}, \quad \ell_{Ri} = (V_{\ell R})_{ij} \ell^I_{Rj}, \\ \nu_{Li} &= (V_{\nu L})_{ij} \nu^I_{Lj}. \end{aligned} \tag{2.21}$$

Since the Standard Model neutrinos are massless, $V_{\nu L}$ is arbitrary.

The charged current interactions (that is the interactions of the charged $SU(2)_\mathrm{L}$ gauge bosons $W^\pm_\mu = \frac{1}{\sqrt{2}}(W^1_\mu \mp i W^2_\mu)$) for quarks, which in the interaction basis are described by (2.8), have a complicated form in the mass basis:

$$-\mathcal{L}_{W^\pm} = \frac{g}{\sqrt{2}} \overline{u_{Li}} \gamma^\mu (V_{uL} V^\dagger_{dL})_{ij} d_{Lj} W^+_\mu + \text{h.c.}. \tag{2.22}$$

The unitary $3 \times 3$ matrix,

$$V_{\mathrm{CKM}} = V_{uL} V^\dagger_{dL}, \quad (V_{\mathrm{CKM}} V^\dagger_{\mathrm{CKM}} = 1), \tag{2.23}$$

is the Cabibbo-Kobayashi-Maskawa (CKM) *mixing matrix* for quarks [24,1]. A unitary $3 \times 3$ matrix depends on nine parameters: three real angles and six phases.

The form of the matrix is not unique. Usually, the following two conventions are employed:

($i$) There is freedom in defining $V_{\mathrm{CKM}}$ in that we can permute between the various generations. This freedom is fixed by ordering the up quarks and the down quarks by their masses, i.e. $m_{u_1} < m_{u_2} < m_{u_3}$ and $m_{d_1} < m_{d_2} < m_{d_3}$. Usually, we call $(u_1, u_2, u_3) \to (u, c, t)$ and $(d_1, d_2, d_3) \to (d, s, b)$, and the elements of $V_{\rm CKM}$ are written as follows:

$$V_{\rm CKM} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}. \tag{2.24}$$

(*ii*) There is further freedom in the phase structure of $V_{\rm CKM}$. Let us define $P_f$ ($f = u, d, \ell$) to be diagonal unitary (phase) matrices. Then, if instead of using $V_{fL}$ and $V_{fR}$ for the rotation (2.21) to the mass basis we use $\tilde{V}_{fL}$ and $\tilde{V}_{fR}$, defined by $\tilde{V}_{fL} = P_f V_{fL}$ and $\tilde{V}_{fR} = P_f V_{fR}$, we still maintain a legitimate mass basis since $M_f^{\rm diag}$ remains unchanged by such transformations. However, $V_{\rm CKM}$ does change:

$$V_{\rm CKM} \to P_u V_{\rm CKM} P_d^*. \tag{2.25}$$

This freedom is fixed by demanding that $V_{\rm CKM}$ will have the minimal number of phases. In the three generation case $V_{\rm CKM}$ has a single phase. (There are five phase differences between the elements of $P_u$ and $P_d$ and, therefore, five of the six phases in the CKM matrix can be removed.) This is the Kobayashi-Maskawa phase $\delta_{\rm KM}$ which is the single source of *CP violation* in the Standard Model [1]. For example, the elements of the CKM matrix can be written as follows (this is the standard parametrization [25,26]):

$$V_{\rm CKM} = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta_{\rm KM}} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta_{\rm KM}} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta_{\rm KM}} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta_{\rm KM}} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta_{\rm KM}} & c_{23}c_{13} \end{pmatrix}, \tag{2.26}$$

where $c_{ij} \equiv \cos\theta_{ij}$ and $s_{ij} \equiv \sin\theta_{ij}$. The three $\sin\theta_{ij}$ are the three real mixing parameters.

As a result of the fact that $V_{\rm CKM}$ is not diagonal, the $W^\pm$ gauge bosons couple to quark (mass eigenstates) of different generations. Within the Standard Model, this is the only source of *flavor changing* interactions. In principle, there could be additional sources of flavor mixing (and of CP violation) in the lepton sector and in $Z^0$ interactions. We now explain why, within the Standard Model, this does not happen.

*Mixing in the lepton sector*: An analysis similar to the above applies also to the left-handed leptons. The mixing matrix is [27] $V_{\rm MNS} = V_{\nu L} V_{\ell L}^\dagger$. However, we can use the arbitrariness of $V_{\nu L}$ (related to the masslessness of neutrinos) to choose $V_{\nu L} = V_{\ell L}$, and the mixing matrix becomes a unit matrix. We conclude that the masslessness of neutrinos (if true) implies that there is no mixing in the lepton sector. If neutrinos have masses then the leptonic charged current interactions will exhibit mixing and CP violation.

*Mixing in neutral current interactions*: We study the interactions of the neutral $Z$-boson, $Z^\mu = \cos\theta_W W_3^\mu - \sin\theta_W B^\mu$ ($\tan\theta_W \equiv g'/g$) with, for example, left-handed down quarks. The $W_3$-interactions are given in (2.8), while the $B$ interactions are given by

$$-\mathcal{L}_B(Q_L) = -\frac{g'}{6}\overline{Q^I_{Li}}\gamma^\mu Q^I_{Li} B_\mu. \tag{2.27}$$

In the mass basis, we have then

$$\begin{aligned}-\mathcal{L}_Z =& \frac{g}{\cos\theta_W}\left(-\frac{1}{2}+\frac{1}{3}\sin^2\theta_W\right)\overline{d_{Li}}\gamma^\mu(V^\dagger_{dL}V_{dL})_{ij}d_{Lj}Z_\mu \\ =& \frac{g}{\cos\theta_W}\left(-\frac{1}{2}+\frac{1}{3}\sin^2\theta_W\right)\overline{d_{Li}}\gamma^\mu d_{Li}Z_\mu.\end{aligned} \tag{2.28}$$

We learn that the neutral current interactions remain universal in the mass basis and there are no additional flavor parameters in their description. This situation goes beyond the Standard Model to all models where all left-handed quarks are in $SU(2)_\mathrm{L}$ doublets and all right-handed ones in singlets. The $Z$-boson does have flavor changing couplings in models where this is not the case.

Examining the mass basis one can easily identify the flavor parameters. In the quark sector, we have six quark masses, three mixing angles (the number of real parameters in $V_\mathrm{CKM}$) and the single phase $\delta_\mathrm{KM}$ mentioned above. In the lepton sector, we have the three charged lepton masses.

We have also learnt now some of the special features of CP violation in the Standard Model:

(i) CP is explicitly broken.

(ii) There is a single source of CP violation, that is $\delta_\mathrm{KM}$.

(iii) CP violation appears only in the charged current interactions of quarks.

(iv) CP violation is closely related to flavor changing interactions.

### *2.3. The CKM Matrix and the Unitarity Triangles*

In the mass basis, CP violation is related to the CKM matrix. The fact that there are only three real and one imaginary physical parameters in $V_\mathrm{CKM}$ can be made manifest by choosing an explicit parametrization. One example was given above, in eq. (2.26), with the four parameters $(s_{12}, s_{23}, s_{13}, \delta_{\rm KM})$. Another, very useful, example is the Wolfenstein parametrization of $V_{\rm CKM}$, where the four mixing parameters are $(\lambda, A, \rho, \eta)$ with $\lambda = |V_{us}| = 0.22$ playing the role of an expansion parameter and $\eta$ representing the CP violating phase [28]:

$$V = \begin{pmatrix} 1-\frac{\lambda^2}{2} & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1-\frac{\lambda^2}{2} & A\lambda^2 \\ A\lambda^3(1-\rho-i\eta) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}(\lambda^4). \tag{2.29}$$

Various parametrizations differ in the way that the freedom of phase rotation, eq. (2.25), is used to leave a single phase in $V_{\rm CKM}$. One can define, however, a CP violating quantity in $V_{\rm CKM}$ that is independent of the parametrization [23]. This quantity is called $J$ and defined through

$$\mathcal{I}m[V_{ij}V_{kl}V_{il}^*V_{kj}^*] = J \sum_{m,n=1}^{3} \epsilon_{ikm}\epsilon_{jln}, \quad (i,j,k,l = 1,2,3). \tag{2.30}$$

CP is violated in the Standard Model only if $J \neq 0$.

The usefulness of $J$ may not be clear from its formal definition in (2.30), but does give useful insights once the *unitarity triangles* are introduced. The unitarity of the CKM matrix leads to various relations among the matrix elements, *e.g.*

$$V_{ud}V_{us}^* + V_{cd}V_{cs}^* + V_{td}V_{ts}^* = 0, \tag{2.31}$$

$$V_{us}V_{ub}^* + V_{cs}V_{cb}^* + V_{ts}V_{tb}^* = 0, \tag{2.32}$$

$$V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0. \tag{2.33}$$

Each of the three relations (2.31)-(2.33) requires the sum of three complex quantities to vanish and so can be geometrically represented in the complex plane as a triangle. These are "the unitarity triangles", though the term "unitarity triangle" is usually reserved for the relation (2.33) only. It is a surprising feature of the CKM matrix that all unitarity triangles are equal in area: the area of each unitarity triangle equals $|J|/2$ while the sign of $J$ gives the direction of the complex vectors around the triangles. The relation between Jarlskog's measure of CP violation $J$ and the Wolfenstein parameters is given by

$$J \simeq \lambda^6 A^2 \eta. \tag{2.34}$$

The rescaled unitarity triangle is derived from (2.33) by (a) choosing a phase convention such that $(V_{cd}V_{cb}^*)$ is real, and (b) dividing the lengths of all sides by $|V_{cd}V_{cb}^*|$. Step (a) aligns one side of the triangle with the real axis, and step (b) makes the length of this side 1. The form of the triangle is unchanged. Two vertices of the rescaled unitarity triangle are thus fixed at (0,0) and (1,0). The coordinates of the remaining vertex correspond to the Wolfenstein parameters $(\rho,\eta)$. The area of the rescaled unitarity triangle is $|\eta|/2$.

Depicting the rescaled unitarity triangle in the $(\rho,\eta)$ plane, the lengths of the two complex sides are

$$R_u \equiv \sqrt{\rho^2+\eta^2} = \frac{1}{\lambda}\left|\frac{V_{ub}}{V_{cb}}\right|, \quad R_t \equiv \sqrt{(1-\rho)^2+\eta^2} = \frac{1}{\lambda}\left|\frac{V_{td}}{V_{cb}}\right|. \tag{2.35}$$

The three angles of the unitarity triangle are denoted by $\alpha,\beta$ and $\gamma$ [29]:

$$\alpha \equiv \arg\left[-\frac{V_{td}V_{tb}^*}{V_{ud}V_{ub}^*}\right], \quad \beta \equiv \arg\left[-\frac{V_{cd}V_{cb}^*}{V_{td}V_{tb}^*}\right], \quad \gamma \equiv \arg\left[-\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*}\right]. \tag{2.36}$$

They are physical quantities and, we will soon see, can be independently measured by CP asymmetries in $B$ decays. It is also useful to define the two small angles of the unitarity triangles (2.32) and (2.31):

$$\beta_s \equiv \arg\left[-\frac{V_{ts}V_{tb}^*}{V_{cs}V_{cb}^*}\right], \quad \beta_K \equiv \arg\left[-\frac{V_{cs}V_{cd}^*}{V_{us}V_{ud}^*}\right]. \tag{2.37}$$

To make predictions for future measurements of CP violating observables, we need to find the allowed ranges for the CKM phases. There are three ways to determine the CKM parameters (see *e.g.* [30]):

(i) **Direct measurements** are related to SM tree level processes. At present, we have direct measurements of $|V_{ud}|$, $|V_{us}|$, $|V_{ub}|$, $|V_{cd}|$, $|V_{cs}|$, $|V_{cb}|$ and $|V_{tb}|$.

(ii) **CKM Unitarity** ($V_{\rm CKM}^\dagger V_{\rm CKM}=\mathbf{1}$) relates the various matrix elements. At present, these relations are useful to constrain $|V_{td}|$, $|V_{ts}|$, $|V_{tb}|$ and $|V_{cs}|$.

(iii) **Indirect measurements** are related to SM loop processes. At present, we constrain in this way $|V_{tb}V_{td}|$ (from $\Delta m_B$ and $\Delta m_{B_s}$) and $\delta_{\rm KM}$ or, equivalently, $\eta$ (from $\varepsilon_K$).

When all available data is taken into account, we find [31-36]:

$$\lambda = 0.2205 \pm 0.0018, \quad A = 0.826 \pm 0.041, \tag{2.38}$$

$$-0.15 \leq \rho \leq +0.35, \quad +0.20 \leq \eta \leq +0.45, \tag{2.39}$$

$$0.4 \leq \sin 2\beta \leq 0.8, \quad -0.9 \leq \sin 2\alpha \leq 1.0, \quad 0.23 \leq \sin^2 \gamma \leq 1.0. \tag{2.40}$$

Of course, there are correlations between the various parameters. The full information can be described by allowed regions in the $(\rho, \eta)$ or the $(\sin 2\alpha, \sin 2\beta)$ planes (see e.g. [35]). (Recently, it has been shown that $B \to \pi K$ decays can provide bounds on the angle $\gamma$ of the unitarity triangle [37-41]. A bound of $\cos\gamma \leq 0.32$ was derived in [40]. We did not incorporate these bounds into our analysis.)

Eqs. (2.39) and (2.40) show yet another important feature of CP violation in the Standard Model. The fact that $\eta/\rho = \mathcal{O}(1)$ or, equivalently, $\sin\gamma = \mathcal{O}(1)$, implies that *CP is not an approximate symmetry* within the Standard Model. This is not an obvious fact: after all, the two measured CP violating quantities, $\varepsilon_K$ and $\varepsilon'_K$, are very small (orders $10^{-3}$ and $10^{-6}$, respectively). The Standard Model accounts for their smallness by the smallness of the flavor violation, that is the mixing angles, and not by the smallness of CP violation, that is a small phase. Indeed, the Standard Model predicts that in some (yet unmeasured) processes, the CP asymmetry is of order one.

### *2.4. The Uniqueness of the Standard Model Picture of CP Violation*

In the previous subsections, we have learnt several features of CP violation as explained by the Standard Model:

(i) CP is explicitly broken.

(ii) $\delta_{\rm KM}$ is the only source of CP violation.

(iii) CP violation appears only in the charged current interactions of quarks.

(iv) CP violation would vanish in the absence of flavor changing interactions.

(v) CP is not an approximate symmetry ($\delta_{\rm KM} = \mathcal{O}(1)$).

(Non-perturbative corrections to the Standard Model tree-level Lagrangian are expected to induce $\theta_{\rm QCD}$, a CP violating parameter. This second possible source of CP violation is related to strong interactions and is flavor diagonal. The bounds on the electric dipole moment of the neutron imply that $\theta_{\rm QCD} \lesssim 10^{-9}$. The Standard Model offers no natural explanation to the smallness of $\theta_{\rm QCD}$. This puzzle is called 'the strong CP problem'. We assume that it is solved by some type of new physics, such as a Peccei-Quinn symmetry [42], which sets $\theta_{\rm QCD}$ to zero.)

It is important to realize that (a) none of features (i)-(v) is experimentally established and that (b) various reasonable extensions of the Standard Model provide examples where these features do not hold. In particular, it could be that CP violation in Nature has some or all of the following features:

(i) CP is spontaneously broken.

(ii) There are many independent sources of CP violation.

(iii) CP violation appears in lepton interactions and/or in neutral current interactions and/or in new sectors beyond the SM.

(iv) CP violation appears also in flavor diagonal interactions.

(v) CP is an approximate symmetry.

This situation, where the Standard Model has a very unique description of CP violation and experiments have not yet confirmed this description, is the basis for the strong interest, experimental and theoretical, in CP violation. There are two types of unambiguous tests concerning CP violation in the Standard Model: First, since there is a single source of CP violation, all observables are correlated with each other. For example, the CP asymmetries in $B \to \psi K_S$ and in $K \to \pi\nu\bar{\nu}$ are strongly correlated. Second, since CP violation is restricted to flavor changing quark processes, it is predicted to practically vanish in the lepton sector and in flavor diagonal processes. For example, the transverse lepton polarization in semileptonic meson decays, CP violation in $t\bar{t}$ production, and (assuming $\theta_{\rm QCD} = 0$) the electric dipole moment of the neutron are all predicted to be orders of magnitude below the (present and near future) experimental sensitivity.

The experimental investigation of CP violation in $B$ decays will shed light on some but not all of these questions. In particular, it will easily test the question of whether CP is an approximate symmetry: if any $\mathcal{O}(1)$ asymmetry is observed, for example in the $B \to \psi K_S$ mode, then we immediately learn that CP is not an approximate symmetry. It will also test (though probably in a less definitive way) the question of whether the Kobayashi-Maskawa phase is the only source of CP violation. On the other hand, we will learn little on flavor diagonal CP violation and on CP violation outside the quark sector.

It is therefore important to search for CP violation in many different systems.

## 3. CP Violation in Meson Decays

In the previous section, we understood how CP violation arises in the Standard Model. In the next three sections, we would like to understand the implications of this theory for the phenomenology of CP violation in meson decays. Our main focus will be on $B$-meson decays. To do so, we first present a model independent analysis of CP violation in meson decays.

There are three different types of CP violation in meson decays:

(i) CP violation in mixing, which occurs when the two neutral mass eigenstate admixtures cannot be chosen to be CP-eigenstates;

(ii) CP violation in decay, which occurs in both charged and neutral decays, when the amplitude for a decay and its CP-conjugate process have different magnitudes;

(iii) CP violation in the interference of decays with and without mixing, which occurs in decays into final states that are common to $B^0$ and $\bar{B}^0$.

### *3.1. Notations and Formalism*

To define the three types of CP violation in meson decays and to discuss their theoretical calculation and experimental measurement, we first introduce some notations and formalism. We refer specifically to $B$ meson mixing and decays, but most of our discussion applies equally well to $K$, $B_s$ and $D$ mesons.

Our phase convention for the CP transformation law of the neutral $B$ mesons is defined by

$$\mathrm{CP}|B^0\rangle = \omega_B|\bar{B}^0\rangle, \quad \mathrm{CP}|\bar{B}^0\rangle = \omega_B^*|B^0\rangle, \quad (|\omega_B| = 1). \tag{3.1}$$

Physical observables do not depend on the phase factor $\omega_B$. The time evolution of any linear combination of the neutral $B$-meson flavor eigenstates,

$$a|B^0\rangle + b|\bar{B}^0\rangle, \tag{3.2}$$

is governed by the Schrödinger equation,

$$i\frac{d}{dt}\begin{pmatrix} a \\ b \end{pmatrix} = H\begin{pmatrix} a \\ b \end{pmatrix} \equiv \left(M - \frac{i}{2}\Gamma\right)\begin{pmatrix} a \\ b \end{pmatrix}, \tag{3.3}$$

for which $M$ and $\Gamma$ are $2 \times 2$ Hermitian matrices.

The off-diagonal terms in these matrices, $M_{12}$ and $\Gamma_{12}$, are particularly important in the discussion of mixing and CP violation. $M_{12}$ is the dispersive part of the transition amplitude from $B^0$ to $\bar{B}^0$, while $\Gamma_{12}$ is the absorptive part of that amplitude.

The light $B_L$ and heavy $B_H$ mass eigenstates are given by

$$|B_{L,H}\rangle = p|B^0\rangle \pm q|\bar{B}^0\rangle. \tag{3.4}$$

The complex coefficients $q$ and $p$ obey the normalization condition $|q|^2 + |p|^2 = 1$. Note that $\arg(q/p^*)$ is just an overall common phase for $|B_L\rangle$ and $|B_H\rangle$ and has no physical significance. The mass difference and the width difference between the physical states are given by

$$\Delta m \equiv M_H - M_L, \quad \Delta\Gamma \equiv \Gamma_H - \Gamma_L. \tag{3.5}$$

Solving the eigenvalue equation gives

$$(\Delta m)^2 - \frac{1}{4}(\Delta\Gamma)^2 = (4|M_{12}|^2 - |\Gamma_{12}|^2), \qquad \Delta m\Delta\Gamma = 4\mathcal{R}e(M_{12}\Gamma_{12}^*), \tag{3.6}$$

$$\frac{q}{p} = -\frac{2M_{12}^* - i\Gamma_{12}^*}{\Delta m - \frac{i}{2}\Delta\Gamma} = -\frac{\Delta m - \frac{i}{2}\Delta\Gamma}{2M_{12} - i\Gamma_{12}}. \tag{3.7}$$

In the $B$ system, $|\Gamma_{12}| \ll |M_{12}|$ (see discussion below), and then, to leading order in $|\Gamma_{12}/M_{12}|$, eqs. (3.6) and (3.7) can be written as

$$\Delta m_B = 2|M_{12}|, \quad \Delta\Gamma_B = 2\mathcal{R}e(M_{12}\Gamma_{12}^*)/|M_{12}|, \tag{3.8}$$

$$\frac{q}{p} = -\frac{M_{12}^*}{|M_{12}|}. \tag{3.9}$$

To discuss CP violation in mixing, it is useful to write eq. (3.7) to first order in $|\Gamma_{12}/M_{12}|$:

$$\frac{q}{p} = -\frac{M_{12}^*}{|M_{12}|}\left[1 - \frac{1}{2}\mathcal{I}m\left(\frac{\Gamma_{12}}{M_{12}}\right)\right]. \tag{3.10}$$

To discuss CP violation in decay, we need to consider decay amplitudes. The CP transformation law for a final state $f$ is

$$\mathrm{CP}|f\rangle = \omega_f|\bar{f}\rangle, \quad \mathrm{CP}|\bar{f}\rangle = \omega_f^*|f\rangle, \quad (|\omega_f| = 1). \tag{3.11}$$

For a final CP eigenstate $f = \bar{f} = f_{\mathrm{CP}}$, the phase factor $\omega_f$ is replaced by $\eta_{f_{\mathrm{CP}}} = \pm 1$, the CP eigenvalue of the final state. We define the decay amplitudes $A_f$ and $\bar{A}_f$ according to

$$A_f = \langle f|\mathcal{H}_d|B^0\rangle, \quad \bar{A}_f = \langle f|\mathcal{H}_d|\bar{B}^0\rangle, \tag{3.12}$$

where $\mathcal{H}_d$ is the decay Hamiltonian.

CP relates $A_f$ and $\bar{A}_{\bar{f}}$. There are two types of phases that may appear in $A_f$ and $\bar{A}_{\bar{f}}$. Complex parameters in any Lagrangian term that contributes to the amplitude will appear in complex conjugate form in the CP-conjugate amplitude. Thus their phases appear in $A_f$ and $\bar{A}_{\bar{f}}$ with opposite signs. In the Standard Model these phases occur only in the CKM matrix which is part of the electroweak sector of the theory, hence these are often called "weak phases". The weak phase of any single term is convention dependent. However the difference between the weak phases in two different terms in $A_f$ is convention independent because the phase rotations of the initial and final states are the same for every term. A second type of phase can appear in scattering or decay amplitudes even when the Lagrangian is real. Such phases do not violate CP and they appear in $A_f$ and $\bar{A}_{\bar{f}}$ with the same sign. Their origin is the possible contribution from intermediate on-shell states in the decay process, that is an absorptive part of an amplitude that has contributions from coupled channels. Usually the dominant rescattering is due to strong interactions and hence the designation "strong phases" for the phase shifts so induced. Again only the relative strong phases of different terms in a scattering amplitude have physical content, an overall phase rotation of the entire amplitude has no physical consequences. Thus it is useful to write each contribution to $A$ in three parts: its magnitude $A_i$; its weak phase term $e^{i\phi_i}$; and its strong phase term $e^{i\delta_i}$. Then, if several amplitudes contribute to $B \to f$, we have

$$\left|\frac{\bar{A}_{\bar{f}}}{A_f}\right| = \left|\frac{\sum_i A_i e^{i(\delta_i - \phi_i)}}{\sum_i A_i e^{i(\delta_i + \phi_i)}}\right|. \tag{3.13}$$

To discuss CP violation in the interference of decays with and without mixing, we introduce a complex quantity $\lambda_f$ defined by

$$\lambda_f = \frac{q}{p} \frac{\bar{A}_f}{A_f}. \tag{3.14}$$

We further define the CP transformation law for the quark fields in the Hamiltonian (a careful treatment of CP conventions can be found in [43]):

$$q \rightarrow \omega_q \bar{q}, \quad \bar{q} \rightarrow \omega_q^* q, \quad (|\omega_q| = 1). \tag{3.15}$$

The effective Hamiltonian that is relevant to $M_{12}$ is of the form

$$H_{\rm eff}^{\Delta b=2} \propto e^{+2i\phi_B} \left[\bar{d}\gamma^\mu(1-\gamma_5)b\right]^2 + e^{-2i\phi_B} \left[\bar{b}\gamma^\mu(1-\gamma_5)d\right]^2, \tag{3.16}$$

where $2\phi_B$ is a CP violating (weak) phase. (We use the Standard Model $V-A$ amplitude, but the results can be generalized to any Dirac structure.) For the $B$ system, where $|\Gamma_{12}| \ll |M_{12}|$, this leads to

$$q/p = \omega_B \omega_b^* \omega_d e^{-2i\phi_B}. \tag{3.17}$$

(We implicitly assumed that the vacuum insertion approximation gives the correct sign for $M_{12}$. In general, there is a sign$(B_B)$ factor on the right hand side of eq. (3.17) [44].) To understand the phase structure of decay amplitudes, we take as an example the $b \rightarrow q\bar{q}d$ decay ($q = u$ or $c$). The decay Hamiltonian is of the form

$$H_d \propto e^{+i\phi_f} \left[\bar{q}\gamma^\mu(1-\gamma_5)d\right] \left[\bar{b}\gamma_\mu(1-\gamma_5)q\right] + e^{-i\phi_f} \left[\bar{q}\gamma^\mu(1-\gamma_5)b\right] \left[\bar{d}\gamma_\mu(1-\gamma_5)q\right], \tag{3.18}$$

where $\phi_f$ is the appropriate weak phase. (Again, for simplicity we use a $V-A$ structure, but the results hold for any Dirac structure.) Then

$$\bar{A}_{\bar{f}}/A_f = \omega_f \omega_B^* \omega_b \omega_d^* e^{-2i\phi_f}. \tag{3.19}$$

Eqs. (3.17) and (3.19) together imply that for a final CP eigenstate,

$$\lambda_{f_{\rm CP}} = \eta_{f_{\rm CP}} e^{-2i(\phi_B+\phi_f)}. \tag{3.20}$$

**(i) CP violation in mixing:**

$$|q/p| \neq 1. \tag{3.21}$$

This results from the mass eigenstates being different from the CP eigenstates, and requires a relative phase between $M_{12}$ and $\Gamma_{12}$. For the neutral $B$ system, this effect could be observed through the asymmetries in semileptonic decays:

$$a_{\rm SL} = \frac{\Gamma(\bar{B}^0_{\rm phys}(t) \to \ell^+ \nu X) - \Gamma(B^0_{\rm phys}(t) \to \ell^- \nu X)}{\Gamma(\bar{B}^0_{\rm phys}(t) \to \ell^+ \nu X) + \Gamma(B^0_{\rm phys}(t) \to \ell^- \nu X)}. \tag{3.22}$$

In terms of $q$ and $p$,

$$a_{\rm SL} = \frac{1 - |q/p|^4}{1 + |q/p|^4}. \tag{3.23}$$

CP violation in mixing has been observed in the neutral $K$ system ($\mathcal{R}e\ \varepsilon_K \neq 0$).

In the neutral $B$ system, the effect is expected to be small, $\lesssim \mathcal{O}(10^{-2})$. The reason is that, model independently, one expects that $a_{\rm SL} \lesssim \Delta\Gamma_B/\Delta m_B$. (We assume here that $\arg(\Gamma_{12}/M_{12})$ is not particularly close to $\pi/2$.) The difference in width is produced by decay channels common to $B^0$ and $\bar{B}^0$. The branching ratios for such channels are at or below the level of $10^{-3}$. Since various channels contribute with differing signs, one expects that their sum does not exceed the individual level. Hence, we can safely assume that $\Delta\Gamma_B/\Gamma_B = \mathcal{O}(10^{-2})$. On the other hand, it is experimentaly known that $\Delta m_B/\Gamma_B \approx 0.7$.

To calculate $a_{\rm SL}$, we use (3.23) and (3.10), and get:

$$a_{\rm SL} = \mathcal{I}m \frac{\Gamma_{12}}{M_{12}}. \tag{3.24}$$

To predict it in a given model, one needs to calculate $M_{12}$ and $\Gamma_{12}$. This involves large hadronic uncertainties, in particular in the hadronization models for $\Gamma_{12}$.

**(ii) CP violation in decay:**

$$|\bar{A}_{\bar{f}}/A_f| \neq 1. \tag{3.25}$$

This appears as a result of interference among various terms in the decay amplitude, and will not occur unless at least two terms have different weak phases and different strong phases. CP asymmetries in charged $B$ decays,

$$a_{f^\pm} = \frac{\Gamma(B^+ \to f^+) - \Gamma(B^- \to f^-)}{\Gamma(B^+ \to f^+) + \Gamma(B^- \to f^-)}, \tag{3.26}$$

are purely an effect of CP violation in decay. In terms of the decay amplitudes,

$$a_{f^\pm} = \frac{1 - |\bar{A}_{f^-}/A_{f^+}|^2}{1 + |\bar{A}_{f^-}/A_{f^+}|^2}. \tag{3.27}$$

CP violation in decay has been observed in the neutral $K$ system ($\mathcal{R}e\ \varepsilon'_K \neq 0$).

To calculate $a_{f^\pm}$, we use (3.27) and (3.13). For simplicity, we consider decays with contributions from two weak phases and with $A_2 \ll A_1$. We get:

$$a_{f^\pm} = -2(A_2/A_1)\sin(\delta_2 - \delta_1)\sin(\phi_2 - \phi_1). \tag{3.28}$$

The magnitude and strong phase of any amplitude involve long distance strong interaction physics, and our ability to calculate these from first principles is limited. Thus quantities that depend only on the weak phases are much cleaner than those that require knowledge of the relative magnitudes or strong phases of various amplitude contributions, such as CP violation in decay.

**(iii) CP violation in the interference between decays with and without mixing:**

$$|\lambda_{f_{\rm CP}}| = 1, \quad \mathcal{I}m\ \lambda_{f_{\rm CP}} \neq 0. \tag{3.29}$$

Any $\lambda_{f_{\rm CP}} \neq \pm 1$ is a manifestation of CP violation. The special case (3.29) isolates the effects of interest since both CP violation in decay, eq. (3.25), and in mixing, eq. (3.21), lead to $|\lambda_{f_{\rm CP}}| \neq 1$. For the neutral $B$ system, this effect can be observed by comparing decays into final CP eigenstates of a time-evolving neutral $B$ state that begins at time zero as $B^0$ to those of the state that begins as $\bar{B}^0$:

$$a_{f_{\rm CP}} = \frac{\Gamma(\bar{B}^0_{\rm phys}(t) \to f_{\rm CP}) - \Gamma(B^0_{\rm phys}(t) \to f_{\rm CP})}{\Gamma(\bar{B}^0_{\rm phys}(t) \to f_{\rm CP}) + \Gamma(B^0_{\rm phys}(t) \to f_{\rm CP})}. \tag{3.30}$$

This time dependent asymmetry is given (for $|\lambda_{f_{\rm CP}}| = 1$) by

$$a_{f_{\rm CP}} = -\mathcal{I}m\lambda_{f_{\rm CP}}\sin(\Delta m_B t). \tag{3.31}$$

CP violation in the interference of decays with and without mixing has been observed for the neutral $K$ system ($\mathcal{I}m\ \varepsilon_K \neq 0$). It is expected to be an effect of $\mathcal{O}(1)$ in various $B$ decays. For such cases, the contribution from CP violation in mixing is clearly negligible.

For decays that are dominated by a single CP violating phase (for example, $B \to \psi K_S$ and $K_L \to \pi^0 \nu\bar{\nu}$), so that the contribution from CP violation in decay is also negligible, $a_{f_{\rm CP}}$ is cleanly interpreted in terms of purely electroweak parameters. Explicitly, $\mathcal{I}m\lambda_{f_{\rm CP}}$ gives the difference between the phase of the $B-\bar{B}$ mixing amplitude ($2\phi_B$) and twice the phase of the relevant decay amplitude ($2\phi_f$) (see eq. (3.20)):

$$\mathcal{I}m\lambda_{f_{\rm CP}} = -\eta_{f_{\rm CP}} \sin[2(\phi_B + \phi_f)]. \tag{3.32}$$

A summary of the main properties of the different types of CP violation in meson decays is given in the table I.

| Type | Exp. | Theory | Calculation | Uncertainties | Observed in |
|---|---|---|---|---|---|
| mixing | $a_{\rm SL}$ | $\frac{1-\lvert q/p\rvert^4}{1+\lvert q/p\rvert^4}$ | $\mathcal{I}m\frac{\Gamma_{12}}{M_{12}}$ | $\Gamma_{12} \Longrightarrow$ Large | $\mathcal{R}e\ \varepsilon_K$ |
| decay | $a_{f\pm}$ | $\frac{1-\lvert\bar{A}_{f-}/A_{f+}\rvert^2}{1+\lvert\bar{A}_{f-}/A_{f+}\rvert^2}$ | $-2\frac{A_2}{A_1}\sin(\delta_2-\delta_1)\sin(\phi_2-\phi_1)$ | $\delta_i, A_i \Longrightarrow$ Large | $\mathcal{R}e\ \varepsilon'_K$ |
| interference | $a_{f_{CP}}$ | $-\mathcal{I}m\lambda_{f_{CP}}$ | $\eta_{f_{CP}}\sin[2(\phi_B+\phi_f)]$ | Small | $\mathcal{I}m\ \varepsilon_K$ |

Table I. The three types of CP violation in meson decays.

## 4. CP Violation in $K$ Decays

The two CP violating observables that have been measured are related to $K$ meson decays. In this chapter we present these observables and their significance.

### *4.1. Direct and Indirect CP Violation*

The terms indirect CP violation and direct CP violation are commonly used in the literature. While various authors use these terms with different meanings, the most useful definition is the following:

**Indirect CP violation** refers to CP violation in meson decays where the CP violating phases can all be chosen to appear in $\Delta F = 2$ (mixing) amplitudes.

**Direct CP violation** refers to CP violation in meson decays where some CP violating phases necessarily appear in $\Delta F = 1$ (decay) amplitudes.

Examining eqs. (3.21) and (3.7), we learn that CP violation in mixing is a manifestation of indirect CP violation. Examining eqs. (3.25) and (3.12), we learn that CP violation in decay is a manifestation of direct CP violation. Examining eqs. (3.29) and (3.14), we learn that the situation concerning CP violation in the interference of decays with and without mixing is more subtle. For any single measurement of $\mathcal{I}m\lambda_f \neq 0$, the relevant CP violating phase can be chosen by convention to reside in the $\Delta F = 2$ amplitude ($\phi_f = 0$, $\phi_B \neq 0$ in the notation of eq. (3.20)), and then we would call it indirect CP violation. Consider, however, the CP asymmetries for two different final CP eigenstates (for the same decaying meson), $f_a$ and $f_b$. Then, a non-zero difference between $\mathcal{I}m\lambda_{f_a}$ and $\mathcal{I}m\lambda_{f_b}$ requires that there exists CP violation in $\Delta F = 1$ processes ($\phi_{f_a} - \phi_{f_b} \neq 0$), namely direct CP violation.

Experimentally, both direct and indirect CP violation have been established. Below we will see that $\varepsilon_K$ signifies indirect CP violation while $\varepsilon'_K$ signifies direct CP violation.

Theoretically, most models of CP violation (including the Standard Model) have predicted that both types of CP violation exist. There is, however, one class of models, that is *superweak models*, that predict only indirect CP violation. The measurement of $\varepsilon'_K \neq 0$ has excluded this class of models.

### *4.2. The $\varepsilon_K$ and $\varepsilon'_K$ Parameters*

Historically, a different language from the one used by us has been employed to describe CP violation in $K \to \pi\pi$ and $K \to \pi\ell\nu$ decays. In this section we 'translate' the language of $\varepsilon_K$ and $\varepsilon'_K$ to our notations. Doing so will make it easy to understand which type of CP violation is related to each quantity.

The two CP violating quantities measured in neutral $K$ decays are

$$\eta_{00} = \frac{\langle \pi^0\pi^0|\mathcal{H}|K_L\rangle}{\langle \pi^0\pi^0|\mathcal{H}|K_S\rangle}, \quad \eta_{+-} = \frac{\langle \pi^+\pi^-|\mathcal{H}|K_L\rangle}{\langle \pi^+\pi^-|\mathcal{H}|K_S\rangle}. \tag{4.1}$$

Define, for $(ij) = (00)$ or $(+-)$,

$$A_{ij} = \langle \pi^i\pi^j|\mathcal{H}|K^0\rangle, \quad \bar{A}_{ij} = \langle \pi^i\pi^j|\mathcal{H}|\bar{K}^0\rangle, \tag{4.2}$$

$$\lambda_{ij} = \left(\frac{q}{p}\right)_K \frac{\bar{A}_{ij}}{A_{ij}}. \tag{4.3}$$

Then

$$\eta_{00} = \frac{1-\lambda_{00}}{1+\lambda_{00}}, \quad \eta_{+-} = \frac{1-\lambda_{+-}}{1+\lambda_{+-}}. \tag{4.4}$$

The $\eta_{00}$ and $\eta_{+-}$ parameters get contributions from CP violation in mixing ($|(q/p)|_K \neq 1$) and from the interference of decays with and without mixing ($\mathcal{I}m\lambda_{ij} \neq 0$) at $\mathcal{O}(10^{-3})$ and from CP violation in decay ($|\bar{A}_{ij}/A_{ij}| \neq 1$) at $\mathcal{O}(10^{-6})$.

There are two isospin channels in $K \to \pi\pi$ leading to final $(2\pi)_{I=0}$ and $(2\pi)_{I=2}$ states:

$$\begin{aligned} \langle \pi^0\pi^0| &= \sqrt{\frac{1}{3}}\langle(\pi\pi)_{I=0}| - \sqrt{\frac{2}{3}}\langle(\pi\pi)_{I=2}|, \\ \langle \pi^+\pi^-| &= \sqrt{\frac{2}{3}}\langle(\pi\pi)_{I=0}| + \sqrt{\frac{1}{3}}\langle(\pi\pi)_{I=2}|. \end{aligned} \tag{4.5}$$

The fact that there are two strong phases allows for CP violation in decay. The possible effects are, however, small (on top of the smallness of the relevant CP violating phases) because the final $I=0$ state is dominant (this is the $\Delta I = 1/2$ rule). Defining

$$A_I = \langle(\pi\pi)_I|\mathcal{H}|K^0\rangle, \quad \bar{A}_I = \langle(\pi\pi)_I|\mathcal{H}|\bar{K}^0\rangle, \tag{4.6}$$

we have, experimentally, $|A_2/A_0| \approx 1/20$. Instead of $\eta_{00}$ and $\eta_{+-}$ we may define two combinations, $\varepsilon_K$ and $\varepsilon'_K$, in such a way that the possible effects of CP violation in decay (mixing) are isolated into $\varepsilon'_K$ ($\varepsilon_K$).

The experimental definition of the $\varepsilon_K$ parameter is

$$\varepsilon_K \equiv \frac{1}{3}(\eta_{00} + 2\eta_{+-}). \tag{4.7}$$

To zeroth order in $A_2/A_0$, we have $\eta_{00} = \eta_{+-} = \varepsilon_K$. However, the specific combination (4.7) is chosen in such a way that the following relation holds to *first* order in $A_2/A_0$:

$$\varepsilon_K = \frac{1-\lambda_0}{1+\lambda_0}, \tag{4.8}$$

where

$$\lambda_0 = \left(\frac{q}{p}\right)_K \left(\frac{\bar{A}_0}{A_0}\right). \tag{4.9}$$

Since, by definition, only one strong channel contributes to $\lambda_0$, there is indeed no CP violation in decay in (4.8). It is simple to show that $\mathcal{R}e\ \varepsilon_K \neq 0$ is a manifestation of CP violation in mixing while $\mathcal{I}m\ \varepsilon_K \neq 0$ is a manifestation of CP violation in the interference between decays with and without mixing. Since experimentally $\arg \varepsilon_K \approx \pi/4$, the two contributions are comparable. It is also clear that $\varepsilon_K \neq 0$ is a manifestation of indirect CP violation: it could be described entirely in terms of a CP violating phase in the $M_{12}$ amplitude.

The experimental value of $\varepsilon_K$ is given by [26]

$$|\varepsilon_K| = (2.280 \pm 0.013) \times 10^{-3}. \tag{4.10}$$

The experimental definition of the $\varepsilon'_K$ parameter is

$$\varepsilon'_K \equiv \frac{1}{3}(\eta_{+-} - \eta_{00}). \tag{4.11}$$

The theoretical expression is

$$\varepsilon'_K \approx \frac{1}{6}(\lambda_{00} - \lambda_{+-}). \tag{4.12}$$

Obviously, any type of CP violation which is independent of the final state does not contribute to $\varepsilon'_K$. Consequently, there is no contribution from CP violation in mixing to (4.12). It is simple to show that $\mathcal{R}e\ \varepsilon'_K \neq 0$ is a manifestation of CP violation in decay while $\mathcal{I}m\ \varepsilon'_K \neq 0$ is a manifestation of CP violation in the interference between decays with and without mixing. Following our explanations in the previous section, we learn that $\varepsilon'_K \neq 0$ is a manifestation of direct CP violation: it requires $\phi_2 - \phi_0 \neq 0$ (where $\phi_I$ is the CP violating phase in the $A_I$ amplitude defined in (4.6)).

The quantity that is actually measured in experiment is

$$1 - \left|\frac{\eta_{00}}{\eta_{+-}}\right|^2 = 6\mathcal{R}e(\varepsilon'/\varepsilon). \tag{4.13}$$

The average over the experimental measurements of $\varepsilon'/\varepsilon$ [45-49] is given by:

$$\mathcal{R}e(\varepsilon'/\varepsilon) = (2.11 \pm 0.46) \times 10^{-3}. \tag{4.14}$$

### *4.3. The $\varepsilon_K$ Parameter in the Standard Model*

An approximate expression for $\varepsilon_K$, that is convenient for calculating it, is given by

$$\varepsilon_K = \frac{e^{i\pi/4}}{\sqrt{2}} \frac{\mathcal{I}m M_{12}}{\Delta m_K}. \tag{4.15}$$

A few points concerning this expression are worth emphasizing:

(i) Eq. (4.15) is given in a specific phase convention, where $A_2$ is real. Within the Standard Model, this is a phase convention where $V_{ud}V_{us}^*$ is real, a condition fulfilled in both the standard parametrization of eq. (2.26) and the Wolfenstein parametrization of eq. (2.29).

(ii) The phase of $\pi/4$ is approximate. It is determined by hadronic parameters and therefore is independent of the electroweak model. Specifically,

$$\Delta\Gamma_K \approx -2\Delta m_K \implies \arg(\varepsilon_K) \approx \arctan(-2\Delta m_K/\Delta\Gamma_K) \approx \pi/4. \tag{4.16}$$

(iii) A term of order $2\frac{\mathcal{I}m\ A_0}{\mathcal{R}e\ A_0}\frac{\mathcal{R}e\ M_{12}}{\mathcal{I}m\ M_{12}} \lesssim 0.02$ is neglected when (4.15) is derived.

(iv) There is a large hadronic uncertainty in the calculation of $M_{12}$ coming from long distance contributions. There are, however, good reasons to believe that the long distance contributions are important in $\mathcal{R}e\ M_{12}$ (where they could be even comparable to the short distance contributions), but negligible in $\mathcal{I}m\ M_{12}$. To avoid this uncertainty, one uses $\mathcal{I}m M_{12}/\Delta m_K$, with the experimentally measured value of $\Delta m_K$, instead of $\mathcal{I}m M_{12}/2\mathcal{R}e\ M_{12}$ with the theoretically calculated value of $\mathcal{R}e\ M_{12}$.

(v) The matrix element $\langle \bar{K}^0|(\bar{s}d)_{V-A}(\bar{s}d)_{V-A}|K^0\rangle$ is yet another source of hadronic uncertainty. If both $\mathcal{I}m\ M_{12}$ and $\mathcal{R}e\ M_{12}$ were dominated by short distance contributions, the matrix element would have cancelled in the ratio between them. However, as explained above, this is not the case.

Within the Standard Model, $\mathcal{I}m\ M_{12}$ is accounted for by box diagrams. One gets:

$$\varepsilon_K = e^{i\pi/4} C_\varepsilon B_K \mathcal{I}m\lambda_t\{\mathcal{R}e\lambda_c[\eta_1 S_0(x_c) - \eta_3 S_0(x_c, x_t)] - \mathcal{R}e\lambda_t\eta_2 S_0(x_t)\}, \tag{4.17}$$

where the CKM parameters are $\lambda_i = V_{is}^* V_{id}$, the constant $C_\varepsilon$ is a well known parameter,

$$C_\varepsilon = \frac{G_F^2 f_K^2 m_K m_W^2}{6\sqrt{2}\pi^2\Delta m_K} = 3.8\times 10^4, \tag{4.18}$$

the $\eta_i$ are QCD correction factors [50],

$$\eta_1 = 1.38\pm 0.20, \quad \eta_2 = 0.57\pm 0.01, \quad \eta_3 = 0.47\pm 0.04, \tag{4.19}$$

$S_0$ is a kinematic factor, given approximately by [36]

$$S_0(x_t) = 2.4 \left(\frac{m_t}{170\ GeV}\right)^{1.52}, \quad S_0(x_c) = x_x = 2.6 \times 10^{-4},$$
$$S_0(x_t, x_c) = x_c \left[\ln\frac{x_t}{x_c} - \frac{3x_t}{4(1-x_t)} - \frac{3x_t^2 \ln x_t}{4(1-x_t)^2}\right] = 2.3 \times 10^{-3}, \tag{4.20}$$

and $B_K$ is the ratio between the matrix element of the four quark operator and its value in the vacuum insertion approximation (see [36] for a precise definition),

$$B_K = 0.85 \pm 0.15. \tag{4.21}$$

Note that $\varepsilon_K$ is proportional to $\mathcal{I}m\lambda_t$ and, consequently, to $\eta$. The $\varepsilon_K$ constraint on the Wolfenstein parameters can be written as follows:

$$\eta \left[(1-\rho)|V_{cb}|^2 \eta_2 S_0(x_t) + \eta_3 S_0(x_c, x_t) - \eta_1 S_0(x_c)\right] |V_{cb}|^2 B_K = 1.24 \times 10^{-6}. \tag{4.22}$$

Eq. (4.22) gives hyperbolae in the $\rho - \eta$ plane. The main sources of uncertainty are in the $B_K$ parameter and in the $|V_{cb}|^4$ dependence. When the theoretically reasonable range for the first and the experimentally allowed range for the second are taken into account, one finds that $\eta \gtrsim 0.2$ is required to explain $\varepsilon_K$ (see (2.39)). Hence our statement that CP is not an approximate symmetry of the Standard Model.

### *4.4. The $\varepsilon'_K$ Parameter in the Standard Model*

A convenient approximate expression for $\varepsilon'_K$ is given by:

$$\varepsilon'_K = \frac{i}{\sqrt{2}} \left|\frac{A_2}{A_0}\right| e^{i(\delta_2 - \delta_0)} \sin(\phi_2 - \phi_0). \tag{4.23}$$

We would like to emphasize a few points:

(i) The approximations used in (4.23) are $\lambda_0 = 1$ and $|A_2/A_0| \ll 1$.

(ii) The phase of $\varepsilon'_K$ is determined by hadronic parameters and, therefore, model independent:

$$\arg(\varepsilon'_K) = \pi/2 + \delta_2 - \delta_0 \approx \pi/4. \tag{4.24}$$

The fact that, accidentally, $\arg(\varepsilon_K) \approx \arg(\varepsilon'_K)$, means that

$$\mathcal{R}e(\varepsilon'/\varepsilon) \approx \varepsilon'/\varepsilon. \tag{4.25}$$

(iii) $\mathcal{R}e\ \varepsilon'_K \neq 0$ requires $\delta_2 - \delta_0 \neq 0$, consistent with our statement that it is a manifestation of CP violation in decay. $\varepsilon'_K \neq 0$ requires $\phi_2 - \phi_0 \neq 0$, consistent with our statement that it is a manifestation of direct CP violation.

The calculation of $\varepsilon'/\varepsilon$ within the Standard Model is complicated and suffers from large hadronic uncertainties. Let us first try a very naive order of magnitude estimate. The relevant quark decay process is $s \to d\bar{u}u$. All tree diagrams contribute with the same weak phase, $\phi_T = \arg(V^*_{ud}V_{us})$. Spectator diagrams contribute to both $I=0$ and $I=2$ final states. Penguin diagrams with an intermediate $q = u, c, t$ quarks, contribute with a weak phase $\phi^q_P = \arg(V^*_{qd}V_{qs})$. Strong penguin contribute to the final $I=0$ states only. Electroweak penguins contribute also to final $I=2$ states, but we will ignore them for the moment. (The fact that the top quark is heavy means that the electroweak penguins are important.) A difference in the weak phase between $A_0$ and $A_2$ is then a result of the fact that $A_0$ has contributions from penguin diagrams with intermediate $c$ and $t$ quarks. Consequently, $\varepsilon'_K$ is suppressed by the following factors:

a. $|A_2/A_0| \sim 0.045$;

b. $|A_0^{\rm penguin}/A_0^{\rm tree}| \sim 0.05$.

c. $\sin\beta_K \sim 10^{-3}$.

The last factor appears, however, also in $\varepsilon_K$. Therefore, it cancels in the ratio $\varepsilon'/\varepsilon$. A very rough order of magnitude estimate is then $\varepsilon'/\varepsilon \sim 10^{-3}$. Note that $\varepsilon'/\varepsilon$ is not small because of small CP violating parameters but because of hadronic parameters. Actually, it is independent of $\sin\delta_{\rm KM}$. (In most calculations one uses the experimental value of $\varepsilon_K$ and the theoretical expression for $\varepsilon'_K$. Then the expression for $\varepsilon'/\varepsilon$ depends on $\sin\delta_{\rm KM}$.)

The actual calculation of $\varepsilon'/\varepsilon$ is very complicated. There are several comparable contributions with differing signs. The final result can be written in the form (for recent work, see [51-54,36] and references therein):

$$\varepsilon'/\varepsilon = \mathcal{I}m(V_{td}V^*_{ts})\left[P^{(1/2)} - P^{(3/2)}\right], \tag{4.26}$$

where we omitted a phase factor using the approximation $\arg(\varepsilon_K) = \arg(\varepsilon'_K)$, and where

$$\begin{aligned} P^{(1/2)} &= \frac{G_F \mathcal{R}eA_2}{2|\varepsilon_K|(\mathcal{R}eA_0)^2} \sum y_i \langle Q_i \rangle_0 (1-\Omega_{\eta+\eta'}), \\ P^{(3/2)} &= \frac{G_F}{2|\varepsilon_K|\mathcal{R}eA_0} \sum y_i \langle Q_i \rangle_2. \end{aligned} \tag{4.27}$$

$Q_i$ are four quark operators, and $y_i$ are the Wilson coefficient functions. The most important operators are

$$\begin{aligned} Q_6 &= (\bar{s}_\alpha d_\beta)_{V-A} \sum_{q=u,d,s} (\bar{q}_\beta q_\alpha)_{V+A}, \\ Q_8 &= \frac{3}{2} (\bar{s}_\alpha d_\beta)_{V-A} \sum_{q=u,d,s} e_q (\bar{q}_\beta q_\alpha)_{V+A}. \end{aligned} \tag{4.28}$$

$P^{(3/2)}$ is dominated by electroweak penguin contributions while $P^{(1/2)}$ is dominated by QCD penguin contributions. The latter are suppressed by isospin breaking effects ($m_u \neq m_d$), parametrized by

$$\Omega_{\eta+\eta'} = \frac{\mathcal{R}eA_0}{\mathcal{R}eA_2} \frac{(\mathcal{I}mA_2)_{\text{I.B.}}}{\mathcal{I}mA_0} \approx 0.25 \pm 0.08. \tag{4.29}$$

A crude approximation to (4.26) which emphasizes the sources of hadronic uncertainty is given by

$$\begin{aligned} \varepsilon'/\varepsilon \approx 13\, \mathcal{I}m(V_{td}V_{ts}^*) \left(\frac{110\ MeV}{m_s(2\ GeV)}\right)^2 \left(\frac{\Lambda^{(4)}_{\overline{\text{MS}}}}{340\ MeV}\right) \\ \times \left[ B_6^{(1/2)}(1-\Omega_{\eta+\eta'}) - 0.4 B_8^{(3/2)} \left(\frac{m_t}{165\ GeV}\right)^{2.5} \right]. \end{aligned} \tag{4.30}$$

$B_6^{(1/2)}$ and $B_8^{(3/2)}$ parametrize the hadronic matrix elements:

$$\begin{aligned} \langle Q_6 \rangle_0 &\equiv B_6^{(1/2)} \langle Q_6 \rangle_0^{(\text{vac})}, \quad B_6^{(1/2)} \approx 1.0 \pm 0.3, \\ \langle Q_8 \rangle_2 &\equiv B_8^{(3/2)} \langle Q_8 \rangle_2^{(\text{vac})}, \quad B_8^{(3/2)} \approx 0.8 \pm 0.2. \end{aligned} \tag{4.31}$$

The main sources of uncertainties lie then in the parameters $m_s$, $B_6^{(1/2)}$, $B_8^{(3/2)}$, $\Omega_{\eta+\eta'}$ and $\Lambda^{(4)}_{\overline{\text{MS}}}$. The importance of these uncertainties is increased because of the cancellation between the two contributions in (4.30). Taking the above reasonable ranges for the hadronic parameters (from lattice calculations and $1/N_c$ expansion), one can estimate [53]:

$$\mathcal{R}e(\varepsilon'/\varepsilon)^{\text{SM}} = (7.7^{+6.0}_{-3.5}) \times 10^{-4}. \tag{4.32}$$

The fact that the range in (4.32) is lower than the experimentally allowed range in (4.14) cannot be taken as evidence for new physics. With a more conservative treatment of the theoretical uncertainties, one can stretch the theoretical upper bound to values consistent with the experimental lower bound [55-58,52-53].

*4.5. CP violation in $K \to \pi\nu\bar\nu$*

CP violation in the rare $K \to \pi\nu\bar\nu$ decays is very interesting. It is very different from the CP violation that has been observed in $K \to \pi\pi$ decays which is small and involves theoretical uncertainties. Similar to the CP asymmetry in $B \to \psi K_S$, it is predicted to be large and can be cleanly interpreted. Furthermore, observation of the $K_L \to \pi^0\nu\bar\nu$ decay at the rate predicted by the Standard Model will provide further evidence that CP violation cannot be attributed to mixing ($\Delta S = 2$) processes only, as in superweak models.

Define the decay amplitudes

$$A_{\pi^0\nu\bar\nu} = \langle \pi^0\nu\bar\nu|\mathcal{H}|K^0\rangle, \quad \bar A_{\pi^0\nu\bar\nu} = \langle \pi^0\nu\bar\nu|\mathcal{H}|\bar K^0\rangle, \tag{4.33}$$

and the related $\lambda_{\pi\nu\bar\nu}$ quantity:

$$\lambda_{\pi\nu\bar\nu} = \left(\frac{q}{p}\right)_K \frac{\bar A_{\pi^0\nu\bar\nu}}{A_{\pi^0\nu\bar\nu}}. \tag{4.34}$$

The decay amplitudes of $K_{L,S}$ into a final $\pi^0\nu\bar\nu$ state are then

$$\langle \pi^0\nu\bar\nu|\mathcal{H}|\bar K_{L,S}\rangle = pA_{\pi^0\nu\bar\nu} \mp q\bar A_{\pi^0\nu\bar\nu}, \tag{4.35}$$

and the ratio between the corresponding decay rates is

$$\frac{\Gamma(K_L \to \pi^0\nu\bar\nu)}{\Gamma(K_S \to \pi^0\nu\bar\nu)} = \frac{1 + |\lambda_{\pi\nu\bar\nu}|^2 - 2\mathcal{R}e\lambda_{\pi\nu\bar\nu}}{1 + |\lambda_{\pi\nu\bar\nu}|^2 + 2\mathcal{R}e\lambda_{\pi\nu\bar\nu}}. \tag{4.36}$$

We learn that the $K_L \to \pi^0\nu\bar\nu$ decay rate vanishes in the CP limit ($\lambda_{\pi\nu\bar\nu} = 1$), as expected on general grounds [16]. (The CP conserving contributions were explicitly calculated within the Standard Model [59] and within its extension with massive neutrinos [60], and found to be negligible.)

Since the effects of CP violation in decay and in mixing are expected to be negligibly small, $\lambda_{\pi\nu\bar\nu}$ is, to an excellent approximation, a pure phase. Defining $\theta_K$ to be the relative phase between the $K - \bar{K}$ mixing amplitude and the $s \to d\nu\bar{\nu}$ decay amplitude, namely $\lambda_{\pi\nu\bar{\nu}} = e^{2i\theta_K}$, we get from (4.36):

$$\frac{\Gamma(K_L \to \pi^0\nu\bar{\nu})}{\Gamma(K_S \to \pi^0\nu\bar{\nu})} = \frac{1 - \cos 2\theta_K}{1 + \cos 2\theta_K} = \tan^2\theta_K. \tag{4.37}$$

Using the isospin relation $A(K^0 \to \pi^0\nu\bar{\nu})/A(K^+ \to \pi^+\nu\bar{\nu}) = 1/\sqrt{2}$, we get

$$a_{\pi\nu\bar{\nu}} \equiv \frac{\Gamma(K_L \to \pi^0\nu\bar{\nu})}{\Gamma(K^+ \to \pi^+\nu\bar{\nu})} = \frac{1 - \cos 2\theta_K}{2} = \sin^2\theta_K. \tag{4.38}$$

Note that $a_{\pi\nu\bar{\nu}} \leq 1$, and consequently a measurement of $\Gamma(K^+ \to \pi^+\nu\bar{\nu})$ can be used to set a model independent upper limit on $\Gamma(K_L \to \pi^0\nu\bar{\nu})$ [19].

Within the Standard Model, the $K \to \pi\nu\bar{\nu}$ decays are dominated by short distance $Z$-penguins and box diagrams. The branching ratio for $K^+ \to \pi^+\nu\bar{\nu}$ can be expressed in terms of $\rho$ and $\eta$ [36]:

$$BR(K^+ \to \pi^+\nu\bar{\nu}) = 4.11 \times 10^{-11} A^4 [X(x_t)]^2 \left[\eta^2 + (\rho_0 - \rho)^2\right], \tag{4.39}$$

where

$$\rho_0 = 1 + \frac{P_0(X)}{A^2 X(x_t)}, \tag{4.40}$$

and $X(x_t)$ and $P_0(X)$ represent the electroweak loop contributions in NLO for the top quark and for the charm quark, respectively. The main theoretical uncertainty is related to the strong dependence of the charm contribution on the renormalization scale and the QCD scale, $P_0(X) = 0.42 \pm 0.06$. First evidence for $K^+ \to \pi^+\nu\bar{\nu}$ was presented recently [61]. The large experimental error does not yet give a useful CKM constraint and is consistent with the Standard Model prediction.

The branching ratio for the $K_L \to \pi^0\nu\bar{\nu}$ decay can be expressed in terms of $\eta$ [36]:

$$BR(K_L \to \pi^0\nu\bar{\nu}) = 1.80 \times 10^{-10} [X(x_t)]^2 A^4 \eta^2. \tag{4.41}$$

The present experimental bound, $BR(K_L \to \pi^0\nu\bar{\nu}) \leq 1.6 \times 10^{-6}$ [62] lies about five orders of magnitude above the Standard Model prediction [36] and about two orders of magnitude above the bound that can be deduced using model independent isospin relations [19] from the experimental upper bound on the charged mode.

Note that if the charm contribution to $K^+ \to \pi^+ \nu\bar{\nu}$ were negligible, so that both the charged and the neutral decays were dominated by the intermediate top contributions, then $a_{\pi\nu\bar{\nu}}$ would simply measure $\sin^2\beta$. While the charm contribution makes the evaluation of $a_{\pi\nu\bar{\nu}}$ more complicated, a reasonable order of magnitude estimate is still given by $\sin^2\beta$.

## 5. CP Violation in $D \to K\pi$ Decays

Within the Standard Model, $D-\bar{D}$ mixing is expected to be well below the experimental bound. Furthermore, effects related to CP violation in $D-\bar{D}$ mixing are expected to be negligibly small since this mixing is described to a good approximation by physics of the first two generations. An experimental observation of $D-\bar{D}$ mixing close to the present bound (and, even more convincingly, of related CP violation) will then be evidence for New Physics. The most sensitive experimental searches for $D-\bar{D}$ mixing use $D \to K\pi$ decays [63-67]. We now give the formalism of neutral $D$ decays into final $K^\pm\pi^\mp$ states.

### *5.1. Formalism*

We define the neutral $D$ mass eigenstates:

$$|D_{1,2}\rangle = p|D^0\rangle \pm q|\bar{D}^0\rangle. \tag{5.1}$$

We define the following four decay amplitudes:

$$\begin{aligned} A_{K^+\pi^-} &= \langle K^+\pi^-|\mathcal{H}|D^0\rangle, \quad \bar{A}_{K^+\pi^-} = \langle K^+\pi^-|\mathcal{H}|\bar{D}^0\rangle, \\ A_{K^-\pi^+} &= \langle K^-\pi^+|\mathcal{H}|D^0\rangle, \quad \bar{A}_{K^-\pi^+} = \langle K^-\pi^+|\mathcal{H}|\bar{D}^0\rangle. \end{aligned} \tag{5.2}$$

We introduce the following two quantities:

$$\lambda_{K^+\pi^-} = \left(\frac{q}{p}\right)_D \frac{\bar{A}_{K^+\pi^-}}{A_{K^+\pi^-}}, \quad \lambda_{K^-\pi^+} = \left(\frac{q}{p}\right)_D \frac{\bar{A}_{K^-\pi^+}}{A_{K^-\pi^+}}. \tag{5.3}$$

The following approximations are all experimentally confirmed:

$$x \equiv \frac{\Delta m_D}{\Gamma_D} \ll 1; \quad y \equiv \frac{\Delta\Gamma_D}{2\Gamma_D} \ll 1; \quad |\lambda^{-1}_{K^+\pi^-}| \ll 1; \quad |\lambda_{K^-\pi^+}| \ll 1. \tag{5.4}$$

Using these approximations, the decay rates for the doubly Cabibbo suppressed (DCS) decays are given by

$$\begin{aligned}
\Gamma[D^0(t)\to K^+\pi^-] &= e^{-t}\left|\bar{A}_{K^+\pi^-}\right|^2|q/p|^2\times \\
&\left[\left|\lambda^{-1}_{K^+\pi^-}\right|^2+\mathcal{R}e(\lambda^{-1}_{K^+\pi^-})yt+\mathcal{I}m(\lambda^{-1}_{K^+\pi^-})xt+\frac{1}{4}(x^2+y^2)t^2\right], \\
\Gamma[\bar{D}^0(t)\to K^-\pi^+] &= e^{-t}\left|A_{K^-\pi^+}\right|^2|p/q|^2\times \\
&\left[\left|\lambda_{K^-\pi^+}\right|^2+\mathcal{R}e(\lambda_{K^-\pi^+})yt+\mathcal{I}m(\lambda_{K^-\pi^+})xt+\frac{1}{4}(x^2+y^2)t^2\right].
\end{aligned} \tag{5.5}$$

The time $t$ is given here in units of the $D$-lifetime, $\tau_D=1/\Gamma_D$. Eqs. (5.5) are valid for times $\lesssim \tau_D$.

The decay rates for the Cabibbo favored (CF) modes are given to a good approximation by

$$\begin{aligned}
\Gamma[D^0(t)\to K^-\pi^+] &= e^{-t}\left|A_{K^-\pi^+}\right|^2, \\
\Gamma[\bar{D}^0(t)\to K^+\pi^-] &= e^{-t}\left|\bar{A}_{K^+\pi^-}\right|^2.
\end{aligned} \tag{5.6}$$

*5.2. CP Violation*

We will assume that the CF decays are unaffected by CP violation, that is,

$$\left|A_{K^-\pi^+}\right| = \left|\bar{A}_{K^+\pi^-}\right| \equiv A_{\rm CF}. \tag{5.7}$$

In general, $|q/p|$ is a positive real number and $\lambda^{-1}_{K^+\pi^-}$ and $\lambda_{K^-\pi^+}$ are two independent complex numbers. We now parametrize these quantities in a way that is convenient to separate the three types of CP violation:

$$\begin{aligned}
|q/p| &= R_m, \\
\lambda^{-1}_{K^+\pi^-} &= \frac{R}{R_d R_m}e^{-i(\delta_{K\pi}+\phi_{K\pi})}, \\
\lambda_{K^-\pi^+} &= R R_d R_m e^{-i(\delta_{K\pi}-\phi_{K\pi})}.
\end{aligned} \tag{5.8}$$

CP violation in mixing, that is violation of $|q/p|=1$, is related to $R_m\neq 1$. CP violation in decay, that is violation of $\left|\frac{A_{K^+\pi^-}}{\bar{A}_{K^-\pi^+}}\right| = \left|\frac{\bar{A}_{K^+\pi^-}}{A_{K^-\pi^+}}\right| = 1$, is related to $R_d\neq 1$. CP violation in interference of decays with and without mixing, that is violation of $\frac{\mathcal{I}m(\lambda^{-1}_{K^+\pi^-})}{|\lambda^{-1}_{K^+\pi^-}|} = \frac{\mathcal{I}m(\lambda_{K^-\pi^+})}{|\lambda_{K^-\pi^+}|}$, is related to $\phi_{K\pi}\neq 0$.

We also define

$$
\begin{aligned}
x' &\equiv x\cos\delta_{K\pi} + y\sin\delta_{K\pi},\\
y' &\equiv y\cos\delta_{K\pi} - x\sin\delta_{K\pi}.
\end{aligned}
\tag{5.9}
$$

In the language of eqs. (5.7), (5.8) and (5.9), we can rewrite eq. (5.5) as follows:

$$
\begin{aligned}
\Gamma[D^0(t)\to K^+\pi^-] &= e^{-t}A_{\rm CF}^2\left[\frac{R^2}{R_d^2} + \frac{RR_m}{R_d}(y'c_\phi - x's_\phi)t + \frac{R_m^2}{4}(x^2+y^2)t^2\right],\\
\Gamma[\bar D^0(t)\to K^-\pi^+] &= e^{-t}A_{\rm CF}^2\left[R^2R_d^2 + \frac{RR_d}{R_m}(y'c_\phi + x's_\phi)t + \frac{1}{4R_m^2}(x^2+y^2)t^2\right],
\end{aligned}
\tag{5.10}
$$

where $s_\phi \equiv \sin\phi_{K\pi}$ and $c_\phi \equiv \cos\phi_{K\pi}$. Note that the three different mechanisms of CP violation can be distinguished if the time dependent rates (5.10) are measured:

(i) A different $t^2e^{-t}$ term in the DCS decays of $D^0$ and $\bar D^0$ is an unambiguous signal of CP violation in mixing.

(ii) A different $e^{-t}$ term in the DCS decays of $D^0$ and $\bar D^0$ is an unambiguous signal of CP violation in decay.

(iii) A measurement of all three terms for each of the two decay rates can provide an unambiguous signal of CP violation in interference.

In the presence of new physics, the most likely situation is that we have observable CP violation in interference between decays with and without mixing, while CP violation in mixing and in decays [68] remain unobservably small. In this case, $R_m = 1$, $R_d = 1$, but $\phi_{K\pi} \neq 0$. Furthermore, while $\Delta m_D$ can be enhanced to the level of present experimental sensitivity, $\Delta\Gamma_D$ is likely to remain close to the SM prediction. Neglecting $y$, we have in this case

$$
\begin{aligned}
\Gamma[D^0(t)\to K^+\pi^-] &= e^{-t}A_{\rm CF}^2\left[R^2 - Rx(s_\delta c_\phi + c_\delta s_\phi)t + \frac{x^2}{4}t^2\right],\\
\Gamma[\bar D^0(t)\to K^-\pi^+] &= e^{-t}A_{\rm CF}^2\left[R^2 - Rx(s_\delta c_\phi - c_\delta s_\phi)t + \frac{x^2}{4}t^2\right].
\end{aligned}
\tag{5.11}
$$

The $te^{-t}$ terms in (5.11) are potentially CP violating. There are four possibilities concerning these terms [69,70]:

(i) They vanish: both strong and weak phases play no role.

(ii) They are equal in magnitude and in sign: weak phases play no role.

(iii) They are equal in magnitude but have opposite signs: strong phases play no role.

(iv) They have different magnitudes: both strong and weak phases play a role.

## 6. CP Violation in $B$ Decays in the Standard Model

### *6.1.* $|q/p| \neq 1$

As explained in the previous section, in the $B_d$ system we expect model independently that $\Gamma_{12} \ll M_{12}$. Within any given model we can actually calculate the two quantities from quark diagrams. Within the Standard Model, $M_{12}$ is given by box diagrams. For both the $B_d$ and $B_s$ systems, the long distance contributions are expected to be negligible and the calculation of these diagrams with a high loop momentum is a very good approximation. $\Gamma_{12}$ is calculated from a cut of box diagrams [71]. Since the cut of a diagram always involves on-shell particles and thus long distance physics, the cut of the quark box diagram is a poor approximation to $\Gamma_{12}$. However, it does correctly give the suppression from small electroweak parameters such as the weak coupling. In other words, though the hadronic uncertainties are large and could change the result by order 50%, the cut in the box diagram is expected to give a reasonable order of magnitude estimate of $\Gamma_{12}$. (For $\Gamma_{12}(B_s)$ it has been shown that local quark-hadron duality holds exactly in the simultaneous limit of small velocity and large number of colors. We thus expect an uncertainty of $\mathcal{O}(1/N_C) \sim 30\%$ [72,73]. For $\Gamma_{12}(B_d)$ the small velocity limit is not as good an approximation but an uncertainty of order 50% still seems a reasonable estimate [74].)

Within the Standard Model, $M_{12}$ is dominated by top-mediated box diagrams [75]:

$$M_{12} = \frac{G_F^2}{12\pi^2} m_B m_W^2 \eta_B B_B f_B^2 (V_{tb} V_{td}^*)^2 S_0(x_t), \tag{6.1}$$

where $S_0(x_t)$ is given in eq. (4.20), $\eta_B = 0.55$ is a QCD correction, and $B_B f_B^2$ parametrizes the hadronic matrix element. For $\Gamma_{12}$, we have [76-78]

$$\begin{aligned}\Gamma_{12} = & -\frac{G_F^2}{24\pi} m_B m_b^2 B_B f_B^2 (V_{tb} V_{td}^*)^2 \\ & \times \left[\frac{5}{3}\frac{m_B^2}{(m_b+m_d)^2}\frac{B_S}{B_B}(K_2-K_1) + \frac{4}{3}(2K_1+K_2) + 8(K_1+K_2)\frac{m_c^2}{m_b^2}\frac{V_{cb}V_{cd}^*}{V_{tb}V_{td}^*}\right],\end{aligned} \tag{6.2}$$

where $K_1 = -0.39$ and $K_2 = 1.25$ [78] are combinations of Wilson coefficients and $B_S$ parametrizes the $(S-P)^2$ matrix element. Note that new physics is not expected to affect $\Gamma_{12}$ significantly because it usually takes place at a high energy scale and is relevant to the short distance part only. Therefore, the Standard Model estimate in eq. (6.2) remains valid model independently. Combining (6.1) and (6.2), one gets

$$\frac{\Gamma_{12}}{M_{12}} = -5.0 \times 10^{-3} \left( 1.4 \frac{B_S}{B_B} + 0.24 + 2.5 \frac{m_c^2}{m_b^2} \frac{V_{cb} V_{cd}^*}{V_{tb} V_{td}^*} \right). \tag{6.3}$$

We learn that $|\Gamma_{12}/M_{12}| = \mathcal{O}(m_b^2/m_t^2)$, which confirms our model independent order of magnitude estimate, $|\Gamma_{12}/M_{12}| \lesssim 10^{-2}$. As concerns the imaginary part of this ratio, we have

$$a_{\rm SL} = \mathcal{I}m \frac{\Gamma_{12}}{M_{12}} = -1.1 \times 10^{-3} \sin\beta = -(2-5) \times 10^{-4}. \tag{6.4}$$

The strong suppression of $a_{\rm SL}$ compared to $|\Gamma_{12}/M_{12}|$ comes from the fact that the leading contribution to $\Gamma_{12}$ has the same phase as $M_{12}$. Consequently, $a_{\rm SL} = \mathcal{O}(m_c^2/m_t^2)$. The CKM factor, $\mathcal{I}m \frac{V_{cb}V_{cd}^*}{V_{tb}V_{td}^*} = \sin\beta$, is of order one. In contrast, for the $B_s$ system, where (6.3) holds except that $V_{td}$ ($V_{cd}$) is replaced by $V_{ts}$ ($V_{cs}$), there is an additional suppression from $\mathcal{I}m \frac{V_{cb}V_{cs}^*}{V_{tb}V_{ts}^*} = \sin\beta_s \sim 10^{-2}$ (see the corresponding unitarity triangle).

*6.2.* $|\bar{A}_{\bar{f}}/A_f| \neq 1$

In the previous subsection we estimated the effect of CP violation in mixing to be of $\mathcal{O}(10^{-3})$ within the Standard Model, and $\leq \mathcal{O}(|\Gamma_{12}/M_{12}|) \sim 10^{-2}$ model independently (for recent discussions, see [79-80,14]). In semileptonic decays, CP violation in mixing is the leading effect and therefore it can be measured through $a_{\rm SL}$. In purely hadronic $B$ decays, however, CP violation in decay and in the interference of decays with and without mixing is $\geq \mathcal{O}(10^{-2})$. We can therefore safely neglect CP violation in mixing in the following discussion and use

$$\frac{q}{p} = -\frac{M_{12}^*}{|M_{12}|} = \frac{V_{tb}^* V_{td}}{V_{tb} V_{td}^*} \omega_B. \tag{6.5}$$

(From here on we omit the convention-dependent quark phases $\omega_q$ defined in eq. (3.15). Our final expressions for physical quantities are of course unaffected by such omission.)

A crucial question is then whether CP violation in decay is comparable to the CP violation in the interference of decays with and without mixing or negligible. In the first case, we can use the corresponding charged $B$ decays to observe effects of CP violation in decay. In the latter case, CP asymmetries in neutral $B$ decays are subject to clean theoretical interpretation: we will either have precise measurements of CKM parameters or be provided with unambiguous evidence for new physics. The question of the relative size of CP violation in decay can only be answered on a channel by channel basis, which is what we do in this section.

Channels that have contributions from tree diagrams only depend each on a single CKM combination:

$$\begin{aligned} A(c\bar{u}d) &= T_{c\bar{u}d} V_{cb} V_{ud}^*, \\ A(c\bar{u}s) &= T_{c\bar{u}s} V_{cb} V_{us}^*, \\ A(u\bar{c}d) &= T_{u\bar{c}d} V_{ub} V_{cd}^*, \\ A(u\bar{c}s) &= T_{u\bar{c}s} V_{ub} V_{cs}^*. \end{aligned} \tag{6.6}$$

The subdivision of tree processes into spectator, exchange and annihilation diagrams is unimportant in this respect since they all carry the same weak phase. For such modes, $|\bar{A}/A| = 1$. It is possible that $\mathcal{I}m\lambda \neq 0$, but since the final states are not CP eigenstates, a clean theoretical interpretation is difficult. We do not discuss these modes here any further.

Most channels have contributions from both tree- and three types of penguin-diagrams, the latter classified according to the identity of the quark in the loop, as diagrams with different intermediate quarks may have both different strong phases and different weak phases [81]. Consider first $b \to q\bar{q}s$ decays, with $q = c$ or $u$. Using (2.32), we can write the CKM dependence of these amplitudes as follows:

$$\begin{aligned} A(c\bar{c}s) &= (T_{c\bar{c}s} + P_s^c - P_s^t) V_{cb} V_{cs}^* + (P_s^u - P_s^t) V_{ub} V_{us}^*, \\ A(u\bar{u}s) &= (P_s^c - P_s^t) V_{cb} V_{cs}^* + (T_{u\bar{u}s} + P_s^u - P_s^t) V_{ub} V_{us}^*, \end{aligned} \tag{6.7}$$

where $T$ stands for the tree amplitude and $P^q$ for a penguin diagram with an intermediate $q$-quark. Next, consider $b \to q\bar{q}d$ decays, with $q = c$ or $u$. Using (2.33), we can write the CKM dependence of these amplitudes as follows:

$$\begin{aligned} A(c\bar{c}d) &= (P_d^t - P_d^u) V_{tb} V_{td}^* + (T_{c\bar{c}d} + P_d^c - P_d^u) V_{cb} V_{cd}^*, \\ A(u\bar{u}d) &= (P_d^t - P_d^c) V_{tb} V_{td}^* + (T_{u\bar{u}d} + P_d^u - P_d^c) V_{ub} V_{ud}^*. \end{aligned} \tag{6.8}$$

Note that in both (6.7) and (6.8) only differences of penguin contributions occur, which makes the cancellation of the ultraviolet divergences of these diagrams explicit.

To estimate the size of CP violation in decay for these channels, we need to know the ratio of the contribution from the difference between a top and light quark strong penguin diagram to the contribution from a tree diagram (with the CKM combination factored out):

$$r_{PT} = \frac{P^t - P^{\rm light}}{T_{q\bar{q}q'}} \approx \frac{\alpha_s}{12\pi} \ln \frac{m_t^2}{m_b^2} = \mathcal{O}(0.03). \tag{6.9}$$

However, this estimate does not include the effect of hadronic matrix elements, which are the probability factor to produce a particular final state particle content from a particular quark content. Since this probability differs for different kinematics, color flow and spin structures, it can be different for tree and penguin contributions and may partially compensate the coupling constant suppression of the penguin term. Recent CLEO results on $BR(B \to K\pi)$ and $BR(B \to \pi\pi)$ [82] suggest that the matrix element of penguin operators is indeed enhanced compared to that of tree operators. The enhancement could be by a factor of a few, leading to

$$r_{PT} \sim \lambda^2 - \lambda. \tag{6.10}$$

(Note that $r_{PT}$ does not depend on the CKM parameters. We use powers of the Wolfenstein parameter $\lambda$ to quantify our estimate for $r_{PT}$ in order to simplify the comparison between the size of CP violation in decay and CP violation in the interference between decays with and without mixing.) Take, for example, the $b \to c\bar{c}s$ decays. Using eqs. (3.28) and (6.7), the size of CP violation in decay is then estimated as follows:

$$1 - \left|\frac{\bar{A}_{c\bar{c}s}}{A_{c\bar{c}\bar{s}}}\right| \lesssim r_{PT}\ \mathcal{I}m \frac{V_{ub}V_{us}^*}{V_{cb}V_{cs}^*} = \mathcal{O}(\lambda^4 - \lambda^3). \tag{6.11}$$

Note that, in the language of eq. (3.28), this estimate includes $(A_2/A_2)\sin(\phi_2 - \phi_1)$ but not $\sin(\delta_2 - \delta_1)$. We only used $\sin(\delta_2 - \delta_1) \leq 1$ but if, for some reason, the difference in strong phases is small, the effect of CP violation in decay will be accordingly suppressed compared to (6.11).

Finally, processes involving only down-type quarks have no contributions from tree diagrams:

$$\begin{aligned} A(s\bar{s}s) &= (P_s^c - P_s^t)V_{cb}V_{cs}^* + (P_s^u - P_s^t)V_{ub}V_{us}^*, \\ A(s\bar{s}d) &= (P_d^t - P_d^u)V_{tb}V_{td}^* + (P_d^c - P_d^u)V_{cb}V_{cd}^*. \end{aligned} \tag{6.12}$$

For the estimate of CP violation in decay for these modes, we need to consider two types of ratios between penguin diagrams:

$$\begin{aligned} r_{P^cP^u} &= \frac{P^u - P^t}{P^c - P^t} \approx 1, \\ r_{P^lP^t} &= \frac{P^c - P^u}{P^t - P^{\rm light}} \approx 0.1. \end{aligned} \tag{6.13}$$

In the $m_c = m_u$ limit, we would have $r_{P^cP^u} = 1$ and $r_{P^lP^t} = 0$. The deviations from these limiting values should then be GIM suppressed [83]. The estimate of the somewhat surprisingly large $r_{P^lP^t}$ is based on refs. [84,85]. We get:

$$1 - \left|\frac{\bar{A}_{s\bar{s}s}}{A_{s\bar{s}s}}\right| \lesssim r_{P^lP^t}\ \mathcal{I}m\frac{V_{ub}V_{us}^*}{V_{cb}V_{cs}^*} = \mathcal{O}(\lambda^2), \tag{6.14}$$

$$1 - \left|\frac{\bar{A}_{s\bar{s}d}}{A_{s\bar{s}\bar{d}}}\right| \lesssim r_{P^uP^c}\ \mathcal{I}m\frac{V_{cb}V_{cd}^*}{V_{tb}V_{td}^*} = \mathcal{O}(0.1). \tag{6.15}$$

As concerns the $b \to d\bar{d}s$ and $b \to d\bar{d}d$ processes, they mix strongly through rescattering effects with the tree mediated $b \to u\bar{u}s$ and $b \to u\bar{u}d$ decays, respectively. It is difficult to estimate these soft rescattering effects and we do not consider these modes here any further.

We thus classify the $B$ decays described in eqs. (6.7), (6.8) and (6.12) into four classes. Classes (i) and (ii) are expected to have relatively small CP violation in decay and hence are particularly interesting for extracting CKM parameters from interference of decays with and without mixing. In classes (iii) and (iv), CP violation in decay could be significant and might be observable in charged $B$ decays.

(i) Decays dominated by a single term: $b \to c\bar{c}s$ and $b \to s\bar{s}s$. The Standard Model predicts very small CP violation in decay: $\mathcal{O}(\lambda^4 - \lambda^3)$ for $b \to c\bar{c}s$ and $\mathcal{O}(\lambda^2)$ for $b \to s\bar{s}s$. Any observation of large CP asymmetries in charged $B$ decays for these channels would be a clue to physics beyond the Standard Model. The corresponding neutral modes have cleanly predicted relationships between CKM parameters and the measured asymmetry from interference between decays with and without mixing. The modes $B \to \psi K$ and $B \to \phi K$ are examples of this class.

(ii) Decays with a small second term: $b \to c\bar{c}d$ and $b \to u\bar{u}d$. The expectation that penguin-only contributions are suppressed compared to tree contributions suggests that these modes will have small effects of CP violation in decay, of $\mathcal{O}(\lambda^2 - \lambda)$, and an approximate prediction for the relationship between measured asymmetries in neutral decays and CKM phases can be made. Examples here are $B \to DD$ and $B \to \pi\pi$.

(iii) Decays with a suppressed tree contribution: $b \to u\bar{u}s$. The tree amplitude is suppressed by small mixing angles, $V_{ub}V_{us}$. The no-tree term may be comparable or even dominate and give large interference effects. An example is $B \to \rho K$.

(iv) Decays with no tree contribution and a small second term: $b \to s\bar{s}d$. Here the interference comes from penguin contributions with different charge 2/3 quarks in the loop and gives CP violation in decay that could be as large as 10%. An example is $B \to KK$.

Note that if the penguin enhancement is significant, then some of the decay modes listed in class (ii) might actually fit better in class (iii). For example, it is possible that $b \to u\bar{u}d$ decays have comparable contributions from tree and penguin amplitudes. On the other hand, this would also mean that some modes listed in class (iii) could be dominated by a single penguin term. For such cases an approximate relationship between measured asymmetries in neutral decays and CKM phases can be made.

A summary of our discussion in this section is given in the table II.

| Quark process | Sample $B^\pm$ mode | $\mathcal{O}(1 - \vert\bar{A}/A\vert)$ |
|---|---|---|
| $b \to c\bar{c}s$ | $\psi K^\pm$ | $r_{PT} \sin\beta_s \sim \lambda^4 - \lambda^3$ |
| $b \to s\bar{s}s$ | $\phi K^\pm$ | $\sin\beta_s \sim \lambda^2$ |
| $b \to u\bar{u}d$ | $\pi^0\pi^\pm$ | $r_{PT} \sin\alpha \sim \lambda^2 - \lambda$ |
| $b \to c\bar{c}d$ | $DD^\pm$ | $r_{PT} \sin\gamma \sim \lambda^2 - \lambda$ |
| $b \to u\bar{u}s$ | $\pi^0 K^\pm$ | $r_{PT}^{-1} \sin\beta_s \sim \lambda - 1$ |
| $b \to s\bar{s}d$ | $\phi\pi^\pm$ | $r_{PP} \sin\beta \sim 0.1$ |

Table II. CP violation in $B$ decays.

### *6.3.* $\mathcal{I}m\lambda_{f_{\rm CP}} \neq 0$

Let us first discuss an example of class (i), $B \to \psi K_S$. A new ingredient in the analysis is the effect of $K - \bar{K}$ mixing. For decays with a single $K_S$ in the final state,

$K-\bar{K}$ mixing is essential because $B^0 \to K^0$ and $\bar{B}^0 \to \bar{K}^0$, and interference is possible only due to $K-\bar{K}$ mixing. This adds a factor of

$$\left(\frac{p}{q}\right)_K = \frac{V_{cs}V_{cd}^*}{V_{cs}^*V_{cd}}\omega_K^* \tag{6.16}$$

into $(\bar{A}/A)$. The quark subprocess in $\bar{B}^0 \to \psi\bar{K}^0$ is $b \to c\bar{c}s$ which is dominated by the $W$-mediated tree diagram:

$$\frac{\bar{A}_{\psi K_S}}{A_{\psi K_S}} = \eta_{\psi K_S}\left(\frac{V_{cb}V_{cs}^*}{V_{cb}^*V_{cs}}\right)\left(\frac{V_{cs}V_{cd}^*}{V_{cs}^*V_{cd}}\right)\omega_B^*. \tag{6.17}$$

The CP-eigenvalue of the state is $\eta_{\psi K_S} = -1$. Combining eqs. (6.5) and (6.17), we find

$$\lambda(B \to \psi K_S) = -\left(\frac{V_{tb}^*V_{td}}{V_{tb}V_{td}^*}\right)\left(\frac{V_{cb}V_{cs}^*}{V_{cb}^*V_{cs}}\right)\left(\frac{V_{cd}^*V_{cs}}{V_{cd}V_{cs}^*}\right) \Longrightarrow \mathcal{I}m\lambda_{\psi K_S} = \sin(2\beta). \tag{6.18}$$

We have seen in the previous section that, for $b \to c\bar{c}s$ decays, we have a very small CP violation in decay, $1-|\bar{A}/A| \sim \lambda^2 r_{PT}$. Consequently, eq. (6.18) is clean of hadronic uncertainties to better than $\mathcal{O}(10^{-2})$. This means that the measuremnet of $a_{\psi K_S}$ can give the theoretically cleanest determination of a CKM parameter, even cleaner than the determination of $|V_{us}|$ from $K \to \pi\ell\nu$. (If BR$(K_L \to \pi\nu\bar{\nu})$ is measured, it will give a comparably clean determination of $\eta$.)

A second example of a theoretically clean mode in class (i) is $B \to \phi K_S$. We showed in the previous section that, for $b \to s\bar{s}s$ decays, we have small CP violation in decay, $1-|\bar{A}/A| = \mathcal{O}(\lambda^2) = \mathcal{O}(0.05)$. We can neglect this effect. The analysis is similar to the $\psi K_S$ case, and the asymmetry is proportional to $\sin(2\beta)$.

The same quark subprocesses give theoretically clean CP asymmetries also in $B_s$ decays. These asymmetries are, however, very small since the relative phase between the mixing amplitude and the decay amplitudes ($\beta_s$ defined in (2.37)) is very small.

The best known example of class (ii) is $B \to \pi\pi$. The quark subprocess is $b \to u\bar{u}d$ which is dominated by the $W$-mediated tree diagram. Neglecting for the moment the second, pure penguin, term we find

$$\frac{\bar{A}_{\pi\pi}}{A_{\pi\pi}} = \eta_{\pi\pi}\frac{V_{ub}V_{ud}^*}{V_{ub}^*V_{ud}}\omega_B^*. \tag{6.19}$$

The CP eigenvalue for two pions is +1. Combining eqs. (6.5) and (6.19), we get

$$\lambda(B \to \pi^+\pi^-) = \left(\frac{V_{tb}^* V_{td}}{V_{tb} V_{td}^*}\right)\left(\frac{V_{ud}^* V_{ub}}{V_{ud} V_{ub}^*}\right) \implies \mathcal{I}m\lambda_{\pi\pi} = \sin(2\alpha). \tag{6.20}$$

The pure penguin term in $A(u\bar{u}d)$ in eq. (6.8) has a weak phase, $\arg(V_{td}^* V_{tb})$, different from the term with the tree contribution, so it modifies both $\mathcal{I}m\lambda_{\pi\pi}$ and (if there are non-trivial strong phases) $|\lambda_{\pi\pi}|$. The recent CLEO results mentioned above suggest that the penguin contribution to $B \to \pi\pi$ channel is significant, probably 10% or more. This then introduces CP violation in decay, unless the strong phases cancel (or are zero, as suggested by factorization arguments). The resulting hadronic uncertainty can be eliminated using isospin analysis [86]. This requires a measurement of the rates for the isospin-related channels $B^+ \to \pi^+\pi^0$ and $B^0 \to \pi^0\pi^0$ as well as the corresponding CP-conjugate processes. The rate for $\pi^0\pi^0$ is expected to be small and the measurement is difficult, but even an upper bound on this rate can be used to limit the magnitude of hadronic uncertainties [87].

Related but slightly more complicated channels with the same underlying quark structure are $B \to \rho^0\pi^0$ and $B \to a_1^0\pi^0$. Again an analysis involving the isospin-related channels can be used to help eliminate hadronic uncertainties from CP violations in the decays [88,89]. Channels such as $\rho\rho$ and $a_1\rho$ could in principle also be studied, using angular analysis to determine the mixture of CP-even and CP-odd contributions.

The analysis of $B \to D^+D^-$ proceeds along very similar lines. The quark subprocess here is $b \to c\bar{c}d$, and so the tree contribution gives

$$\lambda(B \to D^+D^-) = \eta_{D^+D^-}\left(\frac{V_{tb}^* V_{td}}{V_{tb} V_{td}^*}\right)\left(\frac{V_{cd}^* V_{cb}}{V_{cd} V_{cb}^*}\right) \implies \mathcal{I}m\lambda_{DD} = -\sin(2\beta), \tag{6.21}$$

where we used $\eta_{D^+D^-} = +1$. Again, there are hadronic uncertainties due to the pure penguin term in (6.8), but they are estimated to be small.

A summary of our results for CP violation in the interference of decays with and without mixing in $B_q \to f_{\rm CP}$ is given in table III. For each mode, we give the asymmetry that would arise if the dominant contribution were the *only* contribution.

| Quark process | Sample $B_d$ mode | $\mathcal{I}m\lambda_{B_d\to f_{\rm CP}}$ | Sample $B_s$ mode | $\mathcal{I}m\lambda_{B_s\to f_{\rm CP}}$ |
|---|---|---|---|---|
| $b\to c\bar c s$ | $\psi K_S$ | $\sin 2\beta$ | $D_s\overline{D_s}$ | $\sin 2\beta_s$ |
| $b\to s\bar s s$ | $\phi K_S$ | $\sin 2\beta$ | $\phi\eta'$ | 0 |
| $b\to u\bar u d$ | $\pi\pi$ | $\sin 2\alpha$ | $\pi^0 K_S$ | $\sin 2\gamma$ |
| $b\to c\bar c d$ | $D^+D^-$ | $\sin 2\beta$ | $\psi K_S$ | $\sin 2\beta_s$ |
| $b\to u\bar u s$ | $\pi^0 K_S$ | $\sin 2\beta$ | $\phi\pi^0$ | 0 |
| $b\to s\bar s d$ | $\phi\pi$ | 0 | $\phi K_S$ | $\sin 2\beta$ |

Table III. $\mathcal{I}m\lambda(B_q\to f_{\rm CP})$.

In all cases the above discussions have neglected the distinction between strong penguins and electroweak penguins. The CKM phase structure of both types of penguins is the same. The only place where this distinction becomes important is when an isospin argument is used to remove hadronic uncertainties due to penguin contributions. These arguments are based on the fact that gluons have isospin zero, and hence strong penguin processes have definite $\Delta I$. Photons and $Z$-bosons on the other hand contribute to more than one $\Delta I$ transition and hence cannot be separated from tree terms by isospin analysis. In most cases electroweak penguins are small, typically no more than ten percent of the corresponding strong penguins and so their effects can safely be neglected. However in cases (iii) and (iv), where tree contributions are small or absent, their effects may need to be considered. (A full review of the role of electroweak penguins in $B$ decays has been given in ref. [90].)

## 7. CP Violation Can Probe New Physics

We have argued that the Standard Model picture of CP violation is rather unique and highly predictive. We have also stated that reasonable extensions of the Standard Model have a very different picture of CP violation. Experimental results are too few to decide between the various possibilities. But in the near future, we expect many new measurements of CP violating observables. Our discussion of CP violation in the presence of new physics is aimed to demonstrate that, indeed, models of new physics can significantly modify the

Standard Model predictions and that the near future measurements will therefore have a strong impact on the theoretical understanding of CP violation.

To understand how the Standard Model predictions could be modified by New Physics, we focus on CP violation in the interference between decays with and without mixing. As explained above, this type of CP violation may give, due to its theoretical cleanliness, unambiguous evidence for New Physics most easily.

Let us consider five specific CP violating observables.

(i) $\mathcal{I}m\lambda_{\psi K_S}$, the CP asymmetry in $B \to \psi K_S$. This measurement will cleanly determine the relative phase between the $B-\bar{B}$ mixing amplitude and the $b \to c\bar{c}s$ decay amplitude ($\sin 2\beta$ in the Standard Model). The $b \to c\bar{c}s$ decay has Standard Model tree contributions and therefore is very unlikely to be significantly affected by new physics. On the other hand, the mixing amplitude can be easily modified by new physics. We parametrize such a modification by a phase $\theta_d$:

$$2\theta_d = \arg(M_{12}/M_{12}^{\rm SM}) \implies \mathcal{I}m\lambda_{\psi K_S} = \sin[2(\beta+\theta_d)]. \tag{7.1}$$

(ii) $\mathcal{I}m\lambda_{\phi K_S}$, the CP asymmetry in $B \to \phi K_S$. This measurement will cleanly determine the relative phase between the $B-\bar{B}$ mixing amplitude and the $b \to s\bar{s}s$ decay amplitude. The $b \to s\bar{s}s$ decay has only Standard Model penguin contributions and therefore is sensitive to new physics. We parametrize the modification of the decay amplitude by a phase $\theta_A$ [91]:

$$\theta_A = \arg(\bar{A}_{\phi K_S}/\bar{A}_{\phi K_S}^{\rm SM}) \implies \mathcal{I}m\lambda_{\phi K_S} = \sin[2(\beta+\theta_d+\theta_A)]. \tag{7.2}$$

(iii) $a_{\pi\nu\bar{\nu}}$, the CP violating ratio of $K \to \pi\nu\bar{\nu}$ decays, defined in (4.38). This measurement will cleanly determine the relative phase between the $K-\bar{K}$ mixing amplitude and the $s \to d\nu\bar{\nu}$ decay amplitude. The experimentally measured small value of $\varepsilon_K$ requires that the phase of the $K-\bar{K}$ mixing amplitude is not modified from the Standard Model prediction. (More precisely, it requires that the phase in the mixing amplitude is very close to the phase in the $s \to d\bar{u}u$ decay amplitude.) On the other hand, the decay, which in the Standard Model is a loop process with small mixing angles, can be easily modified by new physics.

(iv) $a_{D\to K\pi}$, the CP violating quantity in $D \to K^\pm\pi^\mp$ decays (see (5.5) and (5.11)):

$$a_{D\to K\pi} = \frac{\mathcal{I}m(\lambda_{K^-\pi^+}) - \mathcal{I}m(\lambda^{-1}_{K^+\pi^-})}{|\lambda_{K^-\pi^+}|} \quad\Longrightarrow\quad a_{D\to K\pi} = 2\cos\delta_{K\pi}\sin\phi_{K\pi}. \tag{7.3}$$

It depends on the relative phase between the $D-\bar{D}$ mixing amplitude and the $c \to d\bar{s}u$ and $c \to s\bar{d}u$ decay amplitudes. The two decay channels are tree level and therefore unlikely to be affected by new physics [68]. On the other hand, the mixing amplitude can be easily modified by new physics [69].

(v) $d_N$, the electric dipole moment of the neutron. We did not discuss this quantity so far because, unlike CP violation in meson decays, flavor changing couplings are not necessary for $d_N$. In other words, the CP violation that induces $d_N$ is *flavor diagonal*. It does in general get contributions from flavor changing physics, but it could be induced by sectors that are flavor blind. Within the Standard Model (and ignoring the strong CP angle $\theta_{\rm QCD}$), the contribution from $\delta_{\rm KM}$ arises at the three loop level and is at least six orders of magnitude below the experimental bound. We denote the present 90% C.L upper bound on $d_N$ by $d_N^{\rm exp}$. It is given by [92]

$$d_N^{\rm exp} = 6.3 \times 10^{-26}\ e\ {\rm cm}. \tag{7.4}$$

The main features of the observables that we chose are summarized in Table IV.

| Process | Observable | Mixing | Decay | SM | NP |
|---|---|---|---|---|---|
| $B \to \psi K_S$ | $\mathcal{I}m\lambda_{\psi K_S}$ | $B-\bar{B}$ | $b \to c\bar{c}s$ | $\sin 2\beta$ | $\sin 2(\beta+\theta_d)$ |
| $B \to \phi K_S$ | $\mathcal{I}m\lambda_{\phi K_S}$ | $B-\bar{B}$ | $b \to s\bar{s}s$ | $\sin 2\beta$ | $\sin 2(\beta+\theta_d+\theta_A)$ |
| $K \to \pi\nu\bar{\nu}$ | $a_{\pi\nu\bar{\nu}}$ | $K-\bar{K}$ | $s \to d\nu\bar{\nu}$ | $\sim \sin^2\beta$ | $\sin^2\theta_K$ |
| $D \to K\pi$ | $a_{D\to K\pi}$ | $D-\bar{D}$ | $c \to d\bar{s}u$ | 0 | $\sim \sin\phi_{K\pi}$ |
| $d_N$ | | | | $\lesssim 10^{-6} d_N^{\rm exp}$ | FD phases |

Table IV. Features of various CP violating observables.

The various CP violating observables discussed above are sensitive then to new physics in the mixing amplitudes for the $B-\bar{B}$ and $D-\bar{D}$ systems, in the decay amplitudes for $b \to s\bar{s}s$ and $s \to d\nu\bar{\nu}$ channels and to flavor diagonal CP violation. If information about all these processes becomes available and deviations from the Standard Model predictions are found, we can ask rather detailed questions about the nature of the new physics that is responsible to these deviations:

(i) Is the new physics related to the down sector? the up sector? both?

(ii) Is the new physics related to $\Delta B = 1$ processes? $\Delta B = 2$? both?

(iii) Is the new physics related to the third generation? to all generations?

(iv) Are the new sources of CP violation flavor changing? flavor diagonal? both?

It is no wonder then that with such rich information, flavor and CP violation provide an excellent probe of new physics.

## 8. Supersymmetry

A generic supersymmetric extension of the Standard Model contains a host of new flavor and CP violating parameters. (For recent reviews on supersymmetry see refs. [93-97]. The following chapter is based on [98].) It is an amusing exercise to count the number of parameters. The supersymmetric part of the Lagrangian depends, in addition to the three gauge couplings of $G_{\rm SM}$, on the parameters of the superpotential $W$, which can be written as a function of the scalar matter fields:

$$W = \sum_{i,j} \left( Y^u_{ij} h_u \tilde{q}_{Li} \tilde{u}_{Rj} + Y^d_{ij} h_d \tilde{q}_{Li} \tilde{d}_{Rj} + Y^\ell_{ij} h_d \tilde{L}_{Li} \tilde{\ell}_{Rj} \right) + \mu h_u h_d. \tag{8.1}$$

In addition, we have to add soft supersymmetry breaking terms:

$$\begin{aligned} \mathcal{L}_{\rm soft} \;=\; & -\left( a^u_{ij} h_u \tilde{q}_{Li} \tilde{u}_{Rj} + a^d_{ij} h_d \tilde{q}_{Li} \tilde{d}_{Rj} + a^\ell_{ij} h_d \tilde{L}_{Li} \tilde{\ell}_{Rj} + b h_u h_d + \text{h.c.} \right) \\ & - \sum_{\text{all scalars}} m^{S2}_{ij} A_i \bar{A}_j - \frac{1}{2} \sum_{(a)=1}^{3} \left( \tilde{m}_{(a)} (\lambda\lambda)_{(a)} + \text{h.c.} \right). \end{aligned} \tag{8.2}$$

The three Yukawa matrices $Y^f$ depend on 27 real and 27 imaginary parameters. Similarly, the three $a^f$-matrices depend on 27 real and 27 imaginary parameters. The five $m^{S2}$ hermitian $3 \times 3$ mass-squared matrices for sfermions ($S = \tilde{Q}, \tilde{d}_R, \tilde{u}_R, \tilde{L}, \tilde{\ell}_R$) have 30 real parameters and 15 phases. The gauge and Higgs sectors depend on

$$\theta_{\rm QCD}, \tilde{m}_{(1)}, \tilde{m}_{(2)}, \tilde{m}_{(3)}, g_1, g_2, g_3, \mu, b, m^2_{h_u}, m^2_{h_d}, \tag{8.3}$$

that is 11 real and 5 imaginary parameters. Summing over all sectors, we get 95 real and 74 imaginary parameters. If we switch off all the above parameters but the gauge couplings, we gain global symmetries:

$$G^{\rm SUSY}_{\rm global}(Y^f,\mu,a^f,b,m^2,\tilde m=0)=U(3)^5\times U(1)_{\rm PQ}\times U(1)_R, \tag{8.4}$$

where the $U(1)_{\rm PQ}\times U(1)_R$ charge assignments are:

$$\begin{array}{cccccc} & h_u & h_d & Q\bar u & Q\bar d & L\bar\ell \\ U(1)_{\rm PQ} & 1 & 1 & -1 & -1 & -1\,. \\ U(1)_{\rm R} & 1 & 1 & 1 & 1 & 1 \end{array} \tag{8.5}$$

Consequently, we can remove at most 15 real and 32 imaginary parameters. But even when all the couplings are switched on, there is a global symmtery, that is

$$G^{\rm SUSY}_{\rm global}=U(1)_B\times U(1)_L, \tag{8.6}$$

so that 2 of the 32 imaginary parameters cannot be removed. We are left then with

$$124 \;=\; \begin{cases} 80 & \text{real} \\ 44 & \text{imaginary} \end{cases} \text{ physical parameters.} \tag{8.7}$$

In particular, there are 43 new CP violating phases! In addition to the single Kobayashi-Maskawa of the SM, we can put 3 phases in $M_1, M_2, \mu$ (we used the $U(1)_{\rm PQ}$ and $U(1)_{\rm R}$ to remove the phases from $\mu B^*$ and $M_3$, respectively) and the other 40 phases appear in the mixing matrices of the fermion-sfermion-gaugino couplings. (Of the 80 real parameters, there are 11 absolute values of the parameters in (8.3), 9 fermion masses, 21 sfermion masses, 3 CKM angles and 36 SCKM angles.) Supersymmetry provides a nice example to our statement that reasonable extensions of the Standard Model may have more than one source of CP violation.

The requirement of consistency with experimental data provides strong constraints on many of these parameters. For this reason, the physics of flavor and CP violation has had a profound impact on supersymmetric model building. A discussion of CP violation in this context can hardly avoid addressing the flavor problem itself. Indeed, many of the supersymmetric models that we analyze below were originally aimed at solving flavor problems.

As concerns CP violation, one can distinguish two classes of experimental constraints. First, bounds on nuclear and atomic electric dipole moments determine what is usually called the *supersymmetric CP problem.* Second, the physics of neutral mesons and, most importantly, the small experimental value of $\varepsilon_K$ pose the *supersymmetric* $\varepsilon_K$ *problem.* In the next two subsections we describe the two problems. Then we describe various supersymmetric flavor models and the ways in which they address the supersymmetric CP problem.

Before turning to a detailed discussion, we define two scales that play an important role in supersymmetry: $\Lambda_S$, where the soft supersymmetry breaking terms are generated, and $\Lambda_F$, where flavor dynamics takes place. When $\Lambda_F \gg \Lambda_S$, it is possible that there are no genuinely new sources of flavor and CP violation. This leads to models with exact universality, which we discuss in section 8.3. When $\Lambda_F \lesssim \Lambda_S$, we do not expect, in general, that flavor and CP violation are limited to the Yukawa matrices. One way to suppress CP violation would be to assume that, similarly to the Standard Model, CP violating phases are large, but their effects are screened, possibly by the same physics that explains the various flavor puzzles. Such models, with Abelian or non-Abelian horizontal symmetries, are described in section 8.4. It is also possible that CP violating effects are suppressed because squarks are heavy. This scenario is also discussed in section 8.4. Another option is to assume that CP is an approximate symmetry of the full theory (namely, CP violating phases are all small). We discuss this scenario in section 8.5. A brief discussion of the implications of $\varepsilon'/\varepsilon$ is included in this subsection. Some concluding comments regarding CP violation as a probe of supersymmetric flavor models are given in section 8.6.

### *8.1. The Supersymmetric CP Problem*

One aspect of supersymmetric CP violation involves effects that are flavor preserving. Then, for simplicity, we describe this aspect in a supersymmetric model without additional flavor mixings, *i.e.* the minimal supersymmetric standard model (MSSM) with universal sfermion masses and with the trilinear SUSY-breaking scalar couplings proportional to the corresponding Yukawa couplings. (The generalization to the case of non-universal soft terms is straightforward.) In such a constrained framework, there are four new phases beyond the two phases of the Standard Model ($\delta_{\rm KM}$ and $\theta_{\rm QCD}$). One arises in the bilinear $\mu$-term of the superpotential (8.1), while the other three arise in the soft supersymmetry breaking parameters of (8.2): $\tilde{m}$ (the gaugino mass), $a$ (the trilinear scalar coupling) and $b$ (the bilinear scalar coupling). Only two combinations of the four phases are physical [99,100]. To see this, note that one could treat the various dimensionful parameters in (8.1) and (8.2) as spurions which break the $U(1)_{\rm PQ}\times U(1)_{\rm R}$ symmetry, thus deriving selection rules:

$$\begin{array}{ccccc} & \tilde{m} & A & b & \mu \\ U(1)_{\rm PQ} & 0 & 0 & -2 & -2 \\ U(1)_{\rm R} & -2 & -2 & -2 & 0 \end{array} \tag{8.8}$$

(where we defined $A$ through $a^f = AY^f$). Physical observables can only depend on combinations of the dimensionful parameters that are neutral under both $U(1)$'s. There are three such independent combinations: $\tilde{m}\mu b^*$, $A\mu b^*$ and $A^*\tilde{m}$. However, only two of their phases are independent, say

$$\phi_A = \arg(A^*\tilde{m}), \quad \phi_B = \arg(\tilde{m}\mu b^*). \tag{8.9}$$

In the more general case of non-universal soft terms there is one independent phase $\phi_{A_i}$ for each quark and lepton flavor. Moreover, complex off-diagonal entries in the sfermion mass-squared matrices may represent additional sources of CP violation.

The most significant effect of $\phi_A$ and $\phi_B$ is their contribution to electric dipole moments (EDMs). The electric dipole moment of a fermion $\psi$ is defined as the coefficient $d_\psi$ of the operator

$$\mathcal{L}_{d_\psi} = -\frac{i}{2} d_\psi \bar{\psi}\sigma_{\mu\nu}\gamma_5\psi F^{\mu\nu}. \tag{8.10}$$

For example, the contribution from one-loop gluino diagrams to the down quark EDM is given by [101,102]:

$$d_d = M_d \frac{e\alpha_3}{18\pi\tilde{m}^3}\left(|A|\sin\phi_A + \tan\beta|\mu|\sin\phi_B\right), \tag{8.11}$$

where we have taken $m_Q^2 \sim m_D^2 \sim m_{\tilde{g}}^2 \sim \tilde{m}^2$, for left- and right-handed squark and gluino masses. We define, as usual, $\tan\beta = \langle H_u\rangle/\langle H_d\rangle$. Similar one-loop diagrams give rise to chromoelectric dipole moments. The electric and chromoelectric dipole moments of the light quarks $(u, d, s)$ are the main source of $d_N$ (the EDM of the neutron), giving [103]

$$d_N \sim 2 \left(\frac{100\,GeV}{\tilde{m}}\right)^2 \sin\phi_{A,B} \times 10^{-23}\ e\,\mathrm{cm}, \tag{8.12}$$

where, as above, $\tilde{m}$ represents the overall SUSY scale. In a generic supersymmetric framework, we expect $\tilde{m} = \mathcal{O}(m_Z)$ and $\sin\phi_{A,B} = \mathcal{O}(1)$. Then the constraint (7.4) is generically violated by about two orders of magnitude. This is *the Supersymmetric CP Problem* [101-105].

Eq. (8.12) shows what are the possible ways to solve the supersymmetric CP problem:

(i) Heavy squarks: $\tilde{m} \gtrsim 1\ TeV$;

(ii) Approximate CP: $\sin\phi_{A,B} \ll 1$.

Recently, a third way has been investigated, that is cancellations between various contributions to the electric dipole moments [106-112]. However, there seems to be no symmetry that can guarantee such a cancellation. This is in contrast to the other two mechanisms mentioned above that were shown to arise naturally in specific models. We therefore do not discuss any further this third mechanism.

Finally, we mention that the electric dipole moment of the electron is also a sensitive probe of flavor diagonal CP phases. The present experimental bound, [113-114],

$$|d_e| \leq 4 \times 10^{-27}\ e\ \mathrm{cm}, \tag{8.13}$$

is also violated by about two orders of magnitue for 'natural' values of supersymmetric parameters.

*8.2. The Supersymmetric $\varepsilon_K$ Problem*

The contribution to the CP violating $\varepsilon_K$ parameter in the neutral $K$ system is dominated by diagrams involving $Q$ and $\bar{d}$ squarks in the same loop [115-119]. The corresponding effective four-fermi operator involves fermions of both chiralities, so that its matrix elements are enhanced by $\mathcal{O}(m_K/m_s)^2$ compared to the chirality conserving operators. For $m_{\tilde{g}} \simeq m_Q \simeq m_D = \tilde{m}$ (our results depend only weakly on this assumption) and focusing on the contribution from the first two squark families, one gets [119]:

$$\varepsilon_K = \frac{5\ \alpha_3^2}{162\sqrt{2}} \frac{f_K^2 m_K}{\tilde{m}^2 \Delta m_K} \left[\left(\frac{m_K}{m_s + m_d}\right)^2 + \frac{3}{25}\right] \mathcal{I}m\left\{\frac{(\delta m_Q^2)_{12}}{m_Q^2}\frac{(\delta m_D^2)_{12}}{m_D^2}\right\}, \tag{8.14}$$

where $(\delta m^2_{Q,D})_{12}$ are the off diagonal entries in the squark mass matrices in a basis where the down quark mass matrix and the gluino couplings are diagonal. These flavor violating quantities are often written as $(\delta m^2_{Q,D})_{12} = V^{Q,D}_{11} \delta m^2_{Q,D} V^{Q,D*}_{21}$, where $\delta m^2_{Q,D}$ is the mass splitting among the squarks and $V^{Q,D}$ are the gluino coupling mixing matrices in the mass eigenbasis of quarks and squarks. Note that CP would be violated even if there were two families only [120]. Using the experimental value of $\varepsilon_K$, we get

$$\frac{(\Delta m_K \varepsilon_K)^{\rm SUSY}}{(\Delta m_K \varepsilon_K)^{\rm EXP}} \sim 10^7 \left(\frac{300\ GeV}{\tilde{m}}\right)^2 \left(\frac{m^2_{Q_2} - m^2_{Q_1}}{m^2_Q}\right) \left(\frac{m^2_{D_2} - m^2_{D_1}}{m^2_D}\right) |K^{dL}_{12} K^{dR}_{12}| \sin\phi, \tag{8.15}$$

where $\phi$ is the CP violating phase. In a generic supersymmetric framework, we expect $\tilde{m} = \mathcal{O}(m_Z)$, $\delta m^2_{Q,D}/m^2_{Q,D} = \mathcal{O}(1)$, $K^{Q,D}_{ij} = \mathcal{O}(1)$ and $\sin\phi = \mathcal{O}(1)$. Then the constraint (8.15) is generically violated by about seven orders of magnitude.

Eq. (8.15) also shows what are the possible ways to solve the supersymmetric $\varepsilon_K$ problem:

(i) Heavy squarks: $\tilde{m} \gg 300\ GeV$;

(ii) Universality: $\delta m^2_{Q,D} \ll m^2_{Q,D}$;

(iii) Alignment: $|K^d_{12}| \ll 1$;

(iv) Approximate CP: $\sin\phi \ll 1$.

### 8.3. *Exact Universality*

Both supersymmetric CP problems are solved if, at the scale $\Lambda_S$, the soft supersymmetry breaking terms are universal and the genuine SUSY CP phases $\phi_{A,B}$ vanish. Then the Yukawa matrices represent the only source of flavor and CP violation which is relevant in low energy physics. This situation can naturally arise when supersymmetry breaking is mediated by gauge interactions at a scale $\Lambda_S \ll \Lambda_F$ [121]. In the simplest scenarios, the $A$-terms and the gaugino masses are generated by the same SUSY and $U(1)_R$ breaking source (see eq. (8.8)). Thus, up to very small effects due to the *standard* Yukawa matrices, $\arg(A) = \arg(m_{\tilde{g}})$ so that $\phi_A$ vanishes. In specific models also $\phi_B$ vanishes in a similar way [122,123]. It is also possible that similar boundary conditions occur when supersymmetry breaking is communicated to the observable sector up at the Planck scale.

The situation in this case seems to be less under control from the theoretical point of view. Dilaton dominance in SUSY breaking, though, seems a very interesting direction to explore [124,125].

The most important implication of this type of boundary conditions for soft terms, which we refer to as *exact universality*, is the existence of the SUSY analogue of the GIM mechanism which operates in the SM. The CP violating phase of the CKM matrix can feed into the soft terms via Renormalization Group (RG) evolution only with a strong suppression from light quark masses [99].

With regard to the supersymmetric CP problem, gluino diagrams contribute to quark EDMs as in eq. (8.11), but with a highly suppressed effective phase, *e.g.*

$$\phi_{A_d} \sim (t_S/16\pi^2)^4 Y_t^4 Y_c^2 Y_b^2 J. \qquad (8.16)$$

Here $t_S = \log(\Lambda_S/M_W)$ arises from the RG evolution from $\Lambda_S$ to the electroweak scale, the $Y_i$'s are quark Yukawa couplings (in the mass basis), and $J \simeq 2 \times 10^{-5}$ is defined in eq. (2.30). A similar contribution comes from chargino diagrams. The resulting EDM is $d_N \lesssim 10^{-31}$ $e$ cm. This maximum can be reached only for very large $\tan\beta \sim 60$ while, for small $\tan\beta \sim 1$, $d_N$ is about 5 orders of magnitude smaller. This range of values for $d_N$ is much below the present ($\sim 10^{-25}$ $e$ cm) and foreseen ($\sim 10^{-28}$ $e$ cm) experimental sensitivities [126-129].

With regard to the supersymmetric $\varepsilon_K$ problem, the contribution to $\varepsilon_K$ is proportional to $\mathcal{I}m(V_{td}V_{ts}^*)^2 Y_t^4 (t_S/16\pi^2)^2$, giving the same GIM suppression as in the SM. This contribution turns out to be negligibly small [99]. The supersymmetric contribution to $D - \bar{D}$ mixing is similarly small and we expect no observable effects. For the $B_d$ and $B_s$ systems, the largest SUSY contribution to the mixing comes from box diagrams with intermediate charged Higgs and the up quarks. It can be up to $\mathcal{O}(0.2)$ of the SM amplitude for $\Lambda_S = M_{\rm Pl}$ and $\tan\beta = \mathcal{O}(1)$ [130-133], and much smaller for large $\tan\beta$. The contribution is smaller in models of gauge mediated SUSY breaking where the mass of the charged Higgs boson is typically $\gtrsim 300$ $GeV$ [121] and $t_S \sim 5$. The SUSY contributions to $B_s - \bar{B}_s$ and $B_d - \bar{B}_d$ mixing are, to a good approximation, proportional to $(V_{tb}V_{ts}^*)^2$ and $(V_{tb}V_{td}^*)^2$, respectively, like in the SM. Then, regardless of the size of these contributions, the relation $\Delta m_{B_d}/\Delta m_{B_s} \sim |V_{td}/V_{ts}|^2$ and the CP asymmetries in neutral $B$ decays into final CP eigenstates are the same as in the SM.

### *8.4. Approximate Horizontal Symmetries*

In the class of supersymmetric models with $\Lambda_F \lesssim \Lambda_S$, the soft masses are generically not universal, and we do not expect flavor and CP violation to be limited to the Yukawa matrices. Most models where soft terms arise at the Planck scale ($\Lambda_S \sim M_{\rm Pl}$) belong to this class. It is possible that, similarly to the Standard Model, CP violating phases are large, but their effects are screened, possibly by the same physics that explains the various flavor puzzles. This usually requires Abelian or non-Abelian horizontal symmetries. Two ingredients play a major role here: selection rules that come from the symmetry and holomorphy of Yukawa and $A$-terms that comes from the supersymmetry. With Abelian symmetries, the screening mechanism is provided by *alignment* [134-137], whereby the mixing matrices for gaugino couplings have very small mixing angles, particularly for the first two down squark generations. With non-Abelian symmetries, the screening mechanism is *approximate universality*, where squarks of the two families fit into an irreducible doublet and are, therefore, approximately degenerate [138-145]. In all of these models, it is difficult to avoid $d_N \gtrsim 10^{-28}$ e cm.

As far as the third generation is concerned, the signatures of Abelian and non-Abelian models are similar. In particular, they allow observable deviations from the SM predictions for CP asymmetries in $B$ decays. The recent measurement of $a_{\psi K_S}$ gives first constraints on these contributions [14]. In some cases, non-Abelian models give relations between CKM parameters and consequently predict strong constraints on these CP asymmetries.

For the two light generations, only alignment allows interesting effects. In particular, it predicts large CP violating effects in $D-\bar{D}$ mixing [134,135]. Thus, it allows $a_{D\to K\pi} = \mathcal{O}(1)$.

Finally, it is possible that CP violating effects are suppressed because squarks are heavy. If the masses of the first and second generations squarks $m_i$ are larger than the other soft masses, $m_i^2 \sim 100\,\tilde{m}^2$ then the Supersymmetric CP problem is solved and the $\varepsilon_K$ problem is relaxed (but not eliminated) [141,146]. This does not necessarily lead to naturalness problems, since these two generations are almost decoupled from the Higgs sector.

Notice though that, with the possible exception of $m^2_{\tilde{b}_R}$, third family squark masses cannot naturally be much above $m^2_Z$. If the relevant phases are of $O(1)$, the main contribution to $d_N$ comes from the third family via the two-loop induced three-gluon operator [147], and it is roughly at the present experimental bound when $m_{\tilde{t}_{L,R}} \sim 100\ GeV$.

Models with the first two squark generations heavy have their own signatures of CP violation in neutral meson mixing [148]. The mixing angles relevant to $D-\bar{D}$ mixing are similar, in general, to those of models of alignment (if alignment is invoked to explain $\Delta m_K$ with $m^2_{Q,D} \lesssim 20\ TeV$). However, since the $\tilde{u}$ and $\tilde{c}$ squarks are heavy, the contribution to $D-\bar{D}$ mixing is one to two orders of magnitude below the experimental bound. This may lead to the interesting situation that $D-\bar{D}$ mixing will first be observed through its CP violating part [70]. In the neutral $B$ system, $\mathcal{O}(1)$ shifts from the Standard Model predictions of CP asymmetries in the decays to final CP eigenstates are possible. This can occur even when squark masses of the third family are $\sim 1\ TeV$ [149], since now mixing angles can naturally be larger than in the case of horizontal symmetries (alignment or approximate universality).

### *8.5. Approximate CP Symmetry*

Both supersymmetric CP problems are solved if CP is an approximate symmetry, broken by a small parameter of order $10^{-3}$. This is another possible solution to CP problems in the class of supersymmetric models with $\Lambda_F \lesssim \Lambda_S$. (Of course, some mechanism has also to suppress the real part of the $\Delta S = 2$ amplitude by a sufficient amount.)

If CP is an approximate symmetry, we expect also the SM phase $\delta_{\rm KM}$ to be $\ll 1$. Then the standard box diagrams cannot account for $\varepsilon_K$ which should arise from another source. In supersymmetry with non-universal soft terms, the source could be diagrams involving virtual superpartners, mainly squark-gluino box diagrams. Let us call $(M^K_{12})^{\rm SUSY}$ the supersymmetric contribution to the $K-\bar{K}$ mixing amplitude. Then the requirements $\mathcal{R}e(M^K_{12})^{\rm SUSY} \lesssim \Delta m_K$ and $\mathcal{I}m(M^K_{12})^{\rm SUSY} \sim \varepsilon_K \Delta m_K$ imply that the generic CP phases are $\geq \mathcal{O}(\varepsilon_K) \sim 10^{-3}$.

Of course, $d_N$ constrains the relevant CP violating phases to be $\lesssim 10^{-2}$. If all phases are of the same order, then $d_N$ must be just below or barely compatible with the present experimental bound. A signal should definitely be found if the accuracy is increased by two orders of magnitude.

The main phenomenological implication of these scenarios is that CP asymmetries in $B$ meson decays are small, perhaps $\mathcal{O}(\varepsilon_K)$, rather than $\mathcal{O}(1)$ as expected in the SM. Also the ratio $a_{\pi\nu\bar{\nu}}$ of eq. (4.38) is very small, in contrast to the Standard Model where it is expected to be of $\mathcal{O}(\sin^2\beta)$. Explicit models of approximate CP were presented in refs. [150-153].

The experimntal value of $\varepsilon'/\varepsilon$ has particularly interesting implications on models of approximate CP [154]. In this framework, the standard model cannot account for $\varepsilon'/\varepsilon$. A model of approximate CP would then be excluded if it does not provide sufficiently large contributions from new physics to this parameter. A generic supersymmetric model where the $a^q$ terms in (8.2) are not proportional to the $Y^q$ terms in (8.1) can provide a large contribution [155] related to imaginary part of

$$a^d_{12} \sim m_s|V_{us}|/\tilde{m}. \tag{8.17}$$

In models of non-Abelian flavor symmetries, the contribution is typically not large enough because of cancellation between the $a^d_{12}$ and $a^d_{21}$ terms [156,157]. In models of heavy $\tilde{d}$ and $\tilde{s}$ squarks, the contribution is highly suppressed by the heavy mass scale [154]. Models of alignment can give a contribution that is not much smaller than the estimate in (8.17). If, however, the related CP violating phase is small, then the model can account for $\varepsilon'/\varepsilon$ only if both the model parameters and the hadronic parameters assume rather extreme values [154]. We conclude that most existing models of supersymmetry with approximate CP are excluded (or, at least, strongly disfavored) by the experimental measurement of $\varepsilon'/\varepsilon$. (For other recent works on $\varepsilon'/\varepsilon$ in the supersymmetric framework, see [118-119,54,158-160]

The fact that the Standard Model and the models of approximate CP are both viable at present shows that the mechanism of CP violation has not really been tested experimentally. The only measured CP violating observales, that is $\varepsilon_K$ and $\varepsilon'_K$, are small. Their smallness could be related to the 'accidental' smallness of CP violation for the first two quark generations, as is the case in the Standard Model, or to CP being an approximate symmetry, as is the case in the models discussed here. Future measurements, particularly of processes where the third generation plays a dominant role (such as $a_{\psi K_S}$ or $a_{\pi\nu\bar{\nu}}$), will easily distinguish between the two scenarios. While the Standard Model predicts large CP violating effects for these processes, approximate CP would suppress them too.

*8.6. Some Concluding Remarks*

We can get an intuitive understanding of how the various supersymmetric flavor models discussed in this chapter solve the supersymmetric flavor and CP problems by presenting the general form of the squark mass-squared matrices for each framework. This is summarized in Table V. The implications of each flavor model for the various CP violating observables presented in the previous chapter are given in Table VI.

| Flavor Model | Theory | $\frac{1}{\tilde{m}^2}\tilde{M}^2_{LL} \sim$ | Physical Parameters |
|---|---|---|---|
| Exact Universality | GMSB | $\mathrm{diag}(a,a,a)$ | $\Delta\tilde{m}^2_{12} \sim m_c^2/m_W^2$ |
| Approx. Universality | Non-Abelian $H$ | $\mathrm{diag}(a,a,b)$ | $\Delta\tilde{m}^2_{12} \sim \sin^2\theta_C$ |
| Alignment | Abelian $H$ | $\mathrm{diag}(a,b,c)$ | $(K^d_L)_{12} \ll \sin\theta_C$ |
| Heavy Squarks | Comp.; Anom. $U(1)$ | $\mathrm{diag}(A,B,c)$ | $\tilde{m}^2_{1,2} \sim 100\tilde{m}^2$ |
| Approximate CP | SCPV | | $10^{-3} \lesssim \phi_{\rm CP} \ll 1$ |

Table V. Supersymmetric flavor models.

| Model | $\frac{d_N}{10^{-25}\ e\ \mathrm{cm}}$ | $\theta_d$ | $\theta_A$ | $a_{D\to K\pi}$ | $a_{K\to\pi\nu\bar{\nu}}$ |
|---|---|---|---|---|---|
| Standard Model | $\lesssim 10^{-6}$ | 0 | 0 | 0 | $\mathcal{O}(1)$ |
| Exact Universality | $\lesssim 10^{-6}$ | 0 | 0 | 0 | =SM |
| Approx. Universality | $\gtrsim 10^{-2}$ | $\mathcal{O}(0.2)$ | $\mathcal{O}(1)$ | 0 | ≈SM |
| Alignment | $\gtrsim 10^{-3}$ | $\mathcal{O}(0.2)$ | $\mathcal{O}(1)$ | $\mathcal{O}(1)$ | ≈SM |
| Heavy Squarks | $\sim 10^{-1}$ | $\mathcal{O}(1)$ | $\mathcal{O}(1)$ | $\mathcal{O}(10^{-2})$ | ≈SM |
| Approximate CP | $\sim 10^{-1}$ | $-\beta$ | 0 | $\mathcal{O}(10^{-3})$ | $\mathcal{O}(10^{-5})$ |

Table VI. Phenomenological implicatons of supersymmetric flavor models.

We would like to emphasize the following points:

(i) For supersymmetry to be established, a direct observation of supersymmetric particles is necessary. Once it is discovered, then measurements of CP violating observables will be a very sensitive probe of its flavor structure and, consequently, of the mechanism of dynamical supersymmetry breaking.

(ii) It is easy to distinguish between models of exact universality and models with genuine supersymmetric flavor and CP violation. This can be done through searches of $d_N$ and of CP asymmetries in $B$ decays.

(iii) The neutral $D$ system provides a stringent test of alignment.

(iv) The fact that $K \to \pi\nu\bar{\nu}$ decays are not affected by most supersymmetric flavor models [161,162] is actually an advantage. The Standard Model correlation between $a_{\pi\nu\bar{\nu}}$ and $a_{\psi K_S}$ is a much cleaner test than a comparison of $a_{\psi K_S}$ to the CKM constraints.

(v) Approximate CP has dramatic effects on all observables. My guess is that in lectures given a year from now, it will not appear in the Table as a viable option.

## 9. Left Right Symmetric Models of Spontaneous CP Violation

### *9.1. The Model*

We consider models with a symmetry $G_{\rm LRS} \times D_{\rm LRS}$, where $G_{\rm LRS}$ is the gauge group,

$$G_{\rm LRS} = SU(3)_C \times SU(2)_L \times SU(2)_R \times U(1)_{B-L}, \tag{9.1}$$

and $D_{\rm LRS}$ is a discrete group,

$$D_{\rm LRS} = P \times C. \tag{9.2}$$

Various versions of left-right symmetric models differ in $D_{\rm LRS}$. We are interested here in models where CP is only spontaneously broken, hence our choice of (9.2) [163-168].

The fermion representations consist of three generations of

$$Q_{Li}(3,2,1)_{1/3}, \quad Q_{Ri}(3,1,2)_{1/3}, \quad L_{Li}(1,2,1)_{-1}, \quad L_{Ri}(1,1,2)_{-1}. \tag{9.3}$$

Under $D_{\rm LRS}$, the fermion fields transform as follows:

$$\begin{array}{lcc} P: & Q_L \leftrightarrow Q_R & L_L \leftrightarrow L_R \\ C: & Q_L \leftrightarrow i\sigma_2 (Q_R)^* & L_L \leftrightarrow i\sigma_2 (L_R)^* \end{array} \tag{9.4}$$

The scalar representations consist of three multiplets [169],

$$\Delta_R(1,1,3)_2, \quad \Delta_L(1,3,1)_2, \quad \Phi(1,2,2)_0. \tag{9.5}$$

Under $D_{\rm LRS}$, the scalar fields transform as follows:

$$\begin{array}{lll} P: & \Delta_L \leftrightarrow \Delta_R & \Phi \leftrightarrow \Phi^\dagger \\ C: & \Delta_L \leftrightarrow (\Delta_R)^* & \Phi \leftrightarrow \Phi^T \end{array} \tag{9.6}$$

It is often convenient to write $\Phi$ in a $2\times 2$ matrix form:

$$\Phi = \begin{pmatrix} \phi_1^0 & \phi_1^+ \\ \phi_2^- & \phi_2^0 \end{pmatrix}. \tag{9.7}$$

The spontaneous symmetry breaking occurs in two stages,

$$G_{\rm LRS} \times D_{\rm LRS} \to G_{\rm SM} \to SU(3)_C \times U(1)_{\rm EM}, \tag{9.8}$$

due to the VEVs of the neutral members of the scalar fields:

$$\langle \Delta_R \rangle = \begin{pmatrix} 0 \\ 0 \\ v_R e^{i\beta} \end{pmatrix}, \quad \langle \Delta_L \rangle = \begin{pmatrix} 0 \\ 0 \\ v_L \end{pmatrix}, \quad \langle \Phi \rangle = \begin{pmatrix} k_1 & 0 \\ 0 & k_2 e^{i\alpha} \end{pmatrix}. \tag{9.9}$$

(In general, all four VEVs are complex. There is, however, freedom of rotations by $U(1)_{B-L}$ for $\Delta_L$ and $\Delta_R$ and by $U(1)_{T_{3L}} \times U(1)_{T_{3R}}$ for $\phi_1$ and $\phi_2$, so that only two phases are physical.) These VEVs are assumed to satisfy the hierarchy

$$v_R \gg k_1, k_2 \gg v_L. \tag{9.10}$$

The first stage of symmetry breaking in (9.8) takes place at the scale $v_R$ and the second at $k = \sqrt{k_1^2 + k_2^2}$.

### *9.2. Flavor Parameters*

The quark Yukawa couplings have the following form:

$$\mathcal{L}_{\rm Yuk} = f\overline{Q_L}\Phi Q_R + h\overline{Q_L}\widetilde{\Phi}Q_R + {\rm h.c.}, \tag{9.11}$$

where $\widetilde{\Phi} = \tau_2 \Phi^* \tau_2$. As a consequence of $D_{\rm LRS}$, the Yukawa matrices $f$ and $h$ are *symmetric and real*: $P$ requires that they are hermitian, $C$ requires that they are symmetric, and CP requires that they are real. The resulting mass matrices,

$$\begin{aligned} M_u &= f k_1 + h k_2 e^{-i\alpha}, \\ M_d &= h k_1 + f k_2 e^{i\alpha}, \end{aligned} \tag{9.12}$$

are complex symmetric matrices.

How many independent physical flavor parameters (and, in particular, phases) does this model have? We have two symmetric and complex mass matrices, that is twelve real and twelve imaginary Yukawa parameters. If we set $h = f = 0$, we gain a global $U(3)$ symmetry ($D_{\rm LRS}$ does not allow independent $U(3)_L$ and $U(3)_R$ rotations). This means that we can remove three real and six imaginary parameters. When $h$ and $f$ are different from zero, there is no global symmetry ($U(1)_{B-L}$ is part of the gauge symmetry). We conclude that there are nine real and six imaginary flavor parameters. Six of the real parameters are the six quark masses. To identify the other flavor parameters, note that the symmetric mass matrices can be diagonalized by a unitary transformation of the form

$$V_u M_u V_u^T = M_u^{\rm diag}, \quad V_d M_d V_d^T = M_d^{\rm diag}. \tag{9.13}$$

Consequently, the mixing matrices $V_L$ and $V_R$ describing, respectively, the $W_L$ and $W_R$ interactions,

$$\mathcal{L}_{CC} = \frac{g}{\sqrt{2}} \left( W^+_{L\mu} \overline{u_L} V_L \gamma^\mu d_L + W^+_{R\mu} \overline{u_R} V_R \gamma^\mu d_R \right) + {\rm h.c.}, \tag{9.14}$$

are related:

$$V_L = P^u V_R^* P^d, \tag{9.15}$$

where $P^u$ and $P^d$ are diagonal phase matrices. The three real parameters are then the three mixing angles, which are equal in $V_L$ and $V_R$. The six phases can be arranged in various ways. A convenient choice is to have a single phase in $V_L$, which is then just $\delta_{\rm KM}$ of $V_{\rm CKM}$ of the Standard Model, and five phases in $V_R$. (It is also possible to have $V_L = V_R^*$ with six phases in each.)

*9.3. What is the Low Energy Effective Theory of the LRS Model?*

It is interesting to ask what is the low enrgy effective theory below the scale $v_R$. It is straightforward to show that the light fields are precisely those of the Standard Model: the fermions are chiral under $SU(2)_L$ except for the right-handed neutrinos in $L_{Ri}$ which acquire Majorana masses at the scale $v_R$ due to their coupling to $\Delta_R$. (The left-handed neutrinos acquire very light masses from both the see-saw mechanism and their direct coupling to $\Delta_L$.) Only the one Higgs doublet related to $G_{\rm SM}$ breaking, that is $k_1\phi_1 + k_2 e^{-i\alpha}\phi_2$, remains light. The question is then whether the left-right symmetry constrains Standard Model parameters.

To answer this question, we first argue that phenomenological constraints forbid $r \equiv k_2/k_1 = \mathcal{O}(1)$. (More precisely, it is $r\sin\alpha$ which is constrained to be very small.) Consider eqs. (9.12). They lead to the following equations:

$$\begin{aligned} M_u r e^{i\alpha} - M_d \ &= k_1 h(r^2-1), \\ M_u - M_d r e^{-i\alpha} \ &= k_1 f(1-r^2). \end{aligned} \tag{9.16}$$

The right hand side of these equations is real. Then, the imaginary part of the left-hand side should vanish. Let us put all quark masses to zero, except for $m_t$ and $m_b$. We take then $(M_u)_{33} = m_t e^{i\theta_t}$ and $(M_d)_{33} = m_b e^{i\theta_b}$. We get:

$$\begin{aligned} r m_t \sin(\theta_t + \alpha) - m_b \sin\theta_b \ &= 0, \\ m_t \sin\theta_t - m_b \sin(\theta_b - \alpha) \ &= 0. \end{aligned} \tag{9.17}$$

The second equation implies that $\theta_t \lesssim m_b/m_t$. Then, the first equation gives [170]

$$r\sin\alpha \le m_b/m_t. \tag{9.18}$$

(Note that the model is symmetric under $r \to 1/r$ and $\alpha \to -\alpha$. Therefore, $r\sin\alpha \ge m_t/m_b$ is acceptable and physically equivalent.)

The only source of CP violation in the quark mass matrices is the phase $\alpha$. (The phase $\beta$ in $\langle\Delta_R\rangle$ does not affect quark masses, though it may affect neutrino masses.) Moreover, if one of the $\langle\phi_i^0\rangle$ vanished, then again there would be no CP violation in the quark mass matrices. As a consequence of these two facts, all CP violating phases in the mixing matrices $V_L$ and $V_R$ are proportional to $r\sin\alpha$. Hence the importance of (9.18). In particular, for the Kobayashi-Maskawa phase, one finds [170]

$$\delta_{\rm KM} \sim r\sin\alpha(m_c/m_s) \leq \mathcal{O}(0.1). \tag{9.19}$$

We learn that the low energy effective theory of the left-right symmetric model is the Standard Model with a small value for $\delta_{\rm KM}$.

Phenomenologically, it is difficult, though not impossible, to account for $\varepsilon_K$ with just the Standard Model contribution and a small KM phase. There are then two possibilities:

(i) The left-right symmetry is broken at a very high scale. The low energy theory is to a good approximation just the Standard Model. CP is, however, an approximate symmetry in the kaon sector. The hadronic parameters playing a role in the calculation of $\varepsilon_K$ have to assume rather extreme values.

(ii) The left-right symmetry is broken at low enough scale so that there are significant new contributions to various rare processes. In particular, box diagrams with intermediate $W_R$-boson and tree diagrams with a heavy neutral Higgs dominate $\varepsilon_K$. This sets up an upper bound on the scale $v_R$, of order tens of $TeV$.

### *9.4. Phenomenology of CP Violation*

The smallness of $r\sin\alpha$ does not necessarily mean that CP is an approximate symmetry in the quark sector; the phases in the mixing matrices depend, in addition to $r\sin\alpha$, on quark mass ratios, some of which are large. An explicit calculation shows that the six phases actually divide to two groups: the KM phase and the three phases that appear in $V_R$ in a two generation model (usually denoted by $\delta_1$, $\delta_2$ and $\gamma$) are all small [170], while the two extra phases that appear in the three generation $V_R$ (denoted by $\sigma_1$, $\sigma_2$) are not [171]:

$$\begin{aligned}\delta,\delta_1,\delta_2,\gamma \ &\propto r\sin\alpha(m_c/m_s) \leq \mathcal{O}(0.1),\\ \sigma_1,\sigma_2 \ &\propto r\sin\alpha(m_t/m_b) \leq \mathcal{O}(1).\end{aligned} \tag{9.20}$$

In $\varepsilon_K$, it is mainly $\delta_1$ and $\delta_2$ which play a role. (We here assume that the hadronic parameters are close to their present theoretical estimate and therefore $\varepsilon_K$ cannot be explained in this framework by the Standard Model contribution alone.) Assuming that the $W_L - W_R$ box diagram gives the dominant contribution, one is led to conclude that [170]

$$M(W_R) \lesssim 20\ TeV \tag{9.21}$$

is favored. Note that CP conserving processes provide a lower bound [172,173],

$$M(W_R) \gtrsim 1.6\ TeV. \tag{9.22}$$

The favored range for $M(W_R)$ is then very constrained in this framework.

Taking into account this upper bound and the fact that the $\sigma_i$ phases are enhanced by a factor of about 10 compared to the $\delta_i$ phases, one finds that the left-right symmetric contributions compete with or even dominate over the Standard Model contributions to $B-\bar{B}$ mixing and to $B_s-\bar{B}_s$ mixing [171,174-176]. This means that CP asymmetries in $B$ or $B_s$ decays into final CP eigenstates could be substantially different from the Standard Model prediction. Moreover, the phases in the left-right symmetric contributions to $B-\bar{B}$ and $B_s-\bar{B}_s$ mixing are closely related, predicting correlations between the deviations. The CP asymmetry in semileptonic $B$ decays could also be significantly enhanced [177]. The recent measurement of $a_{\psi K_S}$ gives first constraints on $\sigma_1$ leading to new bounds on $a_{\rm SL}$ [14].

Finally, LRS models could enhance the electric dipole moments of the neutron and of the electron [178-181].

## 10. Multi-Scalar Models

The Standard Model has a single scalar field, $\phi(1,2)_{1/2}$, that is responsible for the spontaneous symmetry breaking, $SU(2)_L \times U(1)_Y \to U(1)_{\rm EM}$. Within the framework of the Standard Model, the complex Yukawa couplings of the scalar doublet to fermions are the only source of flavor physics and of CP violation. However, in the mass basis, the interactions of the Higgs particle are flavor diagonal and CP conserving.

There are several good reasons for the interest in multi-scalar models in the context of flavor and CP violation:

a. If there exist additional scalars and, in particular, $SU(2)_L$-doublets, then not only there are new sources of CP violation, but also the Yukawa interactions in the mass basis as well as the scalar self-interactions may violate CP.

b. CP violation in scalar interactions has very different features from the $W$-mediated CP violation of the Standard Model. For example, it could lead to observable *flavor diagonal* CP violation in top physics or in electric dipole moments, or it could induce transverse lepton polarization in semileptonic meson decays.

c. With more than a single scalar doublet, CP violation could be spontaneous.

Indeed, there is no good reason to assume that the Standard Model doublet is the only scalar in Nature. Most extensions of the Standard Model predict that there exist additional scalars. For example, models with an extended gauge symmetry (such as GUTs and left-right symmetric models) need extra scalars to break the symmetry down to $G_{\rm SM}$; Supersymmetry requires that there exists a scalar partner to each Standard Model fermion. However, scalar masses are generically not protected by a symmetry. Consequently, in models where the low energy effective theory is the Standard Model, we expect in general that the only light scalar is the Standard Model doublet.

The study of multi-scalar models is then best motivated in the following cases:

(i) The scale of new physics is not very high above the electroweak scale. One has to remember, however, that in such cases there is more to the new physics than just extending the scalar sector.

(ii) The scalar is related to the spontaneous breaking of a global symmetry. In some cases, a discrete symmetry is enough to make a scalar light.

We will discuss scalar $SU(2)_L$-doublets and -singlets only. There are two reasons for that. First, the VEVs of higher multiplets need to be very small in order to avoid large deviations from the experimentally successful relation $\rho = 1$. Second, higher multiplets do not couple to the known fermions. (The only exception is an $SU(2)_L$-triplet that can couple to the left-handed leptons.)

### *10.1. Multi Higgs Doublet Models*

The most popular extension of the Higgs sector is the multi Higgs doublet model

(MHDM) and, in particular, the two Higgs doublet model (2HDM). These models have, in addition to the Kobayashi-Maskawa phase of the quark mixing matrix, several new sources of CP violation [182]:

(i) A complex mixing matrix for charged scalars [183].

(ii) Mixing of CP-even and CP-odd neutral scalars [184].

(iii) CP-odd Yukawa couplings (in the quark mass basis).

(iv) Complex quartic scalar couplings.

The CP violation that is relevant to near future experiments always involves fermions. Therefore, we will only discuss the new sources (i), (ii) and (iii).

A generic MHDM, with all dimensionful parameters at the electroweak scale and all dimensionless parameters of order one, leads to severe phenomenological problems. In particular, some of the physical scalars have flavor changing (and CP violating) couplings at tree level, violating bounds on rare processes such as $\Delta m_K$ and $\varepsilon_K$ by several orders of magnitude. There are three possible solutions to these problems:

(I) *Natural flavor conservation (NFC)* [185]: only a single scalar doublet couples to each fermion sector. 2HDM where the same (a different) scalar doublet couples to the up and the down quarks are called type I (II). The absence of flavor changing and/or CP violating Yukawa interactions in this case is based on the same mechanism as within the Standard Model.

(II) *Approximate flavor symmetries (AFS)* [186]: it is quite likely that the smallness and hierarchy in the fermion masses and mixing angles are related to an approximate flavor symmetry, broken by a small parameter. If so, then it is unavoidable that the Yukawa couplings of all scalar doublets are affected by the selection rules related to the flavor symmetry. In such a case, couplings to the light generations and, in particular, off-diagonal couplings, are suppressed.

(III) *Heavy scalars* [187]: all dimensionful parameters that are not constrained by the requirement that the spontaneous breaking of $G_{\rm SM}$ occurs at the electroweak scale are actually higher than this scale, $\Lambda_{\rm NP} \gg \Lambda_{\rm EW}$. Then all the new sources of flavor and CP violation in the scalar sector are suppressed by $\mathcal{O}(\Lambda^2_{\rm EW}/\Lambda^2_{\rm NP})$.

In table VII we summarize the implications of the various multi-scalar models for

CP violation. Note that, if we impose NFC, spontaneous CP violation (SCPV) [184] is impossible in 2HDM [183] and (since the combination of SCPV and NFC leads to $\delta_{\rm KM}=0$ [188]) is phenomenologically excluded in MHDM [189,190]. Explicit models of spontaneous CP violation have been constructed within the frameworks of approximate NFC [191], approximate flavor symmetries [137,152] and heavy scalars [181]. In the supersymmetric framework, one has to add at least two scalar singlets to allow for spontaneous CP violation [192]. Entries marked with '$*$' mean that the number of scalar doublets should be larger than 2 (that is, the answer is 'No' in 2HDM).

| Framework | Model (Example) | SCPV | (i) | (ii) | (iii) |
|---|---|---|---|---|---|
| NFC | MSSM | Excluded | Yes* | Yes* | No |
| AFS | Horizontal Sym. | Yes | Yes* | Yes | $\mathcal{O}(m_q/m_Z)$ |
| Heavy | LRS | Yes | $\mathcal{O}(\frac{\Lambda^2_{\rm EW}}{\Lambda^2_{\rm NP}})^*$ | $\mathcal{O}(\frac{\Lambda^2_{\rm EW}}{\Lambda^2_{\rm NP}})$ | $\mathcal{O}(\frac{\Lambda^2_{\rm EW}}{\Lambda^2_{\rm NP}})$ |

Table VII. Multi Higgs Doublet Models

### *10.2. (i) Charged Scalar Exchange*

We investigate a multi Higgs doublet model (with $n\geq 3$ doublets) with NFC and assume that a different doublet couples to each of the the down, up and lepton sectors:

$$-\mathcal{L}_Y=-\frac{\phi_1^+}{v_1}\overline{U}V_{\rm CKM}M_d^{\rm diag}P_R\,D+\frac{\phi_2^+}{v_2}\overline{U}M_u^{\rm diag}V_{\rm CKM}P_L\,D-\frac{\phi_3^+}{v_3}\overline{\nu}M_\ell P_R\,\ell+{\rm h.c.},\qquad(10.1)$$

where $P_{L,R}=(1\mp\gamma_5)/2$. We denote the physical charged scalars by $H_i^+$ $(i=1,2,\ldots,n-1)$, and the would-be Goldstone boson (eaten by the $W^+$) by $H_n^+$. We define $K$ to be the matrix that rotates the charged scalars from the interaction- to the mass-eigenbasis. Then the Yukawa Lagrangian in the mass basis (for both fermions and scalars) is

$$\mathcal{L}_Y=\frac{G_F^{1/2}}{2^{1/4}}\sum_{i=1}^{n-1}\{H_i^+\overline{U}[Y_iM_u^{\rm diag}V_{\rm CKM}P_L+X_iV_{\rm CKM}M_d^{\rm diag}P_R]D+H_i^+\overline{\nu}[Z_iM_\ell P_R]\ell\}+{\rm h.c.},\qquad(10.2)$$

where

$$X_i=-\frac{K_{i1}^*}{K_{n1}^*},\quad Y_i=-\frac{K_{i2}^*}{K_{n2}^*},\quad Z_i=-\frac{K_{i3}^*}{K_{n3}^*}.\qquad(10.3)$$

CP violation in the charged scalar sector comes from phases in $K$. CP violating effects are largest when the lightest charged scalar is much lighter than the heavier ones [193,194]. Here we assume that all but the lightest charged scalar ($H_1^+$) effectively decouple from the fermions. Then, CP violating observables depend on three parameters:

$$\begin{aligned}
\frac{\mathcal{I}m(XY^*)}{m_H^2} &\equiv \frac{\mathcal{I}m(X_1Y_1^*)}{m_{H_1}^2} \approx \sum_{i=1}^{n-1} \frac{\mathcal{I}m(X_iY_i^*)}{m_{H_i}^2}, \\
\frac{\mathcal{I}m(XZ^*)}{m_H^2} &\equiv \frac{\mathcal{I}m(X_1Z_1^*)}{m_{H_1}^2} \approx \sum_{i=1}^{n-1} \frac{\mathcal{I}m(X_iZ_i^*)}{m_{H_i}^2}, \\
\frac{\mathcal{I}m(YZ^*)}{m_H^2} &\equiv \frac{\mathcal{I}m(Y_1Z_1^*)}{m_{H_1}^2} \approx \sum_{i=1}^{n-1} \frac{\mathcal{I}m(Y_iZ_i^*)}{m_{H_i}^2}.
\end{aligned} \tag{10.4}$$

$\mathcal{I}m(XY^*)$ induces CP violation in the quarks sector, while $\mathcal{I}m(XZ^*)$ and $\mathcal{I}m(YZ^*)$ give CP violation that is observable in semi-leptonic processes.

There is an interesting question of whether charged scalar exchange could be the *only* source of CP violation. In other words, we would like to know whether a model of extended scalar sector with spontaneous CP violation and NFC is viable. In these models, $\delta_{\rm KM} = 0$ and $\varepsilon_K$ has to be accounted for by charged Higgs exchange. This requires very large long distance contributions. The CP violating coupling should fulfill [195-196]

$$\mathcal{I}m(XY^*) \geq \mathcal{O}(40). \tag{10.5}$$

However, the upper bounds on $d_N$ [189] and on BR$(b \to s\gamma)$ [190] require

$$\mathcal{I}m(XY^*) \leq \mathcal{O}(1). \tag{10.6}$$

We conclude that models of SCPV and NFC are excluded. It is, of course, still a viable possibility that CP is explicitly broken, in which case both quark and Higgs mixings provide CP violation.

The bound (10.6) implies that the charged Higgs contribution to $B-\bar{B}$ mixing is numerically small and would modify the Standard Model predictions for CP asymmetries in $B$ decays by no more than $\mathcal{O}(0.02)$ [190]. On the other hand, the contribution to $d_N$ can still be close to the experimental upper bound.

The lepton transverse polarization cannot be generated by vector or axial-vector interactions only [197,198], so it is particularly suited for searching for CP violating scalar contributions. As triple-vector correlation is odd under time-reversal, the experimental observation of such correlation would signal T and – assuming CPT symmetry – CP violation. (It is possible to get non-vanishing T−odd observables even without CP violation (see e.g. [199]). Such "fake" asymmetries can arise from final state interactions (FSI). They can be removed by comparing the measurements in two CP conjugate channels.) The muon transverse polarization in $K \to \pi\, \mu\, \nu$ decays and the tau transverse polarization in $B \to X \tau \nu$ are examples of such observables. The lepton transverse polarization, $P_\perp$, in semileptonic decays is defined as the lepton polarization component along the normal vector of the decay plane,

$$P_\perp = \frac{\vec{s}_\ell \cdot (\vec{p}_\ell \times \vec{p}_X)}{|\vec{p}_\ell \times \vec{p}_X|}\,, \tag{10.7}$$

where $\vec{s}_\ell$ is the lepton spin three-vector and $\vec{p}_\ell$ ($\vec{p}_X$) is the three-momentum of the lepton (hadron). Experimentally, it is useful to define the integrated CP violating asymmetry

$$a_{CP} \equiv \langle P_\perp \rangle = \frac{\Gamma^+ - \Gamma^-}{\Gamma^+ + \Gamma^-}, \tag{10.8}$$

where $\Gamma^+$ ($\Gamma^-$) is the rate of finding the lepton spin parallel (anti-parallel) to the normal vector of the decay plane. A non-zero $a_{CP}$ can arise in our model from the interference between the $W$-mediated and the $H^+$-mediated tree diagrams. For example, in the semitaonic bottom quark decay, the asymmetry is given by $a_{CP} = C_{ps} \frac{\mathcal{I}m(XZ^*)}{m_H^2}$ and could be as large as 0.3 (see *e.g.* [200-202]).

*10.3. (ii) Effects of CP-even and CP-odd Scalar Mixing in Top Physics*

It is possible that the neutral scalars are mixtures of CP-even and CP-odd scalar fields [184,203-206,193-194]. Such a scalar couples to both scalar and pseudoscalar currents:

$$\mathcal{L}_Y = H_i \bar{f}(a_i^f + i b_i^f \gamma_5) f, \tag{10.9}$$

where $H_i$ is the physical Higgs boson and $a_i^f, b_i^f$ are functions of mixing angles in the matrix that diagonalizes the neutral scalar mass matrix. (Specifically, they are proportional to the components of, respectively, $\mathcal{R}e\phi_u$ and $\mathcal{I}m\phi_u$ in $H_i$.) CP violation in processes involving fermions is proportional to $a_i^f b_i^{f*}$. The natural place to look for manifestations of this type of CP violation is top physics, where the large Yukawa couplings allow large asymmetries (see e.g. [207]). Note that unlike our discussion above, the asymmetries here have nothing to do with FCNC processes. Actually, in models with NFC (even if softly broken [204]), the effects discussed here contribute negligibly to $\varepsilon_K$ and to CP asymmetries in $B$ decays. On the other hand, two loop diagrams with intermediate neutral scalar and top quark can induce a CP violating three gluon operator [208,209] that would give $d_N$ close to the experimental bound [209-212].

*10.4. (iii) Flavor Changing Neutral Scalar Exchange*

Natural flavor conservation needs not be exact in models of extended scalar sector [191,204,213,214]. In particular, it is quite likely that the existence of the additional scalars is related to flavor symmetries that explain the smallness and hierarchy in the Yukawa couplings. In this case, the new flavor changing couplings of these scalars are suppressed by the same selection rules as those that are responsible to the smallness of fermion masses and mixing, and there is no need to impose NFC [186,215-219,182]. An explicit framework, with Abelian horizontal symmetries, was presented in [220,135,98]. (For another related study, see [221].) We explain the general idea using these models. We emphasize that in this example the scalar with flavor changing couplings is a Standard Model singlet, and not an extra doublet, but the idea that these couplings are suppressed by approximate horizontal symmetries works in the same way.

The simplest model of ref. [220] extends the SM by supersymmetry and by an Abelian horizontal symmetry $\mathcal{H} = U(1)$ (or $Z_N$). The symmetry $\mathcal{H}$ is broken by a VEV of a single scalar $S$ that is a singlet of the SM gauge group. Consequently, Yukawa couplings that violate $\mathcal{H}$ arise only from nonrenormalizable terms and are therefore suppressed. Explicitly, the quark Yukawa terms have the form

$$\mathcal{L}_Y = X_{ij}^d \left(\frac{S}{M}\right)^{n_{ij}^d} Q_i \bar{d}_j \phi_d + X_{ij}^u \left(\frac{S}{M}\right)^{n_{ij}^u} Q_i \bar{u}_j \phi_u, \tag{10.10}$$

where $M$ is some high energy scale and $n_{ij}^q$ is the horizontal charge of the combination $Q_i \bar{q}_j \phi_q$ (in units of the charge of $S$). The terms (10.10) lead to quark masses and mixing as well as to flavor changing couplings, $Z^q_{ij}$, for the scalar $S$. The magnitude of the latter is then related to that of the effective Yukawa couplings $Y^q_{ij}$:

$$Z^q_{ij} \sim \frac{M^q_{ij}}{\langle S \rangle}. \tag{10.11}$$

Since the order of magnitude of each entry in the quark mass matrices is fixed in these models in terms of quark masses and mixing, the $Z^q_{ij}$ couplings can be estimated in terms of these physical parameters and the scale $\langle S \rangle$. For example, these couplings contribute to $K - \bar{K}$ mixing proportionally to

$$Z^d_{12} Z^{d*}_{21} \sim \frac{m_d m_s}{\langle S \rangle^2}. \tag{10.12}$$

With arbitrary phase factors in the various $Z^q_{ij}$ couplings, the contributions to neutral meson mixing are, in general, CP violating. In particular, there will be a contribution to $\varepsilon_K$ from $\mathcal{I}m(Z^d_{12} Z^{d*}_{21})$. Requiring that the $S$-mediated tree level contribution does not exceed the experimental value of $\varepsilon_K$ gives, for $\mathcal{O}(1)$ phases,

$$M_S \langle S \rangle \gtrsim 1.8 \ TeV^2. \tag{10.13}$$

We learn that (for $M_S \sim \langle S \rangle$) the mass of the $S$-scalar could be as low as 1.5 TeV, some 4 orders of magnitude below the bound corresponding to $\mathcal{O}(1)$ flavor changing couplings.

The flavor changing couplings of the $S$-scalar lead also to a tree level contribution to $B - \bar{B}$ mixing proportional to

$$Z^d_{13} Z^{d*}_{31} \sim \frac{m_d m_b}{\langle S \rangle^2}. \tag{10.14}$$

This means that, for phases of order 1, the neutral scalar exchange accounts for at most a few percent of $B - \bar{B}$ mixing. This cannot be signaled in $\Delta m_B$ (because of the hadronic uncertainties in the calculation) but could be signalled (if $\langle S \rangle$ is at the lower bound) in CP asymmetries in $B^0$ decays.

Finally, the contribution to $D - \bar{D}$ mixing, proportional to

$$Z^u_{12} Z^{u*}_{21} \sim \frac{m_u m_c}{\langle S \rangle^2}, \tag{10.15}$$

is below a percent of the current experimental bound. This is unlikely to be discovered in near-future experiments, even if the new phases maximize the interference effects in the $D^0 \to K^-\pi^+$ decay.

To summarize, models with horizontal symmetries naturally suppress flavor changing couplings of extra scalars. There is no need to invoke NFC even for new scalars at the TeV scale. Furthermore, the magnitude of the flavor changing couplings is related to the observed fermion parameters. Typically, contributions from neutral scalars with flavor changing couplings could dominate $\varepsilon_K$. If they do, then a signal at the few percent level in CP asymmetries in neutral $B$ decays is quite likely

### *10.5. The Superweak Scenario*

The original *superweak* scenario [222] stated that CP violation appears in a new $\Delta S = 2$ interaction while there is no CP violation in the SM $\Delta S = 1$ transitions. Consequently, the only large observable CP violating effect is $\varepsilon_K$, while $\varepsilon'/\varepsilon \sim 10^{-8}$ and EDMs are negligibly small. At present, the term "superweak CP violation" has been used for many different types of models. There are several reasons for this situation:

(i) The work of ref. [222] was concerned only with CP violation in $K$ decays. In extending the idea to other mesons, one may interpret the idea in various ways. On one side, it is possible that the superweak interaction is significant only in $K - \bar{K}$ mixing and (apart from the relaxation of the $\varepsilon_K$-bounds on the CKM parameters) has no effect on mixing of heavier mesons. On the other extreme, one may take the superweak scenario to imply that CP violation comes from $\Delta F = 2$ processes only for all mesons. The most common use of the term 'superweak' refers to the latter option, namely that there is no direct CP violation.

(ii) The scenario proposed in [222] did not employ any specific model. It was actually proposed even before the formulation of the Standard Model. To extend the idea to, for example, the neutral $B$ system, a model is required. Various models give very different predictions for CP asymmetries in $B$ decays.

If one extends the superweak scenario to the $B$ system by assuming that there is CP violation in $\Delta B = 2$ but not in $\Delta B = 1$ transitions, the prediction for CP asymmetries in $B$ decays into final CP eigenstates is that they are equal for all final states [223-225]. Whether these asymmetries are all small or could be large is model dependent. In addition, the asymmetries in charged $B$ decays vanish.

CP violation via neutral scalar exchange is the most commonly studied realization of the superweak idea. In particular, if the complex $Z^q_{ij}$ couplings presented in the previous section were the only source of CP violation, then this model would be superweak. The smallness of the $Z^q_{ij}$ couplings would make the contribution from neutral Higgs mediated diagrams negligible compared to the Standard Model diagrams in $\Delta S = 1$ processes, but the fact that they contribute to mixing at tree level would allow them to dominate the $\Delta S = 2$ processes. Various models (or scenarios) that realize the main features of the superweak idea can be found in refs. [226-227,191,213]. As mentioned above, there is a considerable variation in their predictions for $\varepsilon'/\varepsilon$, $d_N$ and other quantities. If we take the term 'superweak CP violation' to imply that there is only indirect CP violation, or at least that there is no direct CP violation in $K$ decays, then $\varepsilon'/\varepsilon \neq 0$ is inconsistent with this scenario which is therefore excluded.

## 11. Extensions of the Fermion Sector: Down Singlet Quarks

The fermion sector of the Standard Model is described in eq. (2.2). It can be extended by either a fourth, sequential generation or by non-sequential fermions, namely 'exotic' representations, different from those of (2.2). (The four generation model can only be viable if it is further extended to evade bounds related to the neutrino sector [228] and to electroweak precision data [26].) Our discussion in this chapter is focused on non-sequential fermions and their implications on CP asymmetries in neutral $B$ decays and on $K_L \to \pi\nu\bar{\nu}$.

### *11.1. The Theoretical Framework*

We consider a model with extra quarks in vector-like representations of the Standard Model gauge group,

$$d_4(3,1)_{-1/3} \;+\; \bar{d}_4(\bar{3},1)_{+1/3}. \tag{11.1}$$

Such (three pairs of) quark representations appear, for example, in $E_6$ GUTs. The mass matrix in the down sector, $M^d$, is now $4 \times 4$. (Note that the $M^d_{4i}$ entries do not violate $G_{\rm SM}$ and are, therefore, bare mass terms.)

How many independent CP violating parameters are there in $M^d$ and $M^u$? Since $M^d$ ($M^u$) is $4 \times 4$ ($3 \times 3$) and complex, there are 25 real and 25 imaginary parameters in these matrices. If we switch off the mass matrices, there is a global symmetry added to the model,

$$G^{\rm extra\ \it d}_{\rm global}(M^{d,u}=0) = U(3)_Q \times U(4)_{\bar d} \times U(3)_{\bar u} \times U(1)_{d_4}. \tag{11.2}$$

One can remove, at most, 12 real and 23 imaginary parameters. However, the model with the quark mass matrices switched on has still a global symmetry of $U(1)_B$, so one of the imaginary parameters cannot be removed. We conclude that there are 16 flavor parameters: 13 real ones, that is seven masses and six mixing angles, and 3 phases. These three phases are independent sources of CP violation.

### *11.2. Z-Mediated FCNC*

The most important feature of this model for our purposes is that it allows CP violating $Z$-mediated Flavor Changing Neutral Currents (FCNC). To understand how these FCNC arise, it is convenient to work in a basis where the up sector interaction eigenstates are identified with the mass eigenstates. The down sector interaction eigenstates are then related to the mass eigenstates by a $4 \times 4$ unitary matrix $K$. Charged current interactions are described by

$$\mathcal{L}^W_{\rm int} = \frac{g}{\sqrt{2}}(W^+_\mu V_{ij} \bar u_{iL} \gamma^\mu d_{jL} + {\rm h.c.}). \tag{11.3}$$

The charged current mixing matrix $V$ is a $3 \times 4$ sub-matrix of $K$:

$$V_{ij} = K_{ij} \ \ {\rm for} \ \ i = 1,2,3; \ j = 1,2,3,4. \tag{11.4}$$

The $V$ matrix is parameterized, as anticipated above, by six real angles and three phases, instead of three angles and one phase in the original CKM matrix. All three phases may affect CP asymmetries in $B^0$ decays. Neutral current interactions are described by

$$\begin{aligned} \mathcal{L}^Z_{\rm int} &= \frac{g}{\cos\theta_W} Z_\mu (J^{\mu 3} - \sin^2\theta_W J^\mu_{\rm EM}), \\ J^{\mu 3} &= -\frac{1}{2} U_{pq} \bar d_{pL} \gamma^\mu d_{qL} + \frac{1}{2} \delta_{ij} \bar u_{iL} \gamma^\mu u_{jL}. \end{aligned} \tag{11.5}$$

The neutral current mixing matrix for the down sector is $U = V^\dagger V$. As $V$ is not unitary, $U \neq \mathbf{1}$. In particular, its non-diagonal elements do not vanish:

$$U_{pq} = -K_{4p}^* K_{4q} \quad \text{for} \quad p \neq q. \tag{11.6}$$

The three elements which are most relevant to our study are

$$\begin{aligned} U_{ds} &= V_{ud}^* V_{us} + V_{cd}^* V_{cs} + V_{td}^* V_{ts}, \\ U_{db} &= V_{ud}^* V_{ub} + V_{cd}^* V_{cb} + V_{td}^* V_{tb}, \\ U_{sb} &= V_{us}^* V_{ub} + V_{cs}^* V_{cb} + V_{ts}^* V_{tb}. \end{aligned} \tag{11.7}$$

The fact that, in contrast to the Standard Model, the various $U_{pq}$ do not necessarily vanish, allows FCNC at tree level. This may substantially modify the predictions for CP asymmetries.

The flavor changing couplings of the $Z$ contribute to various FCNC processes. Relevant constraints arise from semileptonic FCNC $B$ decays:

$$\frac{\Gamma(B \to \ell^+ \ell^- X)_Z}{\Gamma(B \to \ell^+ \nu X)} = \left[(1/2 - \sin^2\theta_W)^2 + \sin^4\theta_W\right] \frac{|U_{db}|^2 + |U_{sb}|^2}{|V_{ub}|^2 + F_{\rm ps}|V_{cb}|^2}, \tag{11.8}$$

where $F_{\rm ps} \sim 0.5$ is a phase space factor. The experimental upper bound on $\Gamma(B \to \ell^+\ell^- X)$ gives

$$\left|\frac{U_{db}}{V_{cb}}\right| \leq 0.04, \qquad \left|\frac{U_{sb}}{V_{cb}}\right| \leq 0.04. \tag{11.9}$$

Additional constraints come from neutral $B$ mixing:

$$(\Delta m_B)_Z = \frac{\sqrt{2} G_F B_B f_B^2 m_B \eta_B}{3} |U_{db}|^2. \tag{11.10}$$

Using $\sqrt{B_B} f_B \gtrsim 0.16\ GeV$, we get

$$|U_{db}| \lesssim 9 \times 10^{-4}. \tag{11.11}$$

As concerns $\Delta m_{B_s}$, only lower bounds exist and consequently there is no analog bound on $|U_{sb}|$.

Bounds on $U_{ds}$ can be derived from the measurements of $\mathrm{BR}(K_L \to \mu^+\mu^-)$, $\mathrm{BR}(K^+ \to \pi^+\nu\bar{\nu})$, $\varepsilon_K$ and $\varepsilon'/\varepsilon$ yielding, respectively (for recent derivations, see [158,54]),

$$\begin{aligned} |\mathcal{R}e(U_{ds})| &\lesssim\ 10^{-5}, \\ |U_{ds}| &\leq 3\times 10^{-5}, \\ |\mathcal{R}e(U_{ds})\,\mathcal{I}m(U_{ds})| &\lesssim\ 1.3\times 10^{-9}, \\ |\mathcal{I}m(U_{ds})| &\lesssim\ 10^{-5}. \end{aligned} \tag{11.12}$$

(Note that the combination of bounds from $\mathrm{BR}(K_L \to \mu^+\mu^-)$ and from the recently improved $\varepsilon'/\varepsilon$ is stronger than the bounds from $\mathrm{BR}(K^+ \to \pi^+\nu\bar{\nu})$ and $\varepsilon_K$ [54]. The latter bounds are, however, subject to smaller hadronic uncertainties.)

*11.3. CP Asymmetries in B Decays*

The most interesting effects in this model concern CP asymmetries in neutral $B$ decays into final CP eigenstates [229-236,15]. We describe these effects in detail as they illustrate the type of new ingredients that are likely to affect CP asymmetries in neutral $B$ decays and the way in which the SM predictions might be modified. (If there exist light up quarks in exotic representations, they may introduce similar, interesting effects in neutral $D$ decays [69].)

If the $U_{qb}$ elements are not much smaller than the bounds (11.9) and (11.11), they will affect several aspects of physics related to CP asymmetries in $B$ decays.

(i) *Neutral B mixing*: The experimentally measured value of $\Delta m_{B_d}$ (and the lower bound on $\Delta m_{B_s}$) can be explained by Standard Model processes, namely box diagrams with intermediate top quarks. Still, the uncertainties in the theoretical calculations, such as the values of $f_B$ and $V_{td}$ (and the absence of an upper bound on $\Delta m_{B_s}$) allow a situation where SM processes do not give the dominant contributions to either or both of $\Delta m_{B_d}$ and $\Delta m_{B_s}$ [14]. The ratio between the $Z$-mediated tree diagram and the Standard Model box diagram is given by $(q = d, s)$

$$\frac{(\Delta m_{B_q})_{\rm tree}}{(\Delta m_{B_q})_{\rm box}} = \frac{2\sqrt{2}\pi^2}{G_F m_W^2 S_0(x_t)} \left|\frac{U_{qb}}{V_{tq}V_{tb}^*}\right|^2 \approx 150 \left|\frac{U_{qb}}{V_{tq}V_{tb}^*}\right|^2 \lesssim \begin{cases} 5 & q = d \\ 0.25 & q = s \end{cases}. \tag{11.13}$$

(The last inequality is derived under the assumption that the violation of CKM unitarity is not strong. The bound on $(\Delta m_{B_d})_{\rm tree}/(\Delta m_{B_d})_{\rm box}$ is higher if $|V_{td}V_{tb}^*| < 0.005$ holds.)

From (11.9) and (11.13) we learn that the $Z$-mediated tree diagram could give the dominant contribution to $\Delta m_{B_d}$ but at most 25% of $\Delta m_{B_s}$.

(ii) *Unitarity of the* $3 \times 3$ *CKM matrix*: Within the SM, unitarity of the three generation CKM matrix gives:

$$\begin{aligned}
\mathcal{U}_{ds} \equiv&\ V^*_{ud}V_{us} + V^*_{cd}V_{cs} + V^*_{td}V_{ts} = 0, \\
\mathcal{U}_{db} \equiv&\ V^*_{ud}V_{ub} + V^*_{cd}V_{cb} + V^*_{td}V_{tb} = 0, \\
\mathcal{U}_{sb} \equiv&\ V^*_{us}V_{ub} + V^*_{cs}V_{cb} + V^*_{ts}V_{tb} = 0.
\end{aligned} \tag{11.14}$$

Eq. (11.7), however, implies that now (11.14) is replaced by

$$\mathcal{U}_{ds} = U_{ds}, \quad \mathcal{U}_{db} = U_{db}, \quad \mathcal{U}_{sb} = U_{sb}. \tag{11.15}$$

A measure of the violation of (11.14) is given by

$$\left|\frac{U_{ds}}{V_{ud}V^*_{us}}\right| \lesssim 5 \times 10^{-4}, \quad \left|\frac{U_{db}}{V_{td}V^*_{tb}}\right| \lesssim 0.18, \quad \left|\frac{U_{sb}}{V_{ts}V^*_{tb}}\right| \lesssim 0.04. \tag{11.16}$$

The bound on $|U_{db}/(V_{td}V^*_{tb})|$ is even weaker if $|V_{td}|$ is lower than the three generation unitarity bound. We learn that the first of the SM relations in (11.14) is practically maintained, while the third can be violated by at most 4%. However, the $\mathcal{U}_{db} = 0$ constraint may be violated by $\mathcal{O}(0.2)$ effects. The Standard Model unitarity triangle should be replaced by a unitarity *quadrangle*. After the recent measurement of $a_{\psi K_S}$ [13], not only the magnitude of $U_{db}$ but also the phases $\bar{\alpha}$ and $\bar{\beta}$,

$$\bar{\alpha} \equiv \arg\left(\frac{V_{ud}V^*_{ub}}{U^*_{db}}\right), \qquad \bar{\beta} \equiv \arg\left(\frac{U^*_{db}}{V_{cd}V^*_{cb}}\right), \tag{11.17}$$

are constrained [15], but the constraints are not very strong.

(iii) $Z$-mediated $B$ decays: Our main interest in this chapter is in hadronic $B^0$ decays to CP eigenstates, where the quark sub-process is $\bar{b} \to \bar{u}_i u_i \bar{d}_j$, with $u_i = u, c$ and $d_j = d, s$. These decays get new contributions from $Z$-mediated tree diagrams, in addition to the standard $W$-mediated ones. The ratio between the amplitudes is

$$\frac{A_Z}{A_W} = \left[\frac{1}{2} - \frac{2}{3}\sin^2\theta_W\right]\left|\frac{U^*_{jb}}{V_{ij}V^*_{ib}}\right|. \tag{11.18}$$

We find that the $Z$ contributions can be safely neglected in $\bar{b} \to \bar{c}c\bar{s}$ ( $\lesssim 0.013$) and $\bar{b} \to \bar{c}c\bar{d}$ ( $\lesssim 0.03$). On the other hand, it may be significant in $\bar{b} \to \bar{u}u\bar{d}$ ( $\lesssim 0.12$), and in processes with no SM tree contributions, e.g. $\bar{b} \to \bar{s}s\bar{s}$, that may have comparable contributions from penguin and $Z$-mediated tree diagrams.

(iv) *New contributions to* $\Gamma_{12}(B_q)$: The difference in width comes from modes that are common to $B_q$ and $\bar{B}_q$. As discussed above, there are new contributions to such modes from $Z$-mediated FCNC. However, while the new contributions to $M_{12}$ are from tree level diagrams, *i.e.* $\mathcal{O}(g^2)$, those to $\Gamma_{12}$ are still coming from a box-diagram, *i.e.* $\mathcal{O}(g^4)$. Consequently, no significant enhancement of the SM value of $\Gamma_{12}$ is expected, and the relation $\Gamma_{12} \ll M_{12}$ is maintained. (The new contribution could significantly modify the leptonic asymmetry in neutral $B$ decays [237,15] though the asymmetry remains small.)

The fact that $M_{12}(B^0)$ could be dominated by the $Z$-mediated FCNC together with the fact that this new amplitude depends on new CP violating phases means that large deviations from the Standard Model predictions for CP asymmetries are possible. As $\Gamma_{12} \ll M_{12}$ is maintained, future measurements of certain modes will still be subject to a clean theoretical interpretation in terms of the extended electroweak sector parameters.

Let us assume that, indeed, $M_{12}$ is dominated by the new physics. (Generalization to the case that the new contribution is comparable to (but not necessarily dominant over) the Standard Model one is straightforward [231,235].) Then

$$\left(\frac{p}{q}\right)_B \approx \frac{U_{db}^*}{U_{db}}. \tag{11.19}$$

We argued above that $b \to c\bar{c}s$ is still dominated by the $W$ mediated diagram. Furthermore, the first unitarity constraint in (11.14) is practically maintained. Then it is straightforward to evaluate the CP asymmetry in $B \to \psi K_S$. We find that it simply measures an angle of the unitarity quadrangle:

$$a_{CP}(B \to \psi K_S) = -\sin 2\bar{\beta}. \tag{11.20}$$

The new contribution to $b \to c\bar{c}d$ is $\mathcal{O}(3\%)$, which is somewhat smaller than the SM penguins. So we still have, to a good approximation, (taking into account CP-parities)

$$a_{CP}(B \to \psi K_S) \approx -a_{CP}(B \to DD). \tag{11.21}$$

Care has to be taken regarding $b \to u\bar{u}d$ decays. Here, direct CP violation may be large [234] and prevent a clean theoretical interpretation of the asymmetry. Only if the asymmetry is large, so that the shift from the $Z$-mediated contribution to the decay is small, we get

$$a_{CP}(B \to \pi\pi) = -\sin 2\bar{\alpha}. \tag{11.22}$$

The important point about the modification of the SM predictions is then not that the angles $\alpha, \beta$ and $\gamma$ may have very different values from those predicted by the SM, but rather that the CP asymmetries do not measure these angles anymore. As the experimental constraints on $\bar{\alpha}$ and $\bar{\beta}$ are still rather weak [15], a large range is possible for each of the asymmetries. This model demonstrates that there exist extensions of the SM where dramatic deviations from its predictions for CP asymmetries in $B$ decays are not unlikely.

Another interesting point concerns $B_s$ decays. If $B_s - \bar{B}_s$ mixing as well as the $b \to c\bar{c}s$ decay are dominated by the SM diagrams, we have, similar to the SM,

$$a_{CP}(B_s \to \psi\phi) \approx 0. \tag{11.23}$$

As shown in ref. [20], this is a sufficient condition for the angles extracted from $B \to \psi K_S$, $B \to \pi\pi$ and the relative phase between the $B_s - \bar{B}_s$ mixing amplitude and the $b \to u\bar{u}d$ decay amplitude (if it can be deduced from experiment) to sum up to $\pi$ (up to possible effects of direct CP violation). This happens in spite of the fact that the first two asymmetries do not correspond to $\beta$ and $\alpha$ of the unitarity triangle.

### *11.4. The $K_L \to \pi\nu\bar{\nu}$ Decay*

In chapter 7 we argued that the only potentially significant new contribution to $a_{\pi\nu\nu}$ can come from the decay amplitude. $Z$-mediated FCNC provide an explicit example of New Physics that may modify the SM prediction for $a_{\pi\nu\nu}$ of eq. (4.38). Assuming that the $Z$-mediated tree diagram dominates $K \to \pi\nu\bar{\nu}$, we get [229]

$$\sin\theta_K = \mathcal{I}mU_{ds}/|U_{ds}|. \tag{11.24}$$

Bounds on the relevant couplings were given in eq. (11.12) above. We learn that large effects are possible. When $|\mathcal{R}e(U_{ds})|$ and $|\mathcal{I}m(U_{ds})|$ are close to their upper bounds, the branching ratios $BR(K^+ \to \pi^+\nu\bar{\nu})$ and $BR(K_L \to \pi^0\nu\bar{\nu})$ are both $O(10^{-10})$ and $a_{\pi\nu\nu} = \mathcal{O}(1)$. Furthermore, as in this case the SM contribution is small, the measurement of $BR(K^+ \to \pi^+\nu\bar{\nu})$ approximately determines $|U_{ds}|$, and with the additional measurement of $BR(K_L \to \pi^0\nu\bar{\nu})$, $\arg(U_{ds})$ is approximately determined as well.

## 12. Conclusions

Experiments have not yet probed in a significant way the mechanism of CP violation. There is a large number of open questions concerning CP violation. Here are some examples:

- **Why are the measured parameters, $\varepsilon_K$ and $\varepsilon'_K$, small?**

The answer in the Standard Model is that CP violation is screened in processes that are dominated by the first two quark generations by small mixing angles. We have seen examples of new physics, that is supersymmetry with approximate CP, where the reason is the smallness of all CP violating phases.

Observing CP asymmetries of order one, as expected in processes that involve the first and third generations such as $B \to \psi K_S$, $K \to \pi\nu\bar{\nu}$ or even in charged $B$ decays, $B^\pm \to K^\pm\pi^0$, will exclude the approximate CP scenario.

- **What is the number of independent CP violating phases?**

The answer in the Standard Model is *one*, the Kobayashi-Maskawa phase. We have encountered models with a larger number, *e.g.* forty four in the supersymmetric standard model.

If the pattern predicted by the Standard Model, *e.g.* small CP asymmetries in $B_s \to \psi\phi$ and in $D \to K\pi$, a strong correlation between CP violation in $B \to \psi K_S$ and in $K_L \to \pi\nu\bar{\nu}$, equal asymmetries in $B \to \psi K_S$ and in $B \to \phi K_S$, etc., is inconsistent with measurements, then probably there are several independent phases.

- **Why is CP violated?**

The answer in the Standard Model is explicit breaking by complex Yukawa couplings. In left-right-symmetric models, the Lagrangian can be CP symmetric and the breaking is spontaneous.

It will be difficult to answer this question by experimental measurements, unless the correlations predicted by a specific model of spontaneous CP violation will be experimentally confirmed.

• **Is CP violation restricted to flavor changing interactions?**

This is indeed the case in the Standard Model. But in many of its extensions, such as supersymmetry, there is flavor diagonal CP violation.

Observation of an electric dipole moment or of CP violation in $t\bar{t}$ production will provide strong hints for flavor diagonal CP violation.

• **Is CP violation restricted to quark interactions?**

This is the case in the Standard Model but not if neutrinos have masses.

Observation of CP asymmetries in neutrino oscillation experiments will be a direct evidence of CP violation in the lepton sector.

• **Is CP violation restricted to the weak interactions?**

In the Standard Model, CP violation appears in charged current (that is, $W$-mediated) weak interactions only. In multi-scalar models, it appears in scalar interactions. In supersymmetry, it appears in strong interactions.

Observation of transverse lepton polarization in meson decays will provide evidence for CP violation in interactions that are not mediated by vector bosons.

There are more questions that we can ask and answers that we will learn in the near future. But the list above is enough to demonstrate how unique the Standard Model picture of CP violation is, how sensitive is CP violation to new physics, and how important are present and future experiments that will search for CP violation.

## Acknowledgements

I thank Gabriela Barenboim, Sven Bergmann, Galit Eyal, Yuval Grossman, Stephane Plaszczynski, Helen Quinn, Marie-Helene Schune and Joao Silva for useful discussions. Y.N. is supported by a DOE grant DE-FG02-90ER40542, by the Ambrose Monell Foundation, by the Israel Science Foundation founded by the Israel Academy of Sciences and Humanities, and by the Minerva Foundation (Munich).

## References


[1] M. Kobayashi and T. Maskawa, Prog. Theo. Phys. 49 (1973) 652.

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

[3] A.D. Sakharov, ZhETF Pis. Red. 5 (1967) 32; JETP Lett. 5 (1967) 24.

[4] For a review see, A.G. Cohen, D.B. Kaplan and A.E. Nelson, Ann. Rev. Nucl. Part. Sci. 43 (1993) 27.

[5] G.R. Farrar and M.E. Shaposhnikov, Phys. Rev. D50 (1994) 774, hep-ph/9305275.

[6] M.B. Gavela *et al.*, Nucl. Phys. B430 (1994) 382, hep-ph/9406288.

[7] P. Huet and E. Sather, Phys. Rev. D51 (1995) 379, hep-ph/9404302.

[8] A.B. Carter and A.I. Sanda, Phys. Rev. Lett. 45 (1980) 952; Phys. Rev. D23 (1981) 1567.

[9] I.I. Bigi and A.I. Sanda, Nucl. Phys. B193 (1981) 85; Nucl. Phys. B281 (1987) 41.

[10] I. Dunietz and J. Rosner, Phys. Rev. D34 (1986) 1404.

[11] K. Ackerstaff *et al.*, OPAL collaboration, Eur. Phys. J. C5 (1998) 379, hep-ex/9801022.

[12] F. Abe *et al.*, CDF Collaboration, Phys. Rev. Lett. 81 (1998) 5513, hep-ex/9806025.

[13] T. Affolder *et al.*, CDF Collaboration, hep-ex/9909003.

[14] G. Barenboim, G. Eyal and Y. Nir, hep-ph/9905397, Phys. Rev. Lett., in press.

[15] G. Eyal and Y. Nir, JHEP 9909 (1999) 013, hep-ph/9908296.

[16] L.S. Littenberg, Phys. Rev. D39 (1989) 3322.

[17] G. Buchalla and A.J. Buras, Nucl. Phys. B400 (1993) 225.

[18] A.J. Buras, Phys. Lett. B333 (1994) 476, hep-ph/9405368.

[19] Y. Grossman and Y. Nir, Phys. Lett. B398 (1997) 163, hep-ph/9701313.

[20] Y. Nir and D. Silverman, Nucl. Phys. B345 (1990) 301.

[21] C.O. Dib, D. London and Y. Nir, Int. J. Mod. Phys. A6 (1991) 1253.

[22] Y. Nir and H.R. Quinn, Ann. Rev. Nucl. Part. Sci. 42 (1992) 211.

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

[24] N. Cabibbo, Phys. Rev. Lett. 10 (1963) 531.

[25] L.L. Chau and W.Y. Keung, Phys. Rev. Lett. 53 (1984) 1802.

[26] C. Caso *et al.*, Particle Data Group, Eur. Phys. J. C3 (1998) 1.

[27] Z. Maki, M. Nakagawa and S. Sakata, Prog. Theo. Phys. 28 (1962) 247.

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

[29] C.O. Dib, I. Dunietz, F.J. Gilman and Y. Nir, Phys. Rev. D41 (1990) 1522.

[30] H. Harari and Y. Nir, Phys. Lett. 195B (1987) 586.

[31] Y. Nir, in *proceedings of the 1998 European School of High Energy Physics* (University of St. Andrews, 1998), hep-ph/9810520.

[32] S. Plaszczynski and M.-H. Schune, BaBar note 340.

[33] S. Plaszczynski, hep-ph/9804330.

[34] Y. Grossman, Y. Nir, S. Plaszczynski and M.-H. Schune, Nucl. Phys. B511 (1998) 69, hep-ph/9709288.

[35] The BaBar physics book, eds. P. Harrison and H.R. Quinn, SLAC-R-504 (1998).

[36] A.J. Buras, hep-ph/9905437.

[37] R. Fleischer and T. Mannel, Phys. Rev. D57 (1998) 2752, hep-ph/9704423.

[38] M. Neubert and J.L. Rosner, Phys. Lett. B441 (1998) 403, hep-ph/9808493.

[39] A.J. Buras and R. Fleischer, hep-ph/9810260.

[40] M. Neubert, JHEP 02 (1999) 014, hep-ph/9812396; hep-ph/9904321.

[41] R. Fleischer and J. Matias, hep-ph/9906274.

[42] R.D. Peccei and H.R. Quinn, Phys. Rev. Lett. 38 (1977) 1440; Phys. Rev. D16 (1977) 1791.

[43] G.C. Branco, L. Lavoura and J.P. Silva, *CP Violation* (Oxford University Press, 1999).

[44] Y. Grossman, B. Kayser and Y. Nir, Phys. Lett. B415 (1997) 90, hep-ph/9708398.

[45] H. Burkhardt *et al.*, NA31 collaboration, Phys. Lett. B206 (1988) 169.

[46] G.D. Barr *et al.*, NA31 collaboration, Phys. Lett. B317 (1993) 233.

[47] L.K. Gibbons *et al.*, E731 collaboration, Phys. Rev. Lett. 70 (1993) 1203.

[48] A. Alavi-Harati *et al.*, KTeV collaboration, Phys. Rev. Lett. 83 (1999) 22, hep-ex/9905060.

[49] V. Fanti *et al.*, NA48 collaboration, hep-ex/9909022.

[50] S. Herrlich and U. Nierste, Phys. Rev. D52 (1995) 6505, hep-ph/9507262; Nucl. Phys. B476 (1996) 27, hep-ph/9604330.

[51] A.J. Buras, hep-ph/9806471.

[52] Y.-Y. Keum, U. Nierste and A.I. Sanda, Phys. Lett. B457 (1999) 157, hep-ph/9903230.

[53] S. Bosch *et al.*, hep-ph/9904408.

[54] A.J. Buras and L. Silvestrini, Nucl. Phys. B546 (1999) 299, hep-ph/9811471.

[55] S. Bertolini, J.O. Eeg, M. Fabbrichesi and E.I. Lashin, Nucl. Phys. B514 (1998) 93, hep-ph/9706260.

[56] S. Bertolini, J.O. Eeg and M. Fabbrichesi, hep-ph/9802405.

[57] T. Hambye, G.O. Kohler, E.A. Paschos and P.H. Soldan, hep-ph/9906434.

[58] A.A. Belkov, G. Bohm, A.V. Lanyov and A.A. Moshkin, hep-ph/9907335.

[59] G. Buchalla and G. Isidori, Phys. Lett. B140 (1998) 170, hep-ph/9806501.

[60] G. Perez, JHEP 9909 (1999) 019, hep-ph/9907205.

[61] S. Adler *et al.*, E787 Collaboration, Phys. Rev. Lett. 79 (1997) 2204, hep-ex/9708031.

[62] J. Adams *et al.*, KTeV Collaboration, Phys. Lett. B447 (1999) 240, hep-ex/9806007.

[63] E791 collaboration, Phys. Rev. Lett. 77 (1996) 2384, hep-ex/9606016.

[64] E791 collaboration, Phys. Rev. D57 (1998) 13, hep-ex/9608018.

[65] ALEPH collaboration, Phys. Lett. B436 (1998) 211, hep-ex/9811021.

[66] E791 collaboration, Phys. Rev. Lett. 83 (1999) 32, hep-ex/9903012.

[67] CLEO collaboration, hep-ex/9908040.

[68] S. Bergmann and Y. Nir, JHEP 9909 (1999) 031, hep-ph/9909391.

[69] G. Blaylock, A. Seiden and Y. Nir, Phys. Lett. B355 (1995) 555, hep-ph/9504306.

[70] L. Wolfenstein, Phys. Rev. Lett. 75 (1995) 2460, hep-ph/9505285.

[71] I.I. Bigi, V.A. Khoze, N.G. Uraltsev and A.I. Sanda, in *CP Violation*, ed. C. Jarlskog (World Scientific, Singapore, 1992).

[72] R. Aleksan *et al.*, Phys. Lett. B316 (1993) 567.

[73] M. Beneke *et al.*, Phys.Lett. B459 (1999) 631, hep-ph/9808385.

[74] L. Wolfenstein, Phys. Rev. D57 (1998) 5453.

[75] A.J. Buras, G. Buchalla and M.E. Lautenbacher, Rev. Mod. Phys. 68 (1996) 1125, hep-ph/9512380.

[76] J. Hagelin, Nucl. Phys. B193 (1981) 123.

[77] A.J. Buras, W. Slominski and H. Steger, Nucl. Phys. B245 (1984) 369.

[78] M. Beneke, G. Buchalla and I. Dunietz, Phys. Rev. D54 (1996) 4419, hep-ph/9605259.

[79] L. Randall and S. Su, Nucl. Phys. B540 (1999) 37, hep-ph/9807377.

[80] R.N. Cahn and M.P. Worah, Phys. Rev. D60 (1999) 076006, hep-ph/9904480.

[81] M. Bander, S. Silverman and A. Soni, Phys. Rev. Lett. 43 (1979) 242.

[82] R. Godang *et al.*, CLEO Collaboration, Phys. Rev. Lett. 80 (1998) 3456, hep-ex/9711010.

[83] S.L. Glashow, J. Iliopoulos and L. Maiani, Phys. Rev. D2 (1970) 1285.

[84] R. Fleischer, Phys. Lett. B341 (1995) 205, hep-ph/9409290.

[85] A.J. Buras and R. Fleischer, Phys. Lett. B341 (1995) 379, hep-ph/9409244.

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

[87] Y. Grossman and H. Quinn, Phys. Rev. D58 (1998) 017504, hep-ph/9712306.

[88] H.J. Lipkin, Y. Nir, H.R. Quinn and A. Snyder, Phys. Rev. D44 (1991) 1454.

[89] H.R. Quinn and A. Snyder, Phys. Rev. D48 (1993) 2139.

[90] R. Fleischer, Int. J. Mod. Phys. A12 (1997) 2459, hep-ph/9612446.

[91] Y. Grossman and M.P. Worah, Phys. Lett. B395 (1997) 241, hep-ph/9612269.

[92] P.G. Harris *et al.*, Phys. Rev. Lett. 82 (1999) 904.

[93] P. Ramond, Lectures given at TASI 94, hep-ph/9412234.

[94] J.A. Bagger, Lectures given at TASI 95, hep-ph/9604232.

[95] S. Martin, hep-ph/9709356.

[96] J. Louis, I. Brunner and S.J. Huber, hep-ph/9811341.

[97] Y. Shadmi and Y. Shirman, hep-th/9907225.

[98] Y. Grossman, Y. Nir and R. Rattazzi, in *Heavy Flavours II*, eds. A.J. Buras and M. Lindner (World Scientific), hep-ph/9701231.

[99] M. Dugan, B. Grinstein and L.J. Hall, Nucl. Phys. B255 (1985) 413.

[100] S. Dimopoulos ans S. Thomas, Nucl. Phys. B465 (1996) 23, hep-ph/9510220.

[101] W. Buchmuller and D. Wyler, Phys. Lett. B121 (1983) 321.

[102] J. Polchinski and M. Wise, Phys. Lett. B125 (1983) 393.

[103] W. Fischler, S. Paban and S. Thomas, Phys. Lett. B289 (1992) 373.

[104] J. Ellis, S. Ferrara and D. Nanopoulos, Phys. Lett. B114 (1982) 231.

[105] S. Barr, Int. J. Mod. Phys. A8 (1993) 209.

[106] T. Falk and K.A. Olive, Phys. Lett. B439 (1998) 71, hep-ph/9806236; Phys. Lett. B375 (1996) 196, hep-ph/9602299.

[107] T. Ibrahim and P. Nath, Phys. Lett. B418 (1998) 98, hep-ph/9707409; Phys. Rev. D57 (1998) 478, hep-ph/9708456; Phys. Rev. D58 (1998) 111301, hep-ph/9807501.

[108] A. Bartl *et al.*, Phys. Rev. D60 (1999) 073003, hep-ph/9903402.

[109] S.M. Barr and S. Khalil, hep-ph/9903425.

[110] M. Brhlik, G.J. Good and G.L. Kane, Phys. Rev. D59 (1999) 115004, hep-ph/9810457.

[111] M. Brhlik, L. Everett, G.L. Kane and J. Lykken, Phys. Rev. Lett. 83 (1999) 2124, hep-ph/9905215.

[112] S. Pokorski, J. Rosiek and C.A. Savoy, hep-ph/9906206.

[113] E. Commins *et al.*, Phys. Rev. A50 (1994) 2960.

[114] K. Abdullah *et al.*, Phys. Rev. Lett. 65 (1990) 234.

[115] J. Donoghue, H. Nilles and D. Wyler, Phys. Lett. B128 (1983) 55.

[116] F. Gabbiani and A. Masiero, Nucl. Phys. B322 (1989) 235.

[117] J.S. Hagelin, S. Kelley and T. Tanaka, Nucl. Phys. B415 (1994) 293, hep-ph/9304218.

[118] E. Gabrielli, A. Masiero and L. Silvestrini, Phys. Lett. B374 (1996) 80, hep-ph/9509379.

[119] F. Gabbiani, E. Gabrielli, A. Masiero and L. Silvestrini, Nucl. Phys. B477 (1996) 321, hep-ph/9604387.

[120] Y. Nir, Nucl. Phys. B273 (1986) 567.

[121] M. Dine, A. Nelson, Y. Nir and Y. Shirman, Phys. Rev. D53 (1996) 2658, hep-ph/9507378.

[122] M. Dine, A. Nelson and Y. Shirman, Phys. Rev. D51 (1994) 1362, hep-ph/9408384.

[123] M. Dine, Y. Nir and Y. Shirman, Phys. Rev. D55 (1997) 1501, hep-ph/9607397.

[124] V. Kaplunovsky and J. Louis, Phys. Lett. B306 (1993) 269, hep-th/9303040.

[125] R. Barbieri, J. Louis and M. Moretti, Phys. Lett. B312 (1993) 451, hep-ph/9305262; (E) *ibid.* B316 (1993) 632.

[126] A. Romanino and A. Strumia, Nucl. Phys. B490 (1997) 3, hep-ph/9610485.

[127] S. Bertolini and F. Vissani, Phys. Lett. B324 (1994) 164, hep-ph/9311293.

[128] T. Inui, Y. Mimura, N. Sakai and T. Sasaki, Nucl. Phys. B449 (1995) 49, hep-ph/9503383.

[129] S.A. Abel, W.N. Cottingham and I.B. Wittingham, Phys. Lett. B370 (1996) 106, hep-ph/9511326.

[130] S. Bertolini, F. Borzumati, A. Masiero and G. Ridolfi, Nucl. Phys. B353 (1991) 591.

[131] T. Goto, T. Nihei and Y. Okada, Phys. Rev. D53 (1996) 5233, hep-ph/9510286.

[132] T. Nihei, Prog. Theor. Phys. 98 (1997) 1157, hep-ph/9707336.

[133] S. Baek and P. Ko, Phys. Rev. Lett. 83 (1999) 488, hep-ph/9812229.

[134] Y. Nir and N. Seiberg, Phys. Lett. B309 (1993) 337, hep-ph/9304307.

[135] M. Leurer, Y. Nir and N. Seiberg, Nucl. Phys. B420 (1994) 468, hep-ph/9410320.

[136] Y. Grossman and Y. Nir, Nucl. Phys. B448 (1994) 30, hep-ph/9502418.

[137] Y. Nir and R. Rattazzi, Phys. Lett. B382 (1996) 363, hep-ph/9603233.

[138] M. Dine, A. Kagan and R.G. Leigh, Phys. Rev. D48 (1993) 4269, hep-ph/9304299.

[139] P. Pouliot and N. Seiberg, Phys. Lett. B318 (1993) 169, hep-ph/9308363.

[140] L.J. Hall and H. Murayama, Phys. Rev. Lett. 75 (1995) 3985, hep-ph/9508296.

[141] A. Pomarol and D. Tommasini, Nucl. Phys. B466 (1996) 3, hep-ph/9507462.

[142] R. Barbieri, G. Dvali and L.J. Hall, Phys. Lett. B377 (1996) 76, hep-ph/9512388.

[143] C. Carone, L.J. Hall and H. Murayama, Phys. Rev. D54 (1996) 2328, hep-ph/9602364.

[144] Z.G. Berezhiani, hep-ph/9609342.

[145] G. Eyal, Phys. Lett. B441 (1998) 191, hep-ph/9807308.

[146] M. Dine, A. Kagan and S. Samuel, Phys. Lett. B243 (1990) 250.

[147] S. Weinberg, Phys. Rev. Lett. 63 (1989) 2333.

[148] A.G. Cohen, D.B. Kaplan, F. Lepeintre and A.E. Nelson, Phys. Rev. Lett. 78 (1997) 2300, hep-ph/9610252.

[149] A.G. Cohen, D.B. Kaplan and A.E. Nelson, Phys. Lett. B388 (1996) 588, hep-ph/9607394.

[150] S.A. Abel and J.M. Frere, Phys. Rev. D55 (1997) 1623, hep-ph/9608251.

[151] K.S. Babu and S.M. Barr, Phys. Rev. Lett. 72 (1994) 2831, hep-ph/9309249.

[152] G. Eyal and Y. Nir, Nucl. Phys. B528 (1998) 21, hep-ph/9801411.

[153] K.S. Babu, B. Dutta and R.N. Mohapatra, hep-ph/9905464.

[154] G. Eyal, A. Masiero, Y. Nir and L. Silvestrini, hep-ph/9908382.

[155] A. Masiero and H. Murayama, Phys. Rev. Lett. 83 (1999) 907, hep-ph/9903363.

[156] X.G. He, H. Murayama, S. Pakvasa and G. Valencia, hep-ph/9909562.

[157] R. Barbieri, R. Contino and A. Strumia, hep-ph/9908255.

[158] G. Colangelo and G. Isidori, JHEP 9809 (1998) 009, hep-ph/9808487.

[159] S. Khalil, T. Kobayashi and A. Masiero, hep-ph/9903544.

[160] K.S. Babu, B. Dutta and R.N. Mohapatra, hep-ph/9905464.

[161] Y. Nir and M.P. Worah, Phys. Lett. B423 (1998) 319, hep-ph/9711215.

[162] A.J. Buras, A. Romanino and L. Silvestrini, Nucl. Phys. B520 (1998) 3, hep-ph/9712398.

[163] D. Chang, Nucl. Phys. B214 (1983) 435.
[164] G.C. Branco, J.-M. Frère and J.-M. Gérard, Nucl. Phys. B221 (1983) 317.
[165] H. Harari and M. Leurer, Nucl. Phys. B233 (1984) 221.
[166] G. Ecker, W. Grimus and H. Neufeld, Nucl. Phys. B247 (1983) 70.
[167] G. Ecker and W. Grimus, Nucl. Phys. B258 (1985) 328.
[168] M. Leurer, Nucl. Phys. B266 (1986) 147.
[169] R.N. Mohapatra and J.C. Pati, Phys. Rev. D11 (1975) 566.
[170] G. Barenboim, J. Bernabeu and M. Raidal, Nucl. Phys. B478 (1996) 527, hep-ph/9608450.
[171] G. Barenboim, J. Bernabeu and M. Raidal, Nucl. Phys. B511 (1998) 577, hep-ph/9702337.
[172] G. Beall, M. Bander and A. Soni, Phys. Rev. Lett. 48 (1982) 848.
[173] G. Barenboim, J. Bernabeu, J. Prades and M. Raidal, Phys. Rev. D55 (1997) 4213, hep-ph/9611347.
[174] G. Ecker and W. Grimus, Z. Phys. C30 (1986) 293.
[175] D. London and D. Wyler, Phys. Lett. B232 (1989) 503.
[176] G. Barenboim, J. Bernabeu, J. Matias and M. Raidal, hep-ph/9901265.
[177] G. Barenboim, Phys. Lett. B443 (1998) 317, hep-ph/9810325.
[178] G. Beall and A. Soni, Phys. Rev. Lett. 47 (1981) 552.
[179] X.G. He, H.J. McKellar and S. Pakvasa, Phys. Rev. Lett. 61 (1988) 1267.
[180] D. Chang, C.S. Li and T.C. Yuan, Phys. Rev. D42 (1990) 867.
[181] G. Barenboim and J. Bernabeu, Z. Phys. C73 (1997) 321, hep-ph/9603379.
[182] Y.L. Wu and L. Wolfenstein, Phys. Rev. Lett. 73 (1994) 1762, hep-ph/9409421.
[183] S. Weinberg, Phys. Rev. Lett. 37 (1976) 657.
[184] T.D. Lee, Phys. Rev. D8 (1973) 1226.
[185] S.L. Glashow and S. Weinberg, Phys. Rev. D15 (1977) 1958.
[186] T.P. Cheng and M. Sher, Phys. Rev. D35 (1987) 3484.
[187] H. Georgi and D.V. Nanopoulos, Phys. Lett. 82B (1979) 95.
[188] G.C. Branco, Phys. Rev. Lett. 44 (1980) 504.
[189] P. Krawczyk and S. Pokorski, Nucl. Phys. B364 (1991) 11.
[190] Y. Grossman and Y. Nir, Phys. Lett. B313 (1993) 126, hep-ph/9306292.
[191] J. Liu and L. Wolfenstein, Nucl. Phys. B289 (1987) 1.
[192] M. Masip and A. Rasin, Phys. Rev. D58 (1998) 035007, hep-ph/9803271.
[193] S. Weinberg, Phys. Rev. D42 (1990) 860.
[194] L. Lavoura, Int. J. Mod. Phys. A8 (1993) 375.
[195] I.I. Bigi and A.I. Sanda, Phys. Rev. Lett. 58 (1987) 1605.
[196] H.Y. Cheng, Phys. Rev. D42 (1990) 2329.
[197] M. Leurer, Phys. Rev. Lett. 62 (1989) 1967.

[198] P. Castoldi, J.M. Frère and G. Kane, Phys. Rev. D39 (1989) 2633.
[199] M.B. Gavela *et al.*, Phys. Rev. D39 (1989) 1870.
[200] D. Atwood, G. Eilam and A. Soni, Phys. Rev. Lett. 71 (1993) 492, hep-ph/9303268.
[201] Y. Grossman and Z. Ligeti, Phys. Lett. B332 (1994) 373, hep-ph/9403376; Phys. Lett. B347 (1995) 399, hep-ph/9409418.
[202] Y. Grossman, H.E. Haber and Y. Nir, Phys. Lett. B357 (1995) 630, hep-ph/9507213.
[203] C.H. Albright, J. Smith, and S.H.H. Tye, Phys. Rev D21 (1980) 711.
[204] G.C. Branco and M.N. Rebelo, Phys. Lett. B160 (1985) 117.
[205] A. Méndez and A. Pomarol, Phys. Lett. B272 (1991) 313.
[206] A. Pomarol and R. Vega, Nucl. Phys. B413 (1993) 3, hep-ph/9305272.
[207] C.E. Schmidt and M.E. Peskin, Phys. Rev. Lett. 69 (1992) 410.
[208] S. Weinberg, Phys. Rev. Lett. 63 (1989) 2333.
[209] D. Dicus, Phys. Rev. D41 (1990) 999.
[210] J.F. Gunion and D. Wyler, Phys. Lett. B248 (1990) 170.
[211] E. Braaten, C.S. Li and T.C. Yuan, Phys. Rev. Lett. 64 (1990) 1709.
[212] A. De Rújula, M.B. Gavela, O. Pène and F.J. Vegas, Phys. Lett. B245 (1990) 640.
[213] J. Liu and L. Wolfenstein, Phys. Lett. B197 (1987) 536.
[214] H. Haber and Y. Nir, Nucl. Phys. B335 (1990) 363.
[215] A.S. Joshipura and S.D. Rindani, Phys. Lett. B260 (1991) 149.
[216] A. Antaramian, L.J. Hall and A. Rasin, Phys. Rev. Lett. 69 (1992) 1871, hep-ph/9206205.
[217] L.J. Hall and S. Weinberg, Phys. Rev. D48 (1993) 979, hep-ph/9303241.
[218] G.C. Branco, W. Grimus and L. Lavoura, Phys. Lett. B380 (1996) 119, hep-ph/9601383.
[219] D. Atwood, L. Reina and A. Soni, Phys. Rev. D54 (1996) 3269, hep-ph/9603210; Phys. Rev. D55 (1997) 3156, hep-ph/9609279.
[220] M. Leurer, Y. Nir and N. Seiberg, Nucl. Phys. B398 (1993) 319, hep-ph/9212278.
[221] N. Arkani-Hamed, C.D. Carone, L.J. Hall and H. Murayama, Phys. Rev. D54 (1996) 7032, hep-ph/9607298.
[222] L. Wolfenstein, Phys. Rev. Lett. 13 (1964) 562.
[223] B. Winstein, Phys. Rev. Lett. 68 (1992) 1271.
[224] J.M. Soares and L. Wolfenstein, Phys. Rev. D46 (1992) 256.
[225] B. Winstein and L. Wolfenstein, Rev. Mod. Phys. 65 (1993) 1113.
[226] J.-M. Gerard and T. Nakada, Phys. Lett. B261 (1991) 474.
[227] L. Lavoura, Int. J. Mod. Phys. A9 (1994) 1873.
[228] H. Harari and Y. Nir, Nucl. Phys. B292 (1987) 251.
[229] Y. Nir and D. Silverman, Phys. Rev. D42 (1990) 1477.

[230] D. Silverman, Phys. Rev. D45 (1992) 1800; Int. J. Mod. Phys. A11 (1996) 2253, hep-ph/9504387

[231] G.C. Branco, T. Morozumi, P.A. Parada and M.N. Rebelo, Phys. Rev. D48 (1993) 1167.

[232] W-S. Choong and D. Silverman, Phys. Rev. D49 (1994) 2322.

[233] V. Barger, M.S. Berger and R.J.N. Phillips, Phys. Rev. D52 (1995) 1663, hep-ph/9503204.

[234] M. Gronau and D. London, Phys. Rev. D55 (1997) 2845, hep-ph/9608430 (1996).

[235] G. Barenboim and F.J. Botella, Phys. Lett. B433 (1998) 385, hep-ph/9708209.

[236] G. Barenboim, F.J. Botella, G.C. Branco and O. Vives, Phys. Lett. B422 (1998) 277, hep-ph/9709369.

[237] G.C. Branco, P.A. Parada and T. Morozumi, Phys. Lett. B306 (1993) 398.

# SOME BITS OF THE HISTORY OF CP VIOLATION

Val L. Fitch*

Physics Department
Princeton University, Princeton, NJ 08544

It is especially appropriate, here in the Panofsky Auditorium, to begin with the story of my first involvment with neutral K mesons. It was the summer of 1956. Bob Motley and I had just finished taking data at the Brookhaven Cosmotron on the lifetimes of charged K mesons as a function of their decay mode, our own small contribution to the $\tau - \theta$ puzzle becoming a major conumdrum. Panofsky, with his family, was spending the summer at Brookhaven. He came into our lab one day and said, "Let's do an experiment." Motley and I still had data to analyze but Pief was persuasive and we decided we could sandwich some new activity into our schedule. But what experiment to do? Lederman and his group[1] had just discovered the long-lived neutral K, then known as the $\theta_2$. This clearly needed some confirmatory experiments. We decided to try to measure the total cross section of the $\theta_2$ to confirm that it was, indeed a 50-50 mixture of $\mathrm{K}^0$ and $\overline{\mathrm{K}^0}$. Their charged counterparts, the $\mathrm{K}^+$ and $\mathrm{K}^-$, have vastly different cross-sections at low energy and, if the general picture was correct, one could expect the $\theta_2$ to exhibit a cross-section halfway between, after coulomb effects had been taken into account. That is what we proceeded to do, aided by Walter Chesnut, a new post-doc at BNL. Thanks to Pief, it was my first exposure to neutral K mesons.

Our strategy was to detect, using counter techniques, the neutral Ks through their decay either to $\pi\mu\nu$ or $\pi \mathrm{e}\nu$. On both sides of the neutral beam we placed scintillators with the requirement that, on one side or the other we stop the pion and see the $\pi\mu$ e decay sequence while on the opposite side, a single charged particle in time coincidence. It is easy to say now but then, it must be realized, plastic scintillators were not off-the-shelf items – if you wanted some, you made your own. George Clark at MIT was one of the pioneers in the business. He had recently perfected the process and was making wagon wheel-size pieces (36 inches in diameter) for an air-shower experiment

*Supported by TIAA/CREF.

at Volcano ranch in New Mexico. He graciously gave us a wheel of his material and from that we made the pieces for our experiment. Pief and I spent many quality hours that summer polishing the plastic after it had been cut to size. In the end we found Colgate toothpaste to be the best final polishing agent.

We got the counters and the neutral beam constructed that summer and were given some running time in the fall. This was still before scheduling committees had been formalized. At the Cosmotron, George Collins was the director and, to a large degree, controlled the experimental program from his pocket. Of course, he did have advice from an ad hoc kitchen cabinet. Still, the whole approval process was highly informal. I don't remember ever putting together a proposal for this experiment. We just talked to George.* The results of the experiment were published in Physical Review.[2]

Aside from giving you a flavor of what doing accelerator experiments was like in those early days, what does this have to do with the discovery of CP violation? The first few paragraphs of the paper show an early awareness, at least among the cognescenti, that CP was not necesssarily to be taken as a conserved quantity. In the paper we say the following:

"The analysis of this experiment is not affected by the lack of conservation of Parity and of charge conjugation invariance, and is only slightly influenced by a possible lack of time-reversal invariance in the decay of the neutral-K complex. Even though a small admixture of 2 pi decays to the long-lived component is possible in principle, both experiment[3] and theory[4] indicate that this admixture is small."

The principle point I want to make is that, from the very beginning, the 2 $\pi$ decay of the long lived neutral kaon was held as a signature of CP and time reversal noninvariance.

In the fall of 1962 there was considerable discussion at Brookhaven about some results of the Adair group[5] which seem to show an excess of $K_1$s in a hydrogen bubble chamber exposed to a beam of the long-lived $K_2$s. I was working, at the time, at the AGS with David Cassel involved in an experiment to measure the form factor of the pion. Jim Cronin was at the Cosmotron doing an experiment on $\rho$ production. Early in 1963 we learned that, in the near future, a neutral beam was going to become vacant at the AGS. We realized that by using spark chamber techniques (we were both using spark chambers in our respective experiments) one could improve, substantially, the bubble chamber results. Bubble chambers were great for observing the production of short-lived particles in liquid hydrogen but the multiple scattering in the hydrogen and the relatively short lever arm involved in momentum measurements greatly restricted their intrinsic precision. In addition, for certain kinds of experiments, the bubble chamber technique suffered by not being able to acquire target-empty data – necessary to

*Shortly afterward, review committees for experiments were established and the approval process became formalized.

ascertain the background effects of the material in the chamber walls. Furthermore, much higher beam intensies could be tolerated by spark chambers and, because of their intrinsic memory, they could be selectively triggered on interesting events.

We put together a modest proposal on short notice.[6] In this two page document we stated four objectives for the experiment: 1) to check the results of Adair, 2) to study the regeneration phenomena under a variety of conditions in different materials, 3) to set new limits on the decay of the long-lived neutral K to two pions, and 4) to check for the presence of neutral currents in strangeness changing decays. We finished the experiments in which we were previously engaged, moved the apparatus into the beam by late May, 1963, and took data that summer. The rest is history.

It is difficult to visualize the conditions of those early experiments at the AGS from the perspective of recent times. Most of the electronics was designed by us and home-made. At this time it was a hybrid of vacuum tube and discrete transistor circuits. The triggering and recording apparatus was on the main floor of the AGS beside the beam in which the chambers were placed. It was a hot and noisy environment with no regard for the comfort of the experimenters who sat at a table beside the beam with their log book. Targetting the beam in the AGS was internal in those early years − the circulating proton beam intensity was about a thousand times less than that today. Correspondingly, we were largely responsible for our own actions with regard to radiation hazards. To quote the beginning of a famous novel: "It was the best of times. It was the worst of times."

The details of the experiment have been described many times[7,6] and I will not repeat them here. Suffice it to say that, previous to our work, the best limit on the decay of K long to 2 $\pi$ was from a Soviet group[8] who analyzed 597 decays and found no candidates fitting the 2 $\pi$ decay. We found 48 events in 24,000 decays, about 1 in 500. What about the original Adair results? If they had been correct we would have seen about 15 times as many $2\pi$ events in our hydrogen target data than we did, in fact, observe. The Adair experiment was repeated using a bubble chamber with a .07 cm wall instead of the original 1.6 cm. It was shown[9] that the original effect came, to a large extent, from the walls (no empty target data ??).

The original CP paper was published in July of 1964.[10] Almost immediately, within a couple of weeks, theoretical papers began to appear. They were in two categories: those that accepted the results as demonstrating CP (and T) violation and those that attempted to explain the phenomena in terms of a new long-range interaction, as Adair *et al* had done earlier.

The paper of Wu and Yang[11] has endured as a benchmark. It laid the grounds for a phenomenological analysis of the experiment on the basis of CP violation. It has provided a foundation for the interpretation of every succeeding experiment. Aside from factors of 2 the notation and definitions introduced in this paper have remained the same.

Truong[12] had reasons to suggest that the effect might be completely in the $\Delta$ I = 3/2 decay amplitude.

Wolfenstein[13] proposed a new $\Delta$ S = 2 superweak interaction which has existed as a possibility until the recent results which show $\epsilon'$ to be different from zero.[14]

Other papers at the time[15,16] proposed a new particle, a hyperphoton, the existence of which would avoid the necessity of CP and T violation. However, Weinberg gave[17] strong arguments that excluded this possibility. To my knowledge, Weinberg was the first to use the term, "fifth force". It is interesting to note that when the issue was raised more than 20 years later by Fischbach and collaborators[18] the earlier work had been forgotten and was never referred to.

In any event, the first follow-on experiments demonstrating the interference between the CP violating amplitude and the amplitude for coherent regeneration[19] completely ruled out the possibility of any alternate explanation for the CP violation based on a new particle, such as a hyperphoton. Progress in experimentally quantifying the parameters involved in CP violation was initially slow and painful. Eventually, however, a number of remarkable experiments on the neutral K system using increasingly sophisticated and elaborate instruments have pinned down the parameters, originated by Wu and Yang, with remarkable precision. They are tabulated in the PDB. These experiments have culminated in the recent results[14] for $\epsilon'/\epsilon$. Tran Truong was a little bit right in 1964.

A.D. Sakharov,[20] very early, showed that CP violation combined with baryon nonconservation and nonequilibrium dynamics (easily provided in the early times of the big bang) could account for the matter-antimatter asymmetry in the universe. Indeed, one might turn the question around and say that the first evidence, ever, for CP violation was the fact that we exist.

CP violation now has a natural home in the standard model. And exciting times lie ahead. BABAR at SLAC and GEM at KEK promise to explore CP violation in a totally new regime.

# References


[1] K. Lande, E.T. Booth, J. Impeduglia, L.M. Lederman, W. Chinowsky, Phys. Rev. **103** 1901 (1956).

[2] W.K.H. Panofsky, V.L. Fitch, R.M. Motley, and W. Chesnut, Phys. Rev. **109** 1353 (1964).

[3] K. Lande, L.M. Lederman, and W. Chinowsky, Phys. Rev. **105** 1925 (1957).

[4] T.D. Lee, R. Oehme, and C.N. Yang, Phys. Rev. **106**, 340 (1957).

[5] L.B. Liepuner, W. Chinowsky, R. Crittenden, R. Adair, B. Musgrave, and F.T. Shively, Phys. Rev. **132** 2284 (1963).

[6] M. Doncel, A. Hermann, L. Michel, and A. Pais, 1983, Eds. *Symmetries in Physics* (Universitat Autonoma de Barcelona, [distributed by World Scientific Publishing]), pp. 453-471.

[7] V.L. Fitch, Rev. Mod. Phys. **53** 367 (1981).

[8] M.K. Anikina, D.V. Neagu, E.O. Okunov, N.I. Petrov, A.M. Rozanova and V.A. Rusakov, JETP **15** (1962).

[9] C.J.B. Hawkins, Phys. Rev. **156** 1444 (1967).

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

[11] T.T. Wu and C.N. Yang, Phys. Rev. Lett. **13** 380 (1964).

[12] T. Truong, Phys. Rev. Lett. **13** 358 (1964).

[13] L. Wolfenstein, Phys. Rev. Lett. **13** 569 (1964).

[14] A. Alavi-Harati *et al.*, Phys. Rev. Lett. **83** 22-27 (1999); G.D. Barr *et al.*, Phys. Lett. B **317** 233 (1993).

[15] J.S. Bell and J.K. Perring, Phys. Rev. Lett. **13** 349 (1964).

[16] J. Bernstein, N. Cabibbo and T.D. Lee, Phys. Lett. **12** 146 (1964).

[17] S. Weinberg, Phys. Rev. Lett. **13** 562 (1964).

[18] S. Fischbach, S. Sudarsky, A. Szafer, C. Talmadge, and S.H. Aaronson, Phys. Rev. Lett. **56** 3 (1986).

[19] V.L. Fitch, R. Roth, J. Russ, and W. Vernon, Phys. Rev. Lett. **15** 73 (1965).

[20] A.D. Sakharov, JETP Lett. **5** 24 (1967).

# CP VIOLATION AND THREE GENERATIONS

Makoto Kobayashi
KEK
1-1 Oho, Tsukuba, Ibaraki, Japan

ABSTRACT

The development of the notion of flavor before the discovery of the third generation is briefly reviewed. Then the basic mechanism of CP violation in the six-quark model is explained. In particular, phase factors in the CP transformation and the phase redefinition of field operators and their role in CP violation are discussed. Representations of the flavor mixing matrix are also discussed in some detail and an exactly unitary Wolfenstein-type parameterization is given. Finally, a few remarks are made on the present status of the study of CP violation.

# 1 Introduction

CP violation is a very fundamental problem that may be represented by a simple question: “What is the essential difference between particles and antiparticles?” The beginning of the problem goes back to around 1930, when Dirac derived the Dirac equation, from which the existence of the antiparticles for spin-one-half particles was concluded. Obviously, on a macroscopic scale, apparent asymmetry exists between particles and antiparticles, or matter and anti-matter. This must have been a big mystery for the people of that time. In fact, Pauli provided a skeptical view about the interpretation of the Dirac equation in Handbuch der Physik,[1] which was probably written in 1932.

It seems, however, that this mystery was forgotten in the subsequent development of particle physics. Particles and antiparticles had been believed to be symmetrical in the microscopic world until 1964, when CP violation was first discovered in $\mathrm{K}_L \to \pi + \pi$ decay.[2] Parity violation was discovered almost a decade before the discovery of CP violation, and it was also found that charge conjugation invariance is violated. However, at that time it was considered that the nature is invariant under the combined transformation CP. Violation of C implies asymmetry between particles and antiparticles, but if CP is conserved, the difference between particles and antiparticles can be resolved by looking at them through a mirror. So C violation in this case is not an essential difference which creates the asymmetry between matter and antimatter in the universe.

After the discovery of CP violation, Sakharov discussed the necessary conditions for the generation of the matter-dominant universe.[3] Actually, C and CP violations are one of the three necessary conditions, together with the baryon number nonconservation and the nonequiribrium universe. This issue was revived by Yoshimura[4] and Ignatiev, Krasnikov, Kuzmin and Tavkhelidze[5] in connection with grand unified theories some years later.

As for the theoretical understanding of CP violation in the neutral kaon system, it remained at a phenomenological level until the beginning of the 1970’s. Although the attractive idea of the super weak model[6] emerged from these phenomenological analyses, it was not easy to develop theoretical ideas given the limited amount of experimental information that was available. In early 1970’s, however, the development of a renormalizable gauge theory of the weak interactions changed the situation. The requirements of gauge invariance and renormalizability reduced the degrees of arbitrariness in the theoretical understanding of CP violation. Furthermore, subsequent experimental discoveries of new flavors gave a special position to the six-quark model of CP violation.[7]

In the six-quark model, CP violation arises from flavor-mixing. Since flavor-mixing is an important subject of particle physics in its own right, we review the development of the notion of the flavor mixing in the next section. The basic mechanism of CP violation in the six-quark model is discussed in Section 3.

## 2 History of Flavor Mixing

Following the suggestion of Lee and Yang,[8] and the subsequent discovery by Wu[9] that parity was violated, our understanding of the weak interactions developed rapidly. After some initial confusion about the type of interaction at play in $\beta$ decay, the structure of the weak interactions was established as a V-A type interaction.

In this course, the universality of the strength of the weak interactions had already been recognized, in particular in the similarity of the strength of the p-n transitions and muon decay. This fact lead Feynman and Gell-Mann[10] to propose the conserved vector current hypothesis: If the vector part of the weak current is the charged component of an isomultiplet electro-magnetic current, then its conservation guarantees the non-renormalization of the vector coupling of the p-n transitions.

It turned out, however, that there is slight discrepancy between the vector coupling of p-n transition and the muon-decay. Namely, the strength of the p-n transition is a few percent weaker than that for muon-decay. In addition, it was also observed that the coupling constants of the strangeness-changing weak interactions are considerably smaller than their strangeness non-changing counterparts.

In these circumstances, Gell-Mann and Levy[11] noted that the weak current could take the structure of the following form;

$$GV_\alpha + GV_\alpha^{(\Delta S=1)} = \frac{G_\mu}{(1+\epsilon^2)^{\frac{1}{2}}}\bar{p}\gamma_\alpha(n+\epsilon\Lambda) + \cdots .$$

This observation was made in a note-added-in-proof of their famous 1960 paper on the $\sigma$-model.

In 1963, Cabibbo gave a new notion of the universality of the weak interaction by generalizing the isospin current to the SU(3) current:[12]

$$J_\mu = \cos\theta(j_\mu^{(0)} + g_\mu^{(0)}) + \sin\theta(j_\mu^{(1)} + g_\mu^{(1)}),$$

where the superscripts denote $\Delta S$. The mixing angle $\theta$ is now known as the Cabibbo angle. Then the algebraic structure of the weak current began to attract attention, and it eventually lead to the gauge theory of weak interaction.

The current structure of the weak interaction and the success of current algebra are most easily understood in the quark model. However, there was difficulty in understanding the weak interaction field theoretically as the interaction between quarks and intermediate bosons. The difficulty was related to the strangeness changing neutral current, which had to be suppressed at the tree level as well as at an induced current.

In 1970, Glashow, Illiopoulos and Maiani[13] pointed out that this difficulty could be avoided by the introduction of a fourth quark and an extended weak current that is now known as the GIM current. This observation paved the way to the gauge theory of the weak interactions.

What we have seen so far is the main stream history as experienced in the western world. It is interesting to note that there was another stream in the story of the flavor mixing, one from which the present author personally received strong influence.

In 1956, Sakata[14] proposed a model, which we now call the Sakata model, to explain the existence of the many types of strange particles. His claim was that the triplet of the proton, the neutron and the $\Lambda$ particle are the fundamental objects and that all the hadrons are composite states of them.

In addition to reproducing the strange particle spectrum, the Sakata model explained the weak interaction as well. In particular, the $\Delta S = \Delta Q$ rule is quite nicely explained by considering that strangeness-changing processes take place via a $\Lambda$-p transition. In 1959, Gamba, Marshak and Okubo[15] pointed out the following correspondence between p, n, $\Lambda$ triplet and the leptons;

$$\begin{array}{ccc} \mathrm{p} & \mathrm{n} & \Lambda \\ | & | & | \\ \nu & \mathrm{e} & \mu \end{array} .$$

This observation prompted Sakata and his group to propose so-called Nagoya model, in which the triplet baryons are considered to be made of the leptons and an object called B-matter:[16]

$$\mathrm{p} = (\nu \mathrm{B}),\ \mathrm{n} = (e\mathrm{B}),\ \Lambda = (\mu \mathrm{B}).$$

In 1962, Lederman's group revealed that the neutrinos associated with the muon are different from the electron neutrino.[17] When this news reached Japan, two Japanese groups immediately proposed that the proton corresponds to a mixed state of $\nu_{\mathrm{e}}$ and $\nu_\mu$

and therefore the couplings of p-n and p-$\Lambda$ are suppressed by a mixing angle.[18] Furthermore, they pointed out the possible existence of the fourth baryon p' corresponding to the orthogonal two neutrino combination:

$$\begin{aligned} \mathrm{p} &= (\nu_1 \mathrm{B}), \qquad \nu_1 = \cos\theta\, \nu_{\mathrm{e}} + \sin\theta\, \nu_\mu \,, \\ \mathrm{n} &= (\mathrm{eB}), \\ \Lambda &= (\mu \mathrm{B}), \\ \mathrm{p}' &= (\nu_2 \mathrm{B}), \qquad \nu_2 = -\sin\theta\, \nu_{\mathrm{e}} + \cos\theta\, \nu_\mu \,. \end{aligned}$$

This scheme is called the extended Nagoya model.

It should be noted that these papers were published in 1962, even before Cabibbo's paper. This work was developed from a very different view points from that which prevailed in the western world. The Sakata model considered the actual p, n, $\Lambda$ particles as the fundamental triplet and the SU(3) or U(3) symmetry was regarded as an approximate symmetry of the dynamics of those fundamental particles. In contrast, in the western world, the baryons are considered to belong to an octet representation, with p, n, $\Lambda$ being parts of the octet, instead of a triplet, and, at least in the early stages, SU(3) symmetry was considered as something with a more abstract character.

Eventually the octet scheme of the baryons was confirmed experimentally, and two approaches were merged, in some sense. In 1964, the quark model was proposed, although it took several years before most people accepted the actual existence of quarks. On the other hand, in order to reconcile it with the octet nature of the baryons, the Nagoya model was reinterpreted in terms of a new layer of fundamental particles that were distinct from the actual p, n and $\Lambda$ baryons.

The idea of Nagoya model was discussed by the people of Sakata's group from time to time. However, these prescient works of the Japanese groups were not developed in the direction of the gauge theory of the weak interactions.

Meanwhile, in 1971, there was unexpected development. In this year, Niu and his group found a few new kind of events in the emulsion chambers exposed to cosmic rays.[19]

One of those events had two distinct kinks, which suggested the pair production of short-lived unstable particles, and a pair of gamma ray shower which is consistent with $\pi^0$ production. The direction of $\pi^0$ is pointing one of the kinks and energy of $\pi^0$ can be estimated from the shower development. Assuming a two body decay process, they estimated the life time of the unstable particle as around $10^{-14}$ sec. If the charged decay product is the pion or the kaon, then the mass of the parent particle is around 2 GeV.

Hearing this news, Shuzo Ogawa of Hiroshima University (at that time) immediately pointed out that those events could be the productions of new particles which contain the fourth fundamental particle, i.e. the p'. When Ogawa proposed this, he clearly had the extended Nagoya model in his mind.[20]

Following Ogawa's suggestion, several groups in Japan began to study Niu's events and also various general aspects of the quartet models. At the time, I was a graduate student of Nagoya University. I was one of those who are working on the study of Niu's events and became familiar with the quartet model. The rest of this section are my own personal recollections.

In the same year, the Weinberg-Salam-Glashow model of the electroweak interactions began to attract wide attentions, especially because the nonabelian gauge theories were proved to be renormalizable by 'tHooft. In order to extend this type of theory to the quark sector, it was necessary to consider GIM-type quartet models. This might have been thought of as a drawback of the model, but, at least to me, everything looked to be neatly placed.

Actually, I thought that it might be possible to describe everything field theoretically along this line. Of course, this was before the discovery of the asymptotic freedom and, therefore, prior to the notion of quark confinement. Therefore, I could not pin down a specific interaction for the strong interactions, but I thought that the strong interactions were not incompatible with the field theoretical description. In that case, what was left was CP violating interactions.

After I received my Ph.D. from Nagoya University in 1972, I moved to Kyoto University where I started to investigate this problem with Maskawa, who was a research associate of Kyoto University at the time.

The problem was how to accommodate CP violation in the model in a gauge invariant and renormalizable manner. Once the problem was set, however, it was rather straightforward to solve. Rather than to propose a particular model, we tried to seek the logical consequences as much as possible. We quickly realized that the minimal quartet model scheme could not accommodate CP violation and we needed to add new particles. The six-quark model was pointed out as one of such possibilities.

## 3 Six-Quark Model

In this section, we discuss the mechanism of CP violation in the six-quark model. We will start with some very basic observations; one on the phase factor of the CP transformation and the other on the phase redefinition of field operators. Then, the way the six-quark model accommodates CP violations will be explained, and, finally, various representations of the quark mixing matrix will be discussed.

## 3.1 The Phase Factor of the CP Transformation

Let us consider the CP transformation of a certain complex field $\phi$. The CP conjugate of $\phi$ is given by hermitian conjugate $\phi^*$, which, in general, could be multiplied by an arbitrary phase factor:

$$U_{\rm cp}\,\phi\,U_{\rm cp}^{-1} = e^{i\alpha}\,\phi^*.$$

Taking the hermitian conjugate of this relation, we have

$$U_{\rm cp}\,\phi^*\,U_{\rm cp}^{-1} = e^{-i\alpha}\,\phi,$$

from which we can confirm that, when we apply $U_{\rm cp}$ twice, it yields the identity transformation, even if the phase factor has an arbitrary value:

$$U_{\rm cp}^2 = 1.$$

The existence of an arbitrary phase factor implies that there are infinitely many candidates of CP transformation. The question is how to fix the phase factor.

To illustrate how we can do it in the CP conserving case, we consider, as an example, the following interaction Hamiltonian,

$$H^{(1)} = \int d^3x\{g\,\phi\,O + g^*\phi^*O^*\}, \tag{1}$$

where $g$ is a coupling constant and $O$ is a certain operator. For example, in the case of Higgs type interaction, it looks like $O = \bar{\psi}\psi$. Here we assume that only $H^{(1)}$ changes the number of particles described by the field $\phi$. This means that the rest of the Hamiltonian consists of $\phi^*\phi$ combinations.

Now we apply $U_{\rm cp}$ to this interaction Hamiltonian;

$$U_{\rm cp}\,H^{(1)}\,U_{\rm cp}^{-1} = \int d^3x\{g\,e^{i\alpha}\phi^*O^* + g^*e^{-i\alpha}\phi\,O\},$$

where we have assumed that the CP transformation of the operator $O$ is fixed somewhere else, and, for the purpose of simplicity, is given by the following form, without an extra phase factor;

$$U_{\rm cp}\,O\,U_{\rm cp}^{-1} = O^*.$$

We note that, in general, $O$ contains some other fields than $\phi$ and the CP transformation applied to them may have arbitrary phase factors like $\alpha$. If they exist, in order to fix them we have to repeat similar arguments. How it ends up is an interesting question that is left as an exercise.

Now we find that if we choose the phase factor as

$$\alpha = -2\text{arg.}(g),$$

then the Hamiltonian is invariant under such a $U_{\text{cp}}$. This means that the phase factor is fixed by the coupling constant. In other words, the Hamiltonian selects the CP transformation among infinitely many candidates. Only the selected CP transformation is the symmetry of the system, and the other candidates are not symmetries of the system. Note that for the system to be CP symmetric, we need only one CP transformation under which the system is invariant. Note also that a complex coupling constant does not necessarily imply CP violation. The phase of $g_1$ is made harmless by absorbing it into the phase factor $\alpha$ of the CP transformation.

So far we have been considering the CP conserving case. Now we turn to CP violating cases. A simple example of CP violating systems is the following:

$$H^{(2)} = \int d^3x \{g_1 \phi \, O_1 + g_1^* \phi^* O_1^* + g_2 \phi \, O_2 + g_2^* \phi^* O_2^*\}. \tag{2}$$

The difference from the previous example is that the interaction consists of two terms, or two coupling constants.

Now we consider the CP transformation of $H^{(2)}$;

$$\begin{aligned} U_{\text{cp}} \, H^{(2)} \, U_{\text{cp}}^{-1} &= \int d^3x \{g_1 e^{i\alpha} \phi^* O_1^* + g_1^* e^{-i\alpha} \phi \, O_1\} \\ &\quad + \int d^3x \{g_2 e^{i\alpha} \phi^* O_2^* + g_2^* e^{-i\alpha} \phi \, O_2\}, \end{aligned}$$

where we have assumed again that the CP transformations of $O_1$ and $O_2$ are given by their hermitian conjugates without any extra phase factors.

In this case, however, if the two coupling constants have different phases, we can not make the Hamiltonian invariant by choosing the phase factor $\alpha$ properly. If we choose $\alpha$ so that the first interaction is invariant, then the second interaction is not invariant, and vice versa. Whatever $\alpha$ we choose, $U_{\text{cp}}$ is not a symmetry of the system. This is a typical mechanism for CP violation.

The following can be said in general. In CP violating cases, Hamiltonians do not single out $\alpha$. None of the candidate CP transformations can be a symmetry of the system. Therefore, we have no unique CP transformation in CP-violating cases. This means that, when looking at a single interaction, we can not say whether that particular interaction violates CP symmetry or not, because the CP transformation itself is not defined uniquely.

Actually we need not fix $\alpha$. The usual experimental measures of CP violation are defined without reference to the CP transformation. The origin of CP violation is a mismatch of phases of two or more coupling constants, which prevents the singling out of the value of $\alpha$. Experimentally measurable quantities are related to the relative phases of the coupling constants.

## 3.2 Phase Redefinition

Next we discuss the phase convention of complex fields. Although this is closely related to the CP transformation phase factor, it is still an independent notion.

Let us consider the following phase redefinition of a complex field $\phi$;

$$\phi = e^{i\theta}\tilde{\phi}.$$

We consider again $H^{(1)}$ as an example. If we rewrite the total Hamiltonian in terms of $\tilde{\phi}$, the Hamiltonian takes the same form as the original one except for a slight change of the coupling constants $g$ in $H^{(1)}$;

$$\begin{aligned} H^{(1)} &= \int d^3x\{g\,\phi\,O + g^*\phi^*O^*\} \\ &= \int d^3x\{\tilde{g}\,\tilde{\phi}\,O + \tilde{g}^*\tilde{\phi}^*O^*\}, \end{aligned}$$

where

$$\tilde{g} = e^{i\theta}g.$$

Since we are assuming that the rest of the Hamiltonian consists of $\phi^*\phi$, there is no change in them. Only the coupling constants of $\phi$ number changing interactions obey the phase change. Under this phase redefinition, CP transformations with a phase factor changes as

$$U_{\rm cp}\,\tilde{\phi}\,U_{\rm cp}^{-1} = e^{i\alpha}e^{-2i\theta}\tilde{\phi}^*.$$

The new field $\tilde{\phi}$ has the same ability as $\phi$ to describe the particle, and, if we use $\tilde{g}$ instead of $g$, the physics is the same as the original Hamiltonian. Therefore, this phase redefinition is a matter of convention.

Here let us choose $\theta$ as

$$\theta = -\arg.(g),$$

so that $\tilde{g}$ is real;

$$\tilde{g} = e^{i\theta} g \ : \ \text{real} \, .$$

Then under the CP-conserving transformation in which $\alpha = -2\arg.(g)$, the phase factor disappears for $\tilde{\phi}$;

$$U_{\rm cp} \, \tilde{\phi} \, U_{\rm cp}^{-1} = \tilde{\phi}^* .$$

What we learn from this simple example is that when CP is conserved, the coupling constants and phase factors of the CP transformation can be made real using the phase redefinition of the field operators.

In the CP-violating case, however, we can not do this. In the Hamiltonian (2), we can not make both coupling constants real at the same time by choosing phase convention of $\phi$, as long as the two coupling constants have different phases. What we can do is to make one or the other of the coupling constants real.

When $g_1$ is made real, it is sometimes said that the imaginary part of $g_2$ violates CP invariance. But this is not a correct statement, because we did not fix the phase factor of the CP transformation in CP violating case, so that CP transformation itself is not defined uniquely. Since the CP violation arises from relation among various interactions, it is not adequate to say that a particular interaction is CP conserving or not.

For practical purposes, however, it may be convenient to make the coupling constant of the stronger interaction real. For example, if $g_1$ is much larger than $g_2$, then, by making $g_1$ real, we may treat the imaginary part of $g_2$ as the origin of CP violation for some purpose. We must be careful in applying this argument, however, because, in general, the strength of the interaction can not be compared simply by the coupling constants.

In conclusion, we note again that observable CP violating quantities are defined so that they are independent of a particular choice of the phase factor $\alpha$ and also independent of the phase convention of the field operators.

## 3.3 Quark Mixing

On the basis of the observations made in the previous subsections, we consider the mechanism of CP violation in the standard model. Since all the coupling constants of the gauge interactions are real numbers, there exists a chance of complex coupling constants only in the quark mixing which appears in the charged current weak interactions.

We note that there is another possibility for CP violation arising from a non-zero value of the $\theta$ parameter of the QCD vacuum. This is a very different mechanism for CP violation and it is usually thought that $\theta$ is very small, even if it is not zero. So we do not consider this type of CP violation in the following.

The problem we are going to consider is whether the complex parameters appearing in the quark mixing violate CP or not. Let us consider the case where the quark sector consists of $n$-generations of left-handed doublets and all the right-handed quarks are singlets:

$$\underbrace{\begin{pmatrix} u \\ d' \end{pmatrix}_L, \begin{pmatrix} c \\ s' \end{pmatrix}_L, \begin{pmatrix} t \\ b' \end{pmatrix}_L, \cdots .}_{\longleftarrow \quad n \quad \longrightarrow}$$

Here $u, c$ and $t$ denote the mass eigenstates of the u-type quarks, while $d', s'$ and $b'$, which are the doublet partners of the u-type quarks, are not necessarily mass eigenstates. Instead, they can be linear combinations of the d-type mass eigenstates:

$$\begin{pmatrix} d' \\ s' \\ b' \\ \vdots \end{pmatrix} = V \begin{pmatrix} d \\ s \\ b \\ \vdots \end{pmatrix},$$

where $V$ is an $n \times n$ mixing matrix. In this case the charged current weak interactions look like

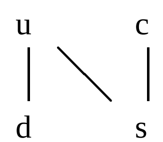
$$H_{\rm cc} = \frac{g}{2\sqrt{2}} W_\mu^+ (\bar{u}, \bar{c}, \bar{t}, \cdots) \gamma^\mu (1 - \gamma^5)\, V \begin{pmatrix} d \\ s \\ b \\ \vdots \end{pmatrix} + h.c. \,.$$

Note that the coupling constants for $\bar{q}qW$ interactions are given by $g$ times the matrix elements of $V$.

In order that the kinetic parts of the quarks are flavor diagonal and properly normalized, the mixing matrix $V$ must be unitary. Since, in general, the elements of a unitary matrix are complex numbers, the charged current weak interactions are a possible source of CP violation.

The difference between the number of parameters of an $n \times n$ unitary matrix and those of an orthogonal matrix, $n^2 - n(n-1)/2 = n(n+1)/2$ can be regarded as the number of parameters arising from the complexness of the unitary matrix. According to the previous discussion, we have to investigate whether these complex numbers can be removed by the phase redefinition of the quark fields. The number of quark fields is $2n$, but the overall phase will have nothing to do with the matrix elements of $V$, because the interaction has the form of $\bar{q}Vq$. Therefore the number of those phases that can be used to absorb the phase factor of the matrix elements of $V$ are $2n-1$. Subtracting $2n-1$ from $n(n+1)/2$, we have $(n-1)(n-2)/2$ as the number of unremovable phase parameters. Therefore, if $n=2$, complex coupling constants would be removed, whereas for $n=3$ they are not. Now we confirm this explicitly.

First we consider the case of two generations. Here we make the phase redefinition in the following way: First we adjust the relative phase of the u and d quark fields so that the corresponding u-d transition is described by a positive real matrix element, i.e. $V_{ud}$ is real and positive. Then we adjust the phase of the s quark field so that $V_{us}$ is real and positive. Finally the phase of the c quark field can be adjusted to make $V_{cs}$ real and positive. Thus, with the phase redefinition for these quark fields, the three matrix elements of $V$ are made real and positive as follows:




$$\Rightarrow \qquad V = \begin{pmatrix} R & \underline{R} \\ * & R \end{pmatrix} .$$

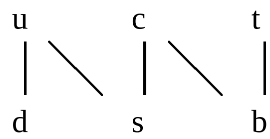
This shows that the matrix elements corresponding to the transitions connected by the lines are positive real numbers. The $R$ symbols in the matrix imply elements that are positive real numbers and the asterisk indicates an element that could, so far, be a complex number. We see, however, that this element must be real as a result of the unitarity relation for the elements of $V$.

There is only one independent parameter and it can be chosen as $V_{us}$, indicated by $\underline{R}$ in the above expression. Parameterizing it by an angle variable, we have the familiar GIM form for $V$;

$$V = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix},$$

where $\theta$ lies in the first quadrant because $\underline{R}$ is positive by the phase convention.

Next we consider the case of three generations. Following a similar method to the previous case, we adjust the phase convention of the quark fields to remove the complex matrix elements as far as possible. Our choice here is such that five relative phases indicated by the following diagram are adjusted so that the corresponding matrix elements are real and positive:




Then the mixing matrix $V$ looks like

$$V = \begin{pmatrix} R & \underline{R} & \underline{*} \\ * & R & \underline{R} \\ * & * & R \end{pmatrix}, \tag{3}$$

where $R$ indicates elements that are real positive numbers and asterisks those that could be complex numbers. From the general structure of the unitary matrix, we can see that, when the diagonal elements are real, a unitary matrix can be parameterized by the elements of the upper-right small triangle indicated by the underlines. Namely the rest of elements can be expressed in terms of the underlined elements by using the unitarity relations and the reality of the diagonal elements.

Let us parameterize the triangle in the following way, as suggested by Wolfenstein:[21]

$$V = \begin{pmatrix} * & \lambda & A\lambda^3(\rho - i\eta) \\ * & * & A\lambda^2 \\ * & * & * \end{pmatrix}.$$

It should be noted that this is not an approximate expression for a small $\lambda$, but a definition of the four parameters, $\lambda$, $A$, $\rho$ and $\eta$. We have three real parameters and one corresponding to the imaginary part, as we expect from the general argument given above.

The rest of matrix elements can be expressed in terms of these four parameters in an exactly unitary manner.[22] The following is the result of such expression:

$$V = \begin{pmatrix} D_1 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda \dfrac{D_2 + A^2\lambda^4(\rho + i\eta)}{D_1} & D_2 & A\lambda^2 \\ A\lambda^3 \dfrac{D_2 - (\rho + i\eta)D_0^2}{D_1 D_3} & -A\lambda^2 \dfrac{D_2 + \lambda^2(\rho + i\eta)}{D_3} & D_3 \end{pmatrix},$$

where

$$\begin{aligned} D_1 &= \sqrt{1 - \lambda^2 - A^2\lambda^6(\rho^2 + \eta^2)}, \\ D_2 &= \frac{-A^2\lambda^6\rho + \sqrt{D_1^2 D_3^2 - A^4\lambda^{12}\eta^2}}{1 - A^2\lambda^6(\rho^2 + \eta^2)}, \end{aligned}$$

$$
\begin{aligned}
D_3 &= \sqrt{1 - A^2\lambda^4 - A^2\lambda^6(\rho^2+\eta^2)}, \\
D_0 &= \sqrt{1 - \lambda^2 - A^2\lambda^4 - A^2\lambda^6(\rho^2+\eta^2)}.
\end{aligned}
$$

The result is slightly complicated but the matrix elements can be expressed in terms of the elementary functions and the derivation is quite straight forward. If we expand these expressions with respect to $\lambda$ we find the familiar expression given by Wolfenstein:

$$
V = \begin{pmatrix} 1-\lambda^2/2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1-\lambda^2/2 & A\lambda^2 \\ A\lambda^3(1-\rho-i\eta) & -A\lambda^2 & 1 \end{pmatrix}.
$$

When $\lambda$ is small, this expression is approximately unitary. We can easily obtain a better approximation for the unitarity from the above exact unitary expression.

Now a few technical remarks are in order. The first concerns the phase convention. So far we have been considering the particular phase convention in which real elements lie in a special pattern. Of course there are other possible phase conventions, some of which will be discussed below. It should be remembered that the physics consequences are the same for any phase convention.

In the original paper with Maskawa, we took another phase convention, in which the first line and the first row are made real, and the mixing matrix is parameterized in the following way;

$$
V = \begin{pmatrix} c_1 & -s_1c_3 & -s_1s_3 \\ s_1c_2 & c_1c_2c_3 - s_2s_3e^{i\delta} & c_1c_2s_3 + s_2c_3e^{i\delta} \\ s_1s_2 & c_1s_2c_3 + s_2s_3e^{i\delta} & c_1s_2s_3 - c_2c_3e^{i\delta} \end{pmatrix},
$$

$$
s_i = \sin\theta_i, \;\; c_i = \cos\theta_i, \;\; i = 1, 2, 3\ .
$$

Although this is one of the natural choices of the phase convention, it is not necessarily convenient in practical applications.

We note that the Particle Data Group is adopting the following representation as a standard one.

$$
V_{PDG} = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta_{13}} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta_{13}} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta_{13}} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta_{13}} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta_{13}} & c_{23}c_{13} \end{pmatrix},
$$

$$
s_{ij} = \sin\theta_{ij}, \;\; c_{ij} = \cos\theta_{ij}, \;\; i, j = 1, 2, 3\ .
$$

This was originally proposed by Chau and Wang[23] and given by successive rotations of two generations. In this representation, the real elements are located as follows:

$$V_{PDG} = \begin{pmatrix} R & R & * \\ * & * & R \\ * & * & R \end{pmatrix}$$

and the elements indicated by the asterisks are actually complex if $\delta_{13}$ is nontrivial. The number of real matrix elements are four and, in contrast to Eq.3, the central element $V_{cs}$ is not real, Therefore this is a different phase convention from the previous one.

Since in $V_{PDG}$, the $V_{us}$ and $V_{cb}$ elements are real and $V_{ub}$ is complex, we can use this representation to define a set of Wolfenstein parameters:

$$V_{PDG} = \begin{pmatrix} * & \hat{\lambda} & \hat{A}\hat{\lambda}^3(\hat{\rho} - i\hat{\eta}) \\ * & * & \hat{A}\hat{\lambda}^2 \\ * & * & * \end{pmatrix},$$

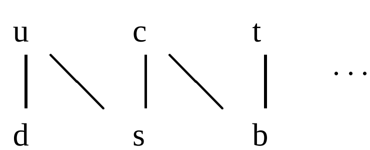
Now we have two sets of Wofenstein parameters defined in two different phase conventions. We should note that although they are defined with the same matrix elements, they are different parameter sets and their values are actually different. The relation between the two sets of parameters is

$$\begin{aligned} \hat{\lambda} &= \lambda, \\ \hat{A} &= A, \\ \hat{\rho} + i\hat{\eta} &= e^{i\xi}(\rho + i\eta), \end{aligned}$$

$$\xi = O(\lambda^6).$$

Namely the values of $\hat{\rho}$ and $\hat{\eta}$ are different from $\rho$ and $\eta$ by a small rotation in the $\rho - \eta$ plane. In this case. the difference is quite small and it is not so important for the purpose of the phenomenological analysis at the present level of experimental accuracy. Nevertheless, this is a good lesson for understanding the role of the phase convention.

Next we consider the generalization to four or more generations. Our previous choice of phase convention can be easily generalized to any number of generations:



In this case we have the following form for the mixing matrix;

$$V = \begin{pmatrix} R & \underline{R} & \underline{*} & \underline{*} & \cdot \\ * & R & \underline{R} & \underline{*} & \cdot \\ * & * & R & \underline{R} & \cdot \\ * & * & * & R & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix},$$

where the notations are the same as for the previous cases. The general pattern can be read off from this expression. Namely, the diagonal elements and their right-hand neighbors are real and positive. Independent parameters can be chosen as indicated by the underlines. From this pattern, we can easily count the number of the independent parameters as follows;

$$\begin{aligned} \text{real part} &: \frac{n(n-1)}{2}, \\ \text{imaginary part} &: \frac{(n-1)(n-2)}{2}. \end{aligned}$$

# 4 Remarks

This section contains a few remarks on what happened after the publication of our original paper in 1973. Of course the six-quark scheme is not an only possible mechanism of violating CP invariance and the point of our paper was that the minimal scheme based on the four-quark GIM-type model was too small to accommodate CP violation. Some new ingredients were needed. If we added some new fields, not necessarily new quarks, it was easy to violate CP. Roughly speaking, this is because the number of possible coupling constants increases with some power of the number of the fields—quadratically or maybe even faster—while the number of phases absorbed into the phase conventions of the field operators increases only linearly. This results in unremovable complex coupling constants with the increasing number of fields.

The special position of the six-quark model emerged from the subsequent experimental discoveries of the third generation particles. In 1975, the first evidence of the $\tau$ lepton was reported. It was after this report that the six-quark model began to attract attention, and a few papers that discussed the model appeared in 1976.[24,25,26] In particular, the first extensive study of the model was made in Ref.26, where the basic mechanism of CP violation in the neutral kaon system was given.

The discovery of the third generation continued. In 1977, the $\Upsilon$ particle was discovered and it soon became clear that it is the bound state of a fifth quark with charge

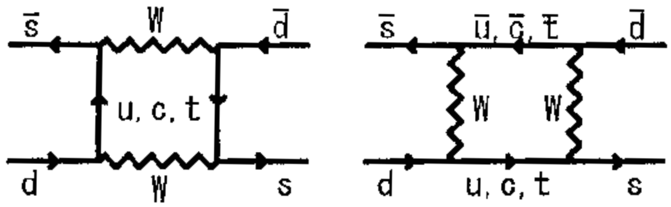



Fig. 1. Box diagrams

$-1/3$ and its antiparticle. After some detours to topless models, the existence of the t quark became a common belief. However we had to wait for long time to have an experimental evidence of the production of the t quark, until CDF group found the t quark in 1994.

Now we turn to what we know about CP violation in the six-quark model. Since this is a subject of other lectures, we will not enter into much detail here.

The present status of our understanding of the neutral K-meson system is summarized as follows. In the usual phase convention, CP violation in the $\mathrm{K} - \bar{\mathrm{K}}$ mixing is determined dominantly by the imaginary part of the dispersive part of the mass matrix of the neutral kaon system:

$$|\mathrm{K_L}\rangle = \frac{1}{\sqrt{2}}\{(1+\epsilon)\ |\mathrm{K}^0\rangle + (1-\epsilon)\ |\bar{\mathrm{K}}^0\rangle\},$$

$$|\mathrm{K_S}\rangle = \frac{1}{\sqrt{2}}\{(1+\epsilon)\ |\mathrm{K}^0\rangle - (1-\epsilon)\ |\bar{\mathrm{K}}^0\rangle\},$$

$$\epsilon \approx \frac{1}{2}\frac{i\mathrm{Im}M_{12}}{M_{12} - \frac{i}{2}\Gamma_{12}},$$

where $M_{12}$ and $\Gamma_{12}$ are the dispersive and absorptive parts of the off-diagonal elements of the mass matrix, respectively. The value of $\mathrm{Im}M_{12}$ is estimated by evaluating the contribution of the box diagram, and then we have

$$\begin{aligned} |\epsilon| \quad \approx \quad & \frac{G_F^2 m_W^2 B_K F_K^2 m_K}{6\sqrt{2}\pi^2 \Delta m}\eta A^2\lambda^6[-E(x_c)\eta_c \\ & + E(x_c, x_t)\eta_{ct} + A^2\lambda^4(1-\rho)E(x_t)\eta_t], \end{aligned}$$

where

$$x_c = m_c^2/m_W^2,\ x_t = m_t^2/m_W^2,$$

and $\eta_c$, $\eta_{ct}$ and $\eta_t$ are QCD correction factors of $O(1)$.

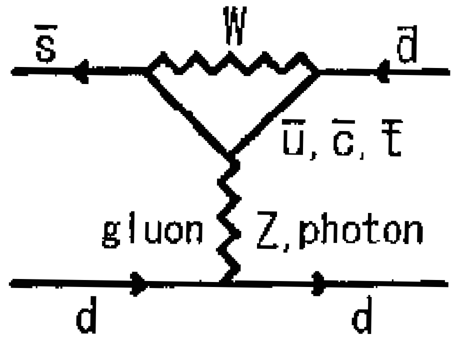



Fig. 2. Penguin diagram

The parameter $B_K$, which represents the bound state effect of the kaon, is still under the study, in particular by lattice QCD calculations. Nevertheless, a comparison of the above expression with the experimental value of $\epsilon$ yields relatively good bounds for the quark mixing parameters, particularly $\rho$ and $\eta$.

Another observable CP violating quantity in the kaon system is the $\epsilon'$ parameter in $\mathrm{K_L} \to \pi + \pi$ decay, which is defined by

$$\begin{aligned} \eta_{+-} &= \frac{A(\mathrm{K_L} \to \pi^+\pi^-)}{A(\mathrm{K_S} \to \pi^+\pi^-)} &= \epsilon + \epsilon', \\ \eta_{00} &= \frac{A(\mathrm{K_L} \to \pi^0\pi^0)}{A(\mathrm{K_S} \to \pi^0\pi^0)} &= \epsilon - 2\epsilon'. \end{aligned}$$

Specifically, $\epsilon'$ is related to the weak phase difference of the decay amplitudes for the two final states with the isospin $I = 0$ and $2$:

$$\sqrt{2}\epsilon' \approx i e^{i(\delta_2 - \delta_0)} \mathrm{Im} \frac{A_2}{A_0},$$

where $A_0$ and $A_2$ are defined by

$$\langle (\pi\pi)_I | T | \mathrm{K}^0 \rangle = e^{i\delta_I} A_I, \qquad I = 0, 2$$

and $\delta_I$ is the phase shift of the $S$-wave $\pi\pi$ scattering for $I = 0$ and $2$. The phase difference arises from so-called Penguin diagrams, because they contribute to $A_0$ and $A_2$ differently. In particular the gluonic Penguin diagram contributes only to the $I = 0$ final state.

Considerable experimental effort has been expended measuring $\epsilon'$. Recently new results are reported from KTeV group at FNAL and NA48 group at CERN:[27][28]

$$\begin{aligned} \mathrm{Re}(\epsilon'/\epsilon) &= (28.0 \pm 4.1) \times 10^{-4} \quad (\mathrm{KTeV}), \\ \mathrm{Re}(\epsilon'/\epsilon) &= (18.5 \pm 7.3) \times 10^{-4} \quad (\mathrm{NA48}). \end{aligned}$$

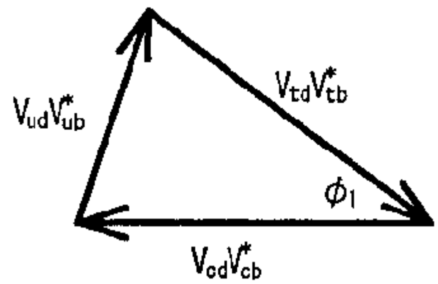



Fig. 3. Unitarity triangle

Even if we take into account these new experimental results, it seems that $\epsilon'$ does not give a stringent constraint on the mixing parameters due to ambiguities in the theoretical calculations.

The experimental value of $\epsilon$, together with the available information about the B-meson system, such as the $\mathrm{B}-\bar{\mathrm{B}}$ mixing and the charm-less decays of the B-meson, constrain the mixing parameters $\rho$ and $\eta$ to a rather small range,

$$\begin{aligned} -0.14 &< \rho < 0.32, \\ 0.23 &< \eta < 0.39. \end{aligned}$$

Existence of a consistent solution for the mixing parameters itself has a non-trivial implication on the validity of the standard model, but more importantly, a charactereistic feature of the six-quark model CP violation resides in this result.

Implications of this constraint can be expressed nicely in the form of so-called unitarity triangle, which is a triangle consisting of vectors in the complex plane corresponding to each term of the following unitarity relation;

$$V_{ud}V^*_{ub} + V_{cd}V^*_{cb} + V_{td}V^*_{tb} = 0.$$

Since all the terms are $O(\lambda^3)$, the shape of the triangle, barring the overall size, is determined by the values of $\rho$ and $\eta$. Drawing the triangle with the above range of the parameters, we note that the triangle is fairly fat. Here it is particularly important that $\eta$ is not small due to the constraint from $\epsilon$.

A fat unitarity triangle implies relatively large phase difference among the matrix elements of $V$, so that we expect fairly large CP asymmetries for certain processes. We may say that CP violations are large in the six-quark model.

Potentially large CP asymmetries are suppressed in $\mathrm{K} \to \pi + \pi$ decays by factors of order $O(10^{-3})$, but should be observable in the B-meson system. The most promising process is $\mathrm{B}(\bar{\mathrm{B}}) \rightarrow \mathrm{J}/\psi + \mathrm{K_S}$ decay in which we expect a large asymmetry in the time profile of two decays starting from $\mathrm{B}$ and $\bar{\mathrm{B}}$ at $t = 0$. The decay comprises two processes, a direct decay and a decay via $\mathrm{B} - \bar{\mathrm{B}}$ mixing. The asymmetry arises because the interference of these two processes takes place differently for $\mathrm{B}$ and $\bar{\mathrm{B}}$. A remarkable fact is that this asymmetry can be related to the angle $\phi_1$ of the above unitarity triangle without much theoretical ambiguity. Besides this "golden" mode, the B-meson system has many other channels in which we can expect large CP asymmetry.

As we have seen, the six-quark model can explain CP violation in the neutral kaon system. But, explaining one CP-violating phenomenon with an adjustable CP violating parameter is not enough to test the model. Still there is a possibility that the quark-mixing parameters are quite different and the CP violation comes from very different origins. Definitely we need a further test of the model. It is particularly important to test the above mentioned characteristic features in the B-meson system.

For this reason, CP violations in the B meson system are currently undergoing extensive experimental study. OPAL at LEP and CDF at Tevatron already reported preliminary results on CP asymmetry in $\mathrm{B}(\bar{\mathrm{B}}) \rightarrow \mathrm{J}/\psi + \mathrm{K_S}$.[29 30] Quite recently, Asymmetric B-factories started operating at SLAC and KEK.

The future direction of the study of CP violation depends on what comes out from these experiments. If the results include some discrepancy from the prediction of the standard model, then we need an immediate cure that must be a breakthrough to "beyond standard" physics. On the other hand, if the results are consistent with the standard model, we will be left with much harder questions: "What determines the Higgs couplings?" "What is the origin of the generations?" etc.. Another challenging problem is the baryogenesis of the early universe. It is likely that we need some other CP violating mechanism in order to explain the matter dominance of the universe. In that case, an interesting question is whether or not there is a common origin for such a new CP violation mechanism and that of the six-quark model. These are not the issue of CP violation alone, but CP violation could be a good clue to attack such fundamental questions.

# References


[1] W. Pauli, "Die Allgemeinen Prinzipien der Wellenmechanik", Handbuch der Physik, Bd.XXIV, Teil 1 (1933).

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

[3] A. D. Sakharov, Pis'ma Zh. Exsp. Theor. Fiz. **5**, 32 (1967) [JETP Lett. **5**, 24 (1967)].

[4] M. Yoshimura, Phys. Rev. Lett. **41**, 281 (1978); **42**, 746(E) (1979).

[5] A. Yu. Ignatiev, N. V. Krasnikov, V. A. Kuzmin and A. N. Tavkhelidze, Phys. Lett. **76B**, 436 (1978).

[6] L. Wolfenstein, Phys. Rev. Lett. **13**, 562 (1964).

[7] M. Kobayashi and T. Maskawa, Progr. Theor. Phys. **49**, 652 (1973).

[8] T .D. Lee and C. N. Yang, Phys. Rev. **104**, 254 (1956).

[9] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes and R. P. Hudson, Phys. Rev. **105**, 1413 (1957); R. L. Garwin, L. M. Lederman and M. Weinrich, Phys. Rev. **105**, 1415 (1957); J. I. Friedman and V. L. Telegdi, Phys. Rev. **105**, 1681 (1957).

[10] R. P. Feynman and M. Gell-Mann, Phys. Rev. **109**, 193 (1958).

[11] M. Gell-Mann and M. Levy, Nuovo Cimento **16**, 705 (1960).

[12] N. Cabibbo, Phys. Rev. Lett. **10**, 531 (1963).

[13] S. L. Glashow, J. Iliopoulos and L. Maiani, Phys. Rev. **D2**, 1285 (1970).

[14] S. Sakata, Progr. Theor. Phys. **16**, 686 (1956).

[15] A. Gamba, R. E. Marshak and S. Okubo, Natl. Acad. Sci. USA **45**, 881 (1959); Y. Yamaguchi, Progr. Theor. Phys. Suppl. No.11, 1 (1959).

[16] Z. Maki, M. Nakagawa, Y. Ohnuki and S. Sakata, Progr. Theor. Phys. **23**, 1174 (1960).

[17] G. Danby *et al.*, Phys. Rev. Lett. **9**, 36 (1962).

[18] Y. Katayama, K. Matumoto, S. Tanaka and E. Yamada, Progr. Theor. Phys. **28**, 675 (1962); Z. Maki, M. Nakagawa and S. Sakata, Progr. Theor. Phys. **28**, 870 (1962).

[19] K. Niu, E. Mikumo and Y. Maeda, Progr. Theor. Phys. **46**, 1644 (1971).

[20] See S. Ogawa, Progr. Theor. Phys. Suppl. No.85, 52 (1985).

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

[22] M. Kobayashi, Progr. Theor. Phys. **92**, 287 (1994).

[23] L. -L. Chau and W. -Y. Keung, Phys. Rev. Lett. **53**, 1802 (1984)

[24] S. Pakvasa and H. Sugawara, Phys. Rev. **D14**, 305 (1976).

[25] L. Maiani, Phys. Lett. **62B**, 183 (1976).

[26] J. Ellis, M. K. Gaillard and D. V. Nanopoulos, Nucl. Phys. **B109**, 213 (1976).

[27] A. Alavi-Harati *et al.*, Phys. Rev. Lett. **83**, 22 (1999).

[28] S. Palestini, hep-ex/9909046.

[29] K. Ackerstaff *et al.*, Eur. Phys. J. **C5**, 379 (1998).

[30] T. Affolder *et al.*, hep-ex/9909003.

SLAC–PUB–8285
December 1999

# HADRONIC EFFECTS IN TWO-BODY $B$ DECAYS

Helen Quinn*
Stanford Linear Accelerator Center
Stanford University, Stanford, CA 94309

ABSTRACT

In these lectures I discuss the impact of soft hadronic physics on predictions for $B$ decays. Unfortunately our tools for calculating these effects are limited; even after the use of the best available tools the resulting theoretical uncertainties are difficult to delimit, and can obscure tests for the presence of beyond-Standard-Model physics. The first lecture reviews what tools are available, the second reviews in more detail two examples of how these tools can be used.

*Support by DOE Contract DE-AC03-76SF00515.

# 1 Lecture 1—Tools

## 1.1 What is the problem?

In these lectures I will follow the notation and definitions given by Yossi Nir in his lectures.[1] (For another excellent set of review lectures on CP Violation, including detailed references to the original literature see lectures by A. J. Buras.[2] For a recent book also covering this topic in detail see "CP Violation" by G. Branco *et al.*[3])

The physics of $B$ meson decays is governed by weak decay processes. Weak decays and any hard QCD effects are calculable by perturbation theory methods, but soft QCD effects not are directly calculable. Such effects are inevitably part of the meson decay process; they define the internal structure of mesons, the branching fractions to few and many-body channels, and the interactions between final-state hadrons once they have formed. Their impact can mask our ability to relate measurements to underlying Standard Model (CKM) parameters. This problem is a familiar one, it is not new in $B$ physics; in fact it is a much worse problem for lighter meson decays. The larger $B$ mass makes some of the physics more calculable, but even in the limit of extremely large $B$ mass there would be some work to do to deal with soft QCD effects.

Hard and soft QCD effects are separated by the scale of the momenta compared to the parameter $\Lambda_{QCD}$. This is the scale at which the strong coupling constant $\alpha_s$, as defined perturbatively, becomes infinite. Physically this scale sets the size of hadrons.* Any freely propagating quark or gluon with momentum small compared to this scale is a fiction—such particles are not observed because of confinement. Said another way: in this regime QCD perturbation theory is not meaningful and nor are Feynman-type diagrams, which are after all just a short-hand for perturbative calculations. Any time you see a line in a diagram for

*The scale $\Lambda_{QCD}$is usually defined as the scale that determines the $q^2$ dependence of the QCD coupling at high energy; in leading order $\alpha_s(q^2) = 12/[(33 - 2N_f)ln(q^2/\Lambda^2_{QCD})]$ where $N_f$ is the number of quark triplets. This scale then defines where the perturbative coupling becomes infinite, which is clearly well below the scale at which perturbation theory is no longer reliable. The physical phenomenon associated with the growth of the coupling at short distance is confinement, and one physical manifestation of that phenomenon is the size of hadrons. It is in this sense that $\Lambda_{QCD}$defines the size scale of hadrons; the two scales are not numerically equal but are related quantities. The relationship cannot be calculated perturbatively, but can be explored in lattice calculations.

a low-momentum quark or gluon you should be suspicious. In reality any such line comes dressed with a multitude of soft gluon emission and absorption processes, and also additional soft quarks and antiquarks. This part of QCD physics is not perturbatively calculable. To incorporate its very real effects we must resort to other tools. Conversely, for quarks or gluons with momenta large compared to the scale $\Lambda_{QCD}$ QCD perturbation theory is an effective and accurate tool.

"Hadronic effects" in my lecture title refers to the soft part of the physics. In my first lecture I will review what tools are available to treat the problem and briefly comment on the uses of these tools. For some further discussion of some of the topics that I treat rather briefly here see A. F. Falk.[4] In the second lecture I will turn to a few specific examples that illustrate in more detail how these tools can be used. Even with the best available tools some residual uncertainties about the impact of soft physics remains. The term "theoretical uncertainty" is used here to characterize impact of this poorly-calculated physics on the extraction of well-defined parameters such as the elements of the CKM matrix. One unfortunate consequence of these uncertainties is that they can mask possible new physics effects, as they obscure the relationship between the data and clean Standard Model predictions.

The goal is then to minimize the parts of the calculation affected (or should I say infected?) by these uncertainties. In addition one hopes that, eventually, comparison of data and calculations for many channels can provide some confidence in the reliability with which the residual uncertainties can be estimated. However it is important to remember that the estimates of these uncertainties are just that, estimates. They may be based on nothing more than a particular theorist's gut feeling about the subject. The models and approximations used are often simply not well-controlled enough for one to know how big the corrections might be. It is not justifiable to treat these estimated uncertainties as if they were statistical errors. This is often done; procedures such combining these uncertainties in quadrature and quoting probabilities for a deviation of twice the theoretical error as if these were statistical standard deviations are all too common.

Figure 1 indicates why the soft physics is usually unavoidable. We can use perturbation theory to calculate the part of the diagram within the inner magnifying lens, that is the short distance parts. The weak decay is short distance because the mass of the decaying $b$-quark is small compared to the $W$-mass, so the $W$ is highly virtual. At the same time (and here the difference with lighter mesons

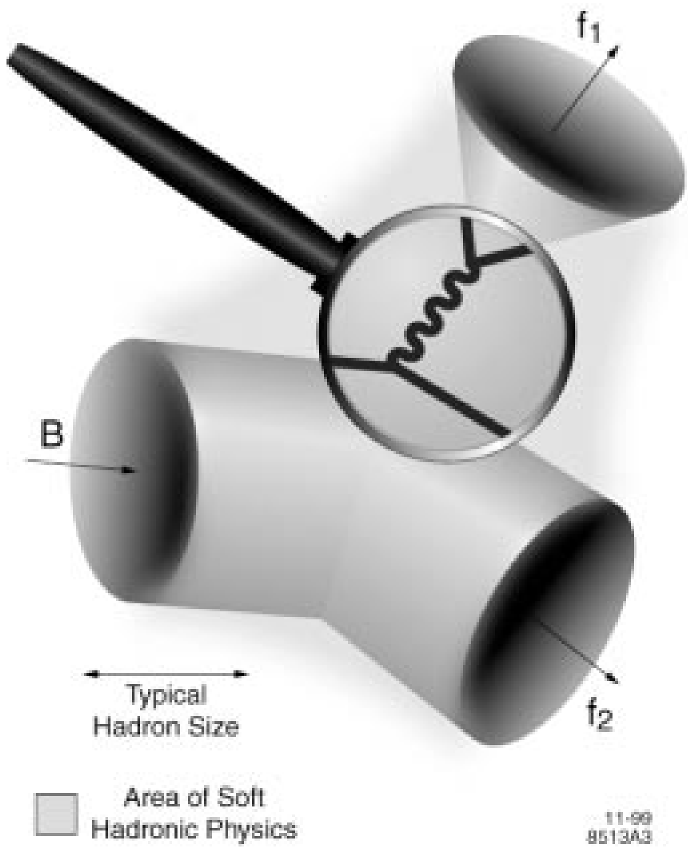



Fig. 1. The hard and soft regions of a typical $B$ Decay. The shaded area contains many (undrawn) soft gluons and quark-antiquark pairs, only the hard process (within the magnifying glass) is perturbatively calculable.

appears), the $b$-mass it is heavy enough that the produced quarks in general have momenta which are large compared to $\Lambda_{QCD}$. In addition hard gluons exchanged between these particles can be included perturbatively. However the initial and final hadron wave-functions, the quantities that describe these hadrons in terms of their quark content, are not perturbatively known, nor do they contain only hard quarks. Even if we take the simplest possible picture of a $B$-meson as a static heavy $b$-quark surrounded by some wave-function distribution for the light quark, that light quark has a typical momentum set by the size of the meson and hence, by definition, of order $\Lambda_{QCD}$. In most calculations of few-body decays this "spectator" quark (so-called because it does not participate in the weak decay except in the case of annihilation diagrams where it is clearly no longer a spectator), is assumed to hadronize as a valence quarks of one of the final mesons. This certainly does the book-keeping of charge etc. correctly, but it gives a deceptively simple diagrammatic picture. Such a quark cannot be included in the hard or short-distance part of the calculation; any estimates that depend on treating it as a freely propagating particle are at some level suspect.

In making the division between hard and soft physics an arbitrary and unphysical scale $\mu$ is introduced into the problem.[5] This scale must be chosen to be large compared to $\Lambda_{QCD}$ but is otherwise unconstrained. As is usual in QCD calculations one ends up with terms of the form $\alpha_s(\mu)ln(km_b/\mu)$ where k is some number (probably of order unity) and the scale $m_b$ enters because it is the quantity that defines the scale of momenta flowing in the hard quark lines. In order to avoid having this logarithm be large, it is convenient to choose $\mu$ of order $m_b$, provided that does not make $\alpha_s(\mu)$ too large. Here is where the $B$ system theoretical analysis is in much better shape than that for charm decays or especially for $K$-decays. The fact that $m_b$ is large compared to $\Lambda_{QCD}$ makes the hard/soft division a relatively clean business in $B$ physics.

In these two lectures I will confine my attention chiefly to two body (or quasi-two-body) decays, for the sake of specificity. The problem of dealing with soft hadronic physics effects is not unique to calculations of two body decays, nor are the general statements made below about methods and symmetry limits special to those decays. Many of the general approaches I mention here also have some applications for inclusive processes and for many body decays. My intent here is not to teach you to use any of the tools that I discuss, but rather to make you aware of them and of their uses, and their limitations. To actually learn to use these tools and approximation methods requires more time than we have available in these two lectures.

As was demonstrated in Yossi Nir's lectures,[1] two-body and quasi-two-body decays to states which are $CP$ eigenstates are of particular interest in neutral $B$ CP violation studies. Such states occur only for the case of two pseudoscalars, or one higher spin (typically spin 1 is studied) and one pseudoscalar, because for both these cases the decay of a spin-zero $B$ can give only one possible relative angular momentum for the two produced particles. Hence a state of definite CP is produced. For two higher spin particles, even when the particle content of the state is CP-self-conjugate, the $B$ decay produces an admixture of CP-even and CP-odd final states because both even and odd relative angular momenta between the two produced particles are allowed. In many cases such systems can be separated into states of definite CP via angular analysis of the decays of the two quasi-stable "final state" particles.[6] Then methods similar to those discussed here for the simpler modes can be applied, once sufficient data is available. Without this separation a "dilution" or cancelation effect occurs in the measured asymmetry; the CP-odd states contribute the same asymmetry as the CP-even ones except for an overall sign, so the two contributions partially cancel each other.

Methods for extracting CKM parameters from asymmetries in production of inclusive final states with a particular CP-self-conjugate quark content have been suggested.[7] These depend on estimates of the CP-even and CP-odd fractions of the decay final states. Such estimates are made at the quark level. They are reliable for the total inclusive rate because hadronization, being a strong interaction process, respects CP symmetry. Typically they suffer from large hadronic uncertainties once any cuts are introduced. Such cuts are unavoidable; they are needed either to define experimental apertures or to discriminate data from backgrounds. I will not discuss such methods further here.

## 1.2 Scales, Exact Limits and Expansions around them.

One of the things that makes the physics of $B$ decays complicated is that many scales can play some role in the problem. Roughly in order of increasing size these are

$$m_u, m_d, m_s, \Lambda_{QCD}, m_c, m_b, \mu$$

where the scale $\mu$ is an unphysical parameter introduced in QCD marking the division between hard and soft QCD effects calculations while $\Lambda_{QCD}$ is the scale that defines the running of the coupling in QCD.

For any significant hierarchy in these scales it is instructive to pursue the limit in which either the small scale is taken to zero, or the larger one to infinity. For example, since $\Lambda_{QCD}/m_b \ll 1$,and $\Lambda_{QCD}/m_c < 1$ the heavy quark limit $m_c/m_b$ fixed, $m_b = \infty$ is a useful approximation to the real world.† It is useful for two reasons. First the theory has additional symmetries which provide exact constraints in this limit. These constraints can be used to limit or relate various model parameters by requiring that the model have the correct limiting behavior.

†For some purposes the limit $m_b - m_c$ fixed, $m_b \to \infty$ may be convenient to consider; it is important to recognize that there are subtle differences between these two variants of heavy quark limits, and to be aware which is used for a particular argument.

Second, one can calculate corrections to this limit as a power series in small quantities, namely ratios $\Lambda_{QCD}/m_b$ and $\Lambda_{QCD}/m_c$. This is called the heavy quark expansion. One has good control over the sizes of neglected corrections and hence over theoretical uncertainties due to these corrections. Unfortunately the second ratio, $\Lambda_{QCD}/m_c \approx 1/3$ is not so small in the real world; quantities where such terms are not suppressed have significant corrections to the limiting behavior. Cases where the leading correction is second order in this ratio are particularly attractive for this approach. Working down the scale hierarchy the following approximations and limits can be considered:

| | | |
|---|---|---|
| $(m_s - m_d)/\Lambda_{QCD} < 1$ | Limit: $m_u = m_d = m_s$ | SU(3) Invariance |
| $(m_d - m_u)/\Lambda_{QCD} \ll 1$ | Limit: $m_u = m_d$ | Isospin Invariance |
| $m_u/\Lambda_{QCD} \ll 1, m_d/\Lambda_{QCD} \ll 1$ | Limit: $m_u = 0, m_d = 0$ | Chiral Invariance. |

Each of these limits can be useful in restricting uncertainties in hadronic physics effects by introducing constrained parameterizations with somewhat controlled corrections. Examples will be discussed in more detail below, and in the second lecture. One point of caution: sometimes the interplay of more than one of these scales can limit the effectiveness of such expansions. Terms which might be treated as small because they contain inverse powers of a large mass cannot be disregarded if at the same time they contain inverse powers of a small mass.

Many of the methods for estimating matrix elements or form factors do not introduce any explicit $\mu$ dependence in them. Thus, at best, the estimates can be valid at only one value of $\mu$. Often we have no good arguments to choose that scale. It is not uncommon for theorists to characterize the uncertainty introduced by this error in matrix element calculations by looking at the variation of the result over the range from $m_b/2 \leq \mu \leq 2m_b$. This choice of range has no theoretical justification. In some model calculations the natural scale for the model is a light hadron mass scale, too small a scale to be acceptable from the QCD point of view. Because of this mismatch between the scale at which the model estimate of the matrix elements can be made and the plateau region of the coefficient calculation it is difficult to characterize the size of the uncertainty in calculations that depend on such models. Methods such as lattice calculation where the matrix element calculation does have explicit scale dependence give much better hope for eventual results with well-controlled uncertainties.[8]

## 1.3 Heavy Quark Limit

This limit is most useful in the context of decays $B \to DX$, particularly the semi-leptonic processes; for example it provides important control over the theoretical uncertainties in the extraction of $V_{cb}$. The best cases are those where the leading correction is quadratic in the quantity $\Lambda_{QCD}/m_c$, since this ratio is not small enough for terms proportional to a single power of it to be a small correction. For channels with no final state charm particles one can use the heavy quark limit to relate $B$ decays to corresponding $D$ decays, for example extracting the behavior of form factors for $B$ decay from those measured in the $D$ decay case. The accuracy of this approach is limited, both by the accuracy with which the $D$ decays are measured and by $\Lambda_{QCD}/m_c$ corrections. There is a large literature on the subject of heavy quark limit calculations, I will not discuss these methods further here.[9]

The heavy quark limit is generally applied for hadronic $B$ decays only in combination with the factorization approximation. In the cases $DX$ it has been shown that factorization is valid in the heavy quark limit for a particular kinematic region.[10] For charmless decays the combination of the two methods adds uncontrolled theoretical uncertainties.

## 1.4 Isospin

Isospin analysis is a useful tool in some $B$ decays, principally for its role in separating gluon-mediated penguin contributions from tree-diagram contributions (see Yossi Nir's lectures for definition of these two types of diagrams). The crucial point is that gluons have isospin zero which limits the isospin amplitudes to which they can contribute. The details of how the isospin information is used depends on the channel. I will review this in some further detail for a couple of channels below and in my second lecture.

For this young an audience it is probably necessary to start a discussion of isospin analysis by defining what is meant by isospin. Isospin is an SU(2) algebra in which the up and down quarks are treated as two identical members of a doublet. Note this strong interaction doublet is similar to, but not the same as, the weak SU(2) (sometimes also called weak isospin) doublet which pairs the up quark with a linear combination of down-type quarks

$$d\cos(\theta_{12})\cos(\theta_{13}) + s\sin(\theta_{12})\cos(\theta_{13}) + b\sin(\theta_{12})\sin(\theta_{13}). \tag{1}$$

Isospin is a symmetry of the strong interactions but not of electroweak, which clearly distinguish quark charges and flavors. It is also broken by quark mass terms. Historically the name isospin came about because physicists were familiar with the SU(2) algebra as the algebra of spin, and from the relationship of the multiplets of this symmetry to the isobars of nuclear physics (nuclei of equal A). From a modern perspective we can understand that hadrons form approximately degenerate isospin multiplets with mass differences small compared to the average mass of the multiplet because most hadron masses are dominated by $\Lambda_{QCD}$. The up-down quark mass difference is small on this scale, even though their mass ratio is far from 1. The exception is that pseudoscalar octet masses scale as $\sqrt{m_q \Lambda_{QCD}}$ ‡ and thus the effect of quark mass differences can give larger isospin breaking in this multiplet.

Isospin-breaking can also be significant in the neutral meson states. Ideally the two neutral quark-antiquark states have $I = 0$ and $(I = 1, I_3 = 0)$: $\eta_{ud}$ and $\pi_0$ for the pseudoscalars, $\omega$ and $\rho$ for the vectors. In actuality, because the up and down quark masses are not identical, the mass eigenstates have small admixtures of the wrong isospin state. This can lead to important contributions that are neglected if the physical particles are treated as having a definite isospin.[11] (The notation $\eta_{ud}$ also serves to warn that, for the pseudo-scalars, the strange-antistrange combination is also mixed into the physical $\eta$ particle; $\eta_{ud}$ means that combination of $\eta$ and $\eta'$ with no strange quark part.)

The photon and the $Z$ each couple to up and down quarks with a well-defined ratio of I=0 and I=1 couplings, for both vector and, in the case of the $Z$, axial vector couplings. These couplings are usually written in terms of coefficients $g_1^X, g_2^X$ for coupling to $u$ and $d$ quarks respectively, with superscripts $X = V, A$ for the vector and axial vector couplings respectively. The combination $(g_1^X \pm g_2^X)/\sqrt{2}$ are the definite isospin couplings. This relationship between coefficients gives a relationship between the amplitudes of definite isospin for a given $Z$-mediated or photon-mediated process if final state interactions are neglected. The final state interactions introduce corrections, including complex phases from absorptive parts, which are in general different in the different isospin states and cannot be

‡This scaling follows from the pseudo-goldstone nature of the pseudoscalar mesons and the PCAC (partially conserved axial current) relationships such as $m_\pi^2 f_\pi^2 = (m_u + m_d) \langle \overline{\psi}\psi \rangle$, since $f_\pi \propto \Lambda_{QCD}$ and $\langle \overline{\psi}\psi \rangle \propto \Lambda_{QCD}^3$ are both quantities whose scale is defined by QCD confinement physics.

calculated from first principles—that is without further assumptions.

Since photons and $Z$ bosons have $I = 1$ as well as $I = 0$ couplings to quark-antiquark states electroweak penguin effects cannot be removed by the same isospin analysis that eliminates QCD penguin effects.[12] Their impact varies from channel to channel, but must be considered. This limits the usefulness of isospin in removing hadronic uncertainties in the extraction of CKM parameters from CP violation in some weak decays. However in many channels the electroweak penguin effects can be shown to be small. Then the uncertainties that they induce in the extraction of CKM parameters are likewise small.

As an example of how isospin enters in $B$-decays let us consider the decays based on the quark process $b \to u\overline{u}d$.[13] These three final quarks can have either $I = 1/2$ or $I = 3/2$, thus we can label the quark transition as $\Delta I = 1/2$ or $\Delta I = 3/2$. The additional (spectator) quark (and hence the charged $B$ and $B_d$) are an isodoublet. Thus, combining this initial isospin with the transition isospin $\Delta I$, we find four possibilities $\Delta I = 1/2, I_f = 0$; $\Delta I = 1/2, I_f = 1$; $\Delta I = 3/2, I_f = 1$ and $\Delta I = 3/2, I_f = 2$ for these decays. The gluonic penguin can contribute only to the first two cases, because the gluon couples only to the $I = 0$ combination of quarks $u\overline{u} + d\overline{d}$. Hence gluonic penguin contributions have $\Delta I = 1/2$ only. Any pure $I = 2$ contribution is thus unaffected by gluonic penguin contributions. Up to corrections from electroweak penguins, it has the property $\overline{A}_2/A_2 = 1$ in the Standard Model. Thus, if this contribution can be isolated, it can provide a relatively clean estimate of the related CKM parameter in channels where the electroweak penguin effects can be demonstrated to be small relative to the dominant terms.

Another reason to arrange the calculation in terms of isospin amplitudes is that final state interactions mix states of different charge structure but, since they are strong interaction effects, do not change isospin. Let us expand in the basis of strong interaction eigenstates $\left|i^I\right\rangle$, for which the scattering matrix is diagonal. The diagonal strong interaction scattering matrix contains an independent strong phase for each entry

$$\left\langle j^I|\mathcal{H}|i^I\right\rangle = \delta_{ij} e^{2i\delta_i^I}. \tag{2}$$

The eigenstates have definite isospin, but include both two-particle and many-particle components. Thus more than one eigenstate $\left|i^I\right\rangle$ exists for each isospin $I$.

The kinematic structure of each operator is different, thus the states of given isospin produced from the $B$ by two different operators are, in general, different linear combinations of the strong interaction eigenstates; we write $\left\langle i^I|\mathcal{O}_j|B\right\rangle = x_i^I(\mathcal{O}_j)$. The rescattering effect introduces the square root of the scattering matrix[14]; heuristically one sees this by noting that the process is not going from an in state to an out state, but starting "in the middle" from a pointlike local superposition and evolving to an out state. Finally, to consider a given two-body final state $f^I$ one needs the overlap $\left\langle f^I|i^I\right\rangle = a_i^I(f)$. Thus one can write

$$\begin{aligned} A^I(\mathcal{O}_j, f) &= \Sigma_i < f^I|i^I > e^{i\delta_i^I} \left\langle i^I|\mathcal{O}_j|B\right\rangle \\ &= \Sigma_i a_i^I(f) e^{i\delta_i^I} x_i^I(\mathcal{O}_j). \end{aligned} \tag{3}$$

This expression is not very useful since, in general, we cannot calculate any of the quantities in the right-hand side. However, it does serve to destroy a couple of myths that appear now and then in the literature. The first is that the only effect of rescattering is to introduce a phase in the isospin amplitudes. The second is that the strong phase for an amplitude with a given isospin is the same independent of the operator. A little playing with the above expression, say for the cases where there are just three strong eigenstates, will show that neither of these statements is true in general. One sees that they would each be true if there were only a single strong eigenstate excited for each isospin, or if the two-body state of definite isospin were by itself a strong interaction eigenstate. (In general, neither of these conditions is true.) [§]

## 1.5 SU(3) Symmetry

This is another approximate strong interaction symmetry, very much like isospin except that in addition to equal mass up and down quarks the symmetry limit requires the strange quark mass to be degenerate with them. Since the ratio $(m_s - m_d)/\Lambda_{QCD}$ is not so small, SU(3) breaking effects can be large. In $B$ decays the most common use of SU(3), beyond its isospin subgroup, is the application of results due to another SU(2) subgroup of the SU(3), traditionally called U-spin. U-spin treats the $s$ and $d$ quarks as a doublet of identical particles. For example, it relates rates where pions are replaced by kaons, and/or $B_d$ by $B_s$.[15]

[§] The misperception that just one state and hence one phase exists for each isospin is perhaps a holdover from low energy isospin physics, where it is true because the multibody channels are kinematically excluded.

In the factorization approximation, for any quantity where an axial current produces a pseudoscalar meson the SU(3)-breaking effect is known, it is the ratio $f_K/f_\pi$, which is measured to be $1.22 \pm 0.01$. For the vector current producing a vector meson the relevant correction factor is $F_K/F_\pi$. Similar corrections occur for transition matrix elements. These corrections provide, presumably, a good first estimate of SU(3) breaking corrections, though there may be further corrections due to differences in final state scattering effects for the two different mesons. At the $B$ mass scale it is reasonable to assume that these are small corrections. However for many other contributions the use of the ratio $f_K/f_\pi$ (or $F_K/F_\pi$) to estimate the SU(3) breaking is not justified even in factorization approximation, and large theoretical uncertainties remain. In some calculations the two SU(3)-related amplitudes for these cases are allowed independently parameterized magnitudes and the SU(3) symmetry approximation is applied only to identify their strong phases.[16] Once again the corrections to this approximation are expected to be small at the $B$ mass. However I do not know how to quantify the expected size of "small" effects due to SU(3) breaking of the strong-rescattering phase relationships. In any particular case one can test the impact of relaxing this constraint by looking at how the fit for the CKM parameters of interest change with the difference between the two strong phases, but no clear statement prescription for what would be a "reasonable range" of phase differences to allow in such a treatment can be given.

## 1.6 Chiral Symmetry

The chiral limit and chiral perturbation theory are based on the approximation that the up and down quarks are massless in which case the pion is a Goldstone boson. This leads to an expansion of amplitudes for the production of an additional soft pion in terms of the amplitude without that pion and correction terms which occur as powers of the momentum of the soft pion scaled by $\Lambda_{QCD}$. (This scaling defines what is meant by soft in this context.) While this method has some uses in $B$ physics calculations[17] it is not a useful tool for the treatment of two body hadronic decays, since the pions produced in such decays are not soft. I will not discuss chiral expansions further in these lectures.

## 1.7 QCD Sum Rules

These are conditions derived from the analytic structure of QCD perturbation theory.[18] Sum rules typically relate certain matrix elements or derive constraints on their kinematic form in particular limits. Such constraints are useful in limiting the arbitrariness of models, for example those for form factors in semi-leptonic decays. The BaBar Physics Book contains an appendix which discusses this subject. I will not treat it further in these lectures.

## 1.8 Lattice calculation of matrix elements

Ideally we need a method for calculating the long distance contributions, that is the matrix elements, that correctly includes all soft physics. This would also give the correct sensitivity to the hard-soft division scale $\mu$. The method with the best hope of doing this is lattice calculation.[8] QCD sum rules can also be used to extract information about certain properties of form factors, but are not powerful enough to calculate the matrix elements themselves. Unfortunately, for most the cases of interest here, the same thing must be said about the lattice calculation of matrix elements, at least at the current state of the art.

For two-body $B$ decays these matrix elements are three-point functions connecting the initial $B$ to the two final-state particles. In actuality what is calculated on the lattice so far is a less-demanding two-point function, where one of the final particles has been "reduced in".[19] It thus appears in the operator that is evaluated, rather than as a final state particle. This removes all sensitivity of the calculation to final state interaction phases, which are one of the major issues for CP-violation physics.[20] Furthermore, most of the relevant lattice calculations have so far only been made in the "quenched approximation" —which means in the approximation of suppressing any virtual quark-antiquark-loop contributions. As with experiments, lattice calculations then have a statistical uncertainty of their result and in addition non-statistical (or systematic) uncertainties arising from these various simplifying approximations. The former are readily estimated and clearly given in lattice results, the latter are hard to estimate and hence again significant theoretical uncertainties remain in most cases.

Where an unquenched calculation exists results are sometimes significantly different from unquenched results for the same quantity. We have no good understanding of how to quantify these differences prior to making the more difficult unquenched calculations. A growing number of unquenched calculations are appearing, but as yet no true three-body calculations. Again there is a large literature on this subject and I do not the time (nor the expertise) to cover it in detail.[8]

There are a number of quantities relevant to the extraction of CKM parameters from $B$ physics for which the lattice calculations are in much better shape than for the three body matrix elements discussed above. For quantities such as $F_B$, and many the various $B_i$ parameters (parameterizing the ratio of true matrix element to vacuum insertion approximation results for the QCD operators $\mathcal{O}_i$) unquenched calculations are beginning to be feasible. Reliable values (with uncertainties in the few percent range) are expected for most of these quantities within the next few years.

## 1.9 When are these methods useful?

I have summarized a fairly large "bag of tricks" for dealing with hadronic effects. Remembering Feynman's dictum that if you have one good method you don't need any others, the length of the list alone should give you an idea of the state of the problem! The applicability and efficacy of each of these methods varies from channel to channel. In the best cases we do not need any of them, because, as Yossi explained, when amplitudes with only a single weak phase dominate a decay, as is the case for the channel $J/\psi K_S$, the hadronic amplitudes cancel out in the ratio that defines the CP asymmetry. Then none of the uncertainties in calculating the matrix elements matter. Such a mode gives the cleanest relationship between a CKM matrix element phase and a measured asymmetry. Conversely the problems are worst when the same channel receives two comparable-magnitude contributions, say from suppressed tree diagrams and from penguin diagrams, or from two different penguin diagrams in a channel with no tree contributions, and the two contributions enter with different weak-phases, that is with different CKM matrix element coefficients. In each such case the relative strength and the relative strong phases of the two contributions affect the relationship between the measured asymmetry and any CKM parameter. One must then use whatever tools are available to try to make estimates of these effects, and equally important, to constrain the uncertainties in these estimates.

## 1.10 Approximations that do not come from exact limits

In many cases the methods described above are not sufficient to obtain all the desired information. When this is the case one is forced to resort to less-controlled approximations, which generally have some intuitive model as their underpinning. Such methods are very useful, for example to obtain estimates of the expected branching fraction for various channels. The most commonly used approximation is that of factorization, which I will discuss shortly. It is difficult to obtain any good estimate of the theoretical uncertainties introduced by such an approximations. Thus it is very difficult to find convincing evidence for non-Standard-Model contributions from any conflict between such estimates and measured results. However they are part of the standard toolkit for calculating $B$-decay processes and so are worth mention here.

## 1.11 Factorization

This approximation starts from the operator product expansion and provides an estimate of the matrix element of the local four-quark operators. One takes any such operator and finds any possible Fierz-rearrangement that groups the four quark fields into two that can create one of the final-state hadrons from a vacuum state, and two that describe a transition matrix element from the $B$ to the other final state hadron. All final state interactions between the two hadrons are ignored, as are any operators that cannot be arranged in this way. This is a very useful approximation as it allows a few-parameter model to describe many two-body decays, using transition matrix elements measured elsewhere, for example in semileptonic decays.

The idea behind this ansatz is that the region of the phase space where the two-body final state is most likely to be produced is that where two quarks that form a meson are produced moving roughly together and in a color-singlet combination. Since the operator that produces them is local, the state so made is a local color singlet state. Hence, unlike a real finite-sized hadron, it has a very small strong interaction cross section with the other quark-antiquark system. Since the two systems are rapidly moving apart, they are far separated from it before the local state has evolved into its final finite-sized configuration as a hadron. Thus it can be expected that no significant strong interaction rescattering occurs between the two mesons so formed. This "color-transparency" argument is attributed to

Bjorken.[21]

When the two quarks that have the right flavor and tensor structure to form the single meson are not automatically in a color singlet state the color transparency argument is less immediately obvious. Effectively the requirement that the meson is formed projects out the color singlet part of the $\overline{q}_\alpha \Gamma_i q'^\beta$ operator (here $\Gamma_i$ denotes some gamma-matrix structure and $\alpha$ $\beta$ are color indices). The color counting then gives a suppression of $1/N_c$ since the "color-allowed" contribution

$$\Sigma_\alpha \left\langle m_1 | \overline{q}_\alpha \Gamma_i q' | 0 \right\rangle \Sigma_\beta \left\langle m_2 | \overline{q}_\beta \Gamma_i q'^\beta | B \right\rangle \propto N_C^2 \tag{4}$$

whereas the contribution

$$\Sigma_\alpha \Sigma_\beta \left\langle m_1 | \overline{q}_\alpha \Gamma_i q'^\beta | 0 \right\rangle \; \left\langle m_2 | \overline{q}_\beta \Gamma_i q'^\alpha | B \right\rangle \propto N_C \ . \tag{5}$$

This is the "color-suppressed" factorized contribution.

If the argument for neglecting final state interactions is rephrased in the language of strong interaction eigenstates given in the isospin section above, it looks much less attractive. As best I can see, it seems to say that the operators excite a linear combination of strong interaction eigenstates each of which gets a strong phase from rescattering, but in such a way that their vector sum is unchanged. (Another option, that looks even less plausible, is that the $B$-decay forms only a single strong interaction eigenstate involving any two pion component, and that that state has zero rescattering phase.)The general formalism instead suggests that configurations where the two quarks that make the final meson are not produced traveling together can contribute, via rescattering, to the two-body final state, even when naive expectations say that is unlikely. This contribution may indeed be small, but we cannot say how small. Our intuition rejects this possibility just because we know that for any given many-body state the probability of rescattering to two pions is typically small. However, at the $B$-mass, the cross section for two pions in an $s$-wave to scatter into many pions is not expected to be small. Thus the inverse process must also be possible for some configurations of the many particles. The problem is that any way of making the exclusive two body final state is suppressed, either because it involves a small corner of the four-quark phase space where two quarks happen to move together or because it involves a many particle to two particle rescattering. Intuition is generally a remarkably poor guide to discovering which of two unlikely events is more likely. I make this comment just to show how little we actually know—and that models can seem quite plausible in words but have little calculational basis. It is not that I know the color transparency argument is wrong—just that I know no way of proving that it is right either.

There have been a number of papers devoted to the impact of final state interactions, which are neglected in the factorization approximation. Some approach the problem generally, others consider specific channels. Some sample papers on this topic are given in the references.[22]

Recent work by Beneke, Buchalla, Neubert and Sachrajda[23] has introduced a more detailed study of how this factorization idea plays out in a one loop calculation, and at leading order in $\Lambda_{QCD}/m_b$ Their approach depends on certain assumptions, such as the dominance of the simple quark-antiquark state in the composition of the meson wave-function, compared to any contribution where additional soft quarks and antiquarks play a key role. It is not based on a rigorous operator product starting point, even in the infinite $m_b$ limit. They find that there are certain additional contributions that are ignored in the simplest factorization calculations, which means there are more input parameters to be determined in their calculations than in the usual factorization approximation calculations. However once these contributions are added they find that final state interactions are suppressed at the one loop level, because of cancelations of the type one would expect from color-transparency arguments such as that given above. They are currently in the process of extending their study to the level of two-loops.

One problem with the factorization approach is that is gives no scale dependence for the matrix elements. Since the coefficients are scale and renormalization-scheme dependent, naive factorization cannot be precisely true except possibly at some particular scale, and in conjunction with a particular choice of renormalization scheme. A common approach to this problem is to use the induced scale and scheme dependence as an estimate of the theoretical uncertainty of the method. However this is surely not a rigorous argument, firstly because the answer depends on the range of scales allowed, and secondly because it gives no estimate whatsoever of the contributions that are ignored in the factorization approximation. The best one can say is that this dependence sets a lower bound on the theoretical uncertainty. But of course what we really need is an upper rather than a lower bound on uncertainties.

## 1.12 Quark Hadron Duality

This set of theoretical buzz words has two basic versions—global duality and local duality. Global duality is the statement that when averaged appropriately over some range of center of mass energies the rate for a given process predicted by a quark level calculation must be the correct result for the rate at the hadron level. For certain quantities such as the ratio of the hadronic cross section to the $\mu^+\mu^-$ cross section in $e^+e^-$ scattering this can be demonstrated to follow from the analyticity structure of the propagator function $\Pi(s)$.[24]

Local duality is the same idea applied at a given center of mass energy. In $B$ decays we cannot vary the energy, it is the $B$ mass, so to relate the quark quantities we know how to calculate to the hadronic quantities we know how to measure we are forced to make this stronger assumption. There is no good justification for the truth of this assumption, nor is there any good way to estimate the size of the uncertainty it introduces. Even within the assumption of local duality there is a weaker and a stronger form. The weaker assumption is to apply duality arguments to calculate rates for a particular class of inclusive decays, the stronger assumption is to rely on details of the quark-level kinematics to predict the hadron-level properties. In fact at the end points and in resonance regions of the spectrum this last approximation must be wrong, because quark kinematics does not know about resonance widths and hadron masses, etc. As soon as one goes from a truly inclusive prediction to one that takes into account any experimental acceptance cuts the predictions tend to become dependent on this strongest form of the quark-hadron duality assumption, and the theoretical uncertainties increase accordingly.

## 1.13 Parameterized Amplitudes and Models

Another way that one can proceed is to introduce parameters for each diagram or each isospin amplitude. One then obtains constraints by relating the parameters describing similar contributions in different processes, via symmetries such as isospin and SU(3). Conversely one can use models to calculate the value of the parameters for each type of contribution. Here the hope is that, with enough channels studied, these parameterized amplitudes will eventually become sufficiently constrained to be predictive. The goal is that the estimates be reliable enough to make relatively definite predictions about some of the interesting quantities, and set relatively reliable constraints on the theoretical corrections to a given calculation. It is certainly true that with enough data from enough channels we can begin to get a better control. Whether that control will become good enough that we could unambiguously identify a non-Standard-Model contribution in channels where more than one amplitude contributes remains to be seen. The history of calculations of hadronic effects in $K$-decay processes, or even $D$-decays, does not give grounds for optimism. Here we are working in a very different kinematic regime and the asymptotic freedom of QCD and the heavy quark limit begin to work in our favor. Time alone will tell how well we can do.

# 2 Lecture 2—Examples

In this lecture I will review some examples where the tools of isospin analysis and SU(3) discussed in the previous lecture may be useful. I will also make a few comments on the impact of hadronic effects in extracting the magnitude of CKM matrix elements, such as $V_{ub}$.

## 2.1 Isospin analysis for $b \to u\overline{u}d$ channels

### 2.1.1 Two pions

In the case of two identical particles in an orbital-angular-momentum zero state (because they are two pseudoscalars coming from a $B$ decay) the set of isospin amplitudes described for this quark content in my first lecture ($\Delta I = 1/2, I_f = 0$; $\Delta I = 1/2, I_f = 1$; $\Delta I = 3/2, I_f = 1$; and $\Delta I = 3/2, I_f = 2$) is reduced. Bose statistics requires a state of even isospin, so that the overall state is even under the interchange of the two pions. Hence the $I_f = 1$ amplitudes are all identically zero. This means only two tree amplitudes, $\Delta I = 1/2, I_f = 0$ and$\Delta I = 3/2, I_f = 2$, and only one penguin amplitude, $\Delta I = 1/2, I_f = 0$, contribute.

Gronau and London[13] showed how a measurement of rates for all three channels $B_d \to \pi^+\pi^-$, $B_d \to \pi^0\pi^0$, $B^+ \to \pi^+\pi^0$ and their CP-conjugates, together with a time dependent asymmetry measurement for the charged pions only, can be used to isolate the weak phase of the $I_f = 2$ contribution. In principal, this method provides a clean measurement of $\sin(2\alpha)$, where $\alpha$ is the angle $\pi - \beta - \gamma$ in the unitarity triangle. Unfortunately the rates for all these channels are low,[25] and the rate for the difficult to measure $\pi^0\pi^0$ channel is expected to be even lower. It appears that the uncertainty of the measurement of this last channel will render the method impotent to obtain a precise result.[26] Put another way, for the foreseeable future the experimental uncertainty on the neutral-pion measurement will be at least as large as the theoretical uncertainty in the shift of the measured charge-channel asymmetry from the simple form $\sin(2\alpha)\sin(\Delta mt)$.

### 2.1.2 $\rho\pi$ channels and the Dalitz plot

For $B_d \to \rho\pi$ three channels contribute, namely the three possible charge assignments for the $\rho$ and the pion, all decaying to the same final state $\pi^+\pi^-\pi^0$. However, by the arguments given in the previous lecture, only two independent QCD-penguin amplitudes exist. One can take the three independent tree amplitudes to be one for each charge channel and the QCD penguin amplitudes to be one each for $I_f = 0$ and $I_f = 1$. (If one plans also to use charged $B$-decay amplitudes to three pions one additional tree amplitude enters; one must measure both the three-charged and the two-neutral, one charged pion final states before significant additional constraints are obtained in this way. The latter is more difficult experimentally, so I will here discuss a study involving only the neutral $B_d$-decays to three pions.)

Five independent amplitudes, one CKM parameter and only three channels looks a bit discouraging. However Art Snyder pointed out to me an important feature of the physics here that could be useful. In some regions of the Dalitz plot more than one of the three channels can contribute. Hence there might be information to be extracted from the interference effects in the overlap regions. Based on his suggestion we made a preliminary study of this channel and found that this is indeed the case. The number of parameters to be fitted requires a large data sample.[27] Further studies made as part of the BaBar Physics workshop confirm this conclusion, and find that, as one might expect, the inclusion of physics backgrounds from other resonances and from non-resonant $B \to 3\pi$ decays, as well as non-$B$ backgrounds make things even more difficult. However the analysis remains an intriguing if distant possibility, so I will describe a little how it works.

The amplitudes for the specific channel decays can be written

$$\begin{aligned} A^{+-} &= (T^{+-} + P_1 + P_0) \\ A^{-+} &= T^{-+} - P_1 + P_0 \\ A^{00} &= T^{00} - P_0 \ . \end{aligned} \tag{6}$$

The assumption made in this approach is that each of the five contributing tree and penguin amplitudes for $B_d \to \rho\pi$ has an independent but fixed (*i.e.* not kinematically varying over the $\rho$ width) strong phase, along with a weak phase given by the Standard Model. The weak phase is different for the tree graph contributions and for the dominant penguin contributions. Further the weak phase of that penguin contribution cancels the weak phase of the mixing. Using the Unitarity of the CKM matrix the sub-dominant penguin contributions can be chosen to have the same weak phase as the tree amplitude; in all further discussion of phase structure these contributions are assumed to be included in the tree terms.[28] (Note however that in making numerical estimates these two types of contributions must be considered separately.)

The additional feature of this mode is that the full $B_d \to \pi^+\pi^-\pi^0$ is a sum of the three specific $\rho$-charge amplitudes. It thus contains known kinematically-varying strong phases that arise from the Breit-Wigner form of the $\rho$ resonances (more precisely stated from the $\pi\pi$ scattering phases shifts in the $\rho$ resonance region, which are parameterized by this form). Thus the amplitude for $B_d$ decay is given by

$$\begin{aligned} A(B_d \to \pi^+\pi^-\pi^0) &= f(k^+,k^0)(T^{+-} + P_1 + P_0) \\ &+ f(k^-,k^0)(T^{-+} - P_1 + P_0) + f(k^+,k^-0)(T^{00} - P_0) \end{aligned} \quad (7)$$

where $f(k_i, K_j)$ is the Breit Wigner function

$$\begin{aligned} f(k_i,k_j) &= \frac{cos(\theta)}{s - m_\rho^2 + i\Pi(s)} \\ \Pi(s) &= \frac{m_\rho^2}{\sqrt{s}}\left(\frac{p}{p_0}\right) 3\Gamma_\rho(m_\rho^2) \end{aligned} \quad (8)$$

where the $k_i$ are the momenta of the two pions, $s = (k_1 + k_2)^2$, and $\theta$ is the angle in the $\rho$ rest frame between $k_1$ and the direction opposite that of the boost from the $B$ rest frame. The function $\Pi(s)$ parameterizes the $\rho$ resonance shape. It is defined to give the correct threshold phase-space behavior and to incorporate the measured $\rho$ width, variations in the parameterization of this function are one of the sources of residual theoretical uncertainty of this analysis.

The angular dependence is that associated with the decay of a longitudinally-polarized $\rho$ meson of charge $(i + j)$ to two pions. The related amplitude for the $\overline{B}_d$ decay also contributes to the time-dependent rate for the decay of an initially pure $B_d$ or $\overline{B}_d$ state. Interference between the different $\rho$ bands is enhanced by the fact that the $\rho$ is longitudinally polarized and thus the $\cos(\theta)$ form for its decay throws the events towards the corners of the Dalitz plot. This is seen in Fig. 2, which is taken from the BaBar Book and represents a simulation using amplitudes calculated from a particular model.[29]

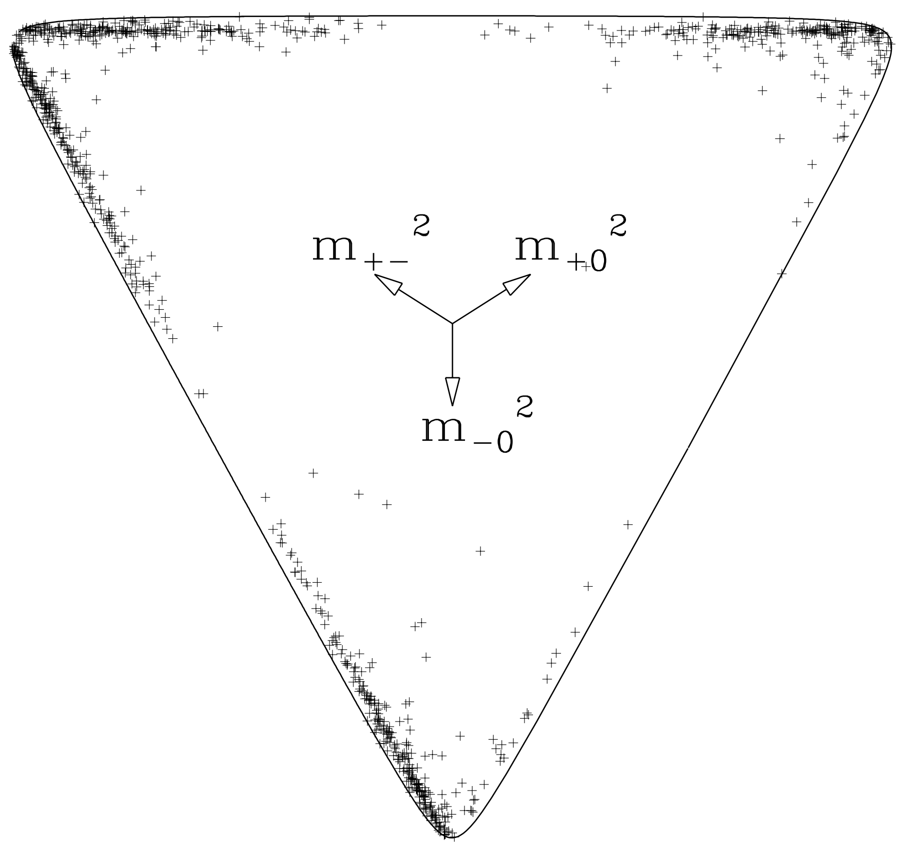



Fig. 2. The $\rho\pi$ contributions to the Dalitz plot for $B \to \pi^+\pi^-\pi^0$ .

The large strong phases from the resonant behavior and the interference of the different charge-channel contributions enhances the CP-violating asymmetry in the regions of the time-dependent Dalitz plot. A multiparameter maximum-likelihood fit to the broad $\rho$-band regions of the time-dependent Dalitz plot is made, with each tree and penguin amplitude parameterized by an arbitrary magnitude and strong phase, and with the weak phases as given by the Standard model. The asymmetries then depend only on one combination of weak phases, $\alpha = \pi - \beta - \gamma$ along with nine other parameters (the magnitude and strong phases of each of the five isospin amplitudes minus one irrelevant overall strong phase). In principal, provided the $\rho^0\pi^0$ contribution is large enough, this fit will allow one to extract not only a value of $\sin(2\alpha)$ free of uncertainties due to penguin contributions, but also $\cos(2\alpha)$, thereby removing some of the discrete ambiguities in the solution for the Unitarity triangle. In a realistic study additional parameters and assumptions must be made to parameterize non-resonant $B$ decays to three pions and also any other resonances that contribute significantly to the three-pion final state. It remains to be seen whether sufficient data can be collected to make this analysis effective when all the contributing channels and background contributions are taken into account. Certainly it will not be easy. It will require many years of data taking at a $B$ factory. Because the final state contains a $\pi^0$ this mode is not accessible to the current TeVatron experiments. Preliminary studies for dedicated hadron collider $B$ experiments suggest this mode may possibly be feasible for study, but further work on signal to background ratios is needed. I still hope that this mode can eventually give us a clean $\alpha$ measurement, but I recognize that the experimental challenge is significant. Some theoretical uncertainties in the value of $\alpha$ extracted in this way remain, due to the contribution of QED penguins, and also due to the assumed constant strong phases for the isospin amplitudes and the sensitivity to the $\rho$-shape. However these effects are estimated to be small. By the time this measurement is made I expect that their impact will be under much better control.

Isospin breaking effects must also be considered as a source of theoretical uncertainties when investigating these modes. The dominant correction comes from the fact that, due to isospin breaking of the quark masses, the physical $\pi^0$ and $\rho^0$ states each have a small admixture of the isospin zero quark combination. The consequence of this effect is largest in the $\pi\pi$ analysis as it reintroduces the $I_f = 1$ amplitude that is otherwise forbidden by Bose statistics. For the $\rho\pi$ channel the impact of isospin breaking has been estimated to be small.

## 2.2 SU(3) in $K\pi$ and $\pi\pi$ and limits on $\gamma$

Here I will briefly describe an analysis to extract the Unitarity triangle angle $\gamma$ from the data on various channels for $B \to K\pi$ and $B \to \pi\pi$. The work I will discuss is that of Neubert and Rosner,[30] and the subsequent paper of Neubert.[31] This analysis provides an interesting example because it uses essentially the entire toolkit of methods, the Operator Product Expansion, diagrammatic classification of contributions, isospin and SU(3), and finally factorization approximation as a way to estimate SU(3) breaking corrections. However a careful selection of the quantities for which the least accurate approximations are used leads to a relatively small theoretical uncertainty for the final result. The simple rule of thumb is that the tool with large fractional uncertainty should, if possible, be restricted to determining a small part of the overall result.

The decays $B \to K\pi$ are interesting because the tree contributions $b \to su\overline{u}$ are Cabibbo suppressed. In fact it appears that the rate is dominated by the QCD penguin contributions. However, as in the $\pi\pi$ case, certain isospin channels do not have any such contribution. The quark transition $b \to u\overline{u}s$ can have $\Delta I = 0, 1$ and thus with the spectator quark added $I_f = 1/2$ or $3/2$. The gluonic penguin contributes only to $\Delta I = 0, I_f = 1/2$. Here electroweak penguin contributions cannot be ignored, as they enter at approximately the same level as the Cabibbo-suppressed tree contributions, and for all isospin amplitudes. A major part of the work then comes in estimating the corrections due to electroweak penguin effects, and the uncertainty on these corrections.

The key to the analysis is to recognize that the $I_f = 3/2$ arises only from tree diagrams and electroweak penguins. The key initial observation is that, in terms of the isospin-based amplitudes $A_{\Delta I, I_f}$

$$\begin{aligned} A(B^+ \to \pi^+ K^0) &= A_{0,1/2} + A_{1,1/2} + A_{1,3/2} \\ -\sqrt{2}A(B^+ \to \pi^+ K^0) &= A_{0,1/2} + A_{1,1/2} - 2A_{1,3/2} \ . \end{aligned} \tag{9}$$

Gluonic penguin diagrams contribute only to $A_{0,1/2}$. Neubert and Rosner define the following quantities

$$R_* = \frac{Br(B^+ \to K^0\pi^+) + Br(B^- \to \overline{K}^0\pi^-)}{2(Br(B^0 \to K^0\pi^+) + Br(\overline{B}^0 \to K^-\pi^0))} = (1 - \Delta_*)^2 \ . \tag{10}$$

One can make the weak and strong phase dependence explicit by writing

$$A(B^+ \to K^0\pi^+) = Pe^{i\phi_P}(e^{i\pi} + e^{i\gamma}e^{i\eta}\epsilon_a) \tag{11}$$

where $\phi_P$ and $\eta$ are strong phases and, in terms of the diagrams,

$$\begin{aligned} P &= |\lambda_c(P_c - P_t - 1/3P_{EW,t})| \\ \epsilon_a &= 7|\lambda_u(P_u - P_c - A)|/P \ . \end{aligned} \tag{12}$$

Similarly one can write the ratio

$$-3A_{1,3/2}/P = \epsilon_{3/2} e^{\phi}_{3/2} (e^{i\gamma} + q e^{i\omega}) \tag{13}$$

expanded so that the weak and strong phase structure of each term is made explicit. The notation is chosen so that the quantities $P$, $q$, $\epsilon_a$ and $\epsilon_{3/2}$ are real and all phases are explicit. Here $qe^{i\omega}$ is the ratio of electroweak penguin type contributions to the tree type contributions to $A_{3/2}$. Only the top-type diagram gives a significant electroweak penguin contribution and that enters with a coefficient $\lambda_t = -\lambda_c - \lambda_u$ but the $\lambda_u$ contribution is dropped in the above as it is too small to matter here.

I find it convenient to introduce the quantities

$$\begin{aligned} r_{3/2} &= \frac{-3A_{1,3/2}}{A(B^+ \to K^0\pi^+)} = \frac{-\epsilon_{3/2} e^{\phi_{3/2} - \phi_P} (e^{i\gamma} + q e^{i\omega})}{1 - e^{i\gamma} e^{i\eta} \epsilon_a} \\ \xi &= \frac{1 - a(0+)}{1 + a(0+)} = \left| \frac{A(B^- \to \overline{K}^0 \pi^-)}{A(B^+ \to K^0 \pi^+)} \right|^2 . \end{aligned} \tag{14}$$

Then one can write

$$R_*^{-1} = \frac{|1 + r_{3/2}|^2 + \xi |1 + \overline{r}_{3/2}|^2}{1 + \xi} . \tag{15}$$

In the above equations the CP-conjugated amplitudes are obtained from their CP partners by simply changing the sign of the weak phase $\gamma$ everywhere (since $e^{i\pi} = e^{-i\pi}$).

A major point of introducing all this notation is that the quantities $\epsilon_a$, $\epsilon_{3/2}$ and $qe^{i\omega}$ are all small, the first two because they are suppressed by the ratio $|\lambda_u/\lambda_c|$ and the last because it is a ratio of electroweak penguin to tree, albeit enhanced by the inverse CKM ratio $|\lambda_c/\lambda_u|$. Useful results can be obtained keeping only the leading effects of these quantities. Relatively large uncertainties in these quantities translate into only small uncertainties in $R_*$.

This statement (which is mine, not Neubert's) is a bit of a cheat, since the sensitivity to $\gamma$ is not in the value of $R_*$ but in its deviation from 1, which is expected to be small for the same reason. The interest in this problem is sparked by preliminary data from CLEO which give $R_* = 0.47 \pm 0.27$. If the value of $R_*$ deviates significantly from 1 then the above equations can be used to put interesting constraints on the allowed range of gamma, provided we can constrain the quantities $\epsilon_a$, $\epsilon_{3/2}$ and $qe^{i\omega}$. The better we can constrain these parameters, the more likely we are to be able to determine whether beyond Standard Model physics is needed to explain the measurement. Further we will need some information on strong phase differences. However even generous ranges on these quantities may translate into constraints on the allowed range of gamma. So now let us pursue the question of how and how well we can calculate each of these quantities.

The quantities $qe^{i\omega}$ turns out to be cleaner than one would expect. In general two operators contribute for the tree amplitude and four for the electroweak penguin. However two of these latter four give very small contributions to this matrix element and can be neglected. The other two are Fierz-equivalent to the two tree-type operators. Furthermore only one linear combination of these two operators contributes in the SU(3) limit, the matrix element of the other must vanish. This is another application of Bose statistics, this time to the U-spin part of SU(3). Thus even though $qe^{i\omega}$ is a ratio of an electroweak penguin amplitude to a tree-type amplitude each is dominated by a single operator in the SU(3) limit. Furthermore and the two operators (for the two diagrams) are Fierz-equivalent to one-another. This means that only a single strong phase enters—the same for both these contributions, so that the ratio is fixed by the ratio of coefficients in this limit. Thus, Neubert writes $qe^{i\omega} = (1 - \kappa e^{i\Delta_{3/2}})\delta_+$ where $\delta_+$ is given by the ratio of coefficients of the $I = 3/2$ electroweak and tree operators that survive in the SU(3) limit and $\kappa e^{i\Delta_{3/2}}$ is the SU(3) breaking correction to this quantity. One can then estimate such corrections and the uncertainties in them. First one estimates SU(3) breaking correction $\kappa$ by calculating it in the factorization approximation. Neubert estimates this effect to be $(6 \pm 6)\%$. In this approximation $\Delta_{3/2} = 0$. This then gives $qe^{i\omega} \approx \delta_{EW} = (1 - \kappa)\delta_+ = 0.64 \pm 0.15$, where the large percentage error reflects the large theoretical uncertainties inherent in the SU(3) and factorization approximation as well as the smaller but still significant uncertainty in the evaluation of the ratio of operator coefficients that reflects small residual scale and scheme dependence of this ratio. He also includes the effect of allowing non-zero $\Delta_{3/2}$ values in this overall error estimation, noting that allowing a phase $|\Delta_{3/2}| \leq 90^\circ$ would yield only $|\omega| \leq 2.7^\circ$.

For the quantity $\epsilon_{3/2}$ one must again rely on SU(3), which relates the $B^+ \to K\pi, I = 3/2$ tree amplitude to the corresponding tree amplitude for $B^+ \to \pi\pi, I = 2$. The measured charged $B \to \pi K$ rates determines the magnitude of penguin amplitude in the denominator of epsilon, up to corrections of order $\epsilon_a$ which we will discuss below. Here one expects a large SU(3) correction. This is estimated again by calculating the correction in the factorization limit, taking the factorization model parameters $a_1^{ij}$ and $a_2^{ij}$ (where $ij = K\pi$ or $\pi\pi$) and the ratio $f_K/f_\pi$ from fits to data. The only model dependent part of this SU(3) correction calculation is the ratio $F(B \to K)/F(B \to \pi)$ which is 1 in the SU(3) limit. Models all agree with the range $1.1 \pm 0.1$. Since this factor enters the $\epsilon_{3/2}$ factorization calculation with a relatively small coefficient, the impact of its large uncertainty on the overall correction factor is not great. Again one must assign some uncertainty to the difference between the factorization-model based estimate of the SU(3) correction and the actual SU(3) breaking effects, but it is reasonable to expect that this estimate has correctly accounted for the largest part of SU(3) breaking corrections. Including this and all the various sources of uncertainty, both theoretical and experimental, Neubert estimates about a 25% uncertainty in the extracted value of $\epsilon_{3/2}$.

The remaining quantity $\epsilon_a$ is inherently small because it a ratio of Cabibbo-suppressed to Cabibbo-allowed terms. It would be a source of direct CP violation $\xi \neq 1$ and may eventually be constrained by measurement of the CP-asymmetry in $B^\pm \to K^\pm \pi^0$ decays. Another constraint comes from using SU(3) to relate these decays to the $B^\pm \to K^\pm K^0$ (or $\overline{K}^0$) decays. (For an alternate discussion of uncertainty introduced by this approach see M. Gronau and D. Pirjol[32]). Further, $\epsilon_a$ can be re-expressed in terms of a difference of $I = 1/2$ and $I = 3/2$ amplitudes that arises solely due to rescattering effects. Neubert uses all of these arguments to estimate a "reasonable" and a "conservative" (which in this context means a more generous) range for this quantity and then explores how the constraints on gamma vary as one varies $\epsilon_a$ over these ranges.

My point in describing this calculation is not to present the results, which you can read in Neubert's paper, and which indeed will change with time as experimental numbers improve. What I want to show is how the tools of SU(3) limit and factorization can be combined to obtain results which are better than either tool used separately. First the SU(3) limit prediction is calculated. Then the correction to that limit is calculated using the factorization approximation. Thus the uncertainty from factorization in the result is the uncertainty in the correction to SU(3) rather than the uncertainty in the entire effect. This is clearly an improvement over a straightforward use of either uncorrected SU(3) or simple factorization estimates to calculate the entire effect.

Even when such tricks are used to the full extent available still the question remains: how big is the uncertainty in the result after all is said and done? Unfortunately the answer is never clean. But clearly the problem is much reduced if we are debating whether an effect is 6% or twice as big rather than whether it is 50% or twice that. The challenge to theorists is to make the sources of their uncertainties clear, and to do as honest a job as possible of constraining them. Here work remains to be done. Neubert's paper gives an example of a serious attempt to explore such questions in a systematic way, for a particular set of decays.

## 2.3 Theoretical Uncertainties

In the end, whatever the estimates might be, it is important to remember that theoretical uncertainty is not statistical; it is simply wrong to talk about the probabilities of certain results as if these estimates were in fact gaussian-based standard deviations. It is also very misleading to combine different sources of theoretical error by adding them in quadrature, though one sees this done frequently in the literature.

A false division between theoretical uncertainties and systematic errors in an experimental value is often made—at least in the minds of theorists making the initial predictions. A theorist makes a clean prediction with small theoretical errors for a quantity—say, for example, the CP-violating asymmetry in inclusive $b \to u\overline{u}d$ decays. The theorist is happy. However that quantity is in fact impossible to measure, since any real experiment has aperture limitations and in addition must apply cuts to separate the signal from background, in the example above both that from sources other than $B$-decays and that from the dominant $b \to c\overline{q}q'$ decays. The impact of these cuts on the relationship of the measurement to the prediction must be evaluated based on some theoretical models. This is where the large theoretical errors will typically appear.

Experimentalists now often quote their uncertainties by separating out such effects as theoretical uncertainties rather than by including them in the overall systematic uncertainties. My point here is that the magnitude of this theoretical uncertainty typically will have nothing to do with the magnitude of the theoretical uncertainty for this measurement given in the original theoretical predictions. Such experiment-dependent theoretical uncertainties belong neither to the domain of pure theory nor to the domain of experiment, but live at the interface between them. They do, however, suffer the usual disease of theoretical errors—they are not statistical effects. It would be very helpful if theorists making their clean predictions could at least consider and briefly discuss what impact experimental cuts will have on the validity of their prediction. I do not mean the theorist should define specific cuts, but rather should discuss the question of whether the result can survive any cuts at all without serious degradation.

My remarks above are borne out in a well-known way in the case of the extraction of the magnitude of the CKM-parameter $V_{ub}$ from semileptonic $B$-decay data. Two general classes of methods are in the market—one uses exclusive decays and has obvious theoretical uncertainties related to the form factors (that is the QCD matrix elements) that govern the particular decay in question. The other uses inclusive semi-leptonic decays and is thus at first sight very clean. But the experiment must make cuts to remove the $b \to cl\nu$ backgrounds. The prediction for the cut data sample has comparable theoretical uncertainties to the exclusive decay cases. Eventually we need to explore both kinds of methods, since the theoretical uncertainties for the two approaches are essentially different.

Recently new predictions for extracting this same parameter from hadronic measurements have appeared. Again one group of theorists advocates an inclusive approach, and others advocate certain exclusive channels. Both are interesting; both will probably have significant theoretical errors once the real experimental limitations on the inclusive methods are understood.

In all these cases, whether semi-leptonic or hadronic decays are considered, one cannot use any of the more rigorous tools discussed above to estimate the theoretical uncertainties introduced due to experimental cuts or those due to form-factor estimates. One is forced to resort to models. Often the models work at the quark rather than the hadron level and then apply the notion of quark-hadron duality which is the assumption that the two-body hadron kinematics reflects the underlying quark kinematics. This is called "local quark-hadron duality". It is not a justifiable assumption.

Estimates of theoretical errors in such cases tend to be very subjective. There really is no clean way to obtain them. The most common method is to try a few models and take the range of the results as the range of theoretical uncertainties. This is risky, since all the models on the market may contain the same unjustified assumption (for example that a particular form factor can be parameterized as a simple pole). Nonetheless it is common practice and perhaps the best we can do.

My advice is one should simply be aware when this is the nature of the theoretical error estimate and treat the resulting numbers with a sufficient amount of salt. The recent history of statements about errors in estimates of $\epsilon'/\epsilon$ should be a clear object lesson to experimenters on the reliability of theoretical error estimates.

## Acknowledgments

I would like to thank João Silva for a careful reading of these lectures and his helpful suggestions for improvements.

## References

[1] Y. Nir Lectures, Lectures at SLAC Summer Institute (1999), in this volume.

[2] A. J. Buras, Lectures at Lake Louise Winter Institute Feb. 1999, hep-ph/9905437 (1999).

[3] G. Branco, L. Lavoura and J. Silva "CP Violation", Oxford University Press (1999).

[4] A. F. Falk, Chapter 2 of 'The BaBar Physics Book: Physics at an Asymmetric $B$ Factory', (SLAC report SLAC-R-504), hep-ph/9812217.

[5] K.G. Wilson, Phys. Rev. **179**, 1499 (1969).

[6] J. R. Dell'Aquila and C.A. Nelson, Phys. Rev. **D33**, 101 (1986); B. Kayser *et al.*, Phys. Lett. **B237**, 3339 (1990); I. Dunietz *et al.*, Phys. Rev. **D43**, 2193 (1991).

[7] M. Beneke, G. Buchalla, and I. Dunietz, Phys. Lett. **B393**, 123 (1997).

[8] S. Aoki, Talk at *22nd International Lepton Photon Conference*, Stanford August (1999); Stephen R. Sharpe, Talk at *29th ICHEP 98*, Vancouver, Canada, 23-29 Jul 1998, hep-lat/9811006.

[9] M. Neubert, Talk HEP 97, Jerusalem, Israel (1997), hep-ph/9801269.

[10] M. J. Dugan and B. Grinstein, Phys. Lett. **B255**, 583 (1991); J. F. Donoghue and A. A. Petrov, Phys. Lett. **B393**, 149 (1996).

[11] For the effects of isospin breaking in $B$ decays see, for example, S. Gardner hep-ph/9809479; hep-ph/9906269.

[12] R. Fleischer, Z. Phys. **C62**, 81 (1994); M. Gronau, O. F. Hernandez, D. London, J. L. Rosner, Phys. Rev. **D52**, 6374 (1995); A. J. Buras, R. Fleischer, Phys. Lett. **B365**, 390 (1996), R. Fleischer, Phys. Lett. **B365**, 399 (1996).

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

[14] K. M. Watson, Phys. Rev. **88**, 1163 (1952); R. N. Cahn and M. Suzuki, hep-ph/9708208 (1997); M. Suzuki and L. Wolfenstein , Phys. Rev. **D60**, 074019 (1999).

[15] See, for example, M. Gronau, J. L. Rosner, and D. London, Phys. Rev. Lett. **73**, 21 (1994); M. Gronau, O. F. Hernandez, D. London, and J. L. Rosner, Phys. Rev. **D50**, 4529 (1994).

[16] M. Gronau and J. L. Rosner, Phys. Rev. Lett. **76**, 1200 (1996).

[17] For example, see M. B. Wise, Phys. Rev. **D45**, 2188 (1992); G. Burdman and J. F. Donoghue, Phys. Lett. **B280**, 287 (1992); H.-Y. Cheng, C.-Y. Cheung, W. Dimm, G.-L. Lin , Y.C. Lin, T.-M. Yan, H.-L. Yu, Phys. Rev. **D48**, 3204 (1993).

[18] M. A. Shiftman, A. I. Vainshtein, V. I. Zakharov, Nucl. Phys. **B147**, 385 (1979); L. J. Reinders, H. Rubinstein and S. Yazaki, Physics Reports **127**, 1 (1985); P. Ball, V. M. Braun Phys. Rev. **D55**, 5561 (1997); BaBar Physics Book, Appendix D, SLAC Report 504 (1998); P. Ball, V. M. Braun, Phys. Rev. **D58**, 094016 (1998).

[19] H. Lehmann K. Symanzik and W. Zimmerman, Nuovo Cimento **1**, 1425 (1955).

[20] S. Sharpe, private communication.

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

[22] J. F. Donoghue, hep-ph/9607352 (1996); M. Gronau , J. L. Rosner, Phys. Lett. **B439**, 171 (1998); M. Gronau, J. L. Rosner Phys. Rev. **D58**, 113005 (1998); A. F. Falk, A. L. Kagan, Y. Nir, A. A. Petrov, Phys. Rev. **D57**, 4290 (1998); A. F. Falk, Talk at QCD 98, hep-ph/9806538; Alexey A. Petrov, Talk at Kaon 99, hep-ph/9909312.

[23] M. Beneke, G. Buchalla, M. Neubert, C. T. Sachrajda, Phys. Rev. Lett. **83**, 1914 (1999).

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

[25] CLEO Conference Report 99-14, Y. Kwon *et al.*, see www.lns.cornell.edu/public/CONF/1999.

[26] BaBar Physics Book, SLAC Report 504 (1998), Paul Harrison and Helen Quinn editors, Chapter 6.

[27] H. R. Quinn and A. E. Snyder, Phys. Rev. **D48**, 2139 (1993).

[28] H. Quinn in Review of Particle Physics, European Physical Journal C, Volume 3 1-4, p. 558 (1998).

[29] J. Charles, Phys. Rev. **D59**, 054007 (1999).

[30] Matthias Neubert and Jonathan L. Rosner, Phys. Lett. **B441**, 403 (1998); Matthias Neubert and Jonathan L. Rosner, Phys. Rev. Lett. **81**, 5076 (1998).

[31] Matthias Neubert , SLAC-PUB-8122, to be published in the *Proceedings of 17th International Workshop on Weak Interactions and Neutrinos (WIN 99)*, Cape Town, South Africa, 24-30 Jan 1999, hep-ph/9904321.

[32] Michael Gronau and Dan Pirjol, CLNS-99-1604 (1999), hep-ph/9902482.

# CP VIOLATION IN ATOMIC AND NUCLEAR PHYSICS

E. D. Commins
Physics Dept., University of California
and
E. O.Lawrence Berkeley National Laboratory
Berkeley, California 94720

ABSTRACT

CP violation and CPT invariance imply T- violation, which should appear at some level in atomic and nuclear phenomena. In this report we briefly summarize various experiments to search for T-violating strong interactions and weak decays, and discuss in some detail various searches for electric dipole moments of elementary particles, nuclei, atoms, and molecules.

# 1 Introduction

CP violation and CPT invariance imply T-violation, which is expected to occur at some level in atomic and nuclear phenomena. For example if both P and T are violated, an elementary particle, nucleus, atom, or molecule can possess a permanent electric dipole moment (EDM). This happens because CP violation and the usual weak interaction must induce EDM's at some level by means of radiative corrections to the P,C,T conserving electromagnetic interaction. Searches for EDM's began 50 years ago, 6 years before the discovery of P-violation in the weak interactions, 7 years before it was realized that EDM's violate T-invariance, and 14 years before the discovery of CP-violation in kaon decay. EDM searches, which have so far yielded only upper limits, continue unabated today. In the aftermath of the discovery of parity violation in 1956-57, several other types of experiments to search for T-violation were launched: ”direct” searches for T-violation in strong interactions, ”triple correlations” in nuclear beta decay, etc. These have also yielded only upper limits so far. Thus one may say that progress in our field is glacial, at best. In this report we devote most of our attention to electric dipole moments, since in the general view this is the area where positive results are most likely to emerge. But we first summarize very briefly the direct tests of T-invariance in strong interactions, and the triple correlation measurements in nuclear beta decay.

Before we begin our survey, however, let's recall what time reversal invariance means. The (anti-unitary) time reversal operation reverses all momenta and spins, and interchanges initial and final states. Consider one or more particles in an initial state A with linear momenta $\mathbf{p}_i$ and spins $\mathbf{s}_i$. Suppose this system undergoes a transition (by decay or collision) to a final state B consisting of particles with linear momenta $\mathbf{p}_f$, $\mathbf{s}_f$. We can describe the transition by an S-matrix element:

$$S_{BA} = \delta_{BA} - iK_{BA} \tag{1}$$

where K is the transition operator:

$$K_{BA} = \langle B(\mathbf{p}_f, \mathbf{s}_f)|K|A(\mathbf{p}_i, \mathbf{s}_i)\rangle \tag{2}$$

Thus the time-reversed transition matrix element is:

$$K'_{AB} = \langle A(-\mathbf{p}_i, -\mathbf{s}_i)|K|B(-\mathbf{p}_f, -\mathbf{s}_f)\rangle \tag{3}$$

By definition, time reversal invariance holds for the reaction if

$$|K'_{AB}| = |K_{BA}| \tag{4}$$

## 2 Direct tests of T-invariance in the Strong Interaction

Experimental tests of the discrete symmetries P, C, and T in the strong interaction have very limited precision. From numerous studies of parity violation in nuclear forces (a phenomenon that is attributed to the charged- and neutral-weak interactions of nucleons), we can say that P is conserved in strong interactions to about 10 ppm.[1] Comparison of the rates of certain strong reactions such as $p\overline{p} \to \overline{K^0}K^+\pi^-$ and $p\overline{p} \to K^0K^-\pi^+$ tells us that the strong interaction is C-conserving to about 1% precision.[2] As for time reversal, the most direct experimental tests involve measurement of the differential cross sections for a certain nuclear reaction and its inverse[3]:

$$^{27}Al + p \to ^{24} Mg + \alpha, \tag{5}$$

studies of the statistics of compound nucleus energy level distributions,[4] and observation of the transmission of low-energy polarized neutrons through a target of aligned $^{165}$Ho nuclei.[5] In the latter case one searches for the correlation $\boldsymbol{\sigma} \cdot (\mathbf{k} \times \mathbf{I})\mathbf{k} \cdot \mathbf{I}$ where $\boldsymbol{\sigma}$, $\mathbf{k}$, and $\mathbf{I}$ refer to the neutron spin, neutron momentum, and nuclear spin respectively. In all of these cases one is looking for a T-violating, P-conserving (TVPC) effect. The result of these experiments is expressed in terms of the ratio $\alpha_{\mathrm{T}}$ of the TVPC transition matrix element to the usual T, P conserving strong interaction matrix element. One is able to obtain only a very crude limit: $\alpha_{\mathrm{T}} \lesssim 3 \cdot 10^{-3}$, which corresponds to a mass scale $\Lambda_{\mathrm{T}} \gtrsim$10 GeV for new TVPC physics beyond the standard model. Clearly, 10 GeV is a very low mass scale, and intuitively it would seem to be very unlikely.

Recently Ramsey-Musolf[6] has shown that a far more stringent limit: $\Lambda_{\mathrm{TVPC}} \gtrsim$150 TeV, corresponding to $\alpha_{\mathrm{T}} \lesssim 10^{-15}$, can be obtained indirectly by combining the upper limits obtained from neutron and electron electric dipole moment experiments (to be discussed below) with an analysis of radiative corrections based on the standard model. The basic idea is that P, C violating standard model weak interactions at the radiative correction level would necessarily combine with C, T violating strong interactions, if they were to exist, to produce P, T violating EDM's.

## 3 Triple correlations in weak decays

How can we test for T-violation in weak decays? Let us consider neutron beta decay as a typical example:$n \to pe^-\overline{\nu_e}$. Obviously it is impossible to check time reversal invariance by running this reaction backward! However, there is a useful indirect alternative.

Consider once again the S operator:

$$S = I - iK \tag{6}$$

Since S is unitary, we have:

$$\begin{aligned} I &= SS^\dagger \\ &= (I - iK)(I + iK^\dagger) \\ &= I - iK + iK^\dagger + KK^\dagger \end{aligned}$$

If K is sufficiently small, we can neglect $KK^\dagger$, in which case $K = K^\dagger$. Then,

$$K_{BA} = \langle B(\mathbf{p}_f, \mathbf{s}_f)|K|A(\mathbf{p}_i, \mathbf{s}_i)\rangle = \langle A(\mathbf{p}_i, \mathbf{s}_i)|K|B(\mathbf{p}_f, \mathbf{s}_f)\rangle^* \tag{7}$$

Meanwhile from (3) the time reversed transition amplitude is given by:

$$K'_{AB} = \langle A(-\mathbf{p}_i, -\mathbf{s}_i)|K|B(-\mathbf{p}_f, -\mathbf{s}_f)\rangle \tag{8}$$

Therefore, we have time reversal invariance if:

$$|\langle B(-\mathbf{p}_f, -\mathbf{s}_f)|K|A(-\mathbf{p}_i, -\mathbf{s}_i)\rangle| = |\langle B(\mathbf{p}_f, \mathbf{s}_f)|K|A(\mathbf{p}_i, \mathbf{s}_i)\rangle| \tag{9}$$

provided, always, that we can assume $K = K^\dagger$. Of course for the weak interaction itself this is a good assumption, but the matrix element refers to initial and final *physical states*, and we must remember that the final states are always affected to some extent by *final state interaction*. Thus result (9) is valid only to the extent to which we can neglect final state corrections.

With this in mind, let us consider neutron decay as viewed in the neutron rest frame. Fig.1a shows a neutron with spin up (out of the page) decaying to a proton, electron, and anti-neutrino (with linear momenta in the plane of the page). For simplicity we shall ignore the spins of the final particles. Suppose the amplitude corresponding to Fig.1a is $M$. In Fig. 1b, the momenta and spins are the same, but initial and final states are reversed. The corresponding amplitude is $M^*$. Now in Fig. 1c we perform the time reversal operation on the system of Fig. 1b. This means that momenta and spins are reversed, and initial and final states are once more interchanged. The corresponding amplitude is $M^{*\prime}$. Notice that apart from a 180 deg rotation about the neutron spin axis, the final particle momenta are the same as in Fig. 1a, but the neutron spin is reversed. If time reversal invariance holds, the probabilities for Figs. 1a and 1c should be the same;

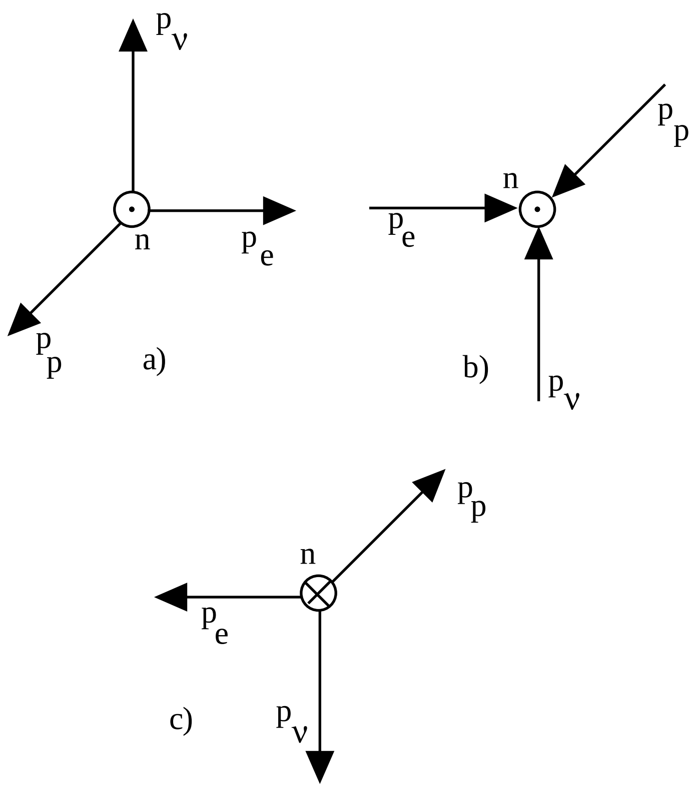



Fig. 1. a) A neutron with spin up (out of page) decays to proton,e$^-$, and $\overline{\nu}$ with momenta in plane of page. Amplitude=$M$. b) Initial and final states interchanged; amplitude=$M^*$. c) Process of 1b time-reversed: Amplitude $M^{*\prime}$.

in other words, there should be no observable correlation of the form:$\boldsymbol{\sigma}_n \cdot \mathbf{p}_e \times \mathbf{p}_{\overline{\nu}}$. Now let's consider the same thing in more detail. Years ago, Jackson, Treiman, and Wyld[7] showed that, neglecting nuclear recoil and final state interactions, the differential transition probability per unit time for beta decay of polarized nuclei with spin 1/2 in the allowed approximation, summed over final spins, is given by:

$$\begin{aligned} dW = & \frac{G_F^2}{(2\pi)^5} \cos^2 \theta_C \delta(E_e + E_\nu - \Delta) \\ & \times F(Z, E_e) d^3\mathbf{p}_e d^3\mathbf{p}_\nu \\ & \times \xi[1 + a\frac{\mathbf{p}_e \cdot \mathbf{p}_\nu}{E_e E_\nu} + \frac{\langle \mathbf{J}_i \rangle}{J_i}(A\frac{\mathbf{p}_e}{E_e} + B\frac{\mathbf{p}_\nu}{E_\nu} + D\frac{\mathbf{p}_e \times \mathbf{p}_\nu}{E_e E_\nu})] \end{aligned} \qquad (10)$$

where $\hbar = c = 1$, $G_F$ is Fermi's constant, $\theta_C$ is the Cabibbo angle, $\Delta$ is the maximum electron energy, $F(Z, E_e)$ is the Coulomb correction factor, and $\mathbf{J}_i$ is the initial nuclear spin. Also,

$$\xi = |C_V|^2|\langle 1\rangle|^2 + |C_A|^2|\langle\sigma\rangle|^2 \tag{11}$$

and $C_{V,A}$ are the vector and axial vector coupling constants, respectively; $\langle 1\rangle$ and $\langle\sigma\rangle$ are the Fermi and Gamow-Teller reduced matrix elements, respectively; a is the neutrino-electron momentum correlation coefficient, A is the beta asymmetry parameter, B is the analogous parameter for the neutrino, and:

$$\xi D = i(C_V C_A * - C_A C_V *)\langle 1\rangle\langle\sigma\rangle(\frac{J_i}{J_i+1})^{1/2} \tag{12}$$

It is the ”triple correlation” coefficient D that is of interest here. In the approximation where we neglect final state interaction, D can only be non-zero if there is T-violation since T-invariance requires that $C_V$ and $C_A$ be relatively real.

The D coefficient has been studied experimentally in neutron decay and in the superallowed beta decay $^{19}Ne \rightarrow ^{19}F + e^+ + \nu_e$. The only significant final state correction to D in each of these decays arises from interference between the Coulomb interaction of the final nucleus and emitted electron, and the ”weak magnetism” portion of the weak vector amplitude.[8] It can be shown that:

$$\begin{aligned} D_n^{WM} &= \frac{E_e^2}{p_e m_n}(-0.032 + 0.040\frac{m_e^2}{E_e}) \\ &= -5.7\cdot 10^{-5}\, at\, max\, p_e \end{aligned} \tag{13}$$

and

$$D_{neon}^{WM} \approx 3\cdot 10^{-4} \tag{14}$$

The present experimental limits on D are summarized in Table 1. The recent N.I.S.T. neutron experiment was designed to yield a substantial improvement in precision compared to the older I.L.L. and Kurtchatov results, but due to various experimental difficulties, the results have so far been disappointing. Theoretical limits on D (see Table 2) are very-model dependent. According to the standard model, and to a number of currently popular models for physics beyond the standard model, D is far too small to be observed in any practical experiment, now or in the foreseeable future. The only existing models that can yield predictions for D close to present experimental limits are those that invoke the existence of leptoquarks.

We have just considered experiments in nuclear beta decay where one sums over all final spins. However, if one can detect the spin of the final electron or positron, the following T-odd terms in the transition probability are observable:

$R\boldsymbol{\sigma}_n \cdot (\boldsymbol{\sigma}_e \times \boldsymbol{p}_e)$

$L\boldsymbol{\sigma}_e \cdot (\mathbf{p}_e \times \mathbf{p}_\nu)$.

The correlation coefficient L has never been measured in any decay. R has been observed in $^8$Li decay:

$$R(^8Li) = (0.9 \pm 2.2) \cdot 10^{-3} \tag{15}$$

and a measurement in neutron decay has been discussed. However, R=0 for pure V, A interaction regardless of the relative phase of V and A amplitudes. Non-zero R requires interference between scalar and axial vector, and/or interference between tensor and vector couplings. Khriplovich[9] has shown that indirect limits on T-odd scalar and tensor couplings in beta decay obtained from neutron and electron EDM results, together with calculated standard model radiative corrections, are far more sensitive than the limits that can be obtained from direct measurements in beta decay.

Table 1. Measurements of D

| Initial nucleus | D | Ref. |
|---|---|---|
| neutron | $(-1.1 \pm 1.7) \cdot 10^{-3}$ (I.L.L.) | 10 |
| neutron | $(2.2 \pm 3.0) \cdot 10^{-3}$ (Kurtchatov) | 11 |
| neutron | $(-0.1 \pm 1.3 \pm 0.7) \cdot 10^{-3}$ (N.I.S.T.) | 12 |
| $^{19}$Ne | $(4 \pm 8) \cdot 10^{-4}$ (Princeton) | 13 |

Table 2. Theoretical limits on D

| Theoretical model | D |
|---|---|
| Standard model | $< 10^{-12}$ |
| Theta-QCD | $< 10^{-12}$ |
| Supersymmetry | $< 10^{-6}$ |
| Left-right symmetry | $< 10^{-4}$ |
| Exotic fermion | $< 10^{-4}$ |
| Leptoquark models | Limited by expt. |

Finally, consider the decay $K^+ \rightarrow \pi^0 \mu^+ \nu_\mu$. The ”transverse muon polarization” $P_T$ in this decay is the component of muon polarization perpendicular to the plane of the final momenta. Observation of $P_T$, that is, observation of a triple correlation of the form $\boldsymbol{\sigma}_\mu \cdot \mathbf{p}_\pi \times \mathbf{p}_\mu$, would constitute evidence for T-violation, since at the level of precision attainable with present methods, final-state corrections are negligible. According to the standard model, $P_T$ should be $\approx 10^{-7}$. Very recently, the following experimental result was obtained[14]:

$$P_T = -0.0042 \pm 0.0049(stat) \pm 0.0009(syst) \tag{16}$$

# 4 General remarks about electric dipole moments

We have already mentioned that an elementary particle, nucleus, atom, or molecule can only possess an electric dipole moment (EDM) if both P and T are violated. This can be seen[15] in a very simple way as follows. Consider a particle of spin 1/2 and assume that it has an EDM $\mathbf{d}$ as well as a spin magnetic dipole moment $\boldsymbol{\mu}$. Both moments lie along the spin direction, because the spin is the only vector available to orient the particle. We write the Hamiltonians $H_M, H_E$ that describe the interaction of $\boldsymbol{\mu}$ with a magnetic field $\mathbf{B}$, and of $\mathbf{d}$ with an electric field $\mathbf{E}$, in the non-relativistic limit:

$$H_M = -\boldsymbol{\mu} \cdot \mathbf{B} = -\mu \boldsymbol{\sigma} \cdot \mathbf{B} \tag{17}$$

$$H_E = -\mathbf{d} \cdot \mathbf{E} = -d \boldsymbol{\sigma} \cdot \mathbf{E} \tag{18}$$

where $\boldsymbol{\sigma}$ is the Pauli spin operator. Under space inversion (P) the axial vectors $\boldsymbol{\sigma}$ and $\mathbf{B}$ remain unchanged, but the polar vector $\mathbf{E}$ changes sign. Hence under P, $H_M$ is invariant, while $H_E$ is not. Under time reversal, $\boldsymbol{\sigma}$ and $\mathbf{B}$ change sign, while $\mathbf{E}$ remains unchanged. Hence under T, $H_M$ is once again invariant, but $H_E$ changes sign.
Note that the P,T-violating EDMs of interest to us are not at all the same as the so-called ”permanent” electric dipole moments of polar molecules, so well known to chemists and molecular spectroscopists. The latter are not really permanent at all and involve no violation of P or T. A very clear quantum-mechanical description of their behavior was already given in 1931 by Penney.[16]

It is impossible to form an intuitive classical picture of the EDM of an elementary particle (e.g. an electron); but it is just as impossible to make a classical model of the spin magnetic moment. Of course, the latter has become very familiar after 75 years, and we all have the illusion that we understand it thoroughly. We know that the Dirac equation for the electron contains the implication that the electron possesses a ”normal” spin magnetic moment with the g-value: $g_S = 2$. We also know that a spin 1/2 particle with an anomalous magnetic moment can still be described by a Dirac-like equation provided we add a gauge-invariant, Lorentz-invariant Pauli moment term. In field theory this is described by the well-known Lagrangian density:

$$\mathcal{L}_{Pauli} = -\kappa \frac{\mu_B}{2} \overline{\Psi} \sigma^{\mu\nu} \Psi F_{\mu\nu} \tag{19}$$

Here $\Psi$ is the Dirac field for the fermion in question, $\overline{\Psi}$ is the Dirac conjugate field, $\sigma^{\mu\nu} = \frac{i}{2}(\gamma^\mu \gamma^\nu - \gamma^\nu \gamma^\mu)$ where $\gamma^\mu, \gamma^\nu$ are the usual 4x4 matrices, $F_{\mu\nu}$ is the electromagnetic field tensor, $\mu_B$ is the Bohr magneton, and $\kappa$ is an appropriate constant. This Lagrangian density is of course invariant under P and T. To describe an EDM $\mathbf{d}$ we construct an analogous P-odd, T-odd Lagrangian density by making the replacements: $\sigma^{\mu\nu} \rightarrow \gamma^5 \sigma^{\mu\nu}$ and $\kappa \mu_B \rightarrow id$ where $i = (-1)^{1/2}$ is necessary so that the resulting Hamiltonian shall be Hermitian.Thus[17] we obtain:

$$\mathcal{L}_{EDM} = -i \frac{d}{2} \overline{\Psi} \gamma^5 \sigma^{\mu\nu} \Psi F_{\mu\nu} \tag{20}$$

From this Lagrangian density, one can easily obtain the following single-particle EDM Hamiltonian:

$$H_{EDM} = -d\gamma^0 \mathbf{\Sigma} \cdot \mathbf{E} + id\boldsymbol{\gamma} \cdot \mathbf{B} \tag{21}$$

where

$$\mathbf{\Sigma} = \begin{pmatrix} \boldsymbol{\sigma} & 0 \\ 0 & \boldsymbol{\sigma} \end{pmatrix}$$

In the non-relativistic limit the first term on the right hand side of (21) reduces to the right hand side of (18). However, in general the factor $\gamma^0$ is extremely important, as we shall see. The second term on the right hand side of (21) gives no contribution in the non-relativistic limit.

The EDM Lagrangian density $\mathcal{L}_{EDM}$ is not renormalizable, and yields an EDM only by virtue of loop corrections to the usual fermion-photon interaction of quantum electrodynamics. The possible loop corrections are naturally very model-dependent and uncertain. For, while the parameters that describe CP violation in kaon decay have been measured ever more precisely over the years, the fundamental explanation for CP violation remains very much in doubt, and a wide variety of plausible, if very speculative, models of CP violation have been presented. According to the standard model, the electron EDM $d_e$, the neutron EDM $d_n$, and other EDM's as well are far too small to be detected by any possible experiment, now or in the foreseeable future. However, in various extensions of the standard model (See Table 3) $d_e$ and/or $d_n$ are sufficiently large to be detected by practical experiments. In some models, even the muon EDM $d_\mu$ and/or the tau EDM $d_\tau$ might be detectable, although less sensitive experimental methods are available for them than for $d_e$ or $d_n$.

Table 3. Theoretical predictions for neutron, electron EDMs

| CP Viol Model | Neutron (e cm) | Electron (e cm) |
|---|---|---|
| Standard | $10^{-32}$-$10^{-34}$ | $\approx 10^{-40}$ |
| Minimal SUSY | $10^{-25}$-$10^{-26}$ | $10^{-26}$-$10^{-28}$ |
| SUSY GUT [SO(10)] | $10^{-25}$-$10^{-27}$ | $10^{-26}$-$10^{-28}$ |
| L-R Symmetric | $10^{-25}$-$10^{-26}$ | $10^{-26}$-$10^{-28}$ |
| Multi-Higgs | $10^{-25}$-$10^{-26}$ | $10^{-26}$-$10^{-28}$ |
| Lepton-flavor changing | — | $10^{-27}$-$10^{-29}$ |

Why does the standard model predict that $d_e$ and $d_n$ are so small? In the standard model, the neutron EDM could arise from valence quark EDMs but this cannot occur at the one-loop level. At the two-loop level individual diagrams do have complex phases that contribute to $d_n$ but it has been shown[18] that the sum of these diagrams over all quark flavors yields zero. Thus, according to the standard model, the neutron EDM appears only at the 3-loop level, and even here there are suppressions and cancellations.[19]

In the standard model with massless neutrinos there is no analog to the CKM matrix in the lepton sector, and thus no analogous way to generate CP violation. For the electron EDM to arise here we would require coupling to virtual quarks via virtual $W^\pm$. Naively one might expect a contribution from the two-loop diagram of Fig.2. However, for each contribution $V_{ij}$ from the CKM matrix at one vertex, there is a contribution $V_{ij}^*$ at the other vertex; hence the overall amplitude cannot contain a CP violating phase. (This is also the reason why there is no neutron EDM at the one-loop level).

Next, one can consider contributions to the electron EDM at the 3-loop level.This situation was first analyzed by Hoogeveen[20] but it was subsequently shown by Pospelov and Khriplovich[21] that the various 3-loop diagrams cancel, yielding a net contribution of zero in the absence of gluonic corrections to the quark lines (see Fig. 3).

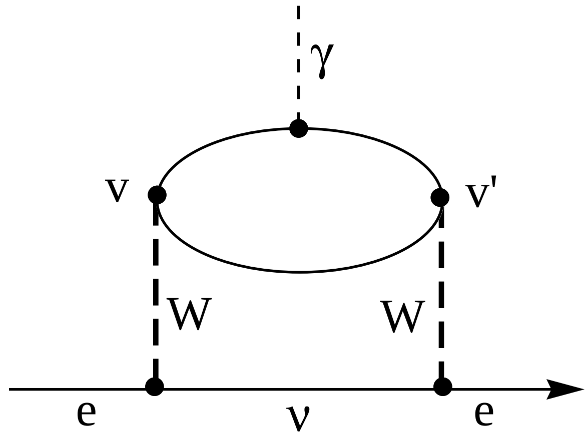



Fig. 2. This two-loop diagram cannot contribute to the electron EDM, because although a factor $V_{ij}$ from the CKM matrix appears at vertex v, a factor $V^*_{ij}$ appears at vertex v'. Thus, there is no net CP-violating phase.

# 5 The neutron EDM

In neutron EDM experiments one employs Ramsey's method[22] of separated oscillating magnetic fields to observe magnetic resonance on the neutron spin-magnetic moment in a fixed static magnetic field $\mathbf{B}$. An electric field $\mathbf{E}$ parallel or anti-parallel to $\mathbf{B}$ is inserted between the oscillating fields. The basic idea is that the interaction $H_E = -\mathbf{d} \cdot \mathbf{E}$ might cause a detectable frequency shift in the magnetic resonance fringe pattern. In the earliest neutron EDM experiments this was done with neutron beams, and the oscillating fields were separated spatially. Modern neutron EDM experiments make use of ultra-cold (UC) neutrons (which typically have kinetic energies of $10^{-7} eV$ or less.) Such experiments take advantage of the fact that UC neutrons undergo critical reflection at any angle of incidence on suitable materials, and can thus be stored in closed vessels. Here, neutron storage bottles with collinear $\mathbf{E}$ and $\mathbf{B}$ fields, and oscillating fields separated *temporally* rather than spatially, are employed. Fig. 4 shows the apparatus employed for the 1990 neutron EDM experiment at Institut Laue-Langevin (I.L.L.) in Grenoble.[23] Ultra-cold neutrons were transported from the source through a guide tube and polarized by transmission through a magnetically saturated $1\mu$ thick iron-cobalt

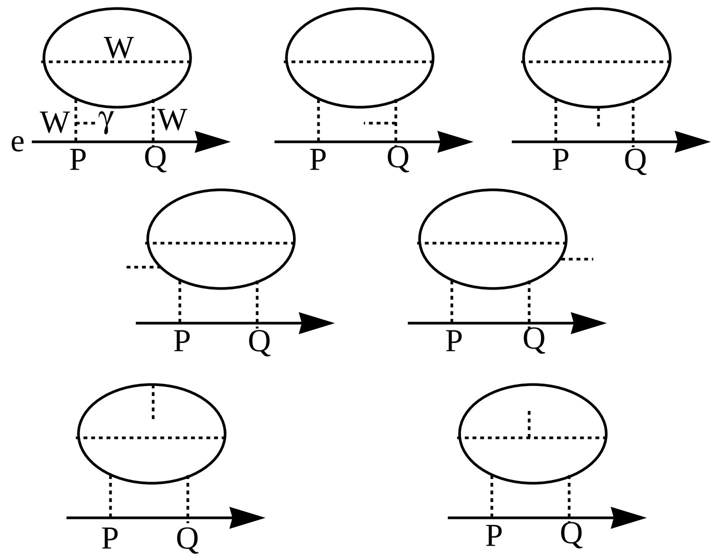



Fig. 3. The sum of the contributions to $d_e$ from these 3-loop diagrams is zero, according to the standard model. If each diagram is disconnected from the lepton line at P,Q one has the 2-loop contributions to the EDM of an (on-mass-shell) W boson. Thus the EDM of a W boson is zero in the 2-loop approximation.

foil. The 5 liter storage bottle was enclosed by two beryllium electrodes 25 cm in diameter and a 10 cm long insulating cylinder of BeO. The static magnetic field was .01 Gauss in magnitude, and the electrodes were used to generate an electric field of 16000 V/cm. Neutrons could be admitted to the bottle by means of a beryllium valve in the grounded electrode. The experiment functioned as follows: the interaction volume was filled with UC polarized neutrons for 10 s. (3 filling time constants) and the beryllium valve was then closed. At this point the neutron number density was approximately 10 $cm^{-3}$. After a 6s. delay to allow the neutron velocities to be randomized, a first Ramsey pulse was applied for 4s., the neutrons were allowed to precess for 70s. in E and B fields, and then a second 4s. Ramsey pulse was applied. Subsequently the valve was opened, neutrons in the appropriate spin state could pass through the polarizing foil (now serving as an analyzer) and finally these neutrons were diverted to a detector

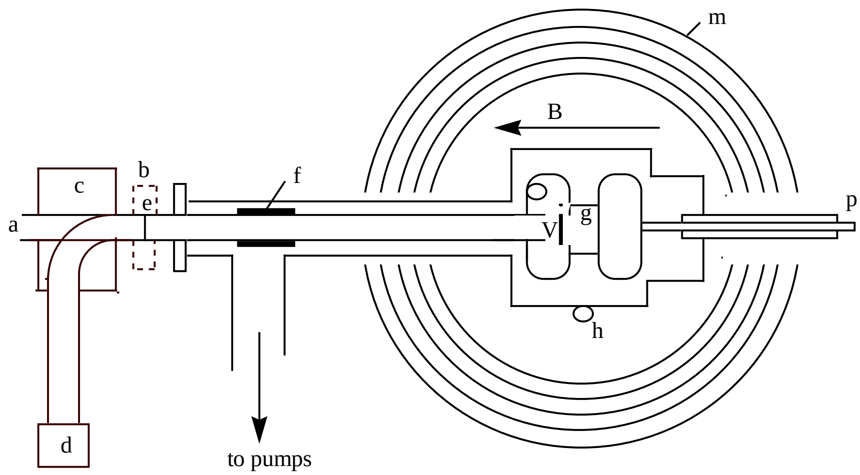



Fig. 4. Schematic diagram of 1990 I.L.L. neutron EDM expt. a:UCN entrance; b: magnet; c: guide changer; d: UCN detector; e: polarizing foil; f: flip coil; g: storage chamber; v: neutron valve; k: rubidium magnetometer; m: 5 layer magnetic shield; p: HV feedthrough.

and counted. The result of this experiment is:

$$d_n = (-3 \pm 5) \cdot 10^{-26}\, e\, cm \tag{22}$$

A comparable result was obtained in a somewhat similar experiment at St. Petersburg.[24]

The 1990 I.L.L. experiment was limited in precision by uncertainties in the magnetic field as determined by 3 separated Rb magnetometers. In a new version of the same experiment[25] that is presently collecting data, the magnetic field inside the bottle is monitored by optically pumped $^{199}$Hg vapor that fills the bottle uniformly (see Fig.5). This and a variety of other improvements make it possible that the current I.L.L. experiment will ultimately yield a result at the $3{\cdot}10^{-26}$ e cm level of precision. (Note, however, that the neutrons are so cold that *they* do not fill the bottle uniformly, but sink gradually toward the bottom! Hence, even in this experiment the magnetometer and the neutrons are not exactly in the same place). New and different experimental methods are needed if $d_n$ is to be measured to much better precision. Following ideas originally presented by Golub and Pendlebury,[26] Golub and Lamoreaux[27] have suggested the following scheme, now in development at Los Alamos National Laboratory. UC

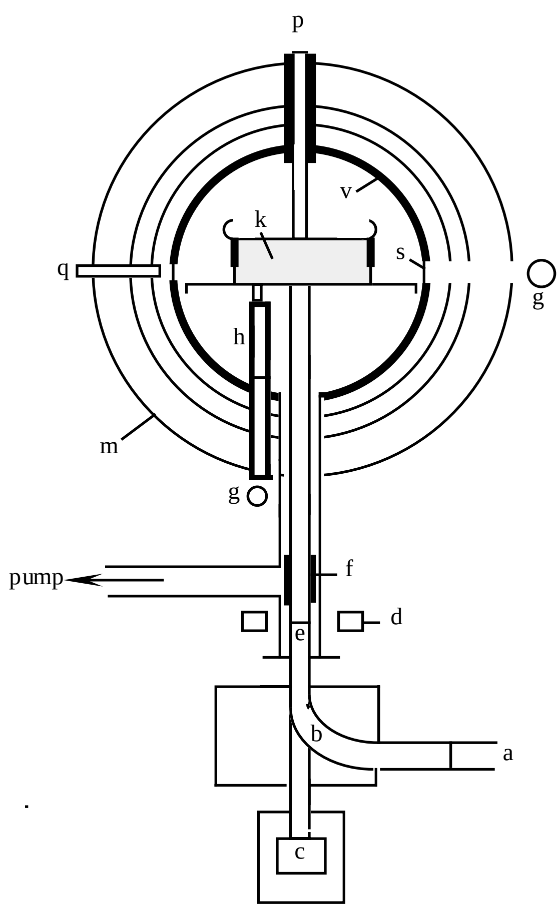



Fig. 5. Schematic diagram of the current I.L.L. neutron EDM experiment. a: UCN source; b: neutron guide change-over; c: UCN detector; d: magnet; e: polarizing foil; f: flip coil; g: Hg UV lamps; h: cell for pre-polarization of Hg atoms; k: storage volume; m: 4-layer magnetic shield; p: HV lead; q: UV light detector; v: vacuum wall; s: optical windows.

neutrons are to be trapped in a storage vessel containing superfluid $^4$He plus a dilute solution of polarized $^3$He. Neutrons at rest or nearly so in superfluid $^4$He can only absorb a $^4$He excitation, the energy $E^*$ and momentum of which lie at the intersection of the $^4$He phonon-roton dispersion curve and the free neutron dispersion curve (see Fig. 6). However, $E^* >> kT$; hence the Boltzmann factor exp(-$E^*$/kT) is much less than unity and the loss rate of UC neutrons by scattering on $^4$He is very small. Therefore it is possible to build up a number density of UC neutrons orders-of-magnitude greater than in previous experiments. The role of $^3$He is as follows. The cross section for absorption of neutrons by the reaction $^3$He(n,p)$^3$H +764 keV is very large, but only in the J=0 spin state of n and $^3$He, (where the spins of these two species are opposed). Thus, observation of this spin-dependent reaction by means of the resulting scintillations in $^4$He provides a means for detecting the precession of the neutron EDM in an applied electric field. As in previous and present neutron EDM experiments, it is critically important to monitor the magnetic field precisely. In principle, this can be done by using the $^3$He as a magnetometer. This proposed experiment has many interesting and subtle technical problems, and may take quite a few years to complete.

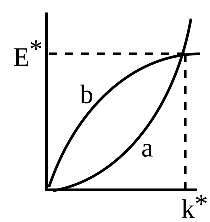



Fig. 6. Sketch of dispersion curves for a) free neutron; b) superfluid $^4$He.

# 6 General remarks about atomic and molecular EDMs

It is obviously impractical to observe the EDM of an electron or bare nucleus by placing it in an external E field, since that particle would quickly be accelerated out of the region of observation. Thus one is led to ask what can be done with a neutral atom or molecule that contains a nucleus or an electron with an EDM d. At first sight, this appears to be an unprofitable approach, because in the point-charge, non-relativistic limit, the atom itself would not possess an EDM $d_a$ (would not exhibit a linear Stark effect), even if d were non-zero. For, neither the atom nor any of its constituents is accelerated in an external electric field $\mathbf{E}$, and in the non-relativistic point-charge limit, where all atomic forces are electrostatic, the *average* electric field at the nucleus or at any electron must be zero. (The electronic and nuclear charges rearrange themselves to cancel the external electric field). This argument is easily cast in quantum-mechanical form and is known as "Schiff's theorem".[28] However, as was first pointed out by Schiff, the theorem does not hold for the nucleus when one takes into account magnetic hyperfine structure, or finite nuclear size if in the latter case the nuclear electric dipole distribution and the nuclear charge distribution are not the same. In this case the nucleus may possess a "Schiff moment" Q, which has dimensions e $cm^3$ and results in non-zero $d_a$.

Sandars[29] showed that Schiff's theorem also fails for an unpaired atomic electron in a paramagnetic atom, when relativistic effects are taken into account (here we recall the all-important factor $\gamma^0$ in the first term on the right hand side of(21)). Sandars demonstrated that in paramagnetic atoms such as the alkalis and thallium, the ratio R=$d_a/d_e$ is of order 10 $Z^3\alpha^2$ where Z is the atomic number and $\alpha$ is the fine structure constant. Thus the "enhancement factor" R can be much larger than unity, and in fact one calculates R=115 for Cs (Z=55) and R=-585 for Tl (Z=81); (see Table 4). Enhancement factors for certain paramagnetic polar molecules (e.g. YbF, PbO, ...) are orders-of-magnitude larger. Sandars' important discovery provides the basis for all modern electron EDM searches - one attempts to observe the linear Stark effect of the appropriate neutral paramagnetic atom or molecule in an external field, and interprets the result in terms of $d_e$ by means of a calculated enhancement factor R.

Table 4. Calculated enhancement factors R

| Atom | Z | State | R |
|---|---|---|---|
| Na | 11 | $3^2S_{1/2}$ | 0.3 |
| Rb | 37 | $5^2S_{1/2}$ | 30 |
| Cs | 55 | $6^2S_{1/2}$ | 115 |
| Fr | 87 | $7^2S_{1/2}$ | 1100 |
| **Tl** | **81** | $\mathbf{6^2P_{1/2}}$ | $\mathbf{-585}$ |

However, it is important to note that an atomic or molecular EDM could arise from other contributions besides a nucleon EDM and/or an electron EDM. P,T-odd nucleon-nucleon (N-N) interactions might generate a nuclear EDM and thus a Schiff moment.[30]

P,T odd electron-nucleon (e-N) interactions might also exist.[31] These as well as the P,T-odd NN interactions could appear in one or several non-derivative coupling forms: ”scalar”, ”tensor”, and ”pseudoscalar” (although the last of these is only capable of making a very small contribution). Finally, as was noted earlier, C,T odd (P even) e-N and N-N interactions and possible T-odd beta decay couplings could cause an EDM through radiative corrections involving the standard-model weak interactions.[6,9] So far, experiments on the diamagnetic atom $^{199}$Hg (see below) provide the best limits on P,T-odd N-N interactions and the tensor form of the P,T-odd eN interaction. Meanwhile an experiment on the paramagnetic atom $^{205}$Tl (see below) provides the best limit on $d_e$ and the scalar form of the P,T-odd eN interaction.

# 7 Diamagnetic atoms

A precise experiment on $^{199}$Hg has been carried out by E.N. Fortson and co-workers at U. Washington, Seattle.[32] (See Fig.7.) $^{199}$Hg was chosen for the following reasons: first, Z=80 is large, and the nuclear-spin-dependent effects that contribute to the Schiff moment scale roughly as $Z^2$. Second, the ground electronic state of Hg is $^1S_0$, (there is no net electronic angular momentum), and the nuclear spin of $^{199}$Hg is I=1/2. Hence nuclear-spin-polarized $^{199}$Hg vapor in a cell has a very long spin relaxation time, because the usual electric quadrupole relaxation mechanisms are absent here. Large polarization in a dense sample ($n \approx 10^{13}$ cm$^{-3}$) is thus quite easily produced by optical pumping with the 254 nm resonance line, and precession induced by an EDM in an external electric field would readily be monitored by the same optical pumping. The main systematic uncertainties arise from possible stray magnetic fields from charging and leakage currents that reverse with electric field. The result is:

$$d(^{199}Hg) = (-1.0 \pm 2.4 \pm 3.6) \cdot 10^{-28}\, e\, cm. \qquad (23)$$

where the first uncertainty is statistical and the second is systematic. Combining these one finds the upper limit:

$$d(^{199}Hg) \leq 8.7 \cdot 10^{-28}\, e\, cm. \qquad (95\%\, confidence) \qquad (24)$$

This is the most precise of all experimental EDM limits. Its implications for various P,T-odd effects are summarized in Table 5. Fortson and co-workers have now embarked on a new version of the experiment, by which they hope to improve the precision by

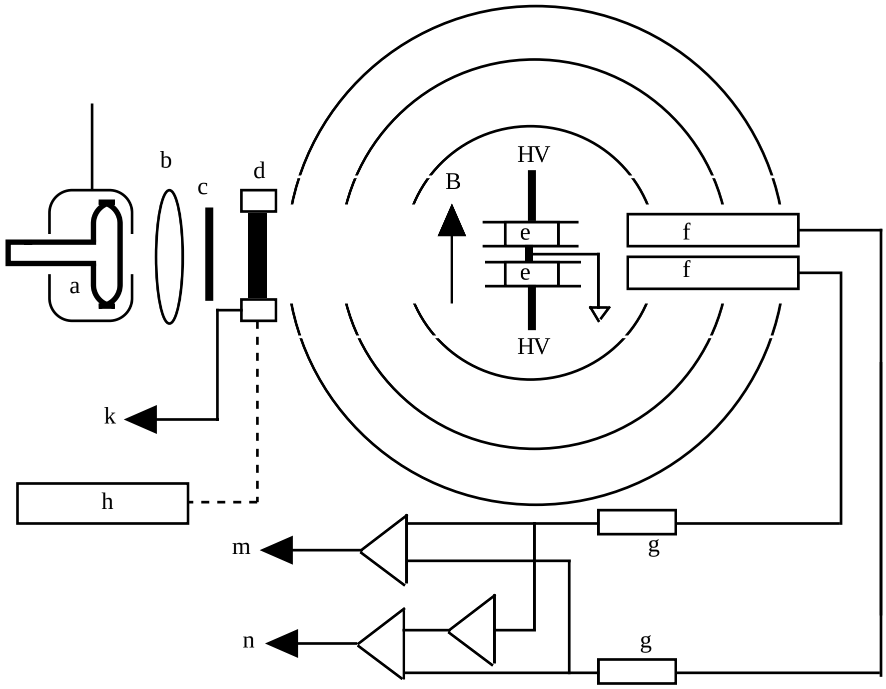



Fig. 7. Schematic diagram of Seattle $^{199}$Hg EDM experiment.a: $^{204}$Hg lamp; b: lens; c: linear polarizer; d: quarter-wave plate; e: optical pumping cells; f: optical detectors; g: phase sensitive detectors; h: stepping motor; k: phase sensitive detector reference; m: magnetic field and gradient correction signals.

an order of magnitude. This gain might be realized by several major improvements, including replacement of the 254 nm light source (formerly a discharge lamp) by a laser; increase in the mercury vapor density, and use of 4 stacked mercury vapor cells to detect and subtract spurious magnetic fields, rather than the two cells previously used.

R. Walsworth of Harvard-Smithsonian, T. Chupp of Michigan, and co-workers[33] have constructed a $^{129}$Xe, $^{3}$He 2-species noble gas maser that may permit a precise EDM experiment on $^{129}$Xe analogous to that on $^{199}$Hg; (see Fig. 8). The apparatus consists of two connected cells containing $^{129}$Xe, $^{3}$He, rubidium vapor, and $N_2$ buffer gas. Polarization of $^{129}$Xe and $^{3}$He is produced by spin exchange with optically pumped Rb vapor in the pump bulb. The spin-polarized xenon and helium diffuse into the maser bulb, where maser oscillations occur at audio frequencies on the nuclear spin-1/2

Zeeman transitions in a static magnetic field. The two-chamber configuration permits excellent frequency stability. The $^3$He functions as a ”co-magnetometer” to monitor the magnetic field, and thus guard against systematic errors arising from leakage and charging currents that might change with the sign of the electric field.

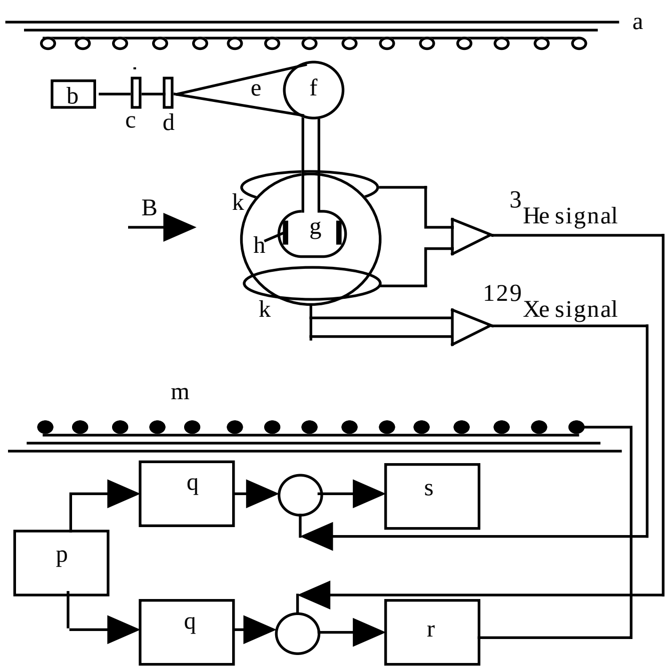



Fig. 8. Schematic diagram of Harvard/Smithsonian-Michigan $^{129}$Xe EDM experiment.a: nested magnetic shields; b: 795 nm light source; c: quarter wave plate; d: beam expander; e: light beam; f: pump bulb; (T=120 C); g: maser bulb; (T=40 C); h: electric field plates; h: pickup coils; m: main solenoid; p: LORAN reference; q: frequency synthesizers; r: phase-sensitive current source; s: A/D converter.

# 8 Paramagnetic atoms

In 1989, L. Hunter and co-workers[34] performed an optical pumping experiment on cesium, and obtained the result:

$$d(Cs) = (-1.8 \pm 6.7 \pm 1.8) \cdot 10^{-24}\, e\, cm \tag{25}$$

which implies

$$|d_e| \leq 9 \cdot 10^{-26}\, e\, cm \tag{26}$$

More recently, electron EDM experiments on Cs have been suggested that would utilize the methods of atom trapping and cooling.[35] Heinzen and co-workers[36] have analyzed the limitations imposed by collisions on the precision that can be achieved by such an experiment. The trapping of cesium atoms in cold solid $^4$He has been studied experimentally, and it may offer interesting possibilities for an EDM search.[37]

At present, however, the best limit on $d_e$ is derived from the Berkeley experiment[38] on $^{205}$Tl. Here the atomic beam magnetic resonance method with separated oscillating fields was employed, (see Fig.9). The experiment was performed in a weak uniform magnetic field that defined the axis of quantization z; (typically $B_z$ =0.4 Gauss). A strong electric field $\mathbf{E}$ (typically 107 kV/cm) was placed between two oscillating rf field regions and was nominally parallel to $\mathbf{B}$. In order to minimize an important possible systematic effect ( the ”$\mathbf{E} \times \mathbf{v}$” effect arising from precession of the atomic magnetic moment in the motional magnetic field) two counterpropagating beams of atomic Tl were utilized, travelling in the vertical direction to minimize the effects of gravity. The final result was:

$$d(^{205}Tl) = (-1.05 \pm .70 \pm .59) \cdot 10^{-24}\, e\, cm \tag{27}$$

Assuming an enhancement factor R = -585 and ignoring all possible contributions to d($^{205}$Tl) except for $d_e$, one obtains from (27) the result:

$$d_e = (1.8 \pm 1.2 \pm 1.0) \cdot 10^{-27}\, e\, cm \tag{28}$$

Some further implications of this result are summarized in Table 5. Analysis of the main sources of noise and systematic uncertainty in this experiment led us to conclude that the overall uncertainty could be reduced by an order-of-magnitude through several basic improvements that have now been implemented (see Fig. 10). The new experiment functions according to the same general plan as the previous one, but utilizes two up-beams issuing from a common oven with two source slits, as well as two down-beams. These beams are separated by 2.5 cm and pass through separate state selectors, collimating slits, rf regions, and analyzer-detector regions, but *opposite* electric fields. However the magnetic field is essentially the same for each beam. Thus the EDM asymmetry is of opposite sign for the two beams, but many sources of noise are highly correlated (”common mode”) for the two beams. Thus when the difference in the signal asymmetries of the two beams is measured, the noise is sharply reduced. In another

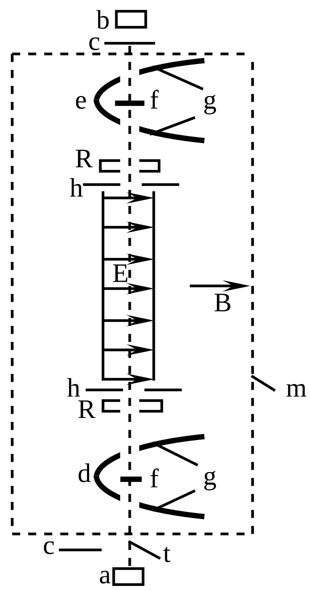



Fig. 9. Schematic diagram of 1994 Berkeley $^{205}$Tl electron EDM experiment. a: up-beam oven; b: down-beam oven; c: automatic beam stops; d: state selector region; e: analyzer region; f: 378 nm laser beams ($\perp$ to page); g: 535 nm fluorescence; h: collimating slits; E: 107 kV/cm electric field; B: static magnetic field; m: 4-layer magnetic shield; R: radio frequency field regions; t: Tl atomic beam.

basic improvement, each up-beam and each down-beam actually consists of atomic sodium as well as atomic thallium issuing simultaneously from each source. Since the atomic number of sodium is Z=11, the enhancement factor is only R(Na) = 0.3. Thus sodium cannot exhibit an observable EDM effect, but is even more sensitive to various systematic effects than is thallium; hence sodium is an effective ”comagnetometer”. At present this experiment has achieved a statistical precision of $3{\cdot}10^{-28}$ e cm in $d_e$, but a subtle systematic effect has so far prevented us from achieving a final result. We hope to overcome this problem in the near future.

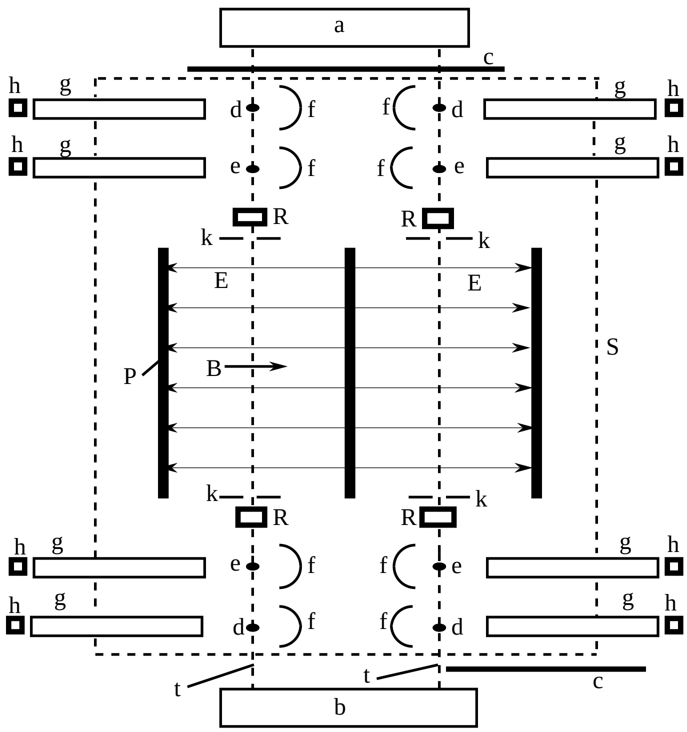



Fig. 10. Schematic diagram of present Berkeley electron EDM experiment. a:down beam oven; b: up beam oven; c: automatic beam stops; d: 378 nm laser beams for Tl (⊥ to page); e: 589 nm laser beams for Na (⊥ to page); f: aluminum reflectors; g: light pipes; h: photodiode detectors; k:collimating slits; P: electric field plates (length 1m., gap 2mm.); R: radio-frequency regions; S: 4-layer magnetic shield; B: uniform magnetic field; E: electric field (120 kV/cm.); t: Na-Tl atomic beams.

# 9 Paramagnetic molecules

Certain paramagnetic polar diatomic molecules are very attractive candidates for experimental electron EDM searches, because their enhancement factors are orders-of-magnitude larger than that of atomic thallium.[39] The list of proposed molecules includes BaF, CdF, YbF, HgF, PbF, PbO, and several others as well. They have large enhancements because the spin-rotational levels of opposite parity that are mixed by P,T-odd EDM interactions lie much closer together (by factors of $\approx 10^3$) than do the analogous levels of paramagnetic atoms. Kozlov and Labzowsky[40] have reviewed calculations of the enhancement factors by a variety of semi-empirical and ab-initio methods.

Consider a typical heavy metal-fluorine molecule MF. Typically M in its normal state has two 6s valence electrons, while the ground configuration of the F atom is $1s^2 ... 2p^5$. In the MF molecule, one of the 6s electrons is transferred to the fluorine, therby completing its p shell and creating an ionic bond with a corresponding molecular electric dipole moment that is typically 3-5 Debye units. The remaining 6s electron moves in a highly polarized orbit in the strong molecular electric field directed along the internuclear axis between $M^+$ and $F^-$. This unpaired electron is the analog of the valence electron in atomic thallium or cesium. If one applies an external E field that is sufficiently strong to generate Stark shifts comparable to the spin-rotational splittings, the internuclear axis becomes aligned with the external E field. For the molecules we have mentioned, this can be achieved with practical laboratory E fields.[41,42] However, the molecular experiments are difficult for diverse reasons, and new results are likely to be slow in coming. One final remark concerning molecules: An experiment[43] on the diamagnetic molecule TlF provides us with the best limit on the EDM of the proton, (see Table 5). However this limit is much less significant than the limit on the neutron EDM.

# 10 The Muon EDM

Until now, the only way to search for the muon EDM $d_\mu$ is to observe the precession of free relativistic muons in a magnetic field. In an important experiment carried out at the CERN muon storage ring more than 20 years ago,[44] the limit:

$$d_{mu} \leq 7 \cdot 10^{-19}\, e\, cm \tag{29}$$

Table 5. Summary of EDM results from neutron, atomic, and molecular expts.

| P,T viol. param. | System | Upper limit | Ref. |
|---|---|---|---|
| $d_n$ | n | $8 \cdot 10^{-26}$ e cm | 23 |
| QCD phase $\Theta_{QCD}$ | n | $4 \cdot 10^{-10}$ | 23 |
| $d_{Diamag.mol}$(TlF) | TlF | $4.6 \cdot 10^{-23}$ e cm | 43 |
| $d_{proton}$ | TlF | $1 \cdot 10^{-23}$ e cm | 43 |
| Schiff moment $Q_S(^{205}Tl)$ | TlF | $1 \cdot 10^{-9}$ e cm | 43 |
| $d_a(^{199}Hg)$ | $^{199}$Hg | $8.7 \cdot 10^{-28}$ e cm | 32 |
| Schiff moment $Q_S^{199}Hg)$ | $^{199}$Hg | $2.2 \cdot 10^{-11}$ e cm | 32 |
| $\eta^{(a)}$ | $^{199}$Hg | $1.6 \cdot 10^{-3}$ | 32 |
| $\eta_q^{(b)}$ | $^{199}$Hg | $3.4 \cdot 10^{-6}$ | 32 |
| $C_T^{(c)}$ | $^{199}$Hg | $1.3 \cdot 10^{-8}$ | 32 |
| Supersym.$\epsilon_q^{SUSY}$ | $^{199}$Hg | $7 \cdot 10^{-3}$ | 32 |
| $d_e$ | $^{205}$Tl | $4 \cdot 10^{-27}$ e cm | 38 |
| $C_S^{(d)}$ | $^{205}$Tl | $4 \cdot 10^{-7}$ | 38 |
| Supersym.$\epsilon_e^{SUSY}$ | $^{205}$Tl | $4 \cdot 10^{-2}$ | 38 |

a) Coupling constant in scalar P,T-odd NN interaction.

b) Coupling constant in scalar P,T-odd quark-quark interaction.

c) Coupling constant in tensor P,T-odd e-N interaction.

d) Coupling constant in scalar P,T-odd e-N interaction.

was established simultaneously with a measurement of the muon g-factor anomaly a = (g-2)/2. Muons from pion decay with an initial polarization $P \geq 95\%$ travelled around the 14 m. dia storage ring in a horizontal plane. A homogeneous vertical magnetic field was applied, and weak vertical focussing was provided by an electrostatic quadrupole field. The precession of the muon spin in these circumstances is described by the famous Bargmann-Michel-Telegdi equation.[45] An EDM causes the precession to be modified slightly because in the muon's rest frame there exists not only a magnetic field, but also a motional electric field to which the EDM is coupled, (See Fig. 11).

A dedicated experiment of the same type has been proposed with the much larger muon storage ring at Brookhaven National Laboratory (which has already been used to improve substantially the precision in determination of g-2). It appears possible in principle to improve the limit on the muon EDM by approximately 4 orders of magnitude.[46]

## 11 Weak dipole moment and EDM of the tau lepton

Finally we mention the tau lepton, which must of course be investigated by the methods of high-energy physics Taus are produced in $e^+e^-$ collisions at colliding beam accelerators:

$$e^+e^- \rightarrow \tau^+\tau^- \rightarrow (A^+\nu) + (B^-\overline{\nu}) \tag{30}$$

The tau production reaction occurs in lowest order by two amplitudes: photon exchange (electromagnetic interaction) and $Z^0$ exchange (neutral weak interaction). When the $e^+e^-$ CM energy is in the vicinity of the Z resonance at 91 GeV, as in various experiments at LEP, the Z amplitude greatly dominates. The mean life of the tau is only $2.9{\cdot}10^{-13}$s., and it decays in a variety of modes represented schematically in (30) by $A^+\nu$ and $B^-\overline{\nu}$. In principle, CP violation could occur at the eeZ vertex, at the $Z\tau\tau$ vertex, and/or in the decays of $\tau^+$ and $\tau^-$. However, if $e^+$ and $e^-$ are unpolarized it can be shown that there are no observable CP violating effects at the eeZ vertex. Furthermore, it seems likely that new CP violation effects would be associated with the vertex at which there is the largest momentum transfer; hence one concentrates on the $Z\tau\tau$ vertex. In order to describe CP violation one then writes the following Lagrangian density:

$$\mathcal{L}_{EDM} = -\frac{i}{2}\overline{\Psi}\gamma^5\sigma^{\mu\nu}\Psi[d_\tau F_{\mu\nu} + \overline{d_\tau}(\partial_\mu Z_\nu - \partial_\nu Z_\mu)] \tag{31}$$

Here, $\Psi$ is the Dirac field describing the tau, $Z_\nu$ is the Z-boson vector potential, and $d_\tau, \overline{d_\tau}$ are the EDM and the "weak dipole moment" (WDM) of the tau, respectively. Strictly speaking the latter quantities depend on the square of the 4-momentum transfer $q^2$; $d_\tau(q^2 = 0)$ is by definition the EDM, while $\overline{d_\tau}(q^2 = m_Z^2)$ is defined as the WDM. While there is no explicit relationship between $d_\tau$ and $\overline{d_\tau}$ they are expected to be roughly the same in most models of CP violation. Near the Z resonance only $\overline{d_\tau}$ term in the Lagrangian density is important.

The transition amplitude $T$ for reaction (30) to a specific final state can be written quite generally as:

$$T = T_{SM} + T_{CP} \tag{32}$$

where $T_{SM}$ is the CP-conserving (standard model) part, while $T_{CP}$ is the CP violating part. Then the differential cross-section is proportional to:

$$|T_{SM} + T_{CP}|^2 = (|T_{SM}|^2 + |T_{CP}|^2) + (T_{SM}T_{CP}^* + cc) \tag{33}$$

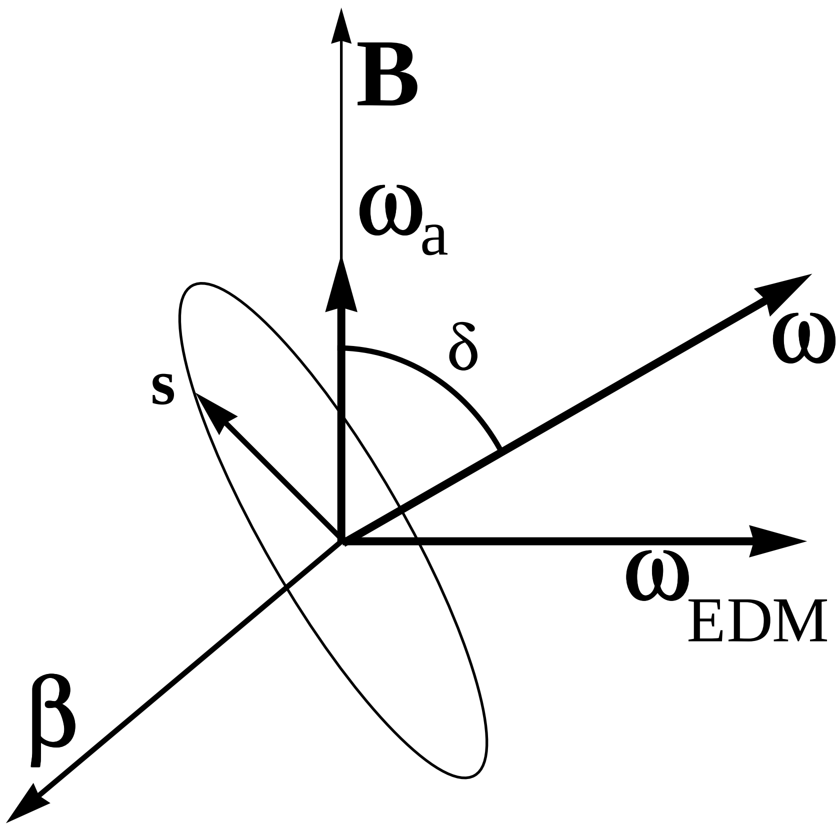



Fig. 11. Schematic diagram (not to scale) of relevant vectors in muon spin-precession experiment, as described by the BMT equation. $\boldsymbol{\beta}$: muon velocity; $\mathbf{B}$: applied magnetic field; $\boldsymbol{\omega}_a$: ang. vel. of precession due to g-2; $\boldsymbol{\omega}_{EDM}$: ang vel of precession due to EDM;$\boldsymbol{\omega}$: resultant; $\delta$: angle between $\boldsymbol{\omega}$ and $\mathbf{B}$; $\mathbf{s}$: muon spin vector, which precesses about $\boldsymbol{\omega}$.

The first quantity in parentheses on the right hand side of (33) is CP-even and is proportional to the total cross-section, or in other words, to the partial width $\Gamma_{\tau\tau}$ for Z decay to $\tau\tau$. The second ("interference") term in parentheses on the right hand side of (33) manifests itself in certain CP-odd observables (correlations) in the differential cross-section.[47] These correlations are related to the spins of the outgoing tau leptons, but the latter are not directly observable. Instead, because of their short lifetime, the taus decay very close to their point of origin, and information about their spins is transferred to the energies and momenta of the decay products $A^+B^-$. The latter quantities must be observed to determine $\overline{d_\tau}$.

Two experiments have been carried out at LEP. One employed the OPAL detector; the other was performed at the ALEPH detector.[48] Combining these results one obtains the limit:

$$\overline{d_\tau} \leq 5.8 \cdot 10^{-18}\, e\, cm \tag{34}$$

More precise limits might be obtained in the future by means of experiments at the proposed "Tau-charm factory". Here it might be possible to employ longitudinally polarized electrons, and thus use more convenient P,T-odd correlations.

# References


[1] E. G. Adelberger and W.C. Haxton, Annu. Rev. Nucl. Part. Sci.**35**, 501 (1985).

[2] L. Dobrzynski et. al., Phys. Lett. **22**, 105 (1966).

[3] E. Blanke et. al., Phys. Rev. Lett. **51**, 355 (1983).

[4] D. Boose, H. L. Harney, and H. A. Wiedenmuller, Phys. Rev. Lett. **56**, 2012 (1986); Z. Phys. A **325**, 363 (1986); J.B. French, V.K. Kota, A. Pandey, and S. Tomsovic, Phys. Rev. Lett. **54**, 2313 (1985); Ann. Phys. (N.Y.) **181**, 198 (1988).

[5] J. E. Koster et. al., Phys. Lett. B **267**, 23 (1991).

[6] M. J. Ramsey-Musolf, Phys. Rev. Lett. **83**, 3997 (1999).

[7] J. D. Jackson, S. B. Treiman, and H.W. Wyld, Nucl. Phys. **4**, 206 (1957).

[8] C. G. Callan and S. B. Treiman, Phys. Rev. **162**, 1494 (1967).

[9] I. B. Khriplovich, JETP Lett. **52**, 461 (1990).

[10] R. I. Steinberg et. al., Phys. Rev. Lett.**33**,41 (1974).

[11] A. V. Vorobiov et. al., Nucl. Instr. Methods A**284**, 127 (1989).

[12] S. Freedman, private communication (1999).

[13] A. L. Hallin et. al. Phys. Rev. Lett. **52**, 337 (1984).

[14] M. Abe et. al. (KEK-E246 Collaboration), Phys. Rev. Lett. **83**, 4253 (1999).

[15] L. Landau, Sov. Phys. JETP **5**, 336 (1957).

[16] W. G. Penney, Phil. Mag. **11**, 602 (1931).

[17] E. E. Salpeter, Phys. Rev. **112**, 1642 (1958).

[18] E. P. Shabalin, Sov. J. Nucl. Phys. **28**, 75 (1978); Sov. Phys. Usp. **26**, 297 (1983); T. Donoghue, Phys. Rev. D **18**, 1632 (1978).

[19] X. He, B. H. J. McKellar, and S. Pakvasa, Int. J. Mod. Phys. A**4**, 5011 (1991); A. Czarnecki and B. Krause, Phys. Rev. Lett. **78**, 4339 (1997).

[20] F. Hoogeveen, Nucl. Phys. B **341**, 322 (1990).

[21] M. Pospelov and I. B. Khriplovich, Sov. J. Nucl. Phys. **53**, 638 (1991).

[22] N. F. Ramsey, Phys. Rev. **76**, 996 (1949); *Molecular Beams* (Oxford University Press, Oxford, 1990), p.365.

[23] K. F. Smith, et. al., Phys. Lett. B **234**, 191 (1990).

[24] I. S. Altarev et. al., Phys. Atomic Nucl. **54**, 1152 (1996).

[25] P. G. Harris et. al., Phys. Rev. Lett. **82**, 904 (1999).

[26] R. Golub, and J. M. Pendlebury, Phys.Lett. A **53**, 133 (1975); Rep. Prog. Phys. **42**, 439 (1979).

[27] R. Golub and S. K. Lamoreaux, Phys. Rep. **237**, 1 (1994).

[28] L. I. Schiff, Phys. Rev. **132**, 2194 (1963).

[29] P. G. H. Sandars, Phys. Lett. **14**, 194 (1965); Phys. Lett.**22**, 290 (1966).

[30] O. P. Sushkov, V. V. Flambaum, and I. B. Khriplovich, Sov. Phys. JETP **60**, 873 (1984); V. V. Flambaum,I. B. Khriplovich, O. P. and Sushkov, Phys. Lett. B **162**, (1985); Nucl. Phys. A **449**, 750 (1986); V. M. Khatsymovsky,I. B. Khriplovich, and A. S. Yelkhovsky, Ann. Phys. **186**, 1 (1988); W. C. Haxton, and E. M. Henley, Phys. Rev. Lett. **51**, 1937 (1983); X. He and B. McKellar, Phys. Lett. B **62**, 97 (1976); I. B. Khriplovich, Sov. Phys. JETP **4**, 25 (1976); V. F. Dmitriev, I.B. Khriplovich, and V. B. Telitzin, Phys. Rev. C **50**,2358 (1994).

[31] C. Bouchiat, Phys. Lett. B **57**, 284 (1975); E. Hinds, C. Loving, and P. G. H. Sandars, Phys. Lett. B**62**, 97 (1976); V. A. Dzuba, V. V. Flambaum, and V. V. Sylvestrov, Phys. Lett. B **154**, 93 (1985); A. M. Martensson-Pendrill, Phys. Rev. Lett. **54**, 1153 (1985); S. M. Barr, Phys. Rev. Lett. **68**, 1822 (1992); Phys. Rev. D**45**, 4148 (1992);I. B. Khriplovich, Nucl. Phys. B **352**, 385 (1991).

[32] J. P. Jacobs, W. M. Klipstein, S. K. Lamoreaux, B. R. Heckel, and E. N. Fortson, Phys. Rev. A **52**, 3521 (1995).

[33] R. E. Stoner et. al. Phys. Rev. Lett. **77**, 3971 (1996).

[34] S. A. Murthy, D. Krause, Z. L. Li, and L. Hunter, L. Phys. Rev. Lett. **63**, 965 (1989).

[35] S. Chu, private communication (1997); D. Weiss, private communication (1999).

[36] M. Bijlsma, J. Verhaar, and D. Heinzen, Phys. Rev. A **49**, R4285 (1994).

[37] S. I. Kanorsky, S. Lang, S. Lucke, S. B. Ross, T. W. Hansch, and A. Weis, Phys. Rev. A **54**, R1010 (1996).

[38] E. D. Commins, S. B. Ross, D. DeMille, and B. C. Regan, Phys. Rev. A **50**, 2960 (1994).

[39] O. P. Sushkov, and V. V. Flambaum, Sov. Phys. JETP **48**, 608 ((1978); V. G. Gorshkov, L. N. Labzowsky, and A. N. Moskalev, Sov. Phys. JETP **49**, 414 (1979); V. V. Flambaum,and I. B. Khriplovich, Phys. Lett. A **110**, 121 (1985); M. G. Kozlov, Sov. Phys. JETP **62**, 1114 (1985); Sov. J. Quantum Electronics **18**, 713 (1988); M. G. Kozlov, V. I. Fomichev, Yu. Yu. Dmitriev, L. N. Labzowsky, and A. V. Titov, J. Phys. B **20**, 4939 (1988); M. G. Kozlov, and V. F. Ezhov, .Phys. Rev. A **49**, 4502 (1994).

[40] M. G. Kozlov and L. N. Labzowsky, J.Phys. B **28**, 1933 (1995).

[41] D. DeMille, Private communication (1997).

[42] E. Hinds, and B. Sauer, Phys. World April, 37 (1997); B. E. Sauer, J. Wang,and E. Hinds, Phys. Rev. Lett. **74**, 1554 (1995); J. Chem Phys. **105**, 7412 (1996).

[43] D. Cho, K. Sangster, and E. Hinds, Phys. Rev. A **44**, 278 (1991).

[44] J. Bailey, et. al. [CERN Muon Storage Ring Collaboration], J.Phys. G **4**, 345 (1978); Nucl. Phys. B **150**, 1 (1979).

[45] V. Bargmann, L. Michel, and V. L. Telegdi, Phys. Rev. Lett. **2**, 435 (1959).

[46] Y. Semertzidis,Private communication (1997).

[47] W. Bernreuther, and O. Nachtmann, Phys. Rev. Lett. **63**, 2787 (1989); W. Bernreuther et. al., Z. Phys. C **43**, 117 (1989); Z. Phys. C **52**, 567 (1991); W. Bernreuther, A. Brandenburg, and P. Overmann, Phys. Lett. B **391**, 413 (1997).

[48] R. Akers et. al., (OPAL Collaboration), Z. Phys. C **66**, 31 (1995); Buskelic, D. et al., (ALEPH Collaboration), Phys. Lett. B **346**, 371 (1995).

# CP VIOLATION AND THE ORIGINS OF MATTER

Andrew G. Cohen
Department of Physics, Boston University
Boston, MA 02215

ABSTRACT

I present a gentle introduction to baryogenesis, the dynamical production of a baryon asymmetry during the early universe. I review the evidence for a cosmic baryon asymmetry and describe some of the elementary ingredients necessary for models of baryon number production.

# 1 Introduction and Experiment

Even though the Universe has a size, age and complexity far beyond our everyday experience, the laws of physics determined in the laboratory can be extrapolated to the vast realms of the cosmos. This program, pursued since the earliest developments in the physical sciences, has seen enormous change over the last century. Especially important for particle physics has been the close interaction between the high energy frontier and the very early universe, and cosmological arguments are now routinely used to constrain the rampant imaginings of particle theorists. One area that is closely connected with the principle topic of this years school, CP violation, is baryogenesis, the dynamical production of a net baryon number during the early universe. This asymmetry, which is well established experimentally, is one of the most important features of the cosmos as a whole, and represents an enormous departure from the CP invariant state of equal matter and antimatter densities, with no net baryon number. The subject has been of concern to particle physicists since the discovery of microscopic CP violation, which encouraged the construction of concrete baryogenesis scenarios. The subject became a standard part of modern cosmology with the introduction of grand unified theories (GUTs), introduced in the 1970s, which establish a possible source for baryon number violation, an essential component of baryogenesis. More recent ideas have attempted to link the baron asymmetry with details of models of electroweak symmetry breaking, and offer the possibility of testing models of baryogenesis in future colliders such as the LHC.

There are many good reviews of baryogenesis at all levels*. Here we give only a brief overview of the subject and encourage further consultation of the references.

## 1.1 Initial Data

One of the fundamental questions concerning the large scale structure of our universe is surprisingly difficult to answer: What is the universe made of? In general terms this question reduces to the value of a single parameter, the total energy density of the universe, which is usually quoted in terms of a "critical" density related to the current Hubble expansion rate:

$$0.01 \lesssim \Omega_0 \equiv \frac{\rho}{\rho_{\rm crit}} \lesssim 3 \,, \qquad (1)$$

*The book[1] by Kolb and Turner is a good (although somewhat dated) starting point. There are many more recent reviews,[2–4] as well as references therein.

where $\rho_{\text{crit}} \equiv 3H_0^2/(8\pi G_N)$, $H_0$ is the Hubble constant and $G_N$ is Newton's gravitational constant. The lower value comes from the visible content of the universe, the mass-energy associated with stars, galaxies, *etc*. The larger value comes from various measurements of large scale structure, especially measurements of the potential associated with gravitating (but not necessarily visible) mass-energy. The discrepancy between these numbers suggests that the majority of the mass-energy of the universe is dark, possibly a completely new kind of material. But even for the visible mass, we have no direct experience of the stuff out of which distant stars are made, although we believe this stuff to be matter similar to that which makes up our own star. The detailed physics of distant stars, such as stellar evolution, spectral lines, *etc*, is convincing evidence that these objects are made of baryons and leptons much as ourselves, but there remains the possibility that they are constructed from *antimatter*, *i.e.* antiquarks and positrons, rather than quarks and electrons. The transformation CP acting on a state of ordinary matter (by which we mean baryons, objects made of quarks carrying a positive baryon number) produces a state of antimatter (with negative baryon number). Thus if all stars in the universe contain matter (in the form of baryons) rather than antimatter (in the form of antibaryons), then this matter antimatter (or baryon) asymmetry represents a departure from CP symmetry as well.

What evidence is there that distant objects are made of matter rather than antimatter? For that matter, how do we know that the earth itself is matter? Matter and antimatter couple electromagnetically with known strength. Contact between matter and antimatter leads naturally to annihilation into photons with characteristic energy of $100$s of MeV. Casual observation easily demonstrates the absence of this radiation when matter (in the form of ourselves, say) comes in contact with another terrestrial object. Thus we easily deduce that the earth (and all its occupants) are made of matter. A similarly pedestrian argument indicates that the moon too is made of matter. Indeed our exploration of nearby space convincingly shows that the solar system is composed of matter.

In fact it is not necessary that a man-made item come into contact with distant objects to establish the nature of such objects. If anything known to be matter is in contact with an unknown object, the absence of gamma radiation from annihilations demonstrates the object is not antimatter. For example micro-meteorites are continuously bombarding the earth without such radiation, and are therefore not antimatter. But these objects also rain upon Mars, which is therefore also not antimatter. This argument can obviously be extended: as long as a sufficiently dense matter trail extends from our solar system, absence of $100$ MeV gamma rays demonstrates the absence of antimatter. This trail extends to distances comparable to the size of our local galactic cluster,[5] the Virgo cluster, a distance of $20$ Mpc.

Unfortunately this region covers only a tiny fraction of the observable universe, which has a characteristic linear size several orders of magnitude larger than that of the Virgo cluster. Constraining the composition of objects beyond our local neighborhood requires a more complex analysis.

Experiments to search for cosmic antimatter from beyond this $20$ Mpc distance have been proposed. The most ambitious of these, the Alpha Magnetic Spectrometer[6–11] (AMS) is scheduled to be deployed aboard the International Space Station sometime in the distant future[†]. This device, essentially a large mass spectrometer, will search for negatively charged nuclei in cosmic rays. The device should place a direct limit on antimatter in cosmic rays coming from a distance of nearly an order of magnitude beyond our local cluster. Although this distance scale remains small compared to the current visible universe, it is a significant step beyond our local cluster.

Lacking further direct experimental evidence against distant regions of antimatter, we must rely on alternative observational and theoretical analyses. Our original argument, the lack of gamma radiation emanating from points of contact between regions of matter and antimatter, fails when the density of both matter and antimatter becomes so small that the expected gamma ray flux falls below a detectable level. However this suggests an improvement on this argument: since the density of matter (and any putative antimatter) is decreasing with the cosmic expansion of the universe, we might expect that the flux of gamma radiation from such points of contact was larger in the early universe than it is today. Thus we might search for radiation from matter antimatter annihilation that occurred not today but sometime in the far past. A search for such radiation would differ from those which already place stringent limits on antimatter in our local neighborhood. Firstly, once produced as gamma rays, radiation would subsequently redshift as the universe expands. Consequently rather than searching for gamma rays with energies of $100$s of MeV, we should search for lower energy radiation. Secondly, when we look out to large redshift (the distant past) on the night sky we are integrating over large portions of the universe. Consequently rather then seeking point sources we should search for a diffuse background of radiation coming from many points of intersection of domains of matter with those of antimatter.

---

[†]A prototype device has flown in the space shuttle. Although the exposure was insufficient to detect antimatter, this brief test has returned interesting cosmic ray physics.[12]

In order to use this technique to place limits on cosmic antimatter we must have some idea of how a diffuse photon spectral flux is related to the properties of domains of antimatter, in particular their size. We already know that such domains should be larger than the $20$ Mpc limit we have in hand. The environment of this photon production, the interface between regions of matter and antimatter in the early universe, involves known principles of physics, and upper limits on the photon flux can be deduced. Although rather complicated in detail, the basic strategy is straightforward:

- The observed uniformity of the cosmic microwave background radiation implies that matter and antimatter must have been extremely uniform at the time when radiation and matter decoupled, a redshift of about $1100$ or a time of about $10^{13}$ seconds. Thus at this time domains of matter and antimatter cannot be separated by voids, and must be in contact with each other. Prior to this time it is conceivable that matter and antimatter domains *are* separated by voids, and thus we do not include any annihilation photons prior to this epoch.
- Annihilation proceeds near matter antimatter boundaries through combustion, converting matter into radiation according to standard annihilation cross-sections. This change of phase in the annihilation region leads to a drop in pressure, and matter and antimatter then flow into this region. This leads to a calculable annihilation flux via the flow of matter and antimatter into this combustion zone. The annihilation process also gives rise to high energy leptons which deposit energy in the matter and antimatter fluids, significantly enhancing the annihilation rate.
- At a redshift of about $20$ (approximately $10^{16}$ s after the big bang) inhomogeneities leading to structure formation begin to become significant. Although this likely does not affect the rate of annihilations significantly, rather than analyze this era in detail it is safer (more conservative) to ignore any further annihilation.
- The spectrum of photons produced prior to a redshift of $20$ continues to evolve due to the expansion of the universe as well as subsequent scattering.

The results of this calculation[13] are shown in Figure 1. The upper curve represents the computed spectral flux of diffuse radiation from domains of antimatter with a characteristic size of $20$ Mpc, the lower limit allowed by other analyses. The lower curve represents the spectral flux for a domain size of $1000$ Mpc, a large fraction of the visible universe. In both cases this calculated flux is substantially larger then the observed diffuse gamma ray background (by balloon and satellite experiments). In particular such a flux would be in serious conflict with the results of the COMPTEL satellite

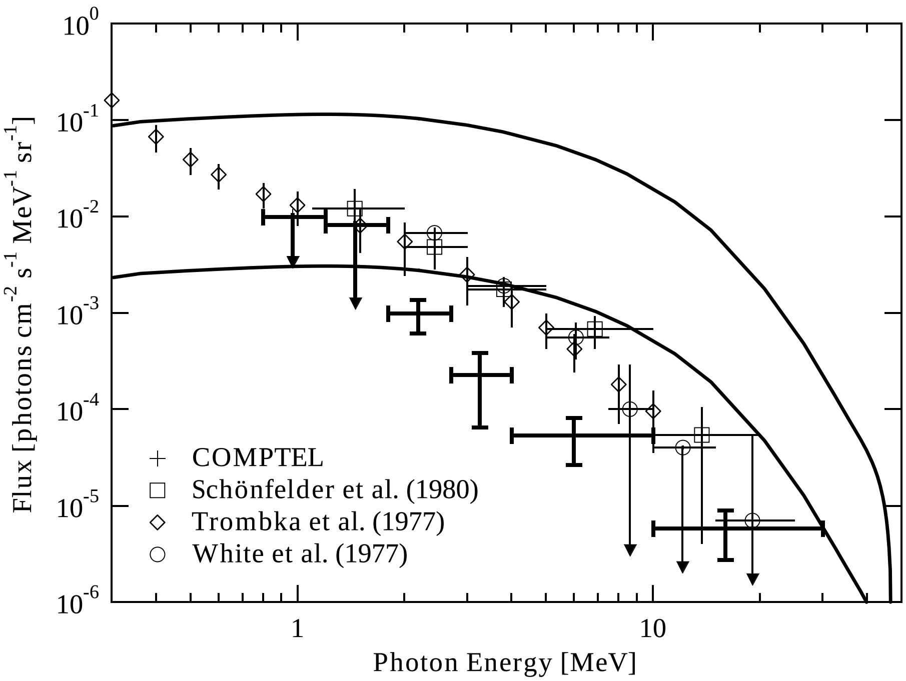



Fig. 1. Data[14] and expectations for the diffuse $\gamma$-ray spectrum.

experiments. We conclude that domains of antimatter of size less than $1000$ Mpc are excluded.

## 1.2 A Baryon Asymmetry

The arguments of the preceding section indicate that the universe contains predominately matter and very little antimatter (or that matter and antimatter have been separated into several near universe-sized domains, a possibility[15,16] we will not consider here.) This asymmetry has been a focus of contemporary cosmology and particle physics principally because of its implied CP violation. To decide the significance of this asymmetry we need a quantitative measure of this departure from baryon antibaryon equality. Normally we will use the baryon density to photon number density ratio:

$$\eta \equiv \frac{n_B + n_{\bar{B}}}{n_\gamma} \ . \tag{2}$$

This choice is motivated partly by the dimensionless nature of the ratio, but more importantly, by the way in which this ratio scales with the expansion of the universe. Provided the expansion is isentropic (and ignoring baryon production or destruction) both the numerator and denominator densities dilute with the cosmic expansion in the same way, inversely proportional to the change in volume, and thus the ratio $\eta$ is time independent.

Since our previous arguments suggest that $n_{\bar{B}}$ is insignificant, we may use the observed (visible) baryon density and the microwave background radiation density to obtain an experimental lower limit on $\eta$

$$10^{-10} \lesssim \eta \; . \tag{3}$$

In fact a more constrained value may be obtained by using some additional theoretical information. The synthesis of the light elements in the early universe depends quite sensitively on the baryon density. Using the best observations on the primordial elements this constrains $\eta$[17]:

$$4 \cdot 10^{-10} \lesssim \eta \lesssim 7 \cdot 10^{-10} \; . \tag{4}$$

Is this value significant? To get a better idea of how large this number is, we might imagine its value in a baryo-symmetric universe. In this case, as the universe cools from temperatures above $1$ GeV where baryons and antibaryons are in thermal equilibrium with a thermal number density proportional to $T^3$, baryon number is kept in thermal equilibrium by baryon antibaryon annihilation. Once the rate for this process becomes slower than the expansion rate, the probability of a subsequent annihilation becomes negligible. Using a typical hadronic cross-section, this equality of rates occurs at a temperature of about $20$ MeV. At this time baryons, in the form of protons and neutrons, have an equilibrium number density proportional to:

$$n_B \propto (T m_N)^{3/2} e^{-m_N/T} \tag{5}$$

and give a value for $\eta$

$$\eta \sim 10^{-20} \; . \tag{6}$$

This value, in gross conflict with the experimental number, cannot be avoided with thermal equilibrium between equal number of baryons and antibaryons, reflecting the efficient and near total annihilation of all matter. However there is a simple path to obtain a much larger value. If the number of baryons exceeds that of antibaryons by even a small amount, than the inability of each baryon to “pair up” with an antibaryon prevents total annihilation. In fact this excess need be only a few parts per billion at high temperature (leading to one extra baryon for each several billion photons) to achieve an adequate value for $\eta$.

But where would such an excess come from? It might appear as an initial condition, set at the beginning of the universe in some way beyond our ken. Note that such an initial condition is irrelevant in the context of inflation; following the reheating phase at the end of inflation all memory of such an initial condition is erased. Without inflation this is a rather unpleasant possibility that we must acknowledge, but we will favor an explanation that does not rely on a *deus ex machina* of this type. What is preferable is a mechanism by which this peculiar excess arises dynamically during the evolution of the universe, a possibility known as *baryogenesis*.

As was first observed by Andrei Sakharov,[18] there are three conditions that must be met in order for baryogenesis to occur:

- Baryon violation. Obviously if the universe is going to evolve a non-zero baryon number from a time when the baryon number vanishes (at the end of inflation, say) than the laws of physics must allow the baryon number to change.
- C and CP violation. Whatever process changes the baryon number must do so in a way that favors baryon production, rather than antibaryon production. Since both C and CP transformations change the sign of the baryon number, the laws of physics must violate both C and CP in order to obtain a positive value. Fortunately nature has provided us with both of these elements. As an example:

$$\frac{\text{Rate}[K_L^0 \to e^+\pi^-\nu]}{\text{Rate}[K_L^0 \to e^-\pi^+\bar{\nu}]} \simeq 1.006 \tag{7}$$

- Departure from thermal equilibrium. Roughly speaking if we populate all levels according to a Boltzmann distribution, since CPT guarantees that each level with a positive baryon number has a corresponding level with a negative baryon number, the total baryon number must vanish. More formally, since $\hat{B}$ is CPT odd and the Hamiltonian CPT even, in thermal equilibrium

$$\langle \hat{B} \rangle = \text{Tr}\hat{B}e^{-\beta\hat{H}} = \text{Tr}\,\Omega_{CPT}\Omega_{CPT}^{-1}\hat{B}e^{-\beta\hat{H}} = -\langle \hat{B} \rangle = 0\,. \tag{8}$$

Discussions of baryogenesis are often, not surprisingly, focused on the origin of these three ingredients. Beginning in the late 1970s it was realized that all three arise in commonly considered extensions of the standard model:

- Baryon Violation. Grand Unified theories, in which quarks and leptons appear in the same representation of a gauge group, naturally give rise to baryon violation.
- C, CP violation. Kaon physics already implies a source of C and CP violation.

- Departure from thermal equilibrium. The universe is known to be expanding and cooling off. This change in the temperature with time *is* a departure from thermal equilibrium.

We will turn to an evaluation each of these items in somewhat more detail.

Baryon violation is severely constrained by its apparent absence in the laboratory: experiments searching for proton decay have already placed a limit on the proton lifetime greater than $10^{32}$ years. How can baryon violation be significant for baryogenesis yet avoid a disastrous instability of the proton? The key is the notion of an accidental symmetry: a symmetry of all possible local operators of dimension four or less constrained by the particle content and gauge invariance of a theory. The significance of accidental symmetries appears when we consider the effects of new physics at high energies. These effects may be incorporated at low energies by including all possible local operators that respect the symmetries of this new physics. By dimensional analysis all operators of dimension higher than four will be suppressed by powers of the ratio of the low energy scale to the high energy scale. Now imagine that new physics at high energies does not respect some symmetry, like baryon number. At low energies we must include all local operators, including those that violate baryon number, an apparently disastrous result. But if the theory has an accidental symmetry, the only such operators are of dimension greater than four (by the definition of accidental symmetry), and thus new physics at high energies which violates this symmetry is suppressed by the high energy scale. In the standard model baryon number is exactly such an accidental symmetry: no baryon violating operators of dimension four or less can be constructed out of the standard particles consistent with the $SU(3) \times SU(2) \times U(1)$ gauge invariance. In fact the leading baryon violating operator in this construction is dimension six. If we then contemplate new physics which violates baryon number at a high energy scale, such as in grand unified theories, baryon violating effects will be suppressed at low energies by two powers of this high energy scale. Thus if the scale is greater than $10^{16}$ GeV, proton decay (a low energy process taking place near 1 GeV) is hugely suppressed.

As already indicated, CP violation is present in the kaon system at a level which appears more than adequate to explain a baryon asymmetry of less than one part in one billion. However CP violation in the standard model arising from a phase in the CKM matrix (which may or may not account for the phenomena observed in the kaon system) is unlikely to be responsible for the baryon asymmetry of the universe. As we will see, the effects of this phase in the early universe are quite small.

If the CP violation in the standard model can not account for the observed baryon asymmetry of the universe, what can? In fact almost *any* new source of CP violation beyond that of the phase in the CKM matrix gives rise to significant effects in the early universe. From a particle physics perspective, this is the principal reason for interest in the cosmic baryon asymmetry: it is a strong indication of physics beyond the standard model.

Lastly, the expansion of the universe which characterizes a departure from thermal equilibrium is governed by the Hubble parameter:

$$\frac{\dot{T}}{T} = -H \tag{9}$$

(at least during periods of constant co-moving entropy.) Today the Hubble parameter is quite small; the characteristic time scale for expansion of the universe is $10$ billion years. Since most microphysical processes lead to thermal equilibrium on much shorter time scales, baryogenesis must take place either at a time when $H$ is much larger, or at a time when Eq. (9) doesn't hold.

# 2 Grand Unification

Together the items of the previous section suggest that baryogenesis occurs at relatively early times, when the universe was hot and baryon violation was important. In particular the ingredients on our list all fit quite naturally into many grand unified theories. In such theories, super-heavy gauge bosons associated with the grand unified gauge group, as well as super-heavy Higgs bosons associated with GUT symmetry breaking, can mediate baryon violating processes. Although suppressed at low energies, at the high temperatures prevalent in the early universe baryon violation rates can be large. In addition, the rapid expansion rate

$$H \sim \frac{T^2}{M_P} \tag{10}$$

allows for significant departure from thermal equilibrium. Finally the interactions associated with new scalar fields that all GUT models must have may include CP violating couplings.

To see how this works in more detail, consider a toy model consisting of bosons $X$ (and $\bar{X}$) which couple to quarks and leptons in a baryon violating, and CP violating, way. For example imagine that the $X$ ($\bar{X}$) boson decays into the two final states $qq$ ($\bar{q}\bar{q}$)

and $\bar{q}\bar{l}$ ($ql$) with branching fractions $r$ ($\bar{r}$) and $1-r$ ($1-\bar{r}$) respectively. The parameters of this toy are constrained by symmetry. For example, CPT insures that the masses of the bosons are equal $m_X = m_{\bar{X}}$, as are the total widths $\Gamma_X = \Gamma_{\bar{X}}$. The baryon number of each final state is conventional: $B(qq) = 2/3$, $B(\bar{q}\bar{l}) = -1/3$, *etc*. Finally C and CP symmetry would imply $r = \bar{r}$. However lacking these symmetries, generically $r$ will differ from $\bar{r}$[‡].

If we now imagine starting with thermal number densities of $X$ and $\bar{X}$ bosons, our CPT constraint insures that these densities are equal $n_{\bar{X}} = n_X$. Using the parameters of introduced in the preceding paragraph we can compute the net baryon number of the quarks and leptons which result from the $X$ and $\bar{X}$ decays:

$$n_B + n_{\bar{B}} = n_X\Big[r\frac{2}{3} + (1-r)(-\frac{1}{3})\Big] + n_{\bar{X}}\Big[\bar{r}(-\frac{2}{3}) + (1-\bar{r})\frac{1}{3}\Big] = n_X(r-\bar{r})\ . \quad (11)$$

Although this formula is correct, it is the answer to the wrong question. If all interactions are in thermal equilibrium, the $X$ and $\bar{X}$ bosons will be replenished at the same time that they decay. That is, the rate for the inverse process, production of $X$ (and $\bar{X}$) bosons through $qq$ or $\bar{q}\bar{l}$ fusion, will have a rate in equilibrium which is precisely the same as the decay rate, when the number densities of all the particles are equal to their thermal equilibrium values. For example, at temperatures small compared to the $X$ boson mass, the production rate of quarks and leptons via $X$ decay is small, since there are very few $X$ bosons in equilibrium, $n_X \propto \exp(-m_X/T)$. Conversely the inverse process, creation of an $X$ boson, is rare since the quarks and leptons are exponentially unlikely to have the energy necessary to produce a real $X$ boson. So in equilibrium the baryon number does not change, and Eq. (11) is not relevant.

This suggests what turns out to be the key to baryon production—we need the number density of $X$ and $\bar{X}$ bosons at $T << m_X$ to be much larger then the exponentially small equilibrium number density. Under these circumstances the $X$ and $\bar{X}$ production processes will be much smaller than the decay processes. If the number density of $X$ and $\bar{X}$ bosons is sufficiently large, we may even ignore the inverse process all together.

How do we arrange this miracle? Clearly we must depart from thermal equilibrium, something we already knew from our discussion of Sakharov's conditions. But as we have also discussed the universal expansion allows such a departure when the rate for an equilibrating process is slow compared to the expansion rate. In this case, we need the processes that keeps the number density of $X$ and $\bar{X}$ bosons in equilibrium to be

‡Of course C and CP violation are not sufficient—interference with a scattering phase is also necessary.

slow compared to the expansion. There are two processes which decrease the number of bosons: the decay of the $X$ and $\bar{X}$ bosons; and annihilation of the $X$ and $\bar{X}$ bosons into other species. Both of these processes can be slow if the couplings of the $X$ boson are weak. Of course "slow" means in comparison with the Hubble expansion rate, $H \sim T^2/M_P$. If this is indeed the case, the number density of $X$ bosons will not track the equilibrium value proportional to $\exp(-m_X/T)$, but instead remain larger. Then once the age of the universe is larger than the lifetime of the $X$ boson, decay will occur, leading to a baryon number according to Eq. (11).

There is one important constraint that we have overlooked. Even though the $X$ and $\bar{X}$ bosons are not re-produced around the time that they decay, there are other processes we must not forget. In particular, there are processes which violate baryon number through the mediation of a (virtual) $X$ boson. In our toy example these may be represented by the effective four-fermion operator $qqql$. This dimension six operator has a coefficient proportional to two inverse powers of the $m_X$ mass, and thus at temperatures low compared to this mass the effects of this operator are small. Nevertheless processes of this type will change the baryon number, tending to equilibrate this number to zero. Therefore we must further require that baryon violating processes such as this one must also be out of equilibrium at the time the $X$ and $\bar{X}$ bosons decay.

The procedure outlined above is usually called a "late decay", or "out-of-equilibrium decay" scenario. Developed extensively from late 1970s through the present, they have provided a framework in which to discuss baryogenesis, and have led to many concrete models that can explain the non-zero value of $\eta$. Although successful in principal, models of GUT baryogenesis often have difficulty obtaining the large baryon asymmetry we observe:

- *Rates:* We have seen that a number of rates must be slow compared to the expansion rate of the universe in order to depart sufficiently from equilibrium. These rates are typically governed by the GUT scale, while the expansion rate is proportional to $T^2/M_P$. The relevant temperature here is that just prior to the decay of the $X$ bosons. Since we need these bosons to be long lived, this temperature is lower than the GUT scale, and the expansion rate is correspondingly slower. Thus the departure from equilibrium is far from automatic and detailed calculations in a specific GUT are necessary to determine whether these conditions can be satisfied.

- *Relics:* One problematic aspect of many GUTS is the presence of possible stable relics. For example some GUTS have exactly stable magnetic monopoles which would be produced in the early universe at temperatures near the GUT scale. Unfortunately these objects are a cosmological disaster: the energy density in the form of monopoles would over-close the universe, in serious conflict with observation. One of the early great successes of inflation was a means for avoiding this catastrophe. At the end of inflation all matter in the universe has been "inflated away", leaving a cold empty space free from all particles (baryons as well as monopoles!). However following the end of inflation, the vacuum energy density in the inflaton field goes into reheating the universe, producing a thermal distribution of particles. If this reheating is fast, energy conservation tells us that the reheat temperature will be close to the original scale of inflation, near or above the GUT scale. Unfortunately this would reintroduce the monopoles. On the other hand if this reheating is slow (as would be the case if the inflaton is weakly coupled) then the energy density in the inflaton field decreases as the universe expands, leading to a much lower reheat temperature. Thus for inflation to solve the monopole crisis, the reheat temperature must be well below the GUT scale, in which case monopoles are not re-introduced during the reheating process. Unfortunately neither are the $X$ and $\bar{X}$ bosons, and thus baryogenesis does not occur.

Neither of these objections are definitive—there are proposals for circumventing them both. For example much of our discussion has focused on small departures from thermal equilibrium. It may be possible to have huge departures, where particle distributions are not even remotely thermal. In this case the analysis of reaction rates is quite different. There may also be many more couplings which allow a greater range of reaction rates. Perhaps these are associated with Yukawa couplings of neutrinos or other sectors of the GUT. These objections do however make these scenarios less compelling. In addition there is another, more philosophical, problem. Often in these models the details of baryogenesis are pushed into very particular aspects of the GUT, physics at scales which are not accessible in the laboratory. Thus in many instances, whether or not GUT baryogenesis occurs is experimentally unanswerable. For these reasons it is advisable to investigate alternatives.

# 3 Electroweak Baryogenesis

In 1985 Kuzmin, Rubakov and Shaposhnikov[19] made the remarkable observation that all three of Sakharov's criteria may be met in the standard model. Firstly, and perhaps most surprisingly, the standard model of the weak interactions does not conserve baryon number!

The non-conservation of baron number in the standard model is a rather subtle effect. At the classical level, the conservation of baryon number is practically obvious—each term in the classical action respects a transformation of the baryon number. Nöther's theorem then applies, and we can construct a four-vector, the baryon number current, which satisfies the continuity equation, that is whose four-divergence vanishes. Nonetheless this naïve argument is wrong: this four vector does *not* have vanishing four-divergence in the full quantum theory.

This situation is not totally unfamiliar. In the simple case of quantum electrodynamics a corresponding phenomena occurs, known as the axial anomaly. QED has a symmetry of the classical action corresponding to an axial rotation of the electron field (that is, a rotation which is opposite on the left and right chirality electron fields). Aside from the electron mass term which we will ignore, this transformation leaves the action unchanged, and the Nöther procedure leads to a covariantly conserved four-vector, the axial current. However as is well known this current is *not* divergenceless:

$$\partial_\mu J^\mu_a = \frac{e^2}{32\pi^2} F_{\mu\nu}\tilde{F}^{\mu\nu} \propto \vec{E}\cdot\vec{B} \ , \qquad (12)$$

where $\tilde{F}^{\mu\nu} \equiv \epsilon^{\mu\nu\alpha\beta}F_{\alpha\beta}/2$. This remarkable equation, which can be derived in a number of different ways, embodies the violation of axial charge due to quantum effects in the theory[§]. Note that ignoring spatial variations this equation implies that the time derivative of the baryon density will be non-zero in the presence of a non-zero $\vec{E}\cdot\vec{B}$. Note that the chiral nature of the current couplings are important for obtaining this result; the current with non-chiral couplings, the electromagnetic current, is strictly conserved.

The situation in the standard model is similar. The baryon current derived via the Nöther procedure is vectorial, and thus would seem an unlikely candidate for an

[§]The axial anomaly in QED has a long and well-known history. Eq. (12) may be obtained for example by evaluating the triangle diagram, by computing the change in the functional integral measure under an axial rotation, or by an exact calculation of the electron propagator in a constant background electric and magnetic field.

anomaly. However the weak interaction couplings *are* chiral, which leads to an equation for the divergence of the baryon current corresponding to Eq. (12):

$$\partial_\mu J_B^\mu = 3\Big[\frac{g^2}{32\pi^2}W^a_{\mu\nu}\tilde{W}^{a\mu\nu} + \frac{g'^2}{32\pi^2}F_{Y\mu\nu}\tilde{F}_Y^{\mu\nu}\Big] = \partial_\mu J_L^\mu \ . \qquad (13)$$

In this equation $W$ and $F_Y$ are the gauge field strengths for the $SU(2)$ and $U(1)$ hypercharge gauge potentials, $g$ and $g'$ are the corresponding gauge couplings and the $3$ arises from a sum over families. We have also noted that the lepton number current has the same divergence as the baryon number current. Consequently the current $J_{B-L} \equiv J_B - J_L$ is divergenceless, and the quantum number $B - L$ is absolutely conserved.

What does Eq. (13) really mean? To gain some understanding of this equation, imagine constructing an electroweak solenoid surrounding an electroweak capacitor, so that we have a region in which the quantity $\vec{E}^a \cdot \vec{B}^a$ is non-zero. In practice this is rather difficult, primarily because we live in the superconducting phase of the weak interactions, and therefore the weak Meissner effect prevents the development of a weak magnetic field. But lets ignore this for the moment. Now perform the following gedanken experiment: start with no weak electromagnetic fields, and the region between the capacitor plates empty. If we solve the Dirac equation for the quarks and leptons, we obtain the usual free particle energy levels. In this language, we fill up the Dirac sea, and leave all positive energy levels unoccupied. Now imagine turning on the weak $\vec{E}^a$ and $\vec{B}^a$ fields adiabatically. In the presence of these slowly varying fields, the energy level solutions to the Dirac equation will flow, while the occupation of any given level does not change. But according to Eq. (13) the baryon number will change with time. This corresponds to the energy of some of the occupied levels in the Dirac sea flowing to positive energy, becoming real particles carrying baryon number. Although surprising at first, this is not very different from ordinary pair production in a background field. What is peculiar is the creation of quarks in a way different from antiquarks, so that a net baryon number is produced.

By itself this effect is intriguing but not sufficient. After all what we are really after is a transition which changes baryon number without changing the state of the gauge field, much as the four-fermion operator in our grand unified example did. That is, what we would like to do is begin our gedanken experiment as above, but at the end of the day turn off the electric and magnetic fields. Naïvely this would leave us with zero baryon number: if we turn the fields off as the time reverse of how we turned them on, we produce baryon number at first, and then remove it later on. Indeed this is what happens with axial charge in the quantum electrodynamics example. But the non-abelian example contains another wrinkle: it is possible to turn the electric and magnetic fields on and then off in a way which leaves a non-zero baryon number!

The trick as realized by 't Hooft[20,21] follows from noticing that, unlike the abelian case, there are a large number of non-trivial gauge potentials which have vanishing electric and magnetic fields. It is possible in our gedanken experiment to begin with one of these potentials, and finish with another, thus tying a "knot" in the gauge field¶. The result is a transition from a state with no weak electric and magnetic fields and no baryon number (a "vacuum"), and ending with no weak electric and magnetic fields but non-zero baryon number. Making such a transition requires a "large" gauge field, one in which the field strength is of order $1/g$. In addition, the total change in baryon number is quantized in units of the number of families, presumably 3.

If we accept this fancy formalism, we have an obvious question: why is the proton stable? If the weak interactions violate baryon number, shouldn't the proton lifetime be a characteristic weak time scale? In fact, the proton is absolutely stable even in the presence of this baryon violation, because each process changes the baryon number by 3. Since the proton is the lightest particle carrying baryon number, its decay would require changing the baryon number by 1, which cannot occur if all baryon violating process change the baryon number by multiples of 3. Thus there is a selection which accounts for the stability of the proton.

What about other baryon violating processes? In fact these too are unimportant. In our gedanken experiment above we ignored the fact that the weak interactions are broken, that we live in a superconducting phase of the weak interactions. But this means that there is a large potential energy cost in creating a weak $\vec{E}^a$ and $\vec{B}^a$ field which interpolates between our states with different baryon number. That is, there is a potential barrier that we must overcome in order to change the baryon number by a weak interaction. Since the gauge field must change by order $1/g$, the height of this barrier (the cost of overcoming the Meissner effect) is

$$E_s \sim \frac{M_Z}{\alpha_{wk}} \sim \text{a few TeV} \tag{14}$$

where $\alpha_{wk}$ is the weak analog of the electromagnetic fine structure constant. The gauge field configuration at the peak of this barrier is called the "sphaleron", and hence this

¶This argument is a bit tricky. In order to discuss the physics of gauge potentials it is necessary to gauge fix. Even after gauge fixing there are gauge potentials which begin in the far past with one "vacuum" potential, and end with a different one.

energy is known as the sphaleron energy.

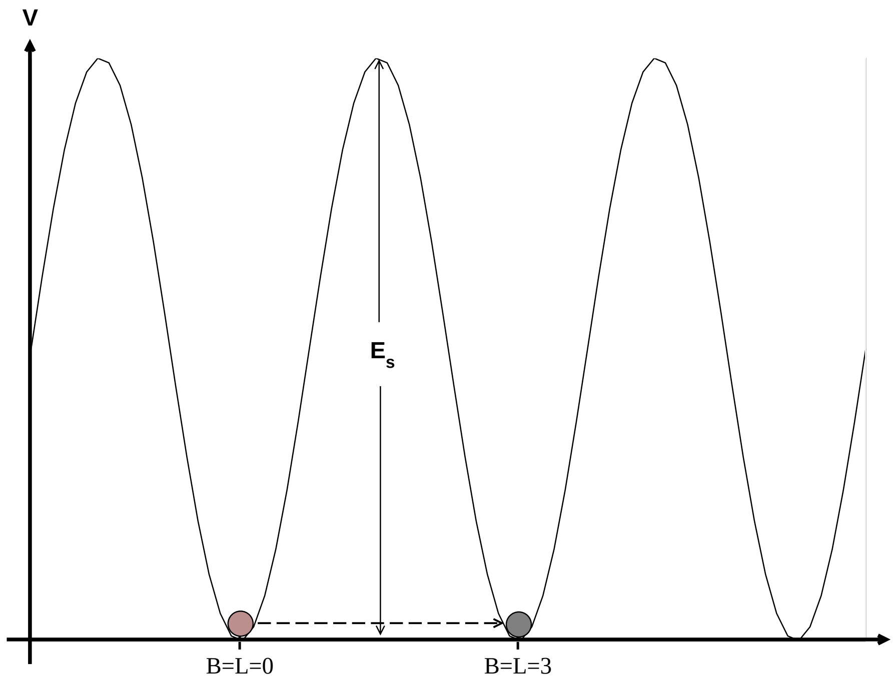



Fig. 2. The potential energy in one direction in gauge field space. This direction has been chosen to go from one zero energy gauge field configuration to another through the pass of lowest energy.

The presence of this barrier means that processes with energies below the barrier height are highly suppressed; they are strictly forbidden classically, but can occur through quantum tunneling. Like all tunneling processes, the probability of such a transition will be proportional to a semi-classical barrier penetration factor:

$$\text{Prob} \propto e^{-4\pi/\alpha_{wk}} \sim 10^{-40} \; , \tag{15}$$

an utterly negligible effect. In contrast to the grand unified case where baryon violation was suppressed at low energies by powers of the ratio of the energy to the grand unification scale, here the baryon violation is exponentially suppressed by the presence of a barrier.

If our interest were only sensitive tests of baryon number conservation in the laboratory, we would safely move on to another area of research. But since our interest is in baryogenesis in the early universe, we must take this picture of baryon violation in the weak interaction by transiting this barrier more seriously. At temperatures comparable or larger than the barrier height we would expect a significant population of states with energies above the barrier. These states could make a transition without the quantum tunneling suppression by simply evolving classically over the top of the barrier. The rate for such a baryon violating process will be controlled by the probability of finding a state with energy at least as large as the sphaleron energy:

$$\Gamma \propto e^{-E_s/T} \ . \tag{16}$$

When the temperature is larger than $E_s$ this exponential is no longer a suppression at all. Hence we expect that at temperatures above a few TeV baryon violation in the weak interactions will occur at a characteristic weak interaction rate. Note that at temperatures of a few TeV weak interactions are extremely rapid compared to the Hubble expansion rate, and thus baryon violating interactions would be in thermal equilibrium.

We come to the first important consequence of baryon violation in the weak interactions: grand unified baryogenesis does not necessarily produce a baryon asymmetry! Even if a late decaying $X$ boson would produce a baryon asymmetry at temperatures near the GUT scale, this asymmetry will be equilibrated away by baryon violating weak interactions. Our discussion of grand unified baryogenesis concluded that baryon violation from virtual $X$ boson exchange must be slow for baryogenesis to succeed, but the real requirement is that *all* baryon violation must be slow; we must take into account *all* sources of baryon violation, including that of the weak interactions.

There is a simple way of avoiding this effect. As indicated in Eq. (13) the baryon and lepton number currents have exactly the same divergence. Hence their difference, the $B - L$ current, is strictly conserved. Therefore if the $X$ boson decay produces a net $B - L$, weak interactions cannot equilibrate this quantum number to zero. The result will be both a net baryon number and a net lepton number. However baryon and lepton number violating weak interactions must be taken into account when calculating the baryon asymmetry produced.

Rather surprisingly we have concluded that baryon violation is present in the standard model, at least at temperatures above a few TeV. In principle this opens the possibility of baryogenesis taking place at temperatures well below the GUT scale. Unfortunately we face another obstacle: departure from thermal equilibrium. As discussed earlier, the expansion rate of the universe at temperatures near a TeV is quite slow: $H \sim T^2/M_P \sim 10^{-16}$ TeV. All standard model interactions lead to reaction rates much larger than this expansion rate, typically of order $\Gamma \sim \alpha_{wk} T \sim 10^{-3}$ TeV. Thus departure from thermal equilibrium is impossible with such a leisurely expansion. Fortunately there are occasions during the early universe in which the smooth variation of the temperature with the expansion, Eq. (9), is invalid. This typically occurs when the equation of state for the content of the universe undergoes an abrupt change, such as during a change in phase structure. For example when the temperature falls below the mass of the electron, electrons and positrons annihilate into photons, converting their energy from a non-relativistic form (the mass-energy of the leptons) into a relativistic form (radiation). But there may be other phase changes in the early universe. With a phase transition there exists the possibility of significant departure from thermal equilibrium, at least if the transition is discontinuous, or first order.

Is there any reason to expect a phase transition in the early universe? At temperatures much higher than a few TeV we have very little idea of the state of the universe; until we probe physics at these high energies in the laboratory we cannot say whether or not phase transitions occur. Of course we are permitted to speculate, and indeed there are many proposals for new physics beyond the standard model which lead to interesting dynamics in the early universe. But beyond speculation, we already expect that there is at least one phase transition in the context of the standard model: the electroweak phase transition.

As we have already mentioned we currently live in a superconducting phase of the electroweak interactions. The $W$ and $Z$ boson masses arise from the interaction of the gauge fields with a non-zero order parameter, an object that carries electroweak quantum numbers and has a non-zero expectation value in the vacuum. The short range nature of the weak force is a consequence of this interaction, just as the electromagnetic interaction is short range in ordinary superconductors. In fact it is this property of the weak interactions which leads us to deduce the existence of a non-zero order parameter. We know the value of the order parameter, the weak vev, is approximately $250$ GeV; we also know that the order parameter is a weak doublet, from the relation between the $W$ and $Z$ masses and the weak mixing angle. However unlike electromagnetic superconductivity where the order parameter is known to be a composite of two electrons, a so-called “Cooper pair”, the weak order parameter remains mysterious. One possibility is that the order parameter is simply some new field with its own physical excitations, the Higgs field. Another is that it is a composite of two fermions, like the Cooper pair. But until we have probed the details of electroweak symmetry breaking in detail, as we hope to do in future collider experiments, we can not say with any confidence what form the detailed physics of this order parameter takes.

One thing we do expect, in analogy with ordinary superconductivity, is the change in phase of the weak interactions at high temperatures. Just as an electromagnetic superconductor becomes non-superconducting as the temperature is increased, so too the weak interactions should revert to an unbroken phase at high temperature. When the temperature is on the order of $100$ GeV, the order parameter should vanish, the weak gauge symmetry will be unbroken and the $W$ and $Z$ (and the quarks and leptons) will become massless. In our discussion of baryon violation in the weak interactions we suggested that at temperatures larger than the sphaleron energy baryon violation would be unsuppressed, as transitions could take place above the barrier. But the barrier itself was a consequence of the Meissner effect, a sign of superconductivity. Indeed Eq. (14) clearly shows the relationship with symmetry breaking: the sphaleron energy is proportional to $M_Z$ which in turn is proportional to the order parameter. At temperatures of a few hundred GeV, well below the sphaleron energy, when the weak symmetry is restored and the order parameter goes to zero, the barrier disappears. Consequently baryon violation will occur rapidly just on the unbroken side of the phase transition.

In order for any of this to play a role in baryogenesis, we require significant non-equilibrium effects at the phase transition. According to the usual classification of phase transitions, such non-equilibrium effects will arise if the phase transition is first order. Under these circumstances the transition itself may proceed in a classic first order form, through the nucleation of bubbles of broken phase[‖]. Indeed as the universe cools from high temperature, we begin with a homogeneous medium in the unbroken phase of the weak interactions. Quarks and leptons are massless, weak interactions are long range (aside from thermal screening effects) and, most importantly, baryon violation is rapid. Calculating the rate for baryon violation requires understanding the details of the classical thermodynamics of the gauge fields, a difficult subject. The result however is relatively simple:

$$\Gamma_{\Delta B} \sim \alpha_{wk}^5 T \tag{17}$$

This is a rather crude approximation; for example there are logarithmic corrections to this relation that may be significant, as well as a potentially large dimensionless coefficient. Nevertheless the exact formula may in principle be obtained numerically in terms of $\alpha_{wk}$ and the temperature.

As the universe cools we eventually reach a moment in which the free energy of the unbroken phase is equal to that of the broken phase, as indicated by the free energy

[‖] This is not the only possibility; for example it may proceed through spinodal decomposition, or some more complicated mechanism. In all these circumstance non-equilibrium phenomena are likely.

curve labeled by $T_c$ in Fig. (3). However if the transition is first order, these two phases are separated by a free energy barrier and the universe, unable to reach the broken phase, remains in the unbroken phase. As the universe continues to expand, the system *supercools*, remaining in the unbroken phase even though the broken phase has a lower free energy. Finally we reach a point where bubbles of the preferred, broken, phase nucleate and begin to grow. Eventually these bubbles percolate, completing the transition.

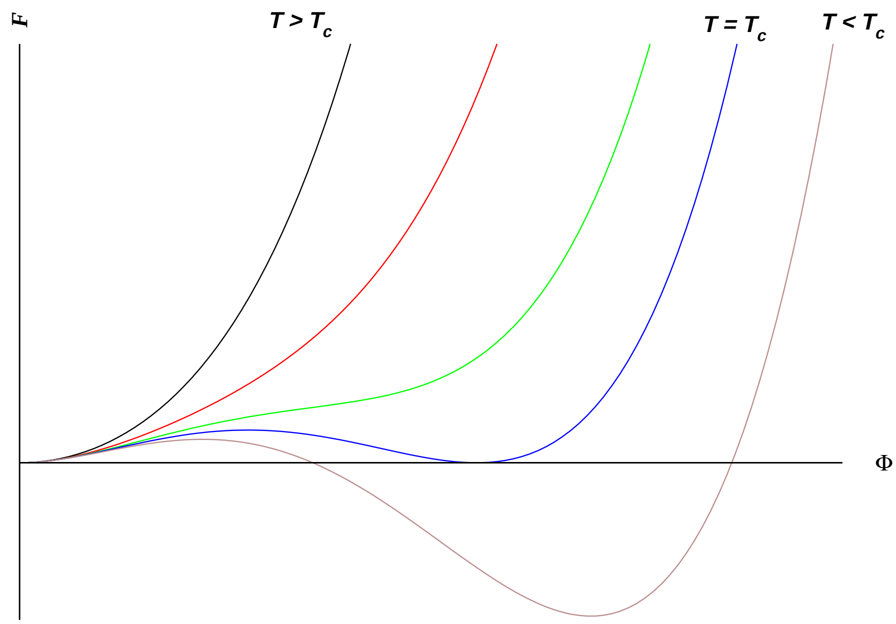



Fig. 3. The free energy versus the order parameter for a classic first order phase transition.

Clearly these expanding bubbles represent a departure from thermal equilibrium. From the point of view of Sakharov's condition the most relevant fact is the discontinuity in the order parameter, the weak vev. In the region outside the bubbles the universe remains in the unbroken phase where the weak order parameter is zero. As discussed previously there is no barrier between the states of different baryon number, and baryon violation is rampant. In the bubble interior the weak vev is non-zero, the $W$ and $Z$ bosons are massive, and *the barrier between states of different baryon number is in place*. In this case the rate of baryon violation is exponentially suppressed according to a Boltzmann factor $\exp(-E_b/T)$ where $E_b$ is the barrier height. Naïvely we might expect $E_b$ to be the sphaleron energy. However the sphaleron energy represented the barrier height at zero temperature; at finite temperature the barrier is generically different, evolving to the zero temperature shape as the universe cools. But $E_b$ is still controlled by the order parameter, the weak vev. If this vev is large, near its vacuum value of $250$ GeV, baryon violation will be essentially shut off in the bubble interior. On the other hand if the vev is too small, baryon violation will proceed rapidly inside the bubble as well as out.

The difference in the weak vev in the bubble interior and the bubble exterior, the discontinuity in the weak order parameter, is a measure of the strength of the transition. If these two values are nearly equal, the phase transition is nearly continuous, a second order transition. If on the other hand the discontinuity is large, the phase transition is said to be strongly first order. For electroweak baryogenesis to occur, baryon violation must be out of thermal equilibrium in the bubble interior, a situation that will transpire only if the vev is sufficiently larger. Thus we need a strongly first order electroweak phase transition.

What do we know about the electroweak phase transition? Unfortunately almost nothing. This is due in small measure to our inability to understand the complex thermal environment in a relativistic quantum field theory. Over the past decade there has been a great deal of progress in simulating field theories at finite temperature, deducing details of phase transitions and reaction rates. However these advances are of little use if we don't know what theory to simulate. The main reason we can't say definitely whether the electroweak phase transition is first or second order, whether it is strongly or weakly first order, or practically anything else about it is simple: we have no idea what physics is responsible for electroweak symmetry breaking.

We do have some theories of electroweak symmetry breaking, and huge effort has been invested in determining the details of the phase transition in these cases. The original theory of electroweak symmetry breaking relied on the introduction of a fundamental weak doublet scalar field, the Higgs field. In this rather simple case, the electroweak phase transition is first order only if the physical Higgs scalar is very light, with a mass well below the current experimental bound. But this theory is not the most popular alternative for electroweak symmetry breaking due to its theoretical shortcomings. Of somewhat greater appeal is the minimal supersymmetric standard model, the MSSM. In this case there are a host of new particles: supersymmetric partners of the quarks, leptons and gauge bosons, as well as two Higgs multiplets. In fact this theory also requires some of these new states to be relatively light in order to obtain a sufficiently strongly first order phase transition. As the LEP bound on the MSSM Higgs mass improves, the region of parameter space for which the phase transition is appropriate is rapidly disappearing.

Should we take this to mean the weak phase transition is probably inappropriate for electroweak baryogenesis to take place? That depends a bit on our philosophy. Given that these are but 2 ideas out of a nearly infinite variety we should not necessarily become disheartened. More importantly there have been analyses of modest alternatives of the above theories: non-supersymmetric theories with multiple Higgs fields, extensions of the MSSM including singlets, and even strongly interacting theories of electroweak symmetry breaking. In most of these cases a sufficiently strong first order phase transition is easy to arrange, if not generic. In fact this is perhaps one of the more positive aspects of electroweak baryogenesis. The physics responsible for electroweak symmetry breaking is intimately related with the possibility of electroweak baryogenesis: some models of electroweak symmetry breaking do not produce a baryon asymmetry (or not one of sufficient size) while others do. This is one of the few places that the forefront of electroweak physics, electroweak symmetry breaking, may have a profound effect on cosmology (or vice versa).

## 3.1 Baryon Production

We now have all of Sakharov's ingredients in place, all in the weak interactions: baryon violation, C and CP violation and a departure from thermal equilibrium. But we still have not explored how these ingredients combine to produce a baryon asymmetry.

Clearly we require all three ingredients to work together—the absence of any one implies the absence of baryogenesis. The non-equilibrium requirement, satisfied by the nucleation and subsequent expansion of bubbles of broken phase, is most importantly realized as a spatial separation of baryon violation: baryon violation is rapid outside the bubble, and non-existent in the bubble interior. C and CP violation, at least in the standard model, take place through the Yukawa couplings in the Lagrangian. That is, C and CP violation appear in the form of non-trivial phases in the couplings of quarks (and possibly leptons in extensions of the standard model) to the Higgs field, the order parameter for electroweak symmetry breaking. But it is precisely this field which represents the electroweak bubbles which appear at the phase transition.

The details of how the baryon asymmetry may be calculated in the context of these expanding bubbles is complicated, and we will not discuss it at any length. The ingredients are clear: the CP violating interaction of quarks and leptons with the expanding bubbles can in principle bias the production of various quantum numbers (including but not limited to baryon and lepton number); all that is required is an interaction that allows the creation or destruction of a net value for such a quantum number. For example, the interaction with the expanding bubble may bias the production of left-chirality top quarks over right-chirality top quarks (to pick a random example). Provided CP violation (either directly or in the form of one of these quantum number asymmetries) biases baryon number in a region outside the bubble where baryon violation is rapid, a net baryon number will be produced. Following our example, an excess of left-chirality top quarks (which have a weak interaction) over right-chirality top quarks (which do not) biases the weak interactions in the direction of increasing baryon number. An important element which complicates the discussion is the transportation of quark and lepton charges from one region of space to another. The transport properties of the plasma are crucial in understanding how the baryon violating interactions, which take place outside the bubble, are biased by CP violation, which is dominant where the Higgs field is changing inside the bubble. Depending on the details of the bubble profile the analysis looks a bit different, although the results are qualitatively similar.

## 3.2 CP Violation

We finally must come to grips with CP violation; now that we understand how it is relevant to electroweak baryogenesis, we can ask what the characteristic size of CP violating effects of the sort described in the last paragraph will be. In fact this question is not as difficult as might be supposed. CP violation in the standard model arises from a non-trivial phase in the Yukawa couplings of the quarks. The only tricky issue is that this phase has no unique location: we may move it from one coupling to another by making field redefinitions. More physically this means that an interaction will only violate CP when the interaction involves enough couplings such that we cannot remove this phase from all these couplings simultaneously. For example, if a process involves only two families of quarks, the CP violating phase may be put in the third family, and this process will be CP conserving.

Since the Yukawa couplings are relatively small (even the top quark coupling), perturbation theory should be an adequate guide to the size of CP violating effects. To estimate this size we must construct an object perturbatively out of the various coupling constants of the standard model in a way which involves an (irremovable) CP violating phase. Clearly there must be a large number (8) of Yukawa couplings from all three families as well as a large number (4) of weak interactions in order to get an irremovable phase. This product of small dimensionless coupling constants is an invariant measure of CP violation in any perturbative process. One such example, involving the largest Yukawa couplings, is

$$\delta_{CP} \sim \alpha_{wk}^2 \lambda_t^4 \lambda_b^2 \lambda_s \lambda_d \sin^2\theta_1 \sin\theta_2 \sin\theta_3 \sin\delta \sim 10^{-16} \ . \tag{18}$$

This remarkably small number, many orders of magnitude smaller than the observed baryon asymmetry, is a consequence of the detailed symmetries of the standard model, where CP violation is intimately connected with flavor violation. As long as the flavor physics of baryogenesis is perturbative, the standard model has no hope of producing a baryon asymmetry large enough. Although we have consistently maintained that the standard model has CP violation, and that this is one of the most interesting reasons to investigate baryogenesis, it now seems that we have been misled, that this CP violation is far too small to be relevant for baryon production in the early universe.

Why did we argue earlier that CP violation in the kaon system, Eq. (7), was so much larger than this perturbative estimate? In fact we have been careful to argue that the estimate of CP violation, Eq. (18), only applies when the standard model Yukawa interactions can be used perturbatively. This is not the case for CP violation in the kaon system. If we wish to compute CP violating effects at kaon energies, $E << 250$ GeV, we must first construct the effective theory appropriate to these energy scales by integrating out modes with energies larger than $E$. This includes for example the $W$ and $Z$, the top and bottom quarks, *etc*. As usual this process introduces inverse powers of these heavy masses, such as $1/M_W^2$ and $1/m_t^2$. Since these masses are proportional to the weak couplings $g$ and $\lambda_t$ appearing above, this effective theory has interactions which can *not* be represented as a power series in couplings (although it is easy enough to construct this effective theory and keep track of the Yukawa couplings), and the estimate Eq. (18) does not apply[**].

But we have now come to the crux of the matter, and if it were not for the interesting physics associated with baryon violation, cosmic expansion, *etc.* that we wished to discuss we could have started (and ended) our discussion of baryogenesis here. The most important message from this analysis is that it is highly unlikely that CP violation from the phase in the CKM matrix has anything at all to do with the cosmic baryon asymmetry. Although we have chosen to mention this in the context of electroweak

[**] A more old-fashioned language for the same phenomena would note the enhancement of perturbative matrix elements by small energy denominators in perturbation theory.

baryogenesis, there is nothing special about this scenario in our analysis of the size of CP violating effects. Everything we have said applies to standard model CP violation in any theory of baryogenesis that takes place at high energies where our perturbative argument applies. This is certainly the case in grand unified baryogenesis as well as electroweak baryogenesis.

Once more, with feeling: standard model CP violation in the form of a phase in the CKM matrix is not likely to produce a significant baryon asymmetry of the universe. Why is this so important? As we have argued there *is* a cosmic baryon asymmetry, and if it didn't come from CP violation in the standard model, where did it come from? The obvious conclusion is that there is CP violation (and hence new physics) beyond the standard model. This is one of the strongest pieces of evidence we have that the standard model is incomplete.

One comment is in order. We have now repeatedly said that standard model CP violation is inadequate for baryogenesis. This is sometimes confused with the (incorrect) statement that the CP violation observed in the kaon system is too small to produce the observed baryon asymmetry. At the moment our knowledge of CP violation is not extensive enough to say definitively that the observed CP violation is associated with a phase in the CKM matrix. It is perfectly possible that CP violation in the kaon system is dominated by physics beyond the standard model. This would likely show up as a discrepancy between CP violation measured in the B system relative to the expectations from the K system.

If the standard model must be augmented with new CP violation to create the baryon asymmetry, what form is this new CP violation likely to take? We don't know. However it is worth noting that CP violation in the standard model, with its intimate connection to flavor symmetries, is rather special. In almost any extension of the standard model, new interactions and new particles allow for new sources of CP violation. Under these circumstances this new CP violation is not constrained by the standard model flavor symmetries and will typically give large effects. Indeed the apparent smallness of CP violation at low energies is a strong constraint on physics beyond the standard model, since most extensions of the standard model lead to large, even unacceptable, CP violating effects.

Most investigations of baryogenesis have focused on models proposed for reasons other than CP violation and the baryon asymmetry. For example, a natural extension of the original fundamental Higgs standard model includes multiple Higgs fields. With one or more new Higgs fields there are new CP violating couplings, the flavor structure of the model is different, and baryogenesis is certainly possible. A particularly popular extension of the standard model, the MSSM, has a number of new CP violating phases, and can easily have large CP violation at the electroweak scale. As we have discussed, the phase transition in this model may be too weak (depending on the latest bounds on the parameters of the Higgs potential) to allow electroweak baryogenesis, but most non-minimal extensions of this model (for example the inclusion of a new singlet superfield), allow a strongly first order phase transition consistent with current supersymmetry bounds. In grand unified models new CP violation may be associated with the scalar fields necessary to break the grand unified symmetry. Many examples of this type have been proposed.

This is in fact the best news from baryogenesis, especially electroweak baryogenesis. By bringing the physics of baryon production down to energies that we are currently probing in the laboratory, we have an opportunity to verify or falsify these ideas in detail. For example CP violation in the extensions of the standard model mentioned above, particularly supersymmetry, lead to observable effects at low energies, both CP conserving and CP violating. If the next round of collider experiments determine the nature of electroweak symmetry breaking, then the nature of the phase transition and its suitability for electroweak baryogenesis may be determined. If new CP violation is observed in experiments like the B factory, or in electric dipole moment experiments, it will be especially interesting to determine the flavor structure of this CP violation and its possible connection with the baryon asymmetry of the universe.

Although we have only touched on two broad areas of baryogenesis, electroweak and grand unified, there are a variety of other interesting ideas, including spontaneous baryogenesis, topological defects, *etc*. One of the more interesting variants, leptogenesis, involves the production of an asymmetry in lepton rather than baryon number. Subsequent production of baryon number then relies upon further processing of the lepton number asymmetry by interactions, like the electroweak interaction we have already discussed. These models are especially timely since the lepton asymmetry may be connected with the physics of neutrinos, an area where we are now beginning to obtain a great deal of experimental information.

The only bad news here, is the rather vague connection between baryogenesis and *specific* laboratory experiments. There is no single smoking gun; new CP violation large enough to produce the observed baryon asymmetry will almost certainly have low energy effects, but not decisively so. And where these effects show up, be it in EDMs, B or D mixing, or top quark physics, is highly model dependent. Without more experimental information constraining our current theoretical ideas, baryogenesis does not suggest that any one experiment is more likely than another to see new CP violation. But these are minor quibbles. Baryogenesis is already a strong indication of new physics to come, and even tells us that this new physics should emerge in one of the most fascinating areas of current research, CP violation.

Baryogenesis has been a fruitful cross-roads between particle physics and cosmology. Uniting ideas of early universe phase transitions, electroweak symmetry breaking and CP violation, it is an area that touches on many of the most exciting experiments that we look forward to in the coming decade. The B factory, the LHC, the Tevatron and even tabletop atomic physics experiments, may provide provide the clues that help explain the presence of matter in the universe. Unraveling the mystery of the cosmic baryon asymmetry remains one of the most exciting tasks for particle physicists and cosmologists alike.

# References


[1] E. W. Kolb and M. S. Turner. *The Early universe*. Addison-Wesley, 1990.

[2] A. G. Cohen, D. B. Kaplan, and A. E. Nelson. Progress in electroweak baryogenesis. *Ann. Rev. Nucl. Part. Sci.*, 43:27–70, 1993.

[3] V. A. Rubakov and M. E. Shaposhnikov. Electroweak baryon number non-conservation in the early universe and in high-energy collisions. *Usp. Fiz. Nauk*, 166:493–537, 1996.

[4] Antonio Riotto and Mark Trodden. Recent progress in baryogenesis. 1999. hep-ph/9901362 submitted to Ann. Rev. Nucl. Part. Sci.

[5] G. Steigman. Observational tests of antimatter cosmologies. *Ann. Rev. Astron. Astrophys.*, 14:339–372, 1976.

[6] S. P. Ahlen. The ams experiment to search for antimatter and dark matter. In *Proceedings of the 5th Annual LeCroy Conference on Electronics for Particle Physics, Chestnut Ridge, NY*, 1995.

[7] S. P. Ahlen. Ams: A magnetic spectrometer for the international space station. In S. J. Ball and Y. A. Kamyshkov, editors, *Proceedings of the International Workshop on Future Prospects of Baryon Instability Search in p decay and n—anti-n Oscillation Experiments, Oak Ridge, TN*, 1996.

[8] R. Battiston. The alpha magnetic spectrometer (ams): Search for antimatter and dark matter on the international space station. *Nucl. Phys. Proc. Suppl.*, 65:19, 1998.

[9] R. Battiston. The alpha magnetic spectrometer (ams). *Nucl. Instrum. Meth.*, A409:458, 1998.

[10] V. Plyaskin. Antimatter and dark matter search with the alpha magnetic spectrometer (ams). *Surveys High Energ. Phys.*, 13:177, 1998.

[11] J. Casaus. The alpha magnetic spectrometer (ams). *Acta Phys. Polon.*, B30:2445, 1999.

[12] U. Bekcer. Alpha magnetic spectrometer ams report on the first flight in space june 2-12. In *Proceedings of the 29th International Conference on High-Energy Physics (ICHEP 98), Vancouver, British Columbia, Canada*, 1998.

[13] A. G. Cohen, A. De Rujula, and S. L. Glashow. A matter-antimatter universe. *Astrophys. J.*, 495:539, 1998.

[14] S.C. Kappadath et al. In *Proc. 12th Int. Cosmic Ray Conf.*, page 25, 1995.

[15] William H. Kinney, Edward W. Kolb, and Michael S. Turner. Ribbons on the cbr sky: A powerful test of a baryon symmetric universe. *Phys. Rev. Lett.*, 79:2620–2623, 1997.

[16] A. G. Cohen and A. De Rujula. Scars on the cbr? *Astrophys. J.*, 496L:63, 1998.

[17] Keith A. Olive, Gary Steigman, and Terry P. Walker. Primordial nucleosynthesis: Theory and observations. 1999. astro-ph/9905320 submitted to Phys. Rep.

[18] A. D. Sakharov. Violation of cp invariance, c asymmetry, and baryon asymmetry of the universe. *JETP Letters*, 5:24, 1967.

[19] V. A. Kuzmin, V. A. Rubakov, and M. E. Shaposhnikov. On the anomalous electroweak baryon number nonconservation in the early universe. *Phys. Lett.*, B155:36, 1985.

[20] G. 't Hooft. Symmetry breaking through bell-jackiw anomalies. *Phys. Rev. Lett.*, 37:8–11, 1976.

[21] G. 't Hooft. Computation of the quantum effects due to a four- dimensional pseudoparticle. *Phys. Rev.*, D14:3432–3450, 1976.