INTRODUCTION

1. Overview

The *Review of Particle Physics* and the abbreviated version, the *Particle Physics Booklet,* are reviews of the field of Particle Physics. This complete *Review* includes a compilation/evaluation of data on particle properties, called the "Particle Listings." These Listings include 2,633 new measurements from 689 papers, in addition to the 21,926 measurements from 6,415 papers that first appeared in previous editions [1].

Both books include Summary Tables with our best values and limits for particle properties such as masses, widths or lifetimes, and branching fractions, as well as an extensive summary of searches for hypothetical particles. In addition, we give a long section of "Reviews, Tables, and Plots" on a wide variety of theoretical and experimental topics, a quick reference for the practicing particle physicist.

The *Review* and the *Booklet* are published in even-numbered years. This edition is an updating through January 2006 (and, in some areas, well into 2006). As described in the section "Using Particle Physics Databases" following this introduction, the content of this *Review* is available on the World-Wide Web, and is updated between printed editions (`http://pdg.lbl.gov/`).

The Summary Tables give our best values of the properties of the particles we consider to be well established, a summary of search limits for hypothetical particles, and a summary of experimental tests of conservation laws.

The Particle Listings contain all the data used to get the values given in the Summary Tables. Other measurements considered recent enough or important enough to mention, but which for one reason or another are not used to get the best values, appear separately just beneath the data we do use for the Summary Tables. The Particle Listings also give information on unconfirmed particles and on particle searches, as well as short "reviews" on subjects of particular interest or controversy.

The Particle Listings were once an archive of all published data on particle properties. This is no longer possible because of the large quantity of data. We refer interested readers to earlier editions for data now considered to be obsolete.

We organize the particles into six categories:

Gauge and Higgs bosons
Leptons
Quarks
Mesons
Baryons
Searches for monopoles, supersymmetry,
compositeness, extra dimensions, *etc.*

The last category only includes searches for particles that do not belong to the previous groups; searches for heavy charged leptons and massive neutrinos, by contrast, are with the leptons.

In Sec. 2 of this Introduction, we list the main areas of responsibility of the authors, and also list our large number of consultants, without whom we would not have been able to produce this *Review.* In Sec. 4, we mention briefly the naming scheme for hadrons. In Sec. 5, we discuss our procedures for choosing among measurements of particle properties

and for obtaining best values of the properties from the measurements.

The accuracy and usefulness of this *Review* depend in large part on interaction between its users and the authors. We appreciate comments, criticisms, and suggestions for improvements of any kind. Please send them to the appropriate author, according to the list of responsibilities in Sec. 2 below, or to the LBNL addresses below.

To order a copy of the *Review* or the *Particle Physics Booklet* from North and South America, Australia, and the Far East, send email to `PDG@LBL.GOV`

or via the web at:
(`http://pdg.lbl.gov/pdgmail`)

or write to:

Particle Data Group, MS 50R6008
Lawrence Berkeley National Laboratory
Berkeley, CA 94720-8166, USA

From all other areas, see
(`http://weblib.cern.ch/publreq.php`)

or write to

CERN Scientific Information Service
CH-1211 Geneva 23
Switzerland

2. Particle Listings responsibilities

* Asterisk indicates the people to contact with questions or comments about Particle Listings sections.

Gauge and Higgs bosons

γ	C. Grab, D.E. Groom*
Gluons	R.M. Barnett,* A.V. Manohar
Graviton	D.E. Groom*
W, Z	C. Caso,* A. Gurtu*
Higgs bosons	K. Hikasa, M.L. Mangano*
Heavy bosons	M. Tanabashi, T. Watari*
Axions	H. Murayama, G. Raffelt*

Leptons	
Neutrinos	M. Goodman, R. Miquel,* K. Nakamura, K.A. Olive, A. Piepke, P. Vogel
e, μ	J. Beringer,* C. Grab
τ	K.G. Hayes, K. Mönig*
Quarks	
Quarks	R.M. Barnett,* A.V. Manohar
Top quark, b'	J.L. Feng,* K. Hagiwara
Free quark	J. Beringer*
Mesons	
π, η	J. Beringer,* C. Grab
Unstable mesons	C. Amsler, M. Doser,* S. Eidelman, T. Gutsche, J.J. Hernández-Rey, A. Masoni, S. Navas, C. Patrignani, N.A. Törnqvist
K (stable)	G. D'Ambrosio, T.G. Trippe*
D (stable)	D. Asner, C.G. Wohl*
B (stable)	Y. Kwon, G. Punzi, J.G. Smith, W.-M. Yao*
Baryons	
Stable baryons	C. Grab, C.G. Wohl*
Unstable baryons	D.M. Manley, C.G. Wohl,* R.L. Workman
Charmed baryons	P.R. Burchat, C.G. Wohl*
Bottom baryons	Y. Kwon, J.G. Smith, G. Punzi, W.-M. Yao*
Miscellaneous searches	
Monopole	D.E. Groom*
Supersymmetry	A. De Gouvea, M.L. Mangano,* K.A. Olive, L. Pape
Technicolor	M. Tanabashi, J. Terning*
Compositeness	M. Tanabashi, J. Terning*
Extra Dimensions	T. Gherghetta*, C. Kolda
WIMPs and Other	J.L. Feng,* K. Hikasa

3. Consultants

The Particle Data Group benefits greatly from the assistance of some 700 physicists who are asked to verify every piece of data entered into this *Review*. Of special value is the advice of the PDG Advisory Committee which meets annually and thoroughly reviews all aspects of our operation. The members of the 2006 committee are:

H. Aihara (Tokyo)
G. Brooijmans (Columbia)
R. Voss (CERN), Chair
M. Whalley (Durham)
P. Zerwas (DESY)

We have especially relied on the expertise of the following people for advice on particular topics:

- M.N. Achasov (BINP, Novosibirsk)
- S.I. Alekhin (COMPAS Group, IHEP, Protvino)
- F. Ambrosino (Naples Univ.)
- M. Antonelli (INFN, Frascati)
- R. Bailey (CERN)
- J.-L. Basdevant (University of Paris)
- D. Bernard (Ecole Polytechnique, France)
- S. Bethke (MPI, Munich)
- S. Bianco (INFN, Frascati)
- I.I. Bigi (Notre Dame University)
- S. Bilenky (Joint Inst. for Nuclear Research, Dubna)
- M. Billing (Cornell University)
- G.C. Blazey (Northern Illinois University)
- A.E. Bondar (BINP, Novosibirsk)
- T. Browder (University of Hawaii)
- O. Bruening (CERN)
- D. Bryman (TRIUMF)
- G. Buchalla (Munich U.)
- M. Chanowitz (LBNL)
- V. Cirigliano (Caltech)
- F. Close (Oxford University)
- E.D. Commins (University of California, Berkeley)
- A. Correa dos Reis (Rio de Janeiro, CBPF)
- J. Cumalat (Colorado U.)
- P. Denes (LBNL)
- L. Di Ciaccio (LAPP, Annecy)
- R. Dixon (FNAL)
- A. Donnachie (University of Manchester)
- R.J. Donahue (LBNL)
- V.P. Druzhinin (BINP, Novosibirsk)
- G. Eigen (Bergen University)

- J. Elias (FNAL)
- M. Erdmann (RWTH Aachen)
- R. Faccini (University of Rome, Italy)
- A. Fasso (SLAC)
- M. Fidecaro (CERN)
- W. Fischer (Brookhaven National Lab)
- P. Franzini (Rome U. & Frascati)
- S.J. Freedman (University of California, Berkeley)
- H. Fritzsch (Ludwig-Maximilians University, Munich)
- M.A. Furman (LBNL)
- P. Gambino (INFN, Turin)
- T. Gershon (U. of Warwick)
- R. Godang (University of Mississippi)
- M.C. Gonzalez-Garcia (SUNY Stony Brook and IFIC Valencia)
- E.M. Gullikson (LBNL)
- R. Hagstrom (Argonne National Lab)
- J. Hardy (Texas A&M U.)
- F. Harris (University of Hawaii)
- S. Heinemeyer (CERN)
- B. Holstein (University of Massachusetts)
- G. Isidori (INFN, Frascati)
- P. Janot (CERN)
- J. Jowett (CERN)
- A. Juste (FNAL)
- R.W. Kadel (LBNL)
- S.G. Karshenboim (VNIIM, St-Petersburg)
- R.W. Kenney (LBNL)
- S. Kettell (Brookhaven National Lab)
- J. Konigsberg (University of Florida)
- R. Van Kooten(Indiana University)
- S. Kretzer (BNL)
- A. Kronfeld (FNAL)
- S.-I. Kurokawa (KEK)
- A.S. Kuzmin (BINP, Novosibirsk)
- H. Lacker (LAL-Orsay)
- G. Landsberg (Brown University)
- U. Langenegger (Heidelberg University)
- L. Lellouch (Marseille, CPT)
- K. Lesko (LBNL)
- E.B. Levichev (BINP, Novosibirsk)
- W. Lewis (Los Alamos National Lab)
- A. Limosani (KEK)
- E. Lisi (INFN Bari)
- F. Di Lodovico (University of Lodon)

- I.B. Logashenko (BINP, Novosibirsk)
- O. Long (University of California, Santa Barbara)
- E. Lorenz (MPI Munich)
- V. Luth (SLAC)
- G.R. Lynch (LBNL)
- H. Mahlke-Krüger (Cornell Univeristy, Ithaca)
- S. Malvezzi (INFN Milan)
- B. Meadows (U. of Cinncinnati)
- P.J. Mohr (NIST)
- D. Morgan (Rutherford Appleton Lab)
- B.M.K. Nefkens (University of California, Los Angeles)
- M. Neubert (Cornell University)
- A. Nyffeler (ETH Zuerich)
- S. Olsen (University of Hawaii)
- F. Parodi (University of Genova)
- M. Paulini (Carnegie Mellon University)
- M.R. Pennington (University of Durham)
- A. Pich (Valencia U., IFIC)
- K. Pitts (University of Illinois, Urbana)
- P. Raimondi (INFN, Frascati)
- J. Richman (University of California, Santa Barbara)
- T. Rizzo (SLAC)
- B.L. Roberts (Boston University)
- J.L. Rosner (University of Chicago)
- P. Roudeau (LAL, Orsay)
- A. Ryd (Cornell U.)
- G. Savard (Argonne Natl. Lab. and U. Chicago)
- G. Schierholz (DESY)
- J.T. Seeman (SLAC)
- S. Sharpe (University of Washington)
- Yu.M. Shatunov (BINP, Novosibirsk)
- B.A. Shwartz (BINP, Novosibirsk)
- P. Skands (FNAL)
- J. Smith (University of Colorado)
- M. Sozzi (Pisa Univ.)
- A. Stocchi (Orsay, LAL)
- S.L. Stone (Syracuse University)
- S.I. Striganov (COMPAS Group, IHEP, Protvino)
- M. Suzuki (LBNL)
- B.N. Taylor (NIST)
- E. Thomson (University of Pennsylvania)
- K.Yu. Todyshev (BINP, Novosibirsk)
- T. Iijima (KEK)
- J.F. de Troconiz (Autonomous University of Madrid)

- G. Unal (LAL, Orsay)
- N. Uraltsev (St. Petersburg, INP)
- R. Voss (CERN)
- A.J. Weinstein (California Institute of Technology)
- C. Weiser (CERN)
- F. Willeke (DESY)
- S. Willenbrock (University of Illinois)
- S. Willocq (University of Massachusetts, Amherst)
- J.E. Wiss (University of Illinois)
- C.Z. Yuan (IHEP, Beijing)
- A.M. Zaitsev (IHEP, Protvino)
- D. Zeppenfeld (University of Karlsruhe, Germany)
- P. Zerwas (DESY)
- C. Zhang (IHEP, Beijing)
- K. Zuber (Oxford University)

4. Naming scheme for hadrons

We introduced in the 1986 edition [2] a new naming scheme for the hadrons. Changes from older terminology affected mainly the heavier mesons made of u, d, and s quarks. Otherwise, the only important change to known hadrons was that the $F^{\pm}$ became the $D_s^{\pm}$. None of the lightest pseudoscalar or vector mesons changed names, nor did the $c\overline{c}$ or $b\overline{b}$ mesons (we do, however, now use χ_c for the $c\overline{c}$ χ states), nor did any of the established baryons. The Summary Tables give both the new and old names whenever a change has occurred.

The scheme is described in "Naming Scheme for Hadrons" (p. 108) of this *Review*.

We give here our conventions on type-setting style. Particle symbols are italic (or slanted) characters: e^-, p, Λ, π^0, K_L, D_s^+, b. Charge is indicated by a superscript: B^-, Δ^{++}. Charge is not normally indicated for p, n, or the quarks, and is optional for neutral isosinglets: η or η^0. Antiparticles and particles are distinguished by charge for charged leptons and mesons: τ^+, K^-. Otherwise, distinct antiparticles are indicated by a bar (overline): $\overline{\nu}_\mu$, $\overline{t}$, $\overline{p}$, $\overline{K}^0$, and $\overline{\Sigma}^+$ (the antiparticle of the Σ^-).

5. Procedures

5.1. ***Selection and treatment of data*** : The Particle Listings contain all relevant data known to us that are published in journals. With very few exceptions, we do not include results from preprints or conference reports. Nor do we include data that are of historical importance only (the Listings are not an archival record). We search every volume of 20 journals through our cutoff date for relevant data. We also include later published papers that are sent to us by the authors (or others).

In the Particle Listings, we clearly separate measurements that are used to calculate or estimate values given in the Summary Tables from measurements that are not used. We give explanatory comments in many such cases. Among the reasons a measurement might be excluded are the following:

- It is superseded by or included in later results.
- No error is given.
- It involves assumptions we question.
- It has a poor signal-to-noise ratio, low statistical significance, or is otherwise of poorer quality than other data available.
- It is clearly inconsistent with other results that appear to be more reliable. Usually we then state the criterion, which sometimes is quite subjective, for selecting "more reliable" data for averaging. See Sec. 5.4.
- It is not independent of other results.
- It is not the best limit (see below).
- It is quoted from a preprint or a conference report.

In some cases, *none* of the measurements is entirely reliable and no average is calculated. For example, the masses of many of the baryon resonances, obtained from partial-wave analyses, are quoted as estimated ranges thought to probably include the true values, rather than as averages with errors. This is discussed in the Baryon Particle Listings.

For upper limits, we normally quote in the Summary Tables the strongest limit. We do not average or combine upper limits except in a very few cases where they may be re-expressed as measured numbers with Gaussian errors.

As is customary, we assume that particle and antiparticle share the same spin, mass, and mean life. The Tests of Conservation Laws table, following the Summary Tables, lists tests of CPT as well as other conservation laws.

We use the following indicators in the Particle Listings to tell how we get values from the tabulated measurements:

- OUR AVERAGE—From a weighted average of selected data.
- OUR FIT—From a constrained or overdetermined multiparameter fit of selected data.
- OUR EVALUATION—Not from a direct measurement, but evaluated from measurements of related quantities.
- OUR ESTIMATE—Based on the observed range of the data. Not from a formal statistical procedure.
- OUR LIMIT—For special cases where the limit is evaluated by us from measured ratios or other data. Not from a direct measurement.

An experimentalist who sees indications of a particle will of course want to know what has been seen in that region in the past. Hence we include in the Particle Listings all reported states that, in our opinion, have sufficient statistical merit and that have not been disproved by more reliable data. However, we promote to the Summary Tables only those states that we feel are well established. This judgment is, of course, somewhat subjective and no precise criteria can be given. For more detailed discussions, see the minireviews in the Particle Listings.

5.2. ***Averages and fits*** : We divide this discussion on obtaining averages and errors into three sections: (1) treatment of errors; (2) unconstrained averaging; (3) constrained fits.

5.2.1. *Treatment of errors:* In what follows, the "error" δx means that the range $x \pm \delta x$ is intended to be a 68.3% confidence interval about the central value x. We treat this error as if it were Gaussian. Thus when the error *is* Gaussian, δx is the usual one standard deviation (1σ). Many experimenters now give statistical and systematic errors separately, in which case we usually quote both errors, with the statistical error first. For averages and fits, we then add the the two errors in quadrature and use this combined error for δx.

When experimenters quote asymmetric errors $(\delta x)^+$ and $(\delta x)^-$ for a measurement x, the error that we use for that measurement in making an average or a fit with other measurements is a continuous function of these three quantities. When the resultant average or fit $\overline{x}$ is less than $x - (\delta x)^-$, we use $(\delta x)^-$; when it is greater than $x + (\delta x)^+$, we use $(\delta x)^+$. In between, the error we use is a linear function of x. Since the errors we use are functions of the result, we iterate to get the final result. Asymmetric output errors are determined from the input errors assuming a linear relation between the input and output quantities.

In fitting or averaging, we usually do not include correlations between different measurements, but we try to select data in such a way as to reduce correlations. Correlated errors are, however, treated explicitly when there are a number of results of the form $A_i \pm \sigma_i \pm \Delta$ that have identical systematic errors Δ. In this case, one can first average the $A_i \pm \sigma_i$ and then combine the resulting statistical error with Δ. One obtains, however, the same result by averaging $A_i \pm (\sigma_i^2 + \Delta_i^2)^{1/2}$, where $\Delta_i = \sigma_i \Delta [\sum (1/\sigma_j^2)]^{1/2}$. This procedure has the advantage that, with the modified systematic errors Δ_i, each measurement may be treated as independent and averaged in the usual way with other data. Therefore, when appropriate, we adopt this procedure. We tabulate Δ and invoke an automated procedure that computes Δ_i before averaging and we include a note saying that there are common systematic errors.

Another common case of correlated errors occurs when experimenters measure two quantities and then quote the two and their difference, *e.g.*, m_1, m_2, and $\Delta = m_2 - m_1$. We cannot enter all of m_1, m_2 and Δ into a constrained fit because they are not independent. In some cases, it is a good approximation to ignore the quantity with the largest error and put the other two into the fit. However, in some cases correlations are such that the errors on m_1, m_2 and Δ are comparable and none of the three values can be ignored. In this case, we put all three values into the fit and invoke an automated procedure to increase the errors prior to fitting such that the three quantities can be treated as independent measurements in the constrained fit. We include a note saying that this has been done.

5.2.2. *Unconstrained averaging:* To average data, we use a standard weighted least-squares procedure and in some cases, discussed below, increase the errors with a "scale factor." We begin by assuming that measurements of a given quantity are uncorrelated, and calculate a weighted average and error as

$$\overline{x} \pm \delta\overline{x} = \frac{\sum_i w_i\, x_i}{\sum_i w_i} \pm (\textstyle\sum_i w_i\,)^{-1/2} \ , \tag{1}$$

where

$$w_i = 1/(\delta x_i)^2 \ .$$

Here x_i and δx_i are the value and error reported by the ith experiment, and the sums run over the N experiments. We then calculate $\chi^2 = \sum w_i(\overline{x} - x_i)^2$ and compare it with $N-1$, which is the expectation value of χ^2 if the measurements are from a Gaussian distribution.

If $\chi^2/(N-1)$ is less than or equal to 1, and there are no known problems with the data, we accept the results.

If $\chi^2/(N-1)$ is very large, we may choose not to use the average at all. Alternatively, we may quote the calculated average, but then make an educated guess of the error, a conservative estimate designed to take into account known problems with the data.

Finally, if $\chi^2/(N-1)$ is greater than 1, but not greatly so, we still average the data, but then also do the following:

(a) We increase our quoted error, $\delta\overline{x}$ in Eq. (1), by a scale factor S defined as

$$S = \left[\chi^2/(N-1)\right]^{1/2} \ . \tag{2}$$

Our reasoning is as follows. The large value of the χ^2 is likely to be due to underestimation of errors in at least one of the experiments. Not knowing which of the errors are underestimated, we assume they are all underestimated by the same factor S. If we scale up all the input errors by this factor, the χ^2 becomes $N-1$, and of course the output error $\delta\overline{x}$ scales up by the same factor. See Ref. 3.

When combining data with widely varying errors, we modify this procedure slightly. We evaluate S using only the experiments with smaller errors. Our cutoff or ceiling on δx_i is arbitrarily chosen to be

$$\delta_0 = 3N^{1/2}\ \delta\overline{x} \ ,$$

where $\delta\overline{x}$ is the unscaled error of the mean of all the experiments. Our reasoning is that although the low-precision experiments have little influence on the values $\overline{x}$ and $\delta\overline{x}$, they can make significant contributions to the χ^2, and the contribution of the high-precision experiments thus tends to be obscured. Note that if each experiment has the same error δx_i, then $\delta\overline{x}$ is $\delta x_i/N^{1/2}$, so each δx_i is well below the cutoff. (More often, however, we simply exclude measurements with relatively large errors from averages and fits: new, precise data chase out old, imprecise data.)

Our scaling procedure has the property that if there are two values with comparable errors separated by much more than their stated errors (with or without a number of

other values of lower accuracy), the scaled-up error $\delta\overline{x}$ is approximately half the interval between the two discrepant values.

We emphasize that our scaling procedure for *errors* in no way affects central values. And if you wish to recover the unscaled error $\delta\overline{x}$, simply divide the quoted error by S.

(b) If the number M of experiments with an error smaller than δ_0 is at least three, and if $\chi^2/(M-1)$ is greater than 1.25, we show in the Particle Listings an ideogram of the data. Figure 1 is an example. Sometimes one or two data points lie apart from the main body; other times the data split into two or more groups. We extract no numbers from these ideograms; they are simply visual aids, which the reader may use as he or she sees fit.

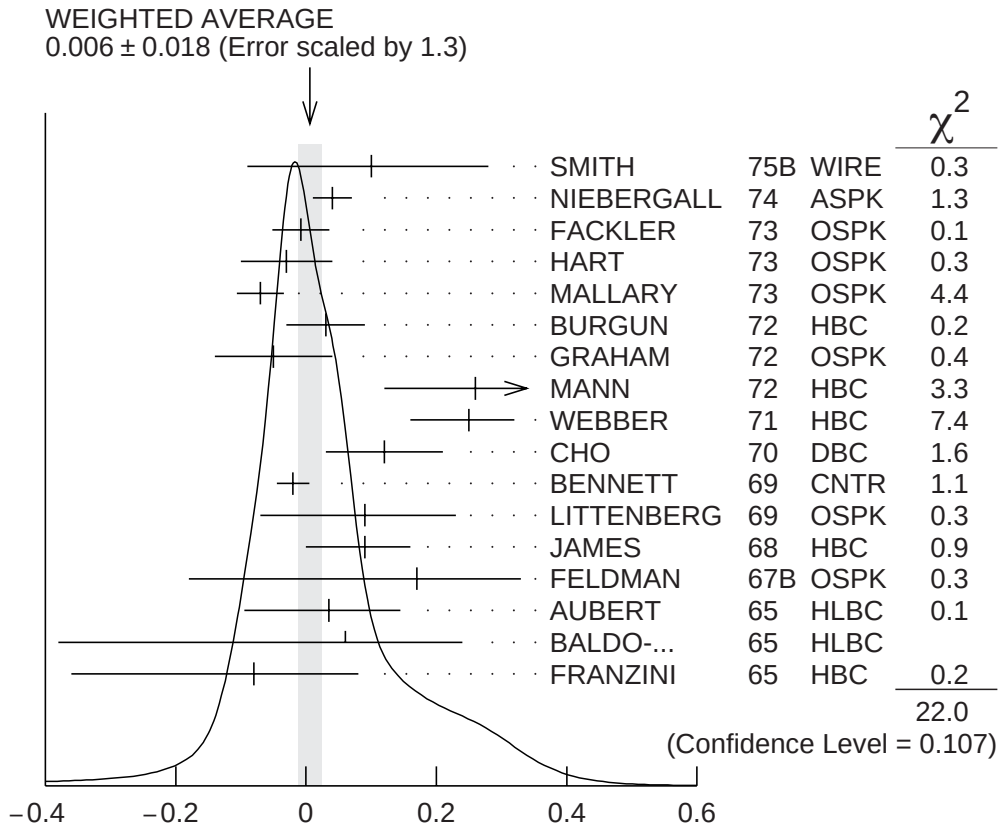


Figure 1: A typical ideogram. The arrow at the top shows the position of the weighted average, while the width of the shaded pattern shows the error in the average after scaling by the factor S. The column on the right gives the χ^2 contribution of each of the experiments. Note that the next-to-last experiment, denoted by the incomplete error flag ($\perp$), is not used in the calculation of S (see the text).

Each measurement in an ideogram is represented by a Gaussian with a central value x_i, error δx_i, and area proportional to $1/\delta x_i$. The choice of $1/\delta x_i$ for the area is somewhat arbitrary. With this choice, the center of gravity of the ideogram corresponds to an average that uses weights $1/\delta x_i$ rather than the $(1/\delta x_i)^2$ actually used in the averages. This may be appropriate when some of the experiments have seriously underestimated

systematic errors. However, since for this choice of area the height of the Gaussian for each measurement is proportional to $(1/\delta x_i)^2$, the peak position of the ideogram will often favor the high-precision measurements at least as much as does the least-squares average. See our 1986 edition [2] for a detailed discussion of the use of ideograms.

5.2.3. *Constrained fits:* In some cases, such as branching ratios or masses and mass differences, a constrained fit may be needed to obtain the best values of a set of parameters. For example, most branching ratios and rate measurements are analyzed by making a simultaneous least-squares fit to all the data and extracting the partial decay fractions P_i, the partial widths Γ_i, the full width Γ (or mean life), and the associated error matrix.

Assume, for example, that a state has m partial decay fractions P_i, where $\sum P_i = 1$. These have been measured in N_r different ratios R_r, where, e.g., $R_1 = P_1/P_2$, $R_2 = P_1/P_3$, *etc.* [We can handle any ratio R of the form $\sum \alpha_i P_i / \sum \beta_i P_i$, where α_i and β_i are constants, usually 1 or 0. The forms $R = P_i P_j$ and $R = (P_i P_j)^{1/2}$ are also allowed.] Further assume that *each* ratio R has been measured by N_k experiments (we designate each experiment with a subscript k, e.g., R_{1k}). We then find the best values of the fractions P_i by minimizing the χ^2 as a function of the $m-1$ independent parameters:

$$\chi^2 = \sum_{r=1}^{N_r} \sum_{k=1}^{N_k} \left(\frac{R_{rk} - R_r}{\delta R_{rk}} \right)^2 , \tag{3}$$

where the R_{rk} are the measured values and R_r are the fitted values of the branching ratios.

In addition to the fitted values $\overline{P}_i$, we calculate an error matrix $\langle \delta\overline{P}_i \; \delta\overline{P}_j \rangle$. We tabulate the diagonal elements of $\delta \overline{P}_i = \langle \delta \overline{P}_i \, \delta \overline{P}_i \rangle^{1/2}$ (except that some errors are scaled as discussed below). In the Particle Listings, we give the complete correlation matrix; we also calculate the fitted value of each ratio, for comparison with the input data, and list it above the relevant input, along with a simple unconstrained average of the same input.

Three comments on the example above:

(1) There was no connection assumed between measurements of the full width and the branching ratios. But often we also have information on partial widths Γ_i as well as the total width Γ. In this case we must introduce Γ as a parameter in the fit, along with the P_i, and we give correlation matrices for the widths in the Particle Listings.

(2) We try to pick those ratios and widths that are as independent and as close to the original data as possible. When one experiment measures all the branching fractions and constrains their sum to be one, we leave one of them (usually the least well-determined one) out of the fit to make the set of input data more nearly independent. We now do allow for correlations between input data.

(3) We calculate scale factors for both the R_r and P_i when the measurements for any R give a larger-than-expected contribution to the χ^2. According to Eq. (3), the double sum for χ^2 is first summed over experiments $k = 1$ to N_k, leaving a single sum over ratios $\chi^2 = \sum \chi^2_r$. One is tempted to define a scale factor for the ratio r as $S^2_r = \chi^2_r / \langle \chi^2_r \rangle$. However, since $\langle \chi^2_r \rangle$ is not a fixed quantity (it is somewhere between N_k and N_{k-1}), we

do not know how to evaluate this expression. Instead we define

$$S_r^2 = \frac{1}{N_k} \sum_{k=1}^{N_k} \frac{\left(R_{rk} - \overline{R}_r\right)^2}{\langle (R_{rk} - \overline{R}_r)^2 \rangle}. \tag{4}$$

With this definition the expected value of S_r^2 is one. We can show that

$$\langle (R_{rk} - \overline{R}_r)^2 \rangle = (\delta R_{rk})^2 - (\delta \overline{R}_r)^2 \ , \tag{5}$$

where $\delta \overline{R}_r$ is the fitted error for ratio r.

The fit is redone using errors for the branching ratios that are scaled by the larger of S_r and unity, from which new and often larger errors $\delta \overline{P}_i'$ are obtained. The scale factors we finally list in such cases are defined by $S_i = \delta \overline{P}_i' / \delta \overline{P}_i$. However, in line with our policy of not letting S affect the central values, we give the values of $\overline{P}_i$ obtained from the original (unscaled) fit.

There is one special case in which the errors that are obtained by the preceding procedure may be changed. When a fitted branching ratio (or rate) $\overline{P}_i$ turns out to be less than three standard deviations ($\delta \overline{P}_i'$) from zero, a new smaller error $(\delta \overline{P}_i'')^-$ is calculated on the low side by requiring the area under the Gaussian between $\overline{P}_i - (\delta \overline{P}_i'')^-$ and $\overline{P}_i$ to be 68.3% of the area between zero and $\overline{P}_i$. A similar correction is made for branching fractions that are within three standard deviations of one. This keeps the quoted errors from overlapping the boundary of the physical region.

5.3. ***Rounding*** : While the results shown in the Particle Listings are usually exactly those published by the experiments, the numbers that appear in the Summary Tables (means, averages and limits) are subject to a set of rounding rules.

The basic rule states that if the three highest order digits of the error lie between 100 and 354, we round to two significant digits. If they lie between 355 and 949, we round to one significant digit. Finally, if they lie between 950 and 999, we round up to 1000 and keep two significant digits. In all cases, the central value is given with a precision that matches that of the error. So, for example, the result (coming from an average) 0.827 ± 0.119 would appear as 0.83 ± 0.12, while 0.827 ± 0.367 would turn into 0.8 ± 0.4.

Rounding is not performed if a result in a Summary Table comes from a single measurement, without any averaging. In that case, the number of digits published in the original paper is kept, unless we feel it inappropriate. Note that, even for a single measurement, when we combine statistical and systematic errors in quadrature, rounding rules apply to the result of the combination. It should be noted also that most of the limits in the Summary Tables come from a single source (the best limit) and, therefore, are not subject to rounding.

Finally, we should point out that in several instances, when a group of results come from a single fit to a set of data, we have chosen to keep two significant digits for all the results. This happens, for instance, for several properties of the W and Z bosons and the τ lepton.

5.4. ***Discussion*** : The problem of averaging data containing discrepant values is nicely discussed by Taylor in Ref. 4. He considers a number of algorithms that attempt to incorporate inconsistent data into a meaningful average. However, it is difficult to develop a procedure that handles simultaneously in a reasonable way two basic types of situations: (a) data that lie apart from the main body of the data are incorrect (contain unreported errors); and (b) the opposite—it is the main body of data that is incorrect. Unfortunately, as Taylor shows, case (b) is not infrequent. He concludes that the choice of procedure is less significant than the initial choice of data to include or exclude.

We place much emphasis on this choice of data. Often we solicit the help of outside experts (consultants). Sometimes, however, it is simply impossible to determine which of a set of discrepant measurements are correct. Our scale-factor technique is an attempt to address this ignorance by increasing the error. In effect, we are saying that present experiments do not allow a precise determination of this quantity because of unresolvable discrepancies, and one must await further measurements. The reader is warned of this situation by the size of the scale factor, and if he or she desires can go back to the literature (via the Particle Listings) and redo the average with a different choice of data.

Our situation is less severe than most of the cases Taylor considers, such as estimates of the fundamental constants like $\hbar$, *etc.* Most of the errors in his case are dominated by systematic effects. For our data, statistical errors are often at least as large as systematic errors, and statistical errors are usually easier to estimate. A notable exception occurs in partial-wave analyses, where different techniques applied to the same data yield different results. In this case, as stated earlier, we often do not make an average but just quote a range of values.

A brief history of early Particle Data Group averages is given in Ref. 3. Figure 2 shows some histories of our values of a few particle properties. Sometimes large changes occur. These usually reflect the introduction of significant new data or the discarding of older data. Older data are discarded in favor of newer data when it is felt that the newer data have smaller systematic errors, or have more checks on systematic errors, or have made corrections unknown at the time of the older experiments, or simply have much smaller errors. Sometimes, the scale factor becomes large near the time at which a large jump takes place, reflecting the uncertainty introduced by the new and inconsistent data. By and large, however, a full scan of our history plots shows a dull progression toward greater precision at central values quite consistent with the first data points shown.

We conclude that the reliability of the combination of experimental data and our averaging procedures is usually good, but it is important to be aware that fluctuations outside of the quoted errors can and do occur.

ACKNOWLEDGMENTS

The publication of the *Review of Particle Physics* is supported by the Director, Office of Science, Office of High Energy and Nuclear Physics, the Division of High Energy Physics of the U.S. Department of Energy under Contract No. DE–AC02–05CH11231; by the U.S. National Science Foundation under Agreement No. PHY-0355084; by the European Laboratory for Particle Physics (CERN); by an implementing arrangement between the governments of Japan (Monbusho) and the United States (DOE) on cooperative research and development; and by the Italian National Institute of Nuclear Physics (INFN).

We thank all those who have assisted in the many phases of preparing this *Review*. We particularly thank the many who have responded to our requests for verification of data entered in the Listings, and those who have made suggestions or pointed out errors.

REFERENCES

1. The previous edition was Particle Data Group: S. Eidelman *et al.*, Phys. Lett. **B592**, 1 (2004).
2. Particle Data Group: M. Aguilar-Benitez *et al.*, Phys. Lett. **170B** (1986).
3. A.H. Rosenfeld, Ann. Rev. Nucl. Sci. **25**, 555 (1975).
4. B.N. Taylor, "Numerical Comparisons of Several Algorithms for Treating Inconsistent Data in a Least-Squares Adjustment of the Fundamental Constants," U.S. National Bureau of Standards NBSIR 81-2426 (1982).

Revised September 2005 by P. Kreitz (SLAC) with the substantial assistance in the Physics Education Section from L. Wolf (SLAC).

This annotated list provides a highly selective set of online resources that are useful to the particle physics community. It describes the scope, size, and organization of the resources so that efficient choices can be made amongst many sites which may appear similar. A resource is excluded if it provides information primarily of interest to only one institution. Because this list must be fixed in print, it is important to consult the updated version of this compilation which includes newly added resources and hypertext links to more complete information at:

http://www.slac.stanford.edu/library/pdg/

Accelerator physics resources have not been included but will be referenced in the next edition of this work.

My thanks to Betty Armstrong and Piotr Zyla, Particle Data Group, Travis Brooks and Kim Sutton, SLAC Library, and the many particle physics Web site and database maintainers who have all given me their generous assistance. Please send comments and corrections by e-mail to `pkreitz@slac.stanford.edu`.

1. Particles & Properties Data:

- REVIEW OF PARTICLE PHYSICS (RPP): A biennial comprehensive review summarizing much of the known data about the field of particle physics produced by the international Particle Data Group (PDG). Includes a compilation/evaluation of data on particle properties, summary tables with best values and limits for particle properties, extensive summaries of searches for hypothetical particles, and a long section of reviews, tables, and plots on a wide variety of theoretical and experimental topics of interest to particle and astrophysicists. The linked table of contents provides access to particle listings, reviews, summary tables, errata, indices, *etc.* The current printed version is S. Eidelman, *et al.*, Physics Letters **B592**, 1 (2004). On the web there is a partial update for the 2006 edition at:

  `http://pdg.lbl.gov/`

- PARTICLE PHYSICS BOOKLET: A pocket-sized 300-page booklet containing the Summary Tables and abbreviated versions of some of the other sections of the full *Review of Particle Physics*. This is extracted from the most recent edition of the full *Review of Particle Physics*. Contains images in an easy-to-read print useful for classroom studies. The next edition will be Summer 2006. Until the new edition is published and available via the Web, students, teachers, and researchers should use the full *Review of Particle Physics*:

  `http://pdg.lbl.gov/`

- COMPUTER-READABLE FILES: Currently available from the PDG: Tables of masses, widths, and PDG Monte Carlo particle numbers and cross-section data, including hadronic total and elastic cross sections vs laboratory momenta, and total center-of mass energy. The PDG Monte Carlo particle numbering scheme has been

updated for the recent edition of the RPP and are also available as a MobileDB database. Palm Pilot products include physical constants, astrophysical constants and particle properties. These files are is updated in even-numbered years coinciding with the production of the *Review of Particle Physics*:

`http://pdg.lbl.gov/2005/html/computer_read.html`

- PARTICLE PHYSICS DATA SYSTEM: This site contains an indexed bibliography of particle physics (1895–1995), a database of computerized numerical data extracted from experimental publications, and an index of papers (1895–present) that contain experimental data or data analyses. The Web interface permits simple searching for compilations of integrated cross-section data. The search interface for numerical data on observables in reactions (ReacData or RD), is under construction. Maintained by the COMPAS group at IHEP:

  `http://wwwppds.ihep.su:8001/ppds.html`

- HEPDATA: REACTION DATA DATABASE: A part of the HEPDATA databases at University of Durham/RAL, this database is compiled by the Durham Database Group (UK) with help from the COMPAS Group (Russia) for the PDG. Contains numerical values of HEP reaction data such as total and differential cross sections, fragmentation functions, structure functions, and polarization measurements from a wide range of experiments. Updated at regular intervals. Provides data reviews which contain precompiled reviewed data such as 'Structure Functions in DIS', 'Single Photon Production in Hadronic Interactions', and 'Drell-Yan Cross Sections':

  `http://durpdg.dur.ac.uk/HEPDATA/REAC`

- NIST PHYSICS LABORATORY: This unit of the National Institute of Standards and Technology provides measurement services and research for electronic, optical, and radiation technologies. Three sub-pages, on Physical Reference Data, on Constants, Units & Uncertainty, and on Measurements & Calibrations, are extremely useful. Additional links to other physical properties and data of tangential interest to particle physics are also available from this page:

  `http://physics.nist.gov/`

2. Collaborations & Experiments:

- EXPERIMENTS Database: Contains more than 2,400 past, present, and future experiments in elementary particle physics. Lists both accelerator and non-accelerator experiments. Includes official experiment name and number, location, spokespersons and collaboration lists. Simple searches by: participant, title, experiment number, institution, date approved, accelerator, or detector, return a result that fully describes the experiment, including a complete list of authors, title, description of the experiment's goals and methods, and a link to the experiment's Web page if available. Publication lists distinguish articles in refereed journals, theses, technical or instrumentation papers, and those which make the Topcite at 50+ subsequent citations or more:

`http://www.slac.stanford.edu/spires/experiments/`

- HIGH ENERGY PHYSICS EXPERIMENTS: A HEPiC page of links to experimental collaboration Web pages. Experiments are arranged alphabetically by name or number under 18 major laboratories or in a miscellaneous group of 'Others':

  `http://www.hep.net/experiments/all_sites.html`

- COSMIC RAY/GAMMA RAY/NEUTRINO AND SIMILAR EXPERIMENTS: This is an extensive collection of experimental Web sites organized by focus of study and also by location. Additional sections link to educational materials, organizations and related Web sites, *etc.* Maintained at the Max Planck Institute for Nuclear Physics by Konrad Bernlöhr:

  `www.mpi-hd.mpg.de/hfm/CosmicRay/CosmicRaySites.html`

3. Conferences:

- CONFERENCES: Database of more than 13,700 past, present and future conferences, schools, and meetings of interest to high-energy physics and related fields. Covers 1973 to the future. Each year lists more than 500 events. Search or browse by title, acronym, date, location. Includes information about published proceedings, links to submitted papers from the SPIRES-HEP database, and links to the conference Web site when available. Links to a form with which one can submit a new conference or edit an existing one:

  `http://www.slac.stanford.edu/spires/conferences/`

  `additions.shtml`

  to submit a new conference. Can also search for any conferences occurring by day, month, quarter, or year:

  `http://www.slac.stanford.edu/spires/conferences/`

- CONFERENCES AND CONFERENCES: Lists 150+ current meetings in many fields of physics. Provides post-conference information such as proceedings, *etc.*, in a second list at the end of the subject categories of conferences. Browse through an ASCII list of all conferences or specialized lists arranged by topic: particles/nuclei, quantum, condensed matter, mathematical, interdisciplinary physics, and related fields. Includes links to the conference Web page and the contact. Provides a useful set of links to universities, laboratories and institutions which host major conferences and/or schools:

  `http://www.physics.umd.edu/robot/confer/confmenu.html`

- CERN & HEP EVENTS: A list of current and upcoming conferences, schools, workshops, *etc.*, of interest to high-energy physicists. Organized by year and then by date. Covers from 1993 to 2010. Includes about 27,000 events. To post an event to this list use the Web form at:

  `http://cdsweb.cern.ch/events/`

- EUROPHYSICS MEETINGS LIST: Maintained by the European Physical Society, this lists in chronological order all the current and future meetings, workshops, schools, *etc.*, organized or sponsored by EPS or organized in conjunction with an EPS-sponsored group. Provides a PDF form to electronically submit a notice of a new conference or to print and mail to EPS:

  `http://www.eps.org/ephconf.html`

- PHYSICSWEB EVENTS: Part of the Institute of Physics (IOP) Web site, this site contains approximately a hundred entries for the current year's meetings, workshops, exhibitions and schools. Fills a gap by covering smaller conferences and workshops around the world. Searchable by type of event, e.g., school, workshop, or by date or free text words. Provides a Web form and email address for adding a conference and for signing up to receive email notices of new events added:

  `http://physicsweb.org/events/`

4. Current Notices & Announcement Services:

- See also the conference and event sites above for links to email notification services or event submission forms.

- CURRENT SCIENCE NEWS: Lists news sites from around the world. A few sites are by subscription and so are labelled as available to Stanford only but most are free and publically accessible. Commercial sites often provide headlines and a brief abstract as a free service and require subscription or payment for a complete article:

  `http://www.slac.stanford.edu/library/`

  `eresources/news.html`

- E-PRINT ARCHIVES LISTSERV NOTICES: The Cornell-based E-Print Archives provides daily notices of preprints in the fields of physics, mathematics, nonlinear sciences, computer science, and quantitative biology which have been submitted to the archives as full text electronic documents. Use the Web-accessible listings:

  `http://arXiv.org/`

  or subscribe:

  `http://arXiv.org/help/subscribe`

- HEPJOBS DATABASE: Maintained by Fermilab and SLAC libraries, this database lists jobs in the fields of core interest to the particle physics and astroparticle physics communities. Use this page to post a job or to receive email notices of new job listings:

  `http://www.slac.stanford.edu/spires/jobs/`

- INTERACTIONS.ORG: Provides an email newsletter covering particle physics news and resources from particle physics laboratories worldwide. Subscribe to Interactions.org Newswire:

  `http://www.interactions.org/cms/?pid=1000502`

- NASA ASTROPHYSICS DATA SYSTEM: This page provides access to the tables of contents of the most recent issues of selected journals in the field. It permits the user to select which titles are shown and eliminates the ones the user has already read:

  `http://adsabs.harvard.edu/custom_toc.html`

- PREPRINTS IN PARTICLES AND FIELDS (PPF): A weekly listing averaging 250 new preprints in particle physics and related fields. Contains bibliographic listings for and, in the Web version, full text links to, the new preprints received by and cataloged into the SPIRES High-Energy Physics (HEP) database. Includes that week's titles from the e-print archives as well as preprints and articles received from other sources. Directions for subscribing to an email version can be found on the page listing the most recent week's preprints:

  `http://www.slac.stanford.edu/library/documents/`
  `newppf.html`

- PSIGATE PHYSICS GATEWAY: IoP News and Jobs: Newsfeeds containing the latest jobs and news from the Institute of Physics' (IoP) PhysicsWeb with news headlines from Optics.org, Fibers.org and Nanotechweb.org:

  `http://www.psigate.ac.uk/newsite/awareness_iop.html`

Note: Use the library pages in Section 5.3 below to find additional announcement lists for recently received preprints, books, and proceedings. Use the online journal links in Section 7 below for journal table of contents.

5. Directories:

5.1. *Directories—Research Institutions:*

- HEP and Astrophysics INSTITUTIONS: SPIRES database of over 6,500 high-energy physics and astroparticle physics institutes, laboratories, and university departments in which research on particle physics is performed. Covers six continents and over a hundred countries. Provides an alphabetical list by country or an interface that is searchable by name, acronym, location, *etc.* Includes address, phone and fax numbers, e-mail address, and Web links where available. Has links to the recent HEP papers from each institution. Maintained by SLAC, DESY and Fermilab libraries.

  For the list of institutions by country:

  `http://www.slac.stanford.edu/spires/`
  `institutions/online_institutions.shtml`

  To search the Institutions database:

  `http://www.slac.stanford.edu/spires/institutions/`

- HEP INSTITUTES: Contains almost a thousand institutional addresses used in the CERN Library catalog. Includes, where available, the following: phone and fax

numbers; e-mail addresses; and Web links. Provides free text searching and result sorting by organization, country, or town:

`http://cdsweb.cern.ch/?c=HEP%20Institutes`

- MISCELLANEOUS SITES OF COSMIC-RAY AND ASTROPARTICLE RESEARCH: A listing by continent or region alphabetical listing of institutions involved in astroparticle physics and cosmic-ray research. Part of a lengthy website covering Cosmic Ray, Gamma Ray, Neutrino and Similar Experiments:

  `http://www.mpi-hd.mpg.de/hfm/CosmicRay/`
  `CosmicRaySites.html`

- TOP 500 HEP AND ASTROPHYSICS INSTITUTIONS BY COUNTRY: Lists the 500 major HEP-related organizations and universities that have published the most papers in the past five years, as identified from the SPIRES HEP Database. Provides active links to the home pages and full INSTITUTIONS database records. Listed by country, and then alphabetically by institution:

  `http://www.slac.stanford.edu/spires/inst/`
  `major.shtml`

5.2. *Directories—People:*

- HEPNAMES: Searchable worldwide database of over 42,000 people associated with particle physics, astroparticle physics, synchrotron radiation, and related fields. Provides e-mail addresses, country in which the person is currently working, and a SPIRES HEP database search for their papers. If the person has supplied the following information, it lists the countries in which they did their undergraduate and graduate work, their url, and their graduate students. It also provides listings of Nobel Laureates, country statistics, Lab Directors, *etc.*:

  `http://www.slac.stanford.edu/spires/hepnames/`

- HEP VIRTUAL PHONEBOOK: A list of links to phonebooks and directories of high-energy physics sites and collaborations around the world organized by site. Often provides links to more specialized phone or e-mail listings, such as a department within a university, visiting scientists, or postdocs. Some phonebooks may require passwords or other authentication to access. Maintained by HEPiC, and many linked phonebooks are still active, however this web site was last updated in 2002:

  `http://www.hep.net/sites/directories.html`

- US-HEPFOLK DIRECTORY: A searchable directory and census of U.S. particle physicists updated annually. Contains more than 4,000 U.S. physicists. Searchable by first or last name, by affiliation, and/or by email address. Includes data on the most recent survey results and historical data back to 1995:

  `http://hepfolk.lbl.gov`

5.3. ***Directories—Libraries:***

- Argonne National Laboratory (ANL) Library:
  `http://www.library.anl.gov/`
- Brookhaven National Laboratory (BNL) Library:
  `http://inform.bnl.gov/RESLIB/reslib.html`
- (CERN) European Organization for Nuclear Research Library:
  `http://library.cern.ch/`
- Deutsches Elektronen-Synchrotron (DESY) Library:
  `http://www.desy.de/html/infodienste/bibliothek.html`
- Fermi National Accelerator Laboratory (Fermilab) Library:
  `http://lss.fnal.gov/ird/index.html`
- Idaho National Engineering and Environmental Laboratory (INEEL) Library:
  `http://www.inel.gov/library/`
- (KEK) National Laboratory for High Energy Physics Library:
  `http://www-lib.kek.jp/top-e.html`
- Lawrence Berkeley National Laboratory (LBNL) Library:
  `http://www-library.lbl.gov/teid/`
  `tmLib/aboutus/LibDefault.htm`
- Lawrence Livermore National Laboratory (LLNL) Library:
  `http://www.llnl.gov/library/`
- Los Alamos National Laboratory (LANL) Library:
  `http://lib-www.lanl.gov/`
- Oak Ridge National Laboratory (ORNL) Library:
  `http://www.ornl.gov/Library/library-home.html`
- Pacific Northwest National Laboratory (PNL) Library:
  `http://libraryweb.pnl.gov`
- Sandia National Laboratory Library:
  `http://www.sandia.gov/news-center/resources/tech-library/index.html`
- Stanford Linear Accelerator Center (SLAC) Library:
  `http://www.slac.stanford.edu/library`
- Thomas Jefferson National Accelerator Facility (JLab) Library:
  `http://www.jlab.org/IR/library/index.html`

5.4. *Directories—Publishers:*

- DIRECTORY OF PUBLISHERS AND VENDORS: Outstanding and comprehensive directory of publishers and vendors used by libraries. Organized by publisher name, by subject (e.g. Science, Mathematics, and Technology), and by location. Also provides an email directory.

  `http://acqweb.library.vanderbilt.edu/acqweb/pubr.html`

5.5. *Directories—Scholarly Societies:*

- American Association for the Advancement of Science:

  `http://www.aaas.org/`

- American Association of Physics Teachers:

  `http://www.aapt.org/`

- American Astronomical Society:

  `http://www.aas.org`

- American Institute of Physics:

  `http://www.aip.org/`

- American Mathematical Society:

  `http://www.ams.org/`

- American Physical Society:

  `http://www.aps.org`

- American Physical Society: Scholarly Societies: Use this list to find national and international scientific and professional societies:

  `http://www.aps.org/resources/society.html`

- European Physical Society:

  `http://www.eps.org/publications.html`

- IEEE Nuclear and Plasma Sciences Society:

  `http://ewh.ieee.org/soc/nps/aboutnpss.htm`

- Institute of Physics:

  `http://www.iop.org/`

- International Union of Pure and Applied Physics:

  `http://www.iupap.org/`

- Japan Society of Applied Physics:

  `http://www.jsap.or.jp/english/`

- Physical Society of Japan:

  `http://wwwsoc.nii.ac.jp/jps/`

- Physical Society of the Republic of China:

  `http://psroc.phys.ntu.edu.tw/english/index.html`

- SCHOLARLY SOCIETIES PROJECT: Directory of more than 4,000 scholarly and technical societies with links to their Web sites. Permits searching by subject, country, language, founding dates, and more. Includes acronyms and indicates when a Web site contains both its native language and an English-language version and when it has a permanent URL. Provides direct links to society meeting and conference announcement lists, standards, and full text journals. Maintained by the University of Waterloo:

  `http://www.scholarly-societies.org/`

6. E-Prints/Pre-Prints, Papers, & Reports:

- CERN ARTICLES & PREPRINTS: The CERN document server contains records of more than 600,000 CERN and non-CERN articles, preprints, theses. Includes records for CERN Yellow Reports, internal and technical notes, and official CERN committee documents. Provides access to full text of the documents for about 50 percent of the entries and to the references when available:

  `http://cdsweb.cern.ch/?c=Articles+%26+Preprints&as=0`

- ECONF: Electronic Conference Proceedings Archive: This site offers a fully electronic, Web-accessible archive for the proceedings of scientific conferences in High-Energy Physics and related fields. Conference editors can use the site tools to prepare and post an electronic version of their proceedings. Librarians and other indexers can download metadata from each proceedings. Researchers can browse an entire proceedings via a table of contents or search for papers through a link to the SPIRES HEP Database which indexes the EConf contents:

  `http://www.slac.stanford.edu/econf/`

- HEP DATABASE (SPIRES): Contains over 630,000 bibliographic records for particle physics articles, including journal papers, preprints, e-prints, technical reports, conference papers and theses. Comprehensively indexed with multiple links to full text as well as links to author and institutional information. Covers 1974 to the present with substantial older materials added. Updated daily with links to electronic texts, Durham Reaction Data, *Review of Particle Physics*, *etc.* Searchable by citation, by all authors and authors' affiliations, title, topic, report number, citation (footnotes), e-print archive number, date, journal, *etc.* A joint project of the SLAC and DESY libraries with the collaboration of Fermilab, Durham University (UK), KEK, Kyoto University, and many other research institutions and scholarly societies:

  `http://www.slac.stanford.edu/spires/hep/`

- JACoW: This Joint Accelerator Conference Website is organized by the editorial boards of the Asian, European and American Particle Accelerator Conferences and the CYCLOTRONS, DIPAC, ICALEPCS and LINAC conferences. It contains the

full text of all the papers of these accelerator conferences. Search by conference name, author, title, keyword or full text of the paper:

`http://www.JACoW.org/`

- KISS (KEK INFORMATION SERVICE SYSTEM) FOR PREPRINTS: KEK Library preprint and technical report database. Contains bibliographic records of preprints and technical reports held in the KEK library with links to the full text images of more than 100,000 papers scanned from their worldwide collection of preprints. Particularly useful for older scanned preprints:

  `http://www-lib.kek.jp/KISS/kiss_prepri.html`

- arXiv.org E-PRINT ARCHIVE: The arXiv.org is an automated electronic repository of full text papers in physics, mathematics, computer, and nonlinear sciences and cosmology and quantitative biology. Papers, called pre-prints or e(electronic)-prints, are usually sent by their authors to arXiv in advance of submission to a journal for publication. Primarily covers 1991 to the present but authors are encouraged to post older papers retroactively. Permits searching by author, title, and keyword in abstract. Allows limiting by subfield archive or by date:

  `http://arXiv.org`

- NASA ASTROPHYSICS DATA SYSTEM: The ADS Abstract Service provides a search interface for four bibliographic databases covering: Astronomy and Astrophysics, Instrumentation, Physics and Geophysics, and arXiv Preprints. Contains abstracts from articles and monographs as well as conference proceedings:

  `http://adsabs.harvard.edu/ads_abstracts.html`

- PARTICLE PHYSICS DATA SYSTEM—PPDS: A search interface to the bibliography of the print publication *A Guide to Experimental Elementary Particle Physics Literature* (LBL-90). This bibliography covers the published literature of theoretical and experimental particle physics from 1895 to 1995. The url is sometimes difficult to reach:

  `http://wwwppds.ihep.su:8001/ppds.html`

- DIRECTORY OF MATHEMATICS PREPRINT AND E-PRINT SERVERS: Provides the current home page and email contacts for mathematical preprint and e-print servers throughout the world:

  `http://www.ams.org/global-preprints/`

7. Particle Physics Journals & Reviews:

7.1. *Online Journals and Tables of Contents:*

Please note, some of these journals, publishers, and reviews may limit access to subscribers. If you encounter access problems, check with your institution's library.

- ADVANCES IN THEORETICAL AND MATHEMATICAL PHYSICS (ATMP): Bi-monthly electronic and hard copy publication. Table of contents has links to arXiv.org since this is the first e-journal to be an overlay to arXiv.org, where papers for this journal are submitted:

  `http://www.intlpress.com/journals/ATMP/archive/`

- AMERICAN JOURNAL OF PHYSICS: A monthly publication of the American Association of Physics Teachers on instructional and cultural aspects of physical science:

  `http://ojps.aip.org/ajp`

- APPLIED PHYSICS LETTERS: Weekly publication of short (3 pages maximum) articles:

  `http://ojps.aip.org/aplo/`

- ASTROPHYSICAL JOURNAL: Published three times a month by the American Astronomical Society (AAS). See also AAS entry under Journal Publishers (below):

  `http://www.journals.uchicago.edu/ApJ/`

- CLASSICAL AND QUANTUM GRAVITY: Published 24 times a year by the Institute of Physics (IOP) covering the fields of gravitation and spacetime theory:

  `http://www.iop.org/Journals/cq`

- EUROPEAN PHYSICAL JOURNAL A: HADRONS AND NUCLEI: This monthly journal merges *Il Nuovo Cimento A* and *Zeitschrift fur Physik A* and covers physics and astronomy:

  `http://www.springerlink.com/`

  `openurl.asp?genre=journal&issn=1434-6001`

- EUROPEAN PHYSICAL JOURNAL C: PARTICLES AND FIELDS: This twice monthly journal is the successor to *Zeitschrift fur Physik C*, covering physics and astronomy:

  `http://www.springerlink.com/openurl.asp?genre=`

  `journal&issn=1434-6044`

- INTERNATIONAL JOURNAL OF MODERN PHYSICS C: PHYSICS AND COMPUTERS: Includes both review and research articles. Published ten times per year:

  `http://ejournals.wspc.com.sg/ijmpc/ijmpc.shtml`

- INTERNATIONAL JOURNAL OF MODERN PHYSICS D: GRAVITATION, ASTROPHYSICS AND COSMOLOGY: Includes both review and research articles. Published ten times per year:

  `http://ejournals.wspc.com.sg/ijmpd/ijmpd.shtml`

- INTERNATIONAL JOURNAL OF MODERN PHYSICS E: NUCLEAR PHYSICS: Includes both review and research articles. Bi-monthly:

  `http://ejournals.wspc.com.sg/ijmpe/ijmpe.shtml`

- JAPANESE JOURNAL OF APPLIED PHYSICS: Part 1 is monthly and covers papers, short notes, and review papers. Part 2 is semi-monthly and publishes letters including a special *Express Letters* section:

  `http://www.ipap.jp/jjap/index.htm`

- JOURNAL OF COSMOLOGY AND ASTROPARTICLE PHYSICS: An electronic peer-reviewed journal created by the International School for Advanced Studies (SISSA) and the Institute of Physics. Authors are encouraged to submit media files to enhance the online versions of articles:

  `http://jcap.sissa.it/`

- JOURNAL OF HIGH ENERGY PHYSICS: Electronic and print available. Like *ATMP*, this is a refereed journal written, run, and distributed by electronic means. It accepts email submission notices and 'fetches' the submitted paper from the arXiv.org E-print archives:

  `http://jhep.sissa.it/`

- JOURNAL OF PHYSICS G: NUCLEAR AND PARTICLE PHYSICS: Monthly, published by IOP:

  `http://www.iop.org/EJ/journal/0954-3899`

- JOURNAL OF THE PHYSICAL SOCIETY OF JAPAN: JPSJ ONLINE: Monthly, online since 1996:

  `http://jpsj.ipap.jp/`

- MODERN PHYSICS LETTERS A: Published 40 times a year, this contains research papers in gravitation, cosmology, nuclear physics, and particles and fields. *Brief Review* section for short reports on new findings and developments:

  `http://www.wspc.com.sg/journals/mpla/mpla.html`

- NEW JOURNAL OF PHYSICS: Co-owned by the Institute of Physics and the Deutsche Physikalische Gesellschaft, this journal is funded by article charges from authors of published papers and by scholarly societies, *NJP* is available in a free, electronic form:

  `http://www.iop.org/EJ/njp`

- NUCLEAR INSTRUMENTS AND METHODS IN PHYSICS RESEARCH A: ACCELERATORS, SPECTROMETERS, DETECTORS, AND ASSOCIATED

EQUIPMENT: This journal was formerly part of *Nuclear Instruments and Methods in Physics Research*. Published approximately 36 times per year, this journal covers instrumentation and large scale facilities:

`http://www.sciencedirect.com/science/journal/01689002`

- NUCLEAR PHYSICS A: NUCLEAR AND HADRONIC PHYSICS:

  `http://www.sciencedirect.com/science/journal/03759474`

- NUCLEAR PHYSICS B: PARTICLE PHYSICS, FIELD THEORY, STATISTICAL SYSTEMS, AND MATHEMATICAL PHYSICS:

  `http://www.sciencedirect.com/science/journal/05503213`

- NUCLEAR PHYSICS B: PROCEEDINGS SUPPLEMENTS: Publishes proceedings of international conferences and topical meetings in high-energy physics and related areas:

  `http://www.sciencedirect.com/science/journal/09205632`

- PHYSICAL REVIEW D: PARTICLES, FIELDS, GRAVITATION, AND COSMOLOGY: Published 24 times a year:

  `http://prd.aps.org/`

- PHYSICAL REVIEW SPECIAL TOPICS – ACCELERATORS AND BEAMS: A peer-reviewed electronic journal freely available from the American Physical Society:

  `http://prst-ab.aps.org/`

- PHYSICS LETTERS B: Nuclear and Particle Physics: Published weekly:

  `http://www.sciencedirect.com/science/`
  `journal/03702693`

- PHYSICS—USPEKHI: English edition of *Uspekhi Fizicheskikh Nauk*:

  `http://ufn.ioc.ac.ru/`

- PROGRESS IN PARTICLE AND NUCLEAR PHYSICS: Published four times a year. Many, but not all, articles are at a level suitable for the general nuclear and particle physicist: item `http://www.sciencedirect.com/science/journal/01466410`

- PROGRESS OF THEORETICAL PHYSICS: Published monthly covering all fields of theoretical physics. A supplement is published roughly quarterly containing either long original or review papers or collections of papers on specific topics:

  `http://ptp.ipap.jp/journal/`

7.2. *Journals – Directories:*

- DESY Library Electronic Journals: A comprehensive collection of up-to-date links to electronic journals of interest to particle physics. Provides a further link to tables of contents services. Use the drop down menu of subjects to browse over 350 titles. Each title includes links to the publisher if available. Some access may be limited by your local library's licensing and subscription agreements:

  `http://www-library.desy.de/eljnl.html`

7.3. *Journals – Publishers & Repositories:*

- NASA ASTROPHYSICS DATA SYSTEM: Provides free electronic access to back issues of the *Astrophysical Journal*, *Astrophysical Journal Letters*, and the *Astrophysical Journal Supplement Series* and to many other titles. Often a journal allows the ADS to provide free, full text access after a delay of some period of time which can be several years:

  `http://adswww.harvard.edu/`

- AIP JOURNAL CENTER: The American Institute of Physics' top-level page for their electronic journals may be found at:

  `http://www.aip.org/ojs/service.html`

- AMERICAN PHYSICAL SOCIETY: The top-level page for the APS research journals. From this page one can access their *Physical Review* Online Archive (PROLA) search engine which is free to users:

  `http://publish.aps.org/`

- ELSEVIER SCIENCE: This Web site lists all Elsevier journal titles alphabetically and also enables browsing by subject field. First select Physical Sciences and then on the next page you must select either physics or astronomy (no longer both) and then subsequently select a sub-field of physics or astronomy:

  `http://www.elsevier.com/wps/find/journal_`
  `browse.cws_home`

- EUROPEAN PHYSICAL SOCIETY: This is the top-level page listing the society's journals:

  `http://www.eps.org/publications.html`

- INSTITUTE OF PHYSICS (IOP): Journals: Information: A list of the IOP journals organized by subject. A page organized by title is also available linked to this page:

  `http://www.iop.org/EJ/S/3/418/main/-list=subject`

- SPRINGER PUBLISHING: Physics & Astronomy: From this link, one can reach a subdiscipline in physics or astronomy:

  `http://www.springeronline.com/sgw/cda/`
  `frontpage/0,11855,4-10100-0-0-0,00.html`

7.4. ***Review Publications:***

- LIVING REVIEWS IN RELATIVITY: A peer-refereed, solely online physics journal publishing invited reviews covering all areas of relativity. Provided as a free service to the scientific community. Published in yearly volumes, although articles appear throughout the year. Hyperlinks are kept checked and active and reviews are updated frequently:

  `http://relativity.livingreviews.org/`

- NET ADVANCE OF PHYSICS: A free electronic service providing review articles and tutorials in an encyclopedic format. Covers all areas of physics. Includes e-prints, book announcements, full text of electronic books, and other resources with hypertext links when available. Welcomes contributions of original review articles:

  `http://web.mit.edu/redingtn/www/netadv/welcome.html`

- PHYSICS REPORTS: A review section for *Physics Letters A* and *Physics Letters B*. Each report deals with one subject. The reviews are specialized in nature, more extensive than a literature survey but normally less than book length:

  `http://www.sciencedirect.com/science/journal/03701573`

- REPORTS ON PROGRESS IN PHYSICS: Covers all areas of physics and is published monthly. All papers are free for 30 days from the date of online publication:

  `http://stacks.iop.org/RoPP`

- REVIEWS OF MODERN PHYSICS: Published quarterly, it includes traditional scholarly reviews and shorter colloquium papers intended to describe recent research of interest to a broad audience of physicists:

  `http://rmp.aps.org/`

8. Particle Physics Education Sites:

Please note, each site in this section containing student activities now lists the (U.S. educational system) school grade level(s) that best fit that site. Also listed are the National Science Education Content Standards for teaching science which are relevant to the classroom activities provided at that site. Further explanation of the National Science Educational Content Standards can be found at:

For Grades 5 - 8:

`http://www.nap.edu/readingroom/books/`
`nses/html/6d.html`

For Grades 9 - 12:

`http://www.nap.edu/readingroom/books/`
`nses/html/6e.html`

8.1. *Particle Physics Education: General Sites:*

- ARGONNE NATIONAL LABORATORY K-12 PROGRAMS: Includes links to a variety of information and programs such as ArthmAttack, NEWTON, and the Rube Goldberg Machine Contest.

  `http://www.dep.anl.gov/p_k-12/`

- CONTEMPORARY PHYSICS EDUCATION PROJECT (CPEP): Provides charts, brochures, Web links, and classroom activities. Online interactive courses include: Fundamental Particles and Interactions; Plasma Physics and Fusion; and Nuclear Science.

  `http://www.cpepweb.org/`

- DEPARTMENT OF ENERGY (DOE) ONLINE k-12 INSTRUCTIONAL RESOURCES: The U.S. Department of Energy (DOE) brings together a collection of online resources and lesson plans from the education sites of DOE-funded national laboratories such as Stanford Linear Accelerator Center, Lawrence Berkeley, Jefferson Lab, and Brookhaven. In the area of atomic and particle physics, see Jefferson Lab's All About Atoms slide show and clickable interactive table of elements which enables you to find out an element's properties, history, and uses.

Grades: K-12; A variety of lesson plans are available in all scientific disciplines, conforming to National Science Education Content Standards.

`http://www-ed.fnal.gov/doe/`

- FERMILAB EDUCATION OFFICE: Outstanding collection of resources from the "grandmother" of all physics lab educational programs. Sections are organized for students and educators by grade level and for general visitors. See in particular the Phriendly Physics Bibliography.

  `http://www-ed.fnal.gov/`

- PARTICLE PHYSICS EDUCATION SITES: This rich site maintained by the Particle Data Group provides links to many other educational sites. Organizes the links by subject, level, and type of educational experience.

Grades 9-12; National Science Education Content Standards: Varies according to site.

`http://particleadventure.org/particleadventure/other/othersites.html`

- PHYSICAL SCIENCE: EDUCATIONAL HOTLISTS: Created by the outstanding Franklin Institute Science Museum, these hotlists contain a pre-screened list of resources for science educators, students, and enthusiasts. The criteria for inclusion is that a site stimulates creative thinking and learning about science. The excellent Physical Science list contains useful links for physics, physicists, optics, material science, applied design and engineering, sites for museums, 'doing science,' and inventors and engineers. Teacher resources include The Physical Science Activity

Manual which contains 34 hands-on activities for the classroom. Included are Newtonian Physics for grades 9-12 and activities such as Floating Objects which may be appropriate for grades 5-8. The Project Labs offer student-centered experiments in the areas of general science, physical science, and the natural, biological and environmental sciences.

Grades K-12; National Science Education Content Standards: Varies according to site visited from the listing.

`http://sln.fi.edu/tfi/hotlists/hotlists.html`

- PHYSLINK.COM: EDUCATION: This site provides sub-lists of online resources in the following areas: History of Physics and Astronomy; Essays on the interface between science, art, religion and philosophy; Astronomy; Graduate School and Student Advice; Software (reviews); References and Learning Sites for Educators; Youth Science; and New Theories.

  `http://www.physlink.com/Education/Index.cfm`

8.2. *Particle Physics Education: Background Knowledge:*

- ALBERT EINSTEIN ONLINE: A meta-Einstein site with links to dozens of resources by and about this scientist. Organized into Overviews; Moments (recollections of Einstein by others); Physics; Writings; Quotes; Pictures; and Miscellaneous.

  `http://www.westegg.com/einstein/`

- ANTIMATTER: MIRROR OF THE UNIVERSE: Find out what antimatter is, where it is made, the history behind its discovery, and how it is a part of our lives. This award-winning site, sponsored by the European Organisation for Nuclear Research, (CERN), explains to big kids and little kids alike the truth (and fiction) about antimatter. Features colorful photos and illustrations, a Kids Corner, and CERN physicists answering your questions on antimatter.

Grades 8-12+; National Science Education Content Standards: A, B, D, G

`http://livefromcern.web.cern.ch/livefromcern/antimatter/index.html`

- BIG BANG SCIENCE–EXPLORING THE ORIGINS OF MATTER: In clear, concise, yet elegant language, this Web site, produced by the Particle Physics and Astronomy Research Council of the UK (PPARC), explains what physicists are looking for with their giant instruments called accelerators and particle detectors. Includes a brief history on how scientists came to define what is fundamental in the universe. Big Bang Science focuses on CERN particle detectors and on United Kingdom scientists' contribution to the search for the fundamental building blocks of matter. In addition to information on the how and why of particle physics, this site also shows particle physics as an international collaborative endeavor.

Grades 9-12; National Science Education Content Standards B, D, E, G

`http://hepwww.rl.ac.uk/pub/bigbang/part1.html`

- LIFE, THE UNIVERSE, AND THE ELECTRON: Sponsored by the Institute of Physics (IOP) and the Science Museum, London, this interactive online exhibit celebrates the centenary of the discovery of the electron. Sections explain many aspects of the nature, history, and usefulness of electrons. Clear explanations and beautiful photography.

Grades 9-12; National Science Education Content Standards: A, B, D

`http://www.sciencemuseum.org.uk/on-line/electron/index.asp`

- SLAC VIRTUAL VISITOR'S CENTER: This Stanford Linear Accelerator Center Web site explains basic particle physics, linear and synchrotron accelerators, electron gamma showers, cosmic rays,and the experiments conducted at SLAC, including real-world applications. Intended for the general public as well as teachers and students.

Grades 9-12; National Science Education Content Standards: A, B, D, E, G

`http://www2.slac.stanford.edu/vvc/Default.htm`

- STEPHEN HAWKING'S UNIVERSE: Developed to accompany the Public Broadcast Station (PBS)'s television series hosted by Stephen Hawking, this web site provides a suite of interesting materials. Sections include: TV Schedule/Programs; Strange Stuff Explained; Universes; Cosmological Stars; Unsolved Mysteries; Things to Do in the Dark; Teacher's Guide; About Stephen Hawking and Ask the Experts.

`http://www.pbs.org/wnet/hawking/html/home.html`

- THE WORLD OF BEAMS: A site to visit if you wish to know a little or a lot about laser beams, particle beams, and other kinds of beams. Includes interactive tutorials, such as: What are Beams?, Working with Beams, and Beam Research and Technology. A good resource for physical science units involving energy, structure and properties of matter, and motion and forces for Grades 8-12. The information here is also helpful if you plan to tour any of the national laboratories listed in the "Libraries" section of this guide.

Grades 8-12; National Science Education Content Standards: A, B, E

`http://bc1.lbl.gov/CBP_pages/educational/WoB/home.htm`

8.3. ***Particle Physics Education:***
Particle Physics Lessons and Activities:

- BNL/BSA ONLINE CLASSROOM: The objective of this site, developed by Brookhaven National Laboratory, is to use technology to bring the scientific research of BNL to students and teachers. BNL has created a series of nine units comprising an online, interactive classroom and provided a Multi User Object-Oriented (MOO) virtual classroom that enables group interactivity. Students can test their knowledge of physics by playing the delightfully interactive RHIC Adventure (Secrets of the Nucleus) that focuses on the science of the Relativistic Heavy Ion Collider. Games can be modified to match student knowledge levels from grades 8-12. Lesson plans are available on nuclear physics for high school and on solar neutrinos for K-8. Try The Mystery of the Sun for grades K-8 or Dippin' Dots Neutrinos for grades 9-12. Each lesson includes National Science Education Content Standards.

Grades K-12 (mostly 9-12); National Science Education Content Standards 5-8 and 9-12: A, B, D, E, G

`http://onlineclassroom.bnl.org/`

- CENTER FOR PARTICLE ASTROPHYSICS ON-LINE DEMOS: A good source for do-it-yourself demonstrations about physics and astronomy aimed at middle school students (modifiable for other levels). Demonstrations include: Air-Powered Rockets, Desktop Stars, Lunar Topography, Ping Pong Ball Launcher, Potato Power, and Solar System. Each includes an introduction, teacher and student worksheets, and a list of materials needed. Site has not been updated with new materials, but existing lesson plans are nevertheless well-written and relevant. Parents might be interested in doing some of these projects with their children.

Grades 7-8+; National Science Education Content Standards: A, B, D, E

`http://cfpa.berkeley.edu/Education/DEMOS/DEMOS.html`

- CONTEMPORARY PHYSICS EDUCATION PROJECT (CPEP): This site is especially designed to help teachers bring four areas of physics to their students in an accessible and engaging format. Provides charts, brochures, Web links, and classroom activities. Online interactive courses include: Fundamental Particles and Interactions (includes lesson plans), Plasma Physics and Fusion, and Nuclear Science (includes lesson plans and simple experiments).

Grade Level: 9-12; Some of the experiments may be of interest to grades 5-8; National Science Education Content Standards 9-12: A, B, D, E

`http://www.cpepweb.org/`

- FERMILAB EDUCATION OFFICE: Outstanding collection of resources from the "grandmother" of all physics lab educational programs. Thoughtful unit and lesson plans in both physics and the environment (Fermilab is located on a rare, protected prairie in Illinois). Sections are organized by grade level. Note in particular

pedagogical resources for teachers, LInC Online, and the Lederman Science Center. Offers online guided tours and and science adventures.

Grades K-12; National Science Education Content Standards: A, B, C, D, E, F. Many lesson plans designed to meet Illinois State Standards

`http://www-ed.fnal.gov/`

- GLAST CLASSROOM MATERIALS: The Gamma Ray Large Area Space Telescope (GLAST) project and the National Aeronautics and Space Administration (NASA)'s Education and Public Outreach Office have developed this colorful, in depth, and engaging Web site teaching about the origin and structure of the universe and the fundamental relationship between energy and matter. Includes lesson plans and a teacher resource booklet which are available in PDF format, HTML, or can be ordered in print. Lesson plans are hands-on, student-group oriented and require common household objects. Activities such as: Three Mysteries, Alien Bandstand, Live! From 2-Alpha, and Starmarket build critical thinking and analytical skills as well as address at least one of the physical science standards. Full color posters and other educational materials also available. Provides links to other educational Web sites with classroom resources.

Grades 9-12; National Science Education Content Standards A, B, D, E

`http://glast.sonoma.edu/teachers/teachers.html`

- JEFFERSON LAB SCIENCE EDUCATION: This well-organized, visually attractive Web site from the Thomas Jefferson National Accelerator Facility, supports science and math education in K-12 classrooms. Features hands-on physics activities, math games, and puzzles. Check out the All About Atoms slide show and the interactive Table of Elements. Includes a-question-and-answer page on Atoms, Elements and Molecules and one on Electricity and Magnetism. Science videos are available on loan.

Grades K-12; Lessons follow Virginia State and National Science Education Content Standards

`http://education.jlab.org/`

- THE PARTICLE ADVENTURE: One of the most popular Web sites for learning the fundamentals of matter and force. Created by the Particle Data Group of Lawrence Berkeley National Laboratory. An award winning, interactive tour of the atom, with visits to quarks, neutrinos, antimatter, extra dimensions, dark matter, accelerators and particle detectors. Simple elegant graphics and translations into eleven languages. May be used by teachers or by students alone or in groups.

Grades 9-12; National Science Education Content Standards A, B, D

`http://ParticleAdventure.org`

- QUARKNET: QuarkNet brings the excitement of particle physics research to high school teachers and their students. Teachers join research groups at sixty universities

and labs across the country. These research groups are part of particle physics experiments at CERN, Fermilab, or SLAC. Students learn fundamental physics as they participate in inquiry-oriented investigations and analyze live, online data. QuarkNet is supported in part by the National Science Foundation and the U.S. Department of Energy.

Grades 9-12; National Science Education Content Standards: A, B, E

`http://QuarkNet.fnal.gov`

8.4. *Particle Physics Education: Astronomy Lessons and Experiments:*

- CENTER FOR PARTICLE ASTROPHYSICS ON-LINE DEMOS: A good source for do-it-yourself demonstrations about physics and astronomy aimed at middle school students (modifiable for other levels). Demonstrations include: Air-Powered Rockets, Desktop Stars, Lunar Topography, Ping Pong Ball Launcher, Potato Power, and Solar System. Each includes an introduction, teacher and student worksheets, and a list of materials needed. Site has not been updated with new materials, but existing lesson plans are nevertheless well-written and relevant. Parents might be interested in doing some of these projects with their children.

  Grades 7-8+; National Science Education Content Standards: A, B, D, E

  `http://cfpa.berkeley.edu/Education/DEMOS/DEMOS.html`

- HANDS-ON UNIVERSE: Enables students in middle and high schools to investigate the night sky without having to stay out late. Created by a collaboration of teachers and students including the Lawrence Hall of Science at the U.C., Berkeley, it uses high quality astronomical images to explore central concepts in math, science, and technology. Students analyze real images with image processing software similar to that used by professional astronomers. Lesson plans and activities are specifically tied to National Science Education Content Standards A and D, Science as Inquiry, and Earth and Space Science. Schools or districts much purchase the software, teacher and student booklets and materials. PDF color versions of the booklets are available from the Web, as well as a number of lesson plans and materials that do not require the purchased software.

  Grades 5-8 and 9-12; National Science Education Content Standards A, D

  `http://www.handsonuniverse.org`

- IMAGINE THE UNIVERSE: Created by the Laboratory for High-Energy Astrophysics at NASA/Goddard Space Flight Center, this site features astronomy and astrophysics lesson plans for age 14 and up, teacher's guides, classroom posters, and links to other classroom resources. Activities are linked to National Standards for Science and Math. Lessons include: What is Your Cosmic Connection to the Elements, Life Cycle of Stars, and Gamma-Ray Bursts? Also included in the Teacher's Corner are links to

math-science lesson plans for grades 6-12. The Multimedia Theatre Archive provides more than a dozen movies with free downloadable viewing software.

Grades 9-12; National Science Education Content Standards: A, B, D, G

`http://imagine.gsfc.nasa.gov`

- *Also note:* STARCHILD: Interlinked with Imagine the Universe, above, this site is a lively, age appropriate site for grade school level astronomy lessons.

  `http://starchild.gsfc.nasa.gov/docs/StarChild/StarChild.html`

- SPACE TODAY ONLINE: This news magazine covers space from Earth to the edge of the universe. It provides news, history, encyclopedia-like explanations of terms, activities, people and events, historical summaries, etc. and an outstanding collection of images covering all aspects of space.

  `http://www.spacetoday.org/STO.html`

- WINDOWS TO THE UNIVERSE: Provides a rich array of material for exploring earth and space, physics, geology, and chemistry in K-12 classrooms. Includes numerous, thorough lesson plans on topics ranging from the solar system to atmosphere and weather to physics and chemistry. Student-centered activities such as Building a Magnetometer or Create Your Own Cloud are simple, yet highly engaging. Content standards are detailed for most lesson plans. The People section of this vast but well-organized site traces the history of human scientific inquiry from Archimedes to Stephen Hawking. Three reading levels.

  Grades K-12; National Science Education Content Standards A, B, D, E, G

  `http://www.windows.ucar.edu/`

8.5. *Particle Physics Education: Ask-a-Scientist Sites:*

- ASK A SCIENTIST SERVICE: Questions are answered by volunteer scientists throughout the world. Service provided by the Newton BBS through Argonne National Lab. Submission form permits very specific age information to be included with the question so that the answer can be targeted to the questioner's level of knowledge.

  `http://www.newton.dep.anl.gov/`

- ASK THE EXPERTS: Submit questions via a form to scientists at PhysLink.com. Questions are answered free. Submission form warns that they won't answer questions from homework assignments or help design something for a science fair or competition. Has links to commonly asked questions and to a list of the most active scientists who provide answers.

  `http://www.physlink.com/Education/AskExperts/Index.cfm`

- HOW THINGS WORK: The author of the popular book, *How Things Work: the Physics of Everyday Life*, has created a site that functions as a virtual 'radio call-in

program'. Submit questions about how something works or consult the sixty plus pages of most recent questions which are searchable by date, topic, or keyword.

`http://howthingswork.virginia.edu/`

- MAD SCIENTIST'S NETWORK: ASK A QUESTION: Scientists at this Web site respond to hundreds of questions a week. Be sure to check out their extensive archive of answered questions and use their Science Fair Links for ideas for projects. Also note questions they decline to answer.

`http://www.madsci.org/submit.html`

8.6. *Particle Physics Education: Experiments, Demos,* & *Fun*

- ALL ABOUT LIGHT: From Fermilab, this offers a delightful collection of pages giving classical, relativistic and quantum explanations of light. Advanced placement high school level or above.

Grades 11-12+

`http://www.fnal.gov/pub/inquiring/more/light/index.html`

- CANTEACH: PHYSICAL SCIENCE: Canadian elementary teachers have put together a list of investigations and hands-on physics experiments for elementary level. These well-written physical science lesson plans feature such activities as Making a Pinhole Camera, Air Takes Space, Acid and Basic Test, Growing Crystals, Potential and Kinetic Energy, and Evaporation Painting.

Grades K-4; National Science Education Standards A, B, E

`http://www.canteach.ca/elementary/physical.html`

- THE EDIBLE/INEDIBLE EXPERIMENTS ARCHIVE: Part of the Mad Scientist's Network, this Web site covers astronomy, mathematics, and physics. Each experiment uses common materials and identifies whether the experiment is edible, inedible, partially drinkable, or not all that edible (!).

Grades K-8

`http://www.madsci.org/experiments/`

- HELPING YOUR CHILD LEARN SCIENCE: A wonderful introduction and set of tools for parents of elementary-age children compiled by the U.S. Department of Education. Provides ideas, home experiments, community-based science activities, and more.

Grades K-4

`http://www.ed.gov/pubs/parents/Science/index.html`

- INSULTINGLY STUPID MOVIE PHYSICS: An entertaining and educational site to learn how many movie special effects violate the laws of physics. Includes a rating

system for movie reviews. Heavy on text, with few graphics. Equations are included. A good way to emphasize, at the high school level, the immutability of the laws of physics in the real world. Provides instructions on how to use movie physics in the classroom and a bibliography.

Grades 9-12;

`http://www.intuitor.com/moviephysics`

- PHYSICS/PHYS/SCI DEMOS: This Web site provides over fifty physics demonstrations on the topics of density, motion, force, angular measurement, waves and sound, electricity and magnetism, optics and nuclear physics. Some of the demos feature photographs. Most of the demos are original, although a few were taken from the T.V. program, *Newton's Apple*. The high school teacher who created this site has won both a Presidential Award for Excellence in Mathematics and Science Teaching and the 2003 Classroom Connect Internet Educator of the Year Award. hfill

  Grades to 5-8 and 9-12; National Science Education Content Standards 5-8 and 9-12: A, B, E

  `http://www.darylscience.com/DemoPhys.html`

8.7. *Particle Physics Education: Physics History and Diversity Sites:*

- AIP CENTER FOR HISTORY OF PHYSICS: This site, produced by the American Institute of Physics, aims to preserve and make known the history of modern physics and allied fields including astronomy, geophysics, and optics. Of interest to teachers and students is the Exhibit Hall, with award-winning exhibits including photos and facts about Marie Curie, Einstein, the discovery of the electron, and the invention of the transistor.

Grades 7-12; National Science Education Content Standard: G

`http://www.aip.org/history`

- A CENTURY OF PHYSICS: This is the top-level page for a timeline from the American Physical Society providing a comprehensive history of major physical science developments with a selection of other events from society, art, politics and literature. Links on this page provide a physical timeline, an index, a search system and reproductions of the posters and images.

Grades 7-12; National Science Education Content Standard: G

`http://timeline.aps.org/APS/`

- CONTRIBUTIONS OF 20TH CENTURY WOMEN TO PHYSICS: A great resource for that history of science paper, this archive features descriptions of important contributions to science made by 83 women in the 20th century. Provides historical essays and links to additional documentation such as primary source materials.

Grades 7-12+; National Science Education Content Standard G

http://cwp.library.ucla.edu

- EDUCATION AND OUTREACH COMMITTEE ON THE STATUS OF WOMEN IN PHYSICS: Interested in inspiring a young woman to pursue physics? This American Physical Society site features *Physics in Your Future*, which conveys the exciting possibilities of a career in physics to middle and high school girls. Copies of this four-color booklet are available at no charge to students and their parents, educators, guidance counselors, and groups who work with young women. Also available in PDF. The popular, full color, *Celebrate Women in Physics* poster, is also available at no charge.

  Grades 7-12; National Science Education Content Standard G

  http://www.aps.org/educ/cswp/

- LIFE, THE UNIVERSE, AND THE ELECTRON: Sponsored by the Institute of Physics (IOP) and the Science Museum, London, this interactive online exhibit celebrates the centenary of the discovery of the electron. Sections explain many aspects of the nature, history, and usefulness of electrons. Clear explanations and beautiful photography.

  Grades 9-12; National Science Education Content Standards: A, B, D.

  http://www.sciencemuseum.org.uk/on-line/
  electron/index.asp

- NOBEL LAUREATES IN PHYSICS 1901-PRESENT: Maintained by SLAC, this site provides very comprehensive information on physics laureates. Links to the Nobel Foundation's pages on each laureate. Also lists the location(s) of the laureate's prize-winning work, where, if appropriate, the laureate is currently working, and where she or he was working when the work was done. Links to books, related Web sites, and to the HEP Database for in-depth bibliography. An interesting Quick Facts section provides great trivia about some of the prize winners.

  Grades 7-12; National Science Education Content Standard: G

  http://www.slac.stanford.edu/library/nobel/index.html

- PHYSLINK.COM HISTORY OF PHYSICS AND ASTRONOMY: This site, which is a compendium of other history of physics, astronomy and science sites, organizes that historical world into: general guides, histories of physics, of astronomy and space exploration, and of mathematics, online archives, museums and exhibits, and famous scientists. Serves as a guide to some of the most well known people and events in the physical sciences.

  Grades 7-12; National Science Education Content Standard: G

  http://www.physlink.com/Education/History.cfm

8.8. *Particle Physics Education: Art in Physics:*

Note: This modest collection of physics art links is provided for high school art, photography and literature teachers who may be interested in the intersections between science and technology and art and literature, or who wish to take an interdisciplinary approach to the curriculum in collaborating with their science department colleagues.

- DESY IN A SPECIAL LIGHT: Six luminescent pages of particle physics technology photographed at the Deutsches Electronen Synchrotronen Laboratory (DESY) by Peter Ginter, German photographer, in 1997.

  `http://www.peterginter.de/01technology/desy_01.html`

- HIDDEN CATHEDRALS–SCIENCE OR ART?: This page provides roughly seventeen dramatic color images of the inner workings of particle detectors at the European Organisation for Nuclear Research (CERN) which is the world's largest particle physics center.

  `http://public.web.cern.ch/Public/Content/Chapters/`
  `AboutCERN/HowStudyPrtcles/ScienceOrArt/ScienceOrArt-en.html`

- PHYSICS ICONS: A video by Chip Dalby, SLAC Sci/Arts Media Group, showing particle physics as delicate, experiential art. This meditation on the shifting nature of physics iconography was featured in the New York Museum of Modern Art's P.S.1 exhibit, *Signatures of the Invisible*.

  `http://www-project.slac.stanford.edu/streaming-media/`
  `Sub-Movies.html`

- PRESS PHOTO PRIZE FOR CERN: This article describes the photos done by Peter Ginter for CERN. The photos won a third prize from the World Press Photo of the Year competition in 1998.

  `http://bulletin.cern.ch/9911/art4/Text_E.html`

- ESSAYS AND BOOKS ON ART IN PHYSICS AND SCIENCE:

  "Art and Physics–a Beautiful Friendship"

  `http://bulletin.cern.ch/9949/art1/Text_E.html`

  "Art and Physics" by Leonard Shlain

  `http://www.artandphysics.com/h_main.html`

  "Physics Meets Art and Literature"

  `http://physicsweb.org/article/world/15/11/7`

  "Signatures of the Invisible"

  `http://www.ps1.org/cut/press/signatures.html`

9. Software Directories:

- CERNLIB: CERN PROGRAM LIBRARY: A large collection of general purpose libraries and modules offered in both source code and object code forms from the CERN central computing division. Provides programs applicable to a wide range of physics research problems such as general mathematics, data analysis, detectors simulation, data-handling, *etc.* Also includes links to commercial, free, and other software:

  `http://wwwasd.web.cern.ch/wwwasd/index.html`

- FREEHEP: A collection of software and information about software useful in high-energy physics. Searching can be done by title, subject, date acquired, date updated, or by browsing an alphabetical list of all packages:

  `http://www.slac.stanford.edu/find/fhmain.html`

- FERMITOOLS: Fermilab's software tools program provides a repository of Fermilab-developed software packages of value to the HEP community. Permits searching for packages by title or subject category:

  `http://fermitools.fnal.gov/`

- HEPIC: SOFTWARE & TOOLS USED IN HEP RESEARCH: A meta-level site with links to other sites of HEP-related software and computing tools:

  `http://www.hep.net/resources/software.html`

- GRID PHYSICS NETWORK: The GriPhyN Project is developing grid technologies for scientific and engineering projects that collect and analyze distributed, petabyte-scale datasets. Provides links to project information such as documents, education, workspace, virtual data toolkits, Chimera and Sphinx, as well as people, activities and news and related projects:

  `http://www.griphyn.org/index.php`

- PARTICLE PHYSICS DATA GRID: The Web site for the U.S. collaboration of federal laboratories and universities to build a worldwide distributed computing model for current and future particle and nuclear physics experiments:

  `http://www.ppdg.net/`

10. Specialized Subject Pages:

10.1. *Subject Pages*

- CAMBRIDGE RELATIVITY: PUBLIC HOME PAGE: These pages focus on the non-technical learner and explain aspects of relativity such as: cosmology, black holes, cosmic strings, inflation, and quantum gravity. Provides links to movies, research-level home pages and to Stephen Hawking's Web site:

  `http://www.damtp.cam.ac.uk/user/gr/public/`

- THE OFFICIAL STRING THEORY WEB SITE: Outstanding compilation of information about string theory includes: basics, mathematics, experiments, cosmology,

black holes, people (including interviews with string theorists), history, theater, links to other Web sites and a discussion forum:

`http://superstringtheory.com/`

- RELATIVITY: BOOKMARKS: Presents over 100 links collected into subject or other logical divisions. Unfortunately, the site owner explains in a note that he has not been able to verify the links for awhile. However, it still represents one of the best initial collections on the subject:

  `http://physics.syr.edu/research/relativity/`
  `RELATIVITY.html`

- RELATIVITY ON THE WORLD WIDE WEB: An excellent set of pages offering links and written information about relativity. Organized into: popular science sites; visualization sites; Web tutorials; observational and experimental evidence and rebuttals; course work (divided into undergraduate and graduate levels); software; research frontiers; and further reading:

  `http://math.ucr.edu/home/baez/relativity.html`

- SUPERSTRINGS: An online introduction to superstring theory for the advanced student. Includes further links:

  `http://www.sukidog.com/jpierre/strings/`

- THE ULTIMATE NEUTRINO PAGE: This page provides a gateway to an extremely useful compilation of experimental data and results:

  `http://cupp.oulu.fi/neutrino/`

1. PHYSICAL CONSTANTS

Table 1.1. Reviewed 2005 by P.J. Mohr and B.N. Taylor (NIST). Based mainly on the "CODATA Recommended Values of the Fundamental Physical Constants: 2002" by P.J. Mohr and B.N. Taylor, Rev. Mod. Phys. **77**, 1 (2005). The last group of constants (beginning with the Fermi coupling constant) comes from the Particle Data Group. The figures in parentheses after the values give the 1-standard-deviation uncertainties in the last digits; the corresponding fractional uncertainties in parts per 10^9 (ppb) are given in the last column. This set of constants (aside from the last group) is recommended for international use by CODATA (the Committee on Data for Science and Technology). The full 2002 CODATA set of constants may be found at `http://physics.nist.gov/constants`

Quantity	Symbol, equation	Value	Uncertainty (ppb)
speed of light in vacuum	c	299 792 458 m s^{-1}	exact*
Planck constant	h	$6.626\ 0693(11)\times10^{-34}$ J s	170
Planck constant, reduced	$\hbar \equiv h/2\pi$	$1.054\ 571\ 68(18)\times10^{-34}$ J s	170
		$= 6.582\ 119\ 15(56)\times10^{-22}$ MeV s	85
electron charge magnitude	e	$1.602\ 176\ 53(14)\times10^{-19}$ C $= 4.803\ 204\ 41(41)\times10^{-10}$ esu	85, 85
conversion constant	$\hbar c$	197.326 968(17) MeV fm	85
conversion constant	$(\hbar c)^2$	0.389 379 323(67) GeV^2 mbarn	170
electron mass	m_e	0.510 998 918(44) MeV/c^2 $= 9.109\ 3826(16)\times10^{-31}$ kg	86, 170
proton mass	m_p	938.272 029(80) MeV/c^2 $= 1.672\ 621\ 71(29)\times10^{-27}$ kg	86, 170
		$= 1.007\ 276\ 466\ 88(13)$ u $= 1836.152\ 672\ 61(85)\ m_e$	0.13, 0.46
deuteron mass	m_d	1875.612 82(16) MeV/c^2	86
unified atomic mass unit (u)	(mass ^{12}C atom)/12 = (1 g)/(N_A mol)	931.494 043(80) MeV/c^2 $= 1.660\ 538\ 86(28)\times10^{-27}$ kg	86, 170
permittivity of free space	$\epsilon_0 = 1/\mu_0 c^2$	$8.854\ 187\ 817\ \ldots\ \times10^{-12}$ F m^{-1}	exact
permeability of free space	μ_0	$4\pi \times 10^{-7}$ N A^{-2} $= 12.566\ 370\ 614\ \ldots\ \times10^{-7}$ N A^{-2}	exact
fine-structure constant	$\alpha = e^2/4\pi\epsilon_0\hbar c$	$7.297\ 352\ 568(24)\times10^{-3} = 1/137.035\ 999\ 11(46)^\dagger$	3.3, 3.3
classical electron radius	$r_e = e^2/4\pi\epsilon_0 m_e c^2$	$2.817\ 940\ 325(28)\times10^{-15}$ m	10
(e^- Compton wavelength)/2π	$\lambda\!\!\!^{-}_e = \hbar/m_e c = r_e\alpha^{-1}$	$3.861\ 592\ 678(26)\times10^{-13}$ m	6.7
Bohr radius ($m_{\rm nucleus} = \infty$)	$a_\infty = 4\pi\epsilon_0\hbar^2/m_e e^2 = r_e\alpha^{-2}$	$0.529\ 177\ 2108(18)\times10^{-10}$ m	3.3
wavelength of 1 eV/c particle	hc/(1 eV)	$1.239\ 841\ 91(11)\times10^{-6}$ m	85
Rydberg energy	$hcR_\infty = m_e e^4/2(4\pi\epsilon_0)^2\hbar^2 = m_e c^2\alpha^2/2$	13.605 6923(12) eV	85
Thomson cross section	$\sigma_T = 8\pi r_e^2/3$	0.665 245 873(13) barn	20
Bohr magneton	$\mu_B = e\hbar/2m_e$	$5.788\ 381\ 804(39)\times10^{-11}$ MeV T^{-1}	6.7
nuclear magneton	$\mu_N = e\hbar/2m_p$	$3.152\ 451\ 259(21)\times10^{-14}$ MeV T^{-1}	6.7
electron cyclotron freq./field	$\omega^e_{\rm cycl}/B = e/m_e$	$1.758\ 820\ 12(15)\times10^{11}$ rad s^{-1} T^{-1}	86
proton cyclotron freq./field	$\omega^p_{\rm cycl}/B = e/m_p$	$9.578\ 833\ 76(82)\times10^{7}$ rad s^{-1} T^{-1}	86
gravitational constant‡	G_N	$6.6742(10)\times10^{-11}$ m^3 kg^{-1} s^{-2}	1.5×10^5
		$= 6.7087(10)\times10^{-39}\ \hbar c$ $(\text{GeV}/c^2)^{-2}$	1.5×10^5
standard gravitational accel.	g_n	9.806 65 m s^{-2}	exact
Avogadro constant	N_A	$6.022\ 1415(10)\times10^{23}$ mol^{-1}	170
Boltzmann constant	k	$1.380\ 6505(24)\times10^{-23}$ J K^{-1}	1800
		$= 8.617\ 343(15)\times10^{-5}$ eV K^{-1}	1800
molar volume, ideal gas at STP	$N_A k$(273.15 K)/(101 325 Pa)	$22.413\ 996(39)\times10^{-3}$ m^3 mol^{-1}	1700
Wien displacement law constant	$b = \lambda_{\rm max}T$	$2.897\ 7685(51)\times10^{-3}$ m K	1700
Stefan-Boltzmann constant	$\sigma = \pi^2k^4/60\hbar^3c^2$	$5.670\ 400(40)\times10^{-8}$ W m^{-2} K^{-4}	7000
Fermi coupling constant**	$G_F/(\hbar c)^3$	$1.166\ 37(1)\times10^{-5}$ GeV^{-2}	9000
weak-mixing angle	$\sin^2\widehat{\theta}(M_Z)$ ($\overline{\rm MS}$)	0.23122(15)††	6.5×10^5
$W^\pm$ boson mass	m_W	80.403(29) GeV/c^2	3.6×10^5
Z^0 boson mass	m_Z	91.1876(21) GeV/c^2	2.3×10^4
strong coupling constant	$\alpha_s(m_Z)$	0.1176(20)	1.7×10^7

$\pi = 3.141\ 592\ 653\ 589\ 793\ 238$ e $= 2.718\ 281\ 828\ 459\ 045\ 235$ $\gamma = 0.577\ 215\ 664\ 901\ 532\ 861$

1 in $\equiv 0.0254$ m | 1 G $\equiv 10^{-4}$ T | 1 eV $= 1.602\ 176\ 53(14)\times10^{-19}$ J | kT at 300 K $= [38.681\ 684(68)]^{-1}$ eV

1 Å $\equiv 0.1$ nm | 1 dyne $\equiv 10^{-5}$ N | 1 eV/$c^2 = 1.782\ 661\ 81(15)\times10^{-36}$ kg | 0 °C $\equiv 273.15$ K

1 barn $\equiv 10^{-28}$ m^2 | 1 erg $\equiv 10^{-7}$ J | $2.997\ 924\ 58\times10^9$ esu $= 1$ C | 1 atmosphere $\equiv 760$ Torr $\equiv 101\ 325$ Pa

* The meter is the length of the path traveled by light in vacuum during a time interval of 1/299 792 458 of a second.
† At $Q^2 = 0$. At $Q^2 \approx m_W^2$ the value is $\sim 1/128$.
‡ Absolute lab measurements of G_N have been made only on scales of about 1 cm to 1 m.
** See the discussion in Sec. 10, "Electroweak model and constraints on new physics."
†† The corresponding $\sin^2\theta$ for the effective angle is 0.23152(14).

2. ASTROPHYSICAL CONSTANTS AND PARAMETERS

Table 2.1. Revised May 2006 by M.A. Dobbs (McGill U), D.E. Groom (LBNL), and D. Scott (UBC). The figures in parentheses after some values give the one-standard deviation uncertainties in the last digit(s). Physical constants are from Ref. 1. While every effort has been made to obtain the most accurate current values of the listed quantities, the table does not represent a critical review or adjustment of the constants, and is not intended as a primary reference. The values and uncertainties for the cosmological parameters depend on the exact data sets, priors, and basis parameters used in the fit. Many of the parameters reported in this table are derived parameters or have non-Gaussian likelihoods. Their error bars may be highly correlated with other parameters and care must be taken when extrapolating to higher significance levels. In most cases we report the best fit of a spatially-flat ΛCDM cosmology with a power-law initial spectrum to WMAP3 data alone [2]. For more information see Ref. 3 and the original papers.

Quantity	Symbol, equation	Value	Reference, footnote
speed of light	c	299 792 458 m s^{-1}	defined[4]
Newtonian gravitational constant	G_N	$6.6742(10) \times 10^{-11}$ $m^3\,kg^{-1}\,s^{-2}$	[1, 5]
astronomical unit (mean Earth-Sun distance)	AU	149 597 870 660(20) m	[6, 7]
tropical year (equinox to equinox) (2005.0)	yr	31 556 925.2 s	[6]
sidereal year (fixed star to fixed star) (2005.0)		31 558 149.8 s	[6]
mean sidereal day (2005.0)		$23^h\,56^m\,04\overset{s}{.}090\,53$	[6]
Jansky	Jy	10^{-26} W $m^{-2}\,Hz^{-1}$	
Planck mass	$\sqrt{\hbar c/G_N}$	$1.22090(9) \times 10^{19}$ GeV/c^2 $= 2.17645(16) \times 10^{-8}$ kg	[1]
Planck length	$\sqrt{\hbar G_N/c^3}$	$1.61624(12) \times 10^{-35}$ m	[1]
Hubble length	c/H_0	$\sim 1.2 \times 10^{26}$ m	[8]
parsec (1 AU/1 arc sec)	pc	$3.085\,677\,580\,7(4) \times 10^{16}$ m $= 3.262\ldots$ ly	[9]
light year (deprecated unit)	ly	$0.306\,6\ldots$ pc $= 0.946\,1\ldots \times 10^{16}$ m	
Schwarzschild radius of the Sun	$2G_N M_\odot/c^2$	2.953 250 08 km	[10]
Solar mass	$M_\odot$	$1.988\,44(30) \times 10^{30}$ kg	[11]
Solar equatorial radius	$R_\odot$	6.961×10^8 m	[6]
Solar luminosity	$L_\odot$	$(3.846 \pm 0.008) \times 10^{26}$ W	[12]
Schwarzschild radius of the Earth	$2G_N M_\oplus/c^2$	8.870 056 22 mm	[13]
Earth mass	$M_\oplus$	$5.972\,3(9) \times 10^{24}$ kg	[14]
Earth mean equatorial radius	$R_\oplus$	$6.378\,140 \times 10^6$ m	[6]
luminosity conversion	L	$3.02 \times 10^{28} \times 10^{-0.4\ M_{\rm bol}}$ W ($M_{\rm bol}$ = absolute bolometric magnitude = bolometric magnitude at 10 pc)	[15]
flux conversion	$\mathscr{F}$	$2.52 \times 10^{-8} \times 10^{-0.4\ m_{\rm bol}}$ W m^{-2} ($m_{\rm bol}$ = apparent bolometric magnitude)	from above
Solar velocity around center of Galaxy	$\Theta_\circ$	220(20) km s^{-1}	[16]
Solar distance from Galactic center	$R_\circ$	8.0(5) kpc	[17]
local disk density	$\rho_{\rm disk}$	3–12 $\times 10^{-24}$ g $cm^{-3} \approx$ 2–7 GeV/c^2 cm^{-3}	[18]
local halo density	$\rho_{\rm halo}$	2–13 $\times 10^{-25}$ g $cm^{-3} \approx$ 0.1–0.7 GeV/$c^2 cm^{-3}$	[19]
present day CBR temperature	T_0	2.725 ± 0.001 K	[20]
present day CBR dipole amplitude		3.346 ± 0.017 mK	[21]
Solar velocity with respect to CBR		369 ± 2 km/s towards $(\ell, b) = (263.86^\circ \pm 0.04^\circ, 48.24^\circ \pm 0.10^\circ)$	[21,22]
local group velocity with respect to CBR	$v_{\rm LG}$	627 ± 22 km s^{-1} towards $(\ell, b) = (276^\circ \pm 3^\circ, 30^\circ \pm 3^\circ)$	[23]
entropy density/Boltzmann constant	s/k	2 889.2 $(T/2.725)^3$ cm^{-3}	[15]
number density of CMB photons	n_γ	(410.5 ± 0.5) cm^{-3}	[24]
present day Hubble expansion rate	H_0	$100\,h$ km s^{-1} Mpc^{-1} $= h \times (9.778\,13$ Gyr$)^{-1}$	[25]
present day normalized Hubble expansion rate‡	h	$0.73^{+0.04}_{-0.03}$	[2]
scale factor for cosmological constant	$c^2/3H_0^2$	$2.853 \times 10^{51}\,h^{-2}$ m^2	
critical density of the Universe	$\rho_c = 3H_0^2/8\pi G_N$	$2.775\,366\,27 \times 10^{11}\,h^2$ $M_\odot Mpc^{-3}$ $= 1.878\,37(28) \times 10^{-29}\,h^2$ g cm^{-3} $= 1.053\,69(16) \times 10^{-5}\,h^2$ (GeV/c^2) cm^{-3}	derived
pressureless matter density of the Universe‡	$\Omega_{\rm m} = \rho_{\rm m}/\rho_c$	$0.127^{+0.007}_{-0.009}\,h^{-2} \Rightarrow 0.24^{+0.03}_{-0.04}$	[2]
baryon density of the Universe‡	$\Omega_{\rm b} = \rho_{\rm b}/\rho_c$	$0.0223^{+0.0007}_{-0.0009}\,h^{-2} \Rightarrow 0.042^{+0.003}_{-0.005}$	[2]
dark matter density of the Universe‡	$\Omega_{\rm dm} = \Omega_{\rm m} - \Omega_{\rm b}$	$0.105^{+0.007}_{-0.010}\,h^{-2} \Rightarrow 0.20^{+0.02}_{-0.04}$	
radiation density of the Universe‡	$\Omega_\gamma = \rho_\gamma/\rho_c$	$(2.471 \pm 0.004) \times 10^{-5}\,h^{-2} \Rightarrow (4.6 \pm 0.5) \times 10^{-5}$	[26]
neutrino density of the Universe‡	Ω_ν	$< 0.007\,h^{-2} \Rightarrow < 0.014$ (95% CL)	[27]
dark energy density‡	Ω_Λ	$0.76^{+0.04}_{-0.06}$	[28]

Quantity	Symbol, equation	Value	Reference, footnote
total energy density‡	$\Omega_{\rm tot} = \Omega_{\rm m} + \ldots + \Omega_\Lambda$	$1.003^{+0.013}_{-0.017}$	[2]
baryon-to-photon ratio‡	$\eta = n_{\rm b}/n_\gamma$	$4.7 \times 10^{-10} < \eta < 6.5 \times 10^{-10}$ (95% CL)	[29]
number density of baryons‡	$n_{\rm b}$	$(1.9 \times 10^{-7} < n_{\rm b} < 2.7 \times 10^{-7})\,{\rm cm}^{-3}$ (95% CL)	from η
dark energy equation of state parameter‡	w	$-0.97^{+0.07}_{-0.09}$	[2, 30]
fluctuation amplitude at $8h^{-1}$ Mpc scale‡	σ_8	$0.74^{+0.05}_{-0.06}$	[2]
scalar spectral index from power-law fit to data‡	n_s	$0.951^{+0.015}_{-0.019}$	[2]
running spectral index slope at $k_0 = 0.05$ Mpc^{-1} ‡	$dn_s/d\ln k$	$-0.055^{+0.029}_{-0.035}$	[2,31]
tensor-to-scalar field perturbations ratio at $k_0 = 0.002$ Mpc^{-1} ‡	$r = T/S$	< 0.55 at 95% C.L.	[2]
reionization optical depth‡	τ	0.09 ± 0.03	[2]
age of the Universe‡	t_0	$13.7^{+0.1}_{-0.2}$ Gyr	[2]

‡ See caption for caveats.

References:

1. P.J. Mohr and B.N. Taylor, CODATA 2002; `physics.nist.gov/cuu/Constants`.
2. D.N. Spergel *et al.*, "Wilkinson Microwave Anisotropy Probe (WMAP) Three Year Results: Implications for Cosmology," `astro-ph/0603449`.
3. O. Lahav, A.R. Liddle, "The Cosmological Parameters," this *Review*.
4. B.W. Petley, Nature **303**, 373 (1983).
5. In the context of the scale dependence of field theoretic quantities, it should be remarked that absolute lab measurements of G_N have been performed on scales of 0.01–1.0 m.
6. *The Astronomical Almanac for the year 2005*, U.S. Government Printing Office, Washington, and Her Majesty's Stationary Office, London (2003).
7. JPL Planetary Ephemerides, E. Myles Standish, Jr., private communication (1989).
8. Derived from H_0 [2].
9. 1 AU divided by $\pi/648\,000$; quoted error is from the JPL Planetary Ephemerides value of the AU [7].
10. Product of $2/c^2$ and the heliocentric gravitational constant [6]. The given 9-place accuracy seems consistent with uncertainties in defining Earth's orbital parameters.
11. Obtained from the heliocentric gravitational constant [6] and G_N [1]. The error is the 150 ppm standard deviation of G_N.
12. 1996 mean total solar irradiance (TSI) = 1367.5 ± 2.7 [32]; the solar luminosity is $4\pi \times (1\ {\rm AU})^2$ times this quantity. This value increased by 0.036% between the minima of solar cycles 21 and 22. It was modulated with an amplitude of 0.039% during solar cycle 21 [33].

    Sackmann *et al.* [34] use TSI = 1370 ± 2 W m^{-2}, but conclude that the solar luminosity ($L_\odot = 3.853 \times 10^{26}$ J s^{-1}) has an uncertainty of 1.5%. Their value comes from three 1977–83 papers, and they comment that the error is based on scatter among the reported values, which is substantially in excess of that expected from the individual quoted errors.

    The conclusion of the 1971 review by Thekaekara and Drummond [35] ($1353 \pm 1\%$ W m^{-2}) is often quoted [36]. The conversion to luminosity is not given in the Thekaekara and Drummond paper, and we cannot exactly reproduce the solar luminosity given in Ref. 36.

    Finally, a value based on the 1954 spectral curve due to Johnson [37] ($1395 \pm 1\%$ W m^{-2}, or $L_\odot = 3.92 \times 10^{26}$ J s^{-1}) has been used widely, and may be the basis for the higher value of the solar luminosity and the corresponding lower value of the solar absolute bolometric magnitude (4.72) still common in the literature [15].
13. Product of $2/c^2$, the heliocentric gravitational constant from Ref. 6, and the Earth/Sun mass ratio, also from Ref. 6. The given 9-place accuracy appears to be consistent with uncertainties in actually defining the earth's orbital parameters.
14. Obtained from the geocentric gravitational constant [6] and G_N [1]. The error is the 150 ppm standard deviation of G_N.
15. E.W. Kolb and M.S. Turner, *The Early Universe*, Addison-Wesley (1990).
16. F.J. Kerr and D. Lynden-Bell, Mon. Not. R. Astr. Soc. **221**, 1023–1038 (1985). "On the basis of this review these [$R_\circ = 8.5 \pm 1.1$ kpc and $\Theta_\circ = 220 \pm 20$ km s^{-1}] were adopted by resolution of IAU Commission 33 on 1985 November 21 at Delhi".
17. M.J. Reid, Annu. Rev. Astron. Astrophys. **31**, 345–372 (1993). Note that $\Theta_\circ$ from the 1985 IAU Commission 33 recommendations is adopted in this review, although the new value for $R_\circ$ is smaller.
18. G. Gilmore, R.F.G. Wyse, and K. Kuijken, Ann. Rev. Astron. Astrophys. **27**, 555 (1989).
19. E.I. Gates, G. Gyuk, and M.S. Turner (Astrophys. J. **449**, L133 (1995)) find the local halo density to be $9.2^{+3.8}_{-3.1} \times 10^{-25}$ g cm^{-3}, but also comment that previously published estimates are in the range $1\text{–}10 \times 10^{-25}$ g cm^{-3}.
    The value 0.3 GeV/c^2 has been taken as "standard" in several papers setting limits on WIMP mass limits, *e.g.* in M. Mori *et al.*, Phys. Lett. **B289**, 463 (1992).
20. J. Mather *et al.*, Astrophys. J. **512**, 511 (1999). This paper gives $T_0 = 2.725 \pm 0.002$K at 95%CL. We take 0.001 as the one-standard deviation uncertainty.
21. C.L. Bennett *et al.*, Astrophys. J. Supp. **148**, 1 (2003).
22. G. Hinshaw, "Three-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Temperature Analysis," `astro-ph/0603451`.
23. D. Scott and G.F. Smoot, "Cosmic Microwave Background," this *Review*.
24. $n_\gamma = \dfrac{2\zeta(3)}{\pi^2}\left(\dfrac{k_B T}{\hbar c}\right)^3$, using T_0 from Ref. 20.
25. Conversion using length of tropical year.
26. $\rho_\gamma = \dfrac{\pi^2}{15}\dfrac{(k_B T)^4}{(\hbar c)^3}$, using T_0 from Ref. 20.
27. Based on $\Omega_\nu h^2 = \sum m_{\nu_i}/93$ eV, with $\sum m_{\nu_i} = 0.7$ eV from CMB + LSS + SN data set, Table 10 in Ref. 2.
28. WMAP + $h = 0.72 \pm 0.08$, Table 11 in Ref. 2. Uses different h than tabulated here.
29. B.D. Fields, S. Sarkar, "Big-Bang Nucleosynthesis," this *Review*.
30. WMAP[2] + Supernova Legacy Survey in a flat Universe.
31. From WMAP 3-year data [2] alone, assuming no tensors.
32. R.C. Willson, Science **277**, 1963 (1997); the 0.2% error estimate is from R.C. Willson, private correspondence (1998).
33. R.C. Willson and H.S. Hudson, Nature **332**, 810 (1988).
34. I.-J. Sackmann, A.I. Boothroyd, and K.E. Kraemer, Astrophys. J. **418**, 457 (1993).
35. M.P. Thekaekara and A.J. Drummond, Nature Phys. Sci. **229**, 6 (1971).
36. K.R. Lang, *Astrophysical Formulae*, Springer-Verlag (1974); K.R. Lang, *Astrophysical Data: Planets and Stars*, Springer-Verlag (1992).
37. F.S. Johnson, J. Meterol. **11**, 431 (1954).

3. INTERNATIONAL SYSTEM OF UNITS (SI)

See "The International System of Units (SI)," NIST Special Publication **330**, B.N. Taylor, ed. (USGPO, Washington, DC, 1991); and "Guide for the Use of the International System of Units (SI)," NIST Special Publication **811**, 1995 edition, B.N. Taylor (USGPO, Washington, DC, 1995).

SI prefixes

10^{24}	yotta	(Y)
10^{21}	zetta	(Z)
10^{18}	exa	(E)
10^{15}	peta	(P)
10^{12}	tera	(T)
10^{9}	giga	(G)
10^{6}	mega	(M)
10^{3}	kilo	(k)
10^{2}	hecto	(h)
10	deca	(da)
10^{-1}	deci	(d)
10^{-2}	centi	(c)
10^{-3}	milli	(m)
10^{-6}	micro	(μ)
10^{-9}	nano	(n)
10^{-12}	pico	(p)
10^{-15}	femto	(f)
10^{-18}	atto	(a)
10^{-21}	zepto	(z)
10^{-24}	yocto	(y)

Physical quantity	**Name of unit**	**Symbol**
Base units		
length	meter	m
mass	kilogram	kg
time	second	s
electric current	ampere	A
thermodynamic temperature	kelvin	K
amount of substance	mole	mol
luminous intensity	candela	cd
Derived units with special names		
plane angle	radian	rad
solid angle	steradian	sr
frequency	hertz	Hz
energy	joule	J
force	newton	N
pressure	pascal	Pa
power	watt	W
electric charge	coulomb	C
electric potential	volt	V
electric resistance	ohm	Ω
electric conductance	siemens	S
electric capacitance	farad	F
magnetic flux	weber	Wb
inductance	henry	H
magnetic flux density	tesla	T
luminous flux	lumen	lm
illuminance	lux	lx
celsius temperature	degree celsius	°C
activity (of a radioactive source)*	becquerel	Bq
absorbed dose (of ionizing radiation)*	gray	Gy
dose equivalent*	sievert	Sv

*See our section 29, on "Radioactivity and radiation protection," p. 1.

Table 4.1. Revised 2005 by C.G. Wohl (LBNL) and D.E. Groom (LBNL). Adapted from the Commission on Atomic Weights and Isotopic Abundances, "Atomic Weights of the Elements 1999," Pure and Applied Chemistry **73**, 667 (2001), and G. Audi, A.H. Wapstra, and C. Thibault, Nucl. Phys. **A729**, 337 (2003). The atomic number (top left) is the number of protons in the nucleus. The atomic mass (bottom) is weighted by isotopic abundances in the Earth's surface. Atomic masses are relative to the mass of ^{12}C, defined to be exactly 12 unified atomic mass units (u) (approx. g/mole). Relative isotopic abundances often vary considerably, both in natural and commercial samples; this is reflected in the number of significant figures given. A number in parentheses is the atomic mass of the longest-lived known isotope of that element—no stable isotope exists. The exceptions are Th, Pa, and U, which do have characteristic terrestrial compositions. As of early 2006 element 112 has not been assigned a name, and there are no confirmed elements with Z > 112.

PERIODIC TABLE OF THE ELEMENTS

1 IA												13 IIIA	14 IVA	15 VA	16 VIA	17 VIIA	18 VIIIA
1 H Hydrogen 1.00794	2 IIA																2 He Helium 4.002602
3 Li Lithium 6.941	4 Be Beryllium 9.012182											5 B Boron 10.811	6 C Carbon 12.0107	7 N Nitrogen 14.0067	8 O Oxygen 15.9994	9 F Fluorine 18.9984032	10 Ne Neon 20.1797
11 Na Sodium 22.989770	12 Mg Magnesium 24.3050	3 IIIB	4 IVB	5 VB	6 VIB	7 VIIB	8	9 VIII	10	11 IB	12 IIB	13 Al Aluminum 26.981538	14 Si Silicon 28.0855	15 P Phosph. 30.973761	16 S Sulfur 32.065	17 Cl Chlorine 35.453	18 Ar Argon 39.948
19 K Potassium 39.0983	20 Ca Calcium 40.078	21 Sc Scandium 44.955910	22 Ti Titanium 47.867	23 V Vanadium 50.9415	24 Cr Chromium 51.9961	25 Mn Manganese 54.938049	26 Fe Iron 55.845	27 Co Cobalt 58.933200	28 Ni Nickel 58.6934	29 Cu Copper 63.546	30 Zn Zinc 65.39	31 Ga Gallium 69.723	32 Ge German. 72.64	33 As Arsenic 74.92160	34 Se Selenium 78.96	35 Br Bromine 79.904	36 Kr Krypton 83.80
37 Rb Rubidium 85.4678	38 Sr Strontium 87.62	39 Y Yttrium 88.90585	40 Zr Zirconium 91.224	41 Nb Niobium 92.90638	42 Mo Molybd. 95.94	43 Tc Technet. (97.907216)	44 Ru Ruthen. 101.07	45 Rh Rhodium 102.90550	46 Pd Palladium 106.42	47 Ag Silver 107.8682	48 Cd Cadmium 112.411	49 In Indium 114.818	50 Sn Tin 118.710	51 Sb Antimony 121.760	52 Te Tellurium 127.60	53 I Iodine 126.90447	54 Xe Xenon 131.293
55 Cs Cesium 132.90545	56 Ba Barium 137.327	57–71 Lantha-nides	72 Hf Hafnium 178.49	73 Ta Tantalum 180.9479	74 W Tungsten 183.84	75 Re Rhenium 186.207	76 Os Osmium 190.23	77 Ir Iridium 192.217	78 Pt Platinum 195.078	79 Au Gold 196.96655	80 Hg Mercury 200.59	81 Tl Thallium 204.3833	82 Pb Lead 207.2	83 Bi Bismuth 208.98038	84 Po Polonium (208.982430)	85 At Astatine (209.987148)	86 Rn Radon (222.017578)
87 Fr Francium (223.019736)	88 Ra Radium (226.025410)	89–103 Actinides	104 Rf Rutherford. (261.10877)	105 Db Dubnium (262.1141)	106 Sg Seaborg. (263.1221)	107 Bh Bohrium (262.1246)	108 Hs Hassium (277.1498)	109 Mt Meitner. (268.1387)	110 Ds Darmstadt. (271.1461)	111 Rg Roentgen. (272.1536)	112 (277.1639)						

Lanthanide series	57 La Lanthan. 138.9055	58 Ce Cerium 140.116	59 Pr Praseodym. 140.90765	60 Nd Neodym. 144.24	61 Pm Prometh. (144.912749)	62 Sm Samarium 150.36	63 Eu Europium 151.964	64 Gd Gadolin. 157.25	65 Tb Terbium 158.92534	66 Dy Dyspros. 162.50	67 Ho Holmium 164.93032	68 Er Erbium 167.259	69 Tm Thulium 168.93421	70 Yb Ytterbium 173.04	71 Lu Lutetium 174.967
Actinide series	89 Ac Actinium (227.027752)	90 Th Thorium 232.038055	91 Pa Protactin. 231.035884	92 U Uranium 238.02891	93 Np Neptunium (237.048173)	94 Pu Plutonium (244.064204)	95 Am Americ. (243.061381)	96 Cm Curium (247.070354)	97 Bk Berkelium (247.070307)	98 Cf Californ. (251.079587)	99 Es Einstein. (252.08298)	100 Fm Fermium (257.085105)	101 Md Mendelev. (258.098431)	102 No Nobelium (259.1010)	103 Lr Lawrenc. (262.1096)

5. ELECTRONIC STRUCTURE OF THE ELEMENTS

Table 5.1. Reviewed 2005 by C.G. Wohl (LBNL). The electronic configurations and the ionization energies are from the NIST database, "Ground Levels and Ionization Energies for the Neutral Atoms," W.C. Martin, A. Musgrove, S. Kotochigova, and J.E. Sansonetti (2003), `http://physics.nist.gov` (select "Physical Reference Data"). The electron configuration for, say, iron indicates an argon electronic core (see argon) plus six $3d$ electrons and two $4s$ electrons. The ionization energy is the least energy necessary to remove to infinity one electron from an atom of the element.

	Element		Electron configuration ($3d^5$ = five $3d$ electrons, *etc.*)		Ground state $^{2S+1}L_J$	Ionization energy (eV)
1	H	Hydrogen	$1s$		$^2S_{1/2}$	13.5984
2	He	Helium	$1s^2$		$^1S_0$	24.5874
3	Li	Lithium	(He) $2s$		$^2S_{1/2}$	5.3917
4	Be	Beryllium	(He) $2s^2$		$^1S_0$	9.3227
5	B	Boron	(He) $2s^2$ $2p$		$^2P_{1/2}$	8.2980
6	C	Carbon	(He) $2s^2$ $2p^2$		$^3P_0$	11.2603
7	N	Nitrogen	(He) $2s^2$ $2p^3$		$^4S_{3/2}$	14.5341
8	O	Oxygen	(He) $2s^2$ $2p^4$		$^3P_2$	13.6181
9	F	Fluorine	(He) $2s^2$ $2p^5$		$^2P_{3/2}$	17.4228
10	Ne	Neon	(He) $2s^2$ $2p^6$		$^1S_0$	21.5645
11	Na	Sodium	(Ne) $3s$		$^2S_{1/2}$	5.1391
12	Mg	Magnesium	(Ne) $3s^2$		$^1S_0$	7.6462
13	Al	Aluminum	(Ne) $3s^2$ $3p$		$^2P_{1/2}$	5.9858
14	Si	Silicon	(Ne) $3s^2$ $3p^2$		$^3P_0$	8.1517
15	P	Phosphorus	(Ne) $3s^2$ $3p^3$		$^4S_{3/2}$	10.4867
16	S	Sulfur	(Ne) $3s^2$ $3p^4$		$^3P_2$	10.3600
17	Cl	Chlorine	(Ne) $3s^2$ $3p^5$		$^2P_{3/2}$	12.9676
18	Ar	Argon	(Ne) $3s^2$ $3p^6$		$^1S_0$	15.7596
19	K	Potassium	(Ar) $4s$		$^2S_{1/2}$	4.3407
20	Ca	Calcium	(Ar) $4s^2$		$^1S_0$	6.1132
21	Sc	Scandium	(Ar) $3d$ $4s^2$	Transition elements	$^2D_{3/2}$	6.5615
22	Ti	Titanium	(Ar) $3d^2$ $4s^2$		$^3F_2$	6.8281
23	V	Vanadium	(Ar) $3d^3$ $4s^2$		$^4F_{3/2}$	6.7462
24	Cr	Chromium	(Ar) $3d^5$ $4s$		$^7S_3$	6.7665
25	Mn	Manganese	(Ar) $3d^5$ $4s^2$		$^6S_{5/2}$	7.4340
26	Fe	Iron	(Ar) $3d^6$ $4s^2$		$^5D_4$	7.9024
27	Co	Cobalt	(Ar) $3d^7$ $4s^2$		$^4F_{9/2}$	7.8810
28	Ni	Nickel	(Ar) $3d^8$ $4s^2$		$^3F_4$	7.6398
29	Cu	Copper	(Ar) $3d^{10}$ $4s$		$^2S_{1/2}$	7.7264
30	Zn	Zinc	(Ar) $3d^{10}$ $4s^2$		$^1S_0$	9.3942
31	Ga	Gallium	(Ar) $3d^{10}$ $4s^2$ $4p$		$^2P_{1/2}$	5.9993
32	Ge	Germanium	(Ar) $3d^{10}$ $4s^2$ $4p^2$		$^3P_0$	7.8994
33	As	Arsenic	(Ar) $3d^{10}$ $4s^2$ $4p^3$		$^4S_{3/2}$	9.7886
34	Se	Selenium	(Ar) $3d^{10}$ $4s^2$ $4p^4$		$^3P_2$	9.7524
35	Br	Bromine	(Ar) $3d^{10}$ $4s^2$ $4p^5$		$^2P_{3/2}$	11.8138
36	Kr	Krypton	(Ar) $3d^{10}$ $4s^2$ $4p^6$		$^1S_0$	13.9996
37	Rb	Rubidium	(Kr) $5s$		$^2S_{1/2}$	4.1771
38	Sr	Strontium	(Kr) $5s^2$		$^1S_0$	5.6949
39	Y	Yttrium	(Kr) $4d$ $5s^2$	Transition elements	$^2D_{3/2}$	6.2173
40	Zr	Zirconium	(Kr) $4d^2$ $5s^2$		$^3F_2$	6.6339
41	Nb	Niobium	(Kr) $4d^4$ $5s$		$^6D_{1/2}$	6.7589
42	Mo	Molybdenum	(Kr) $4d^5$ $5s$		$^7S_3$	7.0924
43	Tc	Technetium	(Kr) $4d^5$ $5s^2$		$^6S_{5/2}$	7.28
44	Ru	Ruthenium	(Kr) $4d^7$ $5s$		$^5F_5$	7.3605
45	Rh	Rhodium	(Kr) $4d^8$ $5s$		$^4F_{9/2}$	7.4589
46	Pd	Palladium	(Kr) $4d^{10}$		$^1S_0$	8.3369
47	Ag	Silver	(Kr) $4d^{10}$ $5s$		$^2S_{1/2}$	7.5762
48	Cd	Cadmium	(Kr) $4d^{10}$ $5s^2$		$^1S_0$	8.9938

49	In	Indium	(Kr) $4d^{10}5s^2$	$5p$				$^2P_{1/2}$	5.7864
50	Sn	Tin	(Kr) $4d^{10}5s^2$	$5p^2$				$^3P_0$	7.3439
51	Sb	Antimony	(Kr) $4d^{10}5s^2$	$5p^3$				$^4S_{3/2}$	8.6084
52	Te	Tellurium	(Kr) $4d^{10}5s^2$	$5p^4$				$^3P_2$	9.0096
53	I	Iodine	(Kr) $4d^{10}5s^2$	$5p^5$				$^2P_{3/2}$	10.4513
54	Xe	Xenon	(Kr) $4d^{10}5s^2$	$5p^6$				$^1S_0$	12.1298
55	Cs	Cesium	(Xe)	$6s$				$^2S_{1/2}$	3.8939
56	Ba	Barium	(Xe)	$6s^2$				$^1S_0$	5.2117
57	La	Lanthanum	(Xe) $5d$	$6s^2$				$^2D_{3/2}$	5.5769
58	Ce	Cerium	(Xe) $4f$ $5d$	$6s^2$				$^1G_4$	5.5387
59	Pr	Praseodymium	(Xe) $4f^3$	$6s^2$		L		$^4I_{9/2}$	5.473
60	Nd	Neodymium	(Xe) $4f^4$	$6s^2$		a		$^5I_4$	5.5250
61	Pm	Promethium	(Xe) $4f^5$	$6s^2$		n		$^6H_{5/2}$	5.582
62	Sm	Samarium	(Xe) $4f^6$	$6s^2$		t		$^7F_0$	5.6437
63	Eu	Europium	(Xe) $4f^7$	$6s^2$		h		$^8S_{7/2}$	5.6704
64	Gd	Gadolinium	(Xe) $4f^7$ $5d$	$6s^2$		a		$^9D_2$	6.1498
65	Tb	Terbium	(Xe) $4f^9$	$6s^2$		n		$^6H_{15/2}$	5.8638
66	Dy	Dysprosium	(Xe) $4f^{10}$	$6s^2$		i		$^5I_8$	5.9389
67	Ho	Holmium	(Xe) $4f^{11}$	$6s^2$		d		$^4I_{15/2}$	6.0215
68	Er	Erbium	(Xe) $4f^{12}$	$6s^2$		e		$^3H_6$	6.1077
69	Tm	Thulium	(Xe) $4f^{13}$	$6s^2$		s		$^2F_{7/2}$	6.1843
70	Yb	Ytterbium	(Xe) $4f^{14}$	$6s^2$				$^1S_0$	6.2542
71	Lu	Lutetium	(Xe) $4f^{14}5d$	$6s^2$				$^2D_{3/2}$	5.4259
72	Hf	Hafnium	(Xe) $4f^{14}5d^2$	$6s^2$		T		$^3F_2$	6.8251
73	Ta	Tantalum	(Xe) $4f^{14}5d^3$	$6s^2$		r	e	$^4F_{3/2}$	7.5496
74	W	Tungsten	(Xe) $4f^{14}5d^4$	$6s^2$		a	l	$^5D_0$	7.8640
75	Re	Rhenium	(Xe) $4f^{14}5d^5$	$6s^2$		n	e	$^6S_{5/2}$	7.8335
76	Os	Osmium	(Xe) $4f^{14}5d^6$	$6s^2$		s	m	$^5D_4$	8.4382
77	Ir	Iridium	(Xe) $4f^{14}5d^7$	$6s^2$		i	e	$^4F_{9/2}$	8.9670
78	Pt	Platinum	(Xe) $4f^{14}5d^9$	$6s$		t	n	$^3D_3$	8.9588
79	Au	Gold	(Xe) $4f^{14}5d^{10}6s$			i	t	$2S_{1/2}$	9.2255
80	Hg	Mercury	(Xe) $4f^{14}5d^{10}6s^2$			o	s	$^1S_0$	10.4375
81	Tl	Thallium	(Xe) $4f^{14}5d^{10}6s^2$	$6p$		n		$^2P_{1/2}$	6.1082
82	Pb	Lead	(Xe) $4f^{14}5d^{10}6s^2$	$6p^2$				$^3P_0$	7.4167
83	Bi	Bismuth	(Xe) $4f^{14}5d^{10}6s^2$	$6p^3$				$^4S_{3/2}$	7.2855
84	Po	Polonium	(Xe) $4f^{14}5d^{10}6s^2$	$6p^4$				$^3P_2$	8.414
85	At	Astatine	(Xe) $4f^{14}5d^{10}6s^2$	$6p^5$				$^2P_{3/2}$	
86	Rn	Radon	(Xe) $4f^{14}5d^{10}6s^2$	$6p^6$				$^1S_0$	10.7485
87	Fr	Francium	(Rn)	$7s$				$^2S_{1/2}$	4.0727
88	Ra	Radium	(Rn)	$7s^2$				$^1S_0$	5.2784
89	Ac	Actinium	(Rn) $6d$	$7s^2$				$^2D_{3/2}$	5.17
90	Th	Thorium	(Rn) $6d^2$	$7s^2$				$^3F_2$	6.3067
91	Pa	Protactinium	(Rn)$5f^2$ $6d$	$7s^2$		A		$^4K_{11/2}$*	5.89
92	U	Uranium	(Rn)$5f^3$ $6d$	$7s^2$		c		$^5L_6$*	6.1941
93	Np	Neptunium	(Rn)$5f^4$ $6d$	$7s^2$		t		$^6L_{11/2}$*	6.2657
94	Pu	Plutonium	(Rn)$5f^6$	$7s^2$		i		$^7F_0$	6.0260
95	Am	Americium	(Rn)$5f^7$	$7s^2$		n		$^8S_{7/2}$	5.9738
96	Cm	Curium	(Rn)$5f^7$ $6d$	$7s^2$		i		$^9D_2$	5.9914
97	Bk	Berkelium	(Rn)$5f^9$	$7s^2$		d		$^6H_{15/2}$	6.1979
98	Cf	Californium	(Rn)$5f^{10}$	$7s^2$		e		$^5I_8$	6.2817
99	Es	Einsteinium	(Rn)$5f^{11}$	$7s^2$		s		$^4I_{15/2}$	6.42
100	Fm	Fermium	(Rn)$5f^{12}$	$7s^2$				$^3H_6$	6.50
101	Md	Mendelevium	(Rn)$5f^{13}$	$7s^2$				$^2F_{7/2}$	6.58
102	No	Nobelium	(Rn)$5f^{14}$	$7s^2$				$^1S_0$	6.65
103	Lr	Lawrencium	(Rn)$5f^{14}$	$7s^2$ $7p$?				$^2P_{1/2}$?	4.9?
104	Rf	Rutherfordium	(Rn)$5f^{14}6d^2$	$7s^2$?				$^3F_2$?	6.0?

* The usual LS coupling scheme does not apply for these three elements. See the introductory note to the NIST table from which this table is taken.

6. ATOMIC AND NUCLEAR PROPERTIES OF MATERIALS

Table 6.1. Revised May 2002 by D.E. Groom (LBNL). Gases are evaluated at 20°C and 1 atm (in parentheses) or at STP [square brackets]. Densities and refractive indices without parentheses or brackets are for solids or liquids, or are for cryogenic liquids at the indicated boiling point (BP) at 1 atm. Refractive indices are evaluated at the sodium D line. Data for compounds and mixtures are from Refs. 1 and 2. Futher materials and properties are given in Ref. 3 and at `http://pdg.lbl.gov/AtomicNuclearProperties`.

Material	Z	A	$\langle Z/A\rangle$	Nuclear [a] collision length λ_T {g/cm^2}	Nuclear [a] interaction length λ_I {g/cm^2}	$dE/dx\|_{\rm min}$ [b] $\left\{\frac{\rm MeV}{\rm g/cm^2}\right\}$	Radiation length [c] X_0 {g/cm^2}	{cm}	Density {g/cm^3} ({g/ℓ} for gas)	Liquid boiling point at 1 atm(K)	Refractive index n ($(n-1)\times10^6$ for gas)
H_2 gas	1	1.00794	0.99212	43.3	50.8	(4.103)	61.28 [d]	(731000)	(0.0838)[0.0899]		[139.2]
H_2 liquid	1	1.00794	0.99212	43.3	50.8	4.034	61.28 [d]	866	0.0708	20.39	1.112
D_2	1	2.0140	0.49652	45.7	54.7	(2.052)	122.4	724	0.169[0.179]	23.65	1.128 [138]
He	2	4.002602	0.49968	49.9	65.1	(1.937)	94.32	756	0.1249[0.1786]	4.224	1.024 [34.9]
Li	3	6.941	0.43221	54.6	73.4	1.639	82.76	155	0.534		—
Be	4	9.012182	0.44384	55.8	75.2	1.594	65.19	35.28	1.848		—
C	6	12.011	0.49954	60.2	86.3	1.745	42.70	18.8	2.265 [e]		—
N_2	7	14.00674	0.49976	61.4	87.8	(1.825)	37.99	47.1	0.8073[1.250]	77.36	1.205 [298]
O_2	8	15.9994	0.50002	63.2	91.0	(1.801)	34.24	30.0	1.141[1.428]	90.18	1.22 [296]
F_2	9	18.9984032	0.47372	65.5	95.3	(1.675)	32.93	21.85	1.507[1.696]	85.24	[195]
Ne	10	20.1797	0.49555	66.1	96.6	(1.724)	28.94	24.0	1.204[0.9005]	27.09	1.092 [67.1]
Al	13	26.981539	0.48181	70.6	106.4	1.615	24.01	8.9	2.70		—
Si	14	28.0855	0.49848	70.6	106.0	1.664	21.82	9.36	2.33		3.95
Ar	18	39.948	0.45059	76.4	117.2	(1.519)	19.55	14.0	1.396[1.782]	87.28	1.233 [283]
Ti	22	47.867	0.45948	79.9	124.9	1.476	16.17	3.56	4.54		—
Fe	26	55.845	0.46556	82.8	131.9	1.451	13.84	1.76	7.87		—
Cu	29	63.546	0.45636	85.6	134.9	1.403	12.86	1.43	8.96		—
Ge	32	72.61	0.44071	88.3	140.5	1.371	12.25	2.30	5.323		—
Sn	50	118.710	0.42120	100.2	163	1.264	8.82	1.21	7.31		—
Xe	54	131.29	0.41130	102.8	169	(1.255)	8.48	2.87	2.953[5.858]	165.1	[701]
W	74	183.84	0.40250	110.3	185	1.145	6.76	0.35	19.3		—
Pt	78	195.08	0.39984	113.3	189.7	1.129	6.54	0.305	21.45		—
Pb	82	207.2	0.39575	116.2	194	1.123	6.37	0.56	11.35		—
U	92	238.0289	0.38651	117.0	199	1.082	6.00	≈0.32	≈18.95		—
Air, (20°C, 1 atm.), [STP]			0.49919	62.0	90.0	(1.815)	36.66	[30420]	(1.205)[1.2931]	78.8	(273) [293]
H_2O			0.55509	60.1	83.6	1.991	36.08	36.1	1.00	373.15	1.33
CO_2 gas			0.49989	62.4	89.7	(1.819)	36.2	[18310]	[1.977]		[410]
CO_2 solid (dry ice)			0.49989	62.4	89.7	1.787	36.2	23.2	1.563	sublimes	—
Shielding concrete [f]			0.50274	67.4	99.9	1.711	26.7	10.7	2.5		—
SiO_2 (fused quartz)			0.49926	66.5	97.4	1.699	27.05	12.3	2.20 [g]		1.458
Dimethyl ether, $(CH_3)_2O$			0.54778	59.4	82.9	—	38.89	—	—	248.7	—
Methane, CH_4			0.62333	54.8	73.4	(2.417)	46.22	[64850]	0.4224[0.717]	111.7	[444]
Ethane, C_2H_6			0.59861	55.8	75.7	(2.304)	45.47	[34035]	0.509(1.356) [h]	184.5	(1.038) [h]
Propane, C_3H_8			0.58962	56.2	76.5	(2.262)	45.20	—	(1.879)	231.1	—
Isobutane, $(CH_3)_2CHCH_3$			0.58496	56.4	77.0	(2.239)	45.07	[16930]	[2.67]	261.42	[1900]
Octane, liquid, $CH_3(CH_2)_6CH_3$			0.57778	56.7	77.7	2.123	44.86	63.8	0.703	398.8	1.397
Paraffin wax, $CH_3(CH_2)_{n\approx23}CH_3$			0.57275	56.9	78.2	2.087	44.71	48.1	0.93		—
Nylon, type 6 [i]			0.54790	58.5	81.5	1.974	41.84	36.7	1.14		—
Polycarbonate (Lexan) [j]			0.52697	59.5	83.9	1.886	41.46	34.6	1.20		—
Polyethylene terephthlate (Mylar) [k]			0.52037	60.2	85.7	1.848	39.95	28.7	1.39		—
Polyethylene [l]			0.57034	57.0	78.4	2.076	44.64	≈47.9	0.92–0.95		—
Polyimide film (Kapton) [m]			0.51264	60.3	85.8	1.820	40.56	28.6	1.42		—
Lucite, Plexiglas [n]			0.53937	59.3	83.0	1.929	40.49	≈34.4	1.16–1.20		≈1.49
Polystyrene, scintillator [o]			0.53768	58.5	81.9	1.936	43.72	42.4	1.032		1.581
Polytetrafluoroethylene (Teflon) [p]			0.47992	64.2	93.0	1.671	34.84	15.8	2.20		—
Polyvinyltolulene, scintillator [q]			0.54155	58.3	81.5	1.956	43.83	42.5	1.032		—
Aluminum oxide (Al_2O_3)			0.49038	67.0	98.9	1.647	27.94	7.04	3.97		1.761
Barium fluoride (BaF_2)			0.42207	92.0	145	1.303	9.91	2.05	4.89		1.56
Bismuth germanate (BGO) [r]			0.42065	98.2	157	1.251	7.97	1.12	7.1		2.15
Cesium iodide (CsI)			0.41569	102	167	1.243	8.39	1.85	4.53		1.80
Lithium fluoride (LiF)			0.46262	62.2	88.2	1.614	39.25	14.91	2.632		1.392
Sodium fluoride (NaF)			0.47632	66.9	98.3	1.69	29.87	11.68	2.558		1.336
Sodium iodide (NaI)			0.42697	94.6	151	1.305	9.49	2.59	3.67		1.775
Silica Aerogel [s]			0.50093	66.3	96.9	1.740	27.25	136@ρ=0.2	0.04–0.6		1.0+0.21ρ
NEMA G10 plate [t]				62.6	90.2	1.87	33.0	19.4	1.7		—

Material	Dielectric constant ($\kappa = \epsilon/\epsilon_0$) () is ($\kappa$–1)×$10^6$ for gas	Young's modulus [10^6 psi]	Coeff. of thermal expansion [10^{-6}cm/cm-°C]	Specific heat [cal/g-°C]	Electrical resistivity [$\mu\Omega$cm(@°C)]	Thermal conductivity [cal/cm-°C-sec]
H_2	(253.9)	—	—	—	—	—
He	(64)	—	—	—	—	—
Li	—	—	56	0.86	8.55(0°)	0.17
Be	—	37	12.4	0.436	5.885(0°)	0.38
C	—	0.7	0.6–4.3	0.165	1375(0°)	0.057
N_2	(548.5)	—	—	—	—	—
O_2	(495)	—	—	—	—	—
Ne	(127)	—	—	—	—	—
Al	—	10	23.9	0.215	2.65(20°)	0.53
Si	11.9	16	2.8–7.3	0.162	—	0.20
Ar	(517)	—	—	—	—	—
Ti	—	16.8	8.5	0.126	50(0°)	—
Fe	—	28.5	11.7	0.11	9.71(20°)	0.18
Cu	—	16	16.5	0.092	1.67(20°)	0.94
Ge	16.0	—	5.75	0.073	—	0.14
Sn	—	6	20	0.052	11.5(20°)	0.16
Xe	—	—	—	—	—	—
W	—	50	4.4	0.032	5.5(20°)	0.48
Pt	—	21	8.9	0.032	9.83(0°)	0.17
Pb	—	2.6	29.3	0.038	20.65(20°)	0.083
U	—	—	36.1	0.028	29(20°)	0.064

1. R.M. Sternheimer, M.J. Berger, and S.M. Seltzer, Atomic Data and Nuclear Data Tables **30**, 261–271 (1984).
2. S.M. Seltzer and M.J. Berger, Int. J. Appl. Radiat. **33**, 1189–1218 (1982).
3. D.E. Groom, N.V. Mokhov, and S.I. Striganov, "Muon stopping-power and range tables," Atomic Data and Nuclear Data Tables **78**, 183–356 (2001).
4. S.M. Seltzer and M.J. Berger, Int. J. Appl. Radiat. **35**, 665 (1984) & `http://physics.nist.gov/PhysRefData/Star/Text/contents.html`.

a. σ_T, λ_T and λ_I are energy dependent. Values quoted apply to high energy range, where energy dependence is weak. Mean free path between collisions (λ_T) or inelastic interactions (λ_I), calculated from $\lambda^{-1} = N_A \sum w_j \sigma_j / A_j$, where N is Avogadro's number and w_j is the weight fraction of the jth element in the element, compound, or mixture. $\sigma_{\rm total}$ at 80–240 GeV for neutrons ($\approx \sigma$ for protons) from Murthy *et al.*, Nucl. Phys. **B92**, 269 (1975). This scales approximately as $A^{0.77}$. $\sigma_{\rm inelastic} = \sigma_{\rm total} - \sigma_{\rm elastic} - \sigma_{\rm quasielastic}$; for neutrons at 60–375 GeV from Roberts *et al.*, Nucl. Phys. **B159**, 56 (1979). For protons and other particles, see Carroll *et al.*, Phys. Lett. **80B**, 319 (1979); note that $\sigma_I(p) \approx \sigma_I(n)$. σ_I scales approximately as $A^{0.71}$.

b. For minimum-ionizing muons (results are very slightly different for other particles). Minimum dE/dx from Ref. 3, using density effect correction coefficients from Ref. 1. For electrons and positrons see Ref. 4. Ionization energy loss is discussed in Sec. 27.

c. From Y.S. Tsai, Rev. Mod. Phys. **46**, 815 (1974); X_0 data for all elements up to uranium are given. Corrections for molecular binding applied for H_2 and D_2. For atomic H, $X_0 = 63.05$ g/cm^2.

d. For molecular hydrogen (deuterium). For atomic H, $X_0 = 63.047$ g cm^{-2}.

e. For pure graphite; industrial graphite density may vary 2.1–2.3 g/cm^3.

f. Standard shielding blocks, typical composition O_2 52%, Si 32.5%, Ca 6%, Na 1.5%, Fe 2%, Al 4%, plus reinforcing iron bars. The attenuation length, $\ell = 115 \pm 5$ g/cm^2, is also valid for earth (typical $\rho = 2.15$), from CERN–LRL–RHEL Shielding exp., UCRL–17841 (1968).

g. For typical fused quartz. The specific gravity of crystalline quartz is 2.64.

h. Solid ethane density at −60°C; gaseous refractive index at 0°C, 546 mm pressure.

i. Nylon, Type 6, $(NH(CH_2)_5CO)_n$

j. Polycarbonate (Lexan), $(C_{16}H_{14}O_3)_n$

k. Polyethylene terephthlate, monomer $C_5H_4O_2$

l. Polyethylene, monomer $CH_2 = CH_2$

m. Polymide film (Kapton), $(C_{22}H_{10}N_2O_5)_n$

n. Polymethylmethacralate, monomer $CH_2 = C(CH_3)CO_2CH_3$

o. Polystyrene, monomer $C_6H_5CH{=}CH_2$

p. Teflon, monomer $CF_2 = CF_2$

q. Polyvinyltolulene, monomer 2-$CH_3C_6H_4CH{=}CH_2$

r. Bismuth germanate (BGO), $(Bi_2O_3)_2(GeO_2)_3$

s. 97% SiO_2 + 3% H_2O by weight; see A. R. Buzykaev *et al.*, Nucl. Instrum. Methods **A433**, 396 (1999). Aerogel in the density range 0.04–0.06 g/cm^3 has been used in Čerenkov counters, but aerogel with higher and lower densities has been produced. ρ = density in g/cm^3.

t. G10-plate, typically 60% SiO_2 and 40% epoxy.

7. ELECTROMAGNETIC RELATIONS

Revised September 2005 by H.G. Spieler (LBNL).

Quantity	Gaussian CGS	SI
Conversion factors: Charge: Potential: Magnetic field:	 $2.997\,924\,58 \times 10^9$ esu (1/299.792 458) statvolt (ergs/esu) 10^4 gauss $= 10^4$ dyne/esu	 $= 1$ C $= 1$ A s $= 1$ V $= 1$ J C^{-1} $= 1$ T $= 1$ N A^{-1}m^{-1}
	$\mathbf{F} = q\,(\mathbf{E} + \frac{\mathbf{v}}{c} \times \mathbf{B})$	$\mathbf{F} = q\,(\mathbf{E} + \mathbf{v} \times \mathbf{B})$
	$\boldsymbol{\nabla}\cdot\,\mathbf{D} = 4\pi\rho$ $\boldsymbol{\nabla} \times \mathbf{H} - \frac{1}{c}\frac{\partial \mathbf{D}}{\partial t} = \frac{4\pi}{c}\mathbf{J}$ $\boldsymbol{\nabla}\cdot\,\mathbf{B} = 0$ $\boldsymbol{\nabla} \times \mathbf{E} + \frac{1}{c}\frac{\partial \mathbf{B}}{\partial t} = 0$	$\boldsymbol{\nabla}\cdot\,\mathbf{D} = \rho$ $\boldsymbol{\nabla} \times \mathbf{H} - \frac{\partial \mathbf{D}}{\partial t} = \mathbf{J}$ $\boldsymbol{\nabla}\cdot\,\mathbf{B} = 0$ $\boldsymbol{\nabla} \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = 0$
Constitutive relations:	$\mathbf{D} = \mathbf{E} + 4\pi\mathbf{P}, \quad \mathbf{H} = \mathbf{B} - 4\pi\mathbf{M}$	$\mathbf{D} = \epsilon_0\mathbf{E} + \mathbf{P}, \quad \mathbf{H} = \mathbf{B}/\mu_0 - \mathbf{M}$
Linear media: 	$\mathbf{D} = \epsilon\mathbf{E}, \quad \mathbf{H} = \mathbf{B}/\mu$ 1 1	$\mathbf{D} = \epsilon\mathbf{E}, \quad \mathbf{H} = \mathbf{B}/\mu$ $\epsilon_0 = 8.854\,187\ldots \times 10^{-12}$ F m^{-1} $\mu_0 = 4\pi \times 10^{-7}$ N A^{-2}
	$\mathbf{E} = -\boldsymbol{\nabla}V - \frac{1}{c}\frac{\partial \mathbf{A}}{\partial t}$ $\mathbf{B} = \boldsymbol{\nabla} \times \mathbf{A}$	$\mathbf{E} = -\boldsymbol{\nabla}V - \frac{\partial \mathbf{A}}{\partial t}$ $\mathbf{B} = \boldsymbol{\nabla} \times \mathbf{A}$
	$V = \sum_{\text{charges}} \frac{q_i}{r_i} = \int \frac{\rho\,(\mathbf{r}')}{\lvert\mathbf{r} - \mathbf{r}'\rvert}\, d^3x'$ $\mathbf{A} = \frac{1}{c}\oint \frac{I\,d\boldsymbol{\ell}}{\lvert\mathbf{r} - \mathbf{r}'\rvert} = \frac{1}{c}\int \frac{\mathbf{J}(\mathbf{r}')}{\lvert\mathbf{r} - \mathbf{r}'\rvert}\, d^3x'$	$V = \frac{1}{4\pi\epsilon_0}\sum_{\text{charges}} \frac{q_i}{r_i} = \frac{1}{4\pi\epsilon_0}\int \frac{\rho\,(\mathbf{r}')}{\lvert\mathbf{r} - \mathbf{r}'\rvert}\, d^3x'$ $\mathbf{A} = \frac{\mu_0}{4\pi}\oint \frac{I\,d\boldsymbol{\ell}}{\lvert\mathbf{r} - \mathbf{r}'\rvert} = \frac{\mu_0}{4\pi}\int \frac{\mathbf{J}(\mathbf{r}')}{\lvert\mathbf{r} - \mathbf{r}'\rvert}\, d^3x'$
	$\mathbf{E}'_\parallel = \mathbf{E}_\parallel$ $\mathbf{E}'_\perp = \gamma(\mathbf{E}_\perp + \frac{1}{c}\mathbf{v} \times \mathbf{B})$ $\mathbf{B}'_\parallel = \mathbf{B}_\parallel$ $\mathbf{B}'_\perp = \gamma(\mathbf{B}_\perp - \frac{1}{c}\mathbf{v} \times \mathbf{E})$	$\mathbf{E}'_\parallel = \mathbf{E}_\parallel$ $\mathbf{E}'_\perp = \gamma(\mathbf{E}_\perp + \mathbf{v} \times \mathbf{B})$ $\mathbf{B}'_\parallel = \mathbf{B}_\parallel$ $\mathbf{B}'_\perp = \gamma(\mathbf{B}_\perp - \frac{1}{c^2}\mathbf{v} \times \mathbf{E})$
$\frac{1}{4\pi\epsilon_0} = c^2 \times 10^{-7}$ N A^{-2} $= 8.987\,55\ldots \times 10^9$ m F^{-1} ; $\frac{\mu_0}{4\pi} = 10^{-7}$ N A^{-2} ; $c = \frac{1}{\sqrt{\mu_0\epsilon_0}} = 2.997\,924\;\;58 \times 10^8$ m s^{-1}		

7.1. Impedances (SI units)

ρ = resistivity at room temperature in $10^{-8}\,\Omega$ m:

~ 1.7 for Cu	~ 5.5 for W
~ 2.4 for Au	~ 73 for SS 304
~ 2.8 for Al	~ 100 for Nichrome

(Al alloys may have double the Al value.)

For alternating currents, instantaneous current I, voltage V, angular frequency ω:

$$V = V_0\, e^{j\omega t} = ZI\ . \tag{7.1}$$

Impedance of self-inductance L: $Z = j\omega L$.

Impedance of capacitance C: $Z = 1/j\omega C$.

Impedance of free space: $Z = \sqrt{\mu_0/\epsilon_0} = 376.7\ \Omega$.

High-frequency surface impedance of a good conductor:

$$Z = \frac{(1+j)\,\rho}{\delta}\ , \quad \text{where } \delta = \text{skin depth}\ ; \tag{7.2}$$

$$\delta = \sqrt{\frac{\rho}{\pi\nu\mu}} \approx \frac{6.6\text{ cm}}{\sqrt{\nu\,(\text{Hz})}} \quad \text{for Cu}\ . \tag{7.3}$$

7.2. Capacitors, inductors, and transmission Lines

The capacitance between two parallel plates of area A spaced by the distance d and enclosing a medium with the dielectric constant ε is

$$C = K\varepsilon A/d\,, \tag{7.4}$$

where the correction factor K depends on the extent of the fringing field. If the dielectric fills the capacitor volume without extending beyond the electrodes. the correction factor $K \approx 0.8$ for capacitors of typical geometry.

The inductance at high frequencies of a straight wire whose length ℓ is much greater than the wire diameter d is

$$L \approx 2.0\left[\frac{\text{nH}}{\text{cm}}\right]\cdot\ell\left(\ln\left(\frac{4\ell}{d}\right)-1\right)\ . \tag{7.5}$$

For very short wires, representative of vias in a printed circuit board, the inductance is

$$L(\text{in nH}) \approx \ell/d\,. \tag{7.6}$$

A transmission line is a pair of conductors with inductance L and capacitance C. The characteristic impedance $Z = \sqrt{L/C}$ and the phase velocity $v_p = 1/\sqrt{LC} = 1/\sqrt{\mu\varepsilon}$, which decreases with the inverse square root of the dielectric constant of the medium. Typical coaxial and ribbon cables have a propagation delay of about 5 ns/cm.

The impedance of a coaxial cable with outer diameter D and inner diameter d is

$$Z = 60\,\Omega\cdot\frac{1}{\sqrt{\varepsilon_r}}\ln\frac{D}{d}\,, \tag{7.7}$$

where the relative dielectric constant $\varepsilon_r = \varepsilon/\varepsilon_0$. A pair of parallel wires of diameter d and spacing $a > 2.5\,d$ has the impedance

$$Z = 120\,\Omega\cdot\frac{1}{\sqrt{\varepsilon_r}}\ln\frac{2a}{d}\,. \tag{7.8}$$

This yields the impedance of a wire at a spacing h above a ground plane,

$$Z = 60\,\Omega\cdot\frac{1}{\sqrt{\varepsilon_r}}\ln\frac{4h}{d}\,. \tag{7.9}$$

A common configuration utilizes a thin rectangular conductor above a ground plane with an intermediate dielectric (microstrip). Detailed calculations for this and other transmission line configurations are given by Gunston.*

* M.A.R. Gunston. Microwave Transmission Line Data, Noble Publishing Corp., Atlanta (1997) ISBN 1-884932-57-6, TK6565.T73G85.

7.3. Synchrotron radiation (CGS units)

For a particle of charge e, velocity $v = \beta c$, and energy $E = \gamma mc^2$, traveling in a circular orbit of radius R, the classical energy loss per revolution δE is

$$\delta E = \frac{4\pi}{3}\,\frac{e^2}{R}\,\beta^3\,\gamma^4\ . \tag{7.10}$$

For high-energy electrons or positrons ($\beta \approx 1$), this becomes

$$\delta E\ (\text{in MeV}) \approx 0.0885\ [E(\text{in GeV})]^4/R(\text{in m})\ . \tag{7.11}$$

For $\gamma \gg 1$, the energy radiated per revolution into the photon energy interval $d(\hbar\omega)$ is

$$dI = \frac{8\pi}{9}\alpha\,\gamma\,F(\omega/\omega_c)\,d(\hbar\omega)\ , \tag{7.12}$$

where $\alpha = e^2/\hbar c$ is the fine-structure constant and

$$\omega_c = \frac{3\gamma^3 c}{2R} \tag{7.13}$$

is the critical frequency. The normalized function $F(y)$ is

$$F(y) = \frac{9}{8\pi}\sqrt{3}\,y\int_y^\infty K_{5/3}\,(x)\ dx\ , \tag{7.14}$$

where $K_{5/3}\,(x)$ is a modified Bessel function of the third kind. For electrons or positrons,

$$\hbar\omega_c\,(\text{in keV}) \approx 2.22\ [E(\text{in GeV})]^3/R(\text{in m})\ . \tag{7.15}$$

Fig. 7.1 shows $F(y)$ over the important range of y.

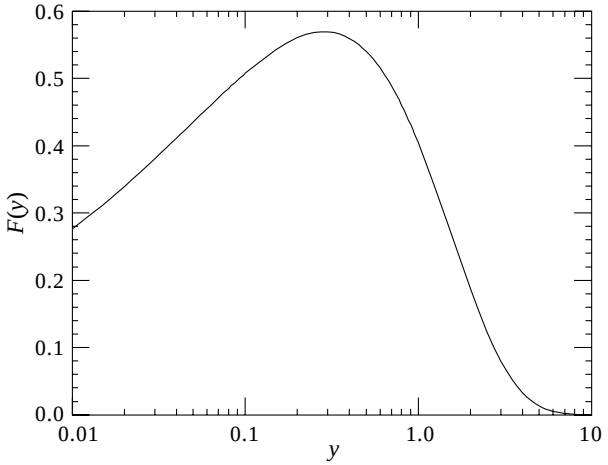


Figure 7.1: The normalized synchrotron radiation spectrum $F(y)$.

For $\gamma \gg 1$ and $\omega \ll \omega_c$,

$$\frac{dI}{d(\hbar\omega)} \approx 3.3\alpha\,(\omega R/c)^{1/3}\ , \tag{7.16}$$

whereas for

$$\gamma \gg 1 \text{ and } \omega \gtrsim 3\omega_c\ ,$$

$$\frac{dI}{d(\hbar\omega)} \approx \sqrt{\frac{3\pi}{2}}\,\alpha\,\gamma\left(\frac{\omega}{\omega_c}\right)^{1/2} e^{-\omega/\omega_c}\left[1+\frac{55}{72}\frac{\omega_c}{\omega}+\ldots\right]\ . \tag{7.17}$$

The radiation is confined to angles $\lesssim 1/\gamma$ relative to the instantaneous direction of motion. For $\gamma \gg 1$, where Eq. (7.12) applies, the mean number of photons emitted per revolution is

$$N_\gamma = \frac{5\pi}{\sqrt{3}}\alpha\gamma\ , \tag{7.18}$$

and the mean energy per photon is

$$\langle\hbar\omega\rangle = \frac{8}{15\sqrt{3}}\hbar\omega_c\ . \tag{7.19}$$

When $\langle\hbar\omega\rangle \gtrsim \mathrm{O}(E)$, quantum corrections are important.

See J.D. Jackson, *Classical Electrodynamics*, 3rd edition (John Wiley & Sons, New York, 1998) for more formulae and details. (Note that earlier editions had ω_c twice as large as Eq. (7.13).

8. NAMING SCHEME FOR HADRONS

Revised 2004 by M. Roos (University of Finland) and C.G. Wohl (LBNL).

8.1. Introduction

We introduced in the 1986 edition [1] a new naming scheme for the hadrons. Changes from older terminology affected mainly the heavier mesons made of the light (u, d, and s) quarks. Old and new names were listed alongside until 1994. Names also change from edition to edition because some characteristic like mass or spin changes. The Summary Tables give both the new and old names whenever a change occurred.

8.2. "Neutral-flavor" mesons ($S = C = B = T = 0$)

Table 8.1 shows the names for mesons having the strangeness and all heavy-flavor quantum numbers equal to zero. The scheme is designed for all ordinary non-exotic mesons, but it will work for many exotic types too, if needed.

Table 8.1: Symbols for mesons with the strangeness and all heavy-flavor quantum numbers equal to zero.

	$J^{PC} =$	$\begin{cases} 0^{-+} \\ 2^{-+} \\ \vdots \end{cases}$	1^{+-} 3^{+-} $\vdots$	1^{--} 2^{--} $\vdots$	0^{++} 1^{++} $\vdots$
$q\overline{q}$ content	$^{2S+1}L_J =$	$^1(L\,\text{even})_J$	$^1(L\,\text{odd})_J$	$^3(L\,\text{even})_J$	$^3(L\,\text{odd})_J$
$u\overline{d}, u\overline{u} - d\overline{d}, d\overline{u}$	$(I=1)$	π	b	ρ	a
$d\overline{d} + u\overline{u}$ and/or $s\overline{s}$	$(I=0)$	η, η'	h, h'	ω, ϕ	f, f'
$c\overline{c}$		η_c	h_c	$\psi^\dagger$	χ_c
$b\overline{b}$		η_b	h_b	Υ	χ_b
$t\overline{t}$		η_t	h_t	θ	χ_t

†The J/ψ remains the J/ψ.

First, we assign names to those states with quantum numbers compatible with being $q\overline{q}$ states. The rows of the Table give the possible $q\overline{q}$ content. The columns give the possible parity/charge-conjugation states,

$$PC = -+,\ +-,\ --, \text{ and } ++\ ;$$

these combinations correspond one-to-one with the angular-momentum state $^{2S+1}L_J$ of the $q\overline{q}$ system being

$$^1(L \text{ even})_J,\ ^1(L \text{ odd})_J,\ ^3(L \text{ even})_J, \text{ or } ^3(L \text{ odd})_J\ .$$

Here S, L, and J are the spin, orbital, and total angular momenta of the $q\overline{q}$ system. The quantum numbers are related by

$P = (-1)^{L+1}$, $C = (-1)^{L+S}$, and G parity $= (-1)^{L+S+I}$,
where of course the C quantum number is only relevant to neutral mesons.

The entries in the Table give the meson names. The spin J is added as a subscript except for pseudoscalar and vector mesons, and the mass is added in parentheses for mesons that decay strongly. However, for the lightest meson resonances, we omit the mass.

Measurements of the mass, quark content (where relevant), and quantum numbers I, J, P, and C (or G) of a meson thus fix its symbol. Conversely, these properties may be inferred unambiguously from the symbol.

If the main symbol cannot be assigned because the quantum numbers are unknown, X is used. Sometimes it is not known whether a meson is mainly the isospin-0 mix of $u\overline{u}$ and $d\overline{d}$ or is mainly $s\overline{s}$. A prime (or pair ω, ϕ) may be used to distinguish two such mixing states.

We follow custom and use spectroscopic names such as $\Upsilon(1S)$ as the primary name for most of those ψ, Υ, and χ states whose spectroscopic identity is known. We use the form $\Upsilon(9460)$ as an alternative, and as the primary name when the spectroscopic identity is not known.

Names are assigned for $t\overline{t}$ mesons, although the top quark is evidently so heavy that it is expected to decay too rapidly for bound states to form.

Gluonium states or other mesons that are not $q\overline{q}$ states are, if the quantum numbers are *not* exotic, to be named just as are the $q\overline{q}$ mesons. Such states will probably be difficult to distinguish from $q\overline{q}$ states and will likely mix with them, and we make no attempt to distinguish those "mostly gluonium" from those "mostly $q\overline{q}$."

An "exotic" meson with J^{PC} quantum numbers that a $q\overline{q}$ system cannot have, namely $J^{PC} = 0^{--}, 0^{+-}, 1^{-+}, 2^{+-}, 3^{-+}, \cdots$, would use the same symbol as does an ordinary meson with all the same quantum numbers as the exotic meson except for the C parity. But then the J subscript may still distinguish it; for example, an isospin-0 1^{-+} meson could be denoted ω_1.

8.3. Mesons with nonzero S, C, B, and/or T

Since the strangeness or a heavy flavor of these mesons is nonzero, none of them are eigenstates of charge conjugation, and in each of them one of the quarks is heavier than the other. The rules are:

1. The main symbol is an upper-case italic letter indicating the heavier quark as follows:

   $$s \to \overline{K} \qquad c \to D \qquad b \to \overline{B} \qquad t \to T\ .$$

   We use the convention that *the flavor and the charge of a quark have the same sign.* Thus the strangeness of the s quark is negative, the charm of the c quark is positive, and the bottom of the b quark is negative. In addition, I_3 of the u and d quarks are positive and negative, respectively. The effect of this convention is as follows: *Any flavor carried by a charged meson has the same sign as its charge.* Thus the K^+, D^+,

and B^+ have positive strangeness, charm, and bottom, respectively, and all have positive I_3. The D_s^+ has positive charm *and* strangeness. Furthermore, the Δ(flavor) $= \Delta Q$ rule, best known for the kaons, applies to every flavor.

2. If the lighter quark is not a u or a d quark, its identity is given by a subscript. The D_s^+ is an example.
3. If the spin-parity is in the "normal" series, $J^P = 0^+, 1^-, 2^+, \cdots$, a superscript "*" is added.
4. The spin is added as a subscript except for pseudoscalar or vector mesons.

8.4. Ordinary (3-quark) baryons

The symbols N, Δ, Λ, Σ, Ξ, and Ω used for more than 30 years for the baryons made of light quarks (u, d, and s quarks) tell the isospin and quark content, and the same information is conveyed by the symbols used for the baryons containing one or more heavy quarks (c and b quarks). The rules are:

1. Baryons with *three* u and/or d quarks are N's (isospin 1/2) or Δ's (isospin 3/2).
2. Baryons with *two* u and/or d quarks are Λ's (isospin 0) or Σ's (isospin 1). If the third quark is a c, b, or t quark, its identity is given by a subscript.
3. Baryons with *one* u or d quark are Ξ's (isospin 1/2). One or two subscripts are used if one or both of the remaining quarks are heavy: thus Ξ_c, Ξ_{cc}, Ξ_b, *etc.**
4. Baryons with *no* u or d quarks are Ω's (isospin 0), and subscripts indicate any heavy-quark content.
5. A baryon that decays strongly has its mass as part of its name. Thus p, Σ^-, Ω^-, Λ_c^+, *etc.*, but $\Delta(1232)^0$, $\Sigma(1385)^-$, $\Xi_c(2645)^+$, *etc.*

In short, the number of u plus d quarks together with the isospin determine the main symbol, and subscripts indicate any content of heavy quarks. A Σ always has isospin 1, an Ω always has isospin 0, *etc.*

8.5. Exotic baryons

In 2003, several experiments reported finding a strangeness $S = +1$, charge $Q = +1$ baryon, and one experiment reported finding an $S = -2$, $Q = -2$ baryon; see the "Exotic Baryons" section of the Data Listings. Baryons with such quantum numbers cannot be made from three quarks, and thus they are exotic. The $S = +1$ baryon, which once would have been called a Z, was quickly dubbed the $\Theta(1540)^+$, and we propose to name the $S = -2$ baryon the $\Phi(1860)$.

Footnote and Reference:

* Sometimes a prime is necessary to distinguish two Ξ_c's in the same SU(n) multiplet. See the "Note on Charmed Baryons" in the Charmed Baryon Listings.

1. Particle Data Group: M. Aguilar-Benitez *et al.,* Phys. Lett. **170B** (1986).

9. QUANTUM CHROMODYNAMICS AND ITS COUPLING

9.1. The QCD Lagrangian

Revised September 2005 by I. Hinchliffe (LBNL).

Quantum Chromodynamics (QCD), the gauge field theory which describes the strong interactions of colored quarks and gluons, is one of the components of the SU(3)×SU(2)×U(1) Standard Model. A quark of specific flavor (such as a charm quark) comes in 3 colors; gluons come in eight colors; hadrons are color-singlet combinations of quarks, anti-quarks, and gluons. The Lagrangian describing the interactions of quarks and gluons is (up to gauge-fixing terms)

$$L_{\rm QCD} = -\frac{1}{4} F^{(a)}_{\mu\nu}\ F^{(a)\mu\nu} + i \sum_q \overline{\psi}^i_q\ \gamma^\mu\ (D_\mu)_{ij}\ \psi^j_q - \sum_q m_q\ \overline{\psi}^i_q\ \psi_{qi}\ , \tag{9.1}$$

$$F^{(a)}_{\mu\nu} = \partial_\mu\ A^a_\nu - \partial_\nu\ A^a_\mu - g_s\ f_{abc}\ A^b_\mu\ A^c_\nu\ , \tag{9.2}$$

$$(D_\mu)_{ij} = \delta_{ij}\ \partial_\mu + i g_s\ \sum_a \frac{\lambda^a_{i,j}}{2} A^a_\mu\ , \tag{9.3}$$

where g_s is the QCD coupling constant, and the f_{abc} are the structure constants of the SU(3) algebra (the λ matrices and values for f_{abc} can be found in "SU(3) Isoscalar Factors and Representation Matrices," Sec. 36 of this *Review*). The $\psi^i_q(x)$ are the 4-component Dirac spinors associated with each quark field of (3) color i and flavor q, and the $A^a_\mu(x)$ are the (8) Yang-Mills (gluon) fields. A complete list of the Feynman rules which derive from this Lagrangian, together with some useful color-algebra identities, can be found in Ref. 1.

The principle of "asymptotic freedom" determines that the renormalized QCD coupling is small only at high energies, and it is only in this domain that high-precision tests—similar to those in QED—can be performed using perturbation theory. Nonetheless, there has been in recent years much progress in understanding and quantifying the predictions of QCD in the nonperturbative domain, for example, in soft hadronic processes and on the lattice [2]. This short review will concentrate on QCD at short distances (large momentum transfers), where perturbation theory is the standard tool. It will discuss the processes that are used to determine the coupling constant of QCD. Other recent reviews of the coupling constant measurements may be consulted for a different perspective [3–6].

9.2. The QCD coupling and renormalization scheme

The renormalization scale dependence of the effective QCD coupling $\alpha_s = g_s^2/4\pi$ is controlled by the β-function:

$$\mu\frac{\partial\alpha_s}{\partial\mu} = 2\beta(\alpha_s) = \ -\frac{\beta_0}{2\pi}\,\alpha_s^2 \ -\frac{\beta_1}{4\pi^2}\,\alpha_s^3 \ -\frac{\beta_2}{64\pi^3}\,\alpha_s^4 \ -\cdots\,, \tag{9.4a}$$

$$\beta_0 = 11 - \frac{2}{3}n_f\ , \tag{9.4b}$$

$$\beta_1 = 51 - \frac{19}{3}n_f\ , \tag{9.4c}$$

$$\beta_2 = 2857 - \frac{5033}{9}n_f + \frac{325}{27}n_f^2\ , \tag{9.4d}$$

where n_f is the number of quarks with mass less than the energy scale μ. The expression for the next term in this series (β_3) can be found in Ref. 8. In solving this differential equation for α_s, a constant of integration is introduced. This constant is the fundamental constant of QCD that must be determined from experiment in addition to the quark masses. The most sensible choice for this constant is the value of α_s at a fixed-reference scale μ_0. It has become standard to choose $\mu_0 = M_Z$. The value at other values of μ can be obtained from $\log(\mu^2/\mu_0^2) = \int_{\alpha_s(\mu_0)}^{\alpha_s(\mu)} \frac{d\alpha}{\beta(\alpha)}$. It is also convenient to introduce the dimensional parameter Λ, since this provides a parameterization of the μ dependence of α_s. The definition of Λ is arbitrary. One way to define it (adopted here) is to write a solution of Eq. (9.4) as an expansion in inverse powers of $\ln(\mu^2)$:

$$\alpha_s(\mu) = \frac{4\pi}{\beta_0\ \ln(\mu^2/\Lambda^2)}\left[1 - \frac{2\beta_1}{\beta_0^2}\ \frac{\ln\left[\ln(\mu^2/\Lambda^2)\right]}{\ln(\mu^2/\Lambda^2)} + \frac{4\beta_1^2}{\beta_0^4\ln^2(\mu^2/\Lambda^2)}\right.$$
$$\left.\times\left(\left(\ln\left[\ln(\mu^2/\Lambda^2)\right] - \frac{1}{2}\right)^2 + \frac{\beta_2\beta_0}{8\beta_1^2} - \frac{5}{4}\right)\right]\ . \tag{9.5}$$

This solution illustrates the *asymptotic freedom* property: $\alpha_s \to 0$ as $\mu \to \infty$ and shows that QCD becomes strongly coupled at $\mu \sim \Lambda$.

Consider a "typical" QCD cross section which, when calculated perturbatively [7], starts at $\mathcal{O}(\alpha_s)$:

$$\sigma = A_1\ \alpha_s + A_2\ \alpha_s^2 + \ \cdots\ . \tag{9.6}$$

The coefficients A_1, A_2 come from calculating the appropriate Feynman diagrams. In performing such calculations, various divergences arise, and these must be regulated in a consistent way. This requires a particular renormalization scheme (RS). The most commonly used one is the modified minimal subtraction ($\overline{\rm MS}$) scheme [9]. This involves continuing momentum integrals from 4 to 4–2ϵ dimensions, and then subtracting off the resulting $1/\epsilon$ poles and also $(\ln 4\pi - \gamma_E)$, which is an artifact of continuing the dimension. (Here γ_E is the Euler-Mascheroni constant.) To preserve the dimensionless nature of the coupling, a mass scale μ must also be introduced: $g \to \mu^\epsilon g$. The finite coefficients

A_i ($i \geq 2$) thus obtained depend implicitly on the renormalization convention used and explicitly on the scale μ.

The first two coefficients (β_0, β_1) in Eq. (9.4) are independent of the choice of RS. In contrast, the coefficients of terms proportional to α_s^n for $n > 3$ *are* RS-dependent. The form given above for β_2 is in the $\overline{\rm MS}$ scheme.

The fundamental theorem of RS dependence is straightforward. Physical quantities, such as the cross section calculated to all orders in perturbation theory, do not depend on the RS. It follows that a truncated series *does* exhibit RS dependence. In practice, QCD cross sections are known to leading order (LO), or to next-to-leading order (NLO), or in some cases, to next-to-next-to-leading order (NNLO); and it is only the latter two cases, which have reduced RS dependence, that are useful for precision tests. At NLO the RS dependence is completely given by one condition which can be taken to be the value of the renormalization scale μ. At NNLO this is not sufficient, and μ is no longer equivalent to a choice of scheme; both must now be specified. One, therefore, has to address the question of what is the "best" choice for μ within a given scheme, usually $\overline{\rm MS}$. There is no definite answer to this question—higher-order corrections do not "fix" the scale, rather they render the theoretical predictions less sensitive to its variation.

One should expect that choosing a scale μ characteristic of the typical energy scale (E) in the process would be most appropriate. In general, a poor choice of scale generates terms of order $\ln(E/\mu)$ in the A_i's. Various methods have been proposed including choosing the scale for which the next-to-leading-order correction vanishes ("Fastest Apparent Convergence [10]"); the scale for which the next-to-leading-order prediction is stationary [11], (*i.e.*, the value of μ where $d\sigma/d\mu = 0$); or the scale dictated by the effective charge scheme [12] or by the BLM scheme [13]. By comparing the values of α_s that different reasonable schemes give, an estimate of theoretical errors can be obtained. It has also been suggested to replace the perturbation series by its Padé approximant [14]. Results obtained using this method have, in certain cases, a reduced scale dependence [15,16]. One can also attempt to determine the scale from data by allowing it to vary and using a fit to determine it. This method can allow a determination of the error due to the scale choice and can give more confidence in the end result [17]. In many of the cases discussed below this scale uncertainty is the dominant error.

An important corollary is that if the higher-order corrections are naturally small, then the additional uncertainties introduced by the μ dependence are likely to be small. There are some processes, however, for which the choice of scheme *can* influence the extracted value of $\alpha_s(M_Z)$. There is no resolution to this problem other than to try to calculate even more terms in the perturbation series. It is important to note that, since the perturbation series is an asymptotic expansion, there is a limit to the precision with which any theoretical quantity can be calculated. In some processes, the highest-order perturbative terms may be comparable in size to nonperturbative corrections (sometimes called higher-twist or renormalon effects, for a discussion see Ref. 18); an estimate of these terms and their uncertainties is required if a value of α_s is to be extracted.

Cases occur where there is more than one large scale, say μ_1 and μ_2. In these cases, terms appear of the form $\log(\mu_1/\mu_2)$. If the ratio μ_1/μ_2 is large, these logarithms can render naive perturbation theory unreliable and a modified perturbation expansion that

takes these terms into account must be used. A few examples are discussed below.

In the cases where the higher-order corrections to a process are known and are large, some caution should be exercised when quoting the value of α_s. In what follows, we will attempt to indicate the size of the theoretical uncertainties on the extracted value of α_s. There are two simple ways to determine this error. First, we can estimate it by comparing the value of $\alpha_s(\mu)$ obtained by fitting data using the QCD formula to highest known order in α_s, and then comparing it with the value obtained using the next-to-highest-order formula (μ is chosen as the typical energy scale in the process). The corresponding Λ's are then obtained by evolving $\alpha_s(\mu)$ to $\mu = M_Z$ using Eq. (9.4) to the same order in α_s as the fit. Alternatively, we can vary the value of μ over a reasonable range, extracting a value of Λ for each choice of μ. This method is by its nature imprecise, since "reasonable" involves a subjective judgment. In either case, if the perturbation series is well behaved, the resulting error on $\alpha_s(M_Z)$ will be small.

In the above discussion we have ignored quark-mass effects, *i.e.*, we have assumed an idealized situation where quarks of mass greater than μ are neglected completely. In this picture, the β-function coefficients change by discrete amounts as flavor thresholds (a quark of mass M) are crossed when integrating the differential equation for α_s. Now imagine an experiment at energy scale μ; for example, this could be $e^+e^- \to$ hadrons at center-of-mass energy μ. If $\mu \gg M$, the mass M is negligible and the process is well described by QCD with n_f massless flavors and its parameter $\alpha_{(n_f)}$ up to terms of order M^2/μ^2. Conversely if $\mu \ll M$, the heavy quark plays no role and the process is well described by QCD with $n_f - 1$ massless flavors and its parameter $\alpha_{(n_f-1)}$ up to terms of order μ^2/M^2. If $\mu \sim M$, the effects of the quark mass are process-dependent and cannot be absorbed into the running coupling. The values of $\alpha_{(n_f)}$ and $\alpha_{(n_f-1)}$ are related so that a physical quantity calculated in both "theories" gives the same result [19]. This implies, for $\mu = M$

$$\alpha_{(n_f)}(M) = \alpha_{(n_f-1)}(M) - \frac{11}{72\pi^2}\alpha^3_{(n_f-1)}(M) + \mathcal{O}\left(\alpha^4_{(n_f-1)}\right) \tag{9.7}$$

which is almost identical to the naive result $\alpha_{(n_f)}(M) = \alpha_{(n_f-1)}(M)$. Here M is the mass of the value of the running quark mass defined in the $\overline{\rm MS}$ scheme (see the note on "Quark Masses" in the Particle Listings for more details), *i.e.*, where $M_{\overline{\rm MS}}(M) = M$.

It also follows that, for a relationship such as Eq. (9.5) to remain valid for all values of μ, Λ must also change as flavor thresholds are crossed, the value corresponds to an effective number of massless quarks: $\Lambda \to \Lambda^{(n_f)}$ [19,20]. The formulae are given in the 1998 edition of this review.

Experiments such as those from deep-inelastic scattering involve a range of energies. After fitting to their measurements to QCD, the resulting fit can be expressed as a value of $\alpha_s(M_Z)$.

Determinations of α_s result from fits to data using NLO and NNLO: LO fits are not useful and their values will not be used in the following. Care must be exercised when comparing results from NLO and NNLO. In order to compare the values of α_s from

various experiments, they must be evolved using the renormalization group to a common scale. For convenience, this is taken to be the mass of the Z boson. The extrapolation is performed using same order in perturbation theory as was used in the analysis. This evolution uses third-order perturbation theory and can introduce additional errors particularly if extrapolation from very small scales is used. The variation in the charm and bottom quark masses ($M_b = 4.3 \pm 0.2$ GeV and $M_c = 1.3 \pm 0.3$ GeV are used [21]) can also introduce errors. These result in a fixed value of $\alpha_s(2\ \text{GeV})$ giving an uncertainty in $\alpha_s(M_Z) = \pm 0.001$ if only perturbative evolution is used. There could be additional errors from nonperturbative effects that enter at low energy.

9.3. QCD in deep-inelastic scattering

The original and still one of the most powerful quantitative tests of perturbative QCD is the breaking of Bjorken scaling in deep-inelastic lepton-hadron scattering. The review 'Structure Functions," (Sec. 16 of this *Review*) describes the basic formalism and reviews the data. α_s is obtained together with the structure functions. The global fit from MRST04 [13] of (Sec. 16) gives $\alpha_s(M_Z) = 0.1205 \pm 0.004$ from NLO and $\alpha_s(M_Z) = 0.1167 \pm 0.004$ from NNLO. Other fits are consistent with these values but cannot be averaged as they use overlapping data sets. The good agreement between the NLO and NNLO fits indicates that the theoretical uncertainties are under control.

Nonsinglet structure functions offer in principle the most precise test of the theory and cleanest way to extract α_s, since the Q^2 evolution is independent of the gluon distribution, which is much more poorly constrained. The CCFR collaboration fit to the Gross-Llewellyn Smith sum rule [23] whose value at leading order is determined by baryon number and which is known to order α_s^3 [24,25] (NNLO); estimates of the order α_s^4 term are available [26].

$$\int_0^1 dx \left(F_3^{\overline{\nu}p}(x,Q^2) + F_3^{\nu p}(x,Q^2) \right) = 3 \left[1 - \frac{\alpha_s}{\pi}(1 + 3.58\frac{\alpha_s}{\pi} + 19.0 \left(\frac{\alpha_s}{\pi}\right)^2) - \Delta HT \right] \quad , \tag{9.8}$$

where the higher-twist contribution ΔHT is estimated to be $(0.09 \pm 0.045)/Q^2$ in [24,27] and to be somewhat smaller by [28]. The CCFR collaboration [29], combines their data with that from other experiments [30] and gives $\alpha_s\ (\sqrt{3}\ \text{GeV}) = 0.28 \pm 0.035\ (\text{expt.}) \pm 0.05\ (\text{sys})\ ^{+0.035}_{-0.03}\ (\text{theory})$. The error from higher-twist terms (assumed to be $\Delta HT = 0.05 \pm 0.05$) dominates the theoretical error. If the higher twist result of [28] is used, the central value increases to 0.31 in agreement with the fit of [31]. This value corresponds to $\alpha_s(M_Z) = 0.118 \pm 0.011$. Fits of the Q^2 evolution [32] of xF_3 using the CCFR data using NNLO and estimates of NNNLO QCD and higher twist terms enables the effect of these terms to be studied.

The spin-dependent structure functions, measured in polarized lepton-nucleon scattering, can also be used to determine α_s. Note that these experiments measure asymmetries and rely on measurements of unpolarized data to extract the spin-dependent structure functions. Here the values of $Q^2 \sim 2.5\ \text{GeV}^2$ are small, particularly for the

E143 data [33], and higher-twist corrections are important. A fit [34] by an experimental group using the measured spin dependent structure functions for several experiments [33,35] as well as their own data has been made. When data from HERMES [36] and SMC are included [37] $\alpha_s(M_Z) = 0.120 \pm 0.009$ is obtained: this is used in the final average.

α_s can also be determined from the Bjorken spin sum rule [38]; a fit gives $\alpha_s(M_Z) = 0.118^{+0.010}_{-0.024}$ [39]; consistent with an earlier determination [40], the larger error being due to the extrapolation into the (unmeasured) small x region. Theoretically, the sum rule is preferable as the perturbative QCD result is known to higher order and these terms are important at the low Q^2 involved. It has been shown that the theoretical errors associated with the choice of scale are considerably reduced by the use of Padé approximants [15], which results in $\alpha_s(1.7\ \text{GeV}) = 0.328 \pm 0.03\ (\text{expt.}) \pm 0.025\ (\text{theory})$ corresponding to $\alpha_s(M_Z) = 0.116\ ^{+0.003}_{-0.005}\ (\text{expt.}) \pm 0.003\ (\text{theory})$. No error is included from the extrapolation into the region of x that is unmeasured. Should data become available at smaller values of x so that this extrapolation could be more tightly constrained. This result is not used in the final average.

9.4. QCD in decays of the τ lepton

The semi-leptonic branching ratio of the tau ($\tau \to \nu_\tau + \text{hadrons}$, R_τ) is an inclusive quantity. It is related to the contribution of hadrons to the imaginary part of the W self energy $\big(\Pi(s)\big)$. It is sensitive to a range of energies since it involves an integral

$$R_\tau \sim \int_0^{m_\tau^2} \frac{ds}{m_\tau^2}(1 - \frac{s}{m_\tau^2})^2 \left((1 + 2s/m_\tau^2) Im\Pi(1) + Im\Pi(0)\right) \tag{9.9}$$

where $Im\Pi(1)$ denotes the vector part and $Im\Pi(0)$ the scalar part. Since the scale involved is low, one must take into account nonperturbative (higher-twist) contributions which are suppressed by powers of the τ mass.

$$\begin{aligned} R_\tau =& 3.058\Bigg[1 + \delta_{EW} + \frac{\alpha_s(m_\tau)}{\pi} + 5.2\Big(\frac{\alpha_s(m_\tau)}{\pi}\Big)^2 + 26.4\Big(\frac{\alpha_s(m_\tau)}{\pi}\Big)^3 \\ &+ a\frac{m^2}{m_\tau^2} + b\frac{m\psi\overline{\psi}}{m_\tau^4} + c\frac{\psi\overline{\psi}\psi\overline{\psi}}{m_\tau^6} + \cdots\Bigg] . \end{aligned} \tag{9.10}$$

$\delta_{EW} = 0.0010$ is the electroweak correction. Here a, b, and c are dimensionless constants and m is a light quark mass. The term of order $1/m_\tau^2$ is a kinematical effect due to the light quark masses and is consequently very small. The nonperturbative terms are estimated using sum rules [41]. In total, they are estimated to be -0.014 ± 0.005 [42,43]. This estimate relies on there being no term of order Λ^2/m_τ^2 $\left(\text{note that } \frac{\alpha_s(m_\tau)}{\pi} \sim (\frac{0.5\ \text{GeV}}{m_\tau})^2\right)$. The a, b, and c can be determined from the data [44] by fitting to moments of the $\Pi(s)$ and separately to the final states accessed by the vector and axial parts of the W coupling. The values so extracted [45,46] are

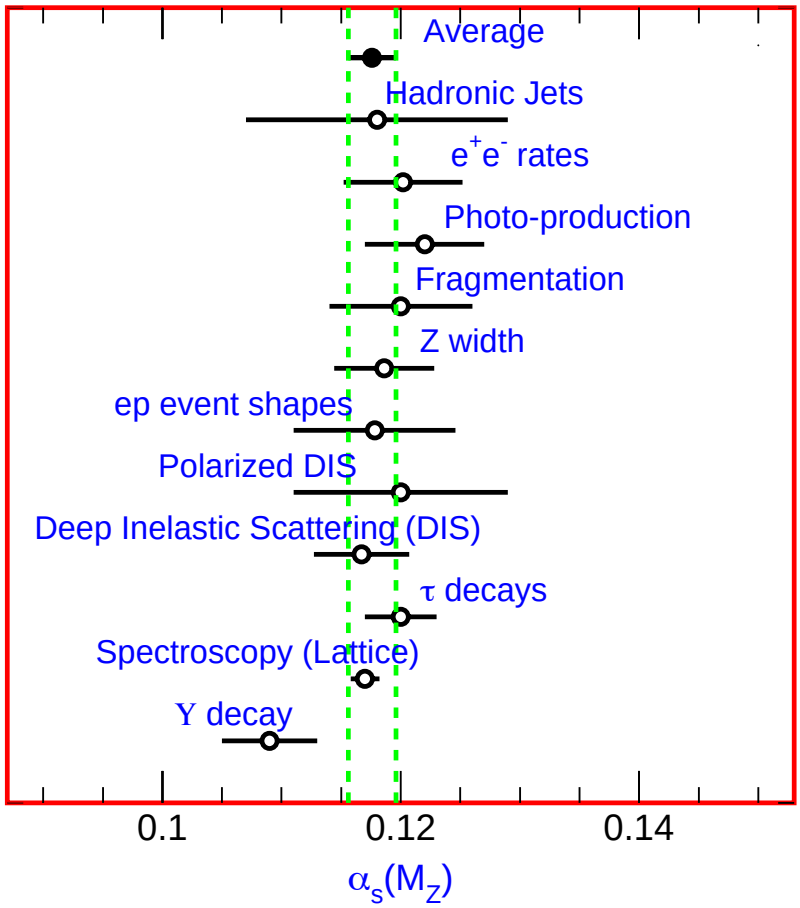


Figure 9.1: Summary of the value of $\alpha_s(M_Z)$ from various processes. The values shown indicate the process and the measured value of α_s extrapolated to $\mu = M_Z$. The error shown is the *total* error including theoretical uncertainties. The average quoted in this report which comes from these measurements is also shown. See text for discussion of errors.

consistent with the theoretical estimates. If the nonperturbative terms are omitted from the fit, the extracted value of $\alpha_s(m_\tau)$ decreases by ~ 0.02.

For $\alpha_s(m_\tau) = 0.35$ the perturbative series for R_τ is $R_\tau \sim 3.058(1{+}0.112{+}0.064{+}0.036)$. The size (estimated error) of the nonperturbative term is 20% (7%) of the size of the order α_s^3 term. The perturbation series is not very well convergent; if the order α_s^3 term is omitted, the extracted value of $\alpha_s(m_\tau)$ increases by 0.05. The order α_s^4 term has been estimated [47] and attempts made to resum the entire series [48,49]. These estimates can be used to obtain an estimate of the errors due to these unknown terms [50,51]. Another approach to estimating this α_s^4 term gives a contribution that is slightly larger than the α_s^3 term [52].

R_τ can be extracted from the semi-leptonic branching ratio from the relation $R_\tau = 1/(\mathrm{B}(\tau \to e\nu\overline{\nu}) - 1.97256)$; where $\mathrm{B}(\tau \to e\nu\overline{\nu})$ is measured directly or extracted from the lifetime, the muon mass, and the muon lifetime assuming universality of lepton couplings. Using the average lifetime of 290.6 ± 1.1 fs and a τ mass of 1776.99 ± 0.29

MeV from the PDG fit gives $R_\tau = 3.645 \pm 0.020$. The direct measurement of $B(\tau \to e\nu\overline{\nu})$ can be combined with $B(\tau \to \mu\nu\overline{\nu})$ to give $B(\tau \to e\nu\overline{\nu}) = 0.1785 \pm 0.0005$ which gives $R_\tau = 3.629 \pm 0.015$. Averaging these yields $\alpha_s(m_\tau) = 0.338 \pm 0.004$ using the experimental error alone. We assign a theoretical error equal to 40% of the contribution from the order α^3 term and all of the nonperturbative contributions. This then gives $\alpha_s(m_\tau) = 0.34 \pm 0.03$ for the final result. This corresponds to $\alpha_s(M_Z) = 0.120 \pm 0.003$. This result is consistent with that obtained by using the moments [53] of the integrand and is used in the average below.

9.5. QCD in high-energy hadron collisions

There are many ways in which perturbative QCD can be tested in high-energy hadron colliders. The quantitative tests are only useful if the process in question has been calculated beyond leading order in QCD perturbation theory. The production of hadronic jets with large transverse momentum in hadron-hadron collisions provides a direct probe of the scattering of quarks and gluons: $qq \to qq$, $qg \to qg$, $gg \to gg$, *etc.* Higher–order QCD calculations of the jet rates [54] and shapes are in impressive agreement with data [55]. This agreement has led to the proposal that these data could be used to provide a determination of α_s [56]. A set of structure functions is assumed and jet data are fitted over a very large range of transverse momenta to the QCD prediction for the underlying scattering process that depends on α_s. The evolution of the coupling over this energy range (40 to 250 GeV) is therefore tested in the analysis. CDF obtains $\alpha_s(M_Z) = 0.1178 \pm 0.0001$ (stat.) ± 0.0085 (syst.) [57]. Estimation of the theoretical errors is not straightforward. The structure functions used depend implicitly on α_s and an iteration procedure must be used to obtain a consistent result; different sets of structure functions yield different correlations between the two values of α_s. CDF includes a scale error of 4% and a structure function error of 5% in the determination of α_s. Ref. 56 estimates the error from unknown higher order QCD corrections to be ± 0.005. Combining these then gives $\alpha_s(M_Z) = 0.118 \pm 0.011$ which is used in the final average. For additional comments on comparisons between these data and theory see Ref. 4. Data are also available on the angular distribution of jets; these are also in agreement with QCD expectations [58,59].

QCD corrections to Drell-Yan type cross sections (*i.e.*, the production in hadron collisions by quark-antiquark annihilation of lepton pairs of invariant mass Q from virtual photons, or of real W or Z bosons), are known [60]. These $\mathcal{O}(\alpha_s)$ QCD corrections are sizable at small values of Q. The correction to W and Z production, as measured in $p\overline{p}$ collisions at $\sqrt{s} = 0.63$ TeV and $\sqrt{s} = 1.8,\ 1.96$ TeV, is of order 30%. The NNLO corrections to this process are known [61].

The production of W and Z bosons and photons at large transverse momentum can also be used to test QCD. The leading-order QCD subprocesses are $q\overline{q} \to Vg$ and $qg \to Vq$ ($V = W, Z, \gamma$). If the parton distributions are taken from other processes and a value of α_s assumed, then an absolute prediction is obtained. Conversely, the data can be used to extract information on quark and gluon distributions and on the value of α_s. The next-to-leading-order QCD corrections are known for photons [62,63], and for W/Z production [64], and so a precision test is possible. Data exist on photon production

from the CDF and DØ collaborations [65,66] and from fixed target experiments [67]. Detailed comparisons with QCD predictions [68] may indicate an excess of the data over the theoretical prediction at low value of transverse momenta, although other authors [69] find smaller excesses.

The UA2 collaboration [70] extracted a value of $\alpha_s(M_W) = 0.123 \pm 0.018$ (stat.) $\pm$ 0.017 (syst.) from the measured ratio $R_W = \dfrac{\sigma(W + 1\text{jet})}{\sigma(W + 0\text{jet})}$. The result depends on the algorithm used to define a jet, and the dominant systematic errors due to fragmentation and corrections for underlying events (the former causes jet energy to be lost, the latter causes it to be increased) are connected to the algorithm. There is also dependence on the parton distribution functions, and hence, α_s appears explicitly in the formula for R_W, and implicitly in the distribution functions. The UA2 result is not used in the final average. Data from CDF and DØ on the W p_t distribution [71] are in agreement with QCD but are not able to determine α_s with sufficient precision to have any weight in a global average.

In the region of low p_t, the fixed order perturbation theory is not applicable; one must sum terms of order $\alpha_s^n \ln^n(p_t/M_W)$ [72]. Data from DØ [73] on the p_t distribution of Z bosons agree well with these predictions.

The production rates of b quarks in $p\overline{p}$ have been used to determine α_s [74]. The next-to-leading-order QCD production processes [75] have been used. By selecting events where the b quarks are back-to-back in azimuth, the next-to-leading-order calculation can be used to compare rates to the measured value and a value of α_s extracted. The errors are dominated by the measurement errors, the choice of μ and the scale at which the structure functions are evaluated, and uncertainties in the choice of structure functions. The last were estimated by varying the structure functions used. The result is $\alpha_s(M_Z) = 0.113^{+0.009}_{-0.013}$, which is not included in the final average, as the measured $b\overline{b}$ cross section is not in very good agreement with perturbative QCD [76] and it is therefore difficult to interpret this result. Recent improvments in the theoretical undestanding [77] and measurements from CDF [78] now show good agreement between the measured cross-sections and the QCD predictions but there is no extraction of α_s using these.

9.6. QCD in heavy-quarkonium decay

Under the assumption that the hadronic and leptonic decay widths of heavy $Q\overline{Q}$ resonances can be factorized into a nonperturbative part—dependent on the confining potential—and a calculable perturbative part, the ratios of partial decay widths allow measurements of α_s at the heavy-quark mass scale. The most precise data come from the decay widths of the 1^{--} $J/\psi(1S)$ and Υ resonances. The total decay width of the Υ is predicted by perturbative QCD [79,80]

$$\begin{aligned} R_\mu(\Upsilon) &= \frac{\Gamma(\Upsilon \to \text{hadrons})}{\Gamma(\Upsilon \to \mu^+\mu^-)} \\ &= \frac{10(\pi^2 - 9)\alpha_s^3(M_b)}{9\pi\alpha_{\text{em}}^2} \end{aligned}$$

$$\times \left[1 + \frac{\alpha_s}{\pi}\left(-14.05 + \frac{3\beta_0}{2}\left(1.161 + \ln\left(\frac{2M_b}{M_\Upsilon}\right)\right)\right)\right] . \qquad (9.11)$$

Data are available for the Υ, Υ', Υ'', and J/ψ. The result is very sensitive to α_s and the data are sufficiently precise ($R_\mu(\Upsilon) = 37.28 \pm 0.75$) [81] that the theoretical errors will dominate. There are theoretical corrections to this formula due to the relativistic nature of the $Q\overline{Q}$ system which have been calculated [80] to order v^2/c^2. These corrections are more severe for the J/ψ. There are also nonperturbative corrections arising from annihilation from higher Fock states ("color octet" contribution) which can only be estimated [82]; again these are more severe for the J/ψ. The Υ gives $\alpha_s(M_b) = 0.183 \pm 0.01$, where the error includes that from the "color octet" term and the choice of scale which together dominate. The ratio of widths $\dfrac{\Upsilon \to \gamma gg}{\Upsilon \to ggg}$ has been measured by the CLEO collaboration and can be used to determine $\alpha_s(M_b) = 0.189 \pm 0.01 \pm 0.01$. The error is dominated by theoretical uncertainties associated with the scale choice; the uncertainty due to the "color octet" piece is not present in this case [83]. The theoretical uncertainties due to the production of photons in fragmentation [84] are small [85]. Higher order QCD calculations of the photon energy distribution are available [86]; this distribution could now be used to further test the theory. The width $\Gamma(\Upsilon \to e^+e^-)$ can also be used to determine α_s by using moments of the quantity $R_b(s) = \dfrac{\sigma(e^+e^- \to b\overline{b})}{\sigma(e^+e^- \to \mu^+\mu^-)}$ defined by $M_n = \int_0^\infty \dfrac{R_b(s)}{s^{n+1}}$ [87]. At large values of n, M_n is dominated by $\Gamma(\Upsilon \to e^+e^-)$. Higher order corrections are available and the method gives $\alpha_s(M_b) = 0.220 \pm 0.027$ [88]. The dominant error is theoretical and is dominated by the choice of scale and by uncertainties in Coulomb corrections that have been resummed in Ref. 89. These various Υ decay measurements can be combined and give $\alpha_s(M_b) = 0.185 \pm 0.01$ corresponding to $\alpha_s(M_Z) = 0.109 \pm 0.004$ which is used in the final average [83]. The mass of charmonium can also be used for determination of α_s after taking into account effects of analytic continuation (π^2-terms summation), and Coulomb summation [90].

9.7. Perturbative QCD in e^+e^- collisions

The total cross section for $e^+e^- \to$ hadrons is obtained (at low values of $\sqrt{s}$) by multiplying the muon-pair cross section by the factor $R = 3\Sigma_q e_q^2$. The higher-order QCD corrections to this quantity have been calculated, and the results can be expressed in terms of the factor:

$$R = R^{(0)}\left[1 + \frac{\alpha_s}{\pi} + C_2\left(\frac{\alpha_s}{\pi}\right)^2 + C_3\left(\frac{\alpha_s}{\pi}\right)^3 + \cdots\right] , \qquad (9.12)$$

where $C_2 = 1.411$ and $C_3 = -12.8$ [91].

$R^{(0)}$ can be obtained from the formula for $d\sigma/d\Omega$ for $e^+e^- \to f\overline{f}$ by integrating over Ω. The formula is given in Sec. 39.2 of this *Review*. This result is only correct in the zero-quark-mass limit. The $\mathcal{O}(\alpha_s)$ corrections are also known for massive quarks [92].

The principal advantage of determining α_s from R in e^+e^- annihilation is that there is no dependence on fragmentation models, jet algorithms, *etc.*

A measurement by CLEO [93] at $\sqrt{s} = 10.52$ GeV yields $\alpha_s(10.52\,\mathrm{GeV}) = 0.20\pm0.01\pm0.06$, which corresponds to $\alpha_s(M_Z) = 0.130\pm0.005\pm0.03$. A comparison of the theoretical prediction of Eq. (9.12) (corrected for the b-quark mass), with all the available data at values of $\sqrt{s}$ between 20 and 65 GeV, gives [94] $\alpha_s(35\ \mathrm{GeV}) = 0.146 \pm 0.030$. The size of the order α_s^3 term is of order 40% of that of the order α_s^2 and 3% of the order α_s. If the order α_s^3 term is not included, a fit to the data yields $\alpha_s\ (35\ \mathrm{GeV}) = 0.142 \pm 0.030$, indicating that the theoretical uncertainty is smaller than the experimental error.

Measurements of the ratio of hadronic to leptonic width of the Z at LEP and SLC, Γ_h/Γ_μ probe the same quantity as R. Using the average of $\Gamma_h/\Gamma_\mu = 20.767 \pm 0.025$ gives $\alpha_s(M_Z) = 0.1226 \pm 0.0038$ [95]. In performing this extraction it is necessary to include the electroweak corrections to the Z width. As these must be calculated beyond leading order, they depend on the top top-quark and Higgs masses. The latter is not yet meausured an is inferred from global fits to the electroweak data. There are additional theoretical errors arising from the choice of QCD scale. While this method has small theoretical uncertainties from QCD itself, it relies sensitively on the electroweak couplings of the Z to quarks [96]. The presence of new physics which changes these couplings via electroweak radiative corrections would invalidate the value of $\alpha_s(M_Z)$. An illustration of the sensitivity can be obtained by comparing this value with the one obtained from the global fits [95] of the various precision measurements at LEP/SLC and the W and top masses: $\alpha_s(M_Z) = 0.1186 \pm 0.0027$. The difference between these two values may be accounted for by systematic uncertainties as large as ±0.003 [95], therefore $\alpha_s(M_Z) = 0.1186 \pm 0.0042$ will be used in the final average.

An alternative method of determining α_s in e^+e^- annihilation is from measuring quantities that are sensitive to the relative rates of two-, three-, and four-jet events. A review should be consulted for more details [97] of the issues mentioned briefly here. In addition to simply counting jets, there are many possible choices of such "shape variables": thrust [98], energy-energy correlations [99], average jet mass, *etc.* All of these are infrared safe, which means they can be reliably calculated in perturbation theory. The starting point for all these quantities is the multijet cross section. For example, at order α_s, for the process $e^+e^- \to q\overline{q}g$: [100]

$$\frac{1}{\sigma}\,\frac{d^2\sigma}{dx_1 dx_2} = \frac{2\alpha_s}{3\pi}\,\frac{x_1^2 + x_2^2}{(1-x_1)(1-x_2)}\ , \tag{9.13}$$

$$x_i = \frac{2E_i}{\sqrt{s}} \tag{9.14}$$

where x_i are the center-of-mass energy fractions of the final-state (massless) quarks. A distribution in a "three-jet" variable, such as those listed above, is obtained by integrating this differential cross section over an appropriate phase space region for a fixed value of the variable. The order α_s^2 corrections to this process have been computed, as well as the 4-jet final states such as $e^+e^- \to qqgg$ [101].

There are many methods used by the e^+e^- experimental groups to determine α_s from the event topology. The jet-counting algorithm, originally introduced by the JADE collaboration [102], has been used by many other groups. Here, particles of momenta p_i and p_j are combined into a pseudo-particle of momentum $p_i + p_j$ if the invariant mass of the pair is less than $y_0\sqrt{s}$. The process is iterated until all pairs of particles or pseudoparticles have a mass-measure that exceeds $y_0\sqrt{s}$; the remaining number is then defined to be the jet multiplicity. The remaining number is then defined to be the number of jets in the event, and can be compared to the QCD prediction. The Durham algorithm is slightly different: in combining a pair of partons, it uses $M^2 = 2\min(E_i^2, E_j^2)(1-\cos\theta_{ij})$ for partons of energies E_i and E_j separated by angle θ_{ij} [103].

There are theoretical ambiguities in the way this process is carried out. Quarks and gluons are massless, whereas the observed hadrons are not, so that the massive jets that result from this scheme cannot be compared directly to the jets of perturbative QCD. Different recombination schemes have been tried, for example combining 3-momenta and then rescaling the energy of the cluster so that it remains massless. These schemes result in the same data giving slightly different values [104,105] of α_s. These differences can be used to determine a systematic error. In addition, since what is observed are hadrons rather than quarks and gluons, a model is needed to describe the evolution of a partonic final state into one involving hadrons, so that detector corrections can be applied. The QCD matrix elements are combined with a parton-fragmentation model. This model can then be used to correct the data for a direct comparison with the parton calculation. The different hadronization models that are used [106–109] model the dynamics that are controlled by nonperturbative QCD effects which we cannot yet calculate. The fragmentation parameters of these Monte Carlos are tuned to get agreement with the observed data. The differences between these models contribute to the systematic errors. The systematic errors from recombination schemes and fragmentation effects dominate over the statistical and other errors of the LEP/SLD experiments.

The scale M at which $\alpha_s(M)$ is to be evaluated is not clear. The invariant mass of a typical jet (or $\sqrt{sy_0}$) is probably a more appropriate choice than the e^+e^- center-of-mass energy. While there is no justification for doing so, if the value is allowed to float in the fit to the data, the fit improves and the data tend to prefer values of order $\sqrt{s}/10$ GeV for some variables [105,110]; the exact value depends on the variable that is fitted.

The perturbative QCD formulae can break down in special kinematical configurations. For example, the thrust (T) distribution contains terms of the type $\alpha_s \ln^2(1-T)$. The higher orders in the perturbation expansion contain terms of order $\alpha_s^n \ln^m(1-T)$. For $T \sim 1$ (the region populated by 2-jet events), the perturbation expansion is unreliable. The terms with $n \leq m$ can be summed to all orders in α_s [111]. If the jet recombination methods are used higher-order terms involve $\alpha_s^n \ln^m(y_0)$, these too can be resummed [112] The resummed results give better agreement with the data at large values of T. Some caution should be exercised in using these resummed results because of the possibility of overcounting; the showering Monte Carlos that are used for the fragmentation corrections also generate some of these leading-log corrections. Different schemes for combining the order α_s^2 and the resummations are available [113]. These different schemes result in

shifts in α_s of order ± 0.002. The use of the resummed results improves the agreement between the data and the theory; for more details see Ref. 114. An average of results at the Z resonance from SLD [105], OPAL [115], L3 [116], ALEPH [117], and DELPHI [118], using the combined α_s^2 and resummation fitting to a large set of shape variables, gives $\alpha_s(M_Z) = 0.122 \pm 0.007$. The errors in the values of $\alpha_s(M_Z)$ from these shape variables are totally dominated by the theoretical uncertainties associated with the choice of scale, and the effects of hadronization Monte Carlos on the different quantities fitted.

Estimates are available for the nonperturbative corrections to the mean value of $1 - T$ [119]. These are of order $1/E$ and involve a single parameter to be determined from experiment. These corrections can then be used as an alternative to those modeled by the fragmentation Monte Carlos. The DELPHI collaboration has fitted its data using an additional parameter to take into account these $1/E$ effects [120] and quotes for the $\overline{\rm MS}$ scheme $\alpha_s = 0.1217 \pm 0.0046$ and a significant $1/E$ term. This term vanishes in the RGI/ECH scheme and the data are well described by pure perturbation theory with consistent $\alpha_s = 0.1201 \pm 0.0020$.

Studies have been carried out at energies between $\sim$130 GeV [121] and $\sim$200 GeV [122]. These can be combined to give $\alpha_s(130\ \text{GeV}) = 0.114 \pm 0.008$ and $\alpha_s(189\ \text{GeV}) = 0.1104 \pm 0.005$. The dominant errors are theoretical and systematic and, most of these are in common at the two energies. These data and those at the Z resonance and below provide clear confirmation of the expected decrease in α_s as the energy is increased.

The LEP QCD working group [123] uses all LEP data Z mass and higher energies to perform a global fit using a large number of shape variables. It determines $\alpha_s(M_Z) = 0.1202 \pm 0.0003\ (\text{stat}) \pm 0.0049\ (\text{syst})$, (result quoted in Ref. 6) the error being dominated by theoretical uncertainties which are the most difficult to quantify.

Similar studies on event shapes have been undertaken at lower energies at TRISTAN, PEP/PETRA, and CLEO. A combined result from various shape parameters by the TOPAZ collaboration gives $\alpha_s(58\ \text{GeV}) = 0.125 \pm 0.009$, using the fixed order QCD result, and $\alpha_s(58\ \text{GeV}) = 0.132 \pm 0.008$ (corresponding to $\alpha_s(M_Z) = 0.123 \pm 0.007$), using the same method as in the SLD and LEP average [124]. The measurements of event shapes at PEP/PETRA are summarized in earlier editions of this note. A recent reevaluation of the JADE data [125] obtained using resummed QCD results with modern models of jet fragmentation and by averaging over several shape variables gives $\alpha_s(22\ \text{GeV}) = 0.151 \pm 0.004\ (\text{expt})\ ^{+0.014}_{-0.012}\ (\text{theory})$ which is used in the final average. These results also attempt to constrain the non-perurbative parameters and show a remarkable agreement with QCD even at low energies [126]. An analysis by the TPC group [127] gives $\alpha_s(29\ \text{GeV}) = 0.160 \pm 0.012$, using the same method as TOPAZ.

The CLEO collaboration fits to the order α_s^2 results for the two jet fraction at $\sqrt{s} = 10.53$ GeV, and obtains $\alpha_s(10.53\ \text{GeV}) = 0.164 \pm 0.004\ (\text{expt.}) \pm 0.014\ (\text{theory})$ [128]. The dominant systematic error arises from the choice of scale (μ), and is determined from the range of α_s that results from fit with $\mu = 10.53$ GeV, and a fit where μ is allowed to vary to get the lowest χ^2. The latter results in $\mu = 1.2$ GeV. Since the quoted result

corresponds to $\alpha_s(1.2\ \text{GeV}) = 0.35$, it is by no means clear that the perturbative QCD expression is reliable and the resulting error should, therefore, be treated with caution. A fit to many different variables as is done in the LEP/SLC analyses would give added confidence to the quoted error.

All these measurements are consistent with the LEP average quoted above which has the smallest statistical error; the systematic errors being mostly theoretical are likely to be strongly correlated between the measurements. The value of $\alpha_s(M_Z) = 0.1202 \pm 0.005$ is used in the final average.

The four jet final states can be used to measure the color factors of QCD, related to the relate strength of the couplings of quarks and gluons to each other. While these factors are not free parameters, the agreement between the measurements and expectations provides more evidence for the validity of QCD. The results are summarized in Ref. 129.

9.8. Scaling violations in fragmentation functions

Measurements of the fragmentation function $d_i(z, E)$, (the probability that a hadron of type i be produced with energy zE in e^+e^- collisions at $\sqrt{s} = 2E$) can be used to determine α_s. (Detailed definitions and a discussion of the properties of fragmentation functions can be found in Sec. 17 of this *Review*). As in the case of scaling violations in structure functions, perturbative QCD predicts only the E dependence. Hence, measurements at different energies are needed to extract a value of α_s. Because the QCD evolution mixes the fragmentation functions for each quark flavor with the gluon fragmentation function, it is necessary to determine each of these before α_s can be extracted.

The ALEPH collaboration has used data from energies ranging from $\sqrt{s} = 22$ GeV to $\sqrt{s} = 91$ GeV. A flavor tag is used to discriminate between different quark species, and the longitudinal and transverse cross sections are used to extract the gluon fragmentation function [130]. The result obtained is $\alpha_s(M_Z) = 0.126 \pm 0.007\ (\text{expt.}) \pm 0.006\ (\text{theory})$ [131]. The theory error is due mainly to the choice of scale. The OPAL collaboration [132] has also extracted the separate fragmentation functions. DELPHI [133] has also performed a similar analysis using data from other experiments at lower energy with the result $\alpha_s(M_Z) = 0.124 \pm 0.007\ (\text{expt.}) \pm 0.009\ (\text{theory})$. The larger theoretical error is due to the larger range of scales that were used in the fit. These results can be combined to give $\alpha_s(M_Z) = 0.125 \pm 0.005\ (\text{expt.}) \pm 0.008\ (\text{theory})$.

A global analysis [134] uses data on the production of π, K, p, and $\overline{p}$ from SLC [135], DELPHI [136], OPAL [137], ALEPH [138], and lower-energy data from the TPC collaboration [139]. A flavor tag and a three-jet analysis is used to disentangle the quark and gluon fragmentation functions. The value $\alpha_s(M_Z) = 0.1172^{+0.0055+0.0017}_{-0.0069-0.0025}$ is obtained. The second error is a theoretical one arising from the choice of scale. The fragmentation functions resulting from this fit are consistent with a recent fit of [140].

It is unclear how to combine the measurements discussed in the two previous paragraphs as much of the data used are common to both. If the theoretical errors dominate then a simple average is appropriate as the methods are different. For want of

a better solution, the naive average of $\alpha_s(M_Z) = 0.1201 \pm 0.006$ is used in the average value quoted below.

9.9. Photon structure functions

e^+e^- can also be used to study photon-photon interactions, which can be used to measure the structure function of a photon [141], by selecting events of the type $e^+e^- \to e^+e^- + hadrons$ which proceeds via two photon scattering. If events are selected where one of the photons is almost on mass shell and the other has a large invariant mass Q, then the latter probes the photon structure function at scale Q; the process is analogous to deep inelastic scattering where a highly virtual photon is used to probe the proton structure. This process was included in earlier versions of this *Review* which can be consulted for details on older measurements [142–145]. A review of the data can be found in [146]. Data are available from LEP [147–151] and from TRISTAN [152,153] which extend the range of Q^2 to of order 300 GeV^2 and x as low as $2 \times 10^{-3}$and show Q^2 dependence of the structure function that is consistent with QCD expectations. There is evidence for a hadronic (non-perturbative) component to the photon structure function that complicates attempts to extract a value of α_s from the data.

Ref. 154 uses data from PETRA, TRISTAN, and LEP to perform a combined fit. The higher data from LEP extend to higher Q^2 (< 780 GeV^2) and enable a measurement: $\alpha_s(m_Z) = 0.1198 \pm 0.0054$ which now is competitive with other results.

Experiments at HERA can also probe the photon structure function by looking at jet production in γp collisions; this is analogous to the jet production in hadron-hadron collisions which is sensitive to hadron structure functions. The data [155] are consistent with theoretical models [156].

9.10. Jet rates in *ep* collisions

At lowest order in α_s, the ep scattering process produces a final state of (1+1) jets, one from the proton fragment and the other from the quark knocked out by the process $e + quark \to e + quark$. At next order in α_s, a gluon can be radiated, and hence a (2+1) jet final state produced. By comparing the rates for these (1+1) and (2+1) jet processes, a value of α_s can be obtained. A NLO QCD calculation is available [157]. The basic methodology is similar to that used in the jet counting experiments in e^+e^- annihilation discussed above. Unlike those measurements, the ones in ep scattering are not at a fixed value of Q^2. In addition to the systematic errors associated with the jet definitions, there are additional ones since the structure functions enter into the rate calculations. A summary of the measurements from HERA can be found in Ref. 158, which clearly demonstrates the evidence for the evolution of $\alpha_s(Q^2)$ with Q^2. Results from H1 [159] $\alpha_s(M_Z) = 0.1175 \pm 0.002$ (expt.) ± 0.0053 (theor.) and ZEUS [160] $\alpha_s(M_Z) = 0.1179 \pm 0.0040$ (expt.) ± 0.005 (theor.) can be combined to give $\alpha_s(M_Z) = 0.1178 \pm 0.0033$ (expt.) ± 0.006 (theor.). which is used in the final average. The theoretical errors arise from scale choice, structure functions, and jet definitions.

Photoproduction of two or more jets via processes such as $\gamma + g \to q\overline{q}$ can also be observed at HERA. The process is similar to jet production in hadron-hadron

collisions. Agreement with perturbative QCD is excellent and ZEUS [161] quotes $\alpha_s(M_Z) = 0.1224 \pm 0.0020$ (expt) ± 0.0050 (theory) which is used in the average below.

9.11. QCD in diffractive events

In approximately 10% of the deep-inelastic scattering events at HERA a rapidity gap is observed [162]; that is events are seen where there are almost no hadrons produced in the direction of the incident proton. This was unexpected; QCD based models of the final state predicted that the rapidity interval between the quark that is hit by the electron and the proton remnant should be populated approximately evenly by the hadrons. Similar phenomena have been observed at the Tevatron in W and jet production. For a review see Ref. 163.

9.12. Lattice QCD

Lattice gauge theory can be used to calculate, using non-perturbative methods, a physical quantity that can be measured experimentally. The value of this quantity can then be used to determine the QCD coupling that enters in the calculation. The main theoretical differance between this approach and those discussed above is that, in the previous cases, precise calculations are restricted to high energy phenomena where perturbation theory can be applied due to the smallness of α_s in the appropriate energy regime. Lattice calculations enable reliable calculations to be done without this restriction. It is important to emphasize that this is exactly the same methodology used in the cases discussed above. The main quantitative difference is that the experimental measurements involved, such as the masses of Υ states, are so precise that their uncertainties have almost no impact on the final comparisons. A discussion of the uncertainties that enter into the QCD tests and determination of α_s is therefore almost exclusively a discussion of the techniques used in the calculations. In addition to α_s, other physical quantities such as the masses of the light quarks can be obtained. For a review of the methodology, see Ref. 164 [165]. For example, the energy levels of a $Q\overline{Q}$ system can be determined and then used to extract α_s. The masses of the $Q\overline{Q}$ states depend only on the quark mass and on α_s. Until a few years ago, calculations have not been performed for three light quark flavors. Results for zero ($n_f = 0$, quenched approximation) and two light flavors were extrapolated to $n_f = 3$. This major limitation has now been removed and a qualitative improvement in the calculations has occurred. Using the mass differences of Υ and Υ' and Υ'' and χ_b, Mason *et al.* [167] extract a value of $\alpha_s(M_Z) = 0.1170 \pm 0.0012$. Many other quantities such as the pion decay constant, and the masses of the B_s meson and Ω baryon are used and the overall consistency is excellent.

There have also been investigations of the running of α_s [173]. These show remarkable agreement with the two loop perturbative result of Eq. (9.5).

There are several sources of error in these estimates of $\alpha_s(M_Z)$. The experimental error associated with the measurements of the particle masses is negligible. The limited statistics of the Monte-Carlo calculation which can be improved only with more computational resources is one dominant error. The conversion from the lattice coupling

constant to the $\overline{\rm MS}$ constant is obtained using a perturbative expansion where one coupling expanded as a power series in the other. The series is known to third order and this leads to the second largest uncertainty [166]. Extra degrees of freedom introduced by calculating (using staggered fermions) on a lattice have to be removed: see Ref. 174 for a discussion of this point and the possible uncertainties related to it. The use of Wilson fermions involves a differant systematic. Results from Ref. 175 using this method with two light quark flavors are $\alpha_s(M_Z) = 0.112 \pm 0.003$, which illustrates the tendency for results using Wilson fermions to be systematically lower than those from staggered fermions.

In this review, we will use only the new result [166] of $\alpha_s(M_Z) = 0.1170 \pm 0.0012$, which is consistent with the value $\alpha_s(M_Z) = 0.121 \pm 0.003$ used in the last version of this review [167].

In addition to the strong coupling constant other quantities can be determined including the light quark masses [168]. Of particular interest are the decay constants of charmed and bottom mesons. These are required, for example, to facilitate the extraction of CKM elements from measurements of charm and bottom decay rates [169,170]. Some of these quantities such as the D-meson decay constant have been found to be in excellent agreement with experiment [171].

9.13. Conclusions

The need for brevity has meant that many other important topics in QCD phenomenology have had to be omitted from this review. One should mention in particular the study of exclusive processes (form factors, elastic scattering, ...), the behavior of quarks and gluons in nuclei, the spin properties of the theory, and QCD effects in hadron spectroscopy.

We have focused on those high-energy processes which currently offer the most quantitative tests of perturbative QCD. Figure 9.1 shows the values of $\alpha_s(M_Z)$ deduced from the various experiments. Figure 9.2 shows the values and the values of Q where they are measured. This figure clearly shows the experimental evidence for the variation of $\alpha_s(Q)$ with Q.

An average of the values in Fig. 9.1 gives $\alpha_s(M_Z) = 0.1176$, with a total χ^2 of 9 for eleven fitted points, showing good consistency among the data. The error on this average, assuming that all of the errors in the contributing results are uncorrelated, is ± 0.0009, and may be an underestimate. Almost all of the values used in the average are dominated by systematic, usually theoretical, errors. Only some of these, notably from the choice of scale, are correlated. The error on the lattice gauge theory result is the smallest and then there are several results with comparable small errors: these are the ones from τ decay, deep inelastic scattering, Υ decay and the Z^0 width. Omitting the lattice-QCD result from the average changes it to $\alpha_s(M_Z) = 0.1185$ or 1σ. All of the results that dominate the average are from NNLO. The NLO results have little weight, there are no LO results used. Almost all of the results have errors that are dominated by theoretical issues, either from unknown higher order perturbative corrections or estimates of non-perturbative contributions. It is therefore prudent be conservative and quote our

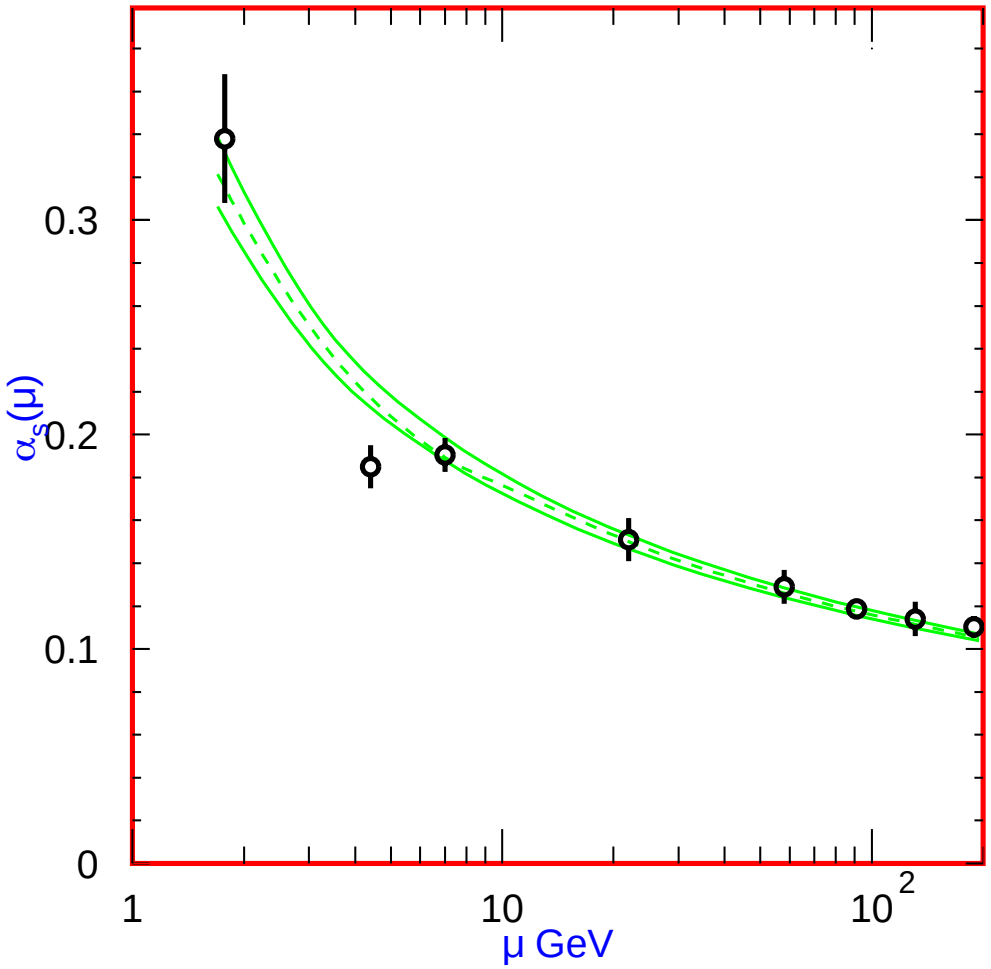


Figure 9.2: Summary of the values of $\alpha_s(\mu)$ at the values of μ where they are measured. The lines show the central values and the $\pm 1\sigma$ limits of our average. The figure clearly shows the decrease in $\alpha_s(\mu)$ with increasing μ. The data are, in increasing order of μ, τ width, Υ decays, deep inelastic scattering, e^+e^- event shapes at 22 GeV from the JADE data, shapes at TRISTAN at 58 GeV, Z width, and e^+e^- event shapes at 135 and 189 GeV.

average value as $\alpha_s(M_Z) = 0.1176 \pm 0.002$. Note that the average has moved by less than 1σ from the last version of this review. Future experiments can be expected to improve the measurements of α_s somewhat.

The value of α_s at any scale corresponding to our average can be obtained from `http://www-theory.lbl.gov/~ianh/alpha/alpha.html` which uses Eq. (9.5) to interpolate.

References:

1. R.K. Ellis *et al.*, "QCD and Collider Physics" (Cambridge 1996).
2. For reviews see, for example, A.S. Kronfeld and P.B. Mackenzie, Ann. Rev. Nucl. and Part. Sci. **43**, 793 (1993);
   H. Wittig, Int. J. Mod. Phys. **A12**, 4477 (1997).

3. For example see, P. Gambino, *International Conference on Lepton Photon Interactions*, Fermilab, USA, (2003); J. Butterworth *International Conference on Lepton Photon Interactions*, Upsala, Sweden, (2005).
4. G Salam at *International Conference on Lepton Photon Interactions*, Upsala, Sweden, (2005).
5. S. Bethke, `hep-ex/0407021`, Nucl. Phys. **B135**, 345 (2004);
   S. Bethke, J. Phys. **G26**, R27 (2000).
6. R.W.L. Jones *et al.*, JHEP **12**, 7 (2003).
7. See, for example, J. Collins "Renormalization: an introduction to renormalization, the renormalization group and the operator product expansion," (Cambridge University Press, Cambridge, 1984). "QCD and Collider Physics" (Cambridge 1996).
8. S.A. Larin *et al.*, Phys. Lett. **B400**, 379 (1997).
9. W.A. Bardeen *et al.*, Phys. Rev. **D18**, 3998 (1978).
10. G. Grunberg, Phys. Lett. **95B**, 70 (1980); Phys. Rev. **D29**, 2315 (1984).
11. P.M. Stevenson, Phys. Rev. **D23**, 2916 (1981); and Nucl. Phys. **B203**, 472 (1982).
12. S. Brodsky and H.J. Lu, SLAC-PUB-6389 (Nov. 1993).
13. S. Brodsky *et al.*, Phys. Rev. **D28**, 228 (1983).
14. M.A. Samuel *et al.*, Phys. Lett. **B323**, 188 (1994);
    M.A. Samuel *et al.*, Phys. Rev. Lett. **74**, 4380 (1995).
15. J. Ellis *et al.*, Phys. Rev. **D54**, 6986 (1996).
16. P.N. Burrows *et al.*, Phys. Lett. **B382**, 157 (1996).
17. P. Abreu *et al.*, Z. Phys. **C54**, 55 (1992).
18. A.H. Mueller, Phys. Lett. **B308**, 355 (1993).
19. W. Bernreuther, Ann. Phys. **151**, 127 (1983); Erratum Nucl. Phys. **B513**, 758 (1998);
    S.A. Larin *et al.*, Nucl. Phys. **B438**, 278 (1995).
20. K.G. Chetyrkin *et al.*, Phys. Rev. Lett. **79**, 2184 (1997);
    K.G. Chetyrkin *et al.*, Nucl. Phys. **B510**, 61 (1998).
21. See the Review on the "Quark Mass" in the Particle Listings for *Review of Particle Physics*.
22. A.D. Martin *et al.*, Phys. Lett. **B604**, 61 (2004).
23. D. Gross and C.H. Llewellyn Smith, Nucl. Phys. **B14**, 337 (1969).
24. J. Chyla and A.L. Kataev, Phys. Lett. **B297**, 385 (1992).
25. S.A. Larin and J.A.M. Vermaseren, Phys. Lett. **B259**, 345 (1991).
26. A.L. Kataev and V.V. Starchenko, Mod. Phys. Lett. **A10**, 235 (1995).
27. V.M. Braun and A.V. Kolesnichenko, Nucl. Phys. **B283**, 723 (1987).
28. M. Dasgupta and B. Webber, Phys. Lett. **B382**, 273 (1993).
29. J. Kim *et al.*, Phys. Rev. Lett. **81**, 3595 (1998).
30. D. Allasia *et al.*, Z. Phys. **C28**, 321 (1985);
    K. Varvell *et al.*, Z. Phys. **C36**, 1 (1987);
    V.V. Ammosov *et al.*, Z. Phys. **C30**, 175 (1986);
    P.C. Bosetti *et al.*, Nucl. Phys. **B142**, 1 (1978).

31. A.L. Kataev *et al.*, Nucl. Phys. **A666 & 667**, 184 (2000); Nucl. Phys. **B573**, 405 (2000).
32. A.L. Kataev *et al.*, hep-ph/0106221;
A.L. Kataev *et al.*, Nucl. Phys. Proc. Suppl. **116**, 105 (2003) hep-ph/0211151.
33. K. Abe *et al.*, Phys. Rev. Lett. **74**, 346 (1995); Phys. Lett. **B364**, 61 (1995); Phys. Rev. Lett. **75**, 25 (1995);
P.L. Anthony *et al.*, Phys. Rev. **D54**, 6620 (1996).
34. B. Adeva *et al.*, Phys. Rev. **D58**, 112002 (1998), Phys. Lett. **B420**, 180 (1998).
35. D. Adams *et al.*, Phys. Lett. **B329**, 399 (1995); Phys. Rev. **D56**, 5330 (1998); Phys. Rev. **D58**, 1112001 (1998);
K. Ackerstaff *et al.*, Phys. Lett. **B464**, 123 (1999).
36. P.L. Anthony *et al.*, Phys. Lett. **B463**, 339 (1999); Phys. Lett. **B493**, 19 (2000).
37. J. Blümlein and H. Böttcher, Nucl. Phys. **B636**, 225 (2002).
38. J.D. Bjorken, Phys. Rev. **148**, 1467 (1966).
39. G. Altarelli *et al.*, Nucl. Phys. **B496**, 337 (1997).
40. J. Ellis and M. Karliner, Phys. Lett. **B341**, 397 (1995).
41. M.A. Shifman *et al.*, Nucl. Phys. **B147**, 385 (1979).
42. S. Narison and A. Pich, Phys. Lett. **B211**, 183 (1988);
E. Braaten *et al.*, Nucl. Phys. **B373**, 581 (1992).
43. M. Neubert, Nucl. Phys. **B463**, 511 (1996).
44. F. Le Diberder and A. Pich, Phys. Lett. **B289**, 165 (1992).
45. R. Barate *et al.*, Z. Phys. **C76**, 1 (1997); Z. Phys. **C76**, 15 (1997);
K. Ackerstaff *et al.*, Eur. Phys. J. **C7**, 571 (1999).
46. T. Coan *et al.*, Phys. Lett. **B356**, 580 (1995).
47. A.L. Kataev and V.V. Starshenko, Mod. Phys. Lett. **A10**, 235 (1995).
48. F. Le Diberder and A. Pich, Phys. Lett. **B286**, 147 (1992).
49. C.J. Maxwell and D.J. Tong, Nucl. Phys. **B481**, 681 (1996).
50. G. Altarelli, Nucl. Phys. **B40**, 59 (1995);
G. Altarelli *et al.*, Z. Phys. **C68**, 257 (1995).
51. S. Narison, Nucl. Phys. **B40**, 47 (1995).
52. S. Narison, hep-ph/0508259.
53. S. Menke, hep-ex/0106011.
54. S.D. Ellis *et al.*, Phys. Rev. Lett. **64**, 2121 (1990);
F. Aversa *et al.*, Phys. Rev. Lett. **65**, 401 (1990);
W.T. Giele *et al.*, Phys. Rev. Lett. **73**, 2019 (1994);
S. Frixione *et al.*, Nucl. Phys. **B467**, 399 (1996).
55. F. Abe *et al.*, Phys. Rev. Lett. **77**, 438 (1996);
B. Abbott *et al.*, Phys. Rev. Lett. **86**, 1707 (2001).
56. W.T. Giele *et al.*, Phys. Rev. **D53**, 120 (1996).
57. T. Affolder *et al.*, Phys. Rev. Lett. **88**, 042001 (2002).
58. UA1 Collaboration: G. Arnison *et al.*, Phys. Lett. **B177**, 244 (1986).
59. F. Abe *et al.*, Phys. Rev. Lett. **77**, 533 (1996);
*ibid.,*erratum Phys. Rev. Lett. **78**, 4307 (1997);

B. Abbott, Phys. Rev. Lett. **80**, 666 (1998);
S. Abachi *et al.*, Phys. Rev. **D53**, 6000 (1996).
60. G. Altarelli *et al.*, Nucl. Phys. **B143**, 521 (1978).
61. R. Hamberg *et al.*, Nucl. Phys. **B359**, 343 (1991).
62. P. Aurenche *et al.*, Phys. Rev. **D42**, 1440 (1990);
P. Aurenche *et al.*, Phys. Lett. **140B**, 87 (1984);
P. Aurenche *et al.*, Nucl. Phys. **B297**, 661 (1988).
63. H. Baer *et al.*, Phys. Lett. **B234**, 127 (1990).
64. H. Baer and M.H. Reno, Phys. Rev. **D43**, 2892 (1991);
P.B. Arnold and M.H. Reno, Nucl. Phys. **B319**, 37 (1989).
65. F. Abe *et al.*, Phys. Rev. Lett. **73**, 2662 (1994).
66. B. Abbott *et al.*, Phys. Rev. Lett. **84**, 2786 (2001);
V.M. Abazov *et al.*, Phys. Rev. Lett. **87**, 251805 (2001).
67. G. Alverson *et al.*, Phys. Rev. **D48**, 5 (1993).
68. L. Apanasevich *et al.*, Phys. Rev. **D59**, 074007 (1999); Phys. Rev. Lett. **81**, 2642 (1998).
69. W. Vogelsang and A. Vogt, Nucl. Phys. **B453**, 334 (1995);
P. Aurenche *et al.*, Eur. Phys. J. **C9**, 107 (1999).
70. J. Alitti *et al.*, Phys. Lett. **B263**, 563 (1991).
71. S. Abachi *et al.*, Phys. Rev. Lett. **75**, 3226 (1995);
J. Womersley, private communication;
J. Huston, in the *Proceedings to the 29th International Conference on High-Energy Physics (ICHEP98)*, Vancouver, Canada (23–29 Jul 1998) `hep-ph/9901352`.
72. R.K. Ellis and S. Veseli, Nucl. Phys. **B511**, 649 (1998);
C.T. Davies *et al.*, Nucl. Phys. **B256**, 413 (1985);
G. Parisi and R. Petronzio, Nucl. Phys. **B154**, 427 (1979);
J.C. Collins *et al.*, Nucl. Phys. **B250**, 199 (1985).
73. DØ Collaboration: B. Abbott *et al.*, Phys. Rev. **D61**, 032004 (2000);
T. Affolder *et al.*, FERMILAB-PUB-99/220.
74. C. Albajar *et al.*, Phys. Lett. **B369**, 46 (1996).
75. M.L. Mangano *et al.*, Nucl. Phys. **B373**, 295 (1992).
76. D. Acosta *et al.*, Phys. Rev. **D65**, 052005 (2002).
77. M. Cacciari *et al.*, JHEP **0407**, 033 (2004).
78. D. Acosta et al. Phys. Rev. D 71, 032001 (2005).
79. R. Barbieri *et al.*, Phys. Lett. **95B**, 93 (1980);
P.B. Mackenzie and G.P. Lepage, Phys. Rev. Lett. **47**, 1244 (1981).
80. G.T. Bodwin *et al.*, Phys. Rev. **D51**, 1125 (1995).
81. *The Review of Particle Physics,*D.E. Groom *et al.*, Eur. Phys. J. **C15**, 1 (2000) and 2001 off-year partial update for the 2002 edition available on the PDG WWW pages (URL: `http://pdg.lbl.gov/`).
82. M. Gremm and A. Kapustin, Phys. Lett. **B407**, 323 (1997).
83. I. Hinchliffe and A.V. Manohar, Ann. Rev. Nucl. Part. Sci. **50**, 643 (2000).
84. S. Catani and F. Hautmann, Nucl. Phys. **B** (Proc. Supp.), vol. **39BC**, 359 (1995).
85. B. Nemati *et al.*, Phys. Rev. **D55**, 5273 (1997).

86. M. Kramer, Phys. Rev. **D60**, 111503 (1999).
87. M. Voloshin, Int. J. Mod. Phys. **A10**, 2865 (1995).
88. M. Jamin and A. Pich, Nucl. Phys. **B507**, 334 (1997).
89. J.H. Kuhn *et al.*, Nucl. Phys. **B534**, 356 (1998).
90. A.H. Hoang and M. Jamin, Phys. Lett. **B594**, 127 (2004).
91. S.G. Gorishny *et al.*, Phys. Lett. **B259**, 144 (1991);
L.R. Surguladze and M.A. Samuel, Phys. Rev. Lett. **66**, 560 (1991).
92. K.G. Chetyrkin and J.H. Kuhn, Phys. Lett. **B308**, 127 (1993).
93. R. Ammar *et al.*, Phys. Rev. **D57**, 1350 (1998).
94. D. Haidt, in *Directions in High Energy Physics,* vol. 14, p. 201, ed. P. Langacker (World Scientific, 1995).
95. LEP electoweak working group, presented at the *International Europhysics Conference on High Energy Physics, EPS05*, Lisboa Portugal (July 2005);
D. Abbaneo, *et al.*, LEPEWWG/2003-01.
96. A. Blondel and C. Verzegrassi, Phys. Lett. **B311**, 346 (1993);
G. Altarelli *et al.*, Nucl. Phys. **B405**, 3 (1993).
97. G. Dissertori *et al.*, "Quantum Chromodynamics: High Energy Experiments and Theory" (Oxford University Press, 2003).
98. E. Farhi, Phys. Rev. Lett. **39**, 1587 (1977).
99. C.L. Basham *et al.*, Phys. Rev. **D17**, 2298 (1978).
100. J. Ellis *et al.*, Nucl. Phys. **B111**, 253 (1976);
ibid.,erratum Nucl. Phys. **B130**, 516 (1977);
P. Hoyer *et al.*, Nucl. Phys. **B161**, 349 (1979).
101. R.K. Ellis *et al.*, Phys. Rev. Lett. **45**, 1226 (1980);
Z. Kunszt and P. Nason, ETH-89-0836 (1989).
102. S. Bethke *et al.*, Phys. Lett. **B213**, 235 (1988), Erratum *ibid.*, B523, 681 (1988).
103. S. Bethke *et al.*, Nucl. Phys. **B370**, 310 (1992).
104. M.Z. Akrawy *et al.*, Z. Phys. **C49**, 375 (1991).
105. K. Abe *et al.*, Phys. Rev. Lett. **71**, 2578 (1993); Phys. Rev. **D51**, 962 (1995).
106. B. Andersson *et al.*, Phys. Reports **97**, 33 (1983).
107. A. Ali *et al.*, Nucl. Phys. **B168**, 409 (1980);
A. Ali and F. Barreiro, Phys. Lett. **118B**, 155 (1982).
108. B.R. Webber, Nucl. Phys. **B238**, 492 (1984);
G. Marchesini *et al.*, Phys. Comm. **67**, 465 (1992).
109. T. Sjostrand and M. Bengtsson, Comp. Phys. Comm. **43**, 367 (1987);
T. Sjostrand, CERN-TH-7112/93 (1993).
110. O. Adriani *et al.*, Phys. Lett. **B284**, 471 (1992);
M. Akrawy *et al.*, Z. Phys. **C47**, 505 (1990);
B. Adeva *et al.*, Phys. Lett. **B248**, 473 (1990);
D. Decamp *et al.*, Phys. Lett. **B255**, 623 (1991).
111. S. Catani *et al.*, Phys. Lett. **B263**, 491 (1991).
112. S. Catani *et al.*, Phys. Lett. **B269**, 432 (1991); G. Dissertori and M. Schmelling, Phys. Lett. B **361**, 167 (1995);

S. Catani *et al.*, Phys. Lett. **B272**, 368 (1991);
N. Brown and J. Stirling, Z. Phys. **C53**, 629 (1992).
113. S. Catani *et al.*, Phys. Lett. **B269**, 432 (1991); Phys. Lett. **B295**, 269 (1992); Nucl. Phys. **B607**, 3 (1993); Phys. Lett. **B269**, 432 (1991).
114. M. Dasgupta and G. P. Salam, J. Phys. **G30**, R143 (2004).
115. P.D. Acton *et al.*, Z. Phys. **C55**, 1 (1992); Z. Phys. **C58**, 386 (1993).
116. O. Adriani *et al.*, Phys. Lett. **B284**, 471 (1992).
117. D. Decamp *et al.*, Phys. Lett. **B255**, 623 (1992); Phys. Lett. **B257**, 479 (1992).
118. P. Abreu *et al.*, Z. Phys. **C59**, 21 (1993); Phys. Lett. **B456**, 322 (1999); M. Acciarri *et al.*, Phys. Lett. **B404**, 390 (1997).
119. Y.L. Dokshitzer and B.R. Webber Phys. Lett. **B352**, 451 (1995); Y.L. Dokshitzer *et al.*, Nucl. Phys. **B511**, 396 (1997); Y.L. Dokshitzer *et al.*, JHEP **9801**, 011 (1998).
120. J. Abdallah *et al.*, [DELPHI Collaboration], Eur. Phys. J. **C29**, 285 (2003).
121. D. Buskulic *et al.*, Z. Phys. **C73**, 409 (1997); Z. Phys. **C73**, 229 (1997).
122. H. Stenzel *et al.* [ALEPH Collaboration], CERN-OPEN-99-303(1999); DELPHI Collaboration: Eur. Phys. J. **C14**, 557 (2000); M. Acciarri *et al.* [L3 Collaboration], Phys. Lett. **B489**, 65 (2000); OPAL Collaboration, PN-403 (1999); all submitted to *International Conference on Lepton Photon Interactions*, Stanford, USA (Aug. 1999); M. Acciarri *et al.* OPAL Collaboration], Phys. Lett. **B371**, 137 (1996); Z. Phys. **C72**, 191 (1996); K. Ackerstaff *et al.*, Z. Phys. **C75**, 193 (1997); ALEPH Collaboration: ALEPH 98-025 (1998).
123. `http://lepqcd.web.cern.ch/LEPQCD/annihilations/Welcome.html`.
124. Y. Ohnishi *et al.*, Phys. Lett. **B313**, 475 (1993).
125. P.A. Movilla Fernandez *et al.*, Eur. Phys. J. **C1**, 461 (1998), `hep-ex/020501`; S. Kluth *et al.*, `hep-ex/0305023`; J. Schieck *et al.*, `hep-ex/0408122`; O. Biebel *et al.*, Phys. Lett. **B459**, 326 (1999).
126. S. Kluth *et al.*, [JADE Collaboration], `hep-ex/0305023`.
127. D.A. Bauer *et al.*, SLAC-PUB-6518.
128. L. Gibbons *et al.*, CLNS 95-1323 (1995).
129. S. Kluth, Nucl. Phys. Proc. Suppl. **133**, 36 (2004).
130. P. Nason and B.R. Webber, Nucl. Phys. **B421**, 473 (1994).
131. D. Buskulic *et al.*, Phys. Lett. **B357**, 487 (1995); *ibid.,*erratum Phys. Lett. **B364**, 247 (1995).
132. R. Akers *et al.*, Z. Phys. **C68**, 203 (1995).
133. P. Abreu *et al.*, Phys. Lett. **B398**, 194 (1997).
134. B.A. Kniehl *et al.*, Phys. Rev. Lett. **85**, 5288 (2000).
135. K. Abe *et al.*, [SLD Collab.], Phys. Rev. **D59**, 052001 (1999).
136. P. Abreu *et al.*, [DELPHI Collab.], Eur. Phys. J. **C5**, 585 (1998).
137. G. Abbiendi *et al.*, [OPAL Collab.], Eur. Phys. J. **C11**, 217 (1999).

138. D. Buskulic *et al.*, [ALEPH Collab.], Z. Phys. **C66**, 355 (1995); R. Barate *et al.*, Eur. Phys. J. **C17**, 1 (2000).
139. H. Aihara, *et al.*, LBL-23737 (1988) (Unpublished).
140. L. Bourhis *et al.*, Eur. Phys. J. **C19**, 89 (2001).
141. E. Witten, Nucl. Phys. **B120**, 189 (1977).
142. C. Berger *et al.*, Nucl. Phys. **B281**, 365 (1987).
143. H. Aihara *et al.*, Z. Phys. **C34**, 1 (1987).
144. M. Althoff *et al.*, Z. Phys. **C31**, 527 (1986).
145. W. Bartel *et al.*, Z. Phys. **C24**, 231 (1984).
146. M. Erdmann, *International Conference on Lepton Photon Interactions*, Rome Italy (Aug. 2001) R. Nisius, `hep-ex/0210059`.
147. K. Ackerstaff *et al.*, Phys. Lett. **B412**, 225 (1997); Phys. Lett. **B411**, 387 (1997).
148. G. Abbiendi *et al.*, [OPAL Collaboration], Eur. Phys. J. **C18**, 15 (2000).
149. R. Barate *et al.*, Phys. Lett. **B458**, 152 (1999).
150. M. Acciarri *et al.*, Phys. Lett. **B436**, 403 (1998); Phys. Lett. **B483**, 373 (2000).
151. P. Abreu *et al.*, Z. Phys. **C69**, 223 (1996).
152. K. Muramatsu *et al.*, Phys. Lett. **B332**, 477 (1994).
153. S.K. Sahu *et al.*, Phys. Lett. **B346**, 208 (1995).
154. S. Albino *et al.*, Phys. Rev. Lett. **89**, 122004 (2002).
155. C. Adloff *et al.*, Eur. Phys. J. **C13**, 397 (2000); J. Breitweg *et al.*, Eur. Phys. J. **C11**, 35 (1999).
156. S. Frixione, Nucl. Phys. **B507**, 295 (1997); B.W. Harris and J.F. Owens, Phys. Rev. **D56**, 4007 (1997); M. Klasen and G. Kramer, Z. Phys. **C72**, 107 (1996).
157. D. Graudenz, Phys. Rev. **D49**, 3921 (1994); J.G. Korner *et al.*, Int. J. Mod. Phys. **A4**, 1781, (1989); S. Catani and M. Seymour, Nucl. Phys. **B485**, 291 (1997); M. Dasgupta and B.R. Webber, Eur. Phys. J. **C1**, 539 (1998); E. Mirkes and D. Zeppenfeld, Phys. Lett. **B380**, 205 (1996).
158. C. Glasman, at DIS2005 Madison, Wisconsin (2005).
159. T. Klugg (H1 collaboration) at DIS2005, Madison, Wisconsin (2005) C. Adloff *et al.*, Eur. Phys. J. **C19**, 289 (2001).
160. ZEUS Collaboration: DESY-05-019, `hep-ex/0502007`.
161. ZEUS Collaboration: S. Chekanov *et al.*, Phys. Lett. **B558**, 41 (2003); E. Tassi at DIS2001 Conference, Bologna (April 2001).
162. M. Derrick *et al.*, Phys. Lett. **B**, 369 (1996); T. Ahmed *et al.*, Nucl. Phys. **B435**, 3 (1995).
163. D.M. Janson *et al.*, `hep-ex/9905537`.
164. P. Weisz, Nucl. Phys. (Proc. Supp.) **B47**, 71 (1996).
165. P. E. L. Rakow, Nucl. Phys. (Proc. Supp.) **B140**, 34 (2005).
166. Q. Mason *et al.*, Phys. Rev. Lett. **95**, 052002 (2005).
167. C. T. H. Davies *et al.*, Phys. Rev. Lett. **92**, 022001 (2004).
168. C. Aubin *et al.*, Phys. Rev. **D70**, 031504 (2004).
169. C. Aubin *et al.*, Phys. Rev. Lett. **94**, 011601 (2005).

170. A. Gray *et al.*, [HPQCD Collaboration], hep-lat/0507015.
171. C. Aubin *et al.*, hep-lat/0506030.
172. A.X. El-Khadra *et al.*, Phys. Rev. Lett. **69**, 729 (1992);
A.X. El-Khadra *et al.*, FNAL 94-091/T (1994);
A.X. El-Khadra *et al.*, hep-ph/9608220.
173. G. de Divitiis *et al.*, Nucl. Phys. **B437**, 447 (1995).
174. S. Durr, hep-lat/0509026.
175. M. Gockeler *et al.*, hep-ph/0502212.

10. ELECTROWEAK MODEL AND CONSTRAINTS ON NEW PHYSICS

Revised September 2005 by J. Erler (U. Mexico) and P. Langacker (Univ. of Pennsylvania).

10.1 Introduction
10.2 Renormalization and radiative corrections
10.3 Cross-section and asymmetry formulae
10.4 Precision flavor physics
10.5 W and Z decays
10.6 Experimental results
10.7 Constraints on new physics

10.1. Introduction

The standard electroweak model (SM) is based on the gauge group [1] SU(2) × U(1), with gauge bosons W^i_μ, $i = 1, 2, 3$, and B_μ for the SU(2) and U(1) factors, respectively, and the corresponding gauge coupling constants g and g'. The left-handed fermion fields $\psi_i = \begin{pmatrix} \nu_i \\ \ell_i^- \end{pmatrix}$ and $\begin{pmatrix} u_i \\ d'_i \end{pmatrix}$ of the i^{th} fermion family transform as doublets under SU(2), where $d'_i \equiv \sum_j V_{ij}\, d_j$, and V is the Cabibbo-Kobayashi-Maskawa mixing matrix. (Constraints on V and tests of universality are discussed in Ref. 2 and in the Section on the Cabibbo-Kobayashi-Maskawa mixing matrix. The extension of the formalism to allow an analogous leptonic mixing matrix is discussed in "Neutrino Mass" in the Particle Listings.) The right-handed fields are SU(2) singlets. In the minimal model there are three fermion families and a single complex Higgs doublet $\phi \equiv \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}$.

After spontaneous symmetry breaking the Lagrangian for the fermion fields is

$$\begin{aligned} \mathscr{L}_F = & \sum_i \overline{\psi}_i \left(i\, \partial\!\!\!/ - m_i - \frac{g m_i H}{2 M_W} \right) \psi_i \\ & - \frac{g}{2\sqrt{2}} \sum_i \overline{\psi}_i\, \gamma^\mu\, (1-\gamma^5)(T^+\, W^+_\mu + T^-\, W^-_\mu)\, \psi_i \\ & - e \sum_i q_i\, \overline{\psi}_i\, \gamma^\mu\, \psi_i\, A_\mu \\ & - \frac{g}{2\cos\theta_W} \sum_i \overline{\psi}_i\, \gamma^\mu (g^i_V - g^i_A \gamma^5)\, \psi_i\, Z_\mu\ . \end{aligned} \tag{10.1}$$

$\theta_W \equiv \tan^{-1}(g'/g)$ is the weak angle; $e = g \sin\theta_W$ is the positron electric charge; and $A \equiv B\cos\theta_W + W^3 \sin\theta_W$ is the (massless) photon field. $W^\pm \equiv (W^1 \mp i W^2)/\sqrt{2}$ and $Z \equiv -B\sin\theta_W + W^3 \cos\theta_W$ are the massive charged and neutral weak boson fields, respectively. T^+ and T^- are the weak isospin raising and lowering operators. The vector

and axial-vector couplings are

$$g_V^i \equiv t_{3L}(i) - 2q_i \sin^2\theta_W \ , \tag{10.2a}$$

$$g_A^i \equiv t_{3L}(i) \ , \tag{10.2b}$$

where $t_{3L}(i)$ is the weak isospin of fermion i ($+1/2$ for u_i and ν_i; $-1/2$ for d_i and e_i) and q_i is the charge of ψ_i in units of e.

The second term in $\mathscr{L}_F$ represents the charged-current weak interaction [3,4]. For example, the coupling of a W to an electron and a neutrino is

$$-\frac{e}{2\sqrt{2}\sin\theta_W}\left[W_\mu^-\ \overline{e}\ \gamma^\mu(1-\gamma^5)\nu + W_\mu^+\ \overline{\nu}\ \gamma^\mu\ (1-\gamma^5)e\right] \ . \tag{10.3}$$

For momenta small compared to M_W, this term gives rise to the effective four-fermion interaction with the Fermi constant given (at tree level, *i.e.*, lowest order in perturbation theory) by $G_F/\sqrt{2} = g^2/8M_W^2$. CP violation is incorporated in the SM by a single observable phase in V_{ij}. The third term in $\mathscr{L}_F$ describes electromagnetic interactions (QED), and the last is the weak neutral-current interaction.

In Eq. (10.1), m_i is the mass of the i^{th} fermion ψ_i. For the quarks these are the current masses. For the light quarks, as described in "The Note on Quark Masses" in the Particle Listings, $\widehat{m}_u \approx 1.5$–$4$ MeV, $\widehat{m}_d \approx 4$–8 MeV, and $\widehat{m}_s \approx 80$–130 MeV. These are running $\overline{\rm MS}$ masses evaluated at the scale $\mu = 2$ GeV. (In this Section we denote quantities defined in the $\overline{\rm MS}$ scheme by a caret; the exception is the strong coupling constant, α_s, which will always correspond to the $\overline{\rm MS}$ definition and where the caret will be dropped.) For the heavier quarks we use QCD sum rule constraints [5] and recalculate their masses in each call of our fits to account for their direct α_s dependence. We find, $\widehat{m}_c(\mu = \widehat{m}_c) = 1.290^{+0.040}_{-0.045}$ GeV and $\widehat{m}_b(\mu = \widehat{m}_b) = 4.207 \pm 0.031$ GeV, with a correlation of 29%. The top quark "pole" mass, $m_t = 172.7 \pm 2.9$ GeV, is an average [6] of published CDF [7] and DØ [8] results from run I and of preliminary results from run II. We are working, however, with $\overline{\rm MS}$ masses in all expressions to minimize theoretical uncertainties, and therefore convert this result to the top quark $\overline{\rm MS}$ mass,

$$\widehat{m}_t(\mu = \widehat{m}_t) = m_t[1 - \frac{4}{3}\frac{\alpha_s}{\pi} + \mathcal{O}(\alpha_s^2)],$$

using the three-loop formula from Ref. 9. This introduces an additional uncertainty which we estimate to 0.6 GeV (the size of the three-loop term). We are assuming that the kinematic mass extracted from the collider events corresponds within this uncertainty to the pole mass. Using the BLM optimized [10] version of the two-loop perturbative QCD formula [11] (as we did in previous editions of this *Review*) gives virtually identical results. Thus, we will use $m_t = 172.7 \pm 2.9 \pm 0.6$ GeV $\approx 172.7 \pm 3.0$ GeV (together with $M_H = 117$ GeV) for the numerical values quoted in Sec. 10.2–Sec. 10.4. In the presence of right-handed neutrinos, Eq. (10.1) gives rise also to Dirac neutrino masses. The possibility of Majorana masses is discussed in "Neutrino mass" in the Particle Listings.

H is the physical neutral Higgs scalar which is the only remaining part of ϕ after spontaneous symmetry breaking. The Yukawa coupling of H to ψ_i, which is flavor diagonal in the minimal model, is $gm_i/2M_W$. In non-minimal models there are additional charged and neutral scalar Higgs particles [12].

10.2. Renormalization and radiative corrections

The SM has three parameters (not counting the Higgs boson mass, M_H, and the fermion masses and mixings). A particularly useful set is:

(*a*) The fine structure constant $\alpha = 1/137.03599911(46)$, determined from the $e^{\pm}$ anomalous magnetic moment, the quantum Hall effect, and other measurements [13]. In most electroweak renormalization schemes, it is convenient to define a running α dependent on the energy scale of the process, with $\alpha^{-1} \sim 137$ appropriate at very low energy. (The running has also been observed directly [14].) For scales above a few hundred MeV this introduces an uncertainty due to the low-energy hadronic contribution to vacuum polarization. In the modified minimal subtraction ($\overline{\rm MS}$) scheme [15] (used for this *Review*), and with $\alpha_s(M_Z) = 0.120$ for the QCD coupling at M_Z, we have $\widehat{\alpha}(m_\tau)^{-1} = 133.445 \pm 0.017$ and $\widehat{\alpha}(M_Z)^{-1} = 127.918 \pm 0.018$. The latter corresponds to a quark sector contribution (without the top) to the conventional (on-shell) QED coupling, $\alpha(M_Z) = \dfrac{\alpha}{1 - \Delta\alpha(M_Z)}$, of $\Delta\alpha^{(5)}_{\rm had}(M_Z) \approx 0.02791 \pm 0.00013$. These values are updated from Ref. 16 with a reduced uncertainty by a factor of 1/3 because they account for the latest results from τ decays (moving $\Delta\alpha^{(5)}_{\rm had}(M_Z)$ up by somewhat less than one standard deviation) and a reanalysis of the CMD 2 collaboration results after correcting a radiative correction [17]. See Ref. 18 for a discussion in the context of the anomalous magnetic moment of the muon. The correlation of the latter with $\widehat{\alpha}(M_Z)$, as well as the non-linear α_s dependence of $\widehat{\alpha}(M_Z)$ and the resulting correlation with the input variable α_s, are fully taken into account in the fits. This is done by using as actual input (fit constraint) instead of $\Delta\alpha^{(5)}_{\rm had}(M_Z)$ the analogous low-energy contribution by the three light quarks, $\Delta\alpha^{(3)}_{\rm had}(1.8\ \text{GeV}) = 0.00577 \pm 0.00010$, and by calculating the perturbative and heavy quark contributions to $\widehat{\alpha}(M_Z)$ in each call of the fits according to Ref. 16. The uncertainty is from e^+e^- annihilation data below 1.8 GeV and τ decay data, from isospin breaking effects (affecting the interpretation of the τ data); from uncalculated higher order perturbative and non-perturbative QCD corrections; and from the $\overline{\rm MS}$ quark masses. Such a short distance mass definition (unlike the pole mass) is free from non-perturbative and renormalon uncertainties. Various recent evaluations of $\Delta\alpha^{(5)}_{\rm had}$ are summarized in Table 10.1, where the relation between the on-shell and $\overline{\rm MS}$ definitions is given by

$$\Delta\widehat{\alpha}(M_Z) - \Delta\alpha(M_Z) = \frac{\alpha}{\pi}\left(\frac{100}{27} - \frac{1}{6} - \frac{7}{4}\ln\frac{M_Z^2}{M_W^2}\right) \approx 0.0072$$

to leading order, where the first term is from fermions and the other two are from $W^{\pm}$ loops which are usually excluded from the on-shell definition. Most of the older results relied on $e^+e^- \to$ hadrons cross-section measurements up to energies of 40 GeV, which were somewhat higher than the QCD prediction, suggested stronger running, and were less precise. The most recent results typically assume the validity of perturbative QCD (PQCD) at scales of 1.8 GeV and above, and are

in reasonable agreement with each other. (Evaluations in the on-shell scheme utilize resonance data from BES [38] as further input.) There is, however, some discrepancy between analyzes based on $e^+e^- \to$ hadrons cross-section data and those based on τ decay spectral functions [18–20]. The latter imply lower central values for the extracted M_H of $\mathcal{O}(10$ GeV). The discrepancy originates from the kinematic region $\sqrt{s} \gtrsim 0.6$ GeV. However, at least some of it appears to be experimental. The $e^+e^- \to \pi^+\pi^-$ cross-sections measured by the SND collaboration [39] are significantly larger than the older results by the CMD collaboration [40]. The data from SND are also about one standard deviation higher than those by CMD 2 [17] but in perfect agreement with information from τ decays. As an alternative to cross-section scans, one can use the high statistics radiative return events [41] at e^+e^- accelerators operating at resonances such as the Φ or the $\Upsilon(4S)$. The method is systematics dominated. The $\pi^+\pi^-$ radiative return results from the Φ obtained by the KLOE collaboration [42] for energies above the ρ peak are significantly lower compared to SND, while CMD 2 lies in between. Results for three and four pion final states are in better agreement. Further improvement of this dominant theoretical uncertainty in the interpretation of precision data will require better measurements of the cross-section for $e^+e^- \to$ hadrons below the charmonium resonances, as well as in the threshold region of the heavy quarks (to improve the precision in $\widehat{m}_c(\widehat{m}_c)$ and $\widehat{m}_b(\widehat{m}_b)$).

(*b*) The Fermi constant, $G_F = 1.16637(1) \times 10^{-5}$ GeV^{-2}, determined from the muon lifetime formula [43,44],

$$\tau_\mu^{-1} = \frac{G_F^2 m_\mu^5}{192\pi^3} F\left(\frac{m_e^2}{m_\mu^2}\right)\left(1 + \frac{3}{5}\frac{m_\mu^2}{M_W^2}\right)$$

$$\times \left[1 + \left(\frac{25}{8} - \frac{\pi^2}{2}\right)\frac{\alpha(m_\mu)}{\pi} + C_2\frac{\alpha^2(m_\mu)}{\pi^2}\right] , \qquad (10.4a)$$

where

$$F(x) = 1 - 8x + 8x^3 - x^4 - 12x^2 \ln x , \qquad (10.4b)$$

$$C_2 = \frac{156815}{5184} - \frac{518}{81}\pi^2 - \frac{895}{36}\zeta(3) + \frac{67}{720}\pi^4 + \frac{53}{6}\pi^2 \ln(2) , \qquad (10.4c)$$

and

$$\alpha(m_\mu)^{-1} = \alpha^{-1} - \frac{2}{3\pi}\ln\left(\frac{m_\mu}{m_e}\right) + \frac{1}{6\pi} \approx 136 . \qquad (10.4d)$$

The $\mathcal{O}(\alpha^2)$ corrections to μ decay have been completed in Ref. 44. The remaining uncertainty in G_F is from the experimental input.

(*c*) The Z-boson mass, $M_Z = 91.1876 \pm 0.0021$ GeV, determined from the Z-lineshape scan at LEP 1 [45].

Table 10.1: Recent evaluations of the on-shell $\Delta\alpha^{(5)}_{\rm had}(M_Z)$. For better comparison we adjusted central values and errors to correspond to a common and fixed value of $\alpha_s(M_Z) = 0.120$. References quoting results without the top quark decoupled are converted to the five flavor definition. Ref. [31] uses $\Lambda_{\rm QCD} = 380 \pm 60$ MeV; for the conversion we assumed $\alpha_s(M_Z) = 0.118 \pm 0.003$.

Reference	Result	Comment
Martin, Zeppenfeld [21]	0.02744 ± 0.00036	PQCD for $\sqrt{s} > 3$ GeV
Eidelman, Jegerlehner [22]	0.02803 ± 0.00065	PQCD for $\sqrt{s} > 40$ GeV
Geshkenbein, Morgunov [23]	0.02780 ± 0.00006	$\mathcal{O}(\alpha_s)$ resonance model
Burkhardt, Pietrzyk [24]	0.0280 ± 0.0007	PQCD for $\sqrt{s} > 40$ GeV
Swartz [25]	0.02754 ± 0.00046	use of fitting function
Alemany *et al.* [26]	0.02816 ± 0.00062	incl. τ decay data
Krasnikov, Rodenberg [27]	0.02737 ± 0.00039	PQCD for $\sqrt{s} > 2.3$ GeV
Davier & Höcker [28]	0.02784 ± 0.00022	PQCD for $\sqrt{s} > 1.8$ GeV
Kühn & Steinhauser [29]	0.02778 ± 0.00016	complete $\mathcal{O}(\alpha_s^2)$
Erler [16]	0.02779 ± 0.00020	conv. from $\overline{\rm MS}$ scheme
Davier & Höcker [30]	0.02770 ± 0.00015	use of QCD sum rules
Groote *et al.* [31]	0.02787 ± 0.00032	use of QCD sum rules
Martin *et al.* [32]	0.02741 ± 0.00019	includes new BES data
Burkhardt, Pietrzyk [33]	0.02763 ± 0.00036	PQCD for $\sqrt{s} > 12$ GeV
de Troconiz, Yndurain [34]	0.02754 ± 0.00010	PQCD for $s > 2$ GeV^2
Jegerlehner [35]	0.02765 ± 0.00013	conv. from MOM scheme
Hagiwara *et al.* [36]	0.02757 ± 0.00023	PQCD for $\sqrt{s} > 11.09$ GeV
Burkhardt, Pietrzyk [37]	0.02760 ± 0.00035	incl. KLOE data

With these inputs, $\sin^2\theta_W$ and the W-boson mass, M_W, can be calculated when values for m_t and M_H are given; conversely (as is done at present), M_H can be constrained by $\sin^2\theta_W$ and M_W. The value of $\sin^2\theta_W$ is extracted from Z-pole observables and neutral-current processes [45–48], and depends on the renormalization prescription. There are a number of popular schemes [49–56] leading to values which differ by small factors depending on m_t and M_H. The notation for these schemes is shown in Table 10.2. Discussion of the schemes follows the table.

(*i*) The on-shell scheme [49] promotes the tree-level formula $\sin^2\theta_W = 1 - M_W^2/M_Z^2$ to a definition of the renormalized $\sin^2\theta_W$ to all orders in perturbation theory, *i.e.*,

Table 10.2: Notations used to indicate the various schemes discussed in the text. Each definition of $\sin\theta_W$ leads to values that differ by small factors depending on m_t and M_H. Approximate values are also given for illustration.

Scheme	Notation and Value
On-shell	$s_W = \sin\theta_W \approx 0.2231$
NOV	$s_{M_Z} = \sin\theta_W \approx 0.2311$
$\overline{\rm MS}$	$\widehat{s}_Z = \sin\theta_W \approx 0.2312$
$\overline{\rm MS}$ ND	$\widehat{s}_{\rm ND} = \sin\theta_W \approx 0.2314$
Effective angle	$\overline{s}_f = \sin\theta_W \approx 0.2315$

$\sin^2\theta_W \to s_W^2 \equiv 1 - M_W^2/M_Z^2$:

$$M_W = \frac{A_0}{s_W(1-\Delta r)^{1/2}} \ , \tag{10.5a}$$

$$M_Z = \frac{M_W}{c_W} \ , \tag{10.5b}$$

where $c_W \equiv \cos\theta_W$, $A_0 = (\pi\alpha/\sqrt{2}G_F)^{1/2} = 37.2805(2)$ GeV, and Δr includes the radiative corrections relating α, $\alpha(M_Z)$, G_F, M_W, and M_Z. One finds $\Delta r \sim \Delta r_0 - \rho_t/\tan^2\theta_W$, where $\Delta r_0 = 1 - \alpha/\widehat{\alpha}(M_Z) = 0.06654(14)$ is due to the running of α, and $\rho_t = 3G_F m_t^2/8\sqrt{2}\pi^2 = 0.00935(m_t/172.7\ {\rm GeV})^2$ represents the dominant (quadratic) m_t dependence. There are additional contributions to Δr from bosonic loops, including those which depend logarithmically on M_H. One has $\Delta r = 0.03630 \mp 0.0011 \pm 0.00014$, where the second uncertainty is from $\alpha(M_Z)$. Thus the value of s_W^2 extracted from M_Z includes an uncertainty (∓ 0.00036) from the currently allowed range of m_t. This scheme is simple conceptually. However, the relatively large ($\sim 3\%$) correction from ρ_t causes large spurious contributions in higher orders.

(*ii*) A more precisely determined quantity $s^2_{M_Z}$ [50] can be obtained from M_Z by removing the (m_t, M_H) dependent term from Δr [51], *i.e.*,

$$s^2_{M_Z} c^2_{M_Z} \equiv \frac{\pi\alpha(M_Z)}{\sqrt{2}\,G_F\,M_Z^2} \ . \tag{10.6}$$

Using $\alpha(M_Z)^{-1} = 128.91 \pm 0.02$ yields $s^2_{M_Z} = 0.23108 \mp 0.00005$. The small uncertainty in $s^2_{M_Z}$ compared to other schemes is because most of the m_t dependence has been removed by definition. However, the m_t uncertainty reemerges when other quantities (*e.g.*, M_W or other Z-pole observables) are predicted in terms of M_Z.

Both s^2_W and $s^2_{M_Z}$ depend not only on the gauge couplings but also on the spontaneous-symmetry breaking, and both definitions are awkward in the presence of any extension of the SM which perturbs the value of M_Z (or M_W). Other definitions are motivated by the tree-level coupling constant definition $\theta_W = \tan^{-1}(g'/g)$.

(*iii*) In particular, the modified minimal subtraction ($\overline{\rm MS}$) scheme introduces the quantity $\sin^2\widehat{\theta}_W(\mu) \equiv \widehat{g}'^2(\mu)/\left[\widehat{g}^{\,2}(\mu) + \widehat{g}'^2(\mu)\right]$, where the couplings $\widehat{g}$ and $\widehat{g}'$ are defined by modified minimal subtraction and the scale μ is conveniently chosen to be M_Z for many electroweak processes. The value of $\widehat{s}^{\,2}_Z = \sin^2\widehat{\theta}_W(M_Z)$ extracted from M_Z is less sensitive than s^2_W to m_t (by a factor of $\tan^2\theta_W$), and is less sensitive to most types of new physics than s^2_W or $s^2_{M_Z}$. It is also very useful for comparing with the predictions of grand unification. There are actually several variant definitions of $\sin^2\widehat{\theta}_W(M_Z)$, differing according to whether or how finite $\alpha\ln(m_t/M_Z)$ terms are decoupled (subtracted from the couplings). One cannot entirely decouple the $\alpha\ln(m_t/M_Z)$ terms from all electroweak quantities because $m_t \gg m_b$ breaks SU(2) symmetry. The scheme that will be adopted here decouples the $\alpha\ln(m_t/M_Z)$ terms from the γ–Z mixing [15,52], essentially eliminating any $\ln(m_t/M_Z)$ dependence in the formulae for asymmetries at the Z-pole when written in terms of $\widehat{s}^{\,2}_Z$. (A similar definition is used for $\widehat{\alpha}$.) The various definitions are related by

$$\widehat{s}^{\,2}_Z = c\,(m_t, M_H) s^2_W = \overline{c}\,(m_t, M_H)\, s^2_{M_Z}\ , \tag{10.7}$$

where $c = 1.0359 \pm 0.0012$ and $\overline{c} = 1.0010 \mp 0.0004$. The quadratic m_t dependence is given by $c \sim 1 + \rho_t/\tan^2\theta_W$ and $\overline{c} \sim 1 - \rho_t/(1 - \tan^2\theta_W)$, respectively. The expressions for M_W and M_Z in the $\overline{\rm MS}$ scheme are

$$M_W = \frac{A_0}{\widehat{s}_Z(1 - \Delta\widehat{r}_W)^{1/2}}\ , \tag{10.8a}$$

$$M_Z = \frac{M_W}{\widehat{\rho}^{1/2}\widehat{c}_Z}\ , \tag{10.8b}$$

and one predicts $\Delta\widehat{r}_W = 0.06969 \pm 0.00004 \pm 0.00014$. $\Delta\widehat{r}_W$ has no quadratic m_t dependence, because shifts in M_W are absorbed into the observed G_F, so that the error in $\Delta\widehat{r}_W$ is dominated by $\Delta r_0 = 1 - \alpha/\widehat{\alpha}(M_Z)$ which induces the second quoted uncertainty. The quadratic m_t dependence has been shifted into $\widehat{\rho} \sim 1 + \rho_t$, where including bosonic loops, $\widehat{\rho} = 1.01043 \pm 0.00034$. Quadratic M_H effects are deferred to two-loop order, while the leading logarithmic M_H effects are dominant only for large M_H values which are currently disfavored by the precision data. As an illustration, the shift in M_W due to a large M_H (for fixed M_Z) is given by

$$\Delta_H M_W = -\frac{11}{96}\frac{\alpha}{\pi}\frac{M_W}{c^2_W - s^2_W}\ln\frac{M^2_H}{M^2_W} + \mathcal{O}(\alpha^2). \tag{10.9}$$

(*iv*) A variant $\overline{\rm MS}$ quantity $\widehat{s}^{\,2}_{\rm ND}$ (used in the 1992 edition of this *Review*) does not decouple the $\alpha\ln(m_t/M_Z)$ terms [53]. It is related to $\widehat{s}^{\,2}_Z$ by

$$\widehat{s}^{\,2}_Z = \widehat{s}^{\,2}_{\rm ND}/\left(1 + \frac{\widehat{\alpha}}{\pi}d\right)\ , \tag{10.10a}$$

$$d = \frac{1}{3}\left(\frac{1}{\widehat{s}^{\,2}} - \frac{8}{3}\right)\left[(1+\frac{\alpha_s}{\pi})\,\ln\frac{m_t}{M_Z} - \frac{15\alpha_s}{8\pi}\right], \qquad (10.10b)$$

Thus, $\widehat{s}^{\,2}_Z - \widehat{s}^{\,2}_{\rm ND} \sim -0.0002$ for $m_t = 172.7$ GeV.

(v) Yet another definition, the effective angle [54–56] $\overline{s}^2_f$ for the Z vector coupling to fermion f, is described in Sec. 10.3.

Experiments are at a level of precision that complete $\mathcal{O}(\alpha)$ radiative corrections must be applied. For neutral-current and Z-pole processes, these corrections are conveniently divided into two classes:

1. QED diagrams involving the emission of real photons or the exchange of virtual photons in loops, but not including vacuum polarization diagrams. These graphs often yield finite and gauge-invariant contributions to observable processes. However, they are dependent on energies, experimental cuts, *etc.*, and must be calculated individually for each experiment.
2. Electroweak corrections, including $\gamma\gamma$, γZ, ZZ, and WW vacuum polarization diagrams, as well as vertex corrections, box graphs, *etc.*, involving virtual W's and Z's. Many of these corrections are absorbed into the renormalized Fermi constant defined in Eq. (10.4). Others modify the tree-level expressions for Z-pole observables and neutral-current amplitudes in several ways [46]. One-loop corrections are included for all processes. In addition, certain two-loop corrections are also important. In particular, two-loop corrections involving the top quark modify ρ_t in $\widehat{\rho}$, Δr, and elsewhere by

$$\rho_t \to \rho_t[1 + R(M_H, m_t)\rho_t/3]\ . \qquad (10.11)$$

$R(M_H, m_t)$ is best described as an expansion in M_Z^2/m_t^2. The unsuppressed terms were first obtained in Ref. 57, and are known analytically [58]. Contributions suppressed by M_Z^2/m_t^2 were first studied in Ref. 59 with the help of small and large Higgs mass expansions, which can be interpolated. These contributions are about as large as the leading ones in Refs. 57 and 58. The complete two-loop calculation of Δr (without further approximation) has been performed in Refs. 60,61 for fermionic and purely bosonic diagrams, respectively. Similarly, the electroweak two-loop calculation for the relation between $\overline{s}^2_\ell$ and s^2_W is complete [62] except for the purely bosonic contribution. For M_H above its lower direct limit, $-17 < R \leq -13$.

Mixed QCD-electroweak contributions to gauge boson self-energies of order $\alpha\alpha_s m_t^2$ [63] and $\alpha\alpha_s^2 m_t^2$ [64] increase the predicted value of m_t by 6%. This is, however, almost entirely an artifact of using the pole mass definition for m_t. The equivalent corrections when using the $\overline{\rm MS}$ definition $\widehat{m}_t(\widehat{m}_t)$ increase m_t by less than 0.5%. The subleading $\alpha\alpha_s$ corrections [65] are also included. Further three-loop corrections of order $\alpha\alpha_s^2$ [66], $\alpha^3 m_t^6$ [67,68], and $\alpha^2\alpha_s m_t^4$ (for $M_H = 0$) [67], are rather small. The same is true for $\alpha^3 M_H^4$ [69] corrections unless M_H approaches 1 TeV.

The leading electroweak two-loop terms for the $Z \to b\bar{b}$-vertex of $\mathcal{O}(\alpha^2 m_t^4)$ have been obtained in Refs. 57,58, and the mixed QCD-electroweak contributions

in Refs. 70,71. The $\mathcal{O}(\alpha\alpha_s)$-vertex corrections involving massless quarks [72] add coherently, resulting in a sizable effect and shift the extracted $\alpha_s(M_Z)$ by $\approx +0.0007$.

Throughout this *Review* we utilize electroweak radiative corrections from the program GAPP [73], which works entirely in the $\overline{\rm MS}$ scheme, and which is independent of the package ZFITTER [56].

10.3. Cross-section and asymmetry formulae

It is convenient to write the four-fermion interactions relevant to ν-hadron, ν-e, and parity violating e-hadron neutral-current processes in a form that is valid in an arbitrary gauge theory (assuming massless left-handed neutrinos). One has

$$-\mathscr{L}^{\nu\text{Hadron}} = \frac{G_F}{\sqrt{2}}\,\overline{\nu}\,\gamma^\mu\,(1-\gamma^5)\nu$$

$$\times\sum_i\left[\epsilon_L(i)\,\overline{q}_i\,\gamma_\mu(1-\gamma^5)q_i + \epsilon_R(i)\,\overline{q}_i\,\gamma_\mu(1+\gamma^5)q_i\right]\,, \tag{10.12}$$

$$-\mathscr{L}^{\nu e} = \frac{G_F}{\sqrt{2}}\,\overline{\nu}_\mu\,\gamma^\mu(1-\gamma^5)\nu_\mu\,\overline{e}\,\gamma_\mu(g_V^{\nu e} - g_A^{\nu e}\gamma^5)e \tag{10.13}$$

(for ν_e-e or $\overline{\nu}_e$-e, the charged-current contribution must be included), and

$$-\mathscr{L}^{e\text{Hadron}} = -\frac{G_F}{\sqrt{2}}$$

$$\times\sum_i\left[C_{1i}\,\overline{e}\,\gamma_\mu\,\gamma^5\,e\,\overline{q}_i\,\gamma^\mu\,q_i + C_{2i}\,\overline{e}\,\gamma_\mu\,e\,\overline{q}_i\,\gamma^\mu\,\gamma^5\,q_i\right]\,. \tag{10.14}$$

(One must add the parity-conserving QED contribution.)

The SM expressions for $\epsilon_{L,R}(i)$, $g_{V,A}^{\nu e}$, and C_{ij} are given in Table 10.3. Note, that $g_{V,A}^{\nu e}$ and the other quantities are coefficients of effective four-Fermi operators, which differ from the quantities defined in Eq. (10.2) in the radiative corrections and in the presence of possible physics beyond the SM.

A precise determination of the on-shell s_W^2, which depends only very weakly on m_t and M_H, is obtained from deep inelastic neutrino scattering from (approximately) isoscalar targets [74]. The ratio $R_\nu \equiv \sigma_{\nu N}^{NC}/\sigma_{\nu N}^{CC}$ of neutral- to charged-current cross-sections has been measured to 1% accuracy by the CDHS [75] and CHARM [76] collaborations at CERN. The CCFR [77] collaboration at Fermilab has obtained an even more precise result and the NOMAD [78] collaboration anticipates a 0.3% measurement, so it is important to obtain theoretical expressions for R_ν and $R_{\overline{\nu}} \equiv \sigma_{\overline{\nu}N}^{NC}/\sigma_{\overline{\nu}N}^{CC}$ to comparable accuracy. Fortunately, most of the uncertainties from the strong interactions and neutrino spectra cancel in the ratio. The largest theoretical uncertainty is associated with the c-threshold, which mainly affects σ^{CC}. Using the slow rescaling prescription [79] the central value of $\sin^2\theta_W$ from CCFR varies as $0.0111(m_c\,[\text{GeV}] - 1.31)$, where m_c is the effective

Table 10.3: Standard Model expressions for the neutral-current parameters for ν-hadron, ν-e, and e-hadron processes. At tree level, $\rho = \kappa = 1$, $\lambda = 0$. If radiative corrections are included, $\rho_{\nu N}^{NC} = 1.0081$, $\widehat{\kappa}_{\nu N}(\langle Q^2\rangle = -12\ \mathrm{GeV}^2) = 0.9978$, $\widehat{\kappa}_{\nu N}(\langle Q^2\rangle = -35\ \mathrm{GeV}^2) = 0.9964$, $\lambda_{uL} = -0.0031$, $\lambda_{dL} = -0.0025$, and $\lambda_{dR} = 2\,\lambda_{uR} = 7.5 \times 10^{-5}$. For ν-e scattering, $\rho_{\nu e} = 1.0127$ and $\widehat{\kappa}_{\nu e} = 0.9965$ (at $\langle Q^2\rangle = 0$.). For atomic parity violation and the SLAC polarized electron experiment, $\rho'_{eq} = 0.9876$, $\rho_{eq} = 1.0006$, $\widehat{\kappa}'_{eq} = 1.0026$, $\widehat{\kappa}_{eq} = 1.0299$, $\lambda_{1d} = -2\,\lambda_{1u} = 3.6 \times 10^{-5}$, $\lambda_{2u} = -0.0121$ and $\lambda_{2d} = 0.0026$. The dominant m_t dependence is given by $\rho \sim 1 + \rho_t$, while $\widehat{\kappa} \sim 1$ ($\overline{\mathrm{MS}}$) or $\kappa \sim 1 + \rho_t/\tan^2\theta_W$ (on-shell).

Quantity	Standard Model Expression
$\epsilon_L(u)$	$\rho_{\nu N}^{NC}\left(\frac{1}{2} - \frac{2}{3}\widehat{\kappa}_{\nu N}\,\widehat{s}_Z^2\right) + \lambda_{uL}$
$\epsilon_L(d)$	$\rho_{\nu N}^{NC}\left(-\frac{1}{2} + \frac{1}{3}\widehat{\kappa}_{\nu N}\,\widehat{s}_Z^2\right) + \lambda_{dL}$
$\epsilon_R(u)$	$\rho_{\nu N}^{NC}\left(-\frac{2}{3}\widehat{\kappa}_{\nu N}\,\widehat{s}_Z^2\right) + \lambda_{uR}$
$\epsilon_R(d)$	$\rho_{\nu N}^{NC}\left(\frac{1}{3}\widehat{\kappa}_{\nu N}\,\widehat{s}_Z^2\right) + \lambda_{dR}$
$g_V^{\nu e}$	$\rho_{\nu e}\left(-\frac{1}{2} + 2\widehat{\kappa}_{\nu e}\,\widehat{s}_Z^2\right)$
$g_A^{\nu e}$	$\rho_{\nu e}\left(-\frac{1}{2}\right)$
C_{1u}	$\rho'_{eq}\left(-\frac{1}{2} + \frac{4}{3}\widehat{\kappa}'_{eq}\,\widehat{s}_Z^2\right) + \lambda_{1u}$
C_{1d}	$\rho'_{eq}\left(\frac{1}{2} - \frac{2}{3}\widehat{\kappa}'_{eq}\,\widehat{s}_Z^2\right) + \lambda_{1d}$
C_{2u}	$\rho_{eq}\left(-\frac{1}{2} + 2\widehat{\kappa}_{eq}\,\widehat{s}_Z^2\right) + \lambda_{2u}$
C_{2d}	$\rho_{eq}\left(\frac{1}{2} - 2\widehat{\kappa}_{eq}\,\widehat{s}_Z^2\right) + \lambda_{2d}$

mass which is numerically close to the $\overline{\mathrm{MS}}$ mass $\widehat{m}_c(\widehat{m}_c)$, but their exact relation is unknown at higher orders. For $m_c = 1.31 \pm 0.24$ GeV (determined from ν-induced dimuon production [80]) this contributes ± 0.003 to the total uncertainty $\Delta \sin^2\theta_W \sim \pm 0.004$. (The experimental uncertainty is also ± 0.003.) This uncertainty largely cancels, however, in the Paschos-Wolfenstein ratio [81],

$$R^- = \frac{\sigma_{\nu N}^{NC} - \sigma_{\bar{\nu} N}^{NC}}{\sigma_{\nu N}^{CC} - \sigma_{\bar{\nu} N}^{CC}}\ . \tag{10.15}$$

It was measured by the NuTeV collaboration [82] for the first time, and required a high-intensity and high-energy anti-neutrino beam.

A simple zeroth-order approximation is

$$R_\nu = g_L^2 + g_R^2 r \ , \tag{10.16a}$$

$$R_{\overline{\nu}} = g_L^2 + \frac{g_R^2}{r} \ , \tag{10.16b}$$

$$R^- = g_L^2 - g_R^2 \ , \tag{10.16c}$$

where

$$g_L^2 \equiv \epsilon_L(u)^2 + \epsilon_L(d)^2 \approx \frac{1}{2} - \sin^2\theta_W + \frac{5}{9}\sin^4\theta_W \ , \tag{10.17a}$$

$$g_R^2 \equiv \epsilon_R(u)^2 + \epsilon_R(d)^2 \approx \frac{5}{9}\sin^4\theta_W \ , \tag{10.17b}$$

and $r \equiv \sigma^{CC}_{\overline{\nu}N}/\sigma^{CC}_{\nu N}$ is the ratio of $\overline{\nu}$ and ν charged-current cross-sections, which can be measured directly. (In the simple parton model, ignoring hadron energy cuts, $r \approx (\frac{1}{3}+\epsilon)/(1+\frac{1}{3}\epsilon)$, where $\epsilon \sim 0.125$ is the ratio of the fraction of the nucleon's momentum carried by anti-quarks to that carried by quarks.) In practice, Eq. (10.16) must be corrected for quark mixing, quark sea effects, c-quark threshold effects, non-isoscalarity, W–Z propagator differences, the finite muon mass, QED and electroweak radiative corrections. Details of the neutrino spectra, experimental cuts, x and Q^2 dependence of structure functions, and longitudinal structure functions enter only at the level of these corrections and therefore lead to very small uncertainties. The CCFR group quotes $s^2_W = 0.2236 \pm 0.0041$ for $(m_t, M_H) = (175, 150)$ GeV with very little sensitivity to (m_t, M_H).

The NuTeV collaboration finds $s^2_W = 0.2277 \pm 0.0016$ (for the same reference values) which is 3.0 σ higher than the SM prediction. The discrepancy is in the left-handed coupling, $g_L^2 = 0.3000 \pm 0.0014$, which is 2.7 σ low, while $g_R^2 = 0.0308 \pm 0.0011$ is 0.6 σ high. Within the SM, we can identify four categories of effects that could cause or contribute to this effect [83]. (i) An asymmetric strange sea [84] by itself is an unlikely explanation, but if this asymmetry takes a positive value it would reduce the discrepancy. A preliminary analysis of dimuon data [85] in the relevant kinematic regime, however, indicates a negative strange asymmetry [86]. On the other hand, Ref. 87 finds the opposite sign, at least in its best fit solution. The two analyzes are not directly comparable, however, since the NuTeV Collaboration [85] used a next-to-leading order fit, but with only a subset of the data. In addition, NuTeV does not constrain its parton distribution functions (PDFs) to yield vanishing net strangeness for the proton. Ref. 87 is a leading-order fit to world data including the net strangeness constraint. (ii) Another possibility is that the PDFs violate isospin symmetry at levels much stronger than generally expected [88]. A minimum χ^2 set of PDFs generalized in this sense [89] shows a reduction in the NuTeV discrepancy in s^2_W by 0.0015. But isospin symmetry violating PDFs are currently not well constrained and within uncertainties the NuTeV anomaly could be accounted for in full or conversely made larger [89]. (iii) Nuclear physics effects by themselves appear too small to explain the NuTeV anomaly [90]. In particular, while nuclear shadowing corrections are likely to affect the interpretation of the NuTeV

result [91] at some level, the NuTeV Collaboration argues that their data are dominated by values of Q^2 at which nuclear shadowing is expected to be relatively small. The model of Ref. 92 indicates that nuclear shadowing effects differ for CC and NC cross-sections as well as ν and $\bar{\nu}$ (both would affect the extraction of s^2_W), but also that $R_{\overline{\nu}}$ is affected more than R_ν, while the anomaly is in the latter. Overall, the model predicts a shift in s^2_W by about 0.001 with a sign corresponding to a reduction of the discrepancy. (iv) The extracted s^2_W may also shift at the level of the quoted uncertainty when analyzed using the most recent set of QED and electroweak radiative corrections [93,94], as well as QCD corrections to the structure functions [95]. However, their precise impact can be estimated only after the NuTeV data have been analyzed with a new set of PDFs including these new radiative corrections while simultaneously allowing isospin breaking and asymmetric strange seas. A step in this direction was taken in Ref. 96 in which QED induced isospin violations were shown to reduce the NuTeV discrepancy by 10–20%. Remaining one- and two-loop radiative corrections have been estimated [94] to induce uncertainties in the extracted s^2_W of ± 0.0004 and ± 0.0003, respectively. In view of these developments and caveats, we consider the NuTeV result and the other neutrino deep inelastic scattering (DIS) data as preliminary until a re-analysis using PDFs including all experimental and theoretical information has been completed. It is well conceivable that various effects add up to bring the NuTeV result in line with the SM prediction. It is likely that the overall uncertainties in g^2_L and g^2_R will increase, but at the same time the older neutrino DIS results may become more precise when analyzed with better PDFs than were available at the time.

The cross-section in the laboratory system for $\nu_\mu e \to \nu_\mu e$ or $\overline{\nu}_\mu e \to \overline{\nu}_\mu e$ elastic scattering is

$$\frac{d\sigma_{\nu_\mu,\overline{\nu}_\mu}}{dy} = \frac{G_F^2 m_e E_\nu}{2\pi}$$

$$\times\Bigg[(g_V^{\nu e} \pm g_A^{\nu e})^2 + (g_V^{\nu e} \mp g_A^{\nu e})^2 (1-y)^2$$

$$-(g_V^{\nu e 2} - g_A^{\nu e 2})\frac{y\, m_e}{E_\nu}\Bigg] , \qquad (10.18)$$

where the upper (lower) sign refers to $\nu_\mu(\overline{\nu}_\mu)$, and $y \equiv T_e/E_\nu$ (which runs from 0 to $(1 + m_e/2E_\nu)^{-1}$) is the ratio of the kinetic energy of the recoil electron to the incident ν or $\overline{\nu}$ energy. For $E_\nu \gg m_e$ this yields a total cross-section

$$\sigma = \frac{G_F^2\, m_e\, E_\nu}{2\pi}\left[(g_V^{\nu e} \pm g_A^{\nu e})^2 + \frac{1}{3}(g_V^{\nu e} \mp g_A^{\nu e})^2\right] . \qquad (10.19)$$

The most accurate leptonic measurements [97–100] of $\sin^2\theta_W$ are from the ratio $R \equiv \sigma_{\nu_\mu e}/\sigma_{\overline{\nu}_\mu e}$ in which many of the systematic uncertainties cancel. Radiative corrections (other than m_t effects) are small compared to the precision of present experiments and have negligible effect on the extracted $\sin^2\theta_W$. The most precise experiment (CHARM II) [99] determined not only $\sin^2\theta_W$ but $g^{\nu e}_{V,A}$ as well. The

cross-sections for ν_e-e and $\overline{\nu}_e$-e may be obtained from Eq. (10.18) by replacing $g_{V,A}^{\nu e}$ by $g_{V,A}^{\nu e} + 1$, where the 1 is due to the charged-current contribution [100,101].

The SLAC polarized-electron experiment [102] measured the parity-violating asymmetry

$$A = \frac{\sigma_R - \sigma_L}{\sigma_R + \sigma_L} \ , \tag{10.20}$$

where $\sigma_{R,L}$ is the cross-section for the deep-inelastic scattering of a right- or left-handed electron: $e_{R,L}N \to e\mathrm{X}$. In the quark parton model

$$\frac{A}{Q^2} = a_1 + a_2 \, \frac{1-(1-y)^2}{1+(1-y)^2} \ , \tag{10.21}$$

where $Q^2 > 0$ is the momentum transfer and y is the fractional energy transfer from the electron to the hadrons. For the deuteron or other isoscalar targets, one has, neglecting the s-quark and anti-quarks,

$$a_1 = \frac{3G_F}{5\sqrt{2}\pi\alpha}\left(C_{1u} - \frac{1}{2}C_{1d}\right) \approx \frac{3G_F}{5\sqrt{2}\pi\alpha}\left(-\frac{3}{4} + \frac{5}{3}\sin^2\theta_W\right) \ , \tag{10.22a}$$

$$a_2 = \frac{3G_F}{5\sqrt{2}\pi\alpha}\left(C_{2u} - \frac{1}{2}C_{2d}\right) \approx \frac{9G_F}{5\sqrt{2}\pi\alpha}\left(\sin^2\theta_W - \frac{1}{4}\right) \ . \tag{10.22b}$$

In another polarized-electron scattering experiment on deuterons, but in the quasi-elastic kinematic regime, the SAMPLE experiment [103] at MIT-Bates extracted the combination $C_{2u} - C_{2d}$ at Q^2 values of 0.1 GeV^2 and 0.038 GeV^2. What was actually determined were nucleon form factors from which the quoted results were obtained by the removal of a multi-quark radiative correction. Other linear combinations of the C_{iq} have been determined in polarized-lepton scattering at CERN in μ-C DIS, at Mainz in e-Be (quasi-elastic), and at Bates in e-C (elastic). See the review articles in Refs. [47,102] for more details.

There are now precise experiments measuring atomic parity violation (APV) [104] in cesium [105,106] (at the 0.4% level [105]) , thallium [107], lead [108], and bismuth [109]. The uncertainties associated with atomic wave functions are quite small for cesium [110], and have been reduced recently to about 0.4%. In the past, the semi-empirical value of the tensor polarizability added another source of theoretical uncertainty [111]. The ratio of the off-diagonal hyperfine amplitude to the polarizability has now been measured directly by the Boulder group [112]. Combined with the precisely known hyperfine amplitude [113] one finds excellent agreement with the earlier results, reducing the overall theory uncertainty to only 0.5% (while slightly increasing the experimental error). An earlier 2.3 σ deviation from the SM (see the year 2000 edition of this *Review*) is now seen at the 1 σ level, after the contributions from the Breit interaction have been reevaluated [114], and after the subsequent inclusion of other large and previously underestimated effects [115] (*e.g.*, from QED radiative corrections), and an update of the SM calculation [116] resulted in a vanishing net effect. The theoretical uncertainties

are 3% for thallium [117] but larger for the other atoms. For heavy atoms one determines the "weak charge"

$$\begin{aligned} Q_W &= -2\left[C_{1u}\,(2Z+N) + C_{1d}(Z+2N)\right] \\ &\approx Z(1-4\sin^2\theta_W) - N\ . \end{aligned} \tag{10.23}$$

The recent Boulder experiment in cesium also observed the parity-violating weak corrections to the nuclear electromagnetic vertex (the anapole moment [118]) .

In the future it could be possible to reduce the theoretical wave function uncertainties by taking the ratios of parity violation in different isotopes [104,119]. There would still be some residual uncertainties from differences in the neutron charge radii, however [120].

The forward-backward asymmetry for $e^+e^- \to \ell^+\ell^-$, $\ell = \mu$ or τ, is defined as

$$A_{FB} \equiv \frac{\sigma_F - \sigma_B}{\sigma_F + \sigma_B}\ , \tag{10.24}$$

where $\sigma_F(\sigma_B)$ is the cross-section for ℓ^- to travel forward (backward) with respect to the e^- direction. A_{FB} and R, the total cross-section relative to pure QED, are given by

$$R = F_1\ , \tag{10.25}$$

$$A_{FB} = 3F_2/4F_1\ , \tag{10.26}$$

where

$$F_1 = 1 - 2\chi_0\, g_V^e\, g_V^\ell\, \cos\delta_R + \chi_0^2\left(g_V^{e2} + g_A^{e2}\right)\left(g_V^{\ell 2} + g_A^{\ell 2}\right), \tag{10.27a}$$

$$F_2 = -2\chi_0\, g_A^e\, g_A^\ell\, \cos\delta_R + 4\chi_0^2\, g_A^e\, g_A^\ell\, g_V^e\, g_V^\ell\ , \tag{10.27b}$$

$$\tan\delta_R = \frac{M_Z\Gamma_Z}{M_Z^2 - s}\ , \tag{10.28}$$

$$\chi_0 = \frac{G_F}{2\sqrt{2}\pi\alpha}\,\frac{sM_Z^2}{\left[(M_Z^2-s)^2 + M_Z^2\Gamma_Z^2\right]^{1/2}}\ , \tag{10.29}$$

and $\sqrt{s}$ is the CM energy. Eq. (10.27) is valid at tree level. If the data are radiatively corrected for QED effects (as described above), then the remaining electroweak corrections can be incorporated [121,122] (in an approximation adequate for existing PEP, PETRA, and TRISTAN data, which are well below the Z-pole) by replacing χ_0 by $\chi(s) \equiv (1+\rho_t)\chi_0(s)\alpha/\alpha(s)$, where $\alpha(s)$ is the running QED coupling, and evaluating g_V in the $\overline{\rm MS}$ scheme. Reviews and formulae for $e^+e^- \to$ hadrons may be found in Ref. 123.

At LEP and SLC, there were high-precision measurements of various Z-pole observables [45,124–130], as summarized in Table 10.5. These include the Z-mass and total width, Γ_Z, and partial widths $\Gamma(f\overline{f})$ for $Z \to f\overline{f}$ where fermion $f = e$, μ, τ, hadrons, b, or c. It is convenient to use the variables M_Z, Γ_Z, $R_{\ell_i} \equiv \Gamma({\rm had})/\Gamma(\ell_i^+\ell_i^-)$ ($\ell_i = e, \mu, \tau$), $\sigma_{\rm had} \equiv 12\pi\Gamma(e^+e^-)\Gamma({\rm had})/M_Z^2\,\Gamma_Z^2$, $R_b \equiv \Gamma(b\overline{b})/\Gamma({\rm had})$, and $R_c \equiv \Gamma(c\overline{c})/\Gamma({\rm had})$, most of which are weakly correlated experimentally. ($\Gamma({\rm had})$ is the partial width into hadrons.)

The three values for R_{ℓ_i} are not inconsistent with lepton universality (although R_τ is somewhat low), but we use the general analysis in which the three observables are treated as independent. Similar remarks apply to A_{FB}^{0,ℓ_i} below ($A_{FB}^{0,\tau}$ is somewhat high). $\mathcal{O}(\alpha^3)$ QED corrections introduce a large anti-correlation (-30%) between Γ_Z and $\sigma_{\rm had}$. The anti-correlation between R_b and R_c is -18% [45]. The R_{ℓ_i} are insensitive to m_t except for the $Z \to b\overline{b}$ vertex and final state corrections and the implicit dependence through $\sin^2\theta_W$. Thus, they are especially useful for constraining α_s. The width for invisible decays [45], $\Gamma({\rm inv}) = \Gamma_Z - 3\Gamma(\ell^+\ell^-) - \Gamma({\rm had}) = 499.0 \pm 1.5$ MeV, can be used to determine the number of neutrino flavors much lighter than $M_Z/2$, $N_\nu = \Gamma({\rm inv})/\Gamma^{\rm theory}(\nu\overline{\nu}) = 2.984 \pm 0.009$ for $(m_t, M_H) = (172.7, 117)$ GeV.

There were also measurements of various Z-pole asymmetries. These include the polarization or left-right asymmetry

$$A_{LR} \equiv \frac{\sigma_L - \sigma_R}{\sigma_L + \sigma_R} \ , \tag{10.30}$$

where $\sigma_L(\sigma_R)$ is the cross-section for a left-(right-)handed incident electron. A_{LR} was measured precisely by the SLD collaboration at the SLC [125], and has the advantages of being extremely sensitive to $\sin^2\theta_W$ and that systematic uncertainties largely cancel. In addition, the SLD collaboration extracted the final-state couplings A_b, A_c [45], A_s [126], A_τ, and A_μ [127] from left-right forward-backward asymmetries, using

$$A_{LR}^{FB}(f) = \frac{\sigma_{LF}^f - \sigma_{LB}^f - \sigma_{RF}^f + \sigma_{RB}^f}{\sigma_{LF}^f + \sigma_{LB}^f + \sigma_{RF}^f + \sigma_{RB}^f} = \frac{3}{4} A_f \ , \tag{10.31}$$

where, for example, σ_{LF} is the cross-section for a left-handed incident electron to produce a fermion f traveling in the forward hemisphere. Similarly, A_τ was measured at LEP [45] through the negative total τ polarization, $\mathcal{P}_\tau$, and A_e was extracted from the angular distribution of $\mathcal{P}_\tau$. An equation such as (10.31) assumes that initial state QED corrections, photon exchange, γ–Z interference, the tiny electroweak boxes, and corrections for $\sqrt{s} \neq M_Z$ are removed from the data, leaving the pure electroweak asymmetries. This allows the use of effective tree-level expressions,

$$A_{LR} = A_e P_e \ , \tag{10.32}$$

$$A_{FB} = \frac{3}{4} A_f \, \frac{A_e + P_e}{1 + P_e A_e} \ , \tag{10.33}$$

where

$$A_f \equiv \frac{2\overline{g}_V^f \, \overline{g}_A^f}{\overline{g}_V^{f2} + \overline{g}_A^{f2}} \ , \tag{10.34}$$

and

$$\overline{g}_V^f = \sqrt{\rho_f} \, (t_{3L}^{(f)} - 2 q_f \kappa_f \sin^2\theta_W) \ , \tag{10.34b}$$

$$\overline{g}_A^f = \sqrt{\rho_f} \, t_{3L}^{(f)} \ . \tag{10.34c}$$

P_e is the initial e^- polarization, so that the second equality in Eq. (10.31) is reproduced for $P_e = 1$, and the Z-pole forward-backward asymmetries at LEP ($P_e = 0$) are given by $A_{FB}^{(0,f)} = \frac{3}{4} A_e A_f$ where $f = e,\ \mu,\ \tau,\ b,\ c,\ s$ [128], and q, and where $A_{FB}^{(0,q)}$ refers to the hadronic charge asymmetry. Corrections for t-channel exchange and s/t-channel interference cause $A_{FB}^{(0,e)}$ to be strongly anti-correlated with R_e (-37%). The correlation between $A_{FB}^{(0,b)}$ and $A_{FB}^{(0,c)}$ amounts to 15%. The initial state coupling, A_e, was also determined through the left-right charge asymmetry [129] and in polarized Bhabba scattering at the SLC [127]. The forward-backward asymmetry, A_{FB}, for e^+e^- final states in $p\bar{p}$ collisions has been measured by CDF [131] and a value for $\overline{s}_\ell^2$ has been extracted. By varying the invariant mass and the scattering angle (and assuming the electron couplings), the effective Z couplings to light quarks, $\overline{g}_{V,A}^{u,d}$, resulted, as well, but with large uncertainties and mutual correlations. A similar analysis has also been reported by the H1 Collaboration at HERA [132].

The electroweak radiative corrections have been absorbed into corrections $\rho_f - 1$ and $\kappa_f - 1$, which depend on the fermion f and on the renormalization scheme. In the on-shell scheme, the quadratic m_t dependence is given by $\rho_f \sim 1 + \rho_t$, $\kappa_f \sim 1 + \rho_t / \tan^2\theta_W$, while in $\overline{\rm MS}$, $\widehat{\rho}_f \sim \widehat{\kappa}_f \sim 1$, for $f \neq b$ ($\widehat{\rho}_b \sim 1 - \frac{4}{3}\rho_t$, $\widehat{\kappa}_b \sim 1 + \frac{2}{3}\rho_t$). In the $\overline{\rm MS}$ scheme the normalization is changed according to $G_F M_Z^2/2\sqrt{2}\pi \to \widehat{\alpha}/4\widehat{s}_Z^2\widehat{c}_Z^2$. (If one continues to normalize amplitudes by $G_F M_Z^2/2\sqrt{2}\pi$, as in the 1996 edition of this *Review*, then $\widehat{\rho}_f$ contains an additional factor of $\widehat{\rho}$.) In practice, additional bosonic and fermionic loops, vertex corrections, leading higher order contributions, *etc.*, must be included. For example, in the $\overline{\rm MS}$ scheme one has $\widehat{\rho}_\ell = 0.9981$, $\widehat{\kappa}_\ell = 1.0013$, $\widehat{\rho}_b = 0.9870$, and $\widehat{\kappa}_b = 1.0067$. It is convenient to define an effective angle $\overline{s}_f^2 \equiv \sin^2\overline{\theta}_{Wf} \equiv \widehat{\kappa}_f \widehat{s}_Z^2 = \kappa_f s_W^2$, in terms of which $\overline{g}_V^f$ and $\overline{g}_A^f$ are given by $\sqrt{\rho_f}$ times their tree-level formulae. Because $\overline{g}_V^\ell$ is very small, not only $A_{LR}^0 = A_e$, $A_{FB}^{(0,\ell)}$, and $\mathcal{P}_\tau$, but also $A_{FB}^{(0,b)}$, $A_{FB}^{(0,c)}$, $A_{FB}^{(0,s)}$, and the hadronic asymmetries are mainly sensitive to $\overline{s}_\ell^2$. One finds that $\widehat{\kappa}_f$ ($f \neq b$) is almost independent of (m_t, M_H), so that one can write

$$\overline{s}_\ell^2 \sim \widehat{s}_Z^2 + 0.00029\ . \tag{10.35}$$

Thus, the asymmetries determine values of $\overline{s}_\ell^2$ and $\widehat{s}_Z^2$ almost independent of m_t, while the κ's for the other schemes are m_t dependent.

LEP 2 [45,130,133] ran at several energies above the Z-pole up to ~ 209 GeV. Measurements were made of a number of observables, including the cross-sections for $e^+e^- \to f\bar{f}$ for $f = q, \mu^-, \tau^-$; the differential cross-sections and A_{FB} for μ and τ; R and A_{FB} for b and c; W branching ratios; and WW, $WW\gamma$, ZZ, single W, and single Z cross-sections. They are in agreement with the SM predictions, with the exceptions of the total hadronic cross-section (1.7 σ high), R_b (2.1 σ low), and $A_{FB}(b)$ (1.6 σ low). Also, the SM Higgs boson was excluded below a mass of 114.4 GeV at the 95% CL [134].

The Z-boson properties are extracted assuming the SM expressions for the γ–Z interference terms. These have also been tested experimentally by performing more general fits [130,135] to the LEP 1 and LEP 2 data. Assuming family universality

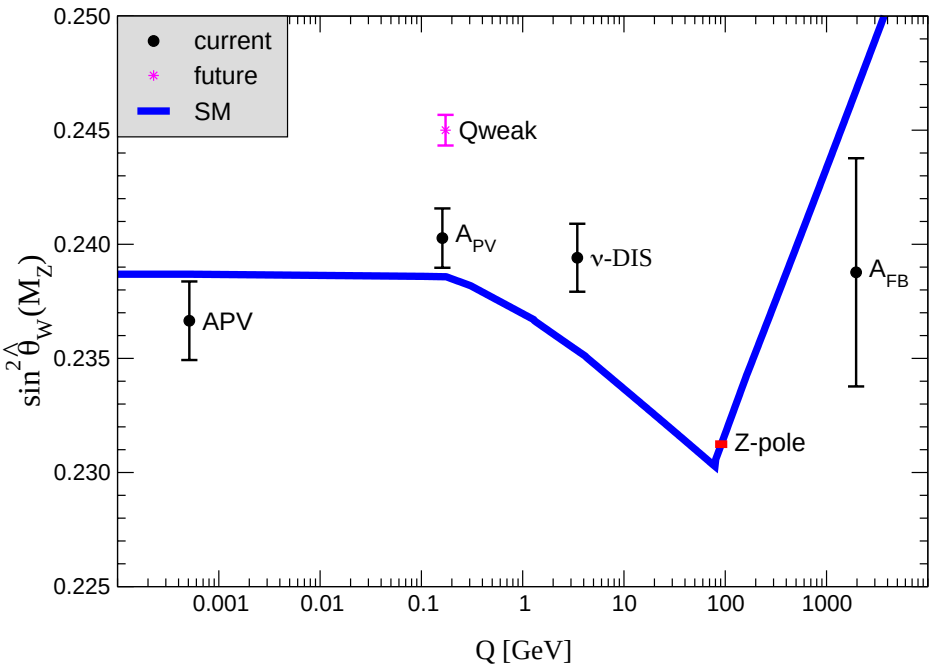


Figure 10.1: Scale dependence of the weak mixing angle defined in the $\overline{\rm MS}$ scheme [137]. The minimum of the curve corresponds to $Q = M_W$, below which we switch to an effective theory with the $W^\pm$ bosons integrated out, and where the β-function for the weak mixing angle changes sign. At the location of the W-boson mass and each fermion mass, there are also discontinuities arising from scheme dependent matching terms which are necessary to ensure that the various effective field theories within a given loop order describe the same physics. However, in the $\overline{\rm MS}$ scheme these are very small numerically and barely visible in the figure provided one decouples quarks at $Q = \widehat{m}_q(\widehat{m}_q)$. The width of the curve reflects the SM uncertainty which is strongly dominated by the experimental error on $\widehat{s}^2_Z$. The theory uncertainty from strong interaction effects is at the level of $\pm 7 \times 10^{-5}$ [137]. See full-color version on color pages at end of book.

this approach introduces three additional parameters relative to the standard fit [45], describing the γ–Z interference contribution to the total hadronic and leptonic cross-sections, $j^{\rm tot}_{\rm had}$ and $j^{\rm tot}_\ell$, and to the leptonic forward-backward asymmetry, $j^{\rm fb}_\ell$. For example,

$$j^{\rm tot}_{\rm had} \sim g^\ell_V g^{\rm had}_V = 0.277 \pm 0.065 \ , \tag{10.36}$$

which is in good agreement with the SM expectation [45] of 0.21 ± 0.01. Similarly, LEP data up to CM energies of 206 GeV are used to constrain the γ–Z interference terms for the heavy quarks. The results for $j^{\rm tot}_b$, $j^{\rm fb}_b$, $j^{\rm tot}_c$, and $j^{\rm fb}_c$ were found in perfect agreement with the SM. These are valuable tests of the SM; but it should be cautioned that new physics is not expected to be described by this set of parameters, since (i) they do not account for extra interactions beyond the standard weak neutral-current, and (ii) the photonic amplitude remains fixed to its SM value.

Strong constraints on anomalous triple and quartic gauge couplings have been obtained at LEP 2 and at the Tevatron, as are described in the Particle Listings.

The parity violating left-right asymmetry, A_{PV}, in fixed target polarized Møller scattering, $e^-e^- \to e^-e^-$, is defined as in Eq. (10.30) but with the opposite sign. It has been measured at low $Q^2 = 0.026$ GeV^2 in the SLAC E158 experiment [136], with the result $A_{PV} = -1.31 \pm 0.14(\text{stat.}) \pm 0.10(\text{syst.}) \times 10^{-7}$. Expressed in terms of the weak mixing angle in the $\overline{\text{MS}}$ scheme, this yields $\widehat{s}^{\,2}(Q^2) = 0.2403 \pm 0.0013$, and established the running of the weak mixing (see Fig. 10.1) at the level of 6.4 standard deviations. In a similar experiment and at about the same Q^2, Qweak at Jefferson Lab [138] will be able to measure $\sin^2\theta_W$ in polarized ep scattering with a relative precision of 0.3%. These experiments will provide the most precise determinations of the weak mixing angle off the Z-peak and will be sensitive to various types of physics beyond the SM.

10.4. W and Z decays

The partial decay width for gauge bosons to decay into massless fermions $f_1\overline{f}_2$ (the numerical values include the small electroweak radiative corrections and final state mass effects) is

$$\Gamma(W^+ \to e^+\nu_e) = \frac{G_F M_W^3}{6\sqrt{2}\pi} \approx 226.29 \pm 0.16 \text{ MeV} \quad , \tag{10.44a}$$

$$\Gamma(W^+ \to u_i\overline{d}_j) = \frac{C G_F M_W^3}{6\sqrt{2}\pi}\,|V_{ij}|^2 \approx (706.24 \pm 0.49)\,|V_{ij}|^2 \text{ MeV} \quad , \tag{10.44b}$$

$$\Gamma(Z \to \psi_i\overline{\psi}_i) = \frac{C G_F M_Z^3}{6\sqrt{2}\pi}\left[g_V^{i2} + g_A^{i2}\right] \tag{10.44c}$$

$$\approx \begin{cases} 300.18 \pm 0.14 \text{ MeV } (u\overline{u}), & 167.21 \pm 0.05 \text{ MeV } (\nu\overline{\nu}), \\ 382.97 \pm 0.14 \text{ MeV } (d\overline{d}), & 83.99 \pm 0.03 \text{ MeV } (e^+e^-), \\ 375.95 \mp 0.10 \text{ MeV } (b\overline{b}). \end{cases}$$

For leptons $C = 1$, while for quarks $C = 3\Big(1 + \alpha_s(M_V)/\pi + 1.409\alpha_s^2/\pi^2 - 12.77\alpha_s^3/\pi^3\Big)$, where the 3 is due to color and the factor in parentheses represents the universal part of the QCD corrections [139] for massless quarks [140]. We also included the leading $\mathcal{O}(\alpha_s^4)$ contribution to hadronic Z decays [141]. The $Z \to f\overline{f}$ widths contain a number of additional corrections: universal (non-singlet) top quark mass contributions [142]; fermion mass effects and further QCD corrections proportional to $\widehat{m}_q^2(M_Z^2)$ [143] which are different for vector and axial-vector partial widths; and singlet contributions starting from two-loop order which are large, strongly top quark mass dependent, family universal, and flavor non-universal [144]. All QCD effects are known and included up to three-loop order. The QED factor $1 + 3\alpha q_f^2/4\pi$, as well as two-loop order $\alpha\alpha_s$ and α^2 self-energy corrections [145] are also included. Working in the on-shell scheme, *i.e.*, expressing the widths in terms of $G_F M_{W,Z}^3$, incorporates the largest radiative corrections from the running QED coupling [49,146]. Electroweak corrections to the

Z-widths are then incorporated by replacing $g^{i2}_{V,A}$ by $\overline{g}^{i2}_{V,A}$. Hence, in the on-shell scheme the Z-widths are proportional to $\rho_i \sim 1 + \rho_t$. The $\overline{\rm MS}$ normalization accounts also for the leading electroweak corrections [54]. There is additional (negative) quadratic m_t dependence in the $Z \to b\overline{b}$ vertex corrections [147] which causes $\Gamma(b\overline{b})$ to decrease with m_t. The dominant effect is to multiply $\Gamma(b\overline{b})$ by the vertex correction $1 + \delta\rho_{b\overline{b}}$, where $\delta\rho_{b\overline{b}} \sim 10^{-2}(-\frac{1}{2}\frac{m_t^2}{M_Z^2} + \frac{1}{5})$. In practice, the corrections are included in ρ_b and κ_b, as discussed before.

For 3 fermion families the total widths are predicted to be

$$\Gamma_Z \approx 2.4956 \pm 0.0007 \text{ GeV} \quad , \tag{10.45}$$

$$\Gamma_W \approx 2.0910 \pm 0.0015 \text{ GeV} \quad . \tag{10.46}$$

We have assumed $\alpha_s(M_Z) = 0.1200$. An uncertainty in α_s of ± 0.0017 introduces an additional uncertainty of 0.05% in the hadronic widths, corresponding to ± 0.8 MeV in Γ_Z. These predictions are to be compared with the experimental results $\Gamma_Z = 2.4952 \pm 0.0023$ GeV [45] and $\Gamma_W = 2.138 \pm 0.044$ GeV (see the Particle Listings for more details).

10.5. Precision flavor physics

In addition to cross-sections, asymmetries, parity violation, W and Z decays, there are a large number of experiments and observables testing the flavor structure of the SM. These are addressed elsewhere in this *Review*, and generally not included in this Section. However, we identify three precision observables with sensitivity to similar types of new physics as the other processes discussed here. The branching fraction of the flavor changing transition $b \to s\gamma$ is of comparatively low precision, but since it is a loop-level process (in the SM) its sensitivity to new physics (and SM parameters, such as heavy quark masses) is enhanced. The τ-lepton lifetime and leptonic branching ratios are primarily sensitive to α_s and not affected significantly by many types of new physics. However, having an independent and reliable low-energy measurement of α_s in a global analysis allows the comparison with the Z-lineshape determination of α_s which shifts easily in the presence of new physics contributions. By far the most precise observable discussed here is the anomalous magnetic moment of the muon (the electron magnetic moment is measured to even greater precision, but its new physics sensitivity is suppressed by terms proportional to m_e^2/M_Z^2). Its combined experimental and theoretical uncertainty is comparable to typical new physics contributions.

The CLEO [148], Belle [149], and BaBar [150] collaborations reported precise measurements of the process $b \to s\gamma$. We extrapolated these results to the full photon spectrum which is defined according to the recommendation in Ref. 151. The results for the branching fractions are then given by,

$$\begin{aligned}
\text{CLEO} &: 3.34 \times 10^{-4}[1 \pm 0.134 \pm 0.076 \pm 0.038 \pm 0.048 \pm 0.006], \\
\text{Belle} &: 3.59 \times 10^{-4}[1 \pm 0.091^{+0.081}_{-0.084} \pm 0.025 \pm 0.020 \pm 0.006], \\
\text{BaBar} &: 4.01 \times 10^{-4}[1 \pm 0.080 \pm 0.091 \pm 0.079 \pm 0.026 \pm 0.006], \\
\text{BaBar} &: 3.57 \times 10^{-4}[1 \pm 0.055^{+0.168}_{-0.122} \pm 0.000 \pm 0.026 \pm 0.000],
\end{aligned}$$

where the first two errors are the statistical and systematic uncertainties (taken uncorrelated). In the case of CLEO, a 3.8% component from the model error of the signal efficiency is moved from the systematic error to the model (third) error. The fourth error accounts for the extrapolation from the finite photon energy cutoff [151–153] (2.0 GeV, 1.815 GeV, and 1.9 GeV, respectively, for CLEO, Belle, and BaBar) to the full theoretical branching ratio. For this we use the results of Ref. 151 for $m_b = 4.70$ GeV which is in good agreement with the more recent Ref. 153. The uncertainty reflects the difference due to choosing $m_b = 4.80$ GeV, instead. The last error is from the correction (0.962 ± 0.006) for the $b \to d\gamma$ component which is common to all inclusive measurements, but absent for the exclusive BaBar measurement in the last line. The last three errors are taken as 100% correlated, resulting in the correlation matrix in Table 10.4. It is advantageous [154] to normalize the result with respect to the semi-leptonic branching fraction, $\mathcal{B}(b \to Xe\nu) = 0.1087 \pm 0.0017$, yielding,

$$R = \frac{\mathcal{B}(b \to s\gamma)}{\mathcal{B}(b \to Xe\nu)} = (3.34 \pm 0.28 \pm 0.37) \times 10^{-3}. \tag{10.47}$$

In the fits we use the variable $\ln R = -5.70 \pm 0.14$ to assure an approximately Gaussian error [155]. The second uncertainty in Eq. (10.47) is an 11% theory uncertainty (excluding parametric errors such as from α_s) in the SM prediction which is based on the next-to-leading order calculations of Refs. 154,156.

Table 10.4: Correlation matrix for measurements of the $b \to s\gamma$ transition.

CLEO	1.000	0.092	0.176	0.048
Belle	0.092	1.000	0.136	0.026
BaBar (inclusive)	0.176	0.136	1.000	0.029
BaBar (exclusive)	0.048	0.026	0.029	1.000

The extraction of α_s from the τ lifetime and leptonic branching ratios is standing out from other determinations, because of a variety of independent reasons: (i) the τ-scale is low, so that upon extrapolation to the Z-scale (where it can be compared to the theoretically clean Z-lineshape determinations) the α_s error shrinks by about an order of magnitude; (ii) yet, this scale is high enough that perturbation theory and the operator product expansion (OPE) can be applied; (iii) these observables are fully inclusive and thus free of fragmentation and hadronization effects that would have to be modeled or measured; (iv) OPE breaking effects are most problematic near the branch cut but there they are suppressed by a double zero at $s = m_\tau^2$; (v) there are enough data [19] to constrain non-perturbative effects both within and breaking the OPE; (vi) a complete three-loop order QCD calculation is available; (vii) large effects associated with the QCD β-function can be resummed [157] (in what has become known as contour

improvement) and these have been computed to even four-loop precision [158]. The largest uncertainty is from the missing perturbative four and higher loop coefficients (appearing in the Adler-D function). The corresponding effects are highly non-linear so that this uncertainty is itself α_s dependent, updated in each call of the fits, and leading to an asymmetric error. The second largest uncertainty is from the missing perturbative five and higher loop coefficients of the QCD β-function; this induces an uncertainty in the contour improvement which is fully correlated with the renormalization group extrapolation from the τ to the Z-scale. The third largest error is from the experimental uncertainty in the lifetime, $\tau_\tau = 290.89 \pm 0.58$ fs, which is from the two leptonic branching ratios and the direct τ_τ. Because of the poor convergence of perturbation theory for strange quark final states, we used for these the experimentally measured branching ratio. Included are also various smaller uncertainties from other sources. In total we obtain a 2% determination of $\alpha_s(M_Z) = 0.1225^{+0.0025}_{-0.0022}$ which updates the result of Ref. 5. For more details, see Ref. 19 where even 1–1.5% uncertainties are advocated (mainly by means of additional assumptions regarding the perturbative four-loop error).

The world average of the muon anomalous magnetic moment*,

$$a_\mu^{\rm exp} = \frac{g_\mu - 2}{2} = (1165920.80 \pm 0.63) \times 10^{-9} \ , \qquad (10.48)$$

is dominated by the 1999, 2000, and 2001 data runs of the E821 collaboration at BNL [159]. The QED contribution has been calculated to four loops [160] (fully analytically to three loops [161,162]) , and the leading logarithms are included to five loops [163,164]. The estimated SM electroweak contribution [165–167], $a_\mu^{\rm EW} = (1.52 \pm 0.03) \times 10^{-9}$, which includes leading two-loop [166] and three-loop [167] corrections, is at the level of the current uncertainty. The limiting factor in the interpretation of the result is the uncertainty from the two-loop hadronic contribution [20], $a_\mu^{\rm had} = (69.54 \pm 0.64) \times 10^{-9}$, which has been obtained using $e^+e^- \to$ hadrons cross-section data (including the KLOE data from radiative returns from the Φ resonance [42] and the very recent SND data [39]) . The latter are dominated by the (reanalyzed) CMD 2 data [17]. This value suggests a 2.3 σ discrepancy between Eq. (10.48) and the SM prediction. In an alternative analysis, the authors of Ref. 18 used τ decay data and isospin symmetry (CVC) to obtain $a_\mu^{\rm had} = (71.10 \pm 0.58) \times 10^{-9}$. This result implies no conflict (0.7 σ) with Eq. (10.48). Thus, there is also a discrepancy between the 2π and 4π spectral functions obtained from the two methods. For example, if one uses the e^+e^- data and CVC to predict the branching ratio for $\tau^- \to \nu_\tau \pi^- \pi^0$ decays one obtains $24.52 \pm 0.31\%$ [20] (this

* In what follows, we summarize the most important aspects of $g_\mu - 2$, and give some details about the evaluation in our fits. For more details see the dedicated contribution by A. Höcker and W. Marciano in this *Review*. There are some small numerical differences (at the level of 0.1 standard deviation), which are well understood and mostly arise because internal consistency of the fits requires the calculation of all observables from analytical expressions and common inputs and fit parameters, so that an independent evaluation is necessary for this Section. Note, that in the spirit of a global analysis based on all available information we have chosen here to average in the τ-decay data, as well.

does not include the SND data) while the average of the measured branching ratios by DELPHI [168], ALEPH, CLEO, L3, and OPAL [18] yields $25.43 \pm 0.09\%$, which is 2.8 σ higher. It is important to understand the origin of this difference, but four observations point to the conclusion that at least some of it is experimental: (i) Including the SND data in the e^+e^- data set (which are consistent with the implications of the τ decay data), this discrepancy decreases to about 2.4 σ (in particular, the KLOE and SND results differ both qualitatively and quantitatively), and would decrease further if the older data are discarded. (ii) The $\tau^- \to \nu_\tau 2\pi^-\pi^+\pi^0$ spectral function also disagrees with the corresponding e^+e^- data at the 4 σ level, which translates to a 23% effect [20] and seems too large to arise from isospin violation. (iii) Isospin violating corrections have been studied in detail in Ref. 169 and found to be largely under control. The largest effect is due to higher-order electroweak corrections [43] but introduces a negligible uncertainty [170]. (iv) Ref. 171 shows on the basis of a QCD sum rule that the spectral functions derived from τ decay data are consistent with values of $\alpha_s(M_Z) \gtrsim 0.120$, in agreement with what we find from the global fit in Sec. 10.6, while the spectral functions from e^+e^- annihilation are consistent only for somewhat lower (disfavored) values. Nevertheless, $a_\mu^{\rm had}$ has been evaluated in Refs. 36,172 excluding the τ decay data with results which are generally in good agreement with each other and other e^+e^- based analyzes. It is argued [172] that CVC breaking effects (*e.g.*, through a relatively large mass difference between the $\rho^\pm$ and ρ^0 vector mesons) may be larger than expected. (This may also be relevant in the context of the NuTeV discrepancy discussed above [172].) Experimentally [19], this mass difference is indeed larger than expected, but then one would also expect a significant width difference which is contrary to observation [19]. Fortunately, due to the suppression at large s (from where the conflicts originate) these problems are less pronounced as far as $a_\mu^{\rm had}$ is concerned. In the following we view all differences in spectral functions as fluctuations and average the results. An additional uncertainty is induced by the hadronic three-loop light-by-light scattering contribution. We use the most recent value [173], $a_\mu^{\rm LBLS} = (+1.36 \pm 0.25) \times 10^{-9}$, which is higher than previous evaluations [174,175]. The sign of this effect is opposite [174] to the one quoted in the 2002 edition of this *Review*, and has subsequently been confirmed by two other groups [175]. Other hadronic effects at three-loop order contribute [176], $a_\mu^{\rm had}\left[\left(\frac{\alpha}{\pi}\right)^3\right] = (-1.00 \pm 0.06) \times 10^{-9}$. Correlations with the two-loop hadronic contribution and with $\Delta\alpha(M_Z)$ (see Sec. 10.2) were considered in Ref. 162, which also contains analytic results for the perturbative QCD contribution. The SM prediction is

$$a_\mu^{\rm theory} = (1165919.52 \pm 0.52) \times 10^{-9} \ , \tag{10.49}$$

where the error is from the hadronic uncertainties excluding parametric ones such as from α_s and the heavy quark masses. We estimate its correlation with $\Delta\alpha(M_Z)$ as 24%. The small overall discrepancy between the experimental and theoretical values could be due to fluctuations or underestimates of the theoretical uncertainties. On the other hand, $g_\mu - 2$ is also affected by many types of new physics, such as supersymmetric models with large $\tan\beta$ and moderately light superparticle masses [177]. Thus, the deviation could also

arise from physics beyond the SM.

Table 10.5: Principal Z-pole and other observables, compared with the SM best fit predictions (see text). The LEP averages of the ALEPH, DELPHI, L3, and OPAL results include common systematic errors and correlations [45]. The heavy flavor results of LEP and SLD are based on common inputs and correlated, as well [45]. The first $\overline{s}^2_\ell(A_{FB}^{(0,q)})$ is the effective angle extracted from the hadronic charge asymmetry, which has some (neglected) correlation with $A_{FB}^{(0,b)}$; the second $\overline{s}^2_\ell(A_{FB}^{(0,q)})$ is from the lepton asymmetry from CDF [131]. The values of $\Gamma(\ell^+\ell^-)$, Γ(had), and Γ(inv) are not independent of Γ_Z, the R_ℓ, and $\sigma_{\rm had}$. The first M_W value is from UA2, CDF, and DØ [178], and based on the two-parameter analysis of Ref. 179; the second one is from LEP 2 [180]. The first M_W and M_Z are correlated, but the effect is negligible due to the tiny M_Z error. The three values of A_e are (i) from A_{LR} for hadronic final states [125]; (ii) from A_{LR} for leptonic final states and from polarized Bhabba scattering [127]; and (iii) from the angular distribution of the τ polarization. The two A_τ values are from SLD and the total τ polarization, respectively. g_L^2 and g_R^2 are from NuTeV [82] and have a very small (-1.7%) residual anti-correlation. The older deep-inelastic scattering (DIS) results from CDHS [75], CHARM [76], and CCFR [77] are included, as well, but not shown in the Table. The world averages for $g^{\nu e}_{V,A}$ are dominated by the CHARM II [99] results, $g_V^{\nu e} = -0.035 \pm 0.017$ and $g_A^{\nu e} = -0.503 \pm 0.017$. A_{PV} is the parity violating asymmetry in Møller scattering. The errors in Q_W, DIS, $b \to s\gamma$, and $g_\mu - 2$ are the total (experimental plus theoretical) uncertainties. The τ_τ value is the τ lifetime world average computed by combining the direct measurements with values derived from the leptonic branching ratios [5]; the theory uncertainty is included in the SM prediction. In all other SM predictions, the uncertainty is from M_Z, M_H, m_t, m_b, m_c, $\widehat{\alpha}(M_Z)$, and α_s, and their correlations have been accounted for. The SM errors in Γ_Z, Γ(had), R_ℓ, and $\sigma_{\rm had}$ are largely dominated by the uncertainty in α_s.

Quantity	Value	Standard Model	Pull
m_t [GeV]	$172.7 \pm 2.9 \pm 0.6$	172.7 ± 2.8	0.0
M_W [GeV]	80.450 ± 0.058	80.376 ± 0.017	1.3
	80.392 ± 0.039		0.4
M_Z [GeV]	91.1876 ± 0.0021	91.1874 ± 0.0021	0.1
Γ_Z [GeV]	2.4952 ± 0.0023	2.4968 ± 0.0011	-0.7
Γ(had) [GeV]	1.7444 ± 0.0020	1.7434 ± 0.0010	—
Γ(inv) [MeV]	499.0 ± 1.5	501.65 ± 0.11	—
$\Gamma(\ell^+\ell^-)$ [MeV]	83.984 ± 0.086	83.996 ± 0.021	—
$\sigma_{\rm had}$ [nb]	41.541 ± 0.037	41.467 ± 0.009	2.0
R_e	20.804 ± 0.050	20.756 ± 0.011	1.0
R_μ	20.785 ± 0.033	20.756 ± 0.011	0.9
R_τ	20.764 ± 0.045	20.801 ± 0.011	-0.8
R_b	0.21629 ± 0.00066	0.21578 ± 0.00010	0.8
R_c	0.1721 ± 0.0030	0.17230 ± 0.00004	-0.1
$A_{FB}^{(0,e)}$	0.0145 ± 0.0025	0.01622 ± 0.00025	-0.7
$A_{FB}^{(0,\mu)}$	0.0169 ± 0.0013		0.5
$A_{FB}^{(0,\tau)}$	0.0188 ± 0.0017		1.5
$A_{FB}^{(0,b)}$	0.0992 ± 0.0016	0.1031 ± 0.0008	-2.4
$A_{FB}^{(0,c)}$	0.0707 ± 0.0035	0.0737 ± 0.0006	-0.8
$A_{FB}^{(0,s)}$	0.0976 ± 0.0114	0.1032 ± 0.0008	-0.5
$\bar{s}_\ell^2(A_{FB}^{(0,q)})$	0.2324 ± 0.0012	0.23152 ± 0.00014	0.7
	0.2238 ± 0.0050		-1.5
A_e	0.15138 ± 0.00216	0.1471 ± 0.0011	2.0
	0.1544 ± 0.0060		1.2
	0.1498 ± 0.0049		0.6
A_μ	0.142 ± 0.015		-0.3
A_τ	0.136 ± 0.015		-0.7
	0.1439 ± 0.0043		-0.7
A_b	0.923 ± 0.020	0.9347 ± 0.0001	-0.6
A_c	0.670 ± 0.027	0.6678 ± 0.0005	0.1
A_s	0.895 ± 0.091	0.9356 ± 0.0001	-0.4
g_L^2	0.30005 ± 0.00137	0.30378 ± 0.00021	-2.7
g_R^2	0.03076 ± 0.00110	0.03006 ± 0.00003	0.6
$g_V^{\nu e}$	-0.040 ± 0.015	-0.0396 ± 0.0003	0.0
$g_A^{\nu e}$	-0.507 ± 0.014	-0.5064 ± 0.0001	0.0
A_{PV}	-1.31 ± 0.17	-1.53 ± 0.02	1.3
Q_W(Cs)	-72.62 ± 0.46	-73.17 ± 0.03	1.2
Q_W(Tl)	-116.6 ± 3.7	-116.78 ± 0.05	0.1
$\frac{\Gamma(b\to s\gamma)}{\Gamma(b\to Xe\nu)}$	$3.35^{+0.50}_{-0.44} \times 10^{-3}$	$(3.22 \pm 0.09) \times 10^{-3}$	0.3
$\frac{1}{2}(g_\mu - 2 - \frac{\alpha}{\pi})$	4511.07 ± 0.82	4509.82 ± 0.10	1.5
τ_τ [fs]	290.89 ± 0.58	291.87 ± 1.76	-0.4

10.6. Experimental results

The values of the principal Z-pole observables are listed in Table 10.5, along with the SM predictions for $M_Z = 91.1874 \pm 0.0021$ GeV, $M_H = 89^{+38}_{-28}$ GeV, $m_t = 172.7 \pm 2.8$ GeV, $\alpha_s(M_Z) = 0.1216 \pm 0.0017$, and $\widehat{\alpha}(M_Z)^{-1} = 127.904 \pm 0.019$ ($\Delta\alpha^{(5)}_{\rm had} \approx 0.02802 \pm 0.00015$). The predictions result from a global least-square (χ^2) fit to all data using the minimization package MINUIT [181] and the electroweak library GAPP [73]. In most cases, we treat all input errors (the uncertainties of the values) as Gaussian. The reason is not that we assume that theoretical and systematic errors are intrinsically bell-shaped (which they are not) but because in most cases the input errors are combinations of many different (including statistical) error sources, which should yield approximately Gaussian *combined* errors by the large number theorem. Thus, it suffices if either the statistical components dominate or there are many components of similar size. An exception is the theory dominated error on the τ lifetime, which we recalculate in each χ^2-function call since it depends itself on α_s yielding an asymmetric (and thus non-Gaussian) error bar. Sizes and shapes of the output errors (the uncertainties of the predictions and the SM fit parameters) are fully determined by the fit, and 1σ errors are defined to correspond to $\Delta\chi^2 = \chi^2 - \chi^2_{\rm min} = 1$, and do not necessarily correspond to the 68.3% probability range or the 39.3% probability contour (for 2 parameters).

Table 10.6: Principal SM fit result including mutual correlations.

M_Z[GeV]	91.1874 ± 0.0021	1.00	−0.02	0.00	0.00	−0.01	0.00	0.08
m_t[GeV]	172.7 ± 2.8	−0.02	1.00	0.00	0.00	−0.03	−0.02	0.61
$\widehat{m}_b(\widehat{m}_b)$[GeV]	4.207 ± 0.031	0.00	0.00	1.00	0.29	−0.03	0.01	0.05
$\widehat{m}_c(\widehat{m}_c)$[GeV]	$1.290^{+0.040}_{-0.045}$	0.00	0.00	0.29	1.00	0.09	0.03	0.14
$\alpha_s(M_Z)$	0.1216 ± 0.0017	−0.01	−0.03	−0.03	0.09	1.00	−0.01	−0.02
$\Delta\alpha^{(3)}_{\rm had}(1.8\ {\rm GeV})$	0.00581 ± 0.00010	0.00	−0.02	0.01	0.03	−0.01	1.00	−0.18
M_H[GeV]	89^{+38}_{-28} GeV	0.08	0.61	0.05	0.14	−0.02	−0.18	1.00

The values and predictions of m_t [6–8]; M_W [178–180]; deep inelastic [82], ν_μ-e [97–99], and polarized Møller scattering [136]; the Q_W for cesium [105,106] and thallium [107]; the $b \to s\gamma$ observable [148–150]; the muon anomalous magnetic moment [159]; and the τ lifetime are also listed in Table 10.5. The values of M_W and m_t differ from those in the Particle Listings because they include recent preliminary results. The agreement is excellent. Despite the discrepancies discussed in the following, the goodness of the fit to all data is very good with a χ^2/d.o.f. = 47.5/42. The probability of a larger χ^2 is 26%. Only g^2_L from NuTeV and $A^{(0,b)}_{FB}$ from LEP are currently showing large (2.7 σ and 2.4 σ) deviations. In addition, the hadronic peak cross-section, $\sigma_{\rm had}$ (LEP), and

the A^0_{LR} (SLD) from hadronic final states differ by 2.0 σ. The final result for $g_\mu - 2$ from BNL has moved up, and so has the SM prediction due to the higher value of the light-by-light contribution [173], so that the small net deviation (1.5 σ, see Sec. 10.5) is basically unchanged compared to the 2004 edition of this *Review*. Observables like $R_b = \Gamma(b\overline{b})/\Gamma(\text{had})$, $R_c = \Gamma(c\overline{c})/\Gamma(\text{had})$, and the combined value for M_W which showed significant deviations in the past, are now in reasonable agreement. In particular, R_b, whose measured value deviated by as much as 3.7 σ from the SM prediction, is now in agreement.

A_b can be extracted from $A_{FB}^{(0,b)}$ when $A_e = 0.1501 \pm 0.0016$ is taken from a fit to leptonic asymmetries (using lepton universality). The result, $A_b = 0.881 \pm 0.017$, is 3.1 σ below the SM prediction†, and also 1.6 σ below $A_b = 0.923 \pm 0.020$ obtained from $A_{LR}^{FB}(b)$ at SLD. Thus, it appears that at least some of the problem in $A_{FB}^{(0,b)}$ is experimental. Note, however, that the uncertainty in $A_{FB}^{(0,b)}$ is strongly statistics dominated. The combined value, $A_b = 0.899 \pm 0.013$ deviates by 2.8 σ. It would be extremely difficult to account for this 3.9% deviation by new physics radiative corrections since about a 20% correction to $\widehat{\kappa}_b$ would be necessary to account for the central value of A_b. If this deviation is due to new physics, it is most likely of tree-level type affecting preferentially the third generation. Examples include the decay of a scalar neutrino resonance [182], mixing of the b quark with heavy exotics [183], and a heavy Z' with family-nonuniversal couplings [184]. It is difficult, however, to simultaneously account for R_b, which has been measured on the Z-peak and off-peak [185] at LEP 1. An average of R_b measurements at LEP 2 at energies between 133 and 207 GeV is 2.1 σ below the SM prediction, while $A_{FB}^{(b)}(\text{LEP 2})$ is 1.6 σ low [133].

The left-right asymmetry, $A^0_{LR} = 0.15138 \pm 0.00216$ [125], based on all hadronic data from 1992–1998 differs 2.0 σ from the SM expectation of 0.1471 ± 0.0011. The combined value of $A_\ell = 0.1513 \pm 0.0021$ from SLD (using lepton-family universality and including correlations) is also 2.0 σ above the SM prediction; but there is now experimental agreement between this SLD value and the LEP value, $A_\ell = 0.1481 \pm 0.0027$, obtained from a fit to $A_{FB}^{(0,\ell)}$, $A_e(\mathcal{P}_\tau)$, and $A_\tau(\mathcal{P}_\tau)$, again assuming universality.

The observables in Table 10.5, as well as some other less precise observables, are used in the global fits described below. The correlations on the LEP lineshape and τ polarization, the LEP/SLD heavy flavor observables, the SLD lepton asymmetries, and the deep inelastic and ν-e scattering observables, are included. The theoretical correlations between $\Delta\alpha^{(5)}_{\rm had}$ and $g_\mu - 2$, and between the charm and bottom quark masses, are also accounted for.

The data allow a simultaneous determination of M_H, m_t, $\sin^2\theta_W$, and the strong coupling $\alpha_s(M_Z)$. ($\widehat{m}_c$, $\widehat{m}_b$, and $\Delta\alpha^{(5)}_{\rm had}$ are also allowed to float in the fits, subject to the theoretical constraints [5,16] described in Sec. 10.1–Sec. 10.2. These are correlated with

† Alternatively, one can use $A_\ell = 0.1481 \pm 0.0027$, which is from LEP alone and in excellent agreement with the SM, and obtain $A_b = 0.893 \pm 0.022$ which is 1.9 σ low. This illustrates that some of the discrepancy is related to the one in A_{LR}.

α_s.) α_s is determined mainly from R_ℓ, Γ_Z, $\sigma_{\rm had}$, and τ_τ and is only weakly correlated with the other variables (except for a 9% correlation with $\widehat{m}_c$). The global fit to all data, including the CDF/DØ average, $m_t = 172.7 \pm 3.0$ GeV, yields

$$\begin{aligned} M_H &= 89^{+38}_{-28}\ \text{GeV}\ , \\ m_t &= 172.7 \pm 2.8\ \text{GeV}\ , \\ \widehat{s}^{\,2}_Z &= 0.23122 \pm 0.00015\ , \\ \alpha_s(M_Z) &= 0.1216 \pm 0.0017\ . \end{aligned} \tag{10.50}$$

The complete fit result including the correlation matrix is given in Table 10.6.

In the on-shell scheme one has $s^2_W = 0.22306 \pm 0.00033$, the larger error due to the stronger sensitivity to m_t, while the corresponding effective angle is related by Eq. (10.35), *i.e.*, $\overline{s}^2_\ell = 0.23152 \pm 0.00014$. The m_t pole mass corresponds to $\widehat{m}_t(\widehat{m}_t) = 162.7 \pm 2.7$ GeV. In all fits, the errors include full statistical, systematic, and theoretical uncertainties. The $\widehat{s}^{\,2}_Z$ ($\overline{s}^2_\ell$) error reflects the error on $\overline{s}^2_\ell = 0.23152 \pm 0.00016$ from a fit to the Z-pole asymmetries (including the CDF lepton asymmetry [131]) .

As described at the beginning of Sec. 10.2 and the last paragraph of Sec. 10.5, there is some spread in the experimental e^+e^- spectral functions and also some stress when these are compared with τ-decay spectral functions. These are below or above the 2σ level (depending on what is actually compared) but not larger than the deviations of some other quantities entering our analyzes. The number and size or these deviations are well consistent with what one would expect to happen as a result of random fluctuations. It is nevertheless instructive to study the effect of doubling the uncertainty in $\Delta\alpha^{(3)}_{\rm had}(1.8\ \text{GeV}) = 0.00577 \pm 0.00010$ (see the beginning of Sec. 10.2) on the extracted Higgs mass. The result, $M_H = 87^{+39}_{-29}$ GeV, demonstrates that the uncertainty in $\Delta\alpha_{\rm had}$ is currently of only secondary importance. Note also, that the uncertainty of ± 0.0001 in $\Delta\alpha^{(3)}_{\rm had}(1.8\ \text{GeV})$ corresponds to a shift of ∓ 6 GeV in M_H or less than one fifth of its total uncertainty.

The weak mixing angle can be determined from Z-pole observables, M_W, and from a variety of neutral-current processes spanning a very wide Q^2 range. The results (for the older low-energy neutral-current data see [46,47]) shown in Table 10.7 are in reasonable agreement with each other, indicating the quantitative success of the SM. The largest discrepancy is the value $\widehat{s}^{\,2}_Z = 0.2355 \pm 0.0016$ from DIS which is 2.7 σ above the value 0.23122 ± 0.00015 from the global fit to all data. Similarly, $\widehat{s}^{\,2}_Z = 0.23193 \pm 0.00028$ from the forward-backward asymmetries into bottom and charm quarks, and $\widehat{s}^{\,2}_Z = 0.23067 \pm 0.00029$ from the SLD asymmetries (both when combined with M_Z) are 2.5 σ high and 1.9 σ low, respectively.

The extracted Z-pole value of $\alpha_s(M_Z)$ is based on a formula with negligible theoretical uncertainty (± 0.0005 in $\alpha_s(M_Z)$) if one assumes the exact validity of the SM. One should keep in mind, however, that this value, $\alpha_s = 0.1198 \pm 0.0028$, is very sensitive to such types of new physics as non-universal vertex corrections. In contrast, the value derived from τ decays, $\alpha_s(M_Z) = 0.1225^{+0.0025}_{-0.0022}$, is theory dominated but less sensitive to new

Table 10.7: Values of $\widehat{s}^2_Z$, s^2_W, α_s, and M_H [in GeV] for various (combinations of) observables. Unless indicated otherwise, the top quark mass, $m_t = 172.7 \pm 3.0$ GeV, is used as an additional constraint in the fits. The (†) symbol indicates a fixed parameter.

Data	$\widehat{s}^2_Z$	s^2_W	$\alpha_s(M_Z)$	M_H
All data	0.23122(15)	0.22306(33)	0.1216(17)	89^{+38}_{-28}
All indirect (no m_t)	0.23122(16)	0.22307(41)	0.1216(17)	87^{+107}_{-43}
Z pole (no m_t)	0.23121(17)	0.22310(59)	0.1198(28)	89^{+112}_{-44}
LEP 1 (no m_t)	0.23152(21)	0.22375(67)	0.1213(30)	168^{+232}_{-91}
SLD + M_Z	0.23067(29)	0.22203(56)	0.1216 (†)	28^{+26}_{-16}
$A_{FB}^{(b,c)}$ + M_Z	0.23193(28)	0.22480(76)	0.1216 (†)	349^{+250}_{-148}
M_W + M_Z	0.23089(38)	0.22241(74)	0.1216 (†)	47^{+52}_{-31}
M_Z	0.23134(11)	0.22334(36)	0.1216 (†)	117 (†)
polarized Møller	0.2330(14)	0.2251(14)	0.1216 (†)	117 (†)
DIS (isoscalar)	0.2355(16)	0.2275(16)	0.1216 (†)	117 (†)
Q_W (APV)	0.2290(19)	0.2210(19)	0.1216 (†)	117 (†)
elastic $\nu_\mu(\overline{\nu_\mu})e$	0.2310(77)	0.2230(77)	0.1216 (†)	117 (†)
SLAC eD	0.222(18)	0.213(19)	0.1216 (†)	117 (†)
elastic $\nu_\mu(\overline{\nu_\mu})p$	0.211(33)	0.203(33)	0.1216 (†)	117 (†)

physics. The two values are in remarkable agreement with each other. They are also in perfect agreement with other recent values, such as from jet-event shapes at LEP [186] (0.1202 ± 0.0050) and HERA [187] (0.1186 ± 0.0051), but the τ decay result is somewhat higher than the value, 0.1170 ± 0.0012, from the most recent unquenched lattice calculation of Ref. 188. For more details and other determinations, see our Section 9 on "Quantum Chromodynamics" in this *Review*.

The data indicate a preference for a small Higgs mass. There is a strong correlation between the quadratic m_t and logarithmic M_H terms in $\widehat{\rho}$ in all of the indirect data except for the $Z \to b\overline{b}$ vertex. Therefore, observables (other than R_b) which favor m_t values higher than the Tevatron range favor lower values of M_H. This effect is enhanced by R_b, which has little direct M_H dependence but favors the lower end of the Tevatron m_t range. M_W has additional M_H dependence through $\Delta\widehat{r}_W$ which is not coupled to m_t^2 effects. The strongest individual pulls toward smaller M_H are from M_W and A^0_{LR}, while $A_{FB}^{(0,b)}$ and the NuTeV results favor high values. The difference in χ^2 for the global fit is

$\Delta\chi^2 = \chi^2(M_H = 1000\ \text{GeV}) - \chi^2_{\rm min} = 60$. Hence, the data favor a small value of M_H, as in supersymmetric extensions of the SM. The central value of the global fit result, $M_H = 89^{+38}_{-28}$ GeV, is below the direct lower bound, $M_H \geq 114.4$ GeV (95% CL) [134].

The 90% central confidence range from all precision data is

$$46\ \text{GeV} \leq M_H \leq 154\ \text{GeV}\ .$$

Including the results of the direct searches as an extra contribution to the likelihood function drives the 95% upper limit to $M_H \leq 189$ GeV. As two further refinements, we account for (i) theoretical uncertainties from uncalculated higher order contributions by allowing the T parameter (see next subsection) subject to the constraint $T = 0 \pm 0.02$, (ii) the M_H dependence of the correlation matrix which gives slightly more weight to lower Higgs masses [189]. The resulting limits at 95 (90, 99)% CL are

$$M_H \leq 194\ (176,\ 235)\ \text{GeV}\ ,$$

respectively. The extraction of M_H from the precision data depends strongly on the value used for $\alpha(M_Z)$. Upper limits, however, are more robust due to two compensating effects: the older results indicated more QED running and were less precise, yielding M_H distributions which were broader with centers shifted to smaller values. The hadronic contribution to $\alpha(M_Z)$ is correlated with $g_\mu - 2$ (see Sec. 10.5). The measurement of the latter is higher than the SM prediction, and its inclusion in the fit favors a larger $\alpha(M_Z)$ and a lower M_H (by 3 GeV).

One can also carry out a fit to the indirect data alone, *i.e.*, without including the constraint, $m_t = 172.7 \pm 3.0$ GeV, obtained by CDF and DØ. (The indirect prediction is for the $\overline{\rm MS}$ mass, $\widehat{m}_t(\widehat{m}_t) = 162.4^{+9.6}_{-7.2}$ GeV, which is in the end converted to the pole mass). One obtains $m_t = 172.3^{+10.2}_{-\ 7.6}$ GeV, with almost no change in the $\sin^2\theta_W$ and α_s values, in perfect agreement with the direct CDF/DØ average. The relations between M_H and m_t for various observables are shown in Fig. 10.2.

Using $\alpha(M_Z)$ and $\widehat{s}^2_Z$ as inputs, one can predict $\alpha_s(M_Z)$ assuming grand unification. One predicts [190] $\alpha_s(M_Z) = 0.130 \pm 0.001 \pm 0.01$ for the simplest theories based on the minimal supersymmetric extension of the SM, where the first (second) uncertainty is from the inputs (thresholds). This is slightly larger, but consistent with the experimental $\alpha_s(M_Z) = 0.1216 \pm 0.0017$ from the Z-lineshape and the τ lifetime, as well as with other determinations. Non-supersymmetric unified theories predict the low value $\alpha_s(M_Z) = 0.073 \pm 0.001 \pm 0.001$. See also the note on "Low-Energy Supersymmetry" in the Particle Listings.

One can also determine the radiative correction parameters Δr: from the global fit one obtains $\Delta r = 0.0355 \pm 0.0010$ and $\Delta\widehat{r}_W = 0.06959 \pm 0.00029$. M_W measurements [178–180] (when combined with M_Z) are equivalent to measurements of $\Delta r = 0.0335 \pm 0.0020$, which is 1.2 σ below the result from all indirect data, $\Delta r = 0.0362 \pm 0.0012$. Fig. 10.3 shows the 1 σ contours in the $M_W - m_t$ plane from the direct and indirect determinations, as well as the combined 90% CL region. The indirect determination uses M_Z from LEP 1 as input, which is defined assuming an s dependent decay width. M_W then corresponds to the s

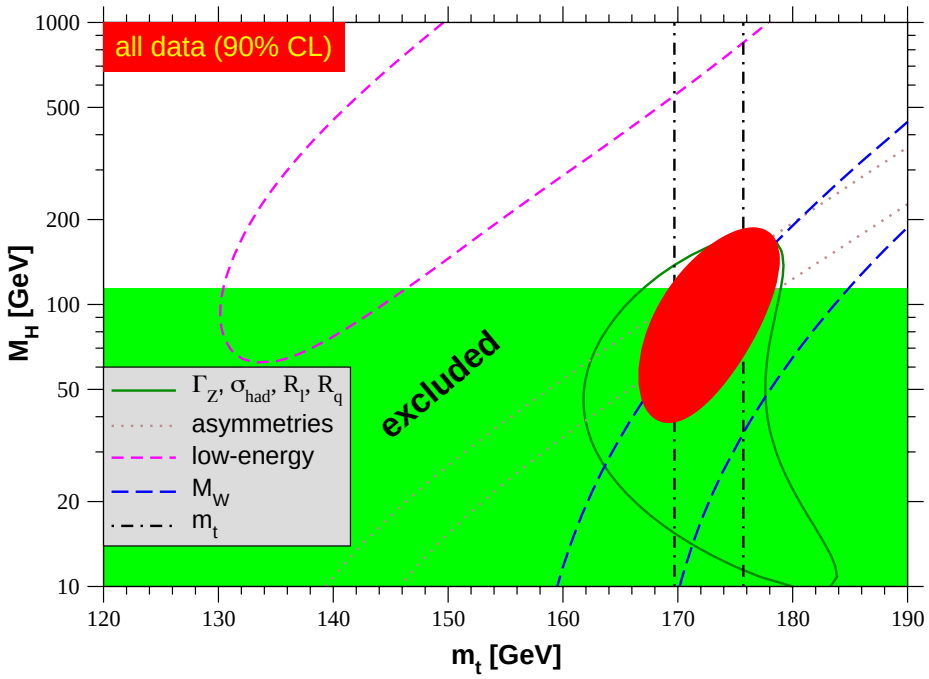


Figure 10.2: One-standard-deviation (39.35%) uncertainties in M_H as a function of m_t for various inputs, and the 90% CL region ($\Delta\chi^2 = 4.605$) allowed by all data. $\alpha_s(M_Z) = 0.120$ is assumed except for the fits including the Z-lineshape data. The 95% direct lower limit from LEP 2 is also shown. See full-color version on color pages at end of book.

dependent width definition, as well, and can be directly compared with the results from the Tevatron and LEP 2 which have been obtained using the same definition. The difference to a constant width definition is formally only of $\mathcal{O}(\alpha^2)$, but is strongly enhanced since the decay channels add up coherently. It is about 34 MeV for M_Z and 27 MeV for M_W. The residual difference between working consistently with one or the other definition is about 3 MeV, *i.e.*, of typical size for non-enhanced $\mathcal{O}(\alpha^2)$ corrections [60–62].

Most of the parameters relevant to ν-hadron, ν-e, e-hadron, and e^+e^- processes are determined uniquely and precisely from the data in "model-independent" fits (*i.e.*, fits which allow for an arbitrary electroweak gauge theory). The values for the parameters defined in Eqs. (10.12)–(10.14) are given in Table 10.8 along with the predictions of the SM. The agreement is reasonable, except for the values of g_L^2 and $\epsilon_L(u,d)$, which reflect the discrepancy in the NuTeV results. (The ν-hadron results without the new NuTeV data can be found in the 1998 edition of this *Review*.). The off Z-pole e^+e^- results are difficult to present in a model-independent way because Z-propagator effects are non-negligible at TRISTAN, PETRA, PEP, and LEP 2 energies. However, assuming e-μ-τ universality, the low-energy lepton asymmetries imply [123] $4(g_A^e)^2 = 0.99 \pm 0.05$, in good agreement with the SM prediction $\simeq 1$.

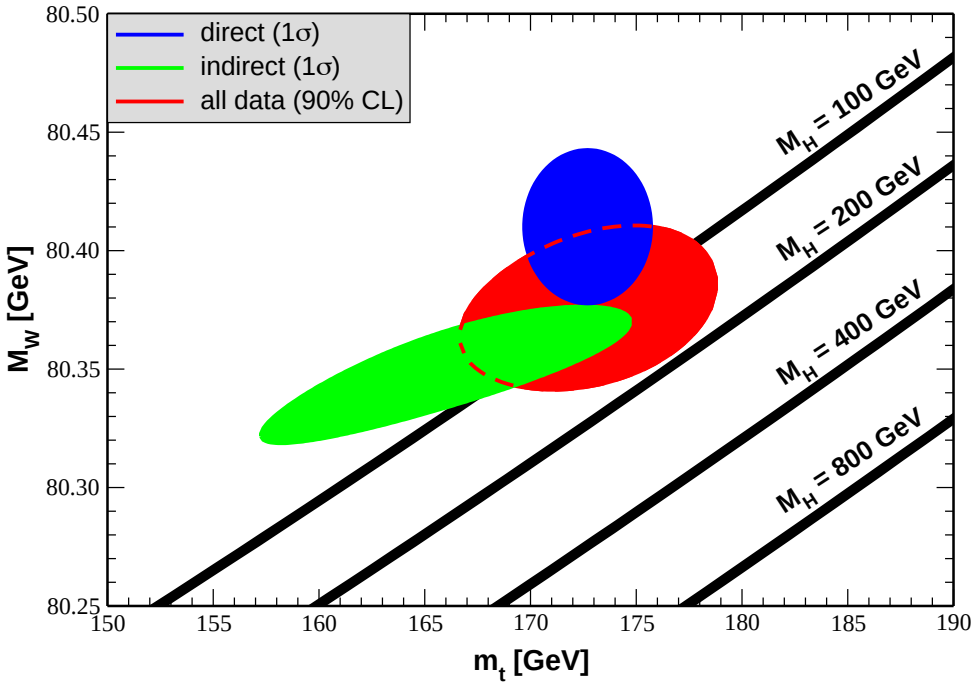


Figure 10.3: One-standard-deviation (39.35%) region in M_W as a function of m_t for the direct and indirect data, and the 90% CL region ($\Delta\chi^2 = 4.605$) allowed by all data. The SM prediction as a function of M_H is also indicated. The widths of the M_H bands reflect the theoretical uncertainty from $\alpha(M_Z)$. See full-color version on color pages at end of book.

10.7. Constraints on new physics

The Z-pole, W-mass, and neutral-current data can be used to search for and set limits on deviations from the SM. In particular, the combination of these indirect data with the direct CDF and DØ average for m_t allows one to set stringent limits on new physics. We will mainly discuss the effects of exotic particles (with heavy masses $M_{\rm new} \gg M_Z$ in an expansion in $M_Z/M_{\rm new}$) on the gauge boson self-energies. (Brief remarks are made on new physics which is not of this type.) Most of the effects on precision measurements can be described by three gauge self-energy parameters S, T, and U. We will define these, as well as related parameters, such as ρ_0, ϵ_i, and $\widehat{\epsilon}_i$, to arise from new physics only. *I.e.*, they are equal to zero ($\rho_0 = 1$) exactly in the SM, and do not include any contributions from m_t or M_H, which are treated separately. Our treatment differs from most of the original papers.

Many extensions of the SM are described by the ρ_0 parameter,

$$\rho_0 \equiv M_W^2/(M_Z^2\,\widehat{c}_Z^{\,2}\,\widehat{\rho}) \ , \tag{10.51}$$

which describes new sources of SU(2) breaking that cannot be accounted for by the SM Higgs doublet or m_t effects. In the presence of $\rho_0 \neq 1$, Eq. (10.51) generalizes Eq. (10.8*b*) while Eq. (10.8*a*) remains unchanged. Provided that the new physics which yields $\rho_0 \neq 1$ is a small perturbation which does not significantly affect the radiative corrections, ρ_0 can be regarded as a phenomenological parameter which

Table 10.8: Values of the model-independent neutral-current parameters, compared with the SM predictions. There is a second $g_{V,A}^{\nu e}$ solution, given approximately by $g_V^{\nu e} \leftrightarrow g_A^{\nu e}$, which is eliminated by e^+e^- data under the assumption that the neutral current is dominated by the exchange of a single Z. The ϵ_L, as well as the ϵ_R, are strongly correlated and non-Gaussian, so that for implementations we recommend the parametrization using g_i^2 and $\theta_i = \tan^{-1}[\epsilon_i(u)/\epsilon_i(d)]$, $i = L$ or R. In the SM predictions, the uncertainty is from M_Z, M_H, m_t, m_b, m_c, $\widehat{\alpha}(M_Z)$, and α_s.

Quantity	Experimental Value	SM	Correlation		
$\epsilon_L(u)$	0.326 ± 0.013	0.3459(1)			
$\epsilon_L(d)$	-0.441 ± 0.010	$-0.4291(1)$		non-	
$\epsilon_R(u)$	$-0.175\ ^{+0.013}_{-0.004}$	$-0.1550(1)$		Gaussian	
$\epsilon_R(d)$	$-0.022\ ^{+0.072}_{-0.047}$	0.0776			
g_L^2	0.3005 ± 0.0012	0.3038(2)	-0.11	-0.21	-0.01
g_R^2	0.0311 ± 0.0010	0.0301		-0.02	-0.03
θ_L	2.51 ± 0.033	2.4631(1)			0.26
θ_R	$4.59\ ^{+0.41}_{-0.28}$	5.1765			
$g_V^{\nu e}$	-0.040 ± 0.015	$-0.0396(3)$			-0.05
$g_A^{\nu e}$	-0.507 ± 0.014	$-0.5064(1)$			
$C_{1u} + C_{1d}$	0.147 ± 0.004	0.1529(1)	0.95	-0.75	-0.10
$C_{1u} - C_{1d}$	-0.604 ± 0.066	$-0.5297(4)$		-0.79	-0.10
$C_{2u} + C_{2d}$	0.72 ± 0.89	-0.0095			-0.11
$C_{2u} - C_{2d}$	-0.071 ± 0.044	$-0.0621(6)$			

multiplies G_F in Eqs. (10.12)–(10.14), (10.29), and Γ_Z in Eq. (10.44). There are enough data to determine ρ_0, M_H, m_t, and α_s, simultaneously. From the global fit,

$$\rho_0 = 1.0002^{+0.0007}_{-0.0004} \ , \tag{10.52}$$

$$114.4 \text{ GeV} \leq M_H \leq 191 \text{ GeV} \ , \tag{10.53}$$

$$m_t = 173.1 \pm 2.9 \text{ GeV} \ , \tag{10.54}$$

$$\alpha_s(M_Z) = 0.1215 \pm 0.0017 \ , \tag{10.55}$$

where the lower limit on M_H is the direct search bound. (If the direct limit is ignored one obtains $M_H = 66^{+85}_{-30}$ GeV and $\rho_0 = 0.9996^{+0.0010}_{-0.0007}$.) The error bar in Eq. (10.52) is highly asymmetric: at the 2 σ level one has $\rho_0 = 1.0002^{+0.0024}_{-0.0009}$ and $M_H \leq 654$ GeV. Clearly, in

the presence of ρ_0 upper limits on M_H become much weaker. The result in Eq. (10.52) is in remarkable agreement with the SM expectation, $\rho_0 = 1$. It can be used to constrain higher-dimensional Higgs representations to have vacuum expectation values of less than a few percent of those of the doublets. Indeed, the relation between M_W and M_Z is modified if there are Higgs multiplets with weak isospin $> 1/2$ with significant vacuum expectation values. In order to calculate to higher orders in such theories one must define a set of four fundamental renormalized parameters which one may conveniently choose to be α, G_F, M_Z, and M_W, since M_W and M_Z are directly measurable. Then $\widehat{s}^2_Z$ and ρ_0 can be considered dependent parameters.

Eq. (10.52) can also be used to constrain other types of new physics. For example, non-degenerate multiplets of heavy fermions or scalars break the vector part of weak SU(2) and lead to a decrease in the value of M_Z/M_W. A non-degenerate SU(2) doublet $\binom{f_1}{f_2}$ yields a positive contribution to ρ_0 [191] of

$$\frac{CG_F}{8\sqrt{2}\pi^2}\Delta m^2 \ , \tag{10.56}$$

where

$$\Delta m^2 \equiv m_1^2 + m_2^2 - \frac{4m_1^2 m_2^2}{m_1^2 - m_2^2}\ln\frac{m_1}{m_2} \geq (m_1 - m_2)^2 \ , \tag{10.57}$$

and $C = 1$ (3) for color singlets (triplets). Thus, in the presence of such multiplets, one has

$$\frac{3G_F}{8\sqrt{2}\pi^2}\sum_i \frac{C_i}{3}\,\Delta m_i^2 = \rho_0 - 1 \ , \tag{10.58}$$

where the sum includes fourth-family quark or lepton doublets, $\binom{t'}{b'}$ or $\binom{E^0}{E^-}$, and scalar doublets such as $\binom{\tilde{t}}{\tilde{b}}$ in Supersymmetry (in the absence of $L-R$ mixing). This implies

$$\sum_i \frac{C_i}{3}\,\Delta m_i^2 \leq (90\ \text{GeV})^2 \tag{10.59}$$

at 95% CL. The corresponding constraints on non-degenerate squark and slepton doublets are even stronger, $\sum_i C_i \Delta m_i^2/3 \leq (64\ \text{GeV})^2$. This is due to the supersymmetric Higgs mass bound, $m_{h^0} < 150$ GeV, and the very strong correlation between m_{h^0} and ρ_0 (84%).

Non-degenerate multiplets usually imply $\rho_0 > 1$. Similarly, heavy Z' bosons decrease the prediction for M_Z due to mixing and generally lead to $\rho_0 > 1$ [192]. On the other hand, additional Higgs doublets which participate in spontaneous symmetry breaking [193], heavy lepton doublets involving Majorana neutrinos [194], and the vacuum expectation values of Higgs triplets or higher-dimensional representations can contribute to ρ_0 with either sign. Allowing for the presence of heavy degenerate chiral multiplets (the S parameter, to be discussed below) affects the determination of ρ_0 from the data, at present leading to a smaller value (for fixed M_H).

A number of authors [195–200] have considered the general effects on neutral-current and Z and W-boson observables of various types of heavy (*i.e.*, $M_{\rm new} \gg M_Z$) physics which contribute to the W and Z self-energies but which do not have any direct coupling to the ordinary fermions. In addition to non-degenerate multiplets, which break the vector part of weak SU(2), these include heavy degenerate multiplets of chiral fermions which break the axial generators. The effects of one degenerate chiral doublet are small, but in Technicolor theories there may be many chiral doublets and therefore significant effects [195].

Such effects can be described by just three parameters, S, T, and U at the (electroweak) one-loop level. (Three additional parameters are needed if the new physics scale is comparable to M_Z [201]. Further generalizations, including effects relevant to LEP 2, are described in Ref. 202.) T is proportional to the difference between the W and Z self-energies at $Q^2 = 0$ (*i.e.*, vector SU(2)-breaking), while S ($S + U$) is associated with the difference between the Z (W) self-energy at $Q^2 = M^2_{Z,W}$ and $Q^2 = 0$ (axial SU(2)-breaking). Denoting the contributions of new physics to the various self-energies by $\Pi^{\rm new}_{ij}$, we have

$$\widehat{\alpha}(M_Z)T \equiv \frac{\Pi^{\rm new}_{WW}(0)}{M^2_W} - \frac{\Pi^{\rm new}_{ZZ}(0)}{M^2_Z} \ , \tag{10.60a}$$

$$\begin{aligned}\frac{\widehat{\alpha}(M_Z)}{4\widehat{s}^2_Z\widehat{c}^2_Z}S \equiv{}& \frac{\Pi^{\rm new}_{ZZ}(M^2_Z) - \Pi^{\rm new}_{ZZ}(0)}{M^2_Z} \\ &- \frac{\widehat{c}^2_Z - \widehat{s}^2_Z}{\widehat{c}_Z\widehat{s}_Z}\ \frac{\Pi^{\rm new}_{Z\gamma}(M^2_Z)}{M^2_Z} - \frac{\Pi^{\rm new}_{\gamma\gamma}(M^2_Z)}{M^2_Z} \ ,\end{aligned} \tag{10.60b}$$

$$\begin{aligned}\frac{\widehat{\alpha}(M_Z)}{4\widehat{s}^2_Z}(S+U) \equiv{}& \frac{\Pi^{\rm new}_{WW}(M^2_W) - \Pi^{\rm new}_{WW}(0)}{M^2_W} \\ &- \frac{\widehat{c}_Z}{\widehat{s}_Z}\ \frac{\Pi^{\rm new}_{Z\gamma}(M^2_Z)}{M^2_Z} - \frac{\Pi^{\rm new}_{\gamma\gamma}(M^2_Z)}{M^2_Z} \ .\end{aligned} \tag{10.60c}$$

S, T, and U are defined with a factor proportional to $\widehat{\alpha}$ removed, so that they are expected to be of order unity in the presence of new physics. In the $\overline{\rm MS}$ scheme as defined in Ref. 52, the last two terms in Eq. (10.60*b*) and Eq. (10.60*c*) can be omitted (as was done in some earlier editions of this *Review*). These three parameters are related to other parameters (S_i, h_i, $\widehat{\epsilon}_i$) defined in Refs. [52,196,197] by

$$\begin{aligned} T &= h_V = \widehat{\epsilon}_1/\alpha \ , \\ S &= h_{AZ} = S_Z = 4\widehat{s}^2_Z\widehat{\epsilon}_3/\alpha \ , \\ U &= h_{AW} - h_{AZ} = S_W - S_Z = -4\widehat{s}^2_Z\widehat{\epsilon}_2/\alpha \ . \end{aligned} \tag{10.61}$$

A heavy non-degenerate multiplet of fermions or scalars contributes positively to T as

$$\rho_0 - 1 = \frac{1}{1-\alpha T} - 1 \simeq \alpha T \ , \tag{10.62}$$

where ρ_0 is given in Eq. (10.58). The effects of non-standard Higgs representations cannot be separated from heavy non-degenerate multiplets unless the new physics has other consequences, such as vertex corrections. Most of the original papers defined T to include the effects of loops only. However, we will redefine T to include all new sources of SU(2) breaking, including non-standard Higgs, so that T and ρ_0 are equivalent by Eq. (10.62).

A multiplet of heavy degenerate chiral fermions yields

$$S = C \sum_i \Big(t_{3L}(i) - t_{3R}(i) \Big)^2 / 3\pi \ , \tag{10.63}$$

where $t_{3L,R}(i)$ is the third component of weak isospin of the left-(right-)handed component of fermion i and C is the number of colors. For example, a heavy degenerate ordinary or mirror family would contribute $2/3\pi$ to S. In Technicolor models with QCD-like dynamics, one expects [195] $S \sim 0.45$ for an iso-doublet of techni-fermions, assuming $N_{TC} = 4$ techni-colors, while $S \sim 1.62$ for a full techni-generation with $N_{TC} = 4$; T is harder to estimate because it is model dependent. In these examples one has $S \geq 0$. However, the QCD-like models are excluded on other grounds (flavor changing neutral-currents, and too-light quarks and pseudo-Goldstone bosons [203]) . In particular, these estimates do not apply to models of walking Technicolor [203], for which S can be smaller or even negative [204]. Other situations in which $S < 0$, such as loops involving scalars or Majorana particles, are also possible [205]. The simplest origin of $S < 0$ would probably be an additional heavy Z' boson [192], which could mimic $S < 0$. Supersymmetric extensions of the SM generally give very small effects. See Refs. 155,206 and the Section on Supersymmetry in this *Review* for a complete set of references.

Most simple types of new physics yield $U = 0$, although there are counter-examples, such as the effects of anomalous triple gauge vertices [197].

The SM expressions for observables are replaced by

$$\begin{aligned} M_Z^2 &= M_{Z0}^2 \ \frac{1 - \alpha T}{1 - G_F M_{Z0}^2 S / 2\sqrt{2}\pi} \ , \\ M_W^2 &= M_{W0}^2 \ \frac{1}{1 - G_F M_{W0}^2 (S+U) / 2\sqrt{2}\pi} \ , \end{aligned} \tag{10.64}$$

where M_{Z0} and M_{W0} are the SM expressions (as functions of m_t and M_H) in the $\overline{\rm MS}$ scheme. Furthermore,

$$\begin{aligned} \Gamma_Z &= \frac{1}{1 - \alpha T} M_Z^3 \beta_Z \ , \\ \Gamma_W &= M_W^3 \beta_W \ , \\ A_i &= \frac{1}{1 - \alpha T} A_{i0} \ , \end{aligned} \tag{10.65}$$

where β_Z and β_W are the SM expressions for the reduced widths Γ_{Z0}/M_{Z0}^3 and Γ_{W0}/M_{W0}^3, M_Z and M_W are the physical masses, and A_i (A_{i0}) is a neutral-current amplitude (in the SM).

The data allow a simultaneous determination of $\widehat{s}^2_Z$ (from the Z-pole asymmetries), S (from M_Z), U (from M_W), T (mainly from Γ_Z), α_s (from R_ℓ, $\sigma_{\rm had}$, and τ_τ), and m_t (from CDF and DØ), with little correlation among the SM parameters:

$$\begin{aligned} S &= -0.13 \pm 0.10\ (-0.08)\ , \\ T &= -0.13 \pm 0.11\ (+0.09)\ , \\ U &= \ \ 0.20 \pm 0.12\ (+0.01)\ , \end{aligned} \tag{10.66}$$

and $\widehat{s}^2_Z = 0.23124 \pm 0.00016$, $\alpha_s(M_Z) = 0.1223 \pm 0.0018$, $m_t = 172.6 \pm 2.9$ GeV, where the uncertainties are from the inputs. The central values assume $M_H = 117$ GeV, and in parentheses we show the change for $M_H = 300$ GeV. As can be seen, the SM parameters (U) can be determined with no (little) M_H dependence. On the other hand, S, T, and M_H cannot be obtained simultaneously, because the Higgs boson loops themselves are resembled approximately by oblique effects. Eqs. (10.66) show that negative (positive) contributions to the S (T) parameter can weaken or entirely remove the strong constraints on M_H from the SM fits. Specific models in which a large M_H is compensated by new physics are reviewed in Ref. 207. The parameters in Eqs. (10.66), which by definition are due to new physics only, all deviate by more than one standard deviation from the SM values of zero. However, these deviations are correlated. Fixing $U = 0$ (as is done in Fig. 10.4) will also move S and T to values compatible with zero within errors,

$$\begin{aligned} S &= -0.07 \pm 0.09\ (-0.07)\ , \\ T &= -0.03 \pm 0.09\ (+0.09)\ . \end{aligned} \tag{10.67}$$

Using Eq. (10.62) the value of ρ_0 corresponding to T is 0.9990 ± 0.0009 (+0.0007), while the one corresponding to Eq. (10.67) is 0.9997 ± 0.0007 (+0.0007). The values of the $\widehat{\epsilon}$ parameters defined in Eq. (10.61) are

$$\begin{aligned} \widehat{\epsilon}_3 &= -0.0011 \pm 0.0008\ (-0.0006)\ , \\ \widehat{\epsilon}_1 &= -0.0010 \pm 0.0009\ (+0.0007)\ , \\ \widehat{\epsilon}_2 &= -0.0017 \pm 0.0010\ (-0.0001)\ . \end{aligned} \tag{10.68}$$

Unlike the original definition, we defined the quantities in Eqs. (10.68) to vanish identically in the absence of new physics and to correspond directly to the parameters S, T, and U in Eqs. (10.66). There is a strong correlation (84%) between the S and T parameters. The allowed region in $S - T$ is shown in Fig. 10.4. From Eqs. (10.66) one obtains $S \leq 0.03\,(-0.04)$ and $T \leq 0.06\,(0.14)$ at 95% CL for $M_H = 117$ GeV (300 GeV). If one fixes $M_H = 600$ GeV and requires the constraint $S \geq 0$ (as is appropriate in QCD-like Technicolor models) then $S \leq 0.09$ (Bayesian) or $S \leq 0.06$ (frequentist). This rules out simple Technicolor models with many techni-doublets and QCD-like dynamics.

An extra generation of ordinary fermions is excluded at the 99.999% CL on the basis of the S parameter alone, corresponding to $N_F = 2.81 \pm 0.24$ for the number of families. This result assumes that there are no new contributions to T or U and therefore that any new families are degenerate. In principle this restriction can be relaxed by allowing

T to vary as well, since $T > 0$ is expected from a non-degenerate extra family. However, the data currently favor $T < 0$, thus strengthening the exclusion limits. A more detailed analysis is required if the extra neutrino (or the extra down-type quark) is close to its direct mass limit [208]. This can drive S to small or even negative values but at the expense of too-large contributions to T. These results are in agreement with a fit to the number of light neutrinos, $N_\nu = 2.986 \pm 0.007$ (which favors a larger value for $\alpha_s(M_Z) = 0.1231 \pm 0.0020$ mainly from R_ℓ and τ_τ). However, the S parameter fits are valid even for a very heavy fourth family neutrino.

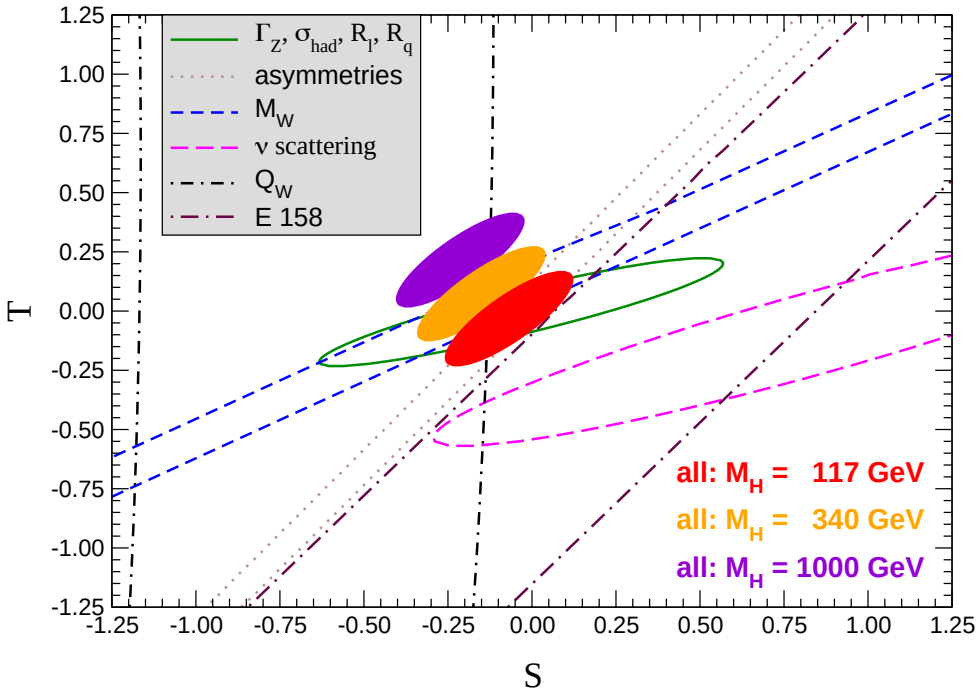


Figure 10.4: 1 σ constraints (39.35 %) on S and T from various inputs combined with M_Z. S and T represent the contributions of new physics only. (Uncertainties from m_t are included in the errors.) The contours assume $M_H = 117$ GeV except for the central and upper 90% CL contours allowed by all data, which are for $M_H = 340$ GeV and 1000 GeV, respectively. Data sets not involving M_W are insensitive to U. Due to higher order effects, however, $U = 0$ has to be assumed in all fits. α_s is constrained using the τ lifetime as additional input in all fits. See full-color version on color pages at end of book.

There is no simple parametrization that is powerful enough to describe the effects of every type of new physics on every possible observable. The S, T, and U formalism describes many types of heavy physics which affect only the gauge self-energies, and it can be applied to all precision observables. However, new physics which couples directly to ordinary fermions, such as heavy Z' bosons [192] or mixing with exotic fermions [209] cannot be fully parametrized in the S, T, and U framework. It is convenient to treat these types of new physics by parameterizations that are specialized to that particular class of

theories (*e.g.*, extra Z' bosons), or to consider specific models (which might contain, *e.g.*, Z' bosons and exotic fermions with correlated parameters). Constraints on various types of new physics are reviewed in Refs. [47,116,210,211].

Fits to Supersymmetric models are described in Refs. 155 and 212. Models involving strong dynamics (such as (extended) Technicolor) for electroweak breaking are considered in Ref. 213. The effects of compactified extra spatial dimensions at the TeV scale are reviewed in Ref. 214, and constraints on Little Higgs models in Ref. 215. Limits on new four-Fermi operators and on leptoquarks using LEP 2 and lower energy data are given in Ref. 130.

An alternate formalism [216] defines parameters, ϵ_1, ϵ_2, ϵ_3, ϵ_b in terms of the specific observables M_W/M_Z, $\Gamma_{\ell\ell}$, $A_{FB}^{(0,\ell)}$, and R_b. The definitions coincide with those for $\widehat{\epsilon}_i$ in Eqs. (10.60) and (10.61) for physics which affects gauge self-energies only, but the ϵ's now parametrize arbitrary types of new physics. However, the ϵ's are not related to other observables unless additional model dependent assumptions are made. Another approach [217–219] parametrizes new physics in terms of gauge-invariant sets of operators. It is especially powerful in studying the effects of new physics on non-Abelian gauge vertices. The most general approach introduces deviation vectors [210]. Each type of new physics defines a deviation vector, the components of which are the deviations of each observable from its SM prediction, normalized to the experimental uncertainty. The length (direction) of the vector represents the strength (type) of new physics.

Table 10.9: 95% CL lower mass limits (in GeV) from low energy and Z pole data on various extra Z' gauge bosons, appearing in models of unification and string theory. (More general parametrizations are described in [224]) . ρ_0 free indicates a completely arbitrary Higgs sector, while $\rho_0 = 1$ restricts to Higgs doublets and singlets with still unspecified charges. The CDF bounds from searches for $\bar{p}p \to e^+e^-, \mu^+\mu^-$ [225] and the LEP 2 $e^+e^- \to f\bar{f}$ [133] bounds are listed in the last two columns, respectively. (The CDF bounds would be weakend if there are open supersymmetric or exotic decay channels [226].)

Z'	ρ_0 free	$\rho_0 = 1$	CDF (direct)	LEP 2
Z_χ	551	545	720	673
Z_ψ	151	146	690	481
Z_η	379	365	715	434
Z_{LR}	570	564	630	804
Z_{SM}	822	809	845	1787
$Z_{\rm string}$	582	578	–	–

One of the best motivated kinds of physics beyond the SM besides Supersymmetry are extra Z' bosons [220]. They do not spoil the observed approximate gauge coupling unification, and appear copiously in many Grand Unified Theories (GUTs), most Superstring models [221], as well as in dynamical symmetry breaking [213,222] and Little

Higgs models [215]. For example, the SO(10) GUT contains an extra U(1) as can be seen from its maximal subgroup, $\mathrm{SU}(5) \times \mathrm{U}(1)_\chi$. Similarly, the E_6 GUT contains the subgroup $\mathrm{SO}(10) \times \mathrm{U}(1)_\psi$. The Z_ψ possesses only axial-vector couplings to the ordinary fermions, and its mass is generally less constrained. The Z_η boson is the linear combination $\sqrt{3/8}\, Z_\chi - \sqrt{5/8}\, Z_\psi$. The Z_{LR} boson occurs in left-right models with gauge group $\mathrm{SU}(3)_C \times \mathrm{SU}(2)_L \times \mathrm{SU}(2)_R \times \mathrm{U}(1)_{B-L} \subset \mathrm{SO}(10)$. The sequential Z_{SM} boson is defined to have the same couplings to fermions as the SM Z-boson. Such a boson is not expected in the context of gauge theories unless it has different couplings to exotic fermions than the ordinary Z. However, it serves as a useful reference case when comparing constraints from various sources. It could also play the role of an excited state of the ordinary Z in models with extra dimensions at the weak scale [214]. Finally, we consider a Superstring motivated Z_{string} boson appearing in a specific model [223]. The potential Z' boson is in general a superposition of the SM Z and the new boson associated with the extra U(1). The mixing angle θ satisfies,

$$\tan^2\theta = \frac{M_{Z_1^0}^2 - M_Z^2}{M_{Z'}^2 - M_{Z_1^0}^2},$$

where $M_{Z_1^0}$ is the SM value for M_Z in the absence of mixing. Note, that $M_Z < M_{Z_1^0}$, and that the SM Z couplings are changed by the mixing. If the Higgs $\mathrm{U}(1)'$ quantum numbers are known, there will be an extra constraint,

$$\theta = C\frac{g_2}{g_1}\frac{M_Z^2}{M_{Z'}^2}\ , \tag{10.69}$$

where $g_{1,2}$ are the U(1) and $\mathrm{U}(1)'$ gauge couplings with $g_2 = \sqrt{\frac{5}{3}}\sin\theta_W\sqrt{\lambda}\, g_1$ and $g_1 = \sqrt{g^2 + g'^2}$. $\lambda \sim 1$ (which we assume) if the GUT group breaks directly to $\mathrm{SU}(3) \times \mathrm{SU}(2) \times \mathrm{U}(1) \times \mathrm{U}(1)'$. C is a function of vacuum expectation values. For minimal Higgs sectors it can be found in Ref. 192. Table 10.9 shows the 95% CL lower mass limits obtained from a somewhat earlier data set [227] for ρ_0 free and $\rho_0 = 1$, respectively. In cases of specific minimal Higgs sectors where C is known, the Z' mass limits are generally pushed into the TeV region. The limits on $|\theta|$ are typically $<$ few $\times 10^{-3}$. For more details see [227,228] and the Section on "The Z' Searches" in this *Review*. Also listed in Table 10.9 are the direct lower limits on Z' production from CDF [225] and LEP 2 bounds [45]. The final LEP 1 value for σ_{had}, some previous values for $Q_W(\mathrm{Cs})$, NuTeV, and $A_{FB}^{0,b}$ (for family-nonuniversal couplings [229]) modify the results and might even suggest the possible existence of a Z' [184,230].

Acknowledgments:

This work was supported in part by CONACyT (México) contract 42026–F, by DGAPA–UNAM contract PAPIIT IN112902, and by the U.S. Department of Energy under Grant No. DOE-EY-76-02-3071.

References:

1. S. Weinberg, Phys. Rev. Lett. **19**, 1264 (1967);
   A. Salam, p. 367 of *Elementary Particle Theory*, ed. N. Svartholm (Almquist and Wiksells, Stockholm, 1969);
   S.L. Glashow, J. Iliopoulos, and L. Maiani, Phys. Rev. **D2**, 1285 (1970).
2. J.L. Rosner, `hep-ph/0410281`;
   CKMfitter Group: J. Charles *et al.*, Eur. Phys. J. **C41**, 1 (2005).
3. For reviews, see G. Barbiellini and C. Santoni, Riv. Nuovo Cimento **9(2)**, 1 (1986);
   E.D. Commins and P.H. Bucksbaum, *Weak Interactions of Leptons and Quarks*, (Cambridge Univ. Press, Cambridge, 1983);
   W. Fetscher and H.J. Gerber, p. 657 of Ref. 4;
   J. Deutsch and P. Quin, p. 706 of Ref. 4;
   J.M. Conrad, M.H. Shaevitz, and T. Bolton, Rev. Mod. Phys. **70**, 1341 (1998).
4. *Precision Tests of the Standard Electroweak Model, ed. P. Langacker (World Scientific, Singapore, 1995).*
5. *J. Erler and M. Luo,* Phys. Lett. **B558**, 125 (2003).
6. *CDF, DØ, and the Tevatron Electroweak Working Group: J.F. Arguin et al.,* `hep-ex/0507091`.
7. CDF: T. Affolder *et al.*, Phys. Rev. **D63**, 032003 (2001).
8. DØ: B. Abbott *et al.*, Phys. Rev. **D60**, 052001 (1999);
   DØ: V.M. Abazov *et al.*, Nature **429**, 638 (2004);
   DØ: V.M. Abazov *et al.*, Phys. Lett. **B606**, 25 (2005).
9. K. Melnikov and T. v. Ritbergen, Phys. Lett. **B482**, 99 (2000).
10. S.J. Brodsky, G.P. Lepage, and P.B. Mackenzie, Phys. Rev. **D28**, 228 (1983).
11. N. Gray *et al.*, Z. Phys. **C48**, 673 (1990).
12. For reviews, see the article on "The Higgs boson" in this *Review*;
    J. Gunion, H.E. Haber, G.L. Kane, and S. Dawson, *The Higgs Hunter's Guide*, (Addison-Wesley, Redwood City, 1990);
    M. Sher, Phys. Reports **179**, 273 (1989);
    M. Carena and H.E. Haber, Prog. Part. Nucl. Phys. **50**, 63 (2003);
    L. Reina, `hep-ph/0512377`.
13. P.J. Mohr and B.N. Taylor, Rev. Mod. Phys. **72**, 351 (2000).
14. TOPAZ: I. Levine *et al.*, Phys. Rev. Lett. **78**, 424 (1997);
    VENUS: S. Okada *et al.*, Phys. Rev. Lett. **81**, 2428 (1998);
    L3: M. Acciarri *et al.*, Phys. Lett. **B476**, 40 (2000);
    L3: P. Achard *et al.*, Phys. Lett. **B623**, 26 (2005);
    OPAL: G. Abbiendi *et al.*, Eur. Phys. J. **C33**, 173 (2004);
    OPAL: G. Abbiendi *et al.*, Eur. Phys. J. **C45**, 1 (2006).
15. S. Fanchiotti, B. Kniehl, and A. Sirlin, Phys. Rev. **D48**, 307 (1993) and references therein.
16. J. Erler, Phys. Rev. **D59**, 054008 (1999).
17. CMD 2: R.R. Akhmetshin *et al.*, Phys. Lett. **B578**, 285 (2004).
18. M. Davier, S. Eidelman, A. Höcker, and Z. Zhang, Eur. Phys. J. **C31**, 503 (2003).
19. ALEPH: S. Schael *et al.*, Phys. Reports **421**, 191 (2005).

20. M. Davier, A. Höcker and Z. Zhang, hep-ph/0507078.
21. A.D. Martin and D. Zeppenfeld, Phys. Lett. **B345**, 558 (1995).
22. S. Eidelman and F. Jegerlehner, Z. Phys. **C67**, 585 (1995).
23. B.V. Geshkenbein and V.L. Morgunov, Phys. Lett. **B340**, 185 (1995);
    B.V. Geshkenbein and V.L. Morgunov, Phys. Lett. **B352**, 456 (1995).
24. H. Burkhardt and B. Pietrzyk, Phys. Lett. **B356**, 398 (1995).
25. M.L. Swartz, Phys. Rev. **D53**, 5268 (1996).
26. R. Alemany, M. Davier, and A. Höcker, Eur. Phys. J. **C2**, 123 (1998).
27. N.V. Krasnikov and R. Rodenberg, Nuovo Cimento **111A**, 217 (1998).
28. M. Davier and A. Höcker, Phys. Lett. **B419**, 419 (1998).
29. J.H. Kühn and M. Steinhauser, Phys. Lett. **B437**, 425 (1998).
30. M. Davier and A. Höcker, Phys. Lett. **B435**, 427 (1998).
31. S. Groote, J.G. Körner, K. Schilcher, N.F. Nasrallah, Phys. Lett. **B440**, 375 (1998).
32. A.D. Martin, J. Outhwaite, and M.G. Ryskin, Phys. Lett. **B492**, 69 (2000).
33. H. Burkhardt and B. Pietrzyk, Phys. Lett. **B513**, 46 (2001).
34. J.F. de Troconiz and F.J. Yndurain, Phys. Rev. **D65**, 093002 (2002).
35. F. Jegerlehner, Nucl. Phys. Proc. Suppl. **126**, 325 (2004).
36. K. Hagiwara, A. D. Martin, D. Nomura and T. Teubner, Phys. Rev. **D69**, 093003 (2004).
37. H. Burkhardt and B. Pietrzyk, Phys. Rev. **D72**, 057501 (2005).
38. BES: J.Z. Bai *et al.*, Phys. Rev. Lett. **88**, 101802 (2002);
    G.S. Huang, hep-ex/0105074.
39. SND: M.N. Achasov *et al.*, hep-ex/0506076.
40. CMD and OLYA: L.M. Barkov *et al.*, Nucl. Phys. **B256**, 365 (1985).
41. S. Binner, J.H. Kühn, and K. Melnikov, Phys. Lett. **B459**, 279 (1999).
42. KLOE: A. Aloisio *et al.*, Phys. Lett. **B606**, 12 (2005).
43. W.J. Marciano and A. Sirlin, Phys. Rev. Lett. **61**, 1815 (1988).
44. T. van Ritbergen and R.G. Stuart, Phys. Rev. Lett. **82**, 488 (1999).
45. ALEPH, DELPHI, L3, OPAL, SLD, LEP Electroweak Working Group, SLD Electroweak and Heavy Flavour Groups: S. Schael *et al.*, hep-ex/0509008.
46. Earlier analyses include U. Amaldi *et al.*, Phys. Rev. **D36**, 1385 (1987);
    G. Costa *et al.*, Nucl. Phys. **B297**, 244 (1988);
    Deep inelastic scattering is considered by G.L. Fogli and D. Haidt, Z. Phys. **C40**, 379 (1988);
    P. Langacker and M. Luo, Phys. Rev. **D44**, 817 (1991);
    For more recent analyses, see Ref. 47.
47. P. Langacker, p. 883 of Ref. 4;
    J. Erler and P. Langacker, Phys. Rev. **D52**, 441 (1995).
48. J. Erler and M.J. Ramsey-Musolf, Prog. Part. Nucl. Phys. **54**, 351 (2005);
    Neutrino scattering is reviewed by J.M. Conrad *et al.* in Ref. 3;
    Nonstandard neutrino interactions are surveyed in Z. Berezhiani and A. Rossi, Phys. Lett. **B535**, 207 (2002);
    S. Davidson, C. Peña-Garay, N. Rius, and A. Santamaria, JHEP **0303**, 011 (2003).

49. A. Sirlin, Phys. Rev. **D22**, 971 (1980);
A. Sirlin, Phys. Rev. **D29**, 89 (1984);
D.C. Kennedy *et al.*, Nucl. Phys. **B321**, 83 (1989);
D.C. Kennedy and B.W. Lynn, Nucl. Phys. **B322**, 1 (1989);
D.Yu. Bardin *et al.*, Z. Phys. **C44**, 493 (1989);
W. Hollik, Fortsch. Phys. **38**, 165 (1990);
For reviews, see the articles by W. Hollik, pp. 37 and 117, and W. Marciano, p. 170 in Ref. 4. Extensive references to other papers are given in Ref. 46.
50. V.A. Novikov, L.B. Okun, and M.I. Vysotsky, Nucl. Phys. **B397**, 35 (1993).
51. W. Hollik in Ref. 49 and references therein.
52. W.J. Marciano and J.L. Rosner, Phys. Rev. Lett. **65**, 2963 (1990).
53. G. Degrassi, S. Fanchiotti, and A. Sirlin, Nucl. Phys. **B351**, 49 (1991).
54. G. Degrassi and A. Sirlin, Nucl. Phys. **B352**, 342 (1991).
55. P. Gambino and A. Sirlin, Phys. Rev. **D49**, 1160 (1994).
56. ZFITTER: D. Bardin *et al.*, Comput. Phys. Commun. **133**, 229 (2001) and references therein;
ZFITTER: A.B. Arbuzov *et al.*, `hep-ph/0507146`.
57. R. Barbieri *et al.*, Phys. Lett. **B288**, 95 (1992) and *ibid.* **312**, 511(E) (1993);
R. Barbieri *et al.*, Nucl. Phys. **B409**, 105 (1993).
58. J. Fleischer, O.V. Tarasov, and F. Jegerlehner, Phys. Lett. **B319**, 249 (1993).
59. G. Degrassi, P. Gambino, and A. Vicini, Phys. Lett. **B383**, 219 (1996);
G. Degrassi, P. Gambino, and A. Sirlin, Phys. Lett. **B394**, 188 (1997).
60. A. Freitas, W. Hollik, W. Walter, and G. Weiglein, Phys. Lett. **B495**, 338 (2000) and *ibid.* **570**, 260(E) (2003);
M. Awramik and M. Czakon, Phys. Lett. **B568**, 48 (2003).
61. A. Freitas, W. Hollik, W. Walter, and G. Weiglein, Nucl. Phys. **B632**, 189 (2002) and *ibid.* **666**, 305(E) (2003);
M. Awramik and M. Czakon, Phys. Rev. Lett. **89**, 241801 (2002);
A. Onishchenko and O. Veretin, Phys. Lett. **B551**, 111 (2003).
62. M. Awramik, M. Czakon, A. Freitas and G. Weiglein, Phys. Rev. Lett. **93**, 201805 (2004);
W. Hollik, U. Meier and S. Uccirati, Nucl. Phys. **B731**, 213 (2005).
63. A. Djouadi and C. Verzegnassi, Phys. Lett. **B195**, 265 (1987);
A. Djouadi, Nuovo Cimento **100A**, 357 (1988).
64. K.G. Chetyrkin, J.H. Kühn, and M. Steinhauser, Phys. Lett. **B351**, 331 (1995);
L. Avdeev *et al.*, Phys. Lett. **B336**, 560 (1994) and **B349**, 597(E) (1995).
65. B.A. Kniehl, J.H. Kühn, and R.G. Stuart, Phys. Lett. **B214**, 621 (1988);
B.A. Kniehl, Nucl. Phys. **B347**, 86 (1990);
F. Halzen and B.A. Kniehl, Nucl. Phys. **B353**, 567 (1991);
A. Djouadi and P. Gambino, Phys. Rev. **D49**, 4705 (1994);
A. Djouadi and P. Gambino, Phys. Rev. **D49**, 3499 (1994) and *ibid.* **53**, 4111(E) (1996).
66. K.G. Chetyrkin, J.H. Kühn, and M. Steinhauser, Phys. Rev. Lett. **75**, 3394 (1995).
67. J.J. van der Bij *et al.*, Phys. Lett. **B498**, 156 (2001).

68. M. Faisst, J.H. Kühn, T. Seidensticker, and O. Veretin, Nucl. Phys. **B665**, 649 (2003).
69. R. Boughezal, J.B. Tausk, and J.J. van der Bij, Nucl. Phys. **B713**, 278 (2005); R. Boughezal, J.B. Tausk, and J.J. van der Bij, Nucl. Phys. **B725**, 3 (2005).
70. J. Fleischer *et al.*, Phys. Lett. **B293**, 437 (1992); K.G. Chetyrkin, A. Kwiatkowski, and M. Steinhauser, Mod. Phys. Lett. **A8**, 2785 (1993).
71. R. Harlander, T. Seidensticker, and M. Steinhauser, Phys. Lett. **B426**, 125 (1998); J. Fleischer *et al.*, Phys. Lett. **B459**, 625 (1999).
72. A. Czarnecki and J.H. Kühn, Phys. Rev. Lett. **77**, 3955 (1996).
73. J. Erler, `hep-ph/0005084`.
74. For reviews, see F. Perrier, p. 385 of Ref. 4; J.M. Conrad *et al.* in Ref. 3.
75. CDHS: H. Abramowicz *et al.*, Phys. Rev. Lett. **57**, 298 (1986); CDHS: A. Blondel *et al.*, Z. Phys. **C45**, 361 (1990).
76. CHARM: J.V. Allaby *et al.*, Phys. Lett. **B177**, 446 (1986); CHARM: J.V. Allaby *et al.*, Z. Phys. **C36**, 611 (1987).
77. CCFR: C.G. Arroyo *et al.*, Phys. Rev. Lett. **72**, 3452 (1994); CCFR: K.S. McFarland *et al.*, Eur. Phys. J. **C1**, 509 (1998).
78. NOMAD: R. Petti *et al.*, `hep-ex/0411032`.
79. R.M. Barnett, Phys. Rev. **D14**, 70 (1976); H. Georgi and H.D. Politzer, Phys. Rev. **D14**, 1829 (1976).
80. LAB-E: S.A. Rabinowitz *et al.*, Phys. Rev. Lett. **70**, 134 (1993).
81. E.A. Paschos and L. Wolfenstein, Phys. Rev. **D7**, 91 (1973).
82. NuTeV: G. P. Zeller *et al.*, Phys. Rev. Lett. **88**, 091802 (2002).
83. For reviews including discussions of possible new physics explanations, see S. Davidson *et al.*, JHEP **0202**, 037 (2002); J.T. Londergan, Nucl. Phys. Proc. Suppl. **141**, 68 (2005).
84. J. Alwall and G. Ingelman, Phys. Rev. **D70**, 111505 (2004); Y. Ding, R.G. Xu and B.Q. Ma, Phys. Lett. **B607**, 101 (2005); M. Wakamatsu, Phys. Rev. **D71**, 057504 (2005); M. Glück, P. Jimenez-Delgado, and E. Reya, Phys. Rev. Lett. **95**, 022002 (2005).
85. NuTeV: M. Goncharov *et al.*, Phys. Rev. **D64**, 112006 (2001); NuTeV: D. Mason *et al.*, `hep-ex/0405037`.
86. NuTeV: G.P. Zeller *et al.*, Phys. Rev. **D65**, 111103 (2002); NuTeV: R. H. Bernstein *et al.*, J. Phys. G **29**, 1919 (2003).
87. S. Kretzer *et al.*, Phys. Rev. Lett. **93**, 041802 (2004).
88. E. Sather, Phys. Lett. **B274**, 433 (1992); E.N. Rodionov, A.W. Thomas, and J.T. Londergan, Mod. Phys. Lett. **A9**, 1799 (1994).
89. A.D. Martin, R.G. Roberts, W.J. Stirling and R.S. Thorne, Eur. Phys. J. **C35**, 325 (2004).

90. S. Kumano, Phys. Rev. **D66**, 111301 (2002);
S.A. Kulagin, Phys. Rev. **D67**, 091301 (2003);
M. Hirai, S. Kumano and T. H. Nagai, Phys. Rev. **D71**, 113007 (2005).
91. G.A. Miller and A.W. Thomas, Int. J. Mod. Phys. A **20**, 95 (2005).
92. S.J. Brodsky, I. Schmidt and J.J. Yang, Phys. Rev. **D70**, 116003 (2004).
93. K.P.O. Diener, S. Dittmaier, and W. Hollik, Phys. Rev. **D69**, 073005 (2004);
A.B. Arbuzov, D.Y. Bardin, and L.V. Kalinovskaya, JHEP **0506**, 078 (2005).
94. K.P.O. Diener, S. Dittmaier, and W. Hollik, Phys. Rev. **D72**, 093002 (2005).
95. B.A. Dobrescu and R.K. Ellis, Phys. Rev. **D69**, 114014 (2004).
96. A.D. Martin, R.G. Roberts, W.J. Stirling, and R.S. Thorne, Eur. Phys. J. **C39**, 155 (2005).
97. CHARM: J. Dorenbosch *et al.*, Z. Phys. **C41**, 567 (1989).
98. CALO: L.A. Ahrens *et al.*, Phys. Rev. **D41**, 3297 (1990).
99. CHARM II: P. Vilain *et al.*, Phys. Lett. **B335**, 246 (1994).
100. See also J. Panman, p. 504 of Ref. 4.
101. ILM: R.C. Allen *et al.*, Phys. Rev. **D47**, 11 (1993);
LSND: L.B. Auerbach *et al.*, Phys. Rev. **D63**, 112001 (2001).
102. SSF: C.Y. Prescott *et al.*, Phys. Lett. **B84**, 524 (1979);
For a review, see P. Souder, p. 599 of Ref. 4.
103. E. J. Beise, M. L. Pitt and D. T. Spayde, Prog. Part. Nucl. Phys. **54**, 289 (2005).
104. For reviews and references to earlier work, see M.A. Bouchiat and L. Pottier, Science **234**, 1203 (1986);
B.P. Masterson and C.E. Wieman, p. 545 of Ref. 4.
105. Cesium (Boulder): C.S. Wood *et al.*, Science **275**, 1759 (1997).
106. Cesium (Paris): J. Guéna, M. Lintz and M.A. Bouchiat, `physics/0412017`.
107. Thallium (Oxford): N.H. Edwards *et al.*, Phys. Rev. Lett. **74**, 2654 (1995);
Thallium (Seattle): P.A. Vetter *et al.*, Phys. Rev. Lett. **74**, 2658 (1995).
108. Lead (Seattle): D.M. Meekhof *et al.*, Phys. Rev. Lett. **71**, 3442 (1993).
109. Bismuth (Oxford): M.J.D. MacPherson *et al.*, Phys. Rev. Lett. **67**, 2784 (1991).
110. V.A. Dzuba, V.V. Flambaum, and O.P. Sushkov, Phys. Lett. **141A**, 147 (1989);
S.A. Blundell, J. Sapirstein, and W.R. Johnson, Phys. Rev. Lett. **65**, 1411 (1990) and Phys. Rev. **D45**, 1602 (1992);
For reviews, see S.A. Blundell, W.R. Johnson, and J. Sapirstein, p. 577 of Ref. 4;
J.S.M. Ginges and V.V. Flambaum, Phys. Reports **397**, 63 (2004);
J. Guena, M. Lintz and M. A. Bouchiat, Mod. Phys. Lett. **A20**, 375 (2005).
111. V.A. Dzuba, V.V. Flambaum, and O.P. Sushkov, Phys. Rev. **A56**, R4357 (1997).
112. S.C. Bennett and C.E. Wieman, Phys. Rev. Lett. **82**, 2484 (1999).
113. M.A. Bouchiat and J. Guéna, J. Phys. (France) **49**, 2037 (1988).
114. A. Derevianko, Phys. Rev. Lett. **85**, 1618 (2000);
V.A. Dzuba, C. Harabati, and W.R. Johnson, Phys. Rev. **A63**, 044103 (2001);
M.G. Kozlov, S.G. Porsev, and I.I. Tupitsyn, Phys. Rev. Lett. **86**, 3260 (2001).
115. A.I. Milstein and O.P. Sushkov, Phys. Rev. **A66**, 022108 (2002);
W.R. Johnson, I. Bednyakov, and G. Soff, Phys. Rev. Lett. **87**, 233001 (2001);
V.A. Dzuba, V.V. Flambaum, and J.S. Ginges, Phys. Rev. **D66**, 076013 (2002);

M.Y. Kuchiev and V.V. Flambaum, Phys. Rev. Lett. **89**, 283002 (2002);
A.I. Milstein, O.P. Sushkov, and I.S. Terekhov, Phys. Rev. Lett. **89**, 283003 (2002);
V.V. Flambaum and J.S.M. Ginges, `physics/0507067`.

116. J. Erler, A. Kurylov, and M.J. Ramsey-Musolf, Phys. Rev. **D68**, 016006 (2003).
117. V.A. Dzuba *et al.*, J. Phys. **B20**, 3297 (1987).
118. Ya.B. Zel'dovich, Sov. Phys. JETP **6**, 1184 (1958);
For recent discussions, see V.V. Flambaum and D.W. Murray, Phys. Rev. **C56**, 1641 (1997);
W.C. Haxton and C.E. Wieman, Ann. Rev. Nucl. Part. Sci. **51**, 261 (2001).
119. J.L. Rosner, Phys. Rev. **D53**, 2724 (1996).
120. S.J. Pollock, E.N. Fortson, and L. Wilets, Phys. Rev. **C46**, 2587 (1992);
B.Q. Chen and P. Vogel, Phys. Rev. **C48**, 1392 (1993).
121. B.W. Lynn and R.G. Stuart, Nucl. Phys. **B253**, 216 (1985).
122. *Physics at LEP*, ed. J. Ellis and R. Peccei, CERN 86–02, Vol. 1.
123. PETRA: S.L. Wu, Phys. Reports **107**, 59 (1984);
C. Kiesling, *Tests of the Standard Theory of Electroweak Interactions*, (Springer-Verlag, New York, 1988);
R. Marshall, Z. Phys. **C43**, 607 (1989);
Y. Mori *et al.*, Phys. Lett. **B218**, 499 (1989);
D. Haidt, p. 203 of Ref. 4.
124. For reviews, see D. Schaile, p. 215, and A. Blondel, p. 277 of Ref. 4.
125. SLD: K. Abe *et al.*, Phys. Rev. Lett. **84**, 5945 (2000).
126. SLD: K. Abe *et al.*, Phys. Rev. Lett. **85**, 5059 (2000).
127. SLD: K. Abe *et al.*, Phys. Rev. Lett. **86**, 1162 (2001).
128. DELPHI: P. Abreu *et al.*, Z. Phys. **C67**, 1 (1995);
OPAL: K. Ackerstaff *et al.*, Z. Phys. **C76**, 387 (1997).
129. SLD: K. Abe *et al.*, Phys. Rev. Lett. **78**, 17 (1997).
130. ALEPH, DELPHI, L3, OPAL, SLD, LEP Electroweak Working Group, SLD Electroweak and Heavy Flavour Groups: J. Alcarez *et al.*, `hep-ex/0511027`.
131. CDF: D. Acosta *et al.*, Phys. Rev. **D71**, 052002 (2005).
132. H1: A. Aktas *et al.*, Phys. Lett. **B632**, 35 (2006).
133. Results of difermion measurements at LEP2 can be found at URL `http://lepewwg.web.cern.c`
134. ALEPH, DELPHI, L3, and OPAL Collaborations, and the LEP Working Group for Higgs Boson Searches: D. Abbaneo *et al.*, Phys. Lett. **B565**, 61 (2003).
135. A. Leike, T. Riemann, and J. Rose, Phys. Lett. **B273**, 513 (1991);
T. Riemann, Phys. Lett. **B293**, 451 (1992).
136. E158: P.L. Anthony *et al.*, Phys. Rev. Lett. **95**, 081601 (2005);
the implications are discussed in A. Czarnecki and W.J. Marciano, Int. J. Mod. Phys. A **15**, 2365 (2000).
137. J. Erler and M.J. Ramsey-Musolf, Phys. Rev. **D72**, 073003 (2005);
for the scale dependence of the weak mixing angle defined in a mass dependent renormalization scheme, see A. Czarnecki and W.J. Marciano, Int. J. Mod. Phys. A **15**, 2365 (2000).

138. Qweak: D.S. Armstrong *et al.*, AIP Conf. Proc. **698**, 172 (2004);
the implications are discussed in Ref. 116.

139. A comprehensive report and further references can be found in K.G. Chetyrkin, J.H. Kühn, and A. Kwiatkowski, Phys. Reports **277**, 189 (1996).

140. J. Schwinger, *Particles, Sources and Fields*, Vol. II, (Addison-Wesley, New York, 1973);
K.G. Chetyrkin, A.L. Kataev, and F.V. Tkachev, Phys. Lett. **B85**, 277 (1979);
M. Dine and J. Sapirstein, Phys. Rev. Lett. **43**, 668 (1979);
W. Celmaster, R.J. Gonsalves, Phys. Rev. Lett. **44**, 560 (1980);
S.G. Gorishnii, A.L. Kataev, and S.A. Larin, Phys. Lett. **B212**, 238 (1988);
S.G. Gorishnii, A.L. Kataev, and S.A. Larin, Phys. Lett. **B259**, 144 (1991);
L.R. Surguladze and M.A. Samuel, Phys. Rev. Lett. **66**, 560 (1991) and *ibid.* 2416(E).

141. A.L. Kataev and V.V. Starshenko, Mod. Phys. Lett. **A10**, 235 (1995).

142. W. Bernreuther and W. Wetzel, Z. Phys. **11**, 113 (1981);
W. Bernreuther and W. Wetzel, Phys. Rev. **D24**, 2724 (1982);
B.A. Kniehl, Phys. Lett. **B237**, 127 (1990);
K.G. Chetyrkin, Phys. Lett. **B307**, 169 (1993);
A.H. Hoang *et al.*, Phys. Lett. **B338**, 330 (1994);
S.A. Larin, T. van Ritbergen, and J.A.M. Vermaseren, Nucl. Phys. **B438**, 278 (1995).

143. T.H. Chang, K.J.F. Gaemers, and W.L. van Neerven, Nucl. Phys. **B202**, 407 (1980);
J. Jersak, E. Laermann, and P.M. Zerwas, Phys. Lett. **B98**, 363 (1981);
J. Jersak, E. Laermann, and P.M. Zerwas, Phys. Rev. **D25**, 1218 (1982);
S.G. Gorishnii, A.L. Kataev, and S.A. Larin, Nuovo Cimento **92**, 117 (1986);
K.G. Chetyrkin and J.H. Kühn, Phys. Lett. **B248**, 359 (1990);
K.G. Chetyrkin, J.H. Kühn, and A. Kwiatkowski, Phys. Lett. **B282**, 221 (1992);
K.G. Chetyrkin and J.H. Kühn, Phys. Lett. **B406**, 102 (1997).

144. B.A. Kniehl and J.H. Kühn, Phys. Lett. **B224**, 229 (1990);
B.A. Kniehl and J.H. Kühn, Nucl. Phys. **B329**, 547 (1990);
K.G. Chetyrkin and A. Kwiatkowski, Phys. Lett. **B305**, 285 (1993);
K.G. Chetyrkin and A. Kwiatkowski, Phys. Lett. **B319**, 307 (1993);
S.A. Larin, T. van Ritbergen, and J.A.M. Vermaseren, Phys. Lett. **B320**, 159 (1994);
K.G. Chetyrkin and O.V. Tarasov, Phys. Lett. **B327**, 114 (1994).

145. A.L. Kataev, Phys. Lett. **B287**, 209 (1992).

146. D. Albert *et al.*, Nucl. Phys. **B166**, 460 (1980);
F. Jegerlehner, Z. Phys. **C32**, 425 (1986);
A. Djouadi, J.H. Kühn, and P.M. Zerwas, Z. Phys. **C46**, 411 (1990);
A. Borrelli *et al.*, Nucl. Phys. **B333**, 357 (1990).

147. A.A. Akhundov, D.Yu. Bardin, and T. Riemann, Nucl. Phys. **B276**, 1 (1986);
W. Beenakker and W. Hollik, Z. Phys. **C40**, 141 (1988);
B.W. Lynn and R.G. Stuart, Phys. Lett. **B352**, 676 (1990);
J. Bernabeu, A. Pich, and A. Santamaria, Nucl. Phys. **B363**, 326 (1991).

148. CLEO: S. Chen *et al.*, Phys. Rev. Lett. **87**, 251807 (2001).
149. Belle: P. Koppenburg *et al.*, Phys. Rev. Lett. **93**, 061803 (2004).
150. BaBar: B. Aubert *et al.*, `hep-ex/0507001`;
BaBar: B. Aubert *et al.*, Phys. Rev. **D72**, 052004 (2005).
151. A.L. Kagan and M. Neubert, Eur. Phys. J. **C7**, 5 (1999).
152. A. Ali and C. Greub, Phys. Lett. **B259**, 182 (1991).
153. I. Bigi and N. Uraltsev, Int. J. Mod. Phys. A **17**, 4709 (2002).
154. A. Czarnecki and W.J. Marciano, Phys. Rev. Lett. **81**, 277 (1998).
155. J. Erler and D.M. Pierce, Nucl. Phys. **B526**, 53 (1998).
156. Y. Nir, Phys. Lett. **B221**, 184 (1989);
K. Adel and Y.P. Yao, Phys. Rev. **D49**, 4945 (1994);
C. Greub, T. Hurth, and D. Wyler, Phys. Rev. **D54**, 3350 (1996);
K.G. Chetyrkin, M. Misiak, and M. Münz, Phys. Lett. **B400**, 206 (1997);
C. Greub and T. Hurth, Phys. Rev. **D56**, 2934 (1997);
M. Ciuchini *et al.*, Nucl. Phys. **B527**, 21 (1998);
M. Ciuchini *et al.*, Nucl. Phys. **B534**, 3 (1998);
F.M. Borzumati and C. Greub, Phys. Rev. **D58**, 074004 (1998);
F.M. Borzumati and C. Greub, Phys. Rev. **D59**, 057501 (1999);
A. Strumia, Nucl. Phys. **B532**, 28 (1998).
157. F. Le Diberder and A. Pich, Phys. Lett. **B286**, 147 (1992).
158. T. van Ritbergen, J.A.M. Vermaseren, and S.A. Larin, Phys. Lett. **B400**, 379 (1997).
159. E821: H.N. Brown *et al.*, Phys. Rev. Lett. **86**, 2227 (2001);
E821: G.W. Bennett, *et al.*, Phys. Rev. Lett. **89**, 101804 (2002);
E821: G.W. Bennett *et al.*, Phys. Rev. Lett. **92**, 161802 (2004).
160. T. Kinoshita and M. Nio, Phys. Rev. **D70**, 113001 (2004).
161. S. Laporta and E. Remiddi, Phys. Lett. **B301**, 440 (1993);
S. Laporta and E. Remiddi, Phys. Lett. **B379**, 283 (1996).
162. J. Erler and M. Luo, Phys. Rev. Lett. **87**, 071804 (2001).
163. T. Kinoshita, Nucl. Phys. Proc. Suppl. **144**, 206 (2005).
164. For reviews, see V.W. Hughes and T. Kinoshita, Rev. Mod. Phys. **71**, S133 (1999);
A. Czarnecki and W.J. Marciano, Phys. Rev. **D64**, 013014 (2001);
T. Kinoshita, J. Phys. **G29**, 9 (2003);
M. Davier and W.J. Marciano, Ann. Rev. Nucl. Part. Sci. **54**, 115 (2004).
165. S.J. Brodsky and J.D. Sullivan, Phys. Rev. **D156**, 1644 (1967);
T. Burnett and M.J. Levine, Phys. Lett. **B24**, 467 (1967);
R. Jackiw and S. Weinberg, Phys. Rev. **D5**, 2473 (1972);
I. Bars and M. Yoshimura, Phys. Rev. **D6**, 374 (1972);
K. Fujikawa, B.W. Lee, and A.I. Sanda, Phys. Rev. **D6**, 2923 (1972);
G. Altarelli, N. Cabibbo, and L. Maiani, Phys. Lett. **B40**, 415 (1972);
W.A. Bardeen, R. Gastmans, and B.E. Laurup, Nucl. Phys. **B46**, 315 (1972).
166. T.V. Kukhto, E.A. Kuraev, A. Schiller, and Z.K. Silagadze, Nucl. Phys. **B371**, 567 (1992);
S. Peris, M. Perrottet, and E. de Rafael, Phys. Lett. **B355**, 523 (1995);

A. Czarnecki, B. Krause, and W.J. Marciano, Phys. Rev. **D52**, 2619 (1995);
A. Czarnecki, B. Krause, and W.J. Marciano, Phys. Rev. Lett. **76**, 3267 (1996).
167. G. Degrassi and G. Giudice, Phys. Rev. **D58**, 053007 (1998).
168. F. Matorras (DELPHI), contributed paper to the *International Europhysics Conference on High Energy Physics* (EPS 2003, Aachen).
169. V. Cirigliano, G. Ecker and H. Neufeld, JHEP **0208**, 002 (2002);
K. Maltman and C.E. Wolfe, Phys. Rev. **D73**, 013004 (2006).
170. J. Erler, Rev. Mex. Fis. **50**, 200 (2004).
171. K. Maltman, `hep-ph/0504201`.
172. S. Ghozzi and F. Jegerlehner, Phys. Lett. **B583**, 222 (2004).
173. K. Melnikov and A. Vainshtein, Phys. Rev. **D70**, 113006 (2004).
174. M. Knecht and A. Nyffeler, Phys. Rev. **D65**, 073034 (2002).
175. M. Hayakawa and T. Kinoshita, `hep-ph/0112102`;
J. Bijnens, E. Pallante and J. Prades, Nucl. Phys. **B626**, 410 (2002).
176. B. Krause, Phys. Lett. **B390**, 392 (1997).
177. J.L. Lopez, D.V. Nanopoulos, and X. Wang, Phys. Rev. **D49**, 366 (1994);
for recent reviews, see Ref. 164.
178. UA2: S. Alitti *et al.*, Phys. Lett. **B276**, 354 (1992);
CDF: T. Affolder *et al.*, Phys. Rev. **D64**, 052001 (2001);
DØ: V. M. Abazov *et al.*, Phys. Rev. **D66**, 012001 (2002).
179. CDF and DØ Collaborations: Phys. Rev. **D70**, 092008 (2004).
180. M. Grünewald, presented at the *International Europhysics Conference on High Energy Physics* (HEPP-EPS 2005, Lisbon).
181. F. James and M. Roos, Comput. Phys. Commun. **10**, 343 (1975).
182. J. Erler, J.L. Feng, and N. Polonsky, Phys. Rev. Lett. **78**, 3063 (1997).
183. D. Choudhury, T.M.P. Tait and C.E.M. Wagner, Phys. Rev. **D65**, 053002 (2002).
184. J. Erler and P. Langacker, Phys. Rev. Lett. **84**, 212 (2000).
185. DELPHI: P. Abreu *et al.*, Eur. Phys. J. **C10**, 415 (1999).
186. S. Bethke, Phys. Reports **403**, 203 (2004).
187. C. Glasman, `hep-ex/0506035`.
188. HPQCD and UKQCD: Q. Mason *et al.*, Phys. Rev. Lett. **95**, 052002 (2005).
189. J. Erler, Phys. Rev. **D63**, 071301 (2001).
190. P. Langacker and N. Polonsky, Phys. Rev. **D52**, 3081 (1995);
J. Bagger, K.T. Matchev, and D. Pierce, Phys. Lett. **B348**, 443 (1995).
191. M. Veltman, Nucl. Phys. **B123**, 89 (1977);
M. Chanowitz, M.A. Furman, and I. Hinchliffe, Phys. Lett. **B78**, 285 (1978).
192. P. Langacker and M. Luo, Phys. Rev. **D45**, 278 (1992) and references therein.
193. A. Denner, R.J. Guth, and J.H. Kühn, Phys. Lett. **B240**, 438 (1990).
194. S. Bertolini and A. Sirlin, Phys. Lett. **B257**, 179 (1991).
195. M. Peskin and T. Takeuchi, Phys. Rev. Lett. **65**, 964 (1990);
M. Peskin and T. Takeuchi, Phys. Rev. **D46**, 381 (1992);
M. Golden and L. Randall, Nucl. Phys. **B361**, 3 (1991).
196. D. Kennedy and P. Langacker, Phys. Rev. Lett. **65**, 2967 (1990);
D. Kennedy and P. Langacker, Phys. Rev. **D44**, 1591 (1991).

197. G. Altarelli and R. Barbieri, Phys. Lett. **B253**, 161 (1990).
198. B. Holdom and J. Terning, Phys. Lett. **B247**, 88 (1990).
199. B.W. Lynn, M.E. Peskin, and R.G. Stuart, p. 90 of Ref. 122.
200. An alternative formulation is given by K. Hagiwara *et al.*, Z. Phys. **C64**, 559 (1994) and *ibid.* **68**, 352(E) (1995);
K. Hagiwara, D. Haidt, and S. Matsumoto, Eur. Phys. J. **C2**, 95 (1998).
201. I. Maksymyk, C.P. Burgess, and D. London, Phys. Rev. **D50**, 529 (1994);
C.P. Burgess *et al.*, Phys. Lett. **B326**, 276 (1994).
202. R. Barbieri, A. Pomarol, R. Rattazzi and A. Strumia, Nucl. Phys. **B703**, 127 (2004).
203. K. Lane, `hep-ph/0202255`.
204. E. Gates and J. Terning, Phys. Rev. Lett. **67**, 1840 (1991);
R. Sundrum and S.D.H. Hsu, Nucl. Phys. **B391**, 127 (1993);
R. Sundrum, Nucl. Phys. **B395**, 60 (1993);
M. Luty and R. Sundrum, Phys. Rev. Lett. **70**, 529 (1993);
T. Appelquist and J. Terning, Phys. Lett. **B315**, 139 (1993);
D.D. Dietrich, F. Sannino and K. Tuominen, Phys. Rev. **D72**, 055001 (2005);
N.D. Christensen and R. Shrock, Phys. Lett. **B632**, 92 (2006);
M. Harada, M. Kurachi and K. Yamawaki, `hep-ph/0509193`.
205. H. Georgi, Nucl. Phys. **B363**, 301 (1991);
M.J. Dugan and L. Randall, Phys. Lett. **B264**, 154 (1991).
206. R. Barbieri *et al.*, Nucl. Phys. **B341**, 309 (1990).
207. M.E. Peskin and J.D. Wells, Phys. Rev. **D64**, 093003 (2001).
208. H.J. He, N. Polonsky, and S. Su, Phys. Rev. **D64**, 053004 (2001);
V.A. Novikov, L.B. Okun, A.N. Rozanov, and M.I. Vysotsky, Sov. Phys. JETP **76**, 127 (2002);
S. S. Bulanov, V. A. Novikov, L. B. Okun, A. N. Rozanov and M. I. Vysotsky, Yad. Fiz. **66**, 2219 (2003) and references therein.
209. For a review, see D. London, p. 951 of Ref. 4;
a recent analysis is M.B. Popovic and E.H. Simmons, Phys. Rev. **D58**, 095007 (1998);
for collider implications, see T.C. Andre and J.L. Rosner, Phys. Rev. **D69**, 035009 (2004).
210. P. Langacker, M. Luo, and A.K. Mann, Rev. Mod. Phys. **64**, 87 (1992);
M. Luo, p. 977 of Ref. 4.
211. F.S. Merritt *et al.*, p. 19 of *Particle Physics: Perspectives and Opportunities: Report of the DPF Committee on Long Term Planning*, ed. R. Peccei *et al.* (World Scientific, Singapore, 1995).
212. G. C. Cho and K. Hagiwara, Nucl. Phys. **B574**, 623 (2000);
G. Altarelli *et al.*, JHEP **0106**, 018 (2001);
A. Kurylov, M.J. Ramsey-Musolf, and S. Su, Nucl. Phys. **B667**, 321 (2003);
A. Kurylov, M.J. Ramsey-Musolf, and S. Su, Phys. Rev. **D68**, 035008 (2003);
W. de Boer and C. Sander, Phys. Lett. **B585**, 276 (2004);
S. Heinemeyer, W. Hollik and G. Weiglein, `hep-ph/0412214`;

J.R. Ellis, K.A. Olive, Y. Santoso, and V.C. Spanos, Phys. Rev. **D69**, 095004 (2004);
S. P. Martin, K. Tobe and J. D. Wells, Phys. Rev. **D71**, 073014 (2005);
G. Marandella, C. Schappacher and A. Strumia, Nucl. Phys. **B715**, 173 (2005).
213. C.T. Hill and E.H. Simmons, Phys. Reports **381**, 235 (2003);
R. S. Chivukula, E. H. Simmons, H. J. He, M. Kurachi and M. Tanabashi, Phys. Rev. **D70**, 075008 (2004).
214. K. Agashe, A. Delgado, M.J. May, and R. Sundrum, JHEP **0308**, 050 (2003);
M. Carena *et al.*, Phys. Rev. **D68**, 035010 (2003);
for reviews, see the articles on "Extra Dimensions" in this *Review* and I. Antoniadis, `hep-th/0102202`.
215. For a review, see M. Perelstein, `hep-ph/0512128`.
216. G. Altarelli, R. Barbieri, and S. Jadach, Nucl. Phys. **B369**, 3 (1992) and **B376**, 444(E) (1992).
217. A. De Rújula *et al.*, Nucl. Phys. **B384**, 3 (1992).
218. K. Hagiwara *et al.*, Phys. Rev. **D48**, 2182 (1993).
219. C.P. Burgess and D. London, Phys. Rev. **D48**, 4337 (1993).
220. For a review, see A. Leike, Phys. Reports **317**, 143 (1999).
221. M. Cvetic and P. Langacker, Phys. Rev. **D54**, 3570 (1996).
222. R.S. Chivukula and E.H. Simmons, Phys. Rev. **D66**, 015006 (2002).
223. S. Chaudhuri *et al.*, Nucl. Phys. **B456**, 89 (1995);
G. Cleaver *et al.*, Phys. Rev. **D59**, 055005 (1999).
224. M. Carena, A. Daleo, B. A. Dobrescu and T. M. P. Tait, Phys. Rev. **D70**, 093009 (2004).
225. CDF: F. Abe *et al.*, Phys. Rev. Lett. **79**, 2192 (1997);
A. Abulencia *et al.*, Phys. Rev. Lett. **95**, 252001 (2005);
preliminary CDF and D0 limits from Run II may be found at URLs `http://www-cdf.fnal.gov/` and `http://www-d0.fnal.gov/`.
226. J. Kang and P. Langacker, Phys. Rev. **D71**, 035014 (2005).
227. J. Erler and P. Langacker, Phys. Lett. **B456**, 68 (1999).
228. T. Appelquist, B.A. Dobrescu, and A.R. Hopper, Phys. Rev. **D68**, 035012 (2003);
R.S. Chivukula, H.J. He, J. Howard, and E.H. Simmons, Phys. Rev. **D69**, 015009 (2004).
229. P. Langacker and M. Plümacher, Phys. Rev. **D62**, 013006 (2000).
230. R. Casalbuoni, S. De Curtis, D. Dominici, and R. Gatto, Phys. Lett. **B460**, 135 (1999);
J.L. Rosner, Phys. Rev. **D61**, 016006 (2000).

11. THE CKM QUARK-MIXING MATRIX

Written January 2006 by A. Ceccucci (CERN), Z. Ligeti (LBNL), and Y. Sakai (KEK).

11.1. Introduction

The masses and mixings of quarks have a common origin in the Standard Model (SM). They arise from the Yukawa interactions with the Higgs condensate,

$$\mathcal{L}_Y = -Y^d_{ij}\,\overline{Q^I_{Li}}\,\phi\,d^I_{Rj} - Y^u_{ij}\,\overline{Q^I_{Li}}\,\epsilon\,\phi^*u^I_{Rj} + \text{h.c.}, \tag{11.1}$$

where $Y^{u,d}$ are 3×3 complex matrices, ϕ is the Higgs field, i,j are generation labels, and ϵ is the 2×2 antisymmetric tensor. Q^I_L are left handed quark doublets, and d^I_R and u^I_R are right handed down- and up-type quark singlets, respectively, in the weak-eigenstate basis. When ϕ acquires a vacuum expectation value, $\langle\phi\rangle = (0, v/\sqrt{2})$, Eq. (11.1) yields mass terms for the quarks. The physical states are obtained by diagonalizing $Y^{u,d}$ by four unitary matrices, $V^{u,d}_{L,R}$, as $M^f_{\rm diag} = V^f_L\,Y^f\,V^{f\dagger}_R(v/\sqrt{2})$, $f = u, d$. As a result, the charged current $W^\pm$ interactions couple to the physical u_{Lj} and d_{Lk} quarks with couplings given by

$$V_{\rm CKM} \equiv V^u_L V^{d\dagger}_L = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}. \tag{11.2}$$

This Cabibbo-Kobayashi-Maskawa (CKM) matrix [1,2] is a 3×3 unitary matrix. It can be parameterized by three mixing angles and a CP-violating phase. Of the many possible parameterizations, a standard choice is [3]

$$V = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix}, \tag{11.3}$$

where $s_{ij} = \sin\theta_{ij}$, $c_{ij} = \cos\theta_{ij}$, and δ is the KM phase [2] responsible for all CP-violating phenomena in flavor changing processes in the SM. The angles θ_{ij} can be chosen to lie in the first quadrant, so $s_{ij}, c_{ij} \geq 0$.

It is known experimentally that $s_{13} \ll s_{23} \ll s_{12} \ll 1$, and it is convenient to exhibit this hierarchy using the Wolfenstein parameterization. We define [4–6]

$$s_{12} = \lambda = \frac{|V_{us}|}{\sqrt{|V_{ud}|^2 + |V_{us}|^2}}\,, \qquad s_{23} = A\lambda^2 = \lambda\left|\frac{V_{cb}}{V_{us}}\right|,$$
$$s_{13}e^{i\delta} = V^*_{ub} = A\lambda^3(\rho + i\eta) = \frac{A\lambda^3(\bar\rho + i\bar\eta)\sqrt{1 - A^2\lambda^4}}{\sqrt{1-\lambda^2}[1 - A^2\lambda^4(\bar\rho + i\bar\eta)]}\,. \tag{11.4}$$

These ensure that $\bar\rho + i\bar\eta = -(V_{ud}V^*_{ub})/(V_{cd}V^*_{cb})$ is phase-convention independent and the CKM matrix written in terms of λ, A, $\bar\rho$ and $\bar\eta$ is unitary to all orders in λ. The definitions of $\bar\rho, \bar\eta$ reproduce all approximate results in the literature. *E.g.*, $\bar\rho = \rho(1 - \lambda^2/2 + \ldots)$ and we can write $V_{\rm CKM}$ to $\mathcal{O}(\lambda^4)$ either in terms of $\bar\rho, \bar\eta$ or, traditionally,

$$V = \begin{pmatrix} 1 - \lambda^2/2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \lambda^2/2 & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}(\lambda^4)\,. \tag{11.5}$$

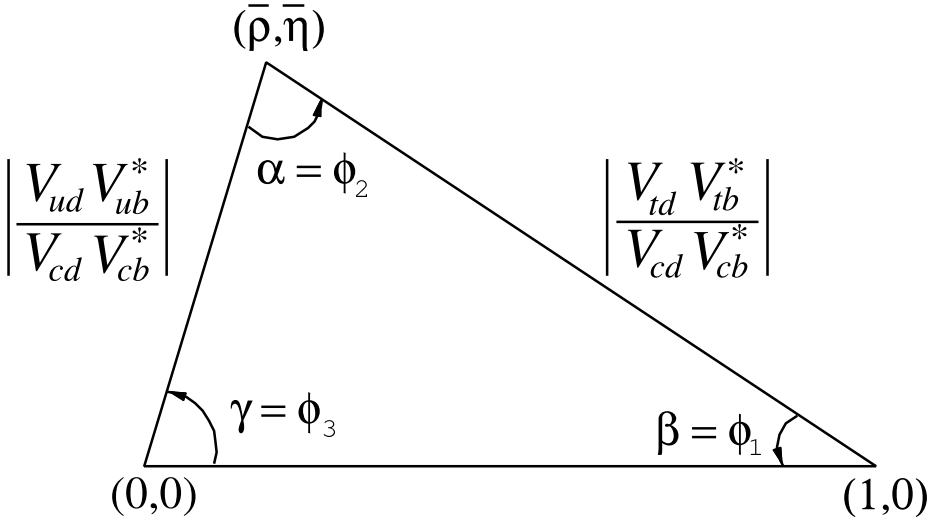


Figure 11.1: Sketch of the unitarity triangle.

The CKM matrix elements are fundamental parameters of the SM, so their precise determination is important. The unitarity of the CKM matrix imposes $\sum_i V_{ij}V_{ik}^* = \delta_{jk}$ and $\sum_j V_{ij}V_{kj}^* = \delta_{ik}$. The six vanishing combinations can be represented as triangles in a complex plane, of which the ones obtained by taking scalar products of neighboring rows or columns are nearly degenerate. The areas of all triangles are the same, half of the Jarlskog invariant, J [7], which is a phase-convention independent measure of CP violation, $\mathrm{Im}\left[V_{ij}V_{kl}V_{il}^*V_{kj}^*\right] = J\sum_{m,n}\varepsilon_{ikm}\varepsilon_{jln}$.

The most commonly used unitarity triangle arises from

$$V_{ud}\,V_{ub}^* + V_{cd}\,V_{cb}^* + V_{td}\,V_{tb}^* = 0\,, \tag{11.6}$$

by dividing each side by the best-known one, $V_{cd}V_{cb}^*$ (see Fig. 1). Its vertices are exactly $(0,0)$, $(1,0)$ and, due to the definition in Eq. (11.4), $(\bar{\rho},\bar{\eta})$. An important goal of flavor physics is to overconstrain the CKM elements, and many measurements can be conveniently displayed and compared in the $\bar{\rho},\bar{\eta}$ plane.

Processes dominated by loop contributions in the SM are sensitive to new physics and can be used to extract CKM elements only if the SM is assumed. In Sec. 11.2 and 11.3 we describe such measurements assuming the SM, and discuss implications for new physics in Sec. 11.5.

11.2. Magnitudes of CKM elements

11.2.1. $|V_{ud}|$:

The most precise determination of $|V_{ud}|$ comes from the study of superallowed $0^+ \to 0^+$ nuclear beta decays, which are pure vector transitions. Taking the average of the nine most precise determinations [8,9] yields [10]

$$|V_{ud}| = 0.97377 \pm 0.00027. \tag{11.7}$$

The error is dominated by theoretical uncertainties stemming from nuclear Coulomb distortions and radiative corrections. A recent measurement of ^{46}V performed using an atomic Penning trap method [9] deviates by 2.7 standard deviations from the average and may indicate a need to remeasure superallowed $0^+ \to 0^+$ transitions with modern atomic physics techniques. A precise determination of $|V_{ud}|$ is also obtained from the measurement of the neutron lifetime. The theoretical uncertainties are very small but the determination is limited by the knowledge of the ratio of the axial-vector and vector couplings, $g_A = G_A/G_V$ [10]. The PIBETA experiment [11] has improved the measurement of the $\pi^+ \to \pi^0 e^+ \nu$ branching ratio to 0.6%, and quote $|V_{ud}| = 0.9728 \pm 0.0030$, in agreement with the more precise results listed above. The interest in this measurement is that the determination of $|V_{ud}|$ is very clean theoretically, because it is a pure vector transition and is free from nuclear structure uncertainties.

11.2.2. $|V_{us}|$:

The magnitude of V_{us} has been extracted traditionally from semileptonic kaon decays. Significant progress on the experimental side has taken place during the last two years. After a high statistics measurement of $\mathcal{B}(K^+ \to \pi^0 e^+ \nu)$ [12], a suite of measurements of neutral kaon branching ratios [13–15], form factors [16], and lifetime [17] followed. An early value of $f_+(0) = 0.961 \pm 0.008$ [18] is still broadly accepted [10,19]. It yields from the kaon semileptonic decay rate

$$|V_{us}| = 0.2257 \pm 0.0021. \tag{11.8}$$

A word of caution is in order, since other calculations of $f_+(0)$ [20,21] differ by as much as 2%, with uncertainties around 1%.

Other determinations of $|V_{us}|$ involve leptonic kaon decays, hyperon decays, and τ decays. The calculation of the ratio of the kaon and pion decay constants enables one to extract $|V_{us}/V_{ud}|$ from $K \to \mu\nu(\gamma)$ and $\pi \to \mu\nu(\gamma)$, where (γ) indicates that radiative decays are included [22]. The KLOE measurement of the $K^+ \to \mu^+\nu(\gamma)$ branching ratio [23], combined with the unquenched lattice QCD calculation, $f_K/f_\pi = 1.198 \pm 0.003^{+0.016}_{-0.005}$ [24], leads to $|V_{us}| = 0.2245^{+0.0012}_{-0.0031}$, where the accuracy is limited by the knowledge of the ratio of the decay constants. Hereafter, the first error is statistical and the second is systematic, unless mentioned otherwise. The determination from hyperon decays was recently updated [25]. These authors focus on the analysis of the vector form factor, protected from first order $SU(3)$ breaking effects by the Ademollo-Gatto theorem [26], and treat the ratio between the axial and vector form factors g_1/f_1 as experimental input, thus avoiding first order $SU(3)$ breaking effects in the axial-vector contribution. They find $|V_{us}| = 0.2250 \pm 0.0027$, although this does not include an estimate of the theoretical uncertainty due to second order $SU(3)$ breaking, contrary to Eq. (11.8). Concerning hadronic τ decays to strange particles, the latest determinations based on OPAL or ALEPH data also yields consistent results [27] with accuracy around 0.003.

11.2.3. $|V_{cd}|$:

The magnitude of V_{cd} can be extracted from semileptonic charm decays if theoretical knowledge of the form factors is available. Recently, the first three-flavor unquenched lattice QCD calculations for $D \to K\ell\nu$ and $D \to \pi\ell\nu$ have been published [28]. Using these estimates and the isospin-averaged charm semileptonic width measured by CLEO-c, one obtains $|V_{cd}| = 0.213 \pm 0.008 \pm 0.021$ [29], where the first uncertainty is experimental and the second is from the theoretical error of the form factor.

This determination is not yet as precise as the one based on neutrino and antineutrino interactions. The difference of the ratio of double-muon to single-muon production by neutrino and antineutrino beams is proportional to the charm cross section off valence d-quarks, and therefore to $|V_{cd}|^2$ times the average semileptonic branching ratio of charm mesons, $\mathcal{B}_\mu$. The method was used first by CDHS [30] and then by CCFR [31,32] and CHARM II [33]. Averaging these results is complicated, not only because it requires assumptions about the scale of the QCD corrections, but also because $\mathcal{B}_\mu$ is an effective quantity, which depends on the specific neutrino beam characteristics. Given that no new experimental input is available, we quote the average provided in the previous edition of this review, $\mathcal{B}_\mu |V_{cd}|^2 = (0.463 \pm 0.034) \times 10^{-2}$ [3]. Analysis cuts make these experiments insensitive to neutrino energies smaller than 30 GeV. Thus, $\mathcal{B}_\mu$ should be computed using only neutrino interactions with visible energy larger than 30 GeV. A recent appraisal [34] based on charm production fractions measured in neutrino interactions [35,36] gives $\mathcal{B}_\mu = 0.088 \pm 0.006$. Data from the CHORUS experiment [37] are now sufficiently precise to extract $\mathcal{B}_\mu$ directly, by comparing the number of charm decays with a muon to the total number of charmed hadrons found in the nuclear emulsions. Requiring the visible energy to be larger than 30 GeV, CHORUS finds $\mathcal{B}_\mu = 0.085 \pm 0.009 \pm 0.006$. To extract $|V_{cd}|$, we use the average of these two determinations, $\mathcal{B}_\mu = 0.0873 \pm 0.0052$, and obtain

$$|V_{cd}| = 0.230 \pm 0.011. \tag{11.9}$$

11.2.4. $|V_{cs}|$:

The determination of $|V_{cs}|$ from neutrino and antineutrino scattering suffers from the uncertainty of the s-quark sea content. Measurements sensitive to $|V_{cs}|$ from on-shell $W^\pm$ decays were performed at LEP-2. The branching ratios of the W depend on the six CKM matrix elements involving quarks with masses smaller than M_W. The W branching ratio to each lepton flavor is $1/\mathcal{B}(W \to \ell\bar{\nu}_\ell) = 3\big[1 + \sum_{u,c,d,s,b} |V_{ij}|^2 \, (1 + \alpha_s(m_W)/\pi)\big]$. The measurement assuming lepton universality, $\mathcal{B}(W \to \ell\bar{\nu}_\ell) = (10.83 \pm 0.07 \pm 0.07)\,\%$ [38], implies $\sum_{u,c,d,s,b} |V_{ij}|^2 = 2.002 \pm 0.027$. This is a precise test of unitarity, but only flavor-tagged measurements determine $|V_{cs}|$. DELPHI measured tagged $W^+ \to c\bar{s}$ decays, obtaining $|V_{cs}| = 0.94^{+0.32}_{-0.26} \pm 0.13$ [39].

The direct determination of $|V_{cs}|$ is possible from semileptonic D or leptonic D_s decays, relying on the calculations of hadronic matrix elements. The measurement in $D_s^+ \to \ell^+\nu$ requires the lattice QCD calculation of the decay constant, f_{D_s}, while in semileptonic D decays one needs the form factors which depend on the invariant mass of the lepton pair, q^2. Unquenched lattice QCD calculations have predicted the normalization and the shape

of the form factor in $D \to K\ell\nu$ and $D \to \pi\ell\nu$ [28]. Using these theoretical results and the isospin averaged semileptonic widths one obtains [29]

$$|V_{cs}| = 0.957 \pm 0.017 \pm 0.093, \tag{11.10}$$

where the first error is experimental and the second one, which is dominant, is from the theoretical error of the form factor.

11.2.5. $|V_{cb}|$:

This matrix element can be determined from exclusive and inclusive semileptonic decays of B mesons to charm. The inclusive determinations use the semileptonic decay rate measurement together with the leptonic energy and the hadronic invariant-mass spectra. The theoretical foundation of the calculation is the operator product expansion (OPE) [40,41]. It expresses the total rate and moments of differential energy and invariant-mass spectra as expansions in α_s and inverse powers of the heavy quark mass. The dependence on m_b, m_c, and the parameters that occur at subleading order is different for different moments, and a large number of measured moments overconstrains all the parameters and tests the consistency of the determination. The precise extraction of $|V_{cb}|$ requires using a "threshold" quark mass definition [42,43]. Inclusive measurements have been performed using B mesons from Z^0 decays at LEP and at e^+e^- machines operated at the $\Upsilon(4S)$. At LEP the large boost of B mesons from Z^0 allows one to determine the moments throughout phase space, which is not possible otherwise, but the large statistics available at the B factories leads to more precise determinations. An average of the measurements and a compilation of the references are provided by Kowalewski and Mannel [44]: $|V_{cb}| = (41.7 \pm 0.7) \times 10^{-3}$.

Exclusive determinations are based on semileptonic B decays to D and D^*. In the $m_{b,c} \gg \Lambda_{\rm QCD}$ limit, all form factors are given by a single Isgur-Wise function [45], which depends on the product of the four-velocities, $w = v \cdot v'$, of the initial and final state hadrons. Heavy quark symmetry determines the normalization of the rate at $w = 1$, the maximum momentum transfer to the leptons, and $|V_{cb}|$ is obtained from an extrapolation to $w = 1$. The exclusive determination, $|V_{cb}| = (40.9 \pm 1.8) \times 10^{-3}$ [44], is less precise than the inclusive one, because the theoretical uncertainty in the form factor and the experimental uncertainty in the rate near $w = 1$ are both about 3%. The $|V_{cb}|$ review quotes the average as [44]

$$|V_{cb}| = (41.6 \pm 0.6) \times 10^{-3}. \tag{11.11}$$

11.2.6. $|V_{ub}|$:

The determination of $|V_{ub}|$ from inclusive $B \to X_u\ell\bar{\nu}$ decay suffers from large $B \to X_c\ell\bar{\nu}$ backgrounds. In most regions of phase space where the charm background is kinematically forbidden the hadronic physics enters via unknown nonperturbative functions, so-called shape functions. (In contrast, the nonperturbative physics for $|V_{cb}|$ is encoded in a few parameters.) At leading order in $\Lambda_{\rm QCD}/m_b$ there is only one shape function, which can be extracted from the photon energy spectrum in $B \to X_s\gamma$ [46,47] and applied to several spectra in $B \to X_u\ell\bar{\nu}$. The subleading shape functions are modeled

in the current calculations. Phase space cuts for which the rate does not depend on the shape function are also possible [48]. The measurements of both hadronic and leptonic system are important for effective choice of phase space. A different approach is to extend the measurements deeper into the $B \to X_c \ell \bar{\nu}$ region to reduce the theoretical uncertainties. Analyses of the electron-energy endpoint from CLEO [49], *BABAR* [50] and Belle [51] quote $B \to X_u e \bar{\nu}$ partial rates for $|\vec{p}_e| \geq 2.0\,\mathrm{GeV}$ and $1.9\,\mathrm{GeV}$, which are well below the charm endpoint. The large and pure B–$\overline{B}$ samples at the B factories permit the selection of $B \to X_u \ell \bar{\nu}$ decays in events where the other B is fully reconstructed [52]. With this full-reconstruction tag method the four-momenta of both the leptonic and the hadronic systems can be measured. It also gives access to a wider kinematic region due to improved signal purity.

To extract $|V_{ub}|$ from an exclusive channel, the form factors have to be known. Experimentally, better signal-to-background ratios are offset by smaller yields. The $B \to \pi \ell \bar{\nu}$ branching ratio is now known to 8%. The first unquenched lattice QCD calculations of the $B \to \pi \ell \bar{\nu}$ form factor for $q^2 > 16\,\mathrm{GeV}^2$ appeared recently [53,54]. Light-cone QCD sum rules are applicable for $q^2 < 14\,\mathrm{GeV}^2$ [55] and yield somewhat smaller $|V_{ub}|$, $(3.3^{+0.6}_{-0.4}) \times 10^{-3}$. The theoretical uncertainties in extracting $|V_{ub}|$ from inclusive and exclusive decays are different. A combination of the determinations is quoted by the V_{cb} and V_{ub} minireview as [44],

$$|V_{ub}| = (4.31 \pm 0.30) \times 10^{-3}, \tag{11.12}$$

which is dominated by the inclusive measurement. In the previous edition of the RPP [56] the average was reported as $|V_{ub}| = (3.67 \pm 0.47) \times 10^{-3}$, with an uncertainty around 13%. The new average is 17% larger, somewhat above the range favored by the measurement of $\sin 2\beta$ discussed below. Given the rapid theoretical and experimental progress, it will be interesting how the determination of $|V_{ub}|$ develops.

11.2.7. $|V_{td}|$ *and* $|V_{ts}|$:

The CKM elements $|V_{td}|$ and $|V_{ts}|$ cannot be measured from tree-level decays of the top quark, so one has to rely on determinations from B–$\overline{B}$ oscillations mediated by box diagrams or loop-mediated rare K and B decays. Theoretical uncertainties in hadronic effects limit the accuracy of the current determinations. These can be reduced by taking ratios of processes that are equal in the flavor $SU(3)$ limit to determine $|V_{td}/V_{ts}|$.

The mass difference of the two neutral B meson mass eigenstates is very well measured, $\Delta m_d = 0.507 \pm 0.004\,\mathrm{ps}^{-1}$ [57]. It is dominated by box diagrams with top quarks. Using the unquenched lattice QCD calculation of the B_d decay constant and bag parameter, $f_{B_d}\sqrt{\widehat{B}_{B_d}} = 244 \pm 11 \pm 24\,\mathrm{MeV}$ [58,59], and assuming $|V_{tb}| = 1$, one finds [60]

$$|V_{td}| = (7.4 \pm 0.8) \times 10^{-3}, \tag{11.13}$$

where the uncertainty is dominated by that of $f_{B_d}\sqrt{\widehat{B}_{B_d}}$. Several uncertainties are reduced in the lattice calculation of the ratio $\xi = (f_{B_s}\sqrt{B_{B_s}})/(f_{B_d}\sqrt{B_{B_d}}) =$

$1.21 \pm 0.04^{+0.04}_{-0.01}$, and therefore the constraint on $|V_{td}/V_{ts}|$ from Δm_d and Δm_s would be more reliable theoretically. So far only a lower bound is available, $\Delta m_s > 16.6\,\mathrm{ps}^{-1}$ at 95% CL [57], which implies $|V_{td}/V_{ts}| < 0.22$ at 95% CL. After completion of this review, new DØ and CDF analyses appeared, obtaining $17\,\mathrm{ps}^{-1} < \Delta m_s < 21\,\mathrm{ps}^{-1}$ at 90% CL [61], and $\Delta m_s = (17.31^{+0.33}_{-0.18} \pm 0.07)\,\mathrm{ps}^{-1}$ [62], respectively. These provide a new theoretically clean and significantly improved constraint, $|V_{td}/V_{ts}| = 0.208^{+0.008}_{+0.006}$ [62].

The inclusive rate $\mathcal{B}(B \to X_s\gamma) = (3.55 \pm 0.26) \times 10^{-4}$ for $E_\gamma > 1.6\,\mathrm{GeV}$ [57] is sensitive to $V_{tb}V_{ts}^*$ via t-quark penguins. However, a large contribution to the rate comes from the $b \to c\bar{c}s$ four-quark operators mixing into the electromagnetic dipole operator. Since any CKM determination from loop processes necessarily assumes the SM, using $V_{cb}V_{cs}^* \approx -V_{tb}V_{ts}^*$, we obtain $|V_{ts}| = (40.6 \pm 2.7) \times 10^{-3}$ [63].

The corresponding exclusive decays suffer from larger theoretical uncertainties due to unknown hadronic form factors, although the experimental accuracy is better. Recently, Belle observed exclusive $B \to (\rho/\omega)\gamma$ signals [64] in $b \to d\gamma$ decays. The theoretically cleanest determination of $|V_{td}/V_{ts}|$ in exclusive radiative B decays will come from $\mathcal{B}(B^0 \to \rho^0\gamma)/\mathcal{B}(B^0 \to K^{*0}\gamma)$, because the poorly known spectator interaction contribution is expected to be smaller in this ratio than that in charged B decay (W-exchange vs. weak annihilation). For now, we average the neutral and charged modes, assuming isospin symmetry, which gives $\mathcal{B}(B \to \rho\gamma)/\mathcal{B}(B \to K^*\gamma) = 0.017 \pm 0.006$ [57]. This ratio is $|V_{td}/V_{ts}|^2/\xi_\gamma^2$, where ξ_γ contains the hadronic physics, and is unity in the $SU(3)$ limit. Using $\xi_\gamma = 1.2 \pm 0.2$ [65], implies $|V_{td}/V_{ts}| = 0.16 \pm 0.04$, where we combined the experimental and theoretical errors in quadrature.

A theoretically clean determination of $|V_{td}V_{ts}^*|$ is possible from $K^+ \to \pi^+\nu\bar{\nu}$ decay [66]. Experimentally, only three events have been observed [67], and the rate is consistent with the SM with large uncertainties. Much more data are needed for a precision measurement.

11.2.8. $|V_{tb}|$:

The direct determination of $|V_{tb}|$ from top decays uses the ratio of branching fractions $R = \mathcal{B}(t \to Wb)/\mathcal{B}(t \to Wq) = |V_{tb}|^2/(\sum_q |V_{tq}|^2) = |V_{tb}|^2$, where $q = b, s, d$. The CDF and DØ measurements performed on data collected during Run II of the Tevatron give $R = 1.12^{+0.27}_{-0.23}$ [68] and $R = 1.03^{+0.19}_{-0.17}$ [69], respectively. Both measurements result in the 95% CL lower limit

$$|V_{tb}| > 0.78\,. \tag{11.14}$$

An attempt at constraining $|V_{tb}|$ from the precision electroweak data was made in [70]. The result, mostly driven by the top-loop contributions to $\Gamma(Z \to b\bar{b})$, gives $|V_{tb}| = 0.77^{+0.18}_{-0.24}$.

11.3. Phases of CKM elements

As can be seen from Fig. 11.1, the angles of the unitarity triangle are

$$\beta = \phi_1 = \arg\left(-\frac{V_{cd}V_{cb}^*}{V_{td}V_{tb}^*}\right),$$
$$\alpha = \phi_2 = \arg\left(-\frac{V_{td}V_{tb}^*}{V_{ud}V_{ub}^*}\right),$$
$$\gamma = \phi_3 = \arg\left(-\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*}\right). \tag{11.15}$$

Since CP violation involves phases of CKM elements, many measurements of CP-violating observables can be used to constraint these angles and the $\bar{\rho}, \bar{\eta}$ parameters.

11.3.1. ϵ and ϵ' :

The measurement of CP violation in K^0–$\overline{K}^0$ mixing, $|\epsilon| = (2.233 \pm 0.015) \times 10^{-3}$ [71], provides important information about the CKM matrix. In the SM [72]

$$|\epsilon| = \frac{G_F^2 f_K^2 m_K m_W^2}{12\sqrt{2}\,\pi^2 \Delta m_K}\, \widehat{B}_K \Big\{ \eta_c S(x_c) \mathrm{Im}[(V_{cs}V_{cd}^*)^2] + \eta_t S(x_t) \mathrm{Im}[(V_{ts}V_{td}^*)^2] + 2\eta_{ct} S(x_c, x_t) \mathrm{Im}(V_{cs}V_{cd}^* V_{ts}V_{td}^*) \Big\}, \tag{11.16}$$

where S is an Inami-Lim function [73], $x_q = m_q^2/m_W^2$, and η_i are perturbative QCD corrections. The constraint from ϵ in the $\bar{\rho}, \bar{\eta}$ plane are bounded by hyperbolas approximately. The dominant uncertainties are due to the bag parameter, for which we use $\widehat{B}_K = 0.79 \pm 0.04 \pm 0.09$ from lattice QCD [74], and the parametric uncertainty proportional to $\sigma(A^4)$ [*i.e.*, $\sigma(|V_{cb}|^4)$].

The measurement of ϵ' provides a qualitative test of the CKM mechanism because its nonzero experimental average, $\mathrm{Re}(\epsilon'/\epsilon) = (1.67 \pm 0.23) \times 10^{-3}$ [71], demonstrated the existence of direct CP violation, a prediction of the KM ansatz. While $\mathrm{Re}(\epsilon'/\epsilon) \propto \mathrm{Im}(V_{td}V_{ts}^*)$, this quantity cannot easily be used to extract CKM parameters, because the electromagnetic penguin contributions tend to cancel the gluonic penguins for large m_t [75], thereby significantly increasing the hadronic uncertainties. Most estimates [76,77] are in agreement with the observed value, indicating a positive value for $\bar{\eta}$. Progress in lattice QCD, in particular finite-volume calculations [78,79], may eventually provide a determination of the $K \to \pi\pi$ matrix elements.

11.3.2. β / ϕ_1 :

11.3.2.1. *Charmonium modes:*

CP violation measurements in B meson decays provide direct information on the angles of the unitarity triangle, shown in Fig. 11.1. These overconstraining measurements serve to improve the determination of the CKM elements or to reveal effects beyond the SM.

The time-dependent CP asymmetry of neutral B decays to a final state f common to B^0 and $\overline{B}^0$ is given by [80,81]

$$\mathcal{A}_f = \frac{\Gamma(\overline{B}^0(t) \to f) - \Gamma(B^0(t) \to f)}{\Gamma(\overline{B}^0(t) \to f) + \Gamma(B^0(t) \to f)} = S_f \sin(\Delta m\, t) - C_f \cos(\Delta m\, t), \tag{11.17}$$

where

$$S_f = \frac{2\,\mathrm{Im}\lambda_f}{1+|\lambda_f|^2}\,, \qquad C_f = \frac{1-|\lambda_f|^2}{1+|\lambda_f|^2}\,, \qquad \lambda_f = \frac{q}{p}\frac{\bar{A}_f}{A_f}\,. \tag{11.18}$$

Here q/p describes B^0–$\overline{B}^0$ mixing and, to a good approximation in the SM, $q/p = V_{tb}^* V_{td}/V_{tb}V_{td}^* = e^{-2i\beta+\mathcal{O}(\lambda^4)}$ in the usual phase convention. A_f ($\bar{A}_f$) is the amplitude of $B^0 \to f$ ($\overline{B}^0 \to f$) decay. If f is a CP eigenstate and amplitudes with one CKM phase dominate a decay, then $|A_f| = |\bar{A}_f|$, $C_f = 0$ and $S_f = \sin(\arg\lambda_f) = \eta_f \sin 2\phi$, where η_f is the CP eigenvalue of f and 2ϕ is the phase difference between the $B^0 \to f$ and $B^0 \to \overline{B}^0 \to f$ decay paths. A contribution of another amplitude to the decay with a different CKM phase makes the value of S_f sensitive to relative strong interaction phases between the decay amplitudes (it also makes $C_f \neq 0$ possible).

The $b \to c\bar{c}s$ decays to CP eigenstates ($B^0 \to$ charmonium $K^0_{S,L}$) are the theoretically cleanest examples, measuring $S_f = -\eta_f \sin 2\beta$. The $b \to sq\bar{q}$ penguin amplitudes have dominantly the same weak phase as the $b \to c\bar{c}s$ tree amplitude. Since only λ^2-suppressed penguin amplitudes introduce a new CP-violating phase, amplitudes with a single weak phase dominate these decays, and we expect $||\bar{A}/A| - 1| < 0.01$. The e^+e^- asymmetric-energy B-factory experiments *BABAR* [82] and Belle [83] provide precise measurements. The world average is [57]

$$\sin 2\beta = 0.687 \pm 0.032\,. \tag{11.19}$$

This measurement has a four-fold ambiguity in β, which can be resolved by a global fit mentioned below. Experimentally, the two-fold ambiguity $\beta \to \pi/2-\beta$ (but not $\beta \to \pi+\beta$) can be resolved by a time-dependent angular analysis of $B^0 \to J/\psi K^{*0}$ [84,85] or a time-dependent Dalitz plot analysis of $B^0 \to \overline{D}^0 h^0$ ($h^0 = \pi^0, \eta, \omega$) with $\overline{D}^0 \to K^0_S\pi^+\pi^-$ [86]. The latter gives the better sensitivity and disfavors the solutions with $\cos 2\beta < 0$ at the 97% CL, consistent with the global CKM fit result.

The $b \to c\bar{c}d$ mediated transitions, such as $B^0 \to J/\psi\pi^0$ and $B^0 \to D^{(*)+}D^{(*)-}$, also measure approximately $\sin 2\beta$. However, the dominant component of $b \to d$ penguin

amplitude has different CKM phase ($V_{tb}^*V_{td}$) than the tree amplitude ($V_{cb}^*V_{cd}$), and their magnitudes are of the same order in λ. Therefore, the effect of penguins could be large, resulting in $S_f \neq -\eta_f \sin 2\beta$ and $C_f \neq 0$. These decay modes have also been measured by *BABAR* and Belle. The world averages [57], $S_{J/\psi\pi^0} = -0.69 \pm 0.25$, $S_{D^+D^-} = -0.29 \pm 0.63$, and $S_{D^{*+}D^{*-}} = -0.75 \pm 0.23$, are consistent with $\sin 2\beta$ obtained from $B^0 \to$ charmonium K^0 decays, and the C_f's are also consistent with zero, although the uncertainties are sizable.

11.3.2.2. *Penguin dominated modes:*

The $b \to s\bar{q}q$ penguin dominated decays have the same CKM phase as the $b \to c\bar{c}s$ tree level decays, up to corrections suppressed by λ^2, since $V_{tb}^*V_{ts} = -V_{cb}^*V_{cs}[1 + \mathcal{O}(\lambda^2)]$. Therefore, decays such as $B^0 \to \phi K^0$ and $\eta' K^0$ provide $\sin 2\beta$ measurements in the SM. If new physics contributes to the $b \to s$ loop diagrams and has a different weak phase, it would give rise to $S_f \neq -\eta_f \sin 2\beta$ and possibly $C_f \neq 0$. Therefore, the main interest in these modes is not simply to measure $\sin 2\beta$, but to search for new physics. Measurements of many other decay modes in this category, such as $B \to \pi^0 K_S^0$, $K_S^0 K_S^0 K_S^0$, etc., have also been performed by *BABAR* and Belle. The results and their uncertainties are summarized in Fig. 12.3 and Table 12.1 of Ref. [81].

11.3.3. α / ϕ_2 :

Since α is the angle between $V_{tb}^*V_{td}$ and $V_{ub}^*V_{ud}$, only time-dependent CP asymmetries in $b \to u\bar{u}d$ dominated modes can directly measure $\sin 2\alpha$, in contrast to $\sin 2\beta$, where several different transitions can be used. Since $b \to d$ penguin amplitudes have a different CKM phase than $b \to u\bar{u}d$ tree amplitudes, and their magnitudes are of the same order in λ, the penguin contribution can be sizable and makes the α determination complicated. To date, α has been measured in $B \to \pi\pi$, $\rho\pi$ and $\rho\rho$ decay modes, among which $B \to \rho\rho$ has the best precision.

11.3.3.1. $B \to \pi\pi$*:*

It is now experimentally well established that there is a sizable contribution of $b \to d$ penguin amplitudes in $B \to \pi\pi$ decays. Then $S_{\pi^+\pi^-}$ in the time-dependent $B^0 \to \pi^+\pi^-$ analysis no longer measures $\sin 2\alpha$, but

$$S_{\pi^+\pi^-} = \sqrt{1 - C_{\pi^+\pi^-}^2}\, \sin(2\alpha + 2\Delta\alpha), \tag{11.20}$$

where $2\Delta\alpha$ is the phase difference between $e^{2i\gamma}\bar{A}_{\pi^+\pi^-}$ and $A_{\pi^+\pi^-}$. The value of $\Delta\alpha$, hence α, can be extracted using the isospin relations among the amplitudes of $B^0 \to \pi^+\pi^-$, $B^0 \to \pi^0\pi^0$, and $B^+ \to \pi^+\pi^0$ decays [87],

$$\frac{1}{\sqrt{2}} A_{\pi^+\pi^-} + A_{\pi^0\pi^0} - A_{\pi^+\pi^0} = 0, \tag{11.21}$$

and a similar one for the $\bar{A}_{\pi\pi}$'s. This method utilizes the fact that a pair of pions in $B \to \pi\pi$ decay must be in a zero angular momentum state, and then, because of

Bose statistics, they must have even isospin. Consequently, $\pi^0\pi^\pm$ is in a pure isospin-2 state, while the penguin amplitudes only contribute to the isospin-0 final state. The latter does not hold for the electroweak penguin amplitudes, but their effect is expected to be small. The isospin analysis uses the world averages [57] $S_{\pi^+\pi^-} = -0.50 \pm 0.12$, $C_{\pi^+\pi^-} = -0.37 \pm 0.10$, the branching fractions of all three modes, and the direct CP asymmetry $C_{\pi^0\pi^0} = -0.28^{+0.39}_{-0.40}$. This analysis leads to 16 mirror solutions for $0 \leq \alpha < 2\pi$. Because of this, and the sizable experimental error of the $B^0 \to \pi^0\pi^0$ rate and CP asymmetry, only a loose constraint on α can be obtained at present, $0^\circ < \alpha < 17^\circ$ and $73^\circ < \alpha < 180^\circ$ at 68% CL.

11.3.3.2. *$B \to \rho\rho$:*

The decay $B^0 \to \rho^+\rho^-$ contains two vector mesons in the final state, which in general is a mixture of CP-even and CP-odd components. Therefore, it was thought that extracting α from this mode would be complicated.

However, the longitudinal polarization fractions (f_L) in $B^+ \to \rho^+\rho^0$ and $B^0 \to \rho^+\rho^-$ decays were measured to be close to unity [88], which implies that the final states are almost purely CP-even. Furthermore, $\mathcal{B}(B^0 \to \rho^0\rho^0) < 1.1 \times 10^{-6}$ at 90% CL [89] is much smaller than $\mathcal{B}(B^0 \to \rho^+\rho^-) = (26.4^{+6.1}_{-6.4}) \times 10^{-6}$ and $\mathcal{B}(B^+ \to \rho^+\rho^0) = (26.2^{+3.6}_{-3.7}) \times 10^{-6}$, which implies that the effect of the penguin diagrams is small. The isospin analysis using the world averages [57], $S_{\rho^+\rho^-} = -0.22 \pm 0.22$ and $C_{\rho^+\rho^-} = -0.02 \pm 0.17$, and the above-mentioned branching fractions of $B \to \rho\rho$ modes gives $\alpha = (96 \pm 13)^\circ$ [90] with a mirror solution at $3\pi/2 - \alpha$. A possible small violation of Eq. (11.21)) due to the finite width of the ρ [91] is neglected.

11.3.3.3. *$B \to \rho\pi$:*

The final state in $B^0 \to \rho^+\pi^-$ decay is not a CP eigenstate, but this decay proceeds via the same quark level diagrams as $B^0 \to \pi^+\pi^-$ and both B^0 and $\overline{B}^0$ can decay to $\rho^+\pi^-$. Consequently, mixing induced CP violations can occur in four decay amplitudes, $B^0 \to \rho^\pm\pi^\mp$ and $\overline{B}^0 \to \rho^\pm\pi^\mp$. The measurements of CP violation parameters for these decays, where $B^0 \to \pi^+\pi^-\pi^0$ decays are treated as quasi-two-body decays, have been made both by *BABAR* [92] and the Belle [93]. However, the isospin analysis is rather complicated and no significant model independent constraint on α has been obtained yet.

The time-dependent Dalitz plot analysis of $B^0 \to \pi^+\pi^-\pi^0$ decays permits the extraction of α with a single discrete ambiguity, $\alpha \to \alpha + \pi$, since one knows the variation of the strong phases in the interference regions of the $\rho^+\pi^-$, $\rho^-\pi^+$, and $\rho^0\pi^0$ amplitudes in the Dalitz plot [94]. The *BABAR* measurement gives $\alpha = (113^{+27}_{-17} \pm 6)^\circ$ [95]. Although this constraint is moderate so far, it favors one of the mirror solutions in $B \to \rho^+\rho^-$.

Combining the above-mentioned three decay modes [90], α is constrained as

$$\alpha = (99^{+13}_{-8})^\circ. \tag{11.22}$$

A different statistical approach [96] gives similar constraint from the combination of these measurements.

11.3.4. γ / ϕ_3 :

By virtue of Eq. (11.15), γ does not depend on CKM elements involving the top quark, so it can be measured in tree level B decays. This is an important distinction from the measurements of α and β, and implies that the direct measurements of γ are unlikely to be affected by physics beyond the SM.

11.3.4.1. $B^\pm \to DK^\pm$:

The interference of $B^- \to D^0 K^-$ ($b \to c\bar{u}s$) and $B^- \to \overline{D}^0 K^-$ ($b \to u\bar{c}s$) transitions can be studied in final states accessible in both D^0 and $\overline{D}^0$ decays [80]. In principle, it is possible to extract the B and D decay amplitudes, the relative strong phases, and the weak phase γ from the data.

A practical complication is that the precision depends sensitively on the ratio of the interfering amplitudes

$$r_B = \left| A(B^- \to \overline{D}^0 K^-)/A(B^- \to D^0 K^-) \right|, \tag{11.23}$$

which is around $0.1 - 0.2$. The original GLW method [97,98] considers D decays to CP eigenstate final states, such as $B^\pm \to D^{(*)}_{CP}(\to \pi^+\pi^-)K^{\pm(*)}$. To alleviate the smallness of r_B and make the interfering amplitudes (which are products of the B and D decay amplitudes) comparable in magnitude, the ADS method [99] considers final states where Cabibbo-allowed $\overline{D}^0$ and doubly Cabibbo-suppressed D^0 decays interfere. Extensive measurements have been made by the B factories using both methods [100], but the constraints on γ are fairly weak so far.

It was realized that both D^0 and $\overline{D}^0$ have large branching fractions to certain three-body final states, such as $K_S\pi^+\pi^-$, and the analysis can be optimized by studying the Dalitz plot dependence of the interferences [101,102]. The best present determination of γ comes from this method. Belle [103] and *BABAR* [104] obtained $\gamma = 68^{+14}_{-15} \pm 13 \pm 11^\circ$ and $\gamma = 67 \pm 28 \pm 13 \pm 11^\circ$, respectively, where the last uncertainty is due to the D decay modeling. The error is sensitive to the central value of the amplitude ratio r_B (and r_B^* for the D^*K mode), for which Belle found somewhat larger central values than *BABAR*. The same values of $r_B^{(*)}$ enter the ADS analyses, and the data can be combined to fit for $r_B^{(*)}$ and γ. The possibility of D^0–$\overline{D}^0$ mixing has been neglected in all measurements, but its effect on γ is far below the present experimental accuracy [105], unless D^0–$\overline{D}^0$ mixing is due to CP-violating new physics, in which case it could be included in the analysis [106].

Combining the GLW, ADS, and Dalitz analyses [90], γ is constrained as

$$\gamma = (63^{+15}_{-12})^\circ. \tag{11.24}$$

The likelihood function of γ is not Gaussian, and the 95% CL range is $40^\circ < \gamma < 110^\circ$. Similar results are found in [96]. New results [107] reported after the completion of this review indicate a decrease of $r_B^{(*)}$, which leads to an increased error on γ.

11.3.4.2. $B \to D^{(*)\pm}\pi^{\mp}$*:*

The interference of $b \to u$ and $b \to c$ transitions can be studied in $\overline{B}^0 \to D^{(*)+}\pi^-$ ($b \to c\bar{u}d$) and $\overline{B}^0 \to B^0 \to D^{(*)+}\pi^-$ ($\bar{b} \to \bar{u}c\bar{d}$) decays and their CP conjugates, since both B^0 and $\overline{B}^0$ decay to $D^{(*)\pm}\pi^{\mp}$ (or $D^{\pm}\rho^{\mp}$, *etc.*). Since there are only tree and no penguin contributions to these decays, in principle, it is possible to extract from the four observable time-dependent rates the magnitudes of the two hadronic amplitudes, their relative strong phase, and the weak phase between the two decay paths, which is $2\beta + \gamma$.

A complication is that the ratio of the interfering amplitudes is very small, $r_{D\pi} = A(B^0 \to D^+\pi^-)/A(\overline{B}^0 \to D^+\pi^-) = \mathcal{O}(0.01)$ (and similarly for $r_{D^*\pi}$ and $r_{D\rho}$), and therefore it has not been possible to measure it. To obtain $2\beta + \gamma$, $SU(3)$ flavor symmetry and dynamical assumptions have been used to relate $A(\overline{B}^0 \to D^-\pi^+)$ to $A(\overline{B}^0 \to D_s^-\pi^+)$, so this measurement is not model independent at present. The amplitude ratio is larger in the analogous B_s decays to $D_s^{\pm}K^{\mp}$, where it will be possible to measure it and model independently extract $\gamma - 2\beta_s$ at LHCb, as proposed in Ref. [108]. Combining the $D^{\pm}\pi^{\mp}$ $D^{*\pm}\pi^{\mp}$ and $D^{\pm}\rho^{\mp}$ measurements [109] gives $\sin(2\beta+\gamma) = 0.8^{+0.18}_{-0.24}$, consistent with the previously discussed results for β and γ.

11.4. Global fit in the Standard Model

Using the independently measured CKM elements mentioned in the previous sections, the unitarity of the CKM matrix can be checked. We obtain $|V_{ud}|^2 + |V_{us}|^2 + |V_{ub}|^2 = 0.9992 \pm 0.0011$ (1st row), $|V_{cd}|^2 + |V_{cs}|^2 + |V_{cb}|^2 = 0.968 \pm 0.181$ (2nd row), and $|V_{ud}|^2 + |V_{cd}|^2 + |V_{td}|^2 = 1.001 \pm 0.005$ (1st column), respectively. For the second row, a more stringent check is obtained from the measurement of $\sum_{u,c,d,s,b} |V_{ij}|^2$ in Sec.11.2.4 minus the sum in the first row above: $|V_{cd}|^2 + |V_{cs}|^2 + |V_{cb}|^2 = 1.003 \pm 0.027$. These provide strong tests of the unitarity of the CKM matrix. The sum of the three angles of the unitarity triangle, $\alpha + \beta + \gamma = (184^{+20}_{-15})^\circ$, is also consistent with the SM expectation.

The CKM matrix elements can be most precisely determined by a global fit that uses all available measurements and imposes the SM constraints (*i.e.*, three generation unitarity). There are several approaches to combining the experimental data. CKMfitter [90,6] and Ref. [110] (which develops [111,112] further) use frequentist statistics with different presentations of the theoretical errors, while UTfit [96,113] uses a Bayesian approach. These approaches provide similar results [114].

The constraints implied by the unitarity of the three generation CKM matrix significantly reduce the allowed range of some of the CKM elements. The fit for the Wolfenstein parameters defined in Eq. (11.4) gives

$$\begin{aligned} \lambda &= 0.2272 \pm 0.0010\,, & A &= 0.818^{+0.007}_{-0.017}\,, \\ \bar{\rho} &= 0.221^{+0.064}_{-0.028}\,, & \bar{\eta} &= 0.340^{+0.017}_{-0.045}\,, \end{aligned} \tag{11.25}$$

These values are obtained using the method of Refs. [6,90], while using the prescription of Refs. [96,113] gives $\lambda = 0.2262 \pm 0.0014$, $A = 0.815 \pm 0.013$, $\bar{\rho} = 0.235 \pm 0.031$, and $\bar{\eta} = 0.349 \pm 0.020$. The allowed ranges of the magnitudes of all nine CKM elements are

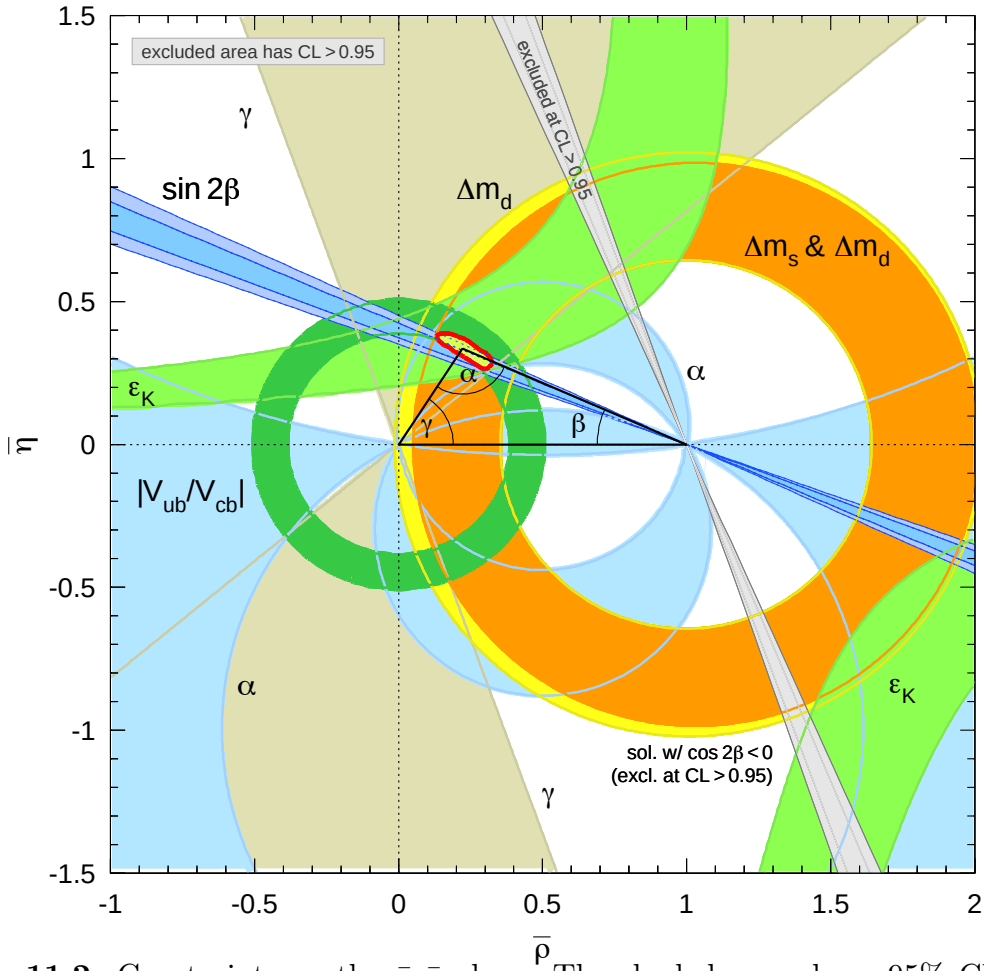


Figure 11.2: Constraints on the $\bar{\rho}, \bar{\eta}$ plane. The shaded areas have 95% CL. See full-color version on color pages at end of book.

$V_{\rm CKM} =$

$$\begin{pmatrix} 0.97383^{+0.00024}_{-0.00023} & 0.2272^{+0.0010}_{-0.0010} & (3.96^{+0.09}_{-0.09}) \times 10^{-3} \\ 0.2271^{+0.0010}_{-0.0010} & 0.97296^{+0.00024}_{-0.00024} & (42.21^{+0.10}_{-0.80}) \times 10^{-3} \\ (8.14^{+0.32}_{-0.64}) \times 10^{-3} & (41.61^{+0.12}_{-0.78}) \times 10^{-3} & 0.999100^{+0.000034}_{-0.000004} \end{pmatrix}, \tag{11.26}$$

and the Jarlskog invariant is $J = (3.08^{+0.16}_{-0.18}) \times 10^{-5}$.

Fig. 11.2 illustrates the constraints on the $\bar{\rho}, \bar{\eta}$ plane from various measurements and the global fit result. The shaded 95% CL regions all overlap consistently around the global fit region, though the consistency of $|V_{ub}/V_{cb}|$ and $\sin 2\beta$ is not very good, as mentioned previously.

11.5. Implications beyond the SM

If the constraints of the SM are lifted, K, B and D decays and mixings are described by many more parameters than just the four CKM parameters and the W, Z and quark masses. The most general effective Lagrangian at lowest order contains around a hundred flavor changing operators, and the observable effects of interactions at the weak-scale or above are encoded in their coefficients. For example, Δm_d, $\Gamma(B \to \rho\gamma)$, and $\Gamma(B \to X_d \ell^+ \ell^-)$ are all proportional to $|V_{td}V_{tb}^*|^2$ in the SM, however, they may receive unrelated contributions from new physics. Similar to the measurements of $\sin 2\beta$ from tree- and loop-dominated modes, such overconstraining measurements of the magnitudes and phases of CKM elements provide excellent sensitivity to new physics. Another very clean test of the SM can come from future measurements of $\mathcal{B}(K_L^0 \to \pi^0 \nu\bar{\nu})$ and $\mathcal{B}(K^+ \to \pi^+ \nu\bar{\nu})$. These loop induced decays are sensitive to new physics and will allow a determination of β independent of its value measured in B decays [115].

Not all CP-violating measurements can be interpreted as constraints on the $\bar{\rho}, \bar{\eta}$ plane. Besides the angles in Eq. (11.15), it is also useful to define $\beta_s = \arg(-V_{ts}V_{tb}^*/V_{cs}V_{cb}^*)$, which is the small, λ^2-suppressed, angle of a "squashed" unitarity triangle obtained by taking the scalar product of the second and third columns. It is the phase between B_s mixing and $b \to c\bar{c}s$ decays, and $\sin 2\beta_s$ can be measured via time-dependent CP violation in $B_s \to J/\psi\phi$, similar to $\sin 2\beta$ in $B \to J/\psi K^0$. Checking if β_s agrees with its SM prediction, $\sin 2\beta_s = 0.036 \pm 0.003$ [116,90], is an equally important test of the theory.

In the kaon sector, both CP-violating observables ϵ and ϵ' are tiny, so models in which all sources of CP violation are small were viable before the B-factory measurements. Since the measurement of $\sin 2\beta$ we know that CP violation can be an $\mathcal{O}(1)$ effect, and it is only flavor mixing that is suppressed between the three quark generations. Thus, many models with spontaneous CP violation are excluded. Model independent statements for the constraints imposed by the CKM measurements on new physics are hard to make, because most models that allow for new flavor physics contain a large number of new parameters. For example, the flavor sector of the MSSM contains 69 CP-conserving parameters and 41 CP-violating phases (*i.e.*, 40 new ones) [117].

In a large class of models the unitarity of the CKM matrix is maintained, and the dominant new physics effect is a contribution to the B^0–$\overline{B}^0$ mixing amplitude [118], which can be parameterized as $M_{12} = M_{12}^{\rm SM}\,(1 + h_d e^{i\sigma_d})$. While the constraints on h_d and σ_d are significant (before the measurements of γ and α they were not), new physics with a generic weak phase may still contribute to M_{12} at order 20% of the SM. Measurements unimportant for the SM CKM fit, such as the CP asymmetry in semileptonic decays, play a role in constraining such extensions of the SM [116]. Similar results for the constraints on new physics in K^0 and B_s^0 mixing are discussed in Refs. [113,119].

The CKM elements are fundamental parameters, so they should be measured as precisely as possible. The overconstraining measurements of CP asymmetries, mixing, semileptonic, and rare decays have started to severely constrain the magnitudes and phases of possible new physics contributions to flavor changing interactions. When new particles are observed at the LHC, it will be important to know the flavor parameters as precisely as possible to understand the underlying physics.

References:

1. N. Cabibbo, Phys. Rev. Lett. **10**, 531 (1963).
2. M. Kobayashi and T. Maskawa, Prog. Theor. Phys. **49**, 652 (1973).
3. F. J. Gilman *et al.*, Phys. Lett. **B592**, 793 (2004).
4. L. Wolfenstein, Phys. Rev. Lett. **51**, 1945 (1983).
5. A. J. Buras *et al.*, Phys. Rev. **D50**, 3433 (1994) [hep-ph/9403384].
6. J. Charles *et al.*, [CKMfitter Group], Eur. Phys. J. **C41**, 1 (2005) [hep-ph/0406184].
7. C. Jarlskog, Phys. Rev. Lett. **55**, 1039 (1985).
8. J. C. Hardy and I. S. Towner, Phys. Rev. Lett. **94**, 092502 (2005) [nucl-th/0412050].
9. G. Savard *et al.*, Phys. Rev. Lett. **95**, 102501 (2005).
10. E. Blucher and W.J. Marciano, "V_{ud}, V_{us}, the Cabibbo Angle and CKM Unitarity," in this *Review*.
11. D. Pocanic *et al.*, Phys. Rev. Lett. **93**, 181803 (2004) [hep-ex/0312030].
12. A. Sher *et al.*, Phys. Rev. Lett. **91**, 261802 (2003) [hep-ex/0305042].
13. T. Alexopoulos *et al.*, [KTeV Collaboration], Phys. Rev. Lett. **93**, 181802 (2004) [hep-ex/0406001]; Phys. Rev. **D70**, 092006 (2004) [hep-ex/0406002].
14. A. Lai *et al.*, [NA48 Collaboration], Phys. Lett. **B602**, 41 (2004) [hep-ex/0410059].
15. F. Ambrosino *et al.*, [KLOE Collaboration], [hep-ex/0508027].
16. A. Lai *et al.*, [NA48 Collaboration], Phys. Lett. **B604**, 1 (2004) [hep-ex/0410065];
    O. P. Yushchenko *et al.*, Phys. Lett. **B589**, 111 (2004) [hep-ex/0404030];
    T. Alexopoulos *et al.*, [KTeV Collaboration], Phys. Rev. **D70**, 092007 (2004) [hep-ex/0406003];
    F. Ambrosino *et al.*, [KLOE Collaboration], hep-ex/0601038.
17. F. Ambrosino *et al.*, [KLOE Collaboration], Phys. Lett. **B626**, 15 (2005) [hep-ex/0507088].
18. H. Leutwyler and M. Roos, Z. Phys. **C25**, 91 (1984).
19. E. Blucher *et al.*, hep-ph/0512039.
20. J. Bijnens and P. Talavera, Nucl. Phys. **B669**, 341 (2003) [hep-ph/0303103];
    M. Jamin *et al.*, JHEP **402**, 047 (2004) [hep-ph/0401080];
    V. Cirigliano *et al.*, JHEP **504**, 6 (2005) [hep-ph/0503108].
21. D. Becirevic *et al.*, Nucl. Phys. **B705**, 339 (2005) [hep-ph/0403217];
    C. Dawson *et al.*, PoS **LAT2005**, 337 (2005) [hep-lat/0510018];
    N. Tsutsui *et al.*, [JLQCD Collaboration], PoS **LAT2005**, 357 (2005) [hep-lat/0510068];
    M. Okamoto [Fermilab Lattice Collab.], hep-lat/0412044.
22. W. J. Marciano, Phys. Rev. Lett. **93**, 231803 (2004) [hep-ph/0402299].
23. F. Ambrosino *et al.*, [KLOE Collaboration], Phys. Lett. **B632**, 76 (2006) [hep-ex/0509045].
24. C. Bernard *et al.*, [MILC Collaboration], PoS **LAT2005**, 025 (2005) [hep-lat/0509137].
25. N. Cabibbo *et al.*, Ann. Rev. Nucl. and Part. Sci. **53**, 39 (2003) [hep-ph/0307298];
    Phys. Rev. Lett. **92**, 251803 (2004) [hep-ph/0307214].
26. M. Ademollo and R. Gatto, Phys. Rev. Lett. **13**, 264 (1964).

27. M. Davier *et al.*, hep-ph/0507078;
E. Gamiz *et al.*, Phys. Rev. Lett. **94**, 011803 (2005) [hep-ph/0408044].
28. C. Aubin *et al.*, [Fermilab Lattice, MILC, and HPQCD Collaboration], Phys. Rev. Lett. **94**, 011601 (2005) [hep-ph/0408306].
29. M. Artuso, hep-ex/0510052.
30. H. Abramowicz *et al.*, Z. Phys. **C15**, 19 (1982).
31. S. A. Rabinowitz *et al.*, Phys. Rev. Lett. **70**, 134 (1993).
32. A. O. Bazarko *et al.*, [CCFR Collaboration], Z. Phys. **C65**, 189 (1995) [hep-ex/9406007].
33. P. Vilain *et al.*, [CHARM II Collaboration], Eur. Phys. J. **C11**, 19 (1999).
34. G. D. Lellis *et al.*, Phys. Reports 399, 227 (2004) [Erratum-*ibid.* **411** (2005) 323].
35. N. Ushida *et al.*, [Fermilab E531 Collaboration], Phys. Lett. **B206**, 380 (1988).
36. T. Bolton, hep-ex/9708014.
37. A. Kayis-Topaksu *et al.*, [CHORUS Collaboration], Phys. Lett. **B626**, 24 (2005).
38. LEP W branching fraction results for this Review of Particle Physics, LEPEWWG/XSEC/2005-01, http://lepewwg.web.cern.ch/LEPEWWG/lepww/4f/Winter05/.
39. P. Abreu *et al.*, [DELPHI Collaboration], Phys. Lett. **B439**, 209 (1998).
40. I. I. Y. Bigi *et al.*, Phys. Rev. Lett. **71**, 496 (1993) [hep-ph/9304225].
41. A. V. Manohar and M. B. Wise, Phys. Rev. **D49**, 1310 (1994) [hep-ph/9308246].
42. I. I. Y. Bigi *et al.*, Phys. Rev. **D56**, 4017 (1997) [hep-ph/9704245].
43. A. H. Hoang *et al.*, Phys. Rev. **D59**, 074017 (1999) [hep-ph/9811239]; Phys. Rev. Lett. **82**, 277 (1999) [hep-ph/9809423].
44. R. Kowalewski and T. Mannel, "Determination of V_{cb} and V_{ub}," in this *Review*.
45. N. Isgur and M. B. Wise, Phys. Lett. **B237**, 527 (1990);
N. Isgur and M. B. Wise, Phys. Lett. **B232**, 113 (1989).
46. M. Neubert, Phys. Rev. **D49**, 3392 (1994) [hep-ph/9311325]; Phys. Rev. **D49**, 4623 (1994) [hep-ph/9312311].
47. I. I. Y. Bigi *et al.*, Int. J. Mod. Phys. **A9**, 2467 (1994) [hep-ph/9312359].
48. C. W. Bauer *et al.*, Phys. Lett. **B479**, 395 (2000) [hep-ph/0002161];
Phys. Rev. **D64**, 113004 (2001) [hep-ph/0107074].
49. A. Bornheim *et al.*, [CLEO Collaboration], Phys. Rev. Lett. **88**, 231803 (2002) [hep-ex/0202019].
50. B. Aubert *et al.*, [*BABAR* Collaboration], hep-ex/0408075.
51. A. Limosani *et al.*, [Belle Collaboration], Phys. Lett. **B621**, 28 (2005) [hep-ex/0504046].
52. I. Bizjak *et al.*, [Belle Collaboration], Phys. Rev. Lett. **95**, 241801 (2005) [hep-ex/0505088];
B. Aubert *et al.*, [*BABAR* Collaboration], Phys. Rev. Lett. **92**, 071802 (2004) [hep-ex/0307062]; hep-ex/0507017.
53. M. Okamoto *et al.*, Nucl. Phys. Proc. Supp. 140, 461 (2005) [hep-lat/0409116].
54. J. Shigemitsu *et al.*, Nucl. Phys. Proc. Suppl. 140, 464 (2005) [hep-lat/0408019].
55. P. Ball and R. Zwicky, Phys. Rev. **D71**, 014015 (2005) [hep-ph/0406232].

56. M. Battaglia and L. Gibbons, "Determination of $|V_{ub}|$," in S. Eidelman *et al.*, [Particle Data Group], Phys. Lett. **B592**, 793 (2004).
57. Heavy Flavor Averaging Group, E. Barberio *et al.*, hep-ex/0603003, and updates at:
http://www.slac.stanford.edu/xorg/hfag/.
58. A. Gray *et al.*, [HPQCD Collaboration], Phys. Rev. Lett. **95**, 212001 (2005) [hep-lat/0507015].
59. S. Aoki *et al.*, [JLQCD Collaboration], Phys. Rev. Lett. **91**, 212001 (2003) [hep-ph/0307039].
60. M. Okamoto, PoS **LAT2005**, 013 (2005) [hep-lat/0510113].
61. V. Abazov *et al.*, [DØ Collaboration], hep-ex/0603029.
62. A. Abulencia *et al.*, [CDF Collaboration], hep-ex/0606027.
63. A. J. Buras *et al.*, Nucl. Phys. **B631**, 219 (2002) [hep-ph/0203135];
M. Neubert, Eur. Phys. J. **C40**, 165 (2005) [hep-ph/0408179];
The quoted value of $|V_{ts}|$ is the average obtained using formulae in these papers.
64. K. Abe *et al.*, [Belle Collaboration], hep-ex/0506079.
65. J. Berryhill, CKM Workshop, UCSD, March 18, 2005,
http://ckm2005.ucsd.edu/agenda/fri3/berryhill.pdf;
B. Grinstein and D. Pirjol, Phys. Rev. **D62**, 093002 (2000) [hep-ph/0002216];
P. Ball and R. Zwicky, Phys. Rev. **D71**, 014029 (2005) [hep-ph/0412079];
A. Ali *et al.*, Phys. Lett. **B595**, 323 (2004) [hep-ph/0405075];
M. Beneke *et al.*, Nucl. Phys. **B612**, 25 (2001) [hep-ph/0106067];
S. W. Bosch and G. Buchalla, Nucl. Phys. **B621**, 459 (2002) [hep-ph/0106081];
Z. Ligeti and M. B. Wise, Phys. Rev. **D60**, 117506 (1999) [hep-ph/9905277];
D. Becirevic *et al.*, JHEP **305**, 7 (2003) [hep-lat/0301020].
66. A. J. Buras *et al.*, hep-ph/0508165.
67. V. V. Anisimovsky *et al.*, [E949 Collaboration], Phys. Rev. Lett. **93**, 031801 (2004) [hep-ex/0403036].
68. D. Acosta *et al.*, [CDF Collaboration], Phys. Rev. Lett. **95**, 102002 (2005) [hep-ex/0505091].
69. V. M. Abazov *et al.*, [DØ Collaboration], hep-ex/0603002.
70. J. Swain and L. Taylor, Phys. Rev. **D58**, 093006 (1998) [hep-ph/9712420].
71. "K_L^0 meson" particle listing, in this *Review*.
72. G. Buchalla *et al.*, Rev. Mod. Phys. **68**, 1125 (1996) [hep-ph/9512380].
73. T. Inami and C. S. Lim, Prog. Theor. Phys. **65**, 297 (1981) [Erratum-ibid. **65**, 1772 (1981)].
74. C. Dawson, PoS **LAT2005**, 007 (2005).
75. J. M. Flynn and L. Randall, Phys. Lett. **B224**, 221 (1989).
76. S. Bertolini *et al.*, Phys. Rev. **D63**, 056009 (2001) [hep-ph/0002234].
77. A. Pich, hep-ph/0410215.
78. L. Lellouch and M. Luscher, Commun. Math. Phys. **219**, 31 (2001) [hep-lat/0003023].
79. C. h. Kim *et al.*, Nucl. Phys. **B727**, 218 (2005) [hep-lat/0507006].

80. A.B. Carter and A.I. Sanda, Phys. Rev. Lett. **45**, 952 (1980); Phys. Rev. **D23**, 1567 (1981).
81. A more detailed discussion and references can be found in: D. Kirkby and Y. Nir, "CP violation in meson decays," in this *Review*.
82. B. Aubert *et al.*, [*BABAR* Collaboration], Phys. Rev. Lett. **94**, 161803 (2005) [hep-ex/0408127].
83. K. Abe *et al.*, [Belle Collaboration], hep-ex/0507037.
84. B. Aubert *et al.*, [*BABAR* Collaboration], Phys. Rev. **D71**, 032005 (2005) [hep-ex/0411016].
85. R. Itoh *et al.*, [Belle Collaboration], Phys. Rev. Lett. **95**, 091601 (2005) [hep-ex/0504030].
86. K Abe *et al.*, [Belle Collaboration], hep-ex/0507065.
87. M. Gronau and D. London, Phys. Rev. Lett. **65**, 3381 (1990).
88. J. Zhang *et al.*, [Belle Collaboration], Phys. Rev. Lett. **91**, 221801 (2003) [hep-ex/0306007]; A. Somov *et al.*, [Belle Collaboration], hep-ex/0601024; B. Aubert *et al.*, [*BABAR* Collaboration], Phys. Rev. Lett. **91**, 171802 (2003) [hep-ex/0307026]; Phys. Rev. **D69**, 031102 (2004) [hep-ex/0311017].
89. B. Aubert *et al.*, [*BABAR* Collaboration], Phys. Rev. Lett. **94**, 131801 (2005) [hep-ex/0412067].
90. A. Höcker *et al.*, Eur. Phys. J. **C21**, 225 (2001) [hep-ph/0104062]; See also Ref. [6] and updates at http://ckmfitter.in2p3.fr/.
91. A. F. Falk *et al.*, Phys. Rev. **D69**, 011502 (2004) [hep-ph/0310242].
92. B. Aubert *et al.*, [*BABAR* Collaboration], Phys. Rev. Lett. **91**, 201802 (2003) [hep-ex/0306030].
93. C. C. Wang *et al.*, [Belle Collaboration], Phys. Rev. Lett. **94**, 121801 (2005) [hep-ex/0408003].
94. H.R. Quinn and A.E. Snyder, Phys. Rev. **D48**, 2139 (1993).
95. B. Aubert *et al.*, [*BABAR* Collaboration], hep-ex/0408099.
96. M. Bona *et al.*, [UTfit Collaboration], JHEP **507**, 28 (2005) [hep-ph/0501199]; and updates at http://www.utfit.org/.
97. M. Gronau and D. London, Phys. Lett. **B253**, 483 (1991).
98. M. Gronau and D. Wyler, Phys. Lett. **B265**, 172 (1991).
99. D. Atwood *et al.*, Phys. Rev. Lett. **78**, 3257 (1997) [hep-ph/9612433]; Phys. Rev. **D63**, 036005 (2001) [hep-ph/0008090].
100. B. Aubert *et al.*, [*BABAR* Collaboration], Phys. Rev. **D71**, 031102 (2005) [hep-ex/0411091]; Phys. Rev. **D72**, 071103 (2005) [hep-ex/0507002]; Phys. Rev. **D72**, 032004 (2005) [hep-ex/0504047]; Phys. Rev. **D72**, 071104 (2005) [hep-ex/0508001]; Phys. Rev. **D73**, 051105 (2006) [hep-ex/0512067]; K. Abe *et al.*, [Belle Collaboration], Phys. Rev. **D73**, 051106 (2006) [hep-ex/0601032]; hep-ex/0307074; hep-ex/0508048.
101. A. Bondar, talk at the Belle analysis workshop, Novosibirsk, September 2002; A. Poluektov *et al.*, [Belle Collaboration], Phys. Rev. **D70**, 072003 (2004) [hep-ex/0406067].

102. A. Giri *et al.*, Phys. Rev. **D68**, 054018 (2003) [hep-ph/0303187].
103. K. Abe *et al.*, [Belle Collaboration], hep-ex/0411049; hep-ex/0504013.
104. B. Aubert *et al.*, [*BABAR* Collaboration], hep-ex/0507101.
105. Y. Grossman *et al.*, Phys. Rev. **D72**, 031501 (2005) [hep-ph/0505270].
106. A. Amorim *et al.*, Phys. Rev. **D59**, 056001 (1999) [hep-ph/9807364].
107. A. Poluektov *et al.*, [Belle Collaboration], hep-ex/0604054.
108. R. Aleksan *et al.*, Z. Phys. **C54**, 653 (1992).
109. B. Aubert *et al.*, [*BABAR* Collaboration], Phys. Rev. **D71**, 112003 (2005) [hep-ex/0504035]; hep-ex/0507075;
K. Abe *et al.*, [Belle Collaboration], Phys. Rev. Lett. **93**, 031802 (2004) [Erratum-ibid. **93**, 059901 (2004)] [hep-ex/0308048], T. Gershon *et al.*, [Belle Collaboration], Phys. Lett. **B624**, 11 (2005) [hep-ex/0408106].
110. G. P. Dubois-Felsmann, *et al.*, hep-ph/0308262.
111. "The *BABAR* physics book: Physics at an asymmetric B factory," (P. F. Harrison and H. R. Quinn, eds.), SLAC-R-0504, 1998.
112. S. Plaszczynski and M. H. Schune, hep-ph/9911280.
113. M. Bona *et al.*, [UTfit Collaboration], JHEP **603**, 080 (2006) [hep-ph/0509219].
114. We thank the CKMfitter and UTfit groups for performing fits and making plots using the inputs in this *Review*.
115. G. Buchalla and A. J. Buras, Phys. Lett. **B333**, 221 (1994) [hep-ph/9405259].
116. S. Laplace *et al.*, Phys. Rev. **D65**, 094040 (2002) [hep-ph/0202010].
117. H. E. Haber, Nucl. Phys. Proc. Supp. 62, 469 (1998) [hep-ph/9709450];
Y. Nir, hep-ph/0109090.
118. J. M. Soares and L. Wolfenstein, Phys. Rev. **D47**, 1021 (1993);
T. Goto *et al.*, Phys. Rev. **D53**, 6662 (1996) [hep-ph/9506311];
J. P. Silva and L. Wolfenstein, Phys. Rev. **D55**, 5331 (1997) [hep-ph/9610208].
119. K. Agashe *et al.*, hep-ph/0509117.

12. *CP* VIOLATION IN MESON DECAYS

Revised November 2005 by D. Kirkby (UC Irvine) and Y. Nir (Weizmann Inst.).

The CP transformation combines charge conjugation C with parity P. Under C, particles and antiparticles are interchanged, by conjugating all internal quantum numbers, *e.g.*, $Q \to -Q$ for electromagnetic charge. Under P, the handedness of space is reversed, $\vec{x} \to -\vec{x}$. Thus, for example, a left-handed electron e^-_L is transformed under CP into a right-handed positron, e^+_R.

If CP were an exact symmetry, the laws of Nature would be the same for matter and for antimatter. We observe that most phenomena are C- and P-symmetric, and therefore, also CP-symmetric. In particular, these symmetries are respected by the gravitational, electromagnetic, and strong interactions. The weak interactions, on the other hand, violate C and P in the strongest possible way. For example, the charged W bosons couple to left-handed electrons, e^-_L, and to their CP-conjugate right-handed positrons, e^+_R, but to neither their C-conjugate left-handed positrons, e^+_L, nor their P-conjugate right-handed electrons, e^-_R. While weak interactions violate C and P separately, CP is still preserved in most weak interaction processes. The CP symmetry is, however, violated in certain rare processes, as discovered in neutral K decays in 1964 [1], and recently observed in neutral B decays [2,3]. A K_L meson decays more often to $\pi^- e^+ \overline{\nu}_e$ than to $\pi^+ e^- \nu_e$, thus allowing electrons and positrons to be unambiguously distinguished, but the decay-rate asymmetry is only at the 0.003 level. The CP-violating effects observed in B decays are larger: the CP asymmetry in $B^0/\overline{B}^0$ meson decays to CP eigenstates like $J/\psi K_S$ is about 0.70. These effects are related to $K^0 - \overline{K}^0$ and $B^0 - \overline{B}^0$ mixing, but CP violation arising solely from decay amplitudes has also been observed, first in $K \to \pi\pi$ decays [4,5,6] and more recently in $B \to K\pi$ decays [7,8]. CP violation has not yet been observed in neutral D or B_s mesons, or in the lepton sector.

In addition to parity and to continuous Lorentz transformations, there is one other spacetime operation that could be a symmetry of the interactions: time reversal T, $t \to -t$. Violations of T symmetry have been observed in neutral K decays [9], and are expected as a corollary of CP violation if the combined CPT transformation is a fundamental symmetry of Nature [10]. All observations indicate that CPT is indeed a symmetry of Nature. Furthermore, one cannot build a Lorentz-invariant quantum field theory with a Hermitian Hamiltonian that violates CPT. (At several points in our discussion, we avoid assumptions about CPT, in order to identify cases where evidence for CP violation relies on assumptions about CPT.)

Within the Standard Model, CP symmetry is broken by complex phases in the Yukawa couplings (that is, the couplings of the Higgs scalar to quarks). When all manipulations to remove unphysical phases in this model are exhausted, one finds that there is a single CP-violating parameter [11]. In the basis of mass eigenstates, this single phase appears in the 3×3 unitary matrix that gives the W-boson couplings to an up-type antiquark and a down-type quark. (If the Standard Model is supplemented with Majorana mass terms for the neutrinos, the analogous mixing matrix for leptons has three CP-violating phases.) The beautifully consistent and economical Standard-Model description of CP violation in terms of Yukawa couplings, known as the Kobayashi-Maskawa (KM) mechanism [11], agrees with all measurements to date. In particular, one can account within this

framework for the three most accurately measured CP-violating observables, ϵ and ϵ' in neutral K decays, and $S_{\psi K}$ in neutral B decays. This agreement implies that the matrix of three-generation quark mixing is, very likely, the dominant source of CP violation in meson decays.

The small number of observations, and the theoretical uncertainties involved in their interpretation, however, leave room for additional sources of CP violation from new physics. Indeed, almost all extensions of the Standard Model imply that there are such additional sources. Moreover, CP violation is a necessary condition for baryogenesis, the process of dynamically generating the matter-antimatter asymmetry of the Universe [12]. Despite the phenomenological success of the KM mechanism, it fails (by several orders of magnitude) to accommodate the observed asymmetry [13]. This discrepancy strongly suggests that Nature provides additional sources of CP violation beyond the KM mechanism. (Recent evidence for neutrino masses implies that CP can be violated also in the lepton sector. This situation makes leptogenesis [14], a scenario where CP-violating phases in the Yukawa couplings of the neutrinos play a crucial role in the generation of the baryon asymmetry, a very attractive possibility.) The expectation of new sources motivates the large ongoing experimental effort to find deviations from the predictions of the KM mechanism.

CP violation can be experimentally searched for in a variety of processes, such as meson decays, electric dipole moments of neutrons, electrons and nuclei, and neutrino oscillations. Meson decays probe flavor-changing CP violation. The search for electric dipole moments may find (or constrain) sources of CP violation that, unlike the KM phase, are not related to flavor changing couplings. Future searches for CP violation in neutrino oscillations might provide further input on leptogenesis.

The present measurements of CP asymmetries provide some of the strongest constraints on the weak couplings of quarks. Future measurements of CP violation in K, D, B, and B_s meson decays will provide additional constraints on the flavor parameters of the Standard Model, and can probe new physics. In this review, we give the formalism and basic physics that are relevant to present and near future measurements of CP violation in meson decays.

Before going into details, we list here the independent CP violating observables where a signal has been established. Experiments have measured to date nine independent CP violating observables. The list of measured observables in B decays is somewhat conservative. We only include observables where the combined significance of Babar and Belle measurements, taking an inflated error in case of inconsistencies, is above 3σ. The list is:

1. Indirect CP violation in $K \to \pi\pi$ decays [1] and in $K \to \pi\ell\nu$ decays is given by

$$\varepsilon_K = (2.28 \pm 0.02) \times 10^{-3}\ e^{i\pi/4}. \tag{12.1}$$

2. Direct CP violation in $K \to \pi\pi$ decays [4,5,6] is given by

$$\varepsilon'/\varepsilon = (1.72 \pm 0.18) \times 10^{-3}. \tag{12.2}$$

3. CP violation in the interference of mixing and decay in the $B \to \psi K_S$ and other, related modes is given by [2,3]:

$$S_{\psi K_S} = +0.69 \pm 0.03. \tag{12.3}$$

4. CP violation in the interference of mixing and decay in the $B \to K^+K^-K_S$ mode is given by [15,16]

$$S_{K^+K^-K_S} = -0.45 \pm 0.13. \tag{12.4}$$

5. CP violation in the interference of mixing and decay in the $B \to D^{*+}D^{*-}$ mode is given by [17,18]

$$S_{D^{*+}D^{*-}} = -0.75 \pm 0.23. \tag{12.5}$$

6. CP violation in the interference of mixing and decay in the $B \to \eta' K^0$ modes is given by [19,20,21]

$$S_{\eta' K_S} = +0.50 \pm 0.09(0.13). \tag{12.6}$$

7. CP violation in the interference of mixing and decay in the $B \to f_0 K_S$ mode is given by [22,20]

$$S_{f_0 K_S} = -0.75 \pm 0.24. \tag{12.7}$$

8. Direct CP violation in the $\overline{B}^0 \to K^-\pi^+$ mode is given by [7,8]

$$\mathcal{A}_{K^\mp\pi^\pm} = -0.115 \pm 0.018. \tag{12.8}$$

9. Direct CP violation in the $B \to \rho\pi$ mode is given by [23,24]

$$\mathcal{A}_{\rho\pi}^{-+} = -0.48 \pm 0.14. \tag{12.9}$$

12.1. Formalism

The phenomenology of CP violation is superficially different in K, D, B, and B_s decays. This is primarily because each of these systems is governed by a different balance between decay rates, oscillations, and lifetime splitting. However, the underlying mechanisms of CP violation are identical for all pseudoscalar mesons.

In this section we present a general formalism for, and classification of, CP violation in the decay of a pseudoscalar meson M that might be a charged or neutral K, D, B, or B_s meson. Subsequent sections describe the CP-violating phenomenology, approximations, and alternate formalisms that are specific to each system.

12.1.1. ***Charged- and neutral-meson decays*** : We define decay amplitudes of M (which could be charged or neutral) and its CP conjugate $\overline{M}$ to a multi-particle final state f and its CP conjugate $\overline{f}$ as

$$A_f = \langle f\,|\mathcal{H}|\,M\rangle \quad , \quad \overline{A}_f = \langle f\,|\mathcal{H}|\,\overline{M}\rangle \quad ,$$
$$A_{\overline{f}} = \langle \overline{f}\,|\mathcal{H}|\,M\rangle \quad , \quad \overline{A}_{\overline{f}} = \langle \overline{f}\,|\mathcal{H}|\,\overline{M}\rangle \quad , \tag{12.10}$$

where $\mathcal{H}$ is the Hamiltonian governing weak interactions. The action of CP on these states introduces phases ξ_M and ξ_f that depend on their flavor content, according to

$$CP|M\rangle = e^{+i\xi_M}\,|\overline{M}\rangle \quad , \quad CP|f\rangle = e^{+i\xi_f}\,|\overline{f}\rangle\,, \tag{12.11}$$

with

$$CP|\overline{M}\rangle = e^{-i\xi_M}\,|M\rangle \quad , \quad CP|\overline{f}\rangle = e^{-i\xi_f}\,|f\rangle \tag{12.12}$$

so that $(CP)^2 = 1$. The phases ξ_M and ξ_f are arbitrary and unphysical because of the flavor symmetry of the strong interaction. If CP is conserved by the dynamics, $[CP, \mathcal{H}] = 0$, then A_f and $\overline{A}_{\overline{f}}$ have the same magnitude and an arbitrary unphysical relative phase

$$\overline{A}_{\overline{f}} = e^{i(\xi_f - \xi_M)}\,A_f\,. \tag{12.13}$$

12.1.2. ***Neutral-meson mixing*** : A state that is initially a superposition of M^0 and $\overline{M}^0$, say

$$|\psi(0)\rangle = a(0)|M^0\rangle + b(0)|\overline{M}^0\rangle\,, \tag{12.14}$$

will evolve in time acquiring components that describe all possible decay final states $\{f_1, f_2, \ldots\}$, that is,

$$|\psi(t)\rangle = a(t)|M^0\rangle + b(t)|\overline{M}^0\rangle + c_1(t)|f_1\rangle + c_2(t)|f_2\rangle + \cdots\,. \tag{12.15}$$

If we are interested in computing only the values of $a(t)$ and $b(t)$ (and not the values of all $c_i(t)$), and if the times t in which we are interested are much larger than the typical strong interaction scale, then we can use a much simplified formalism [25]. The simplified time evolution is determined by a 2×2 effective Hamiltonian $\mathbf{H}$ that is not Hermitian, since otherwise the mesons would only oscillate and not decay. Any complex matrix, such as $\mathbf{H}$, can be written in terms of Hermitian matrices $\mathbf{M}$ and $\mathbf{\Gamma}$ as

$$\mathbf{H} = \mathbf{M} - \frac{i}{2}\,\mathbf{\Gamma}\,. \tag{12.16}$$

$\mathbf{M}$ and $\mathbf{\Gamma}$ are associated with $(M^0, \overline{M}^0) \leftrightarrow (M^0, \overline{M}^0)$ transitions via off-shell (dispersive), and on-shell (absorptive) intermediate states, respectively. Diagonal elements of $\mathbf{M}$ and $\mathbf{\Gamma}$ are associated with the flavor-conserving transitions $M^0 \to M^0$ and $\overline{M}^0 \to \overline{M}^0$, while off-diagonal elements are associated with flavor-changing transitions $M^0 \leftrightarrow \overline{M}^0$.

The eigenvectors of $\mathbf{H}$ have well-defined masses and decay widths. To specify the components of the strong interaction eigenstates, M^0 and $\overline{M}^0$, in the light (M_L) and heavy (M_H) mass eigenstates, we introduce three complex parameters: p, q, and, for the case that both CP and CPT are violated in mixing, z:

$$\begin{aligned} |M_L\rangle &\propto p\sqrt{1-z}\,|M^0\rangle + q\sqrt{1+z}\,|\overline{M}^0\rangle \\ |M_H\rangle &\propto p\sqrt{1+z}\,|M^0\rangle - q\sqrt{1-z}\,|\overline{M}^0\rangle \ , \end{aligned} \tag{12.17}$$

with the normalization $|q|^2 + |p|^2 = 1$ when $z = 0$. (Another possible choice, which is in standard usage for K mesons, defines the mass eigenstates according to their lifetimes: K_S for the short-lived and K_L for the long-lived state. The K_L is experimentally found to be the heavier state.)

The real and imaginary parts of the eigenvalues $\omega_{L,H}$ corresponding to $|M_{L,H}\rangle$ represent their masses and decay-widths, respectively. The mass and width splittings are

$$\begin{aligned} \Delta m &\equiv m_H - m_L = \mathcal{R}e(\omega_H - \omega_L) \quad , \\ \Delta\Gamma &\equiv \Gamma_H - \Gamma_L = -2\,\mathcal{I}m(\omega_H - \omega_L) \ . \end{aligned} \tag{12.18}$$

Note that here Δm is positive by definition, while the sign of $\Delta\Gamma$ is to be experimentally determined. The sign of $\Delta\Gamma$ has not yet been established for D, B, and B_s mesons, while $\Delta\Gamma < 0$ is established for K mesons. The Standard Model predicts $\Delta\Gamma < 0$ for B and B_s mesons (for this reason, $\Delta\Gamma = \Gamma_L - \Gamma_H$, which is still a signed quantity, is often used in the B and B_s literature and is the convention used in the PDG experimental summaries). For D mesons, non-perturbative contributions are expected to dominate and, consequently, it is difficult to make a definitive prediction for the size and sign of $\Delta\Gamma$.

Solving the eigenvalue problem for $\mathbf{H}$ yields

$$\left(\frac{q}{p}\right)^2 = \frac{\mathbf{M}_{12}^* - (i/2)\mathbf{\Gamma}_{12}^*}{\mathbf{M}_{12} - (i/2)\mathbf{\Gamma}_{12}} \tag{12.19}$$

and

$$z \equiv \frac{\delta m - (i/2)\delta\Gamma}{\Delta m - (i/2)\Delta\Gamma} \ , \tag{12.20}$$

where

$$\delta m \equiv \mathbf{M}_{11} - \mathbf{M}_{22} \quad , \quad \delta\Gamma \equiv \mathbf{\Gamma}_{11} - \mathbf{\Gamma}_{22} \tag{12.21}$$

are the differences in effective mass and decay-rate expectation values for the strong interaction states M^0 and $\overline{M}^0$.

If either CP or CPT is a symmetry of $\mathbf{H}$ (independently of whether T is conserved or violated), then the values of δm and $\delta\Gamma$ are both zero, and hence $z = 0$. We also find that

$$\omega_H - \omega_L = 2\sqrt{\left(\mathbf{M}_{12} - \frac{i}{2}\mathbf{\Gamma}_{12}\right)\left(\mathbf{M}_{12}^* - \frac{i}{2}\mathbf{\Gamma}_{12}^*\right)} \ . \tag{12.22}$$

If either CP or T is a symmetry of $\mathbf{H}$ (independently of whether CPT is conserved or violated), then $\mathbf{\Gamma}_{12}/\mathbf{M}_{12}$ is real, leading to

$$\left(\frac{q}{p}\right)^2 = e^{2i\xi_M} \quad \Rightarrow \quad \left|\frac{q}{p}\right| = 1 \,, \tag{12.23}$$

where ξ_M is the arbitrary unphysical phase introduced in Eq. (12.12). If, and only if, CP is a symmetry of $\mathbf{H}$ (independently of CPT and T), then both of the above conditions hold, with the result that the mass eigenstates are orthogonal

$$\langle M_H | M_L \rangle = |p|^2 - |q|^2 = 0 \,. \tag{12.24}$$

12.1.3. ***CP-violating observables*** : All CP-violating observables in M and $\overline{M}$ decays to final states f and $\overline{f}$ can be expressed in terms of phase-convention-independent combinations of A_f, $\overline{A}_f$, $A_{\overline{f}}$, and $\overline{A}_{\overline{f}}$, together with, for neutral-meson decays only, q/p. CP violation in charged-meson decays depends only on the combination $|\overline{A}_{\overline{f}}/A_f|$, while CP violation in neutral-meson decays is complicated by $M^0 \leftrightarrow \overline{M}^0$ oscillations, and depends, additionally, on $|q/p|$ and on $\lambda_f \equiv (q/p)(\overline{A}_f/A_f)$.

The decay-rates of the two neutral K mass eigenstates, K_S and K_L, are different enough ($\Gamma_S/\Gamma_L \sim 500$) that one can, in most cases, actually study their decays independently. For neutral D, B, and B_s mesons, however, values of $\Delta\Gamma/\Gamma$ (where $\Gamma \equiv (\Gamma_H + \Gamma_L)/2$) are relatively small, and so both mass eigenstates must be considered in their evolution. We denote the state of an initially pure $|M^0\rangle$ or $|\overline{M}^0\rangle$ after an elapsed proper time t as $|M^0_{\rm phys}(t)\rangle$ or $|\overline{M}^0_{\rm phys}(t)\rangle$, respectively. Using the effective Hamiltonian approximation, but not assuming CPT is a good symmetry, we obtain

$$\begin{aligned}
|M^0_{\rm phys}(t)\rangle &= \left(g_+(t) + z\, g_-(t)\right)|M^0\rangle - \sqrt{1-z^2}\,\frac{q}{p}\; g_-(t)|\overline{M}^0\rangle, \\
|\overline{M}^0_{\rm phys}(t)\rangle &= \left(g_+(t) - z\, g_-(t)\right)|\overline{M}^0\rangle - \sqrt{1-z^2}\,\frac{p}{q}\; g_-(t)|M^0\rangle \,,
\end{aligned} \tag{12.25}$$

where

$$g_\pm(t) \equiv \frac{1}{2}\left(e^{-im_H t - \frac{1}{2}\Gamma_H t} \pm e^{-im_L t - \frac{1}{2}\Gamma_L t}\right) \tag{12.26}$$

and $z = 0$ if either CPT or CP is conserved.

Defining $x \equiv \Delta m/\Gamma$ and $y \equiv \Delta\Gamma/(2\Gamma)$, and assuming $z = 0$, one obtains the following time-dependent decay rates:

$$\frac{d\Gamma\left[M^0_{\rm phys}(t) \to f\right]/dt}{e^{-\Gamma t}\mathcal{N}_f} =$$

$$\left(|A_f|^2 + |(q/p)\overline{A}_f|^2\right)\cosh(y\Gamma t) + \left(|A_f|^2 - |(q/p)\overline{A}_f|^2\right)\cos(x\Gamma t)$$
$$+ 2\,\mathcal{R}e((q/p)A_f^*\overline{A}_f)\sinh(y\Gamma t) - 2\,\mathcal{I}m((q/p)A_f^*\overline{A}_f)\sin(x\Gamma t)\ , \tag{12.27}$$

$$\frac{d\Gamma\left[\overline{M}^0_{\rm phys}(t)\to f\right]/dt}{e^{-\Gamma t}\mathcal{N}_f} =$$
$$\left(|(p/q)A_f|^2 + |\overline{A}_f|^2\right)\cosh(y\Gamma t) - \left(|(p/q)A_f|^2 - |\overline{A}_f|^2\right)\cos(x\Gamma t)$$
$$+ 2\,\mathcal{R}e((p/q)A_f\overline{A}_f^*)\sinh(y\Gamma t) - 2\,\mathcal{I}m((p/q)A_f\overline{A}_f^*)\sin(x\Gamma t)\ , \tag{12.28}$$

where $\mathcal{N}_f$ is a common, time-independent, normalization factor. Decay rates to the CP-conjugate final state $\overline{f}$ are obtained analogously, with $\mathcal{N}_f = \mathcal{N}_{\overline{f}}$ and the substitutions $A_f \to A_{\overline{f}}$ and $\overline{A}_f \to \overline{A}_{\overline{f}}$ in Eqs. (12.27,12.28). Terms proportional to $|A_f|^2$ or $|\overline{A}_f|^2$ are associated with decays that occur without any net $M \leftrightarrow \overline{M}$ oscillation, while terms proportional to $|(q/p)\overline{A}_f|^2$ or $|(p/q)A_f|^2$ are associated with decays following a net oscillation. The $\sinh(y\Gamma t)$ and $\sin(x\Gamma t)$ terms of Eqs. (12.27,12.28) are associated with the interference between these two cases. Note that, in multi-body decays, amplitudes are functions of phase-space variables. Interference may be present in some regions but not others, and is strongly influenced by resonant substructure.

When neutral pseudoscalar mesons are produced coherently in pairs from the decay of a vector resonance, $V \to M^0\overline{M}^0$ (for example, $\Upsilon(4S) \to B^0\overline{B}^0$ or $\phi \to K^0\overline{K}^0$), the time-dependence of their subsequent decays to final states f_1 and f_2 has a similar form to Eqs. (12.27,12.28):

$$\frac{d\Gamma\left[V_{\rm phys}(t_1,t_2)\to f_1 f_2\right]/dt}{e^{-\Gamma|\Delta t|}\mathcal{N}_{f_1 f_2}} =$$
$$\left(|a_+|^2 + |a_-|^2\right)\cosh(y\Gamma\Delta t) + \left(|a_+|^2 - |a_-|^2\right)\cos(x\Gamma\Delta t)$$
$$- 2\,\mathcal{R}e(a_+^* a_-)\sinh(y\Gamma\Delta t) + 2\,\mathcal{I}m(a_+^* a_-)\sin(x\Gamma\Delta t)\ , \tag{12.29}$$

where $\Delta t \equiv t_2 - t_1$ is the difference in the production times, t_1 and t_2, of f_1 and f_2, respectively, and the dependence on the average decay time and on decay angles has been integrated out. The coefficients in Eq. (12.29) are determined by the amplitudes for no net oscillation from $t_1 \to t_2$, $\overline{A}_{f_1}A_{f_2}$ and $A_{f_1}\overline{A}_{f_2}$, and for a net oscillation, $(q/p)\overline{A}_{f_1}\overline{A}_{f_2}$ and $(p/q)A_{f_1}A_{f_2}$, via

$$a_+ \equiv \overline{A}_{f_1}A_{f_2} - A_{f_1}\overline{A}_{f_2}\ , \tag{12.30}$$
$$a_- \equiv\ -\sqrt{1-z^2}\left(\frac{q}{p}\overline{A}_{f_1}\overline{A}_{f_2} - \frac{p}{q}A_{f_1}A_{f_2}\right) + z\left(\overline{A}_{f_1}A_{f_2} + A_{f_1}\overline{A}_{f_2}\right)\ .$$

Assuming CPT conservation, $z = 0$, and identifying $\Delta t \to t$ and $f_2 \to f$, we find that Eqs. (12.29) and (12.30) reduce to Eq. (12.27) with $A_{f_1} = 0$, $\overline{A}_{f_1} = 1$, or to Eq. (12.28) with $\overline{A}_{f_1} = 0$, $A_{f_1} = 1$. Indeed, such a situation plays an important role in experiments. Final states f_1 with $A_{f_1} = 0$ or $\overline{A}_{f_1} = 0$ are called tagging states, because they identify the decaying pseudoscalar meson as, respectively, $\overline{M}^0$ or M^0. Before one of M^0 or $\overline{M}^0$ decays, they evolve in phase, so that there is always one M^0 and one $\overline{M}^0$ present. A tagging decay of one meson sets the clock for the time evolution of the other: it starts at t_1 as purely M^0 or $\overline{M}^0$, with time evolution that depends only on $t_2 - t_1$.

When f_1 is a state that both M^0 and $\overline{M}^0$ can decay into, then Eq. (12.29) contains interference terms proportional to $A_{f_1}\overline{A}_{f_1} \neq 0$ that are not present in Eqs. (12.27,12.28). Even when f_1 is dominantly produced by M^0 decays rather than $\overline{M}^0$ decays, or vice versa, $A_{f_1}\overline{A}_{f_1}$ can be non-zero owing to doubly-CKM-suppressed decays (with amplitudes suppressed by at least two powers of λ relative to the dominant amplitude, in the language of Section 12.3), and these terms should be considered for precision studies of CP violation in coherent $V \to M^0\overline{M}^0$ decays [26].

12.1.4. ***Classification of CP-violating effects*** : We distinguish three types of CP-violating effects in meson decays:

I. CP violation in decay is defined by

$$|\overline{A}_{\overline{f}}/A_f| \neq 1 \; . \tag{12.31}$$

In charged meson decays, where mixing effects are absent, this is the only possible source of CP asymmetries:

$$\mathcal{A}_{f^\pm} \equiv \frac{\Gamma(M^- \to f^-) - \Gamma(M^+ \to f^+)}{\Gamma(M^- \to f^-) + \Gamma(M^+ \to f^+)} = \frac{|\overline{A}_{f^-}/A_{f^+}|^2 - 1}{|\overline{A}_{f^-}/A_{f^+}|^2 + 1} \; . \tag{12.32}$$

II. CP (and T) violation in mixing is defined by

$$|q/p| \neq 1 \; . \tag{12.33}$$

In charged-current semileptonic neutral meson decays $M, \overline{M} \to \ell^\pm X$ (taking $|A_{\ell^+X}| = |\overline{A}_{\ell^-X}|$ and $A_{\ell^-X} = \overline{A}_{\ell^+X} = 0$, as is the case in the Standard Model, to lowest order in G_F, and in most of its reasonable extensions), this is the only source of CP violation, and can be measured via the asymmetry of "wrong-sign" decays induced by oscillations:

$$\begin{aligned} \mathcal{A}_{\rm SL}(t) &\equiv \frac{d\Gamma/dt\big[\overline{M}^0_{\rm phys}(t) \to \ell^+X\big] - d\Gamma/dt\big[M^0_{\rm phys}(t) \to \ell^-X\big]}{d\Gamma/dt\big[\overline{M}^0_{\rm phys}(t) \to \ell^+X\big] + d\Gamma/dt\big[M^0_{\rm phys}(t) \to \ell^-X\big]} \\ &= \frac{1 - |q/p|^4}{1 + |q/p|^4}. \end{aligned} \tag{12.34}$$

Note that this asymmetry of time-dependent decay rates is actually time-independent.

III. CP violation in interference between a decay without mixing, $M^0 \to f$, and a decay with mixing, $M^0 \to \overline{M}^0 \to f$ (such an effect occurs only in decays to final states that are common to M^0 and $\overline{M}^0$, including all CP eigenstates), is defined by

$$\mathcal{I}m(\lambda_f) \neq 0 \,, \tag{12.35}$$

with

$$\lambda_f \equiv \frac{q}{p}\frac{\overline{A}_f}{A_f} \,. \tag{12.36}$$

This form of CP violation can be observed, for example, using the asymmetry of neutral meson decays into final CP eigenstates f_{CP}

$$\mathcal{A}_{f_{CP}}(t) \equiv \frac{d\Gamma/dt\big[\overline{M}^0_{\rm phys}(t) \to f_{CP}\big] - d\Gamma/dt\big[M^0_{\rm phys}(t) \to f_{CP}\big]}{d\Gamma/dt\big[\overline{M}^0_{\rm phys}(t) \to f_{CP}\big] + d\Gamma/dt\big[M^0_{\rm phys}(t) \to f_{CP}\big]} \,. \tag{12.37}$$

If $\Delta\Gamma = 0$ and $|q/p| = 1$, as expected to a good approximation for B mesons, but not for K mesons, then $\mathcal{A}_{f_{CP}}$ has a particularly simple form (see Eq. (12.72), below). If, in addition, the decay amplitudes fulfill $|\overline{A}_{f_{CP}}| = |A_{f_{CP}}|$, the interference between decays with and without mixing is the only source of the asymmetry and $\mathcal{A}_{f_{CP}}(t) = \mathcal{I}m(\lambda_{f_{CP}})\sin(x\Gamma t)$.

Examples of these three types of CP violation will be given in Sections 12.4, 12.5, and 12.6.

12.2. Theoretical Interpretation: General Considerations

Consider the $M \to f$ decay amplitude A_f, and the CP conjugate process, $\overline{M} \to \overline{f}$, with decay amplitude $\overline{A}_{\overline{f}}$. There are two types of phases that may appear in these decay amplitudes. 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 $\overline{A}_{\overline{f}}$ with opposite signs. In the Standard Model, these phases occur only in the couplings of the $W^\pm$ bosons, and hence, 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. A second type of phase can appear in scattering or decay amplitudes, even when the Lagrangian is real. Their origin is the possible contribution from intermediate on-shell states in the decay process. Since these phases are generated by CP-invariant interactions, they are the same in A_f and $\overline{A}_{\overline{f}}$. Usually the dominant rescattering is due to strong interactions; hence the designation "strong phases" for the phase shifts so induced. Again, only the relative strong phases between different terms in the amplitude are physically meaningful.

The 'weak' and 'strong' phases discussed here appear in addition to the 'spurious' CP-transformation phases of Eq. (12.13). Those spurious phases are due to an arbitrary

choice of phase convention, and do not originate from any dynamics or induce any CP violation. For simplicity, we set them to zero from here on.

It is useful to write each contribution a_i to A_f in three parts: its magnitude $|a_i|$, its weak phase ϕ_i, and its strong phase δ_i. If, for example, there are two such contributions, $A_f = a_1 + a_2$, we have

$$\begin{aligned} A_f &= |a_1| e^{i(\delta_1+\phi_1)} + |a_2| e^{i(\delta_2+\phi_2)}, \\ \overline{A}_{\overline{f}} &= |a_1| e^{i(\delta_1-\phi_1)} + |a_2| e^{i(\delta_2-\phi_2)}. \end{aligned} \tag{12.38}$$

Similarly, for neutral meson decays, it is useful to write

$$\mathbf{M}_{12} = |\mathbf{M}_{12}| e^{i\phi_M} \quad , \quad \mathbf{\Gamma}_{12} = |\mathbf{\Gamma}_{12}| e^{i\phi_\Gamma} \ . \tag{12.39}$$

Each of the phases appearing in Eqs. (12.38,12.39) is convention-dependent, but combinations such as $\delta_1 - \delta_2$, $\phi_1 - \phi_2$, $\phi_M - \phi_\Gamma$, and $\phi_M + \phi_1 - \overline{\phi}_1$ (where $\overline{\phi}_1$ is a weak phase contributing to $\overline{A}_f$) are physical.

It is now straightforward to evaluate the various asymmetries in terms of the theoretical parameters introduced here. We will do so with approximations that are often relevant to the most interesting measured asymmetries.

1. The CP asymmetry in charged meson decays [Eq. (12.32)] is given by

$$\mathcal{A}_{f^\pm} = -\frac{2|a_1 a_2| \sin(\delta_2 - \delta_1) \sin(\phi_2 - \phi_1)}{|a_1|^2 + |a_2|^2 + 2|a_1 a_2| \cos(\delta_2 - \delta_1) \cos(\phi_2 - \phi_1)}. \tag{12.40}$$

The quantity of most interest to theory is the weak phase difference $\phi_2 - \phi_1$. Its extraction from the asymmetry requires, however, that the amplitude ratio $|a_2/a_1|$ and the strong phase difference $\delta_2 - \delta_1$ are known. Both quantities depend on non-perturbative hadronic parameters that are difficult to calculate.

2. In the approximation that $|\mathbf{\Gamma}_{12}/\mathbf{M}_{12}| \ll 1$ (valid for B and B_s mesons), the CP asymmetry in semileptonic neutral-meson decays [Eq. (12.34)] is given by

$$\mathcal{A}_{\rm SL} = -\left|\frac{\mathbf{\Gamma}_{12}}{\mathbf{M}_{12}}\right| \sin(\phi_M - \phi_\Gamma). \tag{12.41}$$

The quantity of most interest to theory is the weak phase $\phi_M - \phi_\Gamma$. Its extraction from the asymmetry requires, however, that $|\mathbf{\Gamma}_{12}/\mathbf{M}_{12}|$ is known. This quantity depends on long distance physics that is difficult to calculate.

3. In the approximations that only a single weak phase contributes to decay, $A_f = |a_f| e^{i(\delta_f + \phi_f)}$, and that $|\mathbf{\Gamma}_{12}/\mathbf{M}_{12}| = 0$, we obtain $|\lambda_f| = 1$, and the CP asymmetries in decays to a final CP eigenstate f [Eq. (12.37)] with eigenvalue $\eta_f = \pm 1$ are given by

$$\mathcal{A}_{f_{CP}}(t) = \mathcal{I}m(\lambda_f) \ \sin(\Delta m t) \ \text{ with } \ \mathcal{I}m(\lambda_f) = \eta_f \sin(\phi_M + 2\phi_f). \tag{12.42}$$

Note that the phase so measured is purely a weak phase, and no hadronic parameters are involved in the extraction of its value from $\mathcal{I}m(\lambda_f)$.

The discussion above allows us to introduce another classification of CP-violating effects:

1. *Indirect CP violation* is consistent with taking $\phi_M \neq 0$ and setting all other CP violating phases to zero. CP violation in mixing (type II) belongs to this class.
2. *Direct CP violation* cannot be accounted for by just $\phi_M \neq 0$. CP violation in decay (type I) belongs to this class.

As concerns type III CP violation, observing $\eta_{f_1}\mathcal{I}m(\lambda_{f_1}) \neq \eta_{f_2}\mathcal{I}m(\lambda_{f_2})$ (for the same decaying meson and two different final CP eigenstates f_1 and f_2) would establish direct CP violation. The significance of this classification is related to theory. In superweak models [27], CP violation appears only in diagrams that contribute to $\mathbf{M}_{12}$, hence they predict that there is no direct CP violation. In most models and, in particular, in the Standard Model, CP violation is both direct and indirect. The experimental observation of $\epsilon' \neq 0$ (see Section 12.4) excluded the superweak scenario.

12.3. Theoretical Interpretation: The KM Mechanism

Of all the Standard Model quark parameters, only the Kobayashi-Maskawa (KM) phase is CP violating. Having a single source of CP violation, the Standard Model is very predictive for CP asymmetries: some vanish, and those that do not are correlated.

To be precise, CP could be violated also by strong interactions. The experimental upper bound on the electric dipole moment of the neutron implies, however, that $\theta_{\rm QCD}$, the non-perturbative parameter that determines the strength of this type of CP violation, is tiny, if not zero. (The smallness of $\theta_{\rm QCD}$ constitutes a theoretical puzzle, known as 'the strong CP problem.') In particular, it is irrelevant to our discussion of meson decays.

The charged current interactions (that is, the $W^\pm$ interactions) for quarks are given by

$$-\mathcal{L}_{W^\pm} = \frac{g}{\sqrt{2}}\,\overline{u_{Li}}\,\gamma^\mu\,(V_{\rm CKM})_{ij}\,d_{Lj}\,W^+_\mu + \text{h.c.} \tag{12.43}$$

Here $i,j = 1,2,3$ are generation numbers. The Cabibbo-Kobayashi-Maskawa (CKM) mixing matrix for quarks is a 3×3 unitary matrix [28]. Ordering the quarks by their masses, *i.e.* $(u_1,u_2,u_3)\to(u,c,t)$ and $(d_1,d_2,d_3)\to(d,s,b)$, the elements of $V_{\rm CKM}$ are written as follows:

$$V_{\rm CKM} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}\,. \tag{12.44}$$

While a general 3×3 unitary matrix depends on three real angles and six phases, the freedom to redefine the phases of the quark mass eigenstates can be used to remove five of the phases, leaving a single physical phase, the Kobayashi-Maskawa phase, that is responsible for all CP violation in meson decays in the Standard Model.

The fact that one can parametrize $V_{\rm CKM}$ by three real and only one imaginary physical parameters can be made manifest by choosing an explicit parametrization. The Wolfenstein parametrization [29,30] is particularly useful:

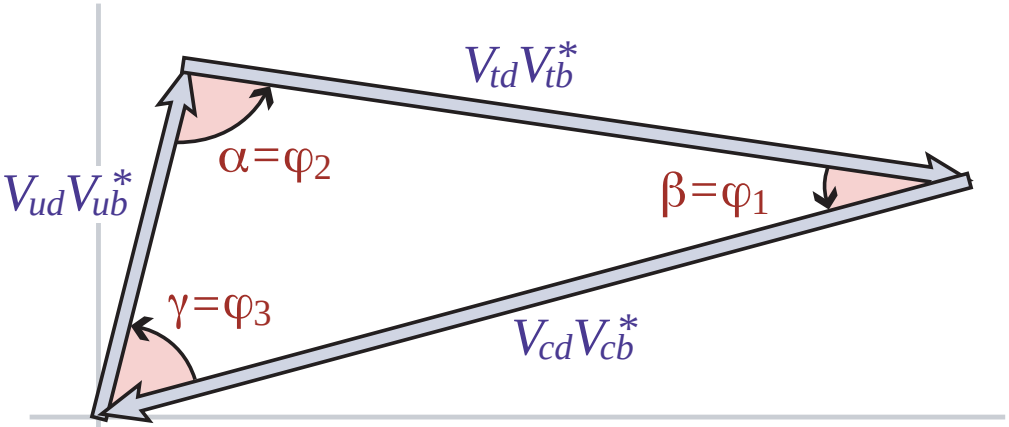


Figure 12.1: Graphical representation of the unitarity constraint $V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0$ as a triangle in the complex plane. See full-color version on color pages at end of book.

$$V_{\rm CKM} = \begin{pmatrix} 1-\frac{1}{2}\lambda^2-\frac{1}{8}\lambda^4 & \lambda & A\lambda^3(\rho-i\eta) \\ -\lambda+\frac{1}{2}A^2\lambda^5[1-2(\rho+i\eta)] & 1-\frac{1}{2}\lambda^2-\frac{1}{8}\lambda^4(1+4A^2) & A\lambda^2 \\ A\lambda^3[1-(1-\frac{1}{2}\lambda^2)(\rho+i\eta)] & -A\lambda^2+\frac{1}{2}A\lambda^4[1-2(\rho+i\eta)] & 1-\frac{1}{2}A^2\lambda^4 \end{pmatrix} . \tag{12.45}$$

Here $\lambda = |V_{us}| = 0.22$ (not to be confused with λ_f) plays the role of an expansion parameter, and η represents the CP violating phase. Terms of $\mathcal{O}(\lambda^6)$ were neglected.

The unitarity of the CKM matrix leads to various relations among the matrix elements; *e.g.*,

$$V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0 \, . \tag{12.46}$$

This relation requires the sum of three complex quantities to vanish and so can be geometrically represented in the complex plane as a triangle (see Fig. 12.1). The angles of this triangle,

$$\begin{aligned} \alpha \equiv \varphi_2 &\equiv \arg\left(-\frac{V_{td}V_{tb}^*}{V_{ud}V_{ub}^*}\right), \\ \beta \equiv \varphi_1 &\equiv \arg\left(-\frac{V_{cd}V_{cb}^*}{V_{td}V_{tb}^*}\right), \\ \gamma \equiv \varphi_3 &\equiv \arg\left(-\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*}\right), \end{aligned} \tag{12.47}$$

are physical quantities and can, in principle, be independently measured by CP asymmetries in B decays. The notations (α, β, γ) and $(\varphi_1, \varphi_2, \varphi_3)$ are both in common usage.

All unitarity triangles that correspond to relations, such as Eq. (12.46) between two different columns or two different rows of the CKM matrix have the same area, commonly denoted by $J/2$ [31]. If CP is violated, J is different from zero and can be taken as the single CP-violating parameter. In the Wolfenstein parametrization of Eq. (12.45), $J \simeq \lambda^6 A^2 \eta$.

12.4. *K* Decays

CP violation was discovered in $K \to \pi\pi$ decays in 1964 [1]. The same mode provided the first evidence for direct CP violation [4–6].

The decay amplitudes actually measured in neutral K decays refer to the mass eigenstates K_L and K_S, rather than to the K and $\overline{K}$ states referred to in Eq. (12.10). The final $\pi^+\pi^-$ and $\pi^0\pi^0$ states are CP-even. In the CP limit, $K_S(K_L)$ would be CP-even (odd) and therefore would (would not) decay to two pions. We define CP-violating amplitude ratios for two-pion final states,

$$\eta_{00} \equiv \frac{\langle \pi^0\pi^0|\mathcal{H}|K_L\rangle}{\langle \pi^0\pi^0|\mathcal{H}|K_S\rangle} \quad , \quad \eta_{+-} \equiv \frac{\langle \pi^+\pi^-|\mathcal{H}|K_L\rangle}{\langle \pi^+\pi^-|\mathcal{H}|K_S\rangle} \; . \tag{12.48}$$

Another important observable is the asymmetry of time-integrated semileptonic decay rates:

$$\delta_L \equiv \frac{\Gamma(K_L \to \ell^+\nu_\ell\pi^-) - \Gamma(K_L \to \ell^-\overline{\nu}_\ell\pi^+)}{\Gamma(K_L \to \ell^+\nu_\ell\pi^-) + \Gamma(K_L \to \ell^-\overline{\nu}_\ell\pi^+)} \; . \tag{12.49}$$

CP violation has been observed as an appearance of K_L decays to two-pion final states [32],

$$|\eta_{00}| = (2.276 \pm 0.014) \times 10^{-3} \qquad |\eta_{+-}| = (2.286 \pm 0.014) \times 10^{-3} \tag{12.50}$$

$$|\eta_{00}/\eta_{+-}| = 0.9950 \pm 0.0008 \; , \tag{12.51}$$

where the phase ϕ_{ij} of the amplitude ratio η_{ij} has been determined both assuming CPT invariance:

$$\begin{aligned} \phi_{00} &= (43.49 \pm 0.06)^\circ \\ \phi_{+-} &= (43.51 \pm 0.05)^\circ \\ \phi_{00} - \phi_{+-} &= (-0.022 \pm 0.020)^\circ \end{aligned} \tag{12.52}$$

and without assuming CPT invariance:

$$\begin{aligned} \phi_{00} &= (43.7 \pm 0.8)^\circ \\ \phi_{+-} &= (43.4 \pm 0.7)^\circ \\ \phi_{00} - \phi_{+-} &= (0.2 \pm 0.4)^\circ \; . \end{aligned} \tag{12.53}$$

CP violation has also been observed in semileptonic K_L decays [32]

$$\delta_L = (3.27 \pm 0.12) \times 10^{-3} \; , \tag{12.54}$$

where δ_L is a weighted average of muon and electron measurements, as well as in K_L decays to $\pi^+\pi^-\gamma$ and $\pi^+\pi^-e^+e^-$ [32]. CP violation in $K \to 3\pi$ decays has not yet been observed [32,33].

Historically, CP violation in neutral K decays has been described in terms of parameters ϵ and ϵ'. The observables η_{00}, η_{+-}, and δ_L are related to these parameters, and to those of Section 12.1, by

$$\begin{aligned} \eta_{00} = \frac{1-\lambda_{\pi^0\pi^0}}{1+\lambda_{\pi^0\pi^0}} &= \epsilon - 2\epsilon' \,, \\ \eta_{+-} = \frac{1-\lambda_{\pi^+\pi^-}}{1+\lambda_{\pi^+\pi^-}} &= \epsilon + \epsilon' \,, \\ \delta_L = \frac{1-|q/p|^2}{1+|q/p|^2} &= \frac{2\mathcal{R}e(\epsilon)}{1+|\epsilon|^2} \,, \end{aligned} \tag{12.55}$$

where, in the last line, we have assumed that $\left|A_{\ell^+\nu_\ell\pi^-}\right| = \left|\overline{A}_{\ell^-\overline{\nu}_\ell\pi^+}\right|$ and $\left|A_{\ell^-\overline{\nu}_\ell\pi^+}\right| = \left|\overline{A}_{\ell^+\nu_\ell\pi^-}\right| = 0$. (The convention-dependent parameter $\tilde{\epsilon} \equiv (1-q/p)/(1+q/p)$, sometimes used in the literature, is, in general, different from ϵ but yields a similar expression, $\delta_L = 2\mathcal{R}e(\tilde{\epsilon})/(1+|\tilde{\epsilon}|^2)$.) A fit to the $K \to \pi\pi$ data yields [32]

$$\begin{aligned} |\epsilon| &= (2.284 \pm 0.014) \times 10^{-3} \,, \\ \mathcal{R}e(\epsilon'/\epsilon) &= (1.67 \pm 0.26) \times 10^{-3} \,. \end{aligned} \tag{12.56}$$

In discussing two-pion final states, it is useful to express the amplitudes $A_{\pi^0\pi^0}$ and $A_{\pi^+\pi^-}$ in terms of their isospin components via

$$\begin{aligned} A_{\pi^0\pi^0} &= \sqrt{\frac{1}{3}}\,|A_0|\,e^{i(\delta_0+\phi_0)} - \sqrt{\frac{2}{3}}\,|A_2|\,e^{i(\delta_2+\phi_2)}, \\ A_{\pi^+\pi^-} &= \sqrt{\frac{2}{3}}\,|A_0|\,e^{i(\delta_0+\phi_0)} + \sqrt{\frac{1}{3}}\,|A_2|\,e^{i(\delta_2+\phi_2)} \,, \end{aligned} \tag{12.57}$$

where we parameterize the amplitude $A_I(\overline{A}_I)$ for $K^0(\overline{K}^0)$ decay into two pions with total isospin $I = 0$ or 2 as

$$\begin{aligned} A_I &\equiv \langle(\pi\pi)_I\,|\mathcal{H}|\,K^0\rangle = |A_I|\,e^{i(\delta_I+\phi_I)} \,, \\ \overline{A}_I &\equiv \langle(\pi\pi)_I\,|\mathcal{H}|\,\overline{K}^0\rangle = |A_I|\,e^{i(\delta_I-\phi_I)} \,. \end{aligned} \tag{12.58}$$

The smallness of $|\eta_{00}|$ and $|\eta_{+-}|$ allows us to approximate

$$\epsilon \simeq \frac{1}{2}(1-\lambda_{(\pi\pi)_{I=0}}), \qquad \epsilon' \simeq \frac{1}{6}\left(\lambda_{\pi^0\pi^0} - \lambda_{\pi^+\pi^-}\right) \,. \tag{12.59}$$

The parameter ϵ represents indirect CP violation, while ϵ' parameterizes direct CP violation: $\mathcal{R}e(\epsilon')$ measures CP violation in decay (type I), $\mathcal{R}e(\epsilon)$ measures CP violation in mixing (type II), and $\mathcal{I}m(\epsilon)$ and $\mathcal{I}m(\epsilon')$ measure the interference between decays with and without mixing (type III).

The following expressions for ϵ and ϵ' are useful for theoretical evaluations:

$$\epsilon \simeq \frac{e^{i\pi/4}}{\sqrt{2}} \frac{\mathcal{I}m(\mathbf{M}_{12})}{\Delta m}, \qquad \epsilon' = \frac{i}{\sqrt{2}} \left|\frac{A_2}{A_0}\right| e^{i(\delta_2-\delta_0)} \sin(\phi_2 - \phi_0). \tag{12.60}$$

The expression for ϵ is only valid in a phase convention where $\phi_2 = 0$, corresponding to a real $V_{ud}V_{us}^*$, and in the approximation that also $\phi_0 = 0$. The phase of ϵ, $\arg(\epsilon) \approx \arctan(-2\Delta m/\Delta\Gamma)$, is independent of the electroweak model and is experimentally determined to be about $\pi/4$. The calculation of ϵ benefits from the fact that $\mathcal{I}m(\mathbf{M}_{12})$ is dominated by short distance physics. Consequently, the main source of uncertainty in theoretical interpretations of ϵ are the values of matrix elements, such as $\langle K^0 | (\overline{s}d)_{V-A} (\overline{s}d)_{V-A} | \overline{K}^0 \rangle$. The expression for ϵ' is valid to first order in $|A_2/A_0| \sim 1/20$. The phase of ϵ' is experimentally determined, $\pi/2 + \delta_2 - \delta_0 \approx \pi/4$, and is independent of the electroweak model. Note that, accidentally, ϵ'/ϵ is real to a good approximation.

A future measurement of much interest is that of CP violation in the rare $K \to \pi\nu\overline{\nu}$ decays. The signal for CP violation is simply observing the $K_L \to \pi^0\nu\overline{\nu}$ decay. The effect here is that of interference between decays with and without mixing (type III) [34]:

$$\frac{\Gamma(K_L \to \pi^0\nu\overline{\nu})}{\Gamma(K^+ \to \pi^+\nu\overline{\nu})} = \frac{1}{2}\left[1 + |\lambda_{\pi\nu\overline{\nu}}|^2 - 2\,\mathcal{R}e(\lambda_{\pi\nu\overline{\nu}})\right] \simeq 1 - \mathcal{R}e(\lambda_{\pi\nu\overline{\nu}}), \tag{12.61}$$

where in the last equation we neglect CP violation in decay and in mixing (expected, model-independently, to be of order 10^{-5} and 10^{-3}, respectively). Such a measurement would be experimentally very challenging and theoretically very rewarding [35]. Similar to the CP asymmetry in $B \to J/\psi K_S$, the CP violation in $K \to \pi\nu\overline{\nu}$ decay is predicted to be large (that is, the ratio in Eq. (12.61) is neither CKM- nor loop-suppressed) and can be very cleanly interpreted.

Within the Standard Model, the $K_L \to \pi^0\nu\overline{\nu}$ decay is dominated by an intermediate top quark contribution and, consequently, can be interpreted in terms of CKM parameters [36]. (For the charged mode, $K^+ \to \pi^+\nu\overline{\nu}$, the contribution from an intermediate charm quark is not negligible, and constitutes a source of hadronic uncertainty.) In particular, $\mathrm{B}(K_L \to \pi^0\nu\overline{\nu})$ provides a theoretically clean way to determine the Wolfenstein parameter η [37]:

$$\mathrm{B}(K_L \to \pi^0\nu\overline{\nu}) = \kappa_L X^2(m_t^2/m_W^2) A^4\eta^2, \tag{12.62}$$

where $\kappa_L = 1.80 \times 10^{-10}$ incorporates the value of the four-fermion matrix element which is deduced, using isospin relations, from $\mathrm{B}(K^+ \to \pi^0 e^+\nu)$, and $X(m_t^2/m_W^2)$ is a known function of the top mass.

12.5. *D* Decays

Unlike the case of neutral K, B, and B_s mixing, D^0–$\overline{D}^0$ mixing has not yet been observed [38]. Long-distance contributions make it difficult to calculate the Standard Model prediction for the D^0–$\overline{D}^0$ mixing parameters. Therefore, the goal of the search for D^0–$\overline{D}^0$ mixing is not to constrain the CKM parameters, but rather to probe new physics. Here CP violation plays an important role. Within the Standard Model, the CP-violating effects are predicted to be negligibly small, since the mixing and the relevant decays are described, to an excellent approximation, by physics of the first two generations. Observation of CP violation in D^0–$\overline{D}^0$ mixing (at a level much higher than $\mathcal{O}(10^{-3})$) will constitute an unambiguous signal of new physics. At present, the most sensitive searches involve the $D \to K^+K^-$ and $D \to K^\pm\pi^\mp$ modes.

The neutral D mesons decay via a singly-Cabibbo-suppressed transition to the CP eigenstate K^+K^-. Since the decay proceeds via a Standard-Model tree diagram, it is very likely unaffected by new physics and, furthermore, dominated by a single weak phase. It is safe then to assume that direct CP violation plays no role here. In addition, given the experimental bounds [39], $x \equiv \Delta m/\Gamma \lesssim 0.03$ and $y \equiv \Delta\Gamma/(2\Gamma) = 0.0045 \pm 0.0065$, we can expand the decay rates to first order in these parameters. Using Eq. (12.27) with these assumptions and approximations yields, for $xt, yt \lesssim \Gamma^{-1}$,

$$\begin{aligned}
&\Gamma[D^0_{\rm phys}(t) \to K^+K^-] \\
&\qquad = \; e^{-\Gamma t}\,|A_{KK}|^2\,\left[1 - |q/p|\,(y\cos\phi_D - x\sin\phi_D)\Gamma t\right] \;, \\
&\Gamma[\overline{D}^0_{\rm phys}(t) \to K^+K^-] \\
&\qquad = \; e^{-\Gamma t}\,|A_{KK}|^2\,\left[1 - |p/q|\,(y\cos\phi_D + x\sin\phi_D)\Gamma t\right] \;,
\end{aligned} \tag{12.63}$$

where ϕ_D is defined via $\lambda_{K^+K^-} = -\,|q/p|\,e^{i\phi_D}$. (In the limit of CP conservation, choosing $\phi_D = 0$ is equivalent to defining the mass eigenstates by their CP eigenvalue: $|D_\mp\rangle = p|D^0\rangle \pm q|\overline{D}^0\rangle$, with $D_-(D_+)$ being the CP-odd (CP-even) state; that is, the state that does not (does) decay into K^+K^-.) Given the small values of x and y, the time dependences of the rates in Eq. (12.63) can be recast into purely exponential forms, but with modified decay-rate parameters [40]:

$$\begin{aligned}
\Gamma_{D^0\to K^+K^-} &= \Gamma \times \left[1 + |q/p|\,(y\cos\phi_D - x\sin\phi_D)\right] \;, \\
\Gamma_{\overline{D}^0\to K^+K^-} &= \Gamma \times \left[1 + |p/q|\,(y\cos\phi_D + x\sin\phi_D)\right] \;.
\end{aligned} \tag{12.64}$$

One can define CP-conserving and CP-violating combinations of these two observables (normalized to the true width Γ):

$$\begin{aligned}
Y &\equiv \frac{\Gamma_{\overline{D}^0\to K^+K^-} + \Gamma_{D^0\to K^+K^-}}{2\Gamma} - 1 \\
&= \frac{|q/p| + |p/q|}{2}\,y\cos\phi_D - \frac{|q/p| - |p/q|}{2}\,x\sin\phi_D \;,
\end{aligned}$$

$$\begin{aligned}\Delta Y &\equiv \frac{\Gamma_{\overline{D}^0\to K^+K^-} - \Gamma_{D^0\to K^+K^-}}{2\Gamma} \\ &= \frac{|q/p| + |p/q|}{2} x \sin\phi_D - \frac{|q/p| - |p/q|}{2} y \cos\phi_D \ .\end{aligned} \tag{12.65}$$

In the limit of CP conservation (and, in particular, within the Standard Model), $Y = y$ and $\Delta Y = 0$.

The $K^\pm\pi^\mp$ states are not CP eigenstates, but they are still common final states for D^0 and $\overline{D}^0$ decays. Since $D^0(\overline{D}^0) \to K^-\pi^+$ is a Cabibbo-favored (doubly-Cabibbo-suppressed) process, these processes are particularly sensitive to x and/or $y = \mathcal{O}(\lambda^2)$. Taking into account that $\left|\lambda_{K^-\pi^+}\right|, \left|\lambda^{-1}_{K^+\pi^-}\right| \ll 1$ and $x, y \ll 1$, assuming that there is no direct CP violation (again, these are Standard Model tree level decays dominated by a single weak phase), and expanding the time-dependent rates for $xt, yt \lesssim \Gamma^{-1}$, one obtains

$$\begin{aligned}&\frac{\Gamma[D^0_{\rm phys}(t) \to K^+\pi^-]}{\Gamma[\overline{D}^0_{\rm phys}(t) \to K^+\pi^-]} \\ &= \ r_d^2 + r_d \left|\frac{q}{p}\right| (y'\cos\phi_D - x'\sin\phi_D)\Gamma t + \left|\frac{q}{p}\right|^2 \frac{y^2+x^2}{4}(\Gamma t)^2 \ , \\ &\frac{\Gamma[\overline{D}^0_{\rm phys}(t) \to K^-\pi^+]}{\Gamma[D^0_{\rm phys}(t) \to K^-\pi^+]} \\ &= \ r_d^2 + r_d \left|\frac{p}{q}\right| (y'\cos\phi_D + x'\sin\phi_D)\Gamma t + \left|\frac{p}{q}\right|^2 \frac{y^2+x^2}{4}(\Gamma t)^2 \ ,\end{aligned} \tag{12.66}$$

where

$$\begin{aligned}y' &\equiv y\cos\delta - x\sin\delta \ , \\ x' &\equiv x\cos\delta + y\sin\delta \ .\end{aligned} \tag{12.67}$$

The weak phase ϕ_D is the same as that of Eq. (12.63) (a consequence of the absence of direct CP violation), δ is a strong phase difference for these processes, and $r_d = \mathcal{O}(\tan^2\theta_c)$ is the amplitude ratio, $r_d = \left|\overline{A}_{K^-\pi^+}/A_{K^-\pi^+}\right| = \left|A_{K^+\pi^-}/\overline{A}_{K^+\pi^-}\right|$, that is, $\lambda_{K^-\pi^+} = r_d(q/p)e^{-i(\delta-\phi_D)}$ and $\lambda^{-1}_{K^+\pi^-} = r_d(p/q)e^{-i(\delta+\phi_D)}$. By fitting to the six coefficients of the various time-dependences, one can extract r_d, $|q/p|$, (x^2+y^2), $y'\cos\phi_D$, and $x'\sin\phi_D$. In particular, finding CP violation, that is, $|q/p| \neq 1$ and/or $\sin\phi_D \neq 0$, would constitute evidence for new physics.

More details on theoretical and experimental aspects of $D^0 - \overline{D}^0$ mixing can be found in [38]. Note that BABAR use $R_D \equiv r_d^2$ and $r_m \equiv |q/p|$. Belle use $R_m \equiv |q/p|$, $y_{CP} \equiv Y$, and $A_\Gamma \equiv -\Delta Y$.

12.6. B and B_s Decays

The upper bound on the CP asymmetry in semileptonic B decays [41] implies that CP violation in $B^0 - \overline{B}^0$ mixing is a small effect (we use $\mathcal{A}_{\rm SL}/2 \approx 1 - |q/p|$, see Eq. (12.34)):

$$\mathcal{A}_{\rm SL} = (-3.0 \pm 7.8) \times 10^{-3} \quad \Longrightarrow \quad |q/p| = 1.0015 \pm 0.0039. \tag{12.68}$$

The Standard Model prediction is

$$\mathcal{A}_{\rm SL} = \mathcal{O}\left(\frac{m_c^2}{m_t^2}\sin\beta\right) \lesssim 0.001. \tag{12.69}$$

In models where $\mathbf{\Gamma}_{12}/\mathbf{M}_{12}$ is approximately real, such as the Standard Model, an upper bound on $\Delta\Gamma/\Delta m \approx \mathcal{R}e(\mathbf{\Gamma}_{12}/\mathbf{M}_{12})$ provides yet another upper bound on the deviation of $|q/p|$ from one. This constraint does not hold if $\mathbf{\Gamma}_{12}/\mathbf{M}_{12}$ is approximately imaginary. (An alternative parameterization uses $q/p = (1-\tilde{\epsilon}_B)/(1+\tilde{\epsilon}_B)$, leading to $\mathcal{A}_{\rm SL} \simeq 4\mathcal{R}e(\tilde{\epsilon}_B)$.)

The small deviation (less than one percent) of $|q/p|$ from 1 implies that, at the present level of experimental precision, CP violation in B mixing is a negligible effect. Thus, for the purpose of analyzing CP asymmetries in hadronic B decays, we can use

$$\lambda_f = e^{-i\phi_{M(B)}}(\overline{A}_f/A_f)\,, \tag{12.70}$$

where $\phi_{M(B)}$ refers to the phase of $\mathbf{M}_{12}$ appearing in Eq. (12.39) that is appropriate for $B^0 - \overline{B}^0$ oscillations. Within the Standard Model, the corresponding phase factor is given by

$$e^{-i\phi_{M(B)}} = (V_{tb}^* V_{td})/(V_{tb} V_{td}^*)\,. \tag{12.71}$$

Some of the most interesting decays involve final states that are common to B^0 and $\overline{B}^0$ [42,43]. It is convenient to rewrite Eq. (12.37) for B decays as [44,45,46]

$$\mathcal{A}_f(t) = S_f \sin(\Delta m t) - C_f \cos(\Delta m t),$$

$$S_f \equiv \frac{2\,\mathcal{I}m(\lambda_f)}{1+\left|\lambda_f\right|^2}, \quad C_f \equiv \frac{1-\left|\lambda_f\right|^2}{1+\left|\lambda_f\right|^2}\,, \tag{12.72}$$

where we assume that $\Delta\Gamma = 0$ and $|q/p| = 1$. An alternative notation in use is $A_f \equiv -C_f$, but this A_f should not be confused with the A_f of Eq. (12.10).

A large class of interesting processes proceed via quark transitions of the form $\overline{b} \to \overline{q}q\overline{q}'$ with $q' = s$ or d. For $q = c$ or u, there are contributions from both tree (t) and penguin (p^{q_u}, where $q_u = u, c, t$ is the quark in the loop) diagrams (see Fig. 12.2) which carry different weak phases:

$$A_f = \left(V_{qb}^* V_{qq'}\right) t_f + \sum_{q_u = u,c,t} \left(V_{q_u b}^* V_{q_u q'}\right) p_f^{q_u}\,. \tag{12.73}$$

(The distinction between tree and penguin contributions is a heuristic one; the separation by the operator that enters is more precise. For a detailed discussion of the more complete operator product approach, which also includes higher order QCD corrections, see, for example, ref. [47].) Using CKM unitarity, these decay amplitudes can always be written in terms of just two CKM combinations. For example, for $f = \pi\pi$, which proceeds via $\overline{b} \to \overline{u}u\overline{d}$ transition, we can write

$$A_{\pi\pi} = (V_{ub}^* V_{ud})\, T_{\pi\pi} + (V_{tb}^* V_{td})\, P_{\pi\pi}^t, \tag{12.74}$$

where $T_{\pi\pi} = t_{\pi\pi} + p_{\pi\pi}^u - p_{\pi\pi}^c$ and $P_{\pi\pi}^t = p_{\pi\pi}^t - p_{\pi\pi}^c$. CP-violating phases in Eq. (12.74) appear only in the CKM elements, so that

$$\frac{\overline{A}_{\pi\pi}}{A_{\pi\pi}} = \frac{\left(V_{ub} V_{ud}^*\right) T_{\pi\pi} + \left(V_{tb} V_{td}^*\right) P_{\pi\pi}^t}{\left(V_{ub}^* V_{ud}\right) T_{\pi\pi} + \left(V_{tb}^* V_{td}\right) P_{\pi\pi}^t}. \tag{12.75}$$

For $f = J/\psi K$, which proceeds via $\overline{b} \to \overline{c}c\overline{s}$ transition, we can write

$$A_{\psi K} = (V_{cb}^* V_{cs})\, T_{\psi K} + (V_{ub}^* V_{us})\, P_{\psi K}^u, \tag{12.76}$$

where $T_{\psi K} = t_{\psi K} + p_{\psi K}^c - p_{\psi K}^t$ and $P_{\psi K}^u = p_{\psi K}^u - p_{\psi K}^t$. A subtlety arises in this decay that is related to the fact that B^0 decays into a final $J/\psi K^0$ state while $\overline{B}^0$ decays into a final $J/\psi \overline{K}^0$ state. A common final state, *e.g.*, $J/\psi K_S$, is reached only via $K^0 - \overline{K}^0$ mixing. Consequently, the phase factor (defined in Eq. (12.39)) corresponding to neutral K mixing, $e^{-i\phi_{M(K)}} = (V_{cd}^* V_{cs})/(V_{cd} V_{cs}^*)$, plays a role:

$$\frac{\overline{A}_{\psi K_S}}{A_{\psi K_S}} = -\frac{\left(V_{cb} V_{cs}^*\right) T_{\psi K} + \left(V_{ub} V_{us}^*\right) P_{\psi K}^u}{\left(V_{cb}^* V_{cs}\right) T_{\psi K} + \left(V_{ub}^* V_{us}\right) P_{\psi K}^u} \times \frac{V_{cd}^* V_{cs}}{V_{cd} V_{cs}^*}. \tag{12.77}$$

For $q = s$ or d, there are only penguin contributions to A_f, that is, $t_f = 0$ in Eq. (12.73). (The tree $\overline{b} \to \overline{u}u\overline{q}'$ transition followed by $\overline{u}u \to \overline{q}q$ rescattering is included below in the P^u terms.) Again, CKM unitarity allows us to write A_f in terms of two CKM combinations. For example, for $f = \phi K_S$, which proceeds via $\overline{b} \to \overline{s}s\overline{s}$ transition, we can write

$$\frac{\overline{A}_{\phi K_S}}{A_{\phi K_S}} = -\frac{\left(V_{cb} V_{cs}^*\right) P_{\phi K}^c + \left(V_{ub} V_{us}^*\right) P_{\phi K}^u}{\left(V_{cb}^* V_{cs}\right) P_{\phi K}^c + \left(V_{ub}^* V_{us}\right) P_{\phi K}^u} \times \frac{V_{cd}^* V_{cs}}{V_{cd} V_{cs}^*}, \tag{12.78}$$

where $P_{\phi K}^c = p_{\phi K}^c - p_{\phi K}^t$ and $P_{\phi K}^u = p_{\phi K}^u - p_{\phi K}^t$.

Since the amplitude A_f involves two different weak phases, the corresponding decays can exhibit both CP violation in the interference of decays with and without mixing, $S_f \neq 0$, and CP violation in decays, $C_f \neq 0$. (At the present level of experimental precision, the contribution to C_f from CP violation in mixing is negligible, see Eq. (12.68).) If the contribution from a second weak phase is suppressed, then the interpretation of S_f in terms of Lagrangian CP-violating parameters is clean, while

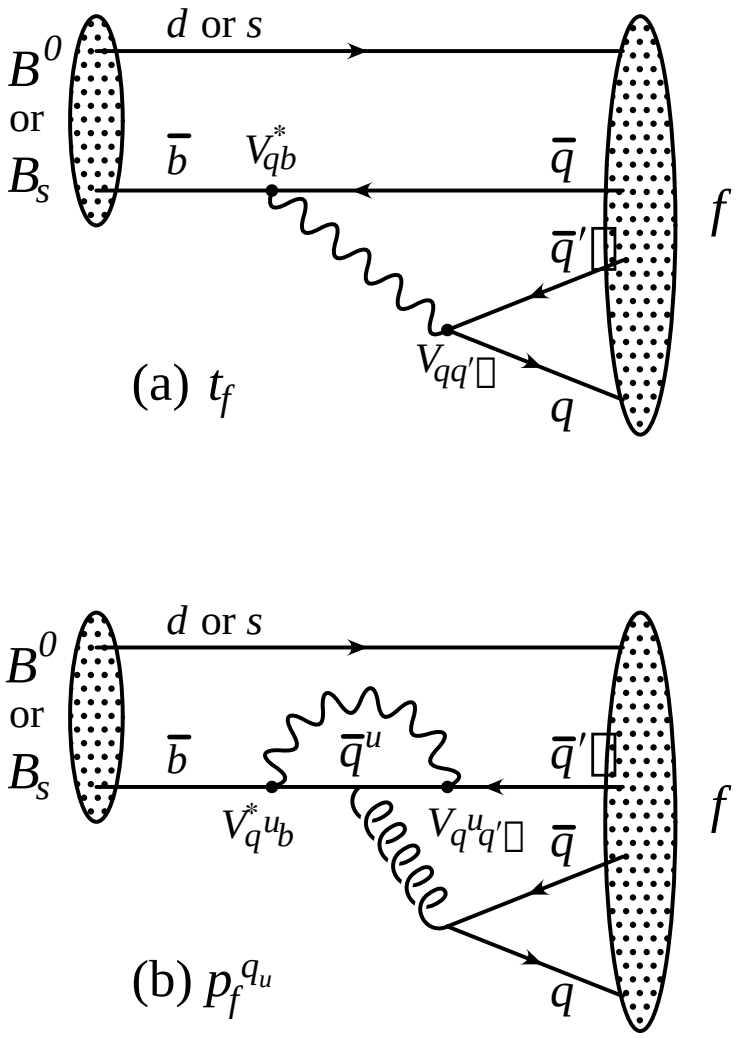


Figure 12.2: Feynman diagrams for (a) tree and (b) penguin amplitudes contributing to $B^0 \to f$ or $B_s \to f$ via a $\overline{b} \to \overline{q}q\overline{q}'$ quark-level process.

C_f is small. If such a second contribution is not suppressed, S_f depends on hadronic parameters and, if the relevant strong phase is large, C_f is large.

A summary of $\overline{b} \to \overline{q}q\overline{q}'$ modes with $q' = s$ or d is given in Table 12.1. The $\overline{b} \to \overline{d}d\overline{q}$ transitions lead to final states that are similar to the $\overline{b} \to \overline{u}u\overline{q}$ transitions and have similar phase dependence. Final states that consist of two vector-mesons ($\psi\phi$ and $\phi\phi$) are not CP eigenstates, and angular analysis is needed to separate the CP-even from the CP-odd contributions.

The cleanliness of the theoretical interpretation of S_f can be assessed from the information in the last column of Table 12.1. In case of small uncertainties, the expression

Table 12.1: Summary of $\overline{b} \to \overline{q}q\overline{q}'$ modes with $q' = s$ or d. The second and third columns give examples of final hadronic states. The fourth column gives the CKM dependence of the amplitude A_f, using the notation of Eqs. (12.74,12.76,12.78), with the dominant term first and the sub-dominant second. The suppression factor of the second term compared to the first is given in the last column. "Loop" refers to a penguin versus tree suppression factor (it is mode-dependent and roughly $\mathcal{O}(0.2-0.3)$) and $\lambda = 0.22$ is the expansion parameter of Eq. (12.45).

$\overline{b} \to \overline{q}q\overline{q}'$	$B^0 \to f$	$B_s \to f$	CKM dependence of A_f	Suppression
$\bar{b} \to \bar{c}c\bar{s}$	ψK_S	$\psi\phi$	$(V_{cb}^* V_{cs})T + (V_{ub}^* V_{us})P^u$	loop $\times\, \lambda^2$
$\bar{b} \to \bar{s}s\bar{s}$	ϕK_S	$\phi\phi$	$(V_{cb}^* V_{cs})P^c + (V_{ub}^* V_{us})P^u$	λ^2
$\bar{b} \to \bar{u}u\bar{s}$	$\pi^0 K_S$	K^+K^-	$(V_{cb}^* V_{cs})P^c + (V_{ub}^* V_{us})T$	λ^2/loop
$\bar{b} \to \bar{c}c\bar{d}$	D^+D^-	ψK_S	$(V_{cb}^* V_{cd})T + (V_{tb}^* V_{td})P^t$	loop
$\bar{b} \to \bar{s}s\bar{d}$	$\phi\pi$	ϕK_S	$(V_{tb}^* V_{td})P^t + (V_{cb}^* V_{cd})P^c$	$\lesssim 1$
$\bar{b} \to \bar{u}u\bar{d}$	$\pi^+\pi^-$	$\pi^0 K_S$	$(V_{ub}^* V_{ud})T + (V_{tb}^* V_{td})P^t$	loop

for S_f in terms of CKM phases can be deduced from the fourth column of Table 12.1 in combination with Eq. (12.71) (and, for $b \to q\overline{q}s$ decays, the example in Eq. (12.77)). Here we consider several interesting examples.

For $B \to J/\psi K_S$ and other $\overline{b} \to \overline{c}c\overline{s}$ processes, we can neglect the P^u contribution to A_f, in the Standard Model, to an approximation that is better than one percent:

$$\lambda_{\psi K_S} = -e^{-2i\beta} \;\Rightarrow\; S_{\psi K_S} = \sin 2\beta, \quad C_{\psi K_S} = 0\,. \tag{12.79}$$

(Below the percent level, several effects have to be taken into account [48].) In the presence of new physics, A_f is still likely to be dominated by the T term, but the mixing amplitude might be modified. We learn that, model independently, $C_f \approx 0$ while S_f cleanly determines the mixing phase $(\phi_M - 2\arg(V_{cb}V_{cd}^*))$. The experimental measurement [49], $S_{\psi K} = 0.685 \pm 0.032$, gave the first precision test of the Kobayashi-Maskawa mechanism, and its consistency with the predictions for $\sin 2\beta$ makes it very likely that this mechanism is indeed the dominant source of CP violation in meson decays.

For $B \to \phi K_S$ and other $\overline{b} \to \overline{s}s\overline{s}$ processes (as well as some $\overline{b} \to \overline{u}u\overline{s}$ processes), we can neglect the sub-dominant contributions, in the Standard Model, to an approximation that is good to order of a few percent:

$$\lambda_{\phi K_S} = -e^{-2i\beta} \;\Rightarrow\; S_{\phi K_S} = \sin 2\beta, \quad C_{\phi K_S} = 0\,. \tag{12.80}$$

In the presence of new physics, both A_f and $\mathbf{M}_{12}$ can get contributions that are comparable in size to those of the Standard Model and carry new weak phases. Such a situation gives several interesting consequences for penguin-dominated $b \to q\overline{q}s$ decays $(q = u, d, s)$ to a final state f:

1. The value of $-\eta_f S_f$ may be different from $S_{\psi K_S}$ by more than a few percent, where η_f is the CP eigenvalue of the final state.
2. The values of $\eta_f S_f$ for different final states f may be different from each other by more than a few percent (for example, $S_{\phi K_S} \neq S_{\eta' K_S}$).
3. The value of C_f may be different from zero by more than a few percent.

While a clear interpretation of such signals in terms of Lagrangian parameters will be difficult because, under these circumstances, hadronic parameters do play a role, any of the above three options will clearly signal new physics. Fig. 12.3 summarizes the present experimental results: none of the possible signatures listed above is unambiguously established, but there is definitely still room for new physics.

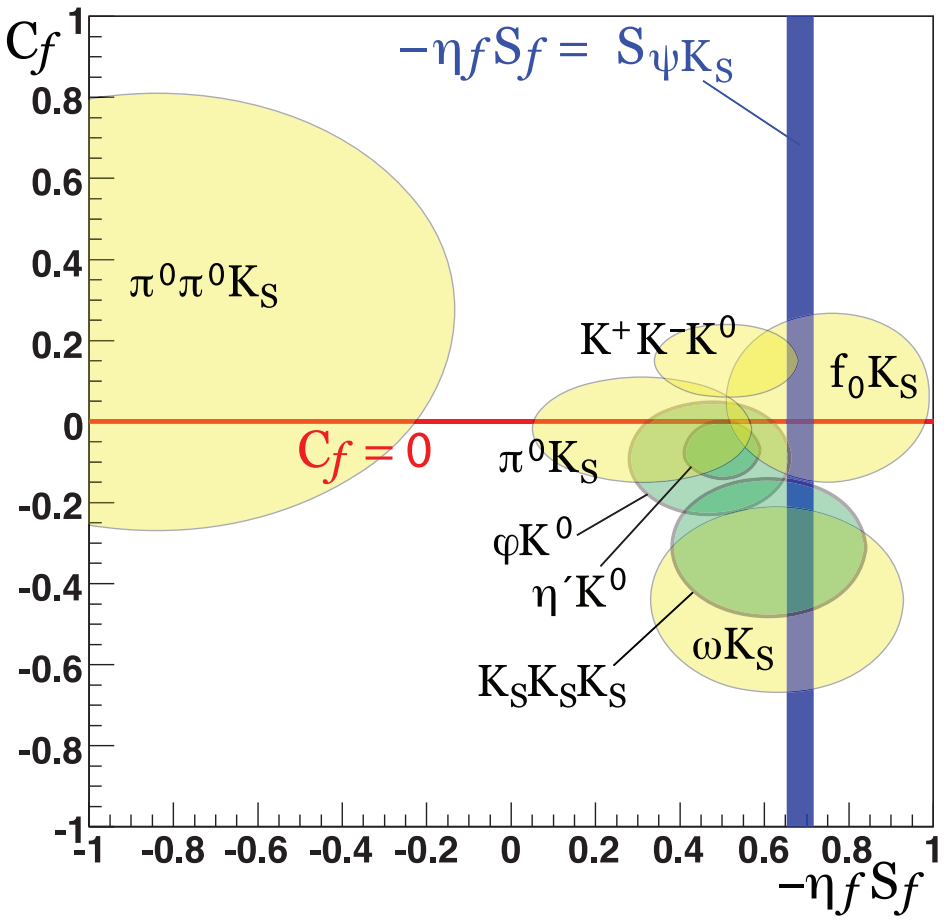


Figure 12.3: Summary of the results [49] of time-dependent analyses of $b \to q\overline{q}s$ decays, which are potentially sensitive to new physics. Sub-dominant corrections are expected to be smallest for the modes shown in green (darker). Results for final states including K^0 mesons combine CP-conjugate K_S and K_L measurements. The final state $K^+K^-K^0$ is not a CP eigenstate; the mixture of CP-even and CP-odd components is taken into account in obtaining an effective value for $\eta_f S_f$. Correlations between C_f and S_f are expected to be small and are not shown. See full-color version on color pages at end of book.

For $B \to \pi\pi$ and other $\overline{b} \to \overline{u}u\overline{d}$ processes, the penguin-to-tree ratio can be estimated using SU(3) relations and experimental data on related $B \to K\pi$ decays. The result is that the suppression is of order $0.2 - 0.3$ and so cannot be neglected. The expressions for $S_{\pi\pi}$ and $C_{\pi\pi}$ to leading order in $R_{PT} \equiv (|V_{tb}V_{td}|\, P^t_{\pi\pi})/(|V_{ub}V_{ud}|\, T_{\pi\pi})$ are:

$$\lambda_{\pi\pi} = e^{2i\alpha}\left[(1 - R_{PT}e^{-i\alpha})/(1 - R_{PT}e^{+i\alpha})\right] \;\Rightarrow$$

$$S_{\pi\pi} \approx \sin 2\alpha + 2\,\mathcal{R}e(R_{PT})\cos 2\alpha \sin\alpha, \quad C_{\pi\pi} \approx 2\,\mathcal{I}m(R_{PT})\sin\alpha. \tag{12.81}$$

Note that R_{PT} is mode-dependent and, in particular, could be different for $\pi^+\pi^-$ and $\pi^0\pi^0$. If strong phases can be neglected then R_{PT} is real, resulting in $C_{\pi\pi} = 0$. The size of $C_{\pi\pi}$ is an indicator of how large the strong phase is. The present experimental range is [49] $C_{\pi\pi} = -0.37 \pm 0.10$. As concerns $S_{\pi\pi}$, it is clear from Eq. (12.81) that the relative size or strong phase of the penguin contribution must be known to extract α. This is the problem of penguin pollution.

The cleanest solution involves isospin relations among the $B \to \pi\pi$ amplitudes [50]:

$$\frac{1}{\sqrt{2}}A_{\pi^+\pi^-} + A_{\pi^0\pi^0} = A_{\pi^+\pi^0}. \tag{12.82}$$

The method exploits the fact that the penguin contribution to $P^t_{\pi\pi}$ is pure $\Delta I = \dfrac{1}{2}$ (this is not true for the electroweak penguins which, however, are expected to be small), while the tree contribution to $T_{\pi\pi}$ contains pieces which are both $\Delta I = \dfrac{1}{2}$ and $\Delta I = \dfrac{3}{2}$. A simple geometric construction then allows one to find R_{PT} and extract α cleanly from $S_{\pi^+\pi^-}$. The key experimental difficulty is that one must measure accurately the separate rates for $B^0, \overline{B}^0 \to \pi^0\pi^0$. It has been noted that an upper bound on the average rate allows one to put a useful upper bound on the deviation of $S_{\pi^+\pi^-}$ from $\sin 2\alpha$ [51,52,53]. Parametrizing the asymmetry by $S_{\pi^+\pi^-}/\sqrt{1 - (C_{\pi^+\pi^-})^2} = \sin(2\alpha + 2\delta_{+-})$, the bound reads

$$\cos 2\delta_{+-} \geq \frac{1}{\sqrt{1 - (C_{\pi^+\pi^-})^2}}\left[1 - \frac{2\mathcal{B}_{00}}{\mathcal{B}_{+0}} + \frac{(\mathcal{B}_{+-} - 2\mathcal{B}_{+0} + 2\mathcal{B}_{00})^2}{4\mathcal{B}_{+-}\mathcal{B}_{+0}}\right], \tag{12.83}$$

where $\mathcal{B}_{ij}$ are the averages over CP-conjugate branching ratios; *e.g.*, $\mathcal{B}_{00} = \dfrac{1}{2}[\mathcal{B}(B^0 \to \pi^0\pi^0) + \mathcal{B}(\overline{B}^0 \to \pi^0\pi^0)]$.

CP asymmetries in $B \to \rho\pi$ and $B \to \rho\rho$ can also be used to determine α. In particular, the $B \to \rho\rho$ measurements are presently very significant in constraining α. The extraction proceeds via isospin analysis similar to that of $B \to \pi\pi$. There are, however, several important differences. First, due to the finite width of the ρ mesons, a final $(\rho\rho)_{I=1}$ state is possible [54]. The effect is, however, small, of order $(\Gamma_\rho/m_\rho)^2 \sim 0.04$. Second, due to the presence of three helicity states for the two vector

mesons, angular analysis is needed to separate the CP-even and CP-odd components. The theoretical expectation is, however, that the CP-odd component is small. This expectation is supported by experiments which find that the $\rho^+\rho^-$ and $\rho^\pm\rho^0$ modes are dominantly longitudinally polarized. Third, an important advantage of the $\rho\rho$ modes is that the penguin contribution is expected to be small due to different hadronic dynamics. This expectation is confirmed by the smallness of the upper bound on $\mathcal{B}(B^0 \to \rho^0\rho^0)$. Thus, $S_{\rho^+\rho^-}$ is not far from $\sin 2\alpha$. Finally, both $S_{\rho^0\rho^0}$ and $C_{\rho^0\rho^0}$ are experimentally accessible, which may allow a precision determination of α. The consistency between the range of α determined by the $B \to \pi\pi, \rho\pi, \rho\rho$ measurements and the range allowed by CKM fits (excluding these direct determinations) provides further support to the Kobayashi-Maskawa mechanism.

An interesting class of decay modes is that of the tree level decays $B^\pm \to D^{(*)0}K^\pm$. These decays provide golden methods for a clean determination of the angle γ [55,56,57,58]. The method uses the decays $B^+ \to D^0K^+$, which proceeds via the quark transition $\overline{b} \to \overline{u}c\overline{s}$, and $B^+ \to \overline{D}^0K^+$, which proceeds via the quark transition $\overline{b} \to \overline{c}u\overline{s}$, with the D^0 and $\overline{D}^0$ decaying into a common final state. The decays into common final states, such $(\pi^0K_S)_DK^+$, involve interference effects between the two amplitudes, with sensitivity to the relative phase, $\delta + \gamma$ (δ is the relevant strong phase). The CP-conjugate processes are sensitive to $\delta - \gamma$. Measurements of branching ratios and CP asymmetries allow an extraction of γ and δ from amplitude triangle relations. The extraction suffers from discrete ambiguities but involves no hadronic uncertainties. However, the smallness of the CKM-suppressed $b \to u$ transitions makes it difficult at present to use the simplest methods [55,56,57] to determine γ. These difficulties are overcome by performing a Dalitz plot analysis for multi-body D decays [58]. The consistency between the range of γ determined by the $B \to DK$ measurements and the range allowed by CKM fits (excluding these direct determinations) provides further support to the Kobayashi-Maskawa mechanism.

For B_s decays, one has to replace Eq. (12.71) with $e^{-i\phi_{M(B_s)}} = (V_{tb}^*V_{ts})/(V_{tb}V_{ts}^*)$. Note that one expects $\Delta\Gamma/\Gamma = \mathcal{O}(0.1)$, and therefore, y should not be put to zero in Eqs. (12.27,12.28), but $|q/p| = 1$ is expected to hold to an even better approximation than for B mesons. The CP asymmetry in $B_s \to J/\psi\phi$ will determine (with angular analysis to disentangle the CP-even and CP-odd components of the final state) $\sin 2\beta_s$, where

$$\beta_s \equiv \arg\left(-\frac{V_{ts}V_{tb}^*}{V_{cs}V_{cb}^*}\right). \tag{12.84}$$

12.7. Summary and Outlook

CP violation has been experimentally established in neutral K and B meson decays:

1. All three types of CP violation have been observed in $K \to \pi\pi$ decays:

$$\begin{aligned}
\mathcal{R}e(\epsilon') &= \frac{1}{6}\left(\left|\frac{\overline{A}_{\pi^0\pi^0}}{A_{\pi^0\pi^0}}\right| - \left|\frac{\overline{A}_{\pi^+\pi^-}}{A_{\pi^+\pi^-}}\right|\right) = (2.5 \pm 0.4) \times 10^{-6} \text{(I)} \\
\mathcal{R}e(\epsilon) &= \frac{1}{2}\left(1 - \left|\frac{q}{p}\right|\right) = (1.657 \pm 0.021) \times 10^{-3} \qquad \text{(II)} \\
\mathcal{I}m(\epsilon) &= -\frac{1}{2}\mathcal{I}m(\lambda_{(\pi\pi)_{I=0}}) = (1.572 \pm 0.022) \times 10^{-3} \qquad \text{(III)}
\end{aligned} \tag{12.85}$$

2. Direct CP violation has been observed in $B \to K^+\pi^-$ decays [59], and CP violation in interference of decays with and without mixing has been observed in $B \to J/\psi K_S$ decays [49] (and related modes):

$$\begin{aligned}
\mathcal{A}_{K^+\pi^-} &= \frac{|\overline{A}_{K^-\pi^+}/A_{K^+\pi^-}|^2 - 1}{|\overline{A}_{K^-\pi^+}/A_{K^+\pi^-}|^2 + 1} = -0.115 \pm 0.018 \text{(I)} \\
S_{\psi K} &= \mathcal{I}m(\lambda_{\psi K}) = 0.685 \pm 0.032 \qquad \text{(III)}
\end{aligned} \tag{12.86}$$

Searches for additional types of CP violation are ongoing in B, D, and K decays, and current limits are consistent with Standard Model expectations.

Based on Standard Model predictions, further observation of direct CP violation in B decays seems promising for the near future, followed later by CP violation observed in B_s decays and in the process $K \to \pi\nu\overline{\nu}$. Observables that are subject to clean theoretical interpretation, such as $S_{\psi K_S}$ and $\mathcal{B}(K_L \to \pi^0\nu\overline{\nu})$, are of particular value for constraining the values of the CKM parameters and probing the flavor sector of extensions to the Standard Model. Other probes of CP violation now being pursued experimentally include the electric dipole moments of the neutron and electron, and the decays of tau leptons. Additional processes that are likely to play an important role in future CP studies include top-quark production and decay, and neutrino oscillations.

All measurements of CP violation to date are consistent with the predictions of the Kobayashi-Maskawa mechanism of the Standard Model. However, a dynamically-generated matter-antimatter asymmetry of the universe requires additional sources of CP violation, and such sources are naturally generated by extensions to the Standard Model. New sources might eventually reveal themselves as small deviations from the predictions of the KM mechanism in meson decay rates, or else might not be observable in meson decays at all, but observable with future probes such as neutrino oscillations or electric dipole moments. We cannot guarantee that new sources of CP violation will ever be found experimentally, but the fundamental nature of CP violation demands a vigorous effort.

A number of excellent reviews of CP violation are available [60–66], where the interested reader may find a detailed discussion of the various topics that are briefly reviewed here.

References:

1. J.H. Christenson *et al.*, Phys. Rev. Lett. **13**, 138 (1964).
2. B. Aubert *et al.*, [BABAR Collab.], Phys. Rev. Lett. **87**, 091801 (2001).
3. K. Abe *et al.*, [Belle Collab.], Phys. Rev. Lett. **87**, 091802 (2001).
4. H. Burkhardt *et al.*, [NA31 Collab.], Phys. Lett. **B206**, 169 (1988).
5. V. Fanti *et al.*, [NA48 Collab.], Phys. Lett. **B465**, 335 (1999).
6. A. Alavi-Harati *et al.*, [KTeV Collab.], Phys. Rev. Lett. **83**, 22 (1999).
7. B. Aubert *et al.*, [BABAR Collab.], Phys. Rev. Lett. **93**, 131801 (2004).
8. K. Abe *et al.*, [Belle Collab.], `arXiv:hep-ex/0507045`.
9. See results on the "Time reversal invariance" within the review on "Tests of conservation laws" in this report.
10. See, for example, R. F. Streater and A. S. Wightman, "CPT, Spin and Statistics, and All That", reprinted by Addison-Wesley, New York (1989).
11. M. Kobayashi and T. Maskawa, Prog. Theor. Phys. **49**, 652 (1973).
12. A.D. Sakharov, Pisma Zh. Eksp. Teor. Fiz. **5**, 32 (1967) [Sov. Phys. JETP Lett. **5**, 24 (1967)].
13. For a review, see *e.g.* A. Riotto, "Theories of baryogenesis," `arXiv:hep-ph/9807454`.
14. M. Fukugita and T. Yanagida, Phys. Lett. **B174**, 45 (1986).
15. K. Abe *et al.*, [Belle Collab.], `arXiv:hep-ex/0409049`.
16. B. Aubert *et al.*, [BABAR Collab.], Phys. Rev. **D71**, 091102 (2005).
17. T. Aushev *et al.*, [Belle Collab.], `arXiv:hep-ex/0411021`.
18. B. Aubert *et al.*, [BABAR Collab.], Phys. Rev. Lett. **95**, 151804 (2005).
19. B. Aubert *et al.*, [BABAR Collab.], Phys. Rev. Lett. **94**, 191802 (2005).
20. K. Abe *et al.*, [Belle Collab.], `arXiv:hep-ex/0507037`.
21. B. Aubert *et al.*, [BABAR Collab.], `arXiv:hep-ex/0507087`.
22. B. Aubert *et al.*, [BABAR Collab.], `arXiv:hep-ex/0408095`.
23. C. C. Wang *et al.*, [Belle Collab.], Phys. Rev. Lett. **94**, 121801 (2005).
24. B. Aubert *et al.*, [BABAR Collab.], `arXiv:hep-ex/0408099`.
25. V. Weisskopf and E. P. Wigner, Z. Phys. **63**, 54 (1930); Z. Phys. **65**, 18 (1930). [See also Appendix A of P.K. Kabir, "The CP Puzzle: Strange Decays of the Neutral Kaon," Academic Press (1968)].
26. O. Long *et al.*, Phys. Rev. **D68**, 034010 (2003).
27. L. Wolfenstein, Phys. Rev. Lett. **13**, 562 (1964).
28. See the review on "Cabibbo-Kobayashi-Maskawa mixing matrix" in this report.
29. L. Wolfenstein, Phys. Rev. Lett. **51**, 1945 (1983).
30. A.J. Buras, M.E. Lautenbacher, and G. Ostermaier, Phys. Rev. **D50**, 3433 (1994).
31. C. Jarlskog, Phys. Rev. Lett. **55**, 1039 (1985).
32. See the K-meson listings in this report.
33. See the review on "CP violation in $K_S \to 3\pi$" in this report.
34. Y. Grossman and Y. Nir, Phys. Lett. **B398**, 163 (1997).

35. L.S. Littenberg, Phys. Rev. **D39**, 3322 (1989).
36. A.J. Buras, Phys. Lett. **B333**, 476 (1994).
37. G. Buchalla and A.J. Buras, Nucl. Phys. **B400**, 225 (1993).
38. See the review on "$D^0 - \overline{D}^0$ mixing" in this report.
39. See the D-meson listings in this report.
40. S. Bergmann *et al.*, Phys. Lett. **B486**, 418 (2000).
41. Heavy Flavor Averaging Group, Oscillations Working Group Summer 2005 Update, `http://www.slac.stanford.edu/xorg/hfag/osc/summer_2005/`.
42. A.B. Carter and A.I. Sanda, Phys. Rev. Lett. **45**, 952 (1980); Phys. Rev. **D23**, 1567 (1981).
43. I.I. Bigi and A.I. Sanda, Nucl. Phys. **B193**, 85 (1981).
44. I. Dunietz and J.L. Rosner, Phys. Rev. **D34**, 1404 (1986).
45. Ya.I. Azimov, N.G. Uraltsev, and V.A. Khoze, Sov. J. Nucl. Phys. **45**, 878 (1987) [Yad. Fiz. **45**, 1412 (1987)].
46. I.I. Bigi and A.I. Sanda, Nucl. Phys. **B281**, 41 (1987).
47. G. Buchalla, A.J. Buras, and M.E. Lautenbacher, Rev. Mod. Phys. **68**, 1125 (1996).
48. Y. Grossman, A.L. Kagan, and Z. Ligeti, Phys. Lett. **B538**, 327 (2002).
49. Heavy Flavor Averaging Group, Unitarity Triangle Working Group Summer 2005 Update, `http://www.slac.stanford.edu/xorg/hfag/triangle/summer2005/`.
50. M. Gronau and D. London, Phys. Rev. Lett. **65**, 3381 (1990).
51. Y. Grossman and H.R. Quinn, Phys. Rev. **D58**, 017504 (1998).
52. J. Charles, Phys. Rev. **D59**, 054007 (1999).
53. M. Gronau *et al.*, Phys. Lett. **B514**, 315 (2001).
54. A. F. Falk *et al.*, Phys. Rev. **D69**, 011502 (2004).
55. M. Gronau and D. London., Phys. Lett. **B253**, 483 (1991).
56. M. Gronau and D. Wyler, Phys. Lett. **B265**, 172 (1991).
57. D. Atwood, I. Dunietz, and A. Soni, Phys. Rev. Lett. **78**, 3257 (1997).
58. A. Giri *et al.*, Phys. Rev. **D68**, 054018 (2003).
59. Heavy Flavor Averaging Group, Rare Decays Working Group Summer 2005 Update, `http://www.slac.stanford.edu/xorg/hfag/rare/leppho05/acp/`.
60. G.C. Branco, L. Lavoura, and J.P. Silva, "CP violation," Oxford University Press, Oxford (1999).
61. I.I. Y. Bigi and A.I. Sanda, "CP Violation," Cambridge Monogr. Part. Phys. Nucl. Phys. Cosmol. **9**, 1 (2000).
62. P.F. Harrison and H.R. Quinn, editors [BABAR Collab.], "The BABAR physics book: Physics at an asymmetric B factory," SLAC-R-0504.
63. K. Anikeev *et al.*, `arXiv:hep-ph/0201071`.
64. K. Kleinknecht, "Uncovering CP Violation," Springer tracts in modern physics **195** (2003).
65. J. Hewett *et al.*, `arXiv:hep-ph/0503261`.

66. A. G. Akeroyd *et al.*, [SuperKEKB Physics Working Group], arXiv:hep-ex/0406071.

13. NEUTRINO MASS, MIXING, AND FLAVOR CHANGE

Revised September 2005 by B. Kayser (Fermilab).

There is now convincing evidence that atmospheric, solar, and reactor neutrinos change from one flavor to another. There is also very strong evidence that accelerator neutrinos do this as well. Barring exotic possibilities, neutrino flavor change implies that neutrinos have masses and that leptons mix. In this review, we discuss the physics of flavor change and the evidence for it, summarize what has been learned so far about neutrino masses and leptonic mixing, consider the relation between neutrinos and their antiparticles, and discuss the open questions about neutrinos to be answered by future experiments.

I. The physics of flavor change: If neutrinos have masses, then there is a spectrum of three or more neutrino mass eigenstates, ν_1, ν_2, ν_3, ..., that are the analogues of the charged-lepton mass eigenstates, e, μ, and τ. If leptons mix, the weak interaction coupling the W boson to a charged lepton and a neutrino can couple any charged-lepton mass eigenstate ℓ_α to any neutrino mass eigenstate ν_i. Here, $\alpha = e, \mu$, or τ, and ℓ_e is the electron, *etc.*. Leptonic W^+ decay can yield a particular ℓ_α^+ in association with any ν_i. The amplitude for this decay to produce the specific combination $\ell_\alpha^+ + \nu_i$ is $U_{\alpha i}^*$, where U is the unitary leptonic mixing matrix [1]. Thus, the neutrino state created in the decay $W^+ \to \ell_\alpha^+ + \nu$ is the state

$$|\nu_\alpha\rangle = \sum_i U_{\alpha i}^* |\nu_i\rangle \ . \tag{13.1}$$

This superposition of neutrino mass eigenstates, produced in association with the charged lepton of "flavor" α, is the state we refer to as the neutrino of flavor α.

While there are only three (known) charged lepton mass eigenstates, the experimental results suggest that perhaps there are more than three neutrino mass eigenstates. If, for example, there are four ν_i, then one linear combination of them,

$$|\nu_s\rangle = \sum_i U_{si}^* |\nu_i\rangle \ , \tag{13.2}$$

does not have a charged-lepton partner, and consequently does not couple to the Standard Model W boson. Indeed, since the decays $Z \to \nu_\alpha \overline{\nu}_\alpha$ of the Standard Model Z boson have been found to yield only three distinct neutrinos ν_α of definite flavor [2], ν_s does not couple to the Z boson either. Such a neutrino, which does not have any Standard Model weak couplings, is referred to as a "sterile" neutrino.

To understand neutrino flavor change, or "oscillation," in vacuum, let us consider how a neutrino born as the ν_α of Eq. (13.1) evolves in time. First, we apply Schrödinger's equation to the ν_i component of ν_α in the rest frame of that component. This tells us that

$$|\nu_i(\tau_i)\rangle = e^{-im_i\tau_i} |\nu_i(0)\rangle \ , \tag{13.3}$$

where m_i is the mass of ν_i, and τ_i is time in the ν_i frame. In terms of the time t and position L in the laboratory frame, the Lorentz-invariant phase factor in Eq. (13.3) may be written

$$e^{-im_i\tau_i} = e^{-i(E_i t - p_i L)} \ . \tag{13.4}$$

Here, E_i and p_i are respectively the energy and momentum of ν_i in the laboratory frame. In practice, our neutrino will be extremely relativistic, so we will be interested in evaluating the phase factor of Eq. (13.4) with $t \approx L$, where it becomes $\exp[-i(E_i - p_i)L]$.

Imagine now that our ν_α has been produced with a definite momentum p, so that all of its mass-eigenstate components have this common momentum. Then the ν_i component has $E_i = \sqrt{p^2 + m_i^2} \approx p + m_i^2/2p$, assuming that all neutrino masses m_i are small compared to the neutrino momentum. The phase factor of Eq. (13.4) is then approximately

$$e^{-i(m_i^2/2p)L} \ . \tag{13.5}$$

From this expression and Eq. (13.1), it follows that after a neutrino born as a ν_α has propagated a distance L, its state vector has become

$$|\nu_\alpha(L)\rangle \approx \sum_i U^*_{\alpha i} e^{-i(m_i^2/2E)L} |\nu_i\rangle \ . \tag{13.6}$$

Here, $E \simeq p$ is the average energy of the various mass eigenstate components of the neutrino. Using the unitarity of U to invert Eq. (13.1), and inserting the result in Eq. (13.6), we find that

$$|\nu_\alpha(L)\rangle \approx \sum_\beta \left[\sum_i U^*_{\alpha i} e^{-i(m_i^2/2E)L} U_{\beta i} \right] |\nu_\beta\rangle \ . \tag{13.7}$$

We see that our ν_α, in traveling the distance L, has turned into a superposition of all the flavors. The probability that it has flavor β, $P(\nu_\alpha \to \nu_\beta)$, is obviously $|\langle \nu_\beta | \nu_\alpha(L)\rangle|^2$. From Eq. (13.7) and the unitarity of U, we easily find that

$$\begin{aligned} P(\nu_\alpha \to \nu_\beta) = & \, \delta_{\alpha\beta} \\ & -4 \sum_{i>j} \Re(U^*_{\alpha i} U_{\beta i} U_{\alpha j} U^*_{\beta j}) \sin^2[1.27\, \Delta m^2_{ij} (L/E)] \\ & +2 \sum_{i>j} \Im(U^*_{\alpha i} U_{\beta i} U_{\alpha j} U^*_{\beta j}) \sin[2.54\, \Delta m^2_{ij} (L/E)] \ . \end{aligned} \tag{13.8}$$

Here, $\Delta m^2_{ij} \equiv m_i^2 - m_j^2$ is in eV^2, L is in km, and E is in GeV. We have used the fact that when the previously omitted factors of $\hbar$ and c are included,

$$\Delta m^2_{ij} (L/4E) \simeq 1.27\, \Delta m^2_{ij} (\mathrm{eV}^2) \frac{L(\mathrm{km})}{E(\mathrm{GeV})} \ . \tag{13.9}$$

The quantum mechanics of neutrino oscillation leading to the result Eq. (13.8) is somewhat subtle. To do justice to the physics requires a more refined treatment [3] than the one we have given. Sophisticated treatments continue to yield new insights [4].

Assuming that CPT invariance holds,

$$P(\overline{\nu}_\alpha \to \overline{\nu}_\beta) = P(\nu_\beta \to \nu_\alpha) \ . \tag{13.10}$$

But, from Eq. (13.8) we see that

$$P(\nu_\beta \to \nu_\alpha; U) = P(\nu_\alpha \to \nu_\beta; U^*) \ . \tag{13.11}$$

Thus, when CPT holds,

$$P(\overline{\nu}_\alpha \to \overline{\nu}_\beta; U) = P(\nu_\alpha \to \nu_\beta; U^*) \ . \tag{13.12}$$

That is, the probability for oscillation of an anti-neutrino is the same as that for a neutrino, except that the mixing matrix U is replaced by its complex conjugate. Thus, if U is not real, the neutrino and anti-neutrino oscillation probabilities can differ by having opposite values of the last term in Eq. (13.8). When CPT holds, any difference between these probabilities indicates a violation of CP invariance.

As we shall see, the squared-mass splittings Δm^2_{ij} called for by the various reported signals of oscillation are quite different from one another. It may be that one splitting, ΔM^2, is much bigger than all the others. If that is the case, then for an oscillation experiment with L/E such that $\Delta M^2 L/E = \mathcal{O}(1)$, Eq. (13.8) simplifies considerably, becoming

$$P(\overset{(-)}{\nu}_\alpha \to \overset{(-)}{\nu}_\beta) \simeq S_{\alpha\beta} \sin^2[1.27\,\Delta M^2 (L/E)] \tag{13.13}$$

for $\beta \neq \alpha$, and

$$P(\overset{(-)}{\nu}_\alpha \to \overset{(-)}{\nu}_\alpha) \simeq 1 - 4\,T_\alpha(1 - T_\alpha) \sin^2[1.27\,\Delta M^2 (L/E)] \ . \tag{13.14}$$

Here,

$$S_{\alpha\beta} \equiv 4 \left| \sum_{i \ \mathrm{Up}} U^*_{\alpha i} U_{\beta i} \right|^2 \tag{13.15}$$

and

$$T_\alpha \equiv \sum_{i \ \mathrm{Up}} |U_{\alpha i}|^2 \ , \tag{13.16}$$

where "i Up" denotes a sum over only those neutrino mass eigenstates that lie *above* ΔM^2 or, alternatively, only those that lie *below* it. The unitarity of U guarantees that summing over either of these two clusters will yield the same results for $S_{\alpha\beta}$ and for $T_\alpha(1 - T_\alpha)$.

The situation described by Eqs. (13.13)–(13.16) may be called "quasi-two-neutrino oscillation." It has also been called "one mass scale dominance" [5]. It corresponds to an experiment whose L/E is such that the experiment can "see" only the big splitting ΔM^2. To this experiment, all the neutrinos above ΔM^2 appear to be a single neutrino, as do all those below ΔM^2.

The relations of Eqs. (13.13)–(13.16) also apply to the special case where, to a good approximation, only two mass eigenstates, and two corresponding flavor eigenstates (or two linear combinations of flavor eigenstates), are relevant. One encounters this case when, for example, only two mass eigenstates couple significantly to the charged lepton with which the neutrino being studied is produced. When only two mass eigenstates count, there is only a single splitting, Δm^2, and, omitting irrelevant phase factors, the unitary mixing matrix U takes the form

$$U = \begin{array}{c} \\ \nu_\alpha \\ \nu_\beta \end{array} \begin{array}{c} \begin{array}{cc} \nu_1 & \nu_2 \end{array} \\ \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \end{array} \ . \tag{13.17}$$

Here, the symbols above and to the left of the matrix label the columns and rows, and θ is referred to as the mixing angle. From Eqs. (13.15) and (13.16), we now have $S_{\alpha\beta} = \sin^2 2\theta$ and $4\,T_\alpha(1 - T_\alpha) = \sin^2 2\theta$, so that Eqs. (13.13) and (13.14) become, respectively,

$$P(\overset{(-)}{\nu}_\alpha \to \overset{(-)}{\nu}_\beta) = \sin^2 2\theta \, \sin^2[1.27\,\Delta m^2(L/E)] \tag{13.18}$$

with $\beta \neq \alpha$, and

$$P(\overset{(-)}{\nu}_\alpha \to \overset{(-)}{\nu}_\alpha) = 1 - \sin^2 2\theta \, \sin^2[1.27\,\Delta m^2(L/E)] \ . \tag{13.19}$$

Many experiments have been analyzed using these two expressions. Some of these experiments actually have been concerned with quasi-two-neutrino oscillation, rather than a genuinely two-neutrino situation. For these experiments, "$\sin^2 2\theta$" and "Δm^2" have the significance that follows from Eqs. (13.13)–(13.16).

When neutrinos travel through matter (*e.g.*in the Sun, Earth, or a supernova), their coherent forward scattering from particles they encounter along the way can significantly modify their propagation [6]. As a result, the probability for changing flavor can be rather different than it is in vacuum [7]. Flavor change that occurs in matter, and that grows out of the interplay between flavor-nonchanging neutrino-matter interactions and neutrino mass and mixing, is known as the Mikheyev-Smirnov-Wolfenstein (MSW) effect.

To a good approximation, one can describe neutrino propagation through matter via a Schrödinger-like equation. This equation governs the evolution of a neutrino state vector with several components, one for each flavor. The effective Hamiltonian in the equation, a matrix $\mathcal{H}$ in neutrino flavor space, differs from its vacuum counterpart by the addition of interaction energies arising from the coherent forward neutrino scattering. For example, the ν_e–ν_e element of $\mathcal{H}$ includes the interaction energy

$$V = \sqrt{2}\,G_F N_e \ , \tag{13.20}$$

arising from W-exchange-induced ν_e forward scattering from ambient electrons. Here, G_F is the Fermi constant, and N_e is the number of electrons per unit volume. In addition, the ν_e–ν_e, ν_μ–ν_μ, and ν_τ–ν_τ elements of $\mathcal{H}$ all contain a common interaction energy growing out of Z-exchange-induced forward scattering. However, when one is not considering the

possibility of transitions to sterile neutrino flavors, this common interaction energy merely adds to $\mathcal{H}$ a multiple of the identity matrix, and such an addition has no effect on flavor transitions.

The effect of matter is illustrated by the propagation of solar neutrinos through solar matter. When combined with information on atmospheric neutrino oscillation, the experimental bounds on short-distance ($L \lesssim 1$ km) oscillation of reactor $\overline{\nu}_e$ [8] tell us that, if there are no sterile neutrinos, then only two neutrino mass eigenstates, ν_1 and ν_2, are significantly involved in the evolution of the solar neutrinos. Correspondingly, only two flavors are involved: the ν_e flavor with which every solar neutrino is born, and the effective flavor ν_x — some linear combination of ν_μ and ν_τ — which it may become. The Hamiltonian $\mathcal{H}$ is then a 2×2 matrix in ν_e–ν_x space. Apart from an irrelevant multiple of the identity, for a distance r from the center of the Sun, $\mathcal{H}$ is given by

$$\begin{aligned} \mathcal{H} &= \mathcal{H}_V + \mathcal{H}_M(r) \\ &= \frac{\Delta m^2_\odot}{4E} \begin{bmatrix} -\cos 2\theta_\odot & \sin 2\theta_\odot \\ \sin 2\theta_\odot & \cos 2\theta_\odot \end{bmatrix} + \begin{bmatrix} V(r) & 0 \\ 0 & 0 \end{bmatrix} \ . \end{aligned} \tag{13.21}$$

Here, the first matrix $\mathcal{H}_V$ is the Hamiltonian in vacuum, and the second matrix $\mathcal{H}_M(r)$ is the modification due to matter. In $\mathcal{H}_V$, $\theta_\odot$ is the solar mixing angle defined by the two-neutrino mixing matrix of Eq. (13.17) with $\theta = \theta_\odot, \nu_\alpha = \nu_e$, and $\nu_\beta = \nu_x$. The splitting $\Delta m^2_\odot$ is $m^2_2 - m^2_1$, and for the present purpose we *define* ν_2 to be the heavier of the two mass eigenstates, so that $\Delta m^2_\odot$ is positive. In $\mathcal{H}_M(r)$, $V(r)$ is the interaction energy of Eq. (13.20) with the electron density $N_e(r)$ evaluated at distance r from the Sun's center.

From Eqs. (13.18–13.19) (with $\theta = \theta_\odot$), we see that two-neutrino oscillation in vacuum cannot distinguish between a mixing angle $\theta_\odot$ and an angle $\theta'_\odot = \pi/2 - \theta_\odot$. But these two mixing angles represent physically different situations. Suppose, for example, that $\theta_\odot < \pi/4$. Then, from Eq. (13.17) we see that if the mixing angle is $\theta_\odot$, the lighter mass eigenstate (defined to be ν_1) is more ν_e than ν_x, while if it is $\theta'_\odot$, then this mass eigenstate is more ν_x than ν_e. While oscillation in vacuum cannot discriminate between these two possibilities, neutrino propagation through solar matter can do so. The neutrino interaction energy V of Eq. (13.20) is of definite, positive sign [9]. Thus, the ν_e–ν_e element of the solar $\mathcal{H}$, $-(\Delta m^2_\odot/4E)\cos 2\theta_\odot + V(r)$, has a different size when the mixing angle is $\theta'_\odot = \pi/2 - \theta_\odot$ than it does when this angle is $\theta_\odot$. As a result, the flavor content of the neutrinos coming from the Sun can be different in the two cases [10].

Solar and long-baseline reactor neutrino data establish that the behavior of solar neutrinos is governed by a Large-Mixing-Angle (LMA) MSW effect (see Sec. II). Let us estimate the probability $P(\nu_e \to \nu_e)$ that a solar neutrino which undergoes the LMA-MSW effect in the Sun still has its original ν_e flavor when it arrives at the Earth. We focus on the neutrinos produced by ^{8}B decay, which are at the high-energy end of the solar neutrino spectrum. At $r \simeq 0$, where the solar neutrinos are created, the electron density $N_e \simeq 6 \times 10^{25}/\text{cm}^3$ [11] yields for the interaction energy V of Eq. (13.20) the value 0.75×10^{-5} eV^2/MeV. Thus, for $\Delta m^2_\odot$ in the favored region, around 8×10^{-5} eV^2, and E a typical ^{8}B neutrino energy ($\sim$6-7 MeV), $\mathcal{H}_M$ dominates over $\mathcal{H}_V$. This

means that, in first approximation, $\mathcal{H}(r \simeq 0)$ is diagonal. Thus, a ^{8}B neutrino is born not only in a ν_e flavor eigenstate, but also, again in first approximation, in an eigenstate of the Hamiltonian $\mathcal{H}(r \simeq 0)$. Since $V > 0$, the neutrino will be in the heavier of the two eigenstates. Now, under the conditions where the LMA-MSW effect occurs, the propagation of a neutrino from $r \simeq 0$ to the outer edge of the Sun is adiabatic. That is, $N_e(r)$ changes sufficiently slowly that we may solve Schrödinger's equation for one r at a time, and then patch together the solutions. This means that our neutrino propagates outward through the Sun as one of the r-dependent eigenstates of the r-dependent $\mathcal{H}(r)$. Since the eigenvalues of $\mathcal{H}(r)$ do not cross at any r, and our neutrino is born in the heavier of the two $r = 0$ eigenstates, it emerges from the Sun in the heavier of the two $\mathcal{H}_V$ eigenstates. The latter is the mass eigenstate we have called ν_2, given according to Eq. (13.17) by

$$\nu_2 = \nu_e \sin\theta_\odot + \nu_x \cos\theta_\odot \ . \tag{13.22}$$

Since this is an eigenstate of the vacuum Hamiltonian, the neutrino remains in it all the way to the surface of the Earth. The probability of observing the neutrino as a ν_e on Earth is then just the probability that ν_2 is a ν_e. That is [cf. Eq. (13.22)] [12],

$$P(\nu_e \to \nu_e) = \sin^2\theta_\odot \ . \tag{13.23}$$

We note that for $\theta_\odot < \pi/4$, this ν_e survival probability is less than 1/2. In contrast, when matter effects are negligible, the energy-averaged survival probability in two-neutrino oscillation cannot be less than 1/2 for any mixing angle [see Eq. (13.19)] [13].

II. The evidence for flavor metamorphosis: The persuasiveness of the evidence that neutrinos actually do change flavor in nature is summarized in Table 13.1. We discuss the different pieces of evidence.

Table 13.1: The persuasiveness of the evidence for neutrino flavor change. The symbol L denotes the distance travelled by the neutrinos. LSND is the Liquid Scintillator Neutrino Detector experiment.

Neutrinos	Evidence for Flavor Change
Atmospheric	Compelling
Accelerator ($L = 250\,\text{km}$)	Very Strong
Solar	Compelling
Reactor ($L \sim 180\,\text{km}$)	Compelling
From Stopped μ^+ Decay (LSND)	Unconfirmed

The atmospheric neutrinos are produced in the Earth's atmosphere by cosmic rays, and then detected in an underground detector. The flux of cosmic rays that lead to neutrinos with energies above a few GeV is isotropic, so that these neutrinos are produced at the same rate all around the Earth. This can easily be shown to imply that at any

underground site, the downward- and upward-going fluxes of multi-GeV neutrinos of a given flavor must be equal. That is, unless some mechanism changes the flux of neutrinos of the given flavor as they propagate, the flux coming down from zenith angle θ_Z must equal that coming up from angle $\pi - \theta_Z$ [14].

The underground Super-Kamiokande (SK) detector finds that for multi-GeV atmospheric muon neutrinos, the θ_Z event distribution looks nothing like the expected $\theta_Z \Leftrightarrow \pi - \theta_Z$ symmetric distribution. For $\cos\theta_Z \gtrsim 0.3$, the observed ν_μ flux coming up from zenith angle $\pi - \theta_Z$ is only about half that coming down from angle θ_Z [15]. Thus, some mechanism does change the ν_μ flux as the neutrinos travel to the detector. The most attractive candidate for this mechanism is the oscillation $\nu_\mu \to \nu_X$ of the muon neutrinos into neutrinos ν_X of another flavor. Since the upward-going muon neutrinos come from the atmosphere on the opposite side of the Earth from the detector, they travel much farther than the downward-going ones to reach the detector. Thus, they have more time to oscillate away into the other flavor, which explains why Flux Up < Flux Down. The null results of short-baseline reactor neutrino experiments [8] imply limits on $P(\overline{\nu}_e \to \overline{\nu}_\mu)$, which, assuming CPT invariance, are also limits on $P(\nu_\mu \to \nu_e)$. From the latter, we know that ν_X is not a ν_e, except possibly a small fraction of the time. Thus, ν_X is a ν_τ, a sterile neutrino ν_s, or sometimes one and sometimes the other. All of the voluminous, detailed SK atmospheric neutrino data are very well described by the hypothesis that the oscillation is purely $\nu_\mu \to \nu_\tau$, and that it is a quasi-two-neutrino oscillation with a splitting $\Delta m^2_{\rm atm}$ and a mixing angle $\theta_{\rm atm}$ that, at 90% CL, are in the ranges [16]

$$1.9 \times 10^{-3}\,{\rm eV}^2 < \Delta m^2_{\rm atm} < 3.0 \times 10^{-3}\,{\rm eV}^2 \ , \tag{13.24}$$

and

$$\sin^2 2\theta_{\rm atm} > 0.90 \ . \tag{13.25}$$

Other experiments yield results consistent with these [17,18]. We note that the constraint (13.24) implies that at least one mass eigenstate ν_i has a mass exceeding 40 meV. From several pieces of evidence, the 90% CL upper limit on the fraction of ν_X that is sterile is 19% [19].

From Eq. (13.14), we see that the quasi-two-neutrino atmospheric $\nu_\mu \to \nu_\tau$ oscillation with dominant splitting $\Delta M^2 = \Delta m^2_{\rm atm}$ should lead to a minimum in the survival probability $P(\nu_\mu \to \nu_\mu)$ when $1.27\Delta m^2_{\rm atm}(L/E) = \pi/2$. For $\Delta m^2_{\rm atm}$ in the range of Eq. (13.24), this minimum is predicted to occur when $L/E \simeq 500\,{\rm km/GeV}$. The Super-Kamiokande experiment has confirmed that a dip in the ν_μ survival probability does occur at this L/E [16,20].

The oscillation interpretation of the atmospheric neutrino data has received support from the KEK to Kamioka (K2K) long-baseline experiment. This experiment produced a ν_μ beam using an accelerator, measured the beam intensity with a complex of near detectors, and then measured the ν_μ flux still in the beam 250 km away using the SK detector. The L/E of this experiment was such that one expected to see an oscillation dominated by the atmospheric squared-mass splitting $\Delta m^2_{\rm atm}$. From its near detector measurements, K2K would have seen $155.9 \pm 0.3\,^{+13.6}_{-15.6}$ events in the distant SK detector if there had been no oscillation. However, it actually saw only 112 events in SK [20].

In addition, the spectrum of ν_μ events observed in SK was distorted relative to the no-oscillation spectrum. The anomalously small number of events and spectral distortion seen by SK are consistent with a neutrino oscillation interpretation, with parameters $\Delta m^2_{\rm atm}$ and $\sin^2 2\theta_{\rm atm}$ compatible with those that fit the atmospheric neutrino data [20].

The neutrinos created in the Sun have been detected on Earth by several experiments, as discussed by K. Nakamura in this *Review*. The nuclear processes that power the Sun make only ν_e, not ν_μ or ν_τ. For years, solar neutrino experiments had been finding that the solar ν_e flux arriving at the Earth is below the one expected from neutrino production calculations. Now, thanks especially to the Sudbury Neutrino Observatory (SNO), we have compelling evidence that the missing ν_e have simply changed into neutrinos of other flavors.

SNO has studied the flux of high-energy solar neutrinos from ^{8}B decay. This experiment detects these neutrinos via the reactions

$$\nu + d \to e^- + p + p \ , \tag{13.26}$$

$$\nu + d \to \nu + p + n \ , \tag{13.27}$$

and

$$\nu + e^- \to \nu + e^- \ . \tag{13.28}$$

The first of these reactions, charged-current deuteron breakup, can be initiated only by a ν_e. Thus, it measures the flux $\phi(\nu_e)$ of ν_e from ^{8}B decay in the Sun. The second reaction, neutral-current deuteron breakup, can be initiated with equal cross sections by neutrinos of all active flavors. Thus, it measures $\phi(\nu_e) + \phi(\nu_{\mu,\tau})$, where $\phi(\nu_{\mu,\tau})$ is the flux of ν_μ and/or ν_τ from the Sun. Finally, the third reaction, neutrino electron elastic scattering, can be triggered by a neutrino of any active flavor, but $\sigma(\nu_{\mu,\tau}\, e \to \nu_{\mu,\tau}\, e) \simeq \sigma(\nu_e\, e \to \nu_e\, e)/6.5$. Thus, this reaction measures $\phi(\nu_e) + \phi(\nu_{\mu,\tau})/6.5$.

Recently, SNO has reported the results of measurements made with increased sensitivity to the neutral-current deuteron breakup, and with an enhanced ability to statistically separate the neutral-current breakups from the charged-current ones [21]. These improvements were achieved by adding salt to the heavy-water-filled SNO detector. From its observed rates for the two deuteron breakup reactions, SNO finds that [21]

$$\frac{\phi(\nu_e)}{\phi(\nu_e) + \phi(\nu_{\mu,\tau})} = 0.340 \pm 0.023\,(\text{stat})\, {}^{+0.029}_{-0.031}\,(\text{syst}) \ . \tag{13.29}$$

Clearly, $\phi(\nu_{\mu,\tau})$ is not zero. This non-vanishing $\nu_{\mu,\tau}$ flux from the Sun is "smoking-gun" evidence that some of the ν_e produced in the solar core do indeed change flavor.

Corroborating information comes from the detection reaction $\nu e^- \to \nu e^-$, studied by both SNO and SK [22].

Change of neutrino flavor, whether in matter or vacuum, does not change the total neutrino flux. Thus, unless some of the solar ν_e are changing into sterile neutrinos, the total active high-energy flux measured by the neutral-current reaction (13.27) should agree with the predicted total ^{8}B solar neutrino flux based on

calculations of neutrino production in the Sun. This predicted total is $(5.69 \pm 0.91) \times 10^6$ cm^{-2}s^{-1} [23]. By comparison, the total active flux measured by reaction (13.27) is $[4.94 \pm 0.21\,(\text{stat})\,^{+0.38}_{-0.34}\,(\text{syst})] \times 10^6$ cm^{-2}s^{-1}, in good agreement. This agreement provides evidence that neutrino production in the Sun is correctly understood, further strengthens the evidence that neutrinos really do change flavor, and strengthens the evidence that the previously-reported deficits of solar ν_e flux are due to this change of flavor.

The strongly favored explanation of solar neutrino flavor change is the LMA-MSW effect. As pointed out after Eq. (13.23), a ν_e survival probability below 1/2, which is indicated by Eq. (13.29), requires that solar matter effects play a significant role [24]. The LMA-MSW interpretation of solar neutrino behavior implies that a substantial fraction of reactor $\overline{\nu}_e$ that travel more than a hundred kilometers should disappear into anti-neutrinos of other flavors. The KamLAND experiment, which studies reactor $\overline{\nu}_e$ that typically travel ~ 180 km to reach the detector, finds that, indeed, the $\overline{\nu}_e$ flux at the detector is only $0.658 \pm 0.044\,(\text{stat}) \pm 0.047\,(\text{syst})$ of what it would be if no $\overline{\nu}_e$ were vanishing [25]. In addition, KamLAND observes the energy spectrum of the $\overline{\nu}_e$ that do reach the detector to be distorted, relative to the spectrum expected in the absence of oscillation. Plotting the ratio of the observed spectrum to the no-oscillation spectrum versus $1/E$, the KamLAND Collaboration finds evidence for the oscillatory L/E dependence that, as illustrated by Eq. (13.19), is a signature feature of neutrino oscillation [25].

The KamLAND data establish that the "solar" mixing angle $\theta_\odot$ is indeed large. In addition, KamLAND helps to confirm the LMA-MSW explanation of solar neutrino behavior since both the KamLAND result and all the solar neutrino data [26] can be described by the same neutrino parameters, in the LMA-MSW region. A global fit to both the solar and KamLAND data constrains these parameters, the solar $\Delta m^2_\odot$ and $\theta_\odot$ defined after Eq. (13.21), to lie in the region shown in Fig. 13.1 [21]. That $\theta_{\rm atm}$, Eq. (13.25), and $\theta_\odot$, Fig. 13.1, are both large, in striking contrast to all quark mixing angles, is very interesting.

While the total active solar neutrino flux measured by SNO via neutral-current deuteron breakup is compatible with the theoretically predicted total ^{8}B neutrino production by the Sun, we have seen that the uncertainties in these quantities are not negligible. It remains possible that some of the solar ν_e that change their flavor become sterile. Taking into account both the solar and KamLAND data, but not assuming the total ^{8}B solar neutrino flux to be known from theory, it has been found that, at $3\,\sigma$, the sterile fraction of the non-ν_e solar neutrino flux at Earth is less than 35% [27].

The neutrinos studied by the LSND experiment [28] come from the decay $\mu^+ \to e^+ \nu_e \overline{\nu}_\mu$ of muons at rest. While this decay does not produce $\overline{\nu}_e$, an excess of $\overline{\nu}_e$ over expected background is reported by the experiment. This excess is interpreted as due to oscillation of some of the $\overline{\nu}_\mu$ produced by μ^+ decay into $\overline{\nu}_e$. The related KArlsruhe Rutherford Medium Energy Neutrino (KARMEN) experiment [29] sees no indication for such an oscillation. However, the LSND and KARMEN experiments are not identical; at LSND the neutrino travels a distance $L \approx 30$ m before detection, while at KARMEN it travels $L \approx 18$ m. The KARMEN results exclude a portion of the neutrino parameter region

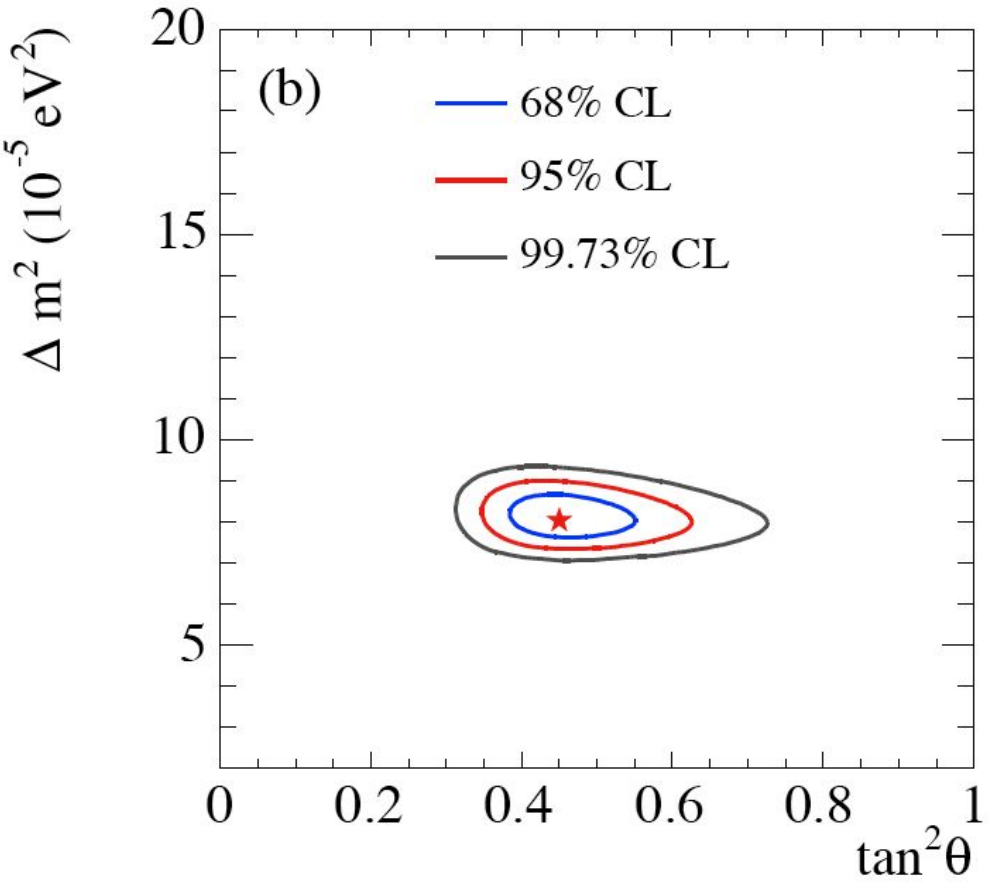


Figure 13.1: The region allowed for the neutrino parameters $\Delta m^2_\odot$ and $\theta_\odot$ by the solar and KamLAND data. The best-fit point, indicated by the star, is $\Delta m^2_\odot = (8.0^{+0.6}_{-0.4}) \times 10^{-5}\,\mathrm{eV}^2$ and $\theta_\odot = (33.9^{+2.4}_{-2.2})^\circ$. [21] See full-color version on color pages at end of book.

favored by LSND, but not all of it. A joint analysis [30] of the results of both experiments finds that a splitting $0.2 \lesssim \Delta m^2_{\mathrm{LSND}} \lesssim 1\,\mathrm{eV}^2$ and mixing $0.003 \lesssim \sin^2 2\theta_{\mathrm{LSND}} \lesssim 0.03$, or a splitting $\Delta m^2_{\mathrm{LSND}} \simeq 7\,\mathrm{eV}^2$ and mixing $\sin^2 2\theta_{\mathrm{LSND}} \simeq 0.004$, might explain both experiments.

The regions of neutrino parameter space favored or excluded by various neutrino oscillation experiments are shown in Fig. 13.2.

III. Neutrino spectra and mixings: If there are only three neutrino mass eigenstates, ν_1, ν_2 and ν_3, then there are only three mass splittings Δm^2_{ij}, and they obviously satisfy

$$\Delta m^2_{32} + \Delta m^2_{21} + \Delta m^2_{13} = 0 \ . \tag{13.30}$$

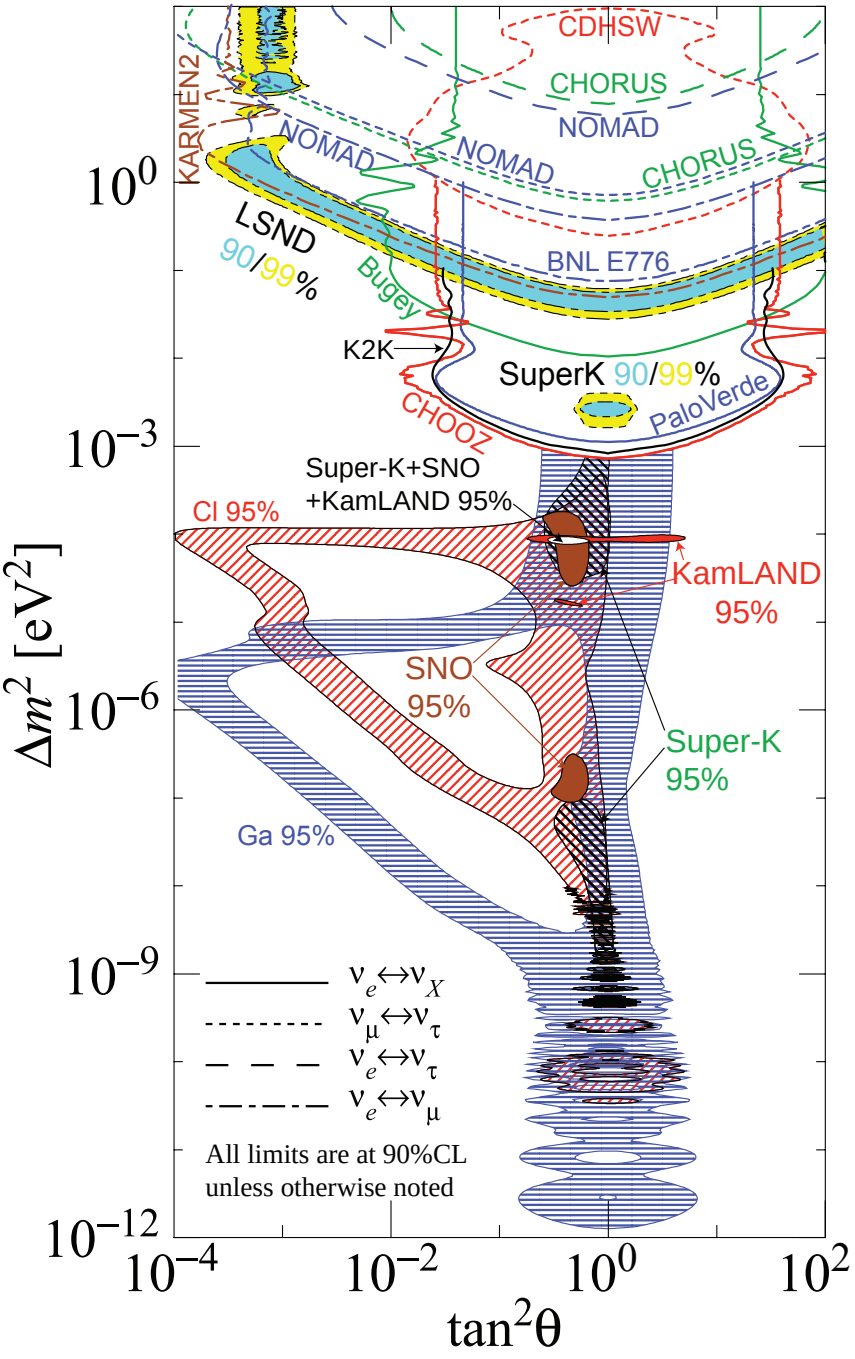


Figure 13.2: The regions of squared-mass splitting and mixing angle favored or excluded by various experiments. This figure was contributed by H. Murayama (University of California, Berkeley). References to the data used in the figure can be found at http://hitoshi.berkeley.edu/neutrino/ref2006.html. See full-color version on color pages at end of book.

However, as we have seen, the Δm^2 values required to explain the flavor changes of the atmospheric, solar, and LSND neutrinos are of three different orders of magnitude. Thus, they cannot possibly obey the constraint of Eq. (13.30). If all of the reported changes of flavor are genuine, then nature must contain at least four neutrino mass eigenstates [31]. As explained in Sec. I, one linear combination of these mass eigenstates would have to be sterile.

If the LSND oscillation is not confirmed, then nature may well contain only three neutrino mass eigenstates. The neutrino spectrum then contains two mass eigenstates separated by the splitting $\Delta m^2_{\odot}$ needed to explain the solar and KamLAND data, and a third eigenstate separated from the first two by the larger splitting $\Delta m^2_{\rm atm}$ called for by the atmospheric and K2K data. Current experiments do not tell us whether the solar pair — the two eigenstates separated by $\Delta m^2_{\odot}$ — is at the bottom or the top of the spectrum. These two possibilities are usually referred to, respectively, as a normal and an inverted spectrum. The study of flavor changes of accelerator-generated neutrinos and anti-neutrinos that pass through matter can discriminate between these two spectra (see Sec. V). If the solar pair is at the bottom, then the spectrum is of the form shown in Fig. 13.3. There we include the approximate flavor content of each mass eigenstate, the flavor-α fraction of eigenstate ν_i being simply $|\langle\nu_\alpha|\nu_i\rangle|^2 = |U_{\alpha i}|^2$. The flavor content shown assumes that the atmospheric mixing angle of Eq. (13.25) is maximal (which gives the best fit to the atmospheric data [20]) , and takes into account the now-established LMA-MSW explanation of solar neutrino behavior. For simplicity, it neglects the small, as-yet-unknown ν_e fraction of ν_3 (see below).

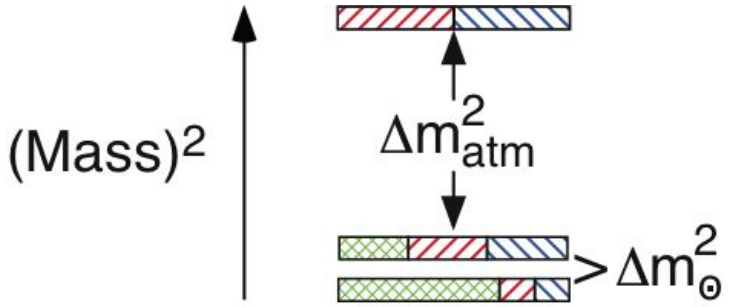


Figure 13.3: A three-neutrino squared-mass spectrum that accounts for the observed flavor changes of solar, reactor, atmospheric, and long-baseline accelerator neutrinos. The ν_e fraction of each mass eigenstate is crosshatched, the ν_μ fraction is indicated by right-leaning hatching, and the ν_τ fraction by left-leaning hatching.

When there are only three neutrino mass eigenstates, and the corresponding three familiar neutrinos of definite flavor, the leptonic mixing matrix U can be written as

$$U = \begin{array}{c} \\ \nu_e \\ \nu_\mu \\ \nu_\tau \end{array} \begin{array}{c} \begin{array}{ccc} \nu_1 & \nu_2 & \nu_3 \end{array} \\ \begin{bmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{bmatrix} \end{array} \times \mathrm{diag}(e^{i\alpha_1/2},\ e^{i\,\alpha_2/2},\ 1)\ . \tag{13.31}$$

Here, ν_1 and ν_2 are the members of the solar pair, with $m_2 > m_1$, and ν_3 is the isolated neutrino, which may be heavier or lighter than the solar pair. Inside the matrix, $c_{ij} \equiv \cos\theta_{ij}$ and $s_{ij} \equiv \sin\theta_{ij}$, where the three θ_{ij}'s are mixing angles. The quantities δ, α_1, and α_2 are CP-violating phases. The phases α_1 and α_2, known as Majorana phases, have physical consequences only if neutrinos are Majorana particles, identical to their antiparticles. Then these phases influence neutrinoless double beta decay [see Sec. IV] and other processes [32]. However, as we see from Eq. (13.8), α_1 and α_2 do not affect neutrino oscillation, regardless of whether neutrinos are Majorana particles. Apart from the phases α_1, α_2, which have no quark analogues, the parametrization of the leptonic mixing matrix in Eq. (13.31) is identical to that [33] advocated for the quark mixing matrix by Gilman, Kleinknecht, and Renk in their article in this *Review*.

From bounds on the short-distance oscillation of reactor $\overline{\nu}_e$ [8] and other data, at $2\,\sigma$, $|U_{e3}|^2 \lesssim 0.032$ [34]. (Thus, the ν_e fraction of ν_3 would have been too small to see in Fig. 13.3; this is the reason it was neglected.) From Eq. (13.31), we see that the bound on $|U_{e3}|^2$ implies that $s^2_{13} \lesssim 0.032$. From Eq. (13.31), we also see that the CP-violating phase δ, which is the sole phase in the U matrix that can produce CP violation in neutrino oscillation, enters U only in combination with s_{13}. Thus, the size of CP violation in oscillation will depend on s_{13}.

Given that s_{13} is small, Eqs. (13.31), (13.14), and (13.16) imply that the atmospheric mixing angle $\theta_{\rm atm}$ extracted from ν_μ disappearance measurements is approximately θ_{23}, while Eqs. (13.31) and (13.17) (with $\nu_\alpha = \nu_e$ and $\theta = \theta_\odot$) imply that $\theta_\odot \simeq \theta_{12}$.

If the LSND oscillation is confirmed, then, as already noted, there must be at least four mass eigenstates. It is found that if there are *exactly* four, a statistically satisfactory fit to all the neutrino data is not possible. However, if there are *at least* four neutrino mass eigenstates, there is no strong reason to believe that there are *exactly* four. The presence of more states may improve the quality of the fit. For example, it has been found that a "3+2" spectrum fits all the short-baseline data significantly better than a 3+1 spectrum [35].

IV. The neutrino-anti-neutrino relation: Unlike quarks and charged leptons, neutrinos may be their own antiparticles. Whether they are depends on the nature of the physics that gives them mass.

In the Standard Model (SM), neutrinos are assumed to be massless. Now that we know they do have masses, it is straightforward to extend the SM to accommodate these masses in the same way that this model accommodates quark and charged lepton masses. When a neutrino ν is assumed to be massless, the SM does not contain the chirally

right-handed neutrino field ν_R, but only the left-handed field ν_L that couples to the W and Z bosons. To accommodate the ν mass in the same manner as quark masses are accommodated, we add ν_R to the Model. Then we may construct the "Dirac mass term"

$$\mathcal{L}_D = -m_D \, \overline{\nu}_L \, \nu_R + h.c. \quad , \tag{13.32}$$

in which m_D is a constant. This term, which mimics the mass terms of quarks and charged leptons, conserves the lepton number L that distinguishes neutrinos and negatively-charged leptons on the one hand from anti-neutrinos and positively-charged leptons on the other. Since everything else in the SM also conserves L, we then have an L-conserving world. In such a world, each neutrino mass eigenstate ν_i differs from its antiparticle $\overline{\nu}_i$, the difference being that $L(\overline{\nu}_i) = -L(\nu_i)$. When $\overline{\nu}_i \neq \nu_i$, we refer to the $\nu_i - \overline{\nu}_i$ complex as a "Dirac neutrino."

Once ν_R has been added to our description of neutrinos, a "Majorana mass term,"

$$\mathcal{L}_M = -m_R \, \overline{\nu_R^c} \, \nu_R + h.c. \quad , \tag{13.33}$$

can be constructed out of ν_R and its charge conjugate, ν_R^c. In this term, m_R is another constant. Since both ν_R and $\overline{\nu_R^c}$ absorb ν and create $\overline{\nu}$, $\mathcal{L}_M$ mixes ν and $\overline{\nu}$. Thus, a Majorana mass term does not conserve L. There is then no conserved lepton number to distinguish a neutrino mass eigenstate ν_i from its antiparticle. Hence, when Majorana mass terms are present, $\overline{\nu}_i = \nu_i$. That is, for a given helicity h, $\overline{\nu}_i(h) = \nu_i(h)$. We then refer to ν_i as a "Majorana neutrino."

Suppose the right-handed neutrinos required by Dirac mass terms have been added to the SM. If we insist that this extended SM conserve L, then, of course, Majorana mass terms are forbidden. However, if we do not impose L conservation, but require only the general principles of gauge invariance and renormalizability, then Majorana mass terms like that of Eq. (13.33) are expected to be present. As a result, L is violated, and neutrinos are Majorana particles [36].

In the see-saw mechanism [37], which is the most popular explanation of why neutrinos — although massive — are nevertheless so light, both Dirac and Majorana mass terms are present. Hence, the neutrinos are Majorana particles. However, while half of them are the familiar light neutrinos, the other half are extremely heavy Majorana particles referred to as the N_i, with masses possibly as large as the GUT scale. The N_i may have played a crucial role in baryogenesis in the early universe, as we shall discuss in Sec. V.

How can the theoretical expectation that L is violated and neutrinos are Majorana particles be confirmed experimentally? The interactions of neutrinos are well described by the SM, and the SM interactions conserve L. If we may neglect any non-SM L-violating interactions, then the only sources of L violation are the neutrino Majorana mass terms. This means that all L-violating effects disappear in the limit of vanishing neutrino masses. Thus, any experimental approach to confirming the violation of L, and the consequent Majorana character of neutrinos, must be able to see an L violation that is going to be very small because of the smallness of the neutrino masses that drive it. One approach that shows great promise is the search for neutrinoless double beta decay ($0\nu\beta\beta$). This is

the process $(A, Z) \to (A, Z+2) + 2e^-$, in which a nucleus containing A nucleons, Z of which are protons, decays to a nucleus containing $Z+2$ protons by emitting two electrons. This process manifestly violates L conservation, so we expect it to be suppressed. However, if (A, Z) is a nucleus that is stable against single β (and α and γ) decay, then it can decay only via the process we are seeking, and the L-conserving two-neutrino process $(A, Z) \to (A, Z+2) + 2e^- + 2\overline{\nu}_e$. The latter decay mode is suppressed by the small phase space associated with the four light particles in the final state, so we have a chance to observe the neutrinoless mode, $(A, Z) \to (A, Z+2) + 2e^-$.

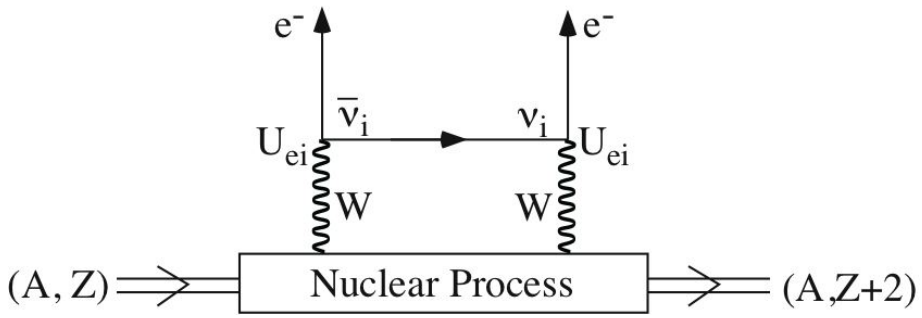


Figure 13.4: The dominant mechanism for $0\nu\beta\beta$. The diagram does not exist unless $\overline{\nu}_i = \nu_i$.

While $0\nu\beta\beta$ can in principle receive contributions from a variety of mechanisms (R-parity-violating supersymmetric couplings, for example), it is easy to show explicitly that the observation of $0\nu\beta\beta$ at any non-vanishing rate would imply that nature contains at least one Majorana neutrino mass term [38]. Now, quarks and charged leptons cannot have Majorana mass terms, because such terms mix fermion and antifermion, and $q \leftrightarrow \overline{q}$ or $\ell \leftrightarrow \overline{\ell}$ would not conserve electric charge. Thus, the discovery of $0\nu\beta\beta$ would demonstrate that the physics of neutrino masses is unlike that of the masses of all other fermions.

The dominant mechanism for $0\nu\beta\beta$ is expected to be the one depicted in Fig. 13.4. There, a pair of virtual W bosons are emitted by the parent nucleus, and then these W bosons exchange one or another of the light neutrino mass eigenstates ν_i to produce the outgoing electrons. The $0\nu\beta\beta$ amplitude is then a sum over the contributions of the different ν_i. It is assumed that the interactions at the two leptonic W vertices are those of the SM.

Since the exchanged ν_i is created together with an e^-, the left-handed SM current that creates it gives it the helicity we associate, in common parlance, with an "anti-neutrino." That is, the ν_i is almost totally right-handed, but has a small left-handed-helicity component, whose amplitude is of order m_i/E, where E is the ν_i energy. At the vertex where this ν_i is absorbed, the absorbing left-handed SM current can absorb only its

small left-handed-helicity component without further suppression. Consequently, the ν_i-exchange contribution to the $0\nu\beta\beta$ amplitude is proportional to m_i. From Fig. 13.4, we see that this contribution is also proportional to U_{ei}^2. Thus, summing over the contributions of all the ν_i, we conclude that the amplitude for $0\nu\beta\beta$ is proportional to the quantity

$$\left| \sum_i m_i \, U_{ei}^2 \right| \equiv | < m_{\beta\beta} > | \quad , \tag{13.34}$$

commonly referred to as the "effective Majorana mass for neutrinoless double beta decay" [39].

That the $0\nu\beta\beta$ amplitude arising from the diagram in Fig. 13.4 is proportional to neutrino mass is no surprise, and illustrates our earlier general discussion. The diagram in Fig. 13.4 is manifestly L-nonconserving. But we are assuming that the interactions in this diagram are L-conserving. Thus, the L-nonconservation in the diagram as a whole must be coming from underlying Majorana neutrino mass terms. Hence, if all the neutrino masses vanish, the L-nonconservation will vanish as well.

To how small an $| < m_{\beta\beta} > |$ should a $0\nu\beta\beta$ search be sensitive? In answering this question, it makes sense to assume there are only three neutrino mass eigenstates — if there are more, $| < m_{\beta\beta} > |$ might be larger. Suppose that there are just three mass eigenstates, and that the solar pair, ν_1 and ν_2, is at the top of the spectrum, so that we have an inverted spectrum. If the various ν_i are not much heavier than demanded by the observed splittings $\Delta m^2_{\rm atm}$ and $\Delta m^2_{\odot}$, then in $| < m_{\beta\beta} > |$, Eq. (13.34), the contribution of ν_3 may be neglected, because both m_3 and $|U_{e3}^2| = s_{13}^2$ are small. From Eqs. (13.34) and (13.31), approximating c_{13} by unity, we then have that

$$| < m_{\beta\beta} > | \simeq m_0 \sqrt{1 - \sin^2 2\theta_{\odot} \sin^2 \left(\frac{\Delta\alpha}{2} \right)} \quad . \tag{13.35}$$

Here, m_0 is the average mass of the members of the solar pair, whose splitting will be invisible in a practical $0\nu\beta\beta$ experiment, and $\Delta\alpha \equiv \alpha_2 - \alpha_1$ is a CP-violating phase. Although $\Delta\alpha$ is completely unknown, we see from Eq. (13.35) that

$$| < m_{\beta\beta} > | \geq m_0 \cos 2\theta_{\odot} \quad . \tag{13.36}$$

Now, in an inverted spectrum, $m_0 \geq \sqrt{\Delta m^2_{\rm atm}}$. At 90% CL, $\sqrt{\Delta m^2_{\rm atm}} > 40$ meV [16], while at 95% CL, $\cos 2\theta_{\odot} > 0.28$ [25]. Thus, if neutrinos are Majorana particles, and the spectrum is as we have assumed, a $0\nu\beta\beta$ experiment sensitive to $| < m_{\beta\beta} > | \gtrsim 10$ meV would have an excellent chance of observing a signal. If the spectrum is inverted, but the ν_i masses are larger than the $\Delta m^2_{\rm atm}$- and $\Delta m^2_{\odot}$-demanded minimum values we have assumed above, then once again $| < m_{\beta\beta} > |$ is larger than 10 meV [40], and an experiment sensitive to 10 meV still has an excellent chance of seeing a signal.

If the solar pair is at the bottom of the spectrum, rather than at the top, then $| < m_{\beta\beta} > |$ is not as tightly constrained, and can be anywhere from the present bound

of 0.3–1.0 eV down to invisibly small [40,41]. For a discussion of the present bounds, see the article by Vogel and Piepke in this *Review* [42].

V. Questions to be answered: The strong evidence for neutrino flavor metamorphosis — hence neutrino mass — opens many questions about the neutrinos. These questions, which hopefully will be answered by future experiments, include the following:

i) How many neutrino species are there? Do sterile neutrinos exist?

This question is being addressed by the MiniBooNE experiment [43], whose purpose is to confirm or refute LSND. MiniBooNE expects to have a result at the end of 2005. If this result is positive, the implications will be far-reaching. We will have learned that either there are more than three neutrino species and at least one of these species is sterile, or else there is an even more amazing departure from what has been our picture of the neutrino world.

ii) What are the masses of the mass eigenstates ν_i?

The sizes of the squared-mass splittings $\Delta m^2_{\odot}$, $\Delta m^2_{\rm atm}$, and, if present, one or more large splittings $\Delta m^2_{\rm LSND}$, can be determined more precisely than they are currently known through future neutrino oscillation measurements. Importantly, if there are only three ν_i, then one can find out whether the solar pair, $\nu_{1,2}$, is at the bottom of the spectrum or at its top by exploiting matter effects in long-baseline neutrino and anti-neutrino oscillations. These matter effects will determine the sign one wishes to learn — that of $\{m^2_3 - [(m^2_2 + m^2_1)/2]\}$ — relative to a sign that is already known — that of the interaction energy of Eq. (13.20). Grand unified theories favor a spectrum with the closely-spaced solar pair at the bottom [44]. The neutrino spectrum would then resemble the spectra of the quarks, to which grand unified theories relate the neutrinos.

While flavor-change experiments can determine a spectral pattern such as the one in Fig. 13.3, they cannot tell us the distance of the entire pattern from the zero of squared-mass. One might discover that distance via study of the β energy spectrum in tritium β decay, if the mass of some ν_i with appreciable coupling to an electron is large enough to be within reach of a feasible experiment. One might also gain some information on the distance from zero by measuring $|< m_{\beta\beta} >|$, the effective Majorana mass for neutrinoless double beta decay [40–42] (see Vogel and Piepke in this *Review*). Finally, one might obtain information on this distance from cosmology or astrophysics. Indeed, from current cosmological data and some cosmological assumptions, it is already concluded that [45]

$$\sum_i m_i < (0.4 - 1.0)\ {\rm eV}\ . \tag{13.37}$$

Here, the sum runs over the masses of all the light neutrino mass eigenstates ν_i that may exist and that were in thermal equilibrium in the early universe. The range quoted in Eq. (13.37) reflects the dependence of this upper bound on the underlying cosmological assumptions and on which data are used [45].

If there are just three ν_i, and their spectrum is either the one shown in Fig. 13.3 or its inverted version, then Eq. (13.37) implies that the mass of the heaviest ν_i, Mass [Heaviest ν_i], cannot exceed (0.2 – 0.4) eV. Moreover, Mass [Heaviest ν_i] obviously cannot be less

than $\sqrt{\Delta m^2_{\rm atm}}$, which in turn is not less than 0.04 eV, as previously noted. Thus, if the cosmological assumptions behind Eq. (13.37) are correct, then

$$0.04\,{\rm eV} < {\rm Mass\,[Heaviest}\ \nu_i] < (0.2-0.4)\,{\rm eV}\ . \tag{13.38}$$

iii) Are the neutrino mass eigenstates Majorana particles?

The confirmed observation of neutrinoless double beta decay would establish that the answer is "yes." If there are only three ν_i, knowledge that the spectrum is inverted and a definitive upper bound on $|<m_{\beta\beta}>|$ that is well below 0.01 eV would establish that it is "no" [see discussion after Eq. (13.36)] [40,41].

iv) What are the mixing angles in the leptonic mixing matrix U?

The solar mixing angle $\theta_\odot$ is already rather well determined.

The atmospheric mixing angle $\theta_{\rm atm}$ is constrained by the most stringent analysis to lie, at 90% CL, in the region where $\sin^2 2\theta_{\rm atm} > 0.92$ [20]. This region is still fairly large: 37° to 53°. A more precise value of $\sin^2 2\theta_{\rm atm}$, and, in particular, its deviation from unity, can be sought in precision long-baseline ν_μ disappearance experiments. If $\sin^2 2\theta_{\rm atm} \neq 1$, so that $\theta_{\rm atm} \neq 45°$, one can determine whether it lies below or above 45° by a reactor $\overline{\nu}_e$ experiment [46,47]. Once we know whether the neutrino spectrum is normal or inverted, this determination will tell us whether the heaviest mass eigenstate is more ν_τ than ν_μ, as naively expected, or more ν_μ than ν_τ [cf. Eq. (13.31)].

A knowledge of the small mixing angle θ_{13} is important not only to help complete our picture of leptonic mixing, but also because, as Eq. (13.31) made clear, all CP-violating effects of the phase δ are proportional to $\sin\theta_{13}$. Thus, a knowledge of the order of magnitude of θ_{13} would help guide the design of experiments to probe CP violation. From Eq. (13.31), we recall that $\sin^2\theta_{13}$ is the ν_e fraction of ν_3. The ν_3 is the isolated neutrino that lies at one end of the atmospheric squared-mass gap $\Delta m^2_{\rm atm}$, so an experiment seeking to measure θ_{13} should have an L/E that makes it sensitive to $\Delta m^2_{\rm atm}$, and should involve ν_e. Possibilities include a sensitive search for the disappearance of reactor $\overline{\nu}_e$ while they travel a distance $L \sim 1$ km, and an accelerator neutrino search for $\nu_\mu \to \nu_e$ or $\nu_e \to \nu_\mu$ with a beamline $L >$ several hundred km.

If LSND is confirmed, then (barring the still more revolutionary) the matrix U is at least 4×4, and contains many more than three angles. A rich program, including short baseline experiments with multiple detectors, will be needed to learn about both the squared-mass spectrum and the mixing matrix.

Given the large sizes of $\theta_{\rm atm}$ and $\theta_\odot$, we already know that leptonic mixing is very different from its quark counterpart, where all the mixing angles are small. This difference, and the striking contrast between the tiny neutrino masses and the very much larger quark masses, suggest that the physics underlying neutrino masses and mixing may be very different from the physics behind quark masses and mixing.

v) Does the behavior of neutrinos violate CP?

From Eqs. (13.8), (13.12), and (13.31), we see that if the CP-violating phase δ and the small mixing angle θ_{13} are both non-vanishing, there will be CP-violating

differences between neutrino and anti-neutrino oscillation probabilities. Observation of these differences would establish that CP violation is not a peculiarity of quarks.

The CP-violating difference $P(\nu_\alpha \to \nu_\beta) - P(\overline{\nu}_\alpha \to \overline{\nu}_\beta)$ between "neutrino" and " anti-neutrino" oscillation probabilities is independent of whether the mass eigenstates ν_i are Majorana or Dirac particles. To study $\nu_\mu \to \nu_e$ with a super-intense but conventionally-generated neutrino beam, for example, one would create the beam via the process $\pi^+ \to \mu^+ \nu_i$, and detect it via $\nu_i + \text{target} \to e^- + \ldots$. To study $\overline{\nu}_\mu \to \overline{\nu}_e$, one would create the beam via $\pi^- \to \mu^- \overline{\nu}_i$, and detect it via $\overline{\nu}_i + \text{target} \to e^+ + \ldots$. Whether $\overline{\nu}_i = \nu_i$ or not, the amplitudes for the latter two processes are proportional to $U_{\mu i}$ and U^*_{ei}, respectively. In contrast, the amplitudes for their $\nu_\mu \to \nu_e$ counterparts are proportional to $U^*_{\mu i}$ and U_{ei}. As this illustrates, Eq. (13.12) relates "neutrino" and "anti-neutrino" oscillation probabilities even when the neutrino mass eigenstates are their own antiparticles.

The baryon asymmetry of the universe could not have developed without some violation of CP during the universe's early history. The one known source of CP violation — the complex phase in the quark mixing matrix — could not have produced sufficiently large effects. Thus, perhaps *leptonic* CP violation is responsible for the baryon asymmetry. The see-saw mechanism predicts very heavy Majorana neutral leptons N_i (see Sec. IV), which would have been produced in the Big Bang. Perhaps CP violation in the leptonic decays of an N_i led to the inequality

$$\Gamma(N_i \to \ell^+ + \ldots) \neq \Gamma(N_i \to \ell^- + \ldots) \ , \tag{13.39}$$

which would have resulted in unequal numbers of ℓ^+ and ℓ^- in the early universe [48]. This leptogenesis could have been followed by nonperturbative SM processes that would have converted the lepton asymmetry, in part, into the observed baryon asymmetry [49].

While the connection between the CP violation that would have led to leptogenesis, and that which we hope to observe in neutrino oscillation, is model-dependent, it is not likely that we have either of these without the other [50]. This makes the search for CP violation in neutrino oscillation very interesting indeed. Depending on the rough size of θ_{13}, this CP violation may be observable with a very intense conventional neutrino beam, or may require a "neutrino factory," whose neutrinos come from the decay of stored muons or radioactive nuclei. The detailed study of CP violation may require a neutrino factory in any case.

vi) Will we encounter the completely unexpected?

The study of neutrinos has been characterized by surprises. It would be surprising if further surprises were not in store. The possibilities include new, non-Standard-Model interactions, unexpectedly large magnetic and electric dipole moments, unexpectedly short lifetimes, and violations of CPT invariance, Lorentz invariance, or the equivalence principle.

The questions we have discussed, and other questions about the world of neutrinos, will be the focus of a major experimental program in the years to come.

Acknowledgements

I am grateful to Susan Kayser for her crucial role in the production of this manuscript.

References:

1. This matrix is sometimes referred to as the Maki-Nakagawa-Sakata matrix, or as the Pontecorvo-Maki-Nakagawa-Sakata matrix, in recognition of the pioneering contributions of these scientists to the physics of mixing and oscillation. See Z. Maki, M. Nakagawa, and S. Sakata, Prog. Theor. Phys. **28**, 870 (1962);
   B. Pontecorvo, Zh. Eksp. Teor. Fiz. **53**, 1717 (1967) [Sov. Phys. JETP **26**, 984 (1968)].
2. D. Karlen in this *Review*.
3. B. Kayser, Phys. Rev. **D24**, 110 (1981);
   F. Boehm and P. Vogel, *Physics of Massive Neutrinos* (Cambridge University Press, Cambridge, 1987) p. 87;
   C. Giunti, C. Kim, and U. Lee, Phys. Rev. **D44**, 3635 (1991);
   J. Rich, Phys. Rev. **D48**, 4318 (1993);
   H. Lipkin, Phys. Lett. **B348**, 604 (1995);
   W. Grimus and P. Stockinger, Phys. Rev. **D54**, 3414 (1996);
   T. Goldman, `hep-ph/9604357`;
   Y. Grossman and H. Lipkin, Phys. Rev. **D55**, 2760 (1997);
   W. Grimus, S. Mohanty, and P. Stockinger, in *Proc. of the 17th Int. Workshop on Weak Interactions and Neutrinos*, eds. C. Dominguez and R. Viollier (World Scientific, Singapore, 2000) p. 355.
4. L. Stodolsky, Phys. Rev. **D58**, 036006 (1998);
   C. Giunti, Phys. Scripta **67**, 29 (2003);
   M. Beuthe, Phys. Rept. **375**, 105 (2003) and Phys. Rev. **D66**, 013003 (2002), and references therein;
   H. Lipkin, Phys. Lett. **B579**, 355 (2004).
5. G. Fogli, E. Lisi, and G. Scioscia, Phys. Rev. **D52**, 5334 (1995).
6. L. Wolfenstein, Phys. Rev. **D17**, 2369 (1978).
7. S. Mikheyev and A. Smirnov, Yad. Fiz. **42**, 1441 (1985) [Sov. J. Nucl. Phys. **42**, 913 (1986)]; Zh. Eksp. Teor. Fiz. **91**, 7, (1986) [Sov. Phys. JETP **64**, 4 (1986)]; Nuovo Cimento **9C**, 17 (1986).
8. The Bugey Collaboration (B. Achkar *et al.*), Nucl. Phys. **B434**, 503 (1995);
   The Palo Verde Collaboration (F. Boehm *et al.*), Phys. Rev. **D64**, 112001 (2001);
   The CHOOZ Collaboration (M. Apollonio *et al.*), Eur. Phys. J. **C27**, 331 (2003).
9. P. Langacker, J. Leveille, and J. Sheiman, Phys. Rev. **D27**, 1228 (1983);
   The corresponding energy for anti-neutrinos is negative.
10. G. L. Fogli, E. Lisi, and D. Montanino, Phys. Rev. **D54**, 2048 (1996);
    A. de Gouvêa, A. Friedland, and H. Murayama, Phys. Lett. **B490**, 125 (2000).
11. J. Bahcall, *Neutrino Astrophysics*, (Cambridge Univ. Press, Cambridge, UK 1989).
12. S. Parke, Phys. Rev. Lett. **57**, 1275 (1986).
13. We thank J. Beacom and A. Smirnov for invaluable conversations on how LMA-MSW works. For an early description, see S. Mikheyev and A. Smirnov, Ref. 7 (first paper).
14. D. Ayres *et al.*, in *Proc. of the 1982 DPF Summer Study on Elementary Particle Physics and Future Facilities*, p. 590;

G. Dass and K. Sarma, Phys. Rev. **D30**, 80 (1984);
J. Flanagan, J. Learned, and S. Pakvasa, Phys. Rev. **D57**, 2649 (1998);
B. Kayser, in *Proc. of the 17th Int. Workshop on Weak Interactions and Neutrinos*, eds. C. Dominguez and R. Viollier (World Scientific, Singapore, 2000) p. 339.

15. The Super-Kamiokande Collaboration (Y. Ashie *et al.*), Phys. Rev. **D71**, 112005 (2005).
16. From the L/E analysis reported by the Super-Kamiokande Collaboration (Y. Ashie *et al.*), Phys. Rev. Lett. **93**, 101801 (2004).
17. MACRO Collaboration (M. Ambrosio *et al.*), Phys. Lett. **B566**, 35 (2003);
MACRO Collaboration, Eur. Phys. J. C, **36**, 323 (2004).
18. M. Sanchez *et al.*, Phys. Rev. **D68**, 113004 (2003);
W. W. M. Allison *et al.*, Phys. Rev. **D72**, 052005 (2005).
19. M. Shiozawa, presented at the XXth Int. Conf. on Neutrino Physics and Astrophysics, Munich, May, 2002.
20. Y. Suzuki, presented at the XXII Int. Symp. on Lepton and Photon Interactions at High Energies (Lepton Photon 2005), Uppsala, Sweden, July, 2005.
21. The SNO Collaboration (B. Aharmim *et al.*), `nucl-ex/0502021`.
22. Y. Koshio, in *2003 Electroweak Interactions and Unified Theories* (Proceedings of the 38th Rencontres de Moriond), ed. J. Trân Thanh Vân (The Gioi, Vietnam, 2003) p. 3;
The Super-Kamiokande Collaboration (J. Hosaka *et al.*), `hep-ex/0508053`.
23. J. Bahcall, S. Basu, and A. Serenelli, Astrophys. J. **621**, L85 (2005).
24. G. Fogli, E. Lisi, A. Marrone, and A. Palazzo, Phys. Lett. **B583**, 149 (2004).
25. The KamLAND Collaboration (T. Araki *et al.*), Phys. Rev. Lett. **94**, 081801 (2005).
26. The latter include the data from the chlorine experiment: B. Cleveland *et al.*, Astrophys. J. **496**, 505 (1998);
and from the gallium experiments: GNO Collaboration (M. Altmann *et al.*), Phys. Lett. **B616**, 174 (2005);
SAGE Collaboration (J. N. Abdurashitov *et al.*), JETP **95**, 181 (2002).
27. J. Bahcall, M.C. Gonzalez-Garcia, and C. Peña-Garay, JHEP **0408**, 016 (2004);
See also P. Holanda and A. Smirnov, `hep-ph/0211264` and Phys. Rev. **D69**, 113002 (2004).
28. The LSND Collaboration (A. Aguilar *et al.*), Phys. Rev. **D64**, 112007 (2001).
29. The KARMEN Collaboration (B. Armbruster *et al.*), Phys. Rev. **D65**, 112001 (2002).
30. E. Church *et al.*, Phys. Rev. **D66**, 013001 (2003).
31. For an alternative possibility entailing CPT violation, see H. Murayama and T. Yanagida, Phys. Lett. **B520**, 263 (2001);
G. Barenboim *et al.*, JHEP **0210**, 001 (2002);
However, after KamLAND, this alternative is disfavored. M. C. Gonzalez-Garcia, M. Maltoni, and T. Schwetz, Phys. Rev. **D68**, 053007 (2003);
G. Barenboim, L. Borissov, and J. Lykken, `hep-ph/0212116 v2`.

32. J. Schechter and J. Valle, Phys. Rev. **D23**, 1666 (1981);
J. Nieves and P. Pal, Phys. Rev. **D64**, 076005 (2001);
A. de Gouvêa, B. Kayser, and R. Mohapatra, Phys. Rev. **D67**, 053004 (2003).
33. L.-L. Chau and W.-Y. Keung, Phys. Rev. Lett. **53**, 1802 (1984);
H. Harari and M. Leurer, Phys. Lett. **B181**, 123 (1986);
F.J. Botella and L.-L. Chau, Phys. Lett. **B168**, 97 (1986);
H. Fritzsch and J. Plankl, Phys. Rev. **D35**, 1732 (1987).
34. G. Fogli, E. Lisi, A. Marrone, and A. Palazzo, `hep-ph/0506083`, to appear in Prog. Part. Nucl. Phys..
35. M. Sorel, J. Conrad, and M. Shaevitz, Phys. Rev. **D70**, 073004 (2004).
36. We thank Belen Gavela for introducing us to this argument.
37. M. Gell-Mann, P. Ramond, and R. Slansky, in: *Supergravity*, eds. D. Freedman and P. van Nieuwenhuizen (North Holland, Amsterdam, 1979) p. 315;
T. Yanagida, in: *Proceedings of the Workshop on Unified Theory and Baryon Number in the Universe,* eds. O. Sawada and A. Sugamoto (KEK, Tsukuba, Japan, 1979);
R. Mohapatra and G. Senjanovic: Phys. Rev. Lett. **44**, 912 (1980) and Phys. Rev. **D23**, 165 (1981);
P. Minkowski, Phys. Lett. **B67**, 421 (1977).
38. J. Schechter and J. Valle, Phys. Rev. **D25**, 2951 (1982).
39. The physics of Majorana neutrinos and $0\nu\beta\beta$ are discussed in S. Bilenky and S. Petcov, Rev. Mod. Phys. **59**, 671 (1987) [Erratum–*ibid.* **61**, 169 (1987)];
B. Kayser, F. Gibrat-Debu, and F. Perrier, *The Physics of Massive Neutrinos* (World Scientific, Singapore, 1989);
B. Kayser, `hep-ph/0504052,` to appear in *Neutrino Physics, Proceedings of the Nobel Symposium 2004*, eds. L. Bergström, O Botner, P. Carlson, P.O. Hulth, and T. Ohlsson.
40. S. Pascoli and S.T. Petcov, Phys. Lett. **B580**, 280 (2003).
41. Analyses of the possible values of $| < m_{\beta\beta} > |$ have been given by H. Murayama and C. Peña-Garay, Phys. Rev. **D69**, 031301 (2004);
S. Pascoli and S. Petcov, Phys. Lett. **B544**, 239 (2002);
S. Bilenky, S. Pascoli, and S. Petcov, Phys. Rev. **D64**, 053010 (2001), and Phys. Rev. **D64**, 113003 (2001);
H. Klapdor-Kleingrothaus, H. Päs, and A. Smirnov, Phys. Rev. **D63**, 073005 (2001);
S. Bilenky *et al.*, Phys. Lett. **B465**, 193 (1999);
References in these papers.
42. See also S. Elliott and P Vogel, Ann. Rev. Nucl. Part. Sci. **52**, 115 (2002), and references therein.
43. The MiniBooNE Collaboration (E. Church *et al.*) FERMILAB-P-0898 (1997), available at `http://library.fnal.gov/archive/test-proposal/0000/fermilab-proposal-0898.shtml`.
44. We thank Carl Albright for discussions of this point.
45. S. Pastor, in the *Proceedings of the Eleventh International Workshop on Neutrino Telescopes,* ed. M. Baldo Ceolin (Italy, 2005) p. 371;

U. Seljak *et al.*, Phys. Rev. **D71**, 103515 (2005).
46. K. McConnel and M. Shaevitz, `hep-ex/0409028`.
47. For an alternative approach to determining $\theta_{\rm atm}$, see M. Gonzalez-Garcia, M. Maltoni and A. Smirnov, Phys. Rev. D **70**, 093005 (2004), and references therein..
48. M. Fukugita and T. Yanagida, Phys. Lett. **B174**, 45 (1986).
49. G. 't Hooft, Phys. Rev. Lett. **37**, 8 (1976);
V. Kuzmin, V. Rubakov, and M. Shaposhnikov, Phys. Lett. **155B**, 36 (1985).
50. S. Pascoli, S. Petcov, and W. Rodejohann, Phys. Rev. **D68**, 093007 (2003);
S. Davidson, S. Pascoli, and S. Petcov, private communications.

14. QUARK MODEL

Revised December 2005 by C. Amsler (University of Zürich), T. DeGrand (University of Colorado, Boulder) and B. Krusche (University of Basel).

14.1. Quantum numbers of the quarks

Quarks are strongly interacting fermions with spin 1/2 and, by convention, positive parity. Then antiquarks have negative parity. Quarks have the additive baryon number 1/3, antiquarks -1/3. Table 14.1 gives the other additive quantum numbers (flavors) for the three generations of quarks. They are related to the charge Q (in units of the elementary charge e) through the generalized Gell-Mann-Nishijima formula

$$\mathsf{Q} = \mathsf{I}_z + \frac{\mathcal{B} + \mathsf{S} + \mathsf{C} + \mathsf{B} + \mathsf{T}}{2} \ , \tag{14.1}$$

where $\mathcal{B}$ is the baryon number. The convention is that the *flavor* of a quark (I_z, S, C, B, or T) has the same sign as its *charge* Q. With this convention, any flavor carried by a charged meson has the same sign as its charge, e.g. the strangeness of the K^+ is +1, the bottomness of the B^+ is +1, and the charm and strangeness of the D_s^- are each −1. Antiquarks have the opposite flavor signs.

Table 14.1: Additive quantum numbers of the quarks.

Property \ Quark	d	u	s	c	b	t
Q – electric charge	$-\frac{1}{3}$	$+\frac{2}{3}$	$-\frac{1}{3}$	$+\frac{2}{3}$	$-\frac{1}{3}$	$+\frac{2}{3}$
I – isospin	$\frac{1}{2}$	$\frac{1}{2}$	0	0	0	0
I_z – isospin z-component	$-\frac{1}{2}$	$+\frac{1}{2}$	0	0	0	0
S – strangeness	0	0	−1	0	0	0
C – charm	0	0	0	+1	0	0
B – bottomness	0	0	0	0	−1	0
T – topness	0	0	0	0	0	+1

14.2. Mesons

Mesons have baryon number $\mathcal{B} = 0$. In the quark model they are $q\overline{q}'$ bound states of quarks q and antiquarks $\overline{q}'$ (the flavors of q and q' may be different). If the orbital angular momentum of the $q\overline{q}'$ state is ℓ, then the parity P is $(-1)^{\ell+1}$. The meson spin J is given by the usual relation $|\ell - s| < J < |\ell + s|$ where s is 0 (antiparallel quark spins) or 1 (parallel quark spins). The charge conjugation, or C-parity $C = (-1)^{\ell+s}$, is defined only for the $q\bar{q}$ states made of quarks and their own antiquarks. The C-parity can be generalized to the G-parity $G = (-1)^{I+\ell+s}$ for mesons made of quarks and their own antiquarks (isospin $\mathsf{I}_z = 0$) and for the charged $u\bar{d}$ and $d\bar{u}$ states (isospin $\mathsf{I} = 1$).

The mesons are classified in J^{PC} multiplets. The $\ell = 0$ states are the pseudoscalars (0^{-+}) and the vectors (1^{--}). The orbital excitations $\ell = 1$ are the scalars (0^{++}), the axial vectors (1^{++}) and (1^{+-}), and the tensors (2^{++}). Assignments for many of the known mesons are given in Tables 14.2 and 14.3. Radial excitations are denoted by the principal quantum number n. The very short lifetime of the t quark makes it likely that bound state hadrons containing t quarks and/or antiquarks do not exist.

States in the natural spin-parity series $P = (-1)^J$ must, according to the above, have $s = 1$ and hence $CP = +1$. Thus mesons with natural spin-parity and $CP = -1$ (0^{+-}, 1^{-+}, 2^{+-}, 3^{-+}, etc) are forbidden in the $q\bar{q}'$ model. The $J^{PC} = 0^{--}$ state is forbidden as well. Mesons with such *exotic* quantum numbers may exist, but would lie outside the $q\bar{q}'$ model (see section below on exotic mesons).

Following SU(3) the nine possible $q\bar{q}'$ combinations containing the light u, d, and s quarks are grouped into an octet and a singlet of light quark mesons:

$$\mathbf{3} \otimes \overline{\mathbf{3}} = \mathbf{8} \oplus \mathbf{1} \ . \tag{14.2}$$

A fourth quark such as charm c can be included by extending SU(3) to SU(4). However, SU(4) is badly broken owing to the much heavier c quark. Nevertheless, in an SU(4) classification the sixteen mesons are grouped into a 15-plet and a singlet:

$$\mathbf{4} \otimes \overline{\mathbf{4}} = \mathbf{15} \oplus \mathbf{1} \ . \tag{14.3}$$

The *weight diagrams* for the ground-state pseudoscalar (0^{-+}) and vector (1^{--}) mesons are depicted in Fig. 14.1. The light quark mesons are members of nonets building the middle plane in Fig. 14.1(a) and (b).

Isoscalar states with the same J^{PC} will mix but mixing between the two light quark isoscalar mesons and the much heavier charmonium or bottomonium states are generally assumed to be negligible. In the following we shall use the generic names a for the $I = 1$, K for the $I = 1/2$, f and f' for the $I = 0$ members of the light quark nonets. Thus the physical isoscalars are mixtures of the SU(3) wave function ψ_8 and ψ_1:

$$f' = \psi_8 \cos\theta - \psi_1 \sin\theta \ , \tag{14.4}$$

$$f = \psi_8 \sin\theta + \psi_1 \cos\theta \ , \tag{14.5}$$

where θ is the nonet mixing angle and

$$\psi_8 = \frac{1}{\sqrt{6}}(u\bar{u} + d\bar{d} - 2s\bar{s}) \ , \tag{14.6}$$

$$\psi_1 = \frac{1}{\sqrt{3}}(u\bar{u} + d\bar{d} + s\bar{s}) \ . \tag{14.7}$$

The mixing angle has to be determined experimentally.

These mixing relations are often rewritten to exhibit the $u\bar{u} + d\bar{d}$ and $s\bar{s}$ components which decouple for the "ideal" mixing angle θ_i such that $\tan\theta_i = 1/\sqrt{2}$ (or θ_i=35.3°). Defining $\alpha = \theta + 54.7°$, one obtains the physical isoscalar in the flavor basis

$$f' = \frac{1}{\sqrt{2}}(u\bar{u} + d\bar{d})\cos\alpha - s\bar{s} \ \sin\alpha \ , \tag{14.8}$$

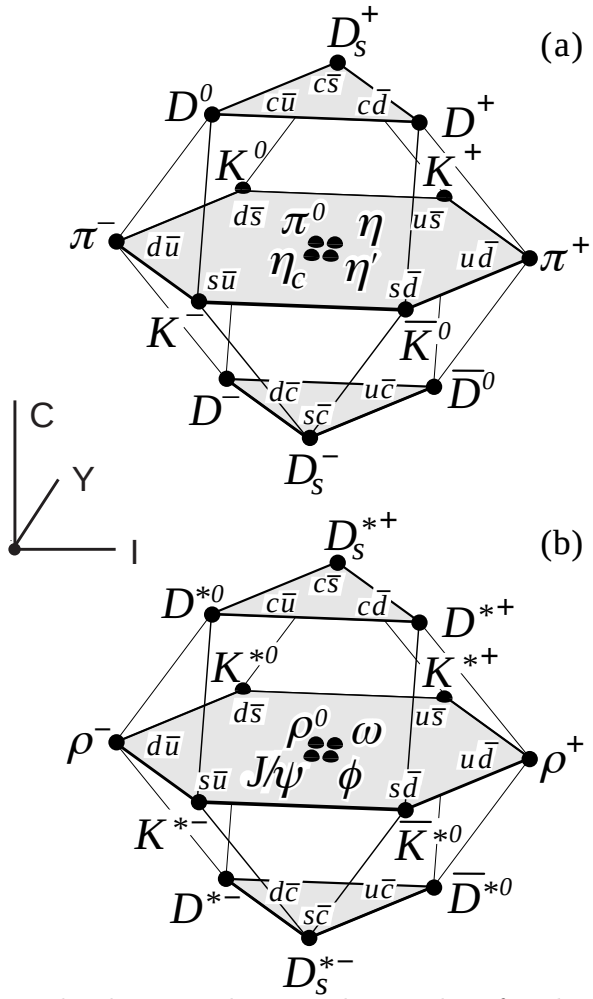


Figure 14.1: SU(4) weight diagram showing the 16-plets for the pseudoscalar (a) and vector mesons (b) made of the u, d, s and c quarks as a function of isospin I, charm C and hypercharge $\mathsf{Y} = \mathsf{S}+\mathcal{B}-\frac{\mathsf{C}}{3}$. The nonets of light mesons occupy the central planes to which the $c\bar{c}$ states have been added.

and its orthogonal partner f (replace α by α – 90°). Thus for ideal mixing ($\alpha_i = 90°$) the f' becomes pure $s\bar{s}$ and the f pure $u\bar{u} + d\bar{d}$. The mixing angle θ can be derived from the mass relation

$$\tan\theta = \frac{4m_K - m_a - 3m_{f'}}{2\sqrt{2}(m_a - m_K)} \ , \tag{14.9}$$

which also determines its sign or, alternatively, from

$$\tan^2\theta = \frac{4m_K - m_a - 3m_{f'}}{-4m_K + m_a + 3m_f} \ . \tag{14.10}$$

Eliminating θ from these equations leads to the sum rule [1]

$$(m_f + m_{f'})(4m_K - m_a) - 3m_f m_{f'} = 8m_K^2 - 8m_K m_a + 3m_a^2. \tag{14.11}$$

This relation is verified for the ground-state vector mesons. We identify the $\phi(1020)$ with the f' and the $\omega(783)$ with the f. Thus

$$\phi(1020) = \psi_8 \cos\theta_V - \psi_1 \sin\theta_V \ , \tag{14.12}$$

$$\omega(782) = \psi_8 \sin\theta_V + \psi_1 \cos\theta_V \ , \tag{14.13}$$

with the vector mixing angle $\theta_V = 35^\circ$ from Eq. (14.9), very close to ideal mixing. Thus $\phi(1020)$ is nearly pure $s\bar{s}$. For ideal mixing Eq. (14.9) and Eq. (14.10) lead to the relations

$$m_K = \frac{m_f + m_{f'}}{2} \ , \quad m_a = m_f \ , \tag{14.14}$$

which are satisfied for the vector mesons. However, for the pseudoscalar (and scalar mesons) Eq. (14.11) is satisfied only approximately. Then Eq. (14.9) and Eq. (14.10) lead to somewhat different values for the mixing angle. Identifying the η with the f' one gets

$$\eta = \psi_8 \cos\theta_P - \psi_1 \sin\theta_P \ , \tag{14.15}$$

$$\eta' = \psi_8 \sin\theta_P + \psi_1 \cos\theta_P \ . \tag{14.16}$$

Following chiral perturbation theory the meson masses in the mass formulae (Eq. (14.9) and Eq. (14.10)) should be replaced by their squares. Table 14.2 lists the mixing angle $\theta_{\rm lin}$ from Eq. (14.10) and the corresponding $\theta_{\rm quad}$ obtained by replacing the meson masses by their squares throughout.

The pseudoscalar mixing angle θ_P can also be measured by comparing the partial widths for radiative J/ψ decay into a vector and a pseudoscalar [2], radiative $\phi(1020)$ decay into η and η' [3], or $\bar{p}p$ annihilation at rest into a pair of vector and pseudoscalar or into two pseudoscalars [4,5]. One obtains a mixing angle between –10° and –20°.

The nonet mixing angles can be measured in $\gamma\gamma$ collisions, *e.g.* for the 0^{-+}, 0^{++} and 2^{++} nonets. In the quark model the amplitude for the coupling of neutral mesons to two photons is proportional to $\sum_i Q_i^2$, where Q_i is the charge of the i-th quark. The 2γ partial width of an isoscalar meson with mass m is then given in terms of the mixing angle α by

$$\Gamma_{2\gamma} = C(5\cos\alpha - \sqrt{2}\sin\alpha)^2 m^3 \ , \tag{14.17}$$

for f' and f ($\alpha \to \alpha - 90^\circ$). The coupling C may depend on the meson mass. It is often assumed to be a constant in the nonet. For the isovector a one then finds $\Gamma_{2\gamma} = 9\ C\ m^3$. Thus the members of an ideally mixed nonet couple to 2γ with partial widths in the ratios $f : f' : a = 25 : 2 : 9$. For tensor mesons one finds from the ratios of the measured 2γ partial widths for the $f_2(1270)$ and $f_2'(1525)$ mesons a mixing angle α_T of $(81\pm 1)^\circ$, or $\theta_T = (27 \pm 1)^\circ$, in accord with the linear mass formula. For the pseudoscalars one finds from the ratios of partial widths $\Gamma(\eta' \to 2\gamma)/\Gamma(\eta \to 2\gamma)$ a mixing angle $\theta_P = (-18 \pm 2)^\circ$ while the ratio $\Gamma(\eta' \to 2\gamma)/\Gamma(\pi^0 \to 2\gamma)$ leads to $\sim -24\ ^\circ$. SU(3) breaking effects for pseudoscalars are discussed in Ref. 6.

Table 14.2: Suggested $q\overline{q}$ quark-model assignments for some of the observed light mesons. Mesons in bold face are included in the Meson Summary Table. The wave functions f and f' are given in the text. The singlet-octet mixing angles from the quadratic and linear mass formulae are also given for the well established nonets. The classification of the 0^{++} mesons is tentative and the mixing angle uncertain due to large uncertainties in some of the masses. Also, the $f_0(1710)$ and $f_0(1370)$ are expected to mix with the $f_0(1500)$. The latter is not in this table as it is hard to accommodate in the scalar nonet. The light scalars $a_0(980)$, $f_0(980)$ and $f_0(600)$ are often considered as meson-meson resonances or four-quark states and are therefore not included in the table. See the "Note on Non-$q\overline{q}$ Mesons" at the end of the Meson Listings.

$n\,^{2s+1}\ell_J$	J^{PC}	I = 1 $u\overline{d}$, $\overline{u}d$, $\frac{1}{\sqrt{2}}(d\overline{d}-u\overline{u})$	I = $\frac{1}{2}$ $u\overline{s}$, $d\overline{s}$; $\overline{d}s$, $-\overline{u}s$	I = 0 f'	I = 0 f	$\theta_{\rm quad}$ [°]	$\theta_{\rm lin}$ [°]
$1\,^1S_0$	0^{-+}	$\boldsymbol{\pi}$	$\boldsymbol{K}$	$\boldsymbol{\eta}$	$\boldsymbol{\eta'(958)}$	−11.5	−24.6
$1\,^3S_1$	1^{--}	$\boldsymbol{\rho(770)}$	$\boldsymbol{K^*(892)}$	$\boldsymbol{\phi(1020)}$	$\boldsymbol{\omega(782)}$	38.7	36.0
$1\,^1P_1$	1^{+-}	$\boldsymbol{b_1(1235)}$	$\boldsymbol{K_{1B}}$†	$h_1(1380)$	$\boldsymbol{h_1(1170)}$		
$1\,^3P_0$	0^{++}	$\boldsymbol{a_0(1450)}$	$\boldsymbol{K_0^*(1430)}$	$\boldsymbol{f_0(1710)}$	$\boldsymbol{f_0(1370)}$		
$1\,^3P_1$	1^{++}	$\boldsymbol{a_1(1260)}$	$\boldsymbol{K_{1A}}$†	$\boldsymbol{f_1(1420)}$	$\boldsymbol{f_1(1285)}$		
$1\,^3P_2$	2^{++}	$\boldsymbol{a_2(1320)}$	$\boldsymbol{K_2^*(1430)}$	$\boldsymbol{f_2'(1525)}$	$\boldsymbol{f_2(1270)}$	29.6	28.0
$1\,^1D_2$	2^{-+}	$\boldsymbol{\pi_2(1670)}$	$\boldsymbol{K_2(1770)}$†	$\eta_2(1870)$	$\boldsymbol{\eta_2(1645)}$		
$1\,^3D_1$	1^{--}	$\boldsymbol{\rho(1700)}$	$\boldsymbol{K^*(1680)}$‡		$\boldsymbol{\omega(1650)}$		
$1\,^3D_2$	2^{--}		$\boldsymbol{K_2(1820)}$‡				
$1\,^3D_3$	3^{--}	$\boldsymbol{\rho_3(1690)}$	$\boldsymbol{K_3^*(1780)}$	$\boldsymbol{\phi_3(1850)}$	$\boldsymbol{\omega_3(1670)}$	32.0	31.0
$1\,^3F_4$	4^{++}	$\boldsymbol{a_4(2040)}$	$\boldsymbol{K_4^*(2045)}$		$\boldsymbol{f_4(2050)}$		
$1\,^3G_5$	5^{--}	$\rho_5(2350)$					
$1\,^3H_6$	6^{++}	$a_6(2450)$			$f_6(2510)$		
$2\,^1S_0$	0^{-+}	$\boldsymbol{\pi(1300)}$	$K(1460)$	$\boldsymbol{\eta(1475)}$	$\boldsymbol{\eta(1295)}$		
$2\,^3S_1$	1^{--}	$\boldsymbol{\rho(1450)}$	$\boldsymbol{K^*(1410)}$‡	$\boldsymbol{\phi(1680)}$	$\boldsymbol{\omega(1420)}$		

† The $1^{+\pm}$ and $2^{-\pm}$ isospin $\frac{1}{2}$ states mix. In particular, the K_{1A} and K_{1B} are nearly equal (45°) mixtures of the $K_1(1270)$ and $K_1(1400)$.
‡ The $K^*(1410)$ could be replaced by the $K^*(1680)$ as the $2\,^3S_1$ state.

Table 14.3: $q\overline{q}$ quark-model assignments for the observed heavy mesons. Mesons in bold face are included in the Meson Summary Table.

$n\,^{2s+1}\ell_J$ J^{PC}	I = 0 $c\overline{c}$	I = 0 $b\overline{b}$	I = $\frac{1}{2}$ $c\overline{u}$, $c\overline{d}$; $\overline{c}u$, $\overline{c}d$	I = 0 $c\overline{s}$; $\overline{c}s$	I = $\frac{1}{2}$ $b\overline{u}$, $b\overline{d}$; $\overline{b}u$, $\overline{b}d$	I = 0 $b\overline{s}$; $\overline{b}s$	I = 0 $b\overline{c}$; $\overline{b}c$
$1\,^1S_0$ 0^{-+}	$\boldsymbol{\eta_c(1S)}$	$\eta_b(1S)$	$\boldsymbol{D}$	$\boldsymbol{D_s^{\pm}}$	$\boldsymbol{B}$	$\boldsymbol{B_s^0}$	$\boldsymbol{B_c^{\pm}}$
$1\,^3S_1$ 1^{--}	$\boldsymbol{J/\psi(1S)}$	$\boldsymbol{\Upsilon(1S)}$	$\boldsymbol{D^*}$	$\boldsymbol{D_s^{*\pm}}$	$\boldsymbol{B^*}$	B_s^*	
$1\,^1P_1$ 1^{+-}	$h_c(1P)$		$\boldsymbol{D_1(2420)}$	$\boldsymbol{D_{s1}(2536)^{\pm}}$			
$1\,^3P_0$ 0^{++}	$\boldsymbol{\chi_{c0}(1P)}$	$\boldsymbol{\chi_{b0}(1P)}$		$\boldsymbol{D_{s0}^*(2317)^{\pm\dagger}}$			
$1\,^3P_1$ 1^{++}	$\boldsymbol{\chi_{c1}(1P)}$	$\boldsymbol{\chi_{b1}(1P)}$		$\boldsymbol{D_{s1}(2460)^{\pm\dagger}}$			
$1\,^3P_2$ 2^{++}	$\boldsymbol{\chi_{c2}(1P)}$	$\boldsymbol{\chi_{b2}(1P)}$	$\boldsymbol{D_2^*(2460)}$	$\boldsymbol{D_{s2}(2573)^{\pm}}$			
$1\,^3D_1$ 1^{--}	$\boldsymbol{\psi(3770)}$						
$2\,^1S_0$ 0^{-+}	$\boldsymbol{\eta_c(2S)}$						
$2\,^3S_1$ 1^{--}	$\boldsymbol{\psi(2S)}$	$\boldsymbol{\Upsilon(2S)}$					
$2\,^3P_{0,1,2}$ $0^{++}, 1^{++}, 2^{++}$		$\boldsymbol{\chi_{b0,1,2}(2P)}$					

† The masses of these states are considerably smaller than most theoretical predictions. They have also been considered as four-quark states (See the "Note on Non-$q\overline{q}$ Mesons" at the end of the Meson Listings). The $D_{s1}(2460)^{\pm}$ and $D_{s1}(2536)^{\pm}$ are mixtures of the $1^{+\pm}$ states.

Table 14.4: SU(3) couplings γ^2 for quarkonium decays as a function of nonet mixing angle α, up to a common multiplicative factor C ($\phi \equiv 54.7^\circ + \theta_P$).

Isospin	Decay channel	γ^2
0	$\pi\pi$	$3\cos^2\alpha$
	$K\overline{K}$	$(\cos\alpha - \sqrt{2}\sin\alpha)^2$
	$\eta\eta$	$(\cos\alpha\cos^2\phi - \sqrt{2}\sin\alpha\sin^2\phi)^2$
	$\eta\eta'$	$\frac{1}{2}\sin^2 2\phi\ (\cos\alpha + \sqrt{2}\sin\alpha)^2$
1	$\eta\pi$	$2\cos^2\phi$
	$\eta'\pi$	$2\sin^2\phi$
	$K\overline{K}$	1
$\frac{1}{2}$	$K\pi$	$\frac{3}{2}$
	$K\eta$	$(\sin\phi - \frac{\cos\phi}{\sqrt{2}})^2$
	$K\eta'$	$(\cos\phi + \frac{\sin\phi}{\sqrt{2}})^2$

The partial width for the decay of a scalar or a tensor meson into a pair of pseudoscalar mesons is model dependent. Following Ref. 7,

$$\Gamma = C \times \gamma^2 \times |F(q)|^2 \times q\ . \tag{14.18}$$

C is a nonet constant, q the momentum of the decay products, $F(q)$ a form factor and γ^2 the SU(3) coupling. The model-dependent form factor may be written as

$$|F(q)|^2 = q^{2\ell} \times \exp(-\frac{q^2}{8\beta^2}), \tag{14.19}$$

where ℓ is the relative angular momentum between the decay products. The decay of a $q\bar{q}$ meson into a pair of mesons involves the creation of a $q\bar{q}$ pair from the vacuum and SU(3) symmetry assumes that the matrix elements for the creation of $s\bar{s}$, $u\bar{u}$ and $d\bar{d}$ pairs are equal. The couplings γ^2 are given in Table 14.4 and their dependence upon the mixing angle α is shown in Fig. 14.2 for isoscalar decays. The generalization to unequal $s\bar{s}$, $u\bar{u}$ and $d\bar{d}$ couplings is given in Ref. 7. An excellent fit to the tensor meson decay widths is obtained assuming SU(3) symmetry, with $\beta \simeq 0.5$ GeV/c, $\theta_V \simeq 26\ ^\circ$ and $\theta_P \simeq$ -17 $^\circ$ [7].

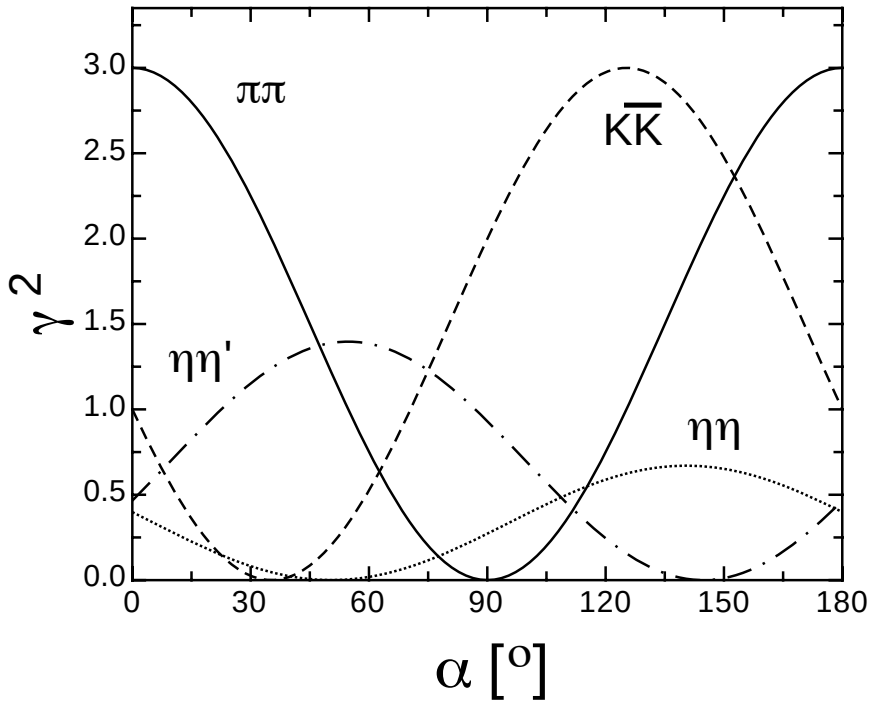


Figure 14.2: SU(3) couplings as a function of mixing angle α for isoscalar decays, up to a common multiplicative factor C and for $\theta_P = -17.3^\circ$ (from Ref. 4).

14.3. Exotic mesons

The existence of a light nonet composed of four quarks with masses below 1 GeV was suggested a long time ago [8]. Coupling two triplets of light quarks u, d and s one obtains nine states, of which the six symmetric ($uu,\ dd,\ ss,\ ud+du,\ us+su, ds+sd$) form the six dimensional representation **6**, while the three antisymmetric ($ud-du,\ us-su,\ ds-sd$) form the three dimensional representation $\overline{\mathbf{3}}$ of SU(3):

$$\mathbf{3} \otimes \mathbf{3} = \mathbf{6} \oplus \overline{\mathbf{3}} \ . \tag{14.20}$$

Combining with spin and color and requiring antisymmetry, one finds that the most deeply bound diquark (and hence the lightest) is the one in the $\overline{\mathbf{3}}$ and spin singlet state. The combination of the diquark with an antidiquark in the **3** representation then gives a light nonet of four-quark scalar states. Letting the number of strange quarks determine the mass splitting one obtains a mass inverted spectrum with a light isosinglet ($ud\bar{u}\bar{d}$), a medium heavy isodoublet (e.g. $ud\bar{s}\bar{d}$) and a heavy isotriplet (e.g. $ds\bar{u}\bar{s}$) + isosinglet (e.g. $us\bar{u}\bar{s}$). It is then tempting to identify the lightest state with the $f_0(600)$, and the heaviest states with the $a_0(980)$, and $f_0(980)$. Then the meson with strangeness $\kappa(800)$ would lie in between.

QCD predicts the existence of isoscalar mesons which contain only gluons, the glueballs. The ground state glueball is predicted by lattice gauge theories to be 0^{++}, the first excited state 2^{++}. Errors on the mass predictions are large. As an example for the glueball mass spectrum we show in Figure 14.3 a recent calculation from the lattice [9](see also [10]) . This group predicts a mass of 1710 MeV for the ground state, with an error of about 100 MeV. Earlier work by other groups produced masses at 1650 MeV [11] and 1550 MeV [12](see also Ref. 13). The first excited state has a mass of about 2.4 GeV and the lightest glueball with exotic quantum numbers (2^{+-}) has a mass of about 4 GeV.

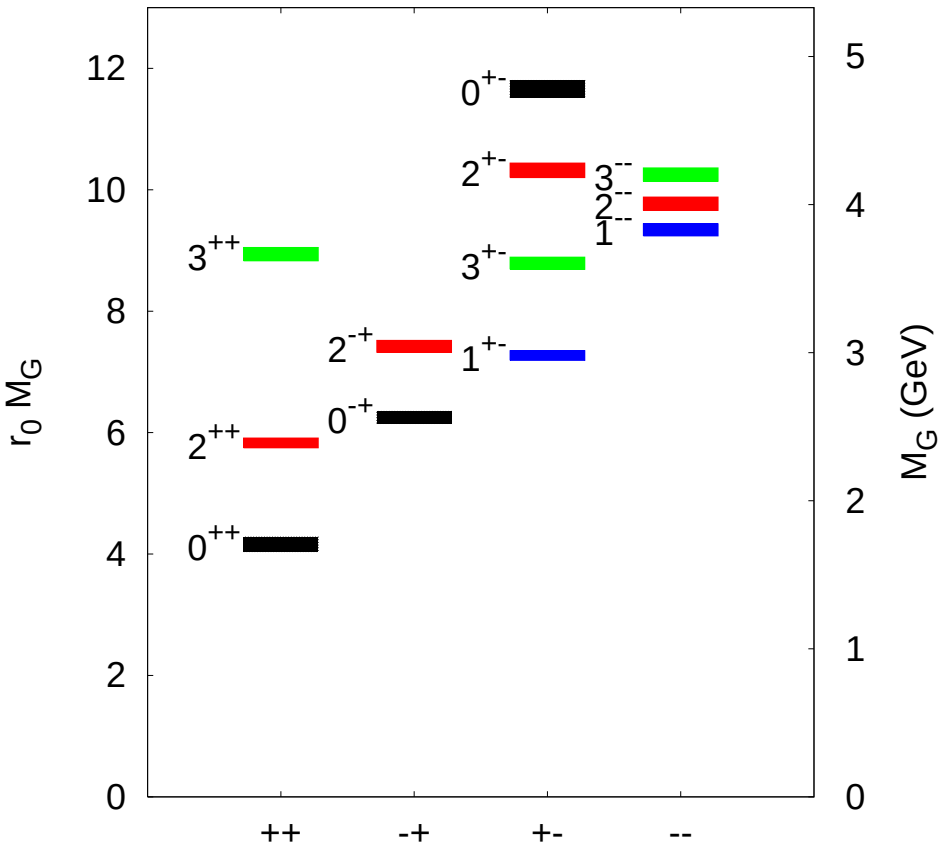


Figure 14.3: Predicted glueball mass spectrum from the lattice (from Ref. 9). See full-color version on color pages at end of book.

These lattice calculations assumed that the quark masses are infinite and neglect $q\bar{q}$ loops. However, one expects that glueballs will mix with nearby $q\bar{q}$ states of the same quantum numbers [7,14]. For example, the two isoscalar 0^{++} mesons will mix with the pure ground state glueball to generate the observed physical states $f_0(1370)$, $f_0(1500)$, and $f_0(1710)$. Experimental evidence is mounting that the $f_0(1500)$ has considerable affinity for glue and that the $f_0(1370)$ and $f_0(1710)$ have large $u\bar{u} + d\bar{d}$ and $s\bar{s}$ components, respectively. The $f_0(1710)$ has also been proposed as a candidate scalar glueball [11](see the "Note on Non-$q\overline{q}$ Mesons" at the end of the Meson Listings and Ref. 15).

Mesons made of $q\bar{q}$ pairs bound by excited gluons g, the hybrid states $q\bar{q}g$, are also predicted. They should lie in the 1.9 GeV mass region, according to gluon flux tube models [16]. Lattice QCD also predicts the lightest hybrid, an exotic 1^{-+}, at a mass of 1.9 GeV [17]. However, the bag model predicts four nonets, among them an exotic 1^{-+} around or above 1.4 GeV [18,19]. There are so far two candidates for exotic states with quantum numbers 1^{-+}, the $\pi_1(1400)$ and $\pi_1(1600)$, which could be hybrids or four-quark states (see the "Note on Non-$q\overline{q}$ Mesons" at the end of the Meson Listings and Ref. 15).

14.4. Baryons: *qqq* states

Baryons are fermions with baryon number $\mathcal{B} = 1$, i.e. in the most general case they are composed of three quarks plus any number of quark - antiquark pairs. Although recently some experimental evidence for ($qqqq\bar{q}$) pentaquark states has been claimed (see review on Possible exotic baryon resonance), so far all established baryons are 3-quark (qqq) configurations. The color part of their state functions is an SU(3) singlet, a completely antisymmetric state of the three colors. Since the quarks are fermions, the state function must be antisymmetric under interchange of any two equal-mass quarks (up and down quarks in the limit of isospin symmetry). Thus it can be written as

$$|\, qqq \,\rangle_A = |\, \text{color} \,\rangle_A \times |\, \text{space, spin, flavor} \,\rangle_S \ , \tag{14.21}$$

where the subscripts S and A indicate symmetry or antisymmetry under interchange of any two equal-mass quarks. Note the contrast with the state function for the three nucleons in ^{3}H or ^{3}He:

$$|\, NNN \,\rangle_A = |\, \text{space, spin, isospin} \,\rangle_A \ . \tag{14.22}$$

This difference has major implications for internal structure, magnetic moments, *etc.* (For a nice discussion, see Ref. 20.)

The "ordinary" baryons are made up of u, d, and s quarks. The three flavors imply an approximate flavor SU(3), which requires that baryons made of these quarks belong to the multiplets on the right side of

$$\mathbf{3} \otimes \mathbf{3} \otimes \mathbf{3} = \mathbf{10}_S \oplus \mathbf{8}_M \oplus \mathbf{8}_M \oplus \mathbf{1}_A \tag{14.23}$$

(see Sec. 37, on "SU(n) Multiplets and Young Diagrams"). Here the subscripts indicate symmetric, mixed-symmetry, or antisymmetric states under interchange of any two quarks. The **1** is a uds state (Λ_1) and the octet contains a similar state (Λ_8). If these have the same spin and parity they can mix. The mechanism is the same as for the mesons (see above). In the ground state multiplet, the SU(3) flavor singlet Λ_1 is forbidden by Fermi statistics. Section 36, on "SU(3) Isoscalar Factors and Representation Matrices", shows how relative decay rates in, say, $\mathbf{10} \to \mathbf{8} \otimes \mathbf{8}$ decays may be calculated.

The addition of the c quark to the light quarks extends the flavor symmetry to SU(4). However, due to the large mass of the c quark, this symmetry is much more strongly broken than the SU(3) of the three light quarks. Figures 14.4(a) and 14.4(b) show the SU(4) baryon multiplets that have as their bottom levels an SU(3) octet, such as the octet that includes the nucleon, or an SU(3) decuplet, such as the decuplet that includes the $\Delta(1232)$. All particles in a given SU(4) multiplet have the same spin and parity. The charmed baryons are discussed in more detail in the "Note on Charmed Baryons" in the Particle Listings. The addition of a b quark extends the flavor symmetry to SU(5), the existence of baryons with t-quarks is very unlikely due to the short lifetime of the top.

For the "ordinary" baryons (no c or b quark), flavor and spin may be combined in an approximate flavor-spin SU(6) in which the six basic states are $d\uparrow$, $d\downarrow$, $\cdots$, $s\downarrow$ ($\uparrow$, $\downarrow$ = spin up, down). Then the baryons belong to the multiplets on the right side of

$$\mathbf{6} \otimes \mathbf{6} \otimes \mathbf{6} = \mathbf{56}_S \oplus \mathbf{70}_M \oplus \mathbf{70}_M \oplus \mathbf{20}_A \ . \tag{14.24}$$

These SU(6) multiplets decompose into flavor SU(3) multiplets as follows:

$$\mathbf{56} = {}^4\mathbf{10} \oplus {}^2\mathbf{8} \tag{14.25a}$$

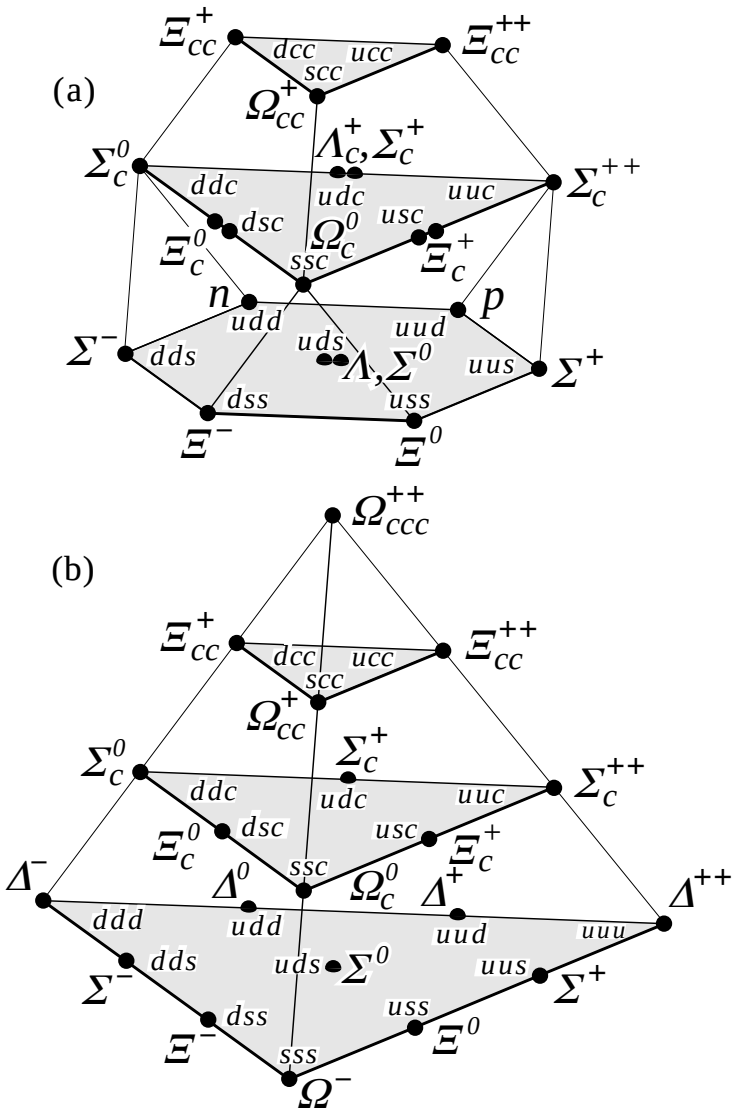


Figure 14.4: SU(4) multiplets of baryons made of u, d, s, and c quarks. (a) The 20-plet with an SU(3) octet. (b) The 20-plet with an SU(3) decuplet.

$$\mathbf{70} = {}^2\mathbf{10} \oplus {}^4\mathbf{8} \oplus {}^2\mathbf{8} \oplus {}^2\mathbf{1} \tag{14.25b}$$

$$\mathbf{20} = {}^2\mathbf{8} \oplus {}^4\mathbf{1} \ , \tag{14.25c}$$

where the superscript $(2S+1)$ gives the net spin S of the quarks for each particle in the SU(3) multiplet. The $J^P = 1/2^+$ octet containing the nucleon and the $J^P = 3/2^+$ decuplet containing the $\Delta(1232)$ together make up the "ground-state" 56-plet in which the orbital angular momenta between the quark pairs are zero (so that the spatial part of the state function is trivially symmetric). The **70** and **20** require some excitation of the spatial part of the state function in order to make the overall state function symmetric. States with nonzero orbital angular momenta are classified in SU(6)⊗O(3) supermultiplets.

It is useful to classify the baryons into bands that have the same number N of quanta of excitation. Each band consists of a number of supermultiplets, specified by (D, L_N^P), where

D is the dimensionality of the SU(6) representation, L is the total quark orbital angular momentum, and P is the total parity. Supermultiplets contained in bands up to N = 12 are given in Ref. 22. The N = 0 band, which contains the nucleon and $\Delta(1232)$, consists only of the $(56,0^+_0)$ supermultiplet. The N = 1 band consists only of the $(70,1^-_1)$ multiplet and contains the negative-parity baryons with masses below about 1.9 GeV. The N = 2 band contains five supermultiplets: $(56,0^+_2)$, $(70,0^+_2)$, $(56,2^+_2)$, $(70,2^+_2)$, and $(20,1^+_2)$.

Table 14.5: N and Δ states in the N=0,1,2 harmonic oscillator bands. L^π denotes angular momentum and parity, S the three-quark spin and 'sym'=A,S,M the symmetry of the spatial wave function.

N	sym	L^π	S	$N(I=1/2)$				$\Delta(I=3/2)$			
2	A	1^+	1/2	$1/2^+$	$3/2^+$						
2	M	2^+	3/2	$1/2^+$	$3/2^+$	$5/2^+$	$7/2^+$				
2	M	2^+	1/2		$3/2^+$	$5/2^+$		$3/2^+$	$5/2^+$		
2	M	0^+	3/2		$3/2^+$						
2	M	0^+	1/2	$1/2^+$			$1/2^+$				
2	S	2^+	3/2				$1/2^+$	$3/2^+$	$5/2^+$	$7/2^+$	
2	S	2^+	1/2		$3/2^+$	$5/2^+$					
2	S	0^+	3/2					$3/2^+$			
2	S	0^+	1/2	$1/2^+$							
1	M	1^-	3/2	$1/2^-$	$3/2^-$	$5/2^-$					
1	M	1^-	1/2	$1/2^-$	$3/2^-$		$1/2^-$	$3/2^-$			
0	S	0^+	3/2					$3/2^+$			
0	S	0^+	1/2	$1/2^+$							

The wave functions of the non-strange baryons in the harmonic oscillator basis are often labeled by $|X^{2S+1}L_\pi J^P\rangle$, where S, L, J, P are as above, $X = N$ or Δ, and $\pi = S, M$ or A denotes the symmetry of the spatial wave function. The possible states for the bands with N=0,1,2 are given in Table 14.5.

In Table 14.6, quark-model assignments are given for many of the established baryons whose SU(6)⊗O(3) compositions are relatively unmixed. One must, however, keep in mind that apart from the mixing of the Λ singlet and octet states also states with same J^π but different L, S combinations can mix. In the quark model with one-gluon exchange motivated interactions the size of the mixing is determined by the relative strength of the tensor term with respect to the contact term (see below). The mixing is more important for the decay patterns of the states than for their positions. An example are the lowest lying $(70, 1^-_1)$ states with J^π=$1/2^-$ and $3/2^-$. The physical states are:

$$|\,S_{11}(1535)\rangle = \cos(\Theta_S)|N^2P_M1/2^-\rangle - \sin(\Theta_S)|N^4P_M1/2^-\rangle \tag{14.26}$$

$$|\,D_{13}(1520)\rangle = \cos(\Theta_D)|N^2P_M3/2^-\rangle - \sin(\Theta)_D|N^4P_M3/2^-\rangle \tag{14.27}$$

and the orthogonal combinations for $S_{11}(1650)$ and $D_{13}(1700)$. The mixing is large for the J^π=$1/2^-$ states ($\Theta_S \approx$ -32^o) but small for the J^π=$3/2^-$ states ($\Theta_D \approx +6^o$) [23], [26].

Table 14.6: Quark-model assignments for some of the known baryons in terms of a flavor-spin SU(6) basis. Only the dominant representation is listed. Assignments for several states, especially for the $\Lambda(1810)$, $\Lambda(2350)$, $\Xi(1820)$, and $\Xi(2030)$, are merely educated guesses. For assignments of the charmed baryons, see the "Note on Charmed Baryons" in the Particle Listings.

J^P	(D, L_N^P)	S	Octet members				Singlets
$1/2^+$	$(56,0_0^+)$	1/2	$N(939)$	$\Lambda(1116)$	$\Sigma(1193)$	$\Xi(1318)$	
$1/2^+$	$(56,0_2^+)$	1/2	$N(1440)$	$\Lambda(1600)$	$\Sigma(1660)$	$\Xi(?)$	
$1/2^-$	$(70,1_1^-)$	1/2	$N(1535)$	$\Lambda(1670)$	$\Sigma(1620)$	$\Xi(?)$	$\Lambda(1405)$
$3/2^-$	$(70,1_1^-)$	1/2	$N(1520)$	$\Lambda(1690)$	$\Sigma(1670)$	$\Xi(1820)$	$\Lambda(1520)$
$1/2^-$	$(70,1_1^-)$	3/2	$N(1650)$	$\Lambda(1800)$	$\Sigma(1750)$	$\Xi(?)$	
$3/2^-$	$(70,1_1^-)$	3/2	$N(1700)$	$\Lambda(?)$	$\Sigma(?)$	$\Xi(?)$	
$5/2^-$	$(70,1_1^-)$	3/2	$N(1675)$	$\Lambda(1830)$	$\Sigma(1775)$	$\Xi(?)$	
$1/2^+$	$(70,0_2^+)$	1/2	$N(1710)$	$\Lambda(1810)$	$\Sigma(1880)$	$\Xi(?)$	$\Lambda(?)$
$3/2^+$	$(56,2_2^+)$	1/2	$N(1720)$	$\Lambda(1890)$	$\Sigma(?)$	$\Xi(?)$	
$5/2^+$	$(56,2_2^+)$	1/2	$N(1680)$	$\Lambda(1820)$	$\Sigma(1915)$	$\Xi(2030)$	
$7/2^-$	$(70,3_3^-)$	1/2	$N(2190)$	$\Lambda(?)$	$\Sigma(?)$	$\Xi(?)$	$\Lambda(2100)$
$9/2^-$	$(70,3_3^-)$	3/2	$N(2250)$	$\Lambda(?)$	$\Sigma(?)$	$\Xi(?)$	
$9/2^+$	$(56,4_4^+)$	1/2	$N(2220)$	$\Lambda(2350)$	$\Sigma(?)$	$\Xi(?)$	
			Decuplet members				
$3/2^+$	$(56,0_0^+)$	3/2	$\Delta(1232)$	$\Sigma(1385)$	$\Xi(1530)$	$\Omega(1672)$	
$3/2^+$	$(56,0_2^+)$	3/2	$\Delta(1600)$	$\Sigma(?)$	$\Xi(?)$	$\Omega(?)$	
$1/2^-$	$(70,1_1^-)$	1/2	$\Delta(1620)$	$\Sigma(?)$	$\Xi(?)$	$\Omega(?)$	
$3/2^-$	$(70,1_1^-)$	1/2	$\Delta(1700)$	$\Sigma(?)$	$\Xi(?)$	$\Omega(?)$	
$5/2^+$	$(56,2_2^+)$	3/2	$\Delta(1905)$	$\Sigma(?)$	$\Xi(?)$	$\Omega(?)$	
$7/2^+$	$(56,2_2^+)$	3/2	$\Delta(1950)$	$\Sigma(2030)$	$\Xi(?)$	$\Omega(?)$	
$11/2^+$	$(56,4_4^+)$	3/2	$\Delta(2420)$	$\Sigma(?)$	$\Xi(?)$	$\Omega(?)$	

All baryons of the ground state multiplets are known. Many of their properties (masses, magnetic moments etc.) are in good agreement with the most basic versions of the quark model, including harmonic (or linear) confinement and a spin-spin interaction, which is responsible for the octet - decuplet mass shifts.

The situation for the excited states is much less clear. There are two main problems which are illustrated in Fig. 14.5, where the experimentally observed excitation spectrum of the nucleon (N and Δ resonances) is compared to the results of a typical quark model calculation [24]. Many more states are predicted than observed but on the other hand states with certain quantum numbers appear in the spectrum at excitation energies much lower than predicted. Up to an excitation energy of 2.4 GeV about 45 N states are predicted, but only 12 are established (four

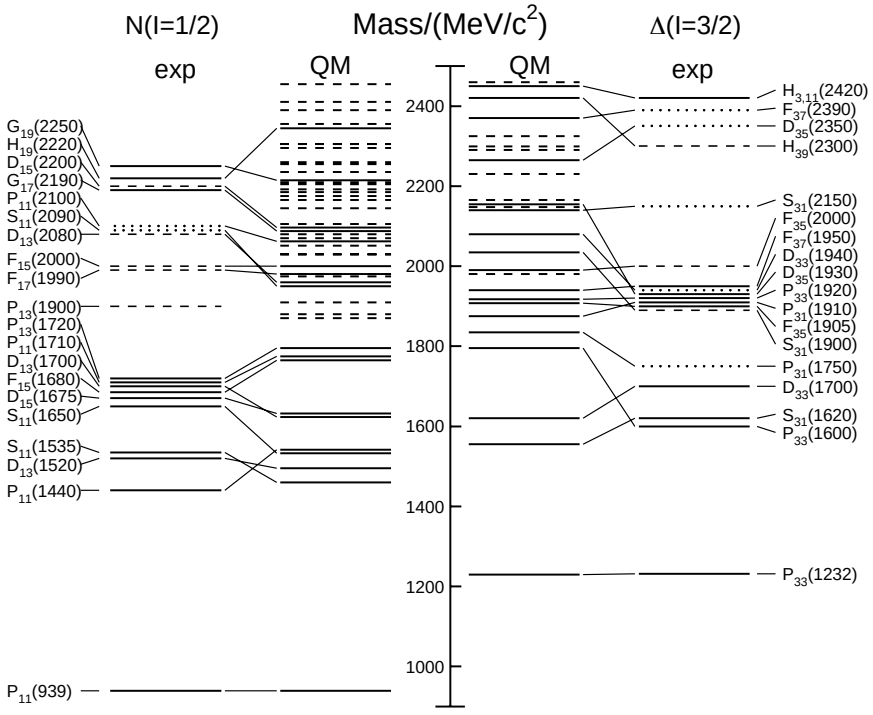


Figure 14.5: Excitation spectrum of the nucleon. Compared are the positions of the excited states identified in experiment, to those predicted by a modern quark model calculation. Left hand side: isospin $I = 1/2$ N-states, right hand side: isospin $I = 3/2$ Δ-states. Experimental: (columns labeled 'exp'), three and four star states are indicated by full lines (two-star dashed lines, one-star dotted lines). At the very left and right of the figure the spectroscopic notation of these states is given. Quark model [24]: (columns labeled 'QM'), all states for the N=1,2 bands, low lying states for the N=3,4,5 bands. Full lines: at least tentative assignment to observed states, dashed lines: so far no observed counterparts. Many of the assignments between predicted and observed states are highly tentative.

or three star; see Note on N and Δ resonances for the rating of the status of resonances) and 7 are tentative (two or one star). Even for the N=1,2 bands up to now only half of the predicted states have been observed. This has been known for a long time as the 'missing resonance' problem [23]. On the other hand, the lowest states from the N=2 band, the $P_{11}(1440)$ and the $P_{33}(1600)$, appear lower than the negative parity states from the N=1 band and much lower than predicted by most models. Also negative parity Δ states from the N=3 band ($S_{31}(1900)$, $D_{33}(1940)$, $D_{35}(1930)$) are too low in energy. Part of the problem could be experimental. Among the negative parity Δ states only the D_{35} has three stars and the uncertainty in the position of the $P_{33}(1600)$ is large (1550 - 1700 MeV). For the missing resonance problem selection rules could play a role [23]. The states are broad and overlapping and most studies of baryon resonances have been done with pion induced reactions, so that there is bias in the data base against resonances, which couple only weakly to the $N\pi$ channel. Quark model predictions for the couplings to other hadronic channels and to photons are given in Ref. 24. A large experimental effort is ongoing at several electron accelerators to study the baryon resonance spectrum with real and virtual photon induced meson production reactions. This includes the search for as yet unobserved states as well as detailed studies of the properties of the low lying states (decay patterns, electromagnetic couplings, magnetic moments etc.) (see Ref. 25 for recent reviews).

In quark models the number of excited states is determined by the effective degrees of freedom while their ordering and decay properties are related to the residual quark - quark interaction. A recent overview of quark models for baryons is given in Ref. 26. The effective degrees of freedom in the standard nonrelativistic quark model are three equivalent valence quarks with one-gluon exchange motivated, flavor independent color magnetic interactions. A different class of models uses interactions which give rise to a quark - diquark clustering of the baryons (for a review see Ref. 27). If there is a tightly bound diquark, only two degrees of freedom are available at low energies and thus *fewer* states are predicted. Furthermore, selection rules in the decay pattern may arise from the quantum numbers of the diquark. *More* states are predicted by collective models of the baryon like the algebraic approach in Ref. 28. In this approach, the quantum numbers of the valence quarks are distributed over a Y-shaped string-like configuration and additional states arise e.g. from vibrations of the strings. *More* states are also predicted in the framework of flux-tube models (see [29]) , which are motivated by lattice QCD. In addition to the quark degrees of freedom, flux-tubes responsible for the confinement of the quarks are considered as degrees of freedom. These models include hybrid baryons containing explicit excitations of the gluon fields. However, since all half integral J^π quantum numbers are possible for ordinary baryons, such 'exotics' will be very hard to identify and probably always mix with ordinary states. So far, the experimentally observed number of states is still far lower even than predicted by the quark - diquark models.

14.5. Dynamics

Many specific quark models exist, but most contain a similar basic set of dynamical ingredients. These include:

i) A confining interaction, which is generally spin-independent (e.g. harmonic oscillator or linear confinement).

ii) Different types of spin-dependent interactions e.g.:

a) color magnetic flavor independent interaction modeled after the effects of gluon exchange in QCD (see e.g. Ref. 31). For example, in the S-wave states, there is a spin-spin hyperfine interaction of the form

$$H_{HF} = -\alpha_S M \sum_{i>j} (\overrightarrow{\sigma}\lambda_a)_i (\overrightarrow{\sigma}\lambda_a)_j \ , \tag{14.28}$$

where M is a constant with units of energy, λ_a $(a = 1, \cdots, 8,)$ is the set of SU(3) unitary spin matrices, defined in Sec. 36, on "SU(3) Isoscalar Factors and Representation Matrices," and the sum runs over constituent quarks or antiquarks. Spin-orbit interactions, although allowed, seem to be small in general, but a tensor term is responsible for the mixing of states with the same J^π but different L, S combinations.

b) flavor dependent short range quark forces from instanton effects (see e.g. Ref. 32). This interaction acts only on scalar, isoscalar pairs of quarks in a relative s-wave state:

$$\langle q^2; S, L, T|W|q^2; S, L, T\rangle = -4g\delta_{S,0}\delta_{L,0}\delta_{I,0}\mathcal{W} \tag{14.29}$$

where $\mathcal{W}$ is the radial matrix element of the contact interaction.

c) flavor dependent spin-spin forces arising from one-boson exchange. The interaction term is of the form:

$$H_{HF} \propto \sum_{i<j} V(\overrightarrow{r}_{ij})\lambda_i^F \cdot \lambda_j^F \overrightarrow{\sigma}_i \cdot \overrightarrow{\sigma}_j \tag{14.30}$$

where the λ_i^F are in flavor space (see e.g. Ref. 33).

iii) A strange quark mass somewhat larger than the up and down quark masses, in order to split the SU(3) multiplets.

iv) In the case of isoscalar mesons, an interaction for mixing $q\overline{q}$ configurations of different flavors (*e.g.*, $u\overline{u} \leftrightarrow d\overline{d} \leftrightarrow s\overline{s}$), in a manner which is generally chosen to be flavor independent.

These four ingredients provide the basic mechanisms that determine the hadron spectrum in the standard quark model. However, a radically different view of the meson and baryon excitation spectrum is being derived in the framework of chiral coupled channel dynamics [30]. Here, the conjecture is that in the u, d, s sector only the ground state baryon-octet $1/2^+$, baryon-decuplet $3/2^+$ and the Goldstone boson 0^- and vector meson 1^- states are genuine qqq or $q\bar{q}$ states. It is then attempted to generate all excited states by coupled channel dynamics.

14.6. Lattice Calculations of Hadronic Spectroscopy

Lattice calculations predict the spectrum of bound states in QCD from first principles, beginning with the Lagrangian of full QCD or of various approximations to it. This is typically done using the Euclidean path integral formulation of quantum field theory, where the analog of a partition function for a field theory containing some generic fields $\phi(x)$, with action $S(\phi)$, is

$$Z = \int [d\phi] \exp(-S(\phi)). \tag{14.31}$$

The expectation value of any observable O is

$$\langle O \rangle = \frac{1}{Z} \int [d\phi] O(\phi) \exp(-S(\phi)). \tag{14.32}$$

The theory is regulated by introducing a space-time lattice, with lattice spacing a. This converts the functional integral Eq. (14.31) into an ordinary integral (of very large dimensionality). The integral is replaced by a Monte Carlo sampling over an ensemble of configurations of field variables, using an algorithm which insures that a field configuration is present in the ensemble with a probability proportional to $\exp(-S(\phi_j))$. Then ensemble averages become sample averages,

$$\langle O \rangle = \frac{1}{N} \sum_{j=1}^{N} O(\phi_j). \tag{14.33}$$

This is all quite similar to the kind of Monte Carlo simulation done by experiments, except that the ensembles of field configurations are created sequentially, as a so-called "Markov chain."

In QCD, the field variables correspond to gauge fields and quark fields. In a lattice calculation, the lattice spacing (which serves as an ultraviolet cutoff) and the (current) quark masses are inputs; hadron masses and other observables are predicted as a function of those masses. The lattice spacing is unphysical, and it is necessary to extrapolate to the limit of zero lattice spacing. Lattice predictions are for dimensionless ratios of dimensionful parameters (like mass ratios) and predictions of dimensionful quantities require using one experimental input to set the scale. Interpolation or extrapolation of lattice results in the light quark masses involves formulas of chiral perturbation theory.

For conventional hadronic states, lattice calculations use the quark model to construct operators which are taken as interpolating fields. This does not mean that the hadronic states have minimal

quark content: the operators create multi-quark states with particular quantum numbers, but they are connected by quark propagators which include all effects of relativity and could include the effects of virtual quark-antiquark pairs in the vacuum.

Constituent gluons do not appear naturally in lattice calculations; instead, gauge fields appear as link variables, which allow color to be parallel transported across the lattice in a gauge covariant way. Calculations of glueballs on the lattice use interpolating fields of the form $O_j \sim \exp i \oint \vec{A} \cdot \vec{dl}$ integrated about some path. The fields look like closed tubes of chromoelectric and chromomagnetic flux. Calculations of exotics are done with interpolating fields involving quark and antiquark creation operators joined by flux tubes.

Calculations with heavy quarks typically use Non-Relativistic QCD (NRQCD) or Heavy Quark Effective Theory (HQET), systematic expansions of the QCD Lagrangian in powers of the heavy quark velocity or the inverse heavy quark mass. Terms in the Lagrangian have obvious quark model analogs but are derived directly from QCD. The heavy quark potential is a derived quantity, measured in simulations.

Lattice calculations are as specialized as the experiments which produce the data in this book and it is not easy to give a blanket answer to the question "how well can lattice calculations predict" any specific quantity. However, let us try:

The cleanest lattice predictions come from measurements of processes in which there is only one particle in the simulation volume. These quantities include masses of hadrons, simple decay constants, like pseudoscalar meson decay constants, and semileptonic form factors (such as the ones appropriate to $B \to Dl\nu$, $Kl\nu$, $\pi l\nu$). The cleanest predictions for masses are for states which have narrow decay widths and are far below any thresholds to open channels, since the effects of final state interactions are not yet under complete control on the lattice. "Difficult" states for the quark model (such as exotics) are also difficult for the lattice because of the lack of simple operators which couple well to them. Technical issues presently prevent lattice practitioners from directly computing matrix elements for weak decays with more than one strongly interacting particle in the final state.

Good-quality modern lattice calculations will present multi-part error budgets with their predictions. Users are advised to read them carefully! A small part of the uncertainty is statistical, from sample size. Typically, the quoted statistical uncertainty includes uncertainty from a fit: it is rare that a simulation measures one global quantity which is the desired observable. Simulations with light dynamical quarks are typically done at mass values heavier than the experimental ones, and it is necessary to extrapolate in the quark mass. They are always done at nonzero lattice spacing, and so it will be necessary to extrapolate to zero lattice spacing. Some theoretical input is needed to do this. Much of the uncertainty in these extrapolations is systematic, from the choice of fitting function. Other systematics include the number of dynamical flavors of quarks actually simulated, and technical issues with how these dynamical quarks are included. The particular choice of a fiducial mass (to normalize other predictions) is not standardized; there are many possible choices, each with its own set of strengths and weaknesses, and determining it usually requires a second lattice simulation from that used to calculate the quantity under consideration.

Of course, there is much more to lattice calculations besides spectroscopy; please refer to the mini-review on Quark Masses in the Quarks section of the Listings for more lattice-based phenomenology.

We conclude with a few "representative" pictures of spectroscopy from recent state-of-the-art simulations. They illustrate (better than any discussion) the size of lattice uncertainties.

A systematic of major historical interest is the "quenched approximation," in which virtual quark-antiquark pairs are simply left out of the simulation. This was done because the addition of

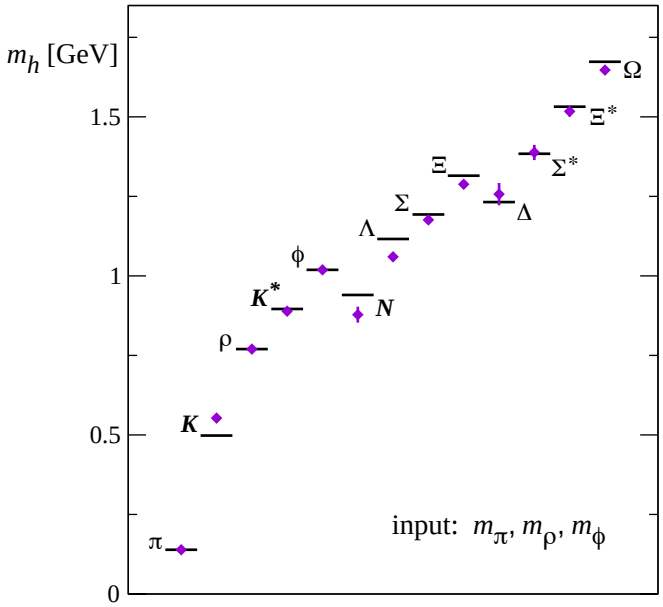


Figure 14.6: Comparison of quenched results with experiment, from Ref. [34] as presented by Ref. [35].

these virtual pairs presented an expensive computational problem. Recent advances in algorithms and computer hardware have rendered it obsolete. A recent quenched simulation [34] of the light hadron spectrum is shown in Fig. 14.6. Note the qualitative quark model features of the lattice result: the equal spacing of the decuplet and the vector meson multiplet. The lattice error bars are small enough to reveal systematic errors in the quenched approximation: note especially, that the kaon and ϕ masses cannot be simultaneously correctly determined, and the nucleon/rho meson mass ratio is too low. No generally-accepted methodology has ever allowed one to correct for quenching effects, short of redoing all calculations with dynamical quarks.

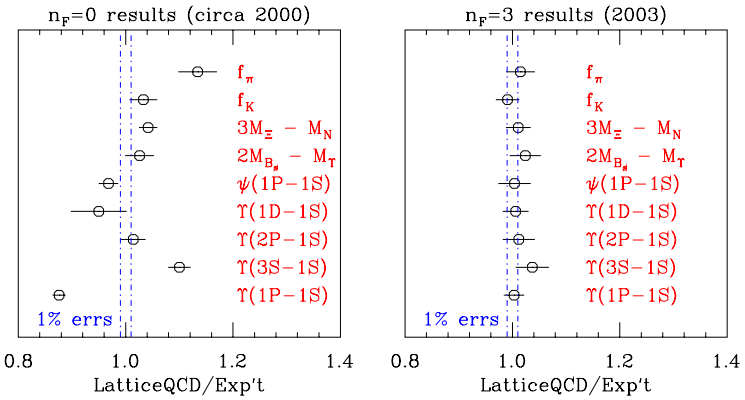


Figure 14.7: Comparison of quenched results with results from simulations with 2+1 flavors of staggered fermions, from Ref. [36].

Fig. 14.7, from Ref. [36], shows a comparison of quenched spectroscopy with results from a simulation with a strange quark at roughly its physical value and two flavors of light dynamical quarks with masses at roughly one fifth of the strange quark's mass. The authors focus on

quantities for which lattice predictions should work well, as we described above. The left panel shows these authors' compilation of quenched results; the right panel, results with dynamical fermions, after a chiral extrapolation.

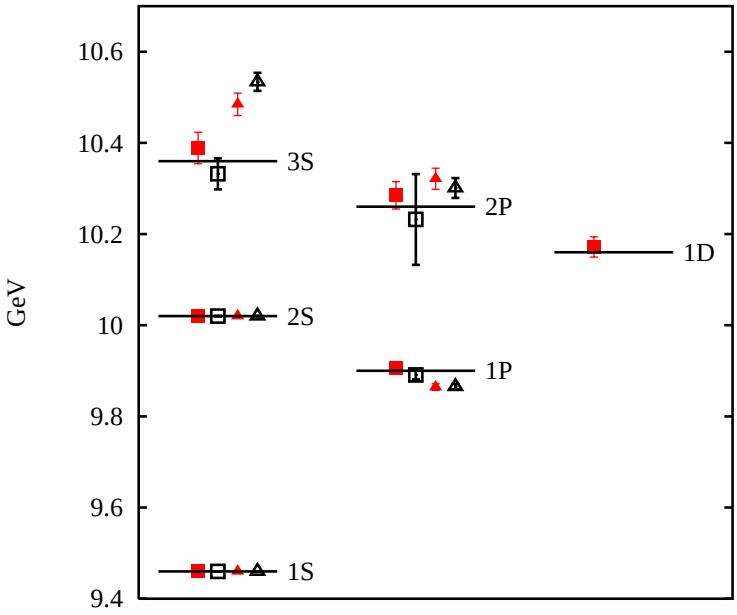


Figure 14.8: The Υ spectrum of radial and orbital levels from Ref. [37]. Closed and open symbols are from coarse and fine lattices respectively. Squares and triangles denote unquenched and quenched results respectively. Lines represent experiment. See full-color version on color pages at end of book.

Fig. 14.8 shows Upsilon ($b\bar{b}$) spectroscopy from Ref. [37]. The calculation uses a discretization of nonrelativistic QCD for its heavy quarks, and includes three flavors of light dynamical staggered fermions. Quenched data are also shown for comparison.

References:

1. J. Schwinger, Phys. Rev. Lett. **12**, 237 (1964).
2. A. Bramon, R. Escribano, and M.D. Scadron, Phys. Lett. **B403**, 339 (1997).
3. A. Aloisio *et al.*, Phys. Lett. **B541**, 45 (2002).
4. C. Amsler *et al.*, Phys. Lett. **B294**, 451 (1992).
5. C. Amsler, Rev. Mod. Phys. **70**, 1293 (1998).
6. T. Feldmann, Int. J. Mod. Phys. **A915**, 159 (2000).
7. C. Amsler and F.E. Close, Phys. Rev. **D53**, 295 (1996).
8. R.L. Jaffe, Phys. Rev. **D 15** 267, 281 (1977).
9. Y. Chen *et al.*, Phys. Rev. **D73**, 014516 (2006).
10. C. Morningstar and M. Peardon, Phys. Rev. **D60**, 034509 (1999).
11. W. J. Lee and D. Weingarten, Phys. Rev. **D61**, 014015 (2000).
12. G. S. Bali, et. al. Phys. Lett. **B309**, 378 (1993).
13. C. Michael, AIP Conf. Proc. **432**, 657 (1998).
14. F.E. Close and A. Kirk, Eur. Phys. J. **C21**, 531 (2001).
15. C. Amsler and N.A. Törnqvist, Phys. Rev. **389**, 61 (2004).
16. N. Isgur and J. Paton, Phys. Rev. **D31**, 2910 (1985).

17. P. Lacock *et al.*, Phys. Lett. **B401**, 308 (1997);
C. Bernard *et al.*, Phys. Rev. **D56**, 7039 (1997).
18. M. Chanowitz and S. Sharpe, Nucl. Phys. **B222**, 211 (1983).
19. T. Barnes, F.E. Close, F. de Viron Nucl. Phys. **B224**, 241 (1983).
20. F.E. Close, in *Quarks and Nuclear Forces* (Springer-Verlag, 1982), p. 56.
21. Particle Data Group, Phys. Lett. **111B** (1982).
22. R.H. Dalitz and L.J. Reinders, in *Hadron Structure as Known from Electromagnetic and Strong Interactions, Proceedings of the Hadron '77 Conference* (Veda, 1979), p. 11.
23. N. Isgur and G. Karl, Phys. Rev. **D18**, 4187 (1978); ibid. **D19**, 2653 (1979); ibid. **D20**, 1191 (1979);
K.-T. Chao, N. Isgur, and G. Karl, Phys. Rev. **D23**, 155 (1981).
24. S. Capstick and W. Roberts, Phys. Rev. **D49**, 4570 (1994); ibid. **D57**, 4301 (1998); ibid. **D58**, 074011 (1998);
S. Capstick, Phys. Rev. **D46**, 2864 (1992).
25. B. Krusche and S. Schadmand, Prog. Part. Nucl. Phys. **51**, 399 (2003);
V.D. Burkert and T.-S.H. Lee, Int. J. Mod. Phys. **E13**, 1035 (2004).
26. S. Capstick and W. Roberts, Prog. Part. Nucl. Phys. **45**, 241 (2000);
see also A.J.G. Hey and R.L. Kelly, Phys. Reports **96**, 71 (1983).
27. M. Anselmino et al., Rev. Mod. Phys. **65**, 1199 (1993).
28. R. Bijker, F. Iachello, and A. Leviatan, Ann. of. Phys. **236** 69 (1994).
29. N. Isgur and J. Paton, Phys. Rev. **D31**, 2910 (1985);
S. Capstick and P.R. Page, Phys. Rev. **C66**, 065204 (2002).
30. M.F.M Lutz and E.E. Kolomeitsev, Nucl. Phys. **A700**, 193 (2002);
M.F.M Lutz and E.E. Kolomeitsev, Nucl. Phys. **A730**, 392 (2004);
E.E. Kolomeitsev and M.F.M Lutz, Phys. Lett. **B585**, 243 (2004).
31. A. De Rujula, H. Georgi, S.L. Glashow, Phys. Rev. **D12**, 147 (1975).
32. U. Löring, K. Kretschmar, B.C. Metsch, H.R. Petry, Eur. Phys. J. **A10** 309 (2001);
U. Löring, B.C. Metsch, H.R. Petry, Eur. Phys. J. **A10** 395 (2001); ibid **A10** 447 (2001).
33. L.Y. Glozman and D.O. Riska, Phys. Rept. **268** 263 (1996);
L.Y. Glozman et al., Phys. Rev. **D58**, 094030 (1998).
34. S. Aoki *et al.* [CP-PACS Collaboration], Phys. Rev. D **67**, 034503 (2003) [arXiv:hep-lat/0206009].
35. M. Luscher, Annales Henri Poincare **4**, S197 (2003) [arXiv:hep-ph/0211220]..
36. C. T. H. Davies *et al.* [HPQCD Collaboration], Phys. Rev. Lett. **92**, 022001 (2004) [arXiv:hep-lat/0304004].
37. A. Gray, I. Allison, C. T. H. Davies, E. Gulez, G. P. Lepage, J. Shigemitsu and M. Wingate, Phys. Rev. D **72**, 094507 (2005) arXiv:hep-lat/0507013.

15. GRAND UNIFIED THEORIES

Revised October 2005 by S. Raby (Ohio State University).

15.1. Grand Unification

15.1.1. *Standard Model: An Introduction* :

In spite of all the successes of the Standard Model [SM], it is unlikely to be the final theory. It leaves many unanswered questions. Why the local gauge interactions $\mathrm{SU}(3)_C \times \mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$, and why 3 families of quarks and leptons? Moreover, why does one family consist of the states $[Q, u^c, d^c; L, e^c]$ transforming as $[(3, 2, 1/3)$, $(\bar{3}, 1, -4/3)$, $(\bar{3}, 1, 2/3)$; $(1, 2, -1)$, $(1, 1, 2)]$, where $Q = (u, d)$, and $L = (\nu, e)$ are $\mathrm{SU}(2)_L$ doublets, and u^c, d^c, e^c are charge conjugate $\mathrm{SU}(2)_L$ singlet fields with the $\mathrm{U}(1)_Y$ quantum numbers given? [We use the convention that electric charge $Q_{EM} = T_{3L} + Y/2$ and all fields are left-handed.] Note the SM gauge interactions of quarks and leptons are completely fixed by their gauge charges. Thus, if we understood the origin of this charge quantization, we would also understand why there are no fractionally charged hadrons. Finally, what is the origin of quark and lepton masses, or the apparent hierarchy of family masses and quark mixing angles? Perhaps if we understood this, we would also know the origin of CP violation, the solution to the strong CP problem, the origin of the cosmological matter-antimatter asymmetry, or the nature of dark matter.

The SM has 19 arbitrary parameters; their values are chosen to fit the data. Three arbitrary gauge couplings: g_3, g, g' (where g, g' are the $\mathrm{SU}(2)_L$, $\mathrm{U}(1)_Y$ couplings, respectively) or equivalently, $\alpha_s = (g_3^2/4\pi)$, $\alpha_{EM} = (e^2/4\pi)$ $(e = g \sin\theta_W)$, and $\sin^2\theta_W = (g')^2/(g^2 + (g')^2)$. In addition, there are 13 parameters associated with the 9 charged fermion masses and the four mixing angles in the CKM matrix. The remaining 3 parameters are v, λ [the Higgs VEV (vacuum expectation value) and quartic coupling] (or equivalently, M_Z, m_h^0), and the QCD θ parameter. In addition, data from neutrino oscillation experiments provide convincing evidence for neutrino masses. With 3 light Majorana neutrinos, there are at least 9 additional parameters in the neutrino sector; 3 masses, 3 mixing angles and 3 phases. In summary, the SM has too many arbitrary parameters, and leaves open too many unresolved questions to be considered complete. These are the problems which grand unified theories hope to address.

15.1.2. *Charge Quantization* :

In the Standard Model, quarks and leptons are on an equal footing; both fundamental particles without substructure. It is now clear that they may be two faces of the same coin; unified, for example, by extending QCD (or $\mathrm{SU}(3)_C$) to include leptons as the fourth color, $\mathrm{SU}(4)_C$ [1]. The complete Pati-Salam gauge group is $\mathrm{SU}(4)_C \times \mathrm{SU}(2)_L \times \mathrm{SU}(2)_R$, with the states of one family $[(Q, L), (Q^c, L^c)]$ transforming as $[(4, 2, 1)$, $(\bar{4}, 1, \bar{2})]$, where $Q^c = (d^c, u^c)$, $L^c = (e^c, \nu^c)$ are doublets under $\mathrm{SU}(2)_R$. Electric charge is now given by the relation $Q_{EM} = T_{3L} + T_{3R} + 1/2(B-L)$, and $\mathrm{SU}(4)_C$ contains the subgroup $\mathrm{SU}(3)_C \times (B-L)$ where B (L) is baryon (lepton) number. Note ν^c has no SM quantum numbers and is thus completely "sterile." It is introduced to complete the $\mathrm{SU}(2)_R$ lepton doublet. This additional state is desirable when considering neutrino masses.

Although quarks and leptons are unified with the states of one family forming two irreducible representations of the gauge group, there are still 3 independent gauge couplings (two if one also imposes parity, *i.e.*, $L \leftrightarrow R$ symmetry). As a result, the three low-energy gauge couplings are still independent arbitrary parameters. This difficulty is resolved by embedding the SM gauge group into the simple unified gauge group, Georgi-Glashow SU(5), with one universal gauge coupling α_G defined at the grand unification scale M_G [2]. Quarks and leptons still sit in two irreducible representations, as before, with a $\mathbf{10} = [Q, u^c, e^c]$ and $\bar{\mathbf{5}} = [d^c, L]$. Nevertheless, the three low energy gauge couplings are now determined in terms of two independent parameters : α_G and M_G. Hence, there is one prediction.

In order to break the electroweak symmetry at the weak scale and give mass to quarks and leptons, Higgs doublets are needed which can sit in either a $\mathbf{5_H}$ or $\bar{\mathbf{5}}_\mathbf{H}$. The additional 3 states are color triplet Higgs scalars. The couplings of these color triplets violate baryon and lepton number, and nucleons decay via the exchange of a single color triplet Higgs scalar. Hence, in order not to violently disagree with the non-observation of nucleon decay, their mass must be greater than $\sim 10^{10-11}$ GeV. Moreover, in supersymmetric GUTs, in order to cancel anomalies, as well as give mass to both up and down quarks, both Higgs multiplets $\mathbf{5_H}$, $\bar{\mathbf{5}}_\mathbf{H}$ are required. As we shall discuss later, nucleon decay now constrains the color triplet Higgs states in a SUSY GUT to have mass significantly greater than M_G.

Complete unification is possible with the symmetry group SO(10), with one universal gauge coupling α_G, and one family of quarks and leptons sitting in the 16-dimensional-spinor representation $\mathbf{16} = [\mathbf{10} + \bar{\mathbf{5}} + \mathbf{1}]$ [3]. The SU(5) singlet $\mathbf{1}$ is identified with ν^c. In Table 15.1 we present the states of one family of quarks and leptons, as they appear in the **16**. It is an amazing and perhaps even profound fact that all the states of a single family of quarks and leptons can be represented digitally as a set of 5 zeros and/or ones or equivalently as the tensor product of 5 "spin" 1/2 states with $\pm = |\pm\frac{1}{2}>$ and with the condition that we have an even number of $|->$ spins. The first three "spins" correspond to $SU(3)_C$ color quantum numbers, while the last two are $SU(2)_L$ weak quantum numbers. In fact, an $SU(3)_C$ rotation just raises one color index and lowers another, thereby changing colors $\{r,\ b,\ y\}$. Similarly an $SU(2)_L$ rotation raises one weak index and lowers another, thereby flipping the weak isospin from up to down or vice versa. In this representation, weak hypercharge Y is given by the simple relation $Y = 2/3(\sum \text{ color spins}) - (\sum \text{ weak spins})$. SU(5) rotations [not in the Standard Model] then raise (or lower) a color index, while at the same time lowering (or raising) a weak index. It is easy to see that such rotations can mix the states $\{Q,\ u^c,\ e^c\}$ and $\{d^c, L\}$ among themselves, and ν^c is a singlet. The new SO(10) rotations [not in SU(5)] are then given by either raising or lowering any two spins. For example, by lowering the two weak indices, ν^c rotates into e^c, etc.

SO(10) has two inequivalent maximal breaking patterns: $SO(10) \to SU(5) \times U(1)_X$ and $SO(10) \to SU(4)_C \times SU(2)_L \times SU(2)_R$. In the first case, we obtain Georgi-Glashow SU(5) if Q_{EM} is given in terms of SU(5) generators alone, or so-called flipped SU(5) [4] if Q_{EM} is partly in $U(1)_X$. In the latter case, we have the Pati-Salam symmetry. If

Table 15.1: The quantum numbers of the **16** dimensional representation of SO(10).

State	Y	Color	Weak
$\boldsymbol{\nu}^c$	0	$+\ +\ +$	$+\,+$
$\boldsymbol{e}^c$	2	$+\ +\ +$	$-\,-$
$\boldsymbol{u}_r$	1/3	$-\ +\ +$	$+\,-$
$\boldsymbol{d}_r$	1/3	$-\ +\ +$	$-\,+$
$\boldsymbol{u}_b$	1/3	$+\ -\ +$	$+\,-$
$\boldsymbol{d}_b$	1/3	$+\ -\ +$	$-\,+$
$\boldsymbol{u}_y$	1/3	$+\ +\ -$	$+\,-$
$\boldsymbol{d}_y$	1/3	$+\ +\ -$	$-\,+$
$\boldsymbol{u}_r^c$	$-4/3$	$+\ -\ -$	$+\,+$
$\boldsymbol{u}_b^c$	$-4/3$	$-\ +\ -$	$+\,+$
$\boldsymbol{u}_y^c$	$-4/3$	$-\ -\ +$	$+\,+$
$\boldsymbol{d}_r^c$	2/3	$+\ -\ -$	$-\,-$
$\boldsymbol{d}_b^c$	2/3	$-\ +\ -$	$-\,-$
$\boldsymbol{d}_y^c$	2/3	$-\ -\ +$	$-\,-$
$\boldsymbol{\nu}$	-1	$-\ -\ -$	$+\,-$
$\boldsymbol{e}$	-1	$-\ -\ -$	$-\,+$

SO(10) breaks directly to the SM at M_G, then we retain the prediction for gauge coupling unification. However, more possibilities for breaking (hence more breaking scales and more parameters) are available in SO(10). Nevertheless with one breaking pattern SO(10) $\to$ SU(5) $\to$ SM, where the last breaking scale is M_G, the predictions from gauge coupling unification are preserved. The Higgs multiplets in minimal SO(10) are contained in the fundamental $\mathbf{10_H} = [\mathbf{5_H},\ \bar{\mathbf{5}}_\mathbf{H}]$ representation. Note, only in SO(10) does the gauge symmetry distinguish quark and lepton multiplets from Higgs multiplets.

Finally, larger symmetry groups have been considered. For example, $E(6)$ has a fundamental representation **27**, which under SO(10) transforms as a $[\mathbf{16} + \mathbf{10} + \mathbf{1}]$. The breaking pattern $E(6) \to \mathrm{SU}(3)_C \times \mathrm{SU}(3)_L \times \mathrm{SU}(3)_R$ is also possible. With the additional permutation symmetry $Z(3)$ interchanging the three SU(3)s, we obtain so-called "trinification [5], " with a universal gauge coupling. The latter breaking pattern has been used in phenomenological analyses of the heterotic string [6]. However, in larger symmetry groups, such as $E(6)$, SU(6), *etc.*, there are now many more states which have not been observed and must be removed from the effective low-energy theory. In particular, three families of **27**s in $E(6)$ contain three Higgs type multiplets transforming as **10**s of SO(10). This makes these larger symmetry groups unattractive starting points

for model building.

15.1.3. ***String Theory and Orbifold GUTs*** :

Orbifold compactification of the heterotic string [7–9], and recent field theoretic constructions known as orbifold GUTs [10], contain grand unified symmetries realized in 5 and 6 dimensions. However, upon compactifying all but four of these extra dimensions, only the MSSM is recovered as a symmetry of the effective four dimensional field theory.[1] These theories can retain many of the nice features of four dimensional SUSY GUTs, such as charge quantization, gauge coupling unification and sometimes even Yukawa unification; while at the same time resolving some of the difficulties of 4d GUTs, in particular problems with unwieldy Higgs sectors necessary for spontaneously breaking the GUT symmetry, problems with doublet-triplet Higgs splitting or rapid proton decay. We will comment further on the corrections to the four dimensional GUT picture due to orbifold GUTs in the following sections.

15.1.4. ***Gauge coupling unification*** :

The biggest paradox of grand unification is to understand how it is possible to have a universal gauge coupling g_G in a grand unified theory [GUT], and yet have three unequal gauge couplings at the weak scale with $g_3 > g > g'$. The solution is given in terms of the concept of an effective field theory [EFT] [16]. The GUT symmetry is spontaneously broken at the scale M_G, and all particles not in the SM obtain mass of order M_G. When calculating Green's functions with external energies $E \gg M_G$, we can neglect the mass of all particles in the loop and hence all particles contribute to the renormalization group running of the universal gauge coupling. However, for $E \ll M_G$, one can consider an effective field theory including only the states with mass $< E \ll M_G$. The gauge symmetry of the EFT is $\mathrm{SU}(3)_C \times \mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$, and the three gauge couplings renormalize independently. The states of the EFT include only those of the SM; 12 gauge bosons, 3 families of quarks and leptons, and one or

[1] Also, in recent years there has been a great deal of progress in constructing three and four family models in Type IIA string theory with intersecting D6 branes [11]. Although these models can incorporate SU(5) or a Pati-Salam symmetry group in four dimensions, they typically have problems with gauge coupling unification. In the former case this is due to charged exotics which affect the RG running, while in the latter case the $\mathrm{SU}(4)\times\mathrm{SU}(2)_L\times\mathrm{SU}(2)_R$ symmetry never unifies. Note, heterotic string theory models also exist whose low energy effective 4d field theory is a SUSY GUT [12]. These models have all the virtues and problems of 4d GUTs. Finally, many heterotic string models have been constructed with the standard model gauge symmetry in 4d and no intermediate GUT symmetry in less than 10d. Recently some minimal 3 family supersymmetric models have been constructed [13,14]. These theories may retain some of the symmetry relations of GUTs, however the unification scale would typically be the string scale, of order 5×10^{17} GeV, which is inconsistent with low energy data. A way out of this problem was discovered in the context of the strongly coupled heterotic string, defined in an effective 11 dimensions [15]. In this case the 4d Planck scale (which controls the value of the string scale) now unifies with the GUT scale.

more Higgs doublets. At M_G, the two effective theories [the GUT itself is most likely the EFT of a more fundamental theory defined at a higher scale] must give identical results; hence we have the boundary conditions $g_3 = g_2 = g_1 \equiv g_G$, where at any scale $\mu < M_G$, we have $g_2 \equiv g$ and $g_1 = \sqrt{5/3}\, g'$. Then using two low-energy couplings, such as $\alpha_s(M_Z)$, $\alpha_{EM}(M_Z)$, the two independent parameters α_G, M_G can be fixed. The third gauge coupling, $\sin^2\theta_W$ in this case, is then predicted. This was the procedure up until about 1991 [17,18]. Subsequently, the uncertainties in $\sin^2\theta_W$ were reduced tenfold. Since then, $\alpha_{EM}(M_Z)$, $\sin^2\theta_W$ have been used as input to predict α_G, M_G, and $\alpha_s(M_Z)$ [19].

We emphasize that the above boundary condition is only valid when using one-loop-renormalization group [RG] running. With precision electroweak data, however, it is necessary to use two-loop-RG running. Hence, one must include one-loop-threshold corrections to gauge coupling boundary conditions at both the weak and GUT scales. In this case, it is always possible to define the GUT scale as the point where $\alpha_1(M_G) = \alpha_2(M_G) \equiv \tilde{\alpha}_G$ and $\alpha_3(M_G) = \tilde{\alpha}_G\,(1 + \epsilon_3)$. The threshold correction ϵ_3 is a logarithmic function of all states with mass of order M_G and $\tilde{\alpha}_G = \alpha_G + \Delta$, where α_G is the GUT coupling constant above M_G, and Δ is a one-loop-threshold correction. To the extent that gauge coupling unification is perturbative, the GUT threshold corrections are small and calculable. This presumes that the GUT scale is sufficiently below the Planck scale or any other strong coupling extension of the GUT, such as a strongly coupled string theory.

Supersymmetric grand unified theories [SUSY GUTs] are an extension of non-SUSY GUTs [20]. The key difference between SUSY GUTs and non-SUSY GUTs is the low-energy effective theory. The low-energy effective field theory in a SUSY GUT is assumed to satisfy $N = 1$ supersymmetry down to scales of order the weak scale, in addition to the SM gauge symmetry. Hence, the spectrum includes all the SM states, plus their supersymmetric partners. It also includes one pair (or more) of Higgs doublets; one to give mass to up-type quarks, and the other to down-type quarks and charged leptons. Two doublets with opposite hypercharge Y are also needed to cancel fermionic triangle anomalies. Finally, it is important to recognize that a low-energy SUSY-breaking scale (the scale at which the SUSY partners of SM particles obtain mass) is necessary to solve the gauge hierarchy problem.

Simple non-SUSY SU(5) is ruled out, initially by the increased accuracy in the measurement of $\sin^2\theta_W$, and by early bounds on the proton lifetime (see below) [18]. However, by now LEP data [19] has conclusively shown that SUSY GUTs is the new Standard Model; by which we mean the theory used to guide the search for new physics beyond the present SM (see Fig. 15.1). SUSY extensions of the SM have the property that their effects decouple as the effective SUSY-breaking scale is increased. Any theory beyond the SM must have this property simply because the SM works so well. However, the SUSY-breaking scale cannot be increased with impunity, since this would reintroduce a gauge hierarchy problem. Unfortunately there is no clear-cut answer to the question, "When is the SUSY-breaking scale too high?" A conservative bound would suggest that the third generation squarks and sleptons must be lighter than about 1 TeV, in order that the one-loop corrections to the Higgs mass from Yukawa interactions remain of order

the Higgs mass bound itself.

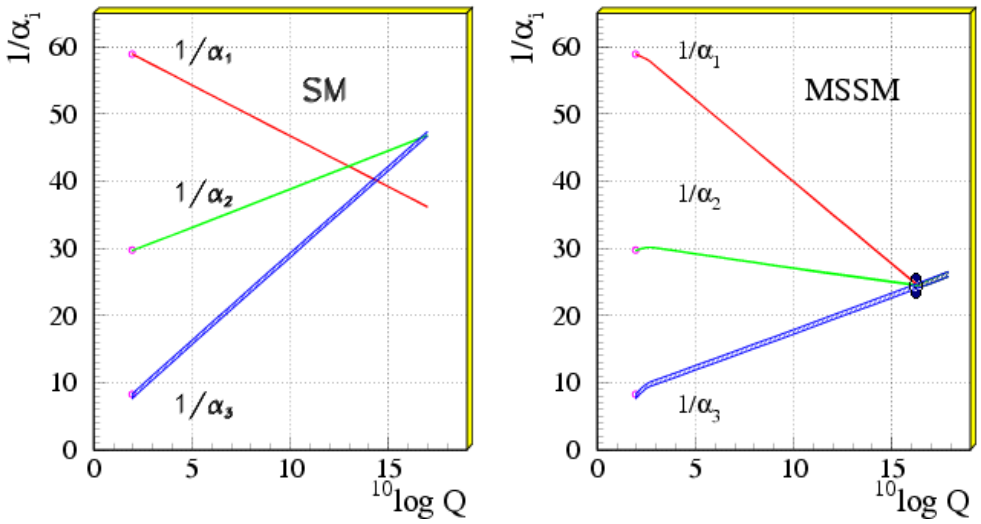


Figure 15.1: Gauge coupling unification in non-SUSY GUTs on the left vs. SUSY GUTs on the right using the LEP data as of 1991. Note, the difference in the running for SUSY is the inclusion of supersymmetric partners of standard model particles at scales of order a TeV (Fig. taken from Ref. 21). Given the present accurate measurements of the three low energy couplings, in particular $\alpha_s(M_Z)$, GUT scale threshold corrections are now needed to precisely fit the low energy data. The dark blob in the plot on the right represents these model dependent corrections.

At present, gauge coupling unification within SUSY GUTs works extremely well. Exact unification at M_G, with two-loop-RG running from M_G to M_Z, and one-loop-threshold corrections at the weak scale, fits to within 3 σ of the present precise low-energy data. A small threshold correction at M_G ($\epsilon_3 \sim -3$ to -4%) is sufficient to fit the low-energy data precisely [22–24]. [2] This may be compared to non-SUSY GUTs, where the fit misses by ~ 12 σ, and a precise fit requires new weak-scale states in incomplete GUT multiplets, or multiple GUT-breaking scales.[3]

[2] This result implicitly assumes universal GUT boundary conditions for soft SUSY-breaking parameters at M_G. In the simplest case, we have a universal gaugino mass $M_{1/2}$, a universal mass for squarks and sleptons m_{16}, and a universal Higgs mass m_{10}, as motivated by SO(10). In some cases, threshold corrections to gauge coupling unification can be exchanged for threshold corrections to soft SUSY parameters. See for example, Ref. 25 and references therein.

[3] Non-SUSY GUTs with a more complicated breaking pattern can still fit the data. For example, non-SUSY SO(10) $\rightarrow$ SU(4)$_C$ $\times$ SU(2)$_L$ $\times$ SU(2)$_R$ $\rightarrow$SM, with the second breaking scale of order an intermediate scale, determined by light neutrino masses using the see-saw mechanism, can fit the low-energy data for gauge couplings [26], and at the same time survive nucleon decay bounds [27], discussed in the following section.

Following the analysis of Ref. 24 let us try to understand the need for the GUT threshold correction and its order of magnitude. The renormalization group equations relate the low energy gauge coupling constants $\alpha_i(M_Z),\ i = 1, 2, 3$ to the value of the unification scale Λ_U and the GUT coupling α_U by the expression

$$\frac{1}{\alpha_i(M_Z)} = \frac{1}{\alpha_U} + \frac{b_i}{2\pi} \log\left(\frac{\Lambda_U}{M_Z}\right) + \delta_i \tag{15.1}$$

where Λ_U is the GUT scale evaluated at one loop and the threshold corrections, δ_i, are given by $\delta_i = \delta_i^{(2)} + \delta_i^{(l)} + \delta_i^{(g)}$ with $\delta_i^{(2)}$ representing two loop running effects, $\delta_i^{(l)}$ the light threshold corrections at the SUSY breaking scale and $\delta_i^{(g)} = \delta_i^{(h)} + \delta_i^{(b)}$ representing GUT scale threshold corrections. Note, in this analysis, the two loop RG running is treated on the same footing as weak and GUT scale threshold corrections. One then obtains the prediction

$$(\alpha_3(M_Z) - \alpha_3^{LO}(M_Z))/\alpha_3^{LO}(M_Z) = -\alpha_3^{LO}(M_Z)\ \delta_s \tag{15.2}$$

where $\alpha_3^{LO}(M_Z)$ is the leading order one loop RG result and $\delta_s = \frac{1}{7}(5\delta_1 - 12\delta_2 + 7\delta_3)$ is the net threshold correction. [A similar formula applies at the GUT scale with the GUT threshold correction, ϵ_3, given by $\epsilon_3 = -\tilde{\alpha}_G\ \delta_s^{(g)}$.] Given the experimental inputs [28]:

$$\begin{aligned} \alpha_{em}^{-1}(M_Z) &= 127.906 \pm 0.019 \\ sin^2\theta_W(M_Z) &= 0.2312 \pm 0.0002 \\ \alpha_3(M_Z) &= 0.1187 \pm 0.0020 \end{aligned} \tag{15.3}$$

and taking into account the light threshold corrections, assuming an ensemble of 10 SUSY spectra [24](corresponding to the Snowmass benchmark points), we have

$$\alpha_3^{LO}(M_Z) \approx 0.118 \tag{15.4}$$

and

$$\begin{aligned} \delta_s^{(2)} &\approx -0.82 \\ \delta_s^{(l)} &\approx -0.50 + \frac{19}{28\pi} \log\frac{M_{SUSY}}{M_Z}. \end{aligned}$$

For $M_{SUSY} = 1$ TeV, we have $\delta_s^{(2)} + \delta_s^{(l)} \approx -0.80$. Since the one loop result $\alpha_3^{LO}(M_Z)$ is very close to the experimental value, we need $\delta_s \approx 0$ or equivalently, $\delta_s^{(g)} \approx 0.80$. This corresponds, at the GUT scale, to $\epsilon_3 \approx -3\%$.[4]

[4] In order to fit the low energy data for gauge coupling constants we require a relative shift in $\alpha_3(M_G)$ of order 3% due to GUT scale threshold corrections. If these GUT scale corrections were not present, however, weak scale threshold corrections of order 9% (due to the larger value of α_3 at M_Z) would be needed to resolve the discrepancy with the data for exact gauge coupling unification at M_G. Leaving out the fact that any consistent GUT necessarily contributes threshold corrections at the GUT scale, it is much more difficult to find the necessary larger corrections at the weak scale. For example, we need $M_{SUSY} \approx 40$ TeV for the necessary GUT scale threshold correction to vanish.

In four dimensional SUSY GUTs, the threshold correction ϵ_3 receives a positive contribution from Higgs doublets and triplets.[5] Thus a larger, negative contribution must come from the GUT breaking sector of the theory. This is certainly possible in specific SO(10) [29] or SU(5) [30] models, but it is clearly a significant constraint on the 4d GUT sector of the theory. In five or six dimensional orbifold GUTs, on the other hand, the "GUT scale" threshold correction comes from the Kaluza-Klein modes between the compactification scale, M_c, and the effective cutoff scale M_*.[6] Thus, in orbifold GUTs, gauge coupling unification at two loops is only consistent with the low energy data with a fixed value for M_c and M_*.[7] Typically, one finds $M_c < M_G = 3 \times 10^{16}$ GeV, where M_G is the 4d GUT scale. Since the grand unified gauge bosons, responsible for nucleon decay, get mass at the compactification scale, the result $M_c < M_G$ for orbifold GUTs has significant consequences for nucleon decay.

A few final comments are in order. We do not consider the scenario of split supersymmetry [33] in this review. In this scenario squarks and sleptons have mass at a scale $\tilde{m} \gg M_Z$, while gauginos and Higgsinos have mass of order the weak scale. Gauge coupling unification occurs at a scale of order 10^{16} GeV, *provided that the scale* $\tilde{m}$ *lies in the range* $10^3 - 10^{11}$ *GeV* [34]. A serious complaint concerning the split SUSY scenario is that it does not provide a solution to the gauge hierarchy problem. Moreover, it is only consistent with grand unification if it also postulates an "intermediate" scale, $\tilde{m}$, for scalar masses. In addition, it is in conflict with $b - \tau$ Yukawa unification, unless $\tan\beta$ is fine-tuned to be close to 1 [34]. [8]

We have also neglected to discuss non-supersymmetric GUTs in four dimensions which still survive once one allows for several scales of GUT symmetry breaking [26]. Finally, it has been shown that non-supersymmetric GUTs in warped 5 dimensional orbifolds can be consistent with gauge coupling unification, assuming that the right-handed top quark and the Higgs doublets are composite-like objects with a compositeness scale of order a

---

[5] Note, the Higgs contribution is given by $\epsilon_3 = \frac{3\tilde{\alpha}_G}{5\pi} \log|\frac{\tilde{M}_t\ \gamma}{M_G}|$ where $\tilde{M}_t$ is the effective color triplet Higgs mass (setting the scale for dimension 5 baryon and lepton number violating operators) and $\gamma = \lambda_b/\lambda_t$ at M_G. Since $\tilde{M}_t$ is necessarily greater than M_G, the Higgs contribution to ϵ_3 is positive.

[6] In string theory, the cutoff scale is the string scale.

[7] It is interesting to note that a ratio $M_*/M_c \sim 100$, needed for gauge coupling unification to work in orbifold GUTs is typically the maximum value for this ratio consistent with perturbativity [31]. In addition, in orbifold GUTs brane-localized gauge kinetic terms may destroy the successes of gauge coupling unification. However, for values of $M_*/M_c = M_*\pi R \gg 1$ the unified bulk gauge kinetic terms can dominate over the brane-localized terms [32].

[8] $b - \tau$ Yukawa unification only works for $\tilde{m} < 10^4$ for $\tan\beta \geq 1.5$. This is because the effective theory between the gaugino mass scale and $\tilde{m}$ includes only one Higgs doublet, as in the standard model. In this case, the large top quark Yukawa coupling tends to increase the ratio λ_b/λ_τ as one runs down in energy below $\tilde{m}$. This is opposite to what happens in MSSM where the large top quark Yukawa coupling decreases the ratio λ_b/λ_τ [35].

TeV [36].

15.1.5. ***Nucleon Decay*** :

Baryon number is necessarily violated in any GUT [37]. In SU(5), nucleons decay via the exchange of gauge bosons with GUT scale masses, resulting in dimension-6 baryon-number-violating operators suppressed by $(1/M_G^2)$. The nucleon lifetime is calculable and given by $\tau_N \propto M_G^4/(\alpha_G^2\, m_p^5)$. The dominant decay mode of the proton (and the baryon-violating decay mode of the neutron), via gauge exchange, is $p \to e^+\, \pi^0$ ($n \to e^+\, \pi^-$). In any simple gauge symmetry, with one universal GUT coupling and scale ($\alpha_G,\ M_G$), the nucleon lifetime from gauge exchange is calculable. Hence, the GUT scale may be directly observed via the extremely rare decay of the nucleon. Experimental searches for nucleon decay began with the Kolar Gold Mine, Homestake, Soudan, NUSEX, Frejus, HPW, and IMB detectors [17]. The present experimental bounds come from Super-Kamiokande and Soudan II. We discuss these results shortly. Non-SUSY GUTs are also ruled out by the non-observation of nucleon decay [18]. In SUSY GUTs, the GUT scale is of order 3×10^{16} GeV, as compared to the GUT scale in non-SUSY GUTs, which is of order 10^{15} GeV. Hence, the dimension-6 baryon-violating operators are significantly suppressed in SUSY GUTs [20] with $\tau_p \sim 10^{34-38}$ yrs.

However, in SUSY GUTs, there are additional sources for baryon-number violation—dimension-4 and -5 operators [38]. Although the notation does not change, when discussing SUSY GUTs, all fields are implicitly bosonic superfields, and the operators considered are the so-called F terms, which contain two fermionic components, and the rest scalars or products of scalars. Within the context of SU(5), the dimension-4 and -5 operators have the form $(\mathbf{10}\ \bar{\mathbf{5}}\ \bar{\mathbf{5}}) \supset (u^c\ d^c\ d^c) + (Q\ L\ d^c) + (e^c\ L\ L)$, and $(\mathbf{10}\ \mathbf{10}\ \mathbf{10}\ \bar{\mathbf{5}})$ $\supset (Q\ Q\ Q\ L) + (u^c\ u^c\ d^c\ e^c) + B$ and L conserving terms, respectively. The dimension-4 operators are renormalizable with dimensionless couplings; similar to Yukawa couplings. On the other hand, the dimension-5 operators have a dimensionful coupling of order $(1/M_G)$.

The dimension-4 operators violate baryon number or lepton number, respectively, but not both. The nucleon lifetime is extremely short if both types of dimension-4 operators are present in the low-energy theory. However, both types can be eliminated by requiring R parity. In SU(5), the Higgs doublets reside in a $\mathbf{5_H}$, $\bar{\mathbf{5}}_\mathbf{H}$, and R parity distinguishes the $\bar{\mathbf{5}}$ (quarks and leptons) from $\bar{\mathbf{5}}_\mathbf{H}$ (Higgs). R parity [39] (or more precisely, its cousin, family reflection symmetry) (see Dimopoulos and Georgi [20] and DRW [40]) takes $F \to -F$, $H \to H$ with $F = \{\mathbf{10},\ \bar{\mathbf{5}}\}$, $H = \{\bar{\mathbf{5}}_\mathbf{H},\ \mathbf{5_H}\}$. This forbids the dimension-4 operator $(\mathbf{10}\ \bar{\mathbf{5}}\ \bar{\mathbf{5}})$, but allows the Yukawa couplings of the form $(\mathbf{10}\ \bar{\mathbf{5}}\ \bar{\mathbf{5}}_\mathbf{H})$ and $(\mathbf{10}\ \mathbf{10}\ \mathbf{5_H})$. It also forbids the dimension-3, lepton-number-violating operator $(\bar{\mathbf{5}}\ \mathbf{5_H})$ $\supset (L\ H_u)$, with a coefficient with dimensions of mass which, like the μ parameter, could be of order the weak scale and the dimension-5, baryon-number-violating operator $(\mathbf{10}\ \mathbf{10}\ \mathbf{10}\ \bar{\mathbf{5}}_\mathbf{H}) \supset (Q\ Q\ Q\ H_d) + \cdots$.

Note, in the MSSM, it is possible to retain R-parity-violating operators at low energy, as long as they violate either baryon number or lepton number only, but not both. Such schemes are natural if one assumes a low-energy symmetry, such as lepton number, baryon number, or a baryon parity [41]. However, these symmetries cannot be embedded

in a GUT. Thus, in a SUSY GUT, only R parity can prevent unwanted dimension four operators. Hence, by naturalness arguments, R parity must be a symmetry in the effective low-energy theory of any SUSY GUT. This does not mean to say that R parity is guaranteed to be satisfied in any GUT.

Note also, R parity distinguishes Higgs multiplets from ordinary families. In SU(5), Higgs and quark/lepton multiplets have identical quantum numbers; while in $E(6)$, Higgs and families are unified within the fundamental **27** representation. Only in SO(10) are Higgs and ordinary families distinguished by their gauge quantum numbers. Moreover, the Z(4) center of SO(10) distinguishes **10**s from **16**s, and can be associated with R parity [42].

In SU(5), dimension-5 baryon-number-violating operators may be forbidden at tree level by additional symmetries. These symmetries are typically broken, however, by the VEVs responsible for the color triplet Higgs masses. Consequently, these dimension-5 operators are generically generated via color triplet Higgsino exchange. Hence, the color triplet partners of Higgs doublets must necessarily obtain mass of order the GUT scale. The dominant decay modes from dimension-5 operators are $p \to K^+ \, \overline{\nu}$ $(n \to K^0 \, \overline{\nu})$. This is due to a simple symmetry argument; the operators $(Q_i \, Q_j \, Q_k \, L_l)$, $(u_i^c \, u_j^c \, d_k^c \, e_l^c)$ (where $i, \; j, \; k, \; l = 1, 2, 3$ are family indices, and color and weak indices are implicit) must be invariant under $\mathrm{SU}(3)_C$ and $\mathrm{SU}(2)_L$. As a result, their color and weak doublet indices must be anti-symmetrized. However, since these operators are given by bosonic superfields, they must be totally symmetric under interchange of all indices. Thus, the first operator vanishes for $i = j = k$, and the second vanishes for $i = j$. Hence, a second or third generation member must exist in the final state [40].

Recent Super-Kamiokande bounds on the proton lifetime severely constrain these dimension-6 and dimension-5 operators with $\tau_{(p \to e^+ \pi^0)} > 5.0 \times 10^{33}$ yrs (79.3 ktyr exposure), $\tau_{(n \to e^+ \pi^-)} > 5 \times 10^{33}$ yrs (61 ktyr), and $\tau_{(p \to K^+ \overline{\nu})} > 1.6 \times 10^{33}$ yrs (79.3 ktyr), $\tau_{(n \to K^0 \overline{\nu})} > 1.7 \times 10^{32}$ yrs (61 ktyr) at (90% CL) based on the listed exposures [43]. These constraints are now sufficient to rule out minimal SUSY SU(5) [44]. [9] Non-minimal Higgs sectors in SU(5) or SO(10) theories still survive [23,30]. The upper bound on the proton lifetime from these theories is approximately a factor of 5 above the experimental bounds. They are, however, being pushed to their theoretical limits. Hence, if SUSY GUTs are correct, nucleon decay should be seen soon.

Is there a way out of this conclusion? Orbifold GUTs and string theories, see Sect. 15.1.3, contain grand unified symmetries realized in higher dimensions. In the process of compactification and GUT symmetry breaking, color triplet Higgs states are removed (projected out of the massless sector of the theory). In addition, the same projections typically rearrange the quark and lepton states so that the massless states which survive emanate from different GUT multiplets. In these models, proton decay due to dimension

[9] This conclusion relies on the mild assumption that the three-by-three matrices diagonalizing squark and slepton mass matrices are not so different from their fermionic partners. It has been shown that if this caveat is violated, then dimension five proton decay in minimal SUSY $SU(5)$ may not be ruled out [45].

5 operators can be severely suppressed or eliminated completely. However, proton decay due to dimension 6 operators may be enhanced, since the gauge bosons mediating proton decay obtain mass at the compactification scale, M_c, which is less than the 4d GUT scale (see the discussion at the end of Section 15.1.4), or suppressed, if the states of one family come from different irreducible representations. Which effect dominates is a model dependent issue. In some complete 5d orbifold GUT models [47,24] the lifetime for the decay $\tau(p \to e^+\pi^0)$ can be near the excluded bound of 5×10^{33} years with, however, large model dependent and/or theoretical uncertainties. In other cases, the modes $p \to K^+\bar{\nu}$ and $p \to K^0\mu^+$ may be dominant [24]. To summarize, in either 4d or orbifold string/field theories, nucleon decay remains a premier signature for SUSY GUTs. Moreover, the observation of nucleon decay may distinguish extra-dimensional orbifold GUTs from four dimensional ones.

Before concluding the topic of baryon-number violation, consider the status of $\Delta B = 2$ neutron- anti-neutron oscillations. Generically, the leading operator for this process is the dimension-9 six-quark operator $G_{(\Delta B=2)}$ $(u^c\ d^c\ d^c\ u^c\ d^c\ d^c)$, with dimensionful coefficient $G_{(\Delta B=2)} \sim 1/M^5$. The present experimental bound $\tau_{n-\overline{n}} \geq 0.86 \times 10^8$ sec. at 90% CL [48] probes only up to the scale $M \leq 10^6$ GeV. For $M \sim M_G$, $n-\overline{n}$ oscillations appear to be unobservable for any GUT (for a recent discussion see Ref. 49).

15.1.6. *Yukawa coupling unification* :

15.1.6.1. *3rd generation, b–τ or t–b–τ unification:*

If quarks and leptons are two sides of the same coin, related by a new grand unified gauge symmetry, then that same symmetry relates the Yukawa couplings (and hence the masses) of quarks and leptons. In SU(5), there are two independent renormalizable Yukawa interactions given by λ_t $(\mathbf{10\ 10\ 5_H}) + \lambda$ $(\mathbf{10\ \bar{5}\ \bar{5}_H})$. These contain the SM interactions λ_t $(\mathbf{Q\ u^c\ H_u}) + \lambda$ $(\mathbf{Q\ d^c\ H_d} + \mathbf{e^c\ L\ H_d})$. Hence, at the GUT scale, we have the tree-level relation, $\lambda_b = \lambda_\tau \equiv \lambda$ [35]. In SO(10), there is only one independent renormalizable Yukawa interaction given by λ $(\mathbf{16\ 16\ 10_H})$, which gives the tree-level relation, $\lambda_t = \lambda_b = \lambda_\tau \equiv \lambda$ [50,51]. Note, in the discussion above, we assume the minimal Higgs content, with Higgs in $\mathbf{5}$, $\bar{\mathbf{5}}$ for SU(5) and $\mathbf{10}$ for SO(10). With Higgs in higher-dimensional representations, there are more possible Yukawa couplings. [58–60]

In order to make contact with the data, one now renormalizes the top, bottom, and τ Yukawa couplings, using two-loop-RG equations, from M_G to M_Z. One then obtains the running quark masses $m_t(M_Z) = \lambda_t(M_Z)\ v_u$, $m_b(M_Z) = \lambda_b(M_Z)\ v_d$, and $m_\tau(M_Z) = \lambda_\tau(M_Z)\ v_d$, where $< H_u^0 > \equiv v_u = \sin\beta\ v/\sqrt{2}$, $< H_d^0 > \equiv v_d = \cos\beta\ v/\sqrt{2}$, $v_u/v_d \equiv \tan\beta$, and $v \sim 246$ GeV is fixed by the Fermi constant, G_μ.

Including one-loop-threshold corrections at M_Z, and additional RG running, one finds the top, bottom, and τ-pole masses. In SUSY, $b-\tau$ unification has two possible solutions, with $\tan\beta \sim 1$ or $40-50$. The small $\tan\beta$ solution is now disfavored by the LEP limit, $\tan\beta > 2.4$ [52]. [10] The large $\tan\beta$ limit overlaps the SO(10) symmetry relation.

[10] However, this bound disappears if one takes $M_{SUSY} = 2$ TeV and $m_t = 180$ GeV [53].

When $\tan\beta$ is large, there are significant weak-scale threshold corrections to down quark and charged lepton masses, from either gluino and/or chargino loops [54]. Yukawa unification (consistent with low energy data) is only possible in a restricted region of SUSY parameter space with important consequences for SUSY searches [55].

15.1.6.2. *Three families:*

Simple Yukawa unification is not possible for the first two generations, of quarks and leptons. Consider the SU(5) GUT scale relation $\lambda_b = \lambda_\tau$. If extended to the first two generations, one would have $\lambda_s = \lambda_\mu$, $\lambda_d = \lambda_e$, which gives $\lambda_s/\lambda_d = \lambda_\mu/\lambda_e$. The last relation is a renormalization group invariant, and is thus satisfied at any scale. In particular, at the weak scale, one obtains $m_s/m_d = m_\mu/m_e$, which is in serious disagreement with the data, namely $m_s/m_d \sim 20$ and $m_\mu/m_e \sim 200$. An elegant solution to this problem was given by Georgi and Jarlskog [56]. Of course, a three-family model must also give the observed CKM mixing in the quark sector. Note, although there are typically many more parameters in the GUT theory above M_G, it is possible to obtain effective low-energy theories with many fewer parameters making strong predictions for quark and lepton masses.

It is important to note that grand unification alone is not sufficient to obtain predictive theories of fermion masses and mixing angles. Other ingredients are needed. In one approach additional global family symmetries are introduced (non-abelian family symmetries can significantly reduce the number of arbitrary parameters in the Yukawa matrices). These family symmetries constrain the set of effective higher dimensional fermion mass operators. In addition, sequential breaking of the family symmetry is correlated with the hierarchy of fermion masses. Three-family models exist which fit all the data, including neutrino masses and mixing [57]. In a completely separate approach for $SO(10)$ models, the Standard Model Higgs bosons are contained in the higher dimensional Higgs representations including the **10**, $\overline{\mathbf{126}}$ and/or **120**. Such theories have been shown to make predictions for neutrino masses and mixing angles [58–60].

15.1.7. ***Neutrino Masses*** :

Atmospheric and solar neutrino oscillations require neutrino masses. Adding three "sterile" neutrinos ν^c with the Yukawa coupling λ_ν ($\nu^{\mathbf{c}}\,\mathbf{L}\,\mathbf{H_u}$), one easily obtains three massive Dirac neutrinos with mass $m_\nu = \lambda_\nu\, v_u$. [11] However, in order to obtain a tau neutrino with mass of order 0.1 eV, one needs $\lambda_{\nu_\tau}/\lambda_\tau \leq 10^{-10}$. The see-saw mechanism, on the other hand, can naturally explain such small neutrino masses [61,62]. Since ν^c has no SM quantum numbers, there is no symmetry (other than global lepton number) which prevents the mass term $\frac{1}{2}\,\nu^c\,M\,\nu^c$. Moreover, one might expect $M \sim M_G$. Heavy "sterile" neutrinos can be integrated out of the theory, defining an effective low-energy theory with only light active Majorana neutrinos, with the effective dimension-5 operator $\frac{1}{2}\,(\mathbf{L}\,\mathbf{H_u})\,\lambda_\nu^T\,M^{-1}\,\lambda_\nu\,(\mathbf{L}\,\mathbf{H_u})$. This then leads to a 3×3 Majorana neutrino mass matrix $\mathbf{m} = m_\nu^T\,M^{-1}\,m_\nu$.

[11] Note, these "sterile" neutrinos are quite naturally identified with the right-handed neutrinos necessarily contained in complete families of $SO(10)$ or Pati-Salam.

Atmospheric neutrino oscillations require neutrino masses with $\Delta m_\nu^2 \sim 3 \times 10^{-3}$ eV^2 with maximal mixing, in the simplest two-neutrino scenario. With hierarchical neutrino masses, $m_{\nu_\tau} = \sqrt{\Delta m_\nu^2} \sim 0.055$ eV. Moreover, via the "see-saw" mechanism, $m_{\nu_\tau} = m_t(m_t)^2/(3M)$. Hence, one finds $M \sim 2 \times 10^{14}$ GeV; remarkably close to the GUT scale. Note we have related the neutrino-Yukawa coupling to the top-quark-Yukawa coupling $\lambda_{\nu_\tau} = \lambda_t$ at M_G, as given in SO(10) or SU(4) $\times$ SU(2)$_L$ $\times$ SU(2)$_R$. However, at low energies they are no longer equal, and we have estimated this RG effect by $\lambda_{\nu_\tau}(M_Z) \approx \lambda_t(M_Z)/\sqrt{3}$.

15.1.8. *Selected Topics* :

15.1.8.1. *Magnetic Monopoles:*

In the broken phase of a GUT, there are typically localized classical solutions carrying magnetic charge under an unbroken U(1) symmetry [63]. These magnetic monopoles with mass of order M_G/α_G are produced during the GUT phase transition in the early universe. The flux of magnetic monopoles is experimentally found to be less than $\sim 10^{-16}$ cm^{-2} s^{-1} sr^{-1} [64]. Many more are predicted however, hence the GUT monopole problem. In fact, one of the original motivations for an inflationary universe is to solve the monopole problem by invoking an epoch of rapid inflation after the GUT phase transition [65]. This would have the effect of diluting the monopole density as long as the reheat temperature is sufficiently below M_G. Other possible solutions to the monopole problem include: sweeping them away by domain walls [66], U(1) electromagnetic symmetry breaking at high temperature [67] or GUT symmetry non-restoration [68]. Parenthetically, it was also shown that GUT monopoles can catalyze nucleon decay [69]. A significantly lower bound on the monopole flux can then be obtained by considering X-ray emission from radio pulsars due to monopole capture and the subsequent nucleon decay catalysis [70].

15.1.8.2. *Baryogenesis via Leptogenesis:*

Baryon-number-violating operators in SU(5) or SO(10) preserve the global symmetry $B-L$. Hence, the value of the cosmological $B-L$ density is an initial condition of the theory, and is typically assumed to be zero. On the other hand, anomalies of the electroweak symmetry violate $B+L$ while also preserving $B-L$. Hence, thermal fluctuations in the early universe, via so-called sphaleron processes, can drive $B+L$ to zero, washing out any net baryon number generated in the early universe at GUT temperatures [71].

One way out of this dilemma is to generate a net $B-L$ dynamically in the early universe. We have just seen that neutrino oscillations suggest a new scale of physics of order 10^{14} GeV. This scale is associated with heavy Majorana neutrinos with mass M. If in the early universe, the decay of the heavy neutrinos is out of equilibrium and violates both lepton number and CP, then a net lepton number may be generated. This lepton number will then be partially converted into baryon number via electroweak processes [72].

15.1.8.3. *GUT symmetry breaking:*

The grand unification symmetry is necessarily broken spontaneously. Scalar potentials (or superpotentials) exist whose vacua spontaneously break SU(5) and SO(10). These potentials are ad hoc (just like the Higgs potential in the SM), and, therefore it is hoped that they may be replaced with better motivated sectors. Gauge coupling unification now tests GUT-breaking sectors, since it is one of the two dominant corrections to the GUT threshold correction ϵ_3. The other dominant correction comes from the Higgs sector and doublet-triplet splitting. This latter contribution is always positive $\epsilon_3 \propto \ln(M_T/M_G)$ (where M_T is an effective color triplet Higgs mass), while the low-energy data requires $\epsilon_3 < 0$. Hence, the GUT-breaking sector must provide a significant (of order -8%) contribution to ϵ_3 to be consistent with the Super-K bound on the proton lifetime [23,29,30,57].

In string theory (and GUTs in extra-dimensions), GUT breaking may occur due to boundary conditions in the compactified dimensions [7,10]. This is still ad hoc. The major benefits are that it does not require complicated GUT-breaking sectors.

15.1.8.4. *Doublet-triplet splitting:*

The Minimal Supersymmetric Standard Model has a μ problem: why is the coefficient of the bilinear Higgs term in the superpotential μ ($\mathbf{H_u}$ $\mathbf{H_d}$) of order the weak scale when, since it violates no low-energy symmetry, it could be as large as M_G? In a SUSY GUT, the μ problem is replaced by the problem of *doublet-triplet* splitting—giving mass of order M_G to the color triplet Higgs, and mass μ to the Higgs doublets. Several mechanisms for natural doublet-triplet splitting have been suggested, such as the sliding singlet, missing partner or missing VEV [73], and pseudo-Nambu-Goldstone boson mechanisms. Particular examples of the missing partner mechanism for SU(5) [30], the missing VEV mechanism for SO(10) [23,57], and the pseudo-Nambu-Goldstone boson mechanism for SU(6) [74], have been shown to be consistent with gauge coupling unification and proton decay. There are also several mechanisms for explaining why μ is of order the SUSY-breaking scale [75]. Finally, for a recent review of the μ problem and some suggested solutions in SUSY GUTs and string theory, see Refs. [76,9] and references therein.

Once again, in string theory (and orbifold GUTs), the act of breaking the GUT symmetry via orbifolding projects certain states out of the theory. It has been shown that it is possible to remove the color triplet Higgs while retaining the Higgs doublets in this process. Hence the doublet-triplet splitting problem is finessed. As discussed earlier (see Section 15.1.5), this has the effect of eliminating the contribution of dimension 5 operators to nucleon decay.

15.2. Conclusion

Grand unification of the strong and electroweak interactions requires that the three low energy gauge couplings unify (up to small threshold corrections) at a unique scale, M_G. Supersymmetric grand unified theories provide, by far, the most predictive and economical framework allowing for perturbative unification.

The three pillars of SUSY GUTs are:

- gauge coupling unification at $M_G \sim 3 \times 10^{16}$ GeV;
- low-energy supersymmetry [with a large SUSY desert], and
- nucleon decay.

The first prediction has already been verified (see Fig. 15.1). Perhaps the next two will soon be seen. Whether or not Yukawa couplings unify is more model dependent. Nevertheless, the "digital" 16-dimensional representation of quarks and leptons in SO(10) is very compelling, and may yet lead to an understanding of fermion masses and mixing angles.

In any event, the experimental verification of the first three pillars of SUSY GUTs would forever change our view of Nature. Moreover, the concomitant evidence for a vast SUSY desert would expose a huge lever arm for discovery. For then it would become clear that experiments probing the TeV scale could reveal physics at the GUT scale and perhaps beyond. Of course, some questions will still remain: Why do we have three families of quarks and leptons? How is the grand unified symmetry and possible family symmetries chosen by Nature? At what scale might stringy physics become relevant? *Etc.*

References:

1. J. Pati and A. Salam, Phys. Rev. **D8**, 1240 (1973);
   For more discussion on the standard charge assignments in this formalism, see A. Davidson, Phys. Rev. **D20**, 776 (1979); and R.N. Mohapatra and R.E. Marshak, Phys. Lett. **B91**, 222 (1980).
2. H. Georgi and S.L. Glashow, Phys. Rev. Lett. **32**, 438 (1974).
3. H. Georgi, Particles and Fields, *Proceedings of the APS Div. of Particles and Fields*, ed. C. Carlson, p. 575 (1975);
   H. Fritzsch and P. Minkowski, Ann. Phys. **93**, 193 (1975).
4. S.M. Barr, Phys. Lett. **B112**, 219 (1982).
5. A. de Rujula *et al.*, p. 88, *5th Workshop on Grand Unification*, ed. K. Kang *et al.*, World Scientific, Singapore (1984);
   See also earlier paper by Y. Achiman and B. Stech, p. 303, "New Phenomena in Lepton-Hadron Physics," ed. D.E.C. Fries and J. Wess, Plenum, NY (1979).
6. B.R. Greene *et al.*, Nucl. Phys. **B278**, 667 (1986);
   ibid., Nucl. Phys. **B292**, 606 (1987);
   B.R. Greene *et al.*, Nucl. Phys. **B325**, 101 (1989).
7. P. Candelas *et al.*, Nucl. Phys. **B258**, 46 (1985);
   L.J. Dixon *et al.*, Nucl. Phys. **B261**, 678 (1985);
   ibid., Nucl. Phys. **B274**, 285 (1986);

L. E. Ibanez *et al.*, Phys. Lett. **B187**, 25 (1987);
ibid., Phys. Lett. **B191**, 282 (1987);
J.E. Kim *et al.*, Nucl. Phys. **B712**, 139 (2005).

8. T.Kobayashi *et al.*, Phys. Lett. **B593**, 262 (2004);
S. Forste *et al.*, Phys. Rev. **D70**, 106008 (2004);
T. Kobayashi *et al.*, Nucl. Phys. **B704**, 3 (2005);
W. Buchmuller *et al.*, Nucl. Phys. **B712**, 139 (2005).
9. E. Witten, [hep-ph/0201018];
M. Dine *et al.*, Phys. Rev. **D66**, 115001 (2002), [hep-ph/0206268].
10. Y. Kawamura, Prog. Theor. Phys. **103**, 613 (2000);
ibid., **105**, 999 (2001);
G. Altarelli *et al.*, Phys. Lett. **B5111**, 257 (2001);
L.J. Hall *et al.*, Phys. Rev. **D64**, 055003 (2001);
A. Hebecker and J. March-Russell, Nucl. Phys. **B613**, 3 (2001);
T. Asaka *et al.*, Phys. Lett. **B523**, 199 (2001);
L.J. Hall *et al.*, Phys. Rev. **D65**, 035008 (2002);
R. Dermisek and A. Mafi, Phys. Rev. **D65**, 055002 (2002);
H.D. Kim and S. Raby, JHEP **0301**, 056 (2003).
11. For a recent review see, R. Blumenhagen *et al.*, "Toward realistic intersecting D-brane models," hep-th/0502005.
12. G. Aldazabal *et al.*, Nucl. Phys. **B452**, 3 (1995);
Z. Kakushadze and S.H.H. Tye, Phys. Rev. **D54**, 7520 (1996);
Z. Kakushadze *et al.*, Int. J. Mod. Phys. **A13**, 2551 (1998).
13. G. B. Cleaver *et al.*, Int. J. Mod. Phys. **A16**, 425 (2001), hep-ph/9904301;
ibid., Nucl. Phys. **B593**, 471 (2001) hep-ph/9910230.
14. V. Braun *et al.*, Phys. Lett. **B618**, 252 (2005), hep-th/0501070;
ibid., JHEP **506**, 039 (2005), [hep-th/0502155].
15. E. Witten, Nucl. Phys. **B471**, 135 (1996), [hep-th/9602070].
16. H. Georgi *et al.*, Phys. Rev. Lett. **33**, 451 (1974);
See also the definition of effective field theories by S. Weinberg, Phys. Lett. **91B**, 51 (1980).
17. See talks on proposed and running nucleon decay experiments, and theoretical talks by P. Langacker, p. 131, and W.J. Marciano and A. Sirlin, p. 151, in *The Second Workshop on Grand Unification*, eds. J.P. Leveille *et al.*, Birkhäuser, Boston (1981).
18. W.J. Marciano, p. 190, *Eighth Workshop on Grand Unification*, ed. K. Wali, World Scientific Publishing Co., Singapore (1987).
19. U. Amaldi *et al.*, Phys. Lett. **B260**, 447 (1991);
J. Ellis *et al.*, Phys. Lett. **B260**, 131 (1991);
P. Langacker and M. Luo, Phys. Rev. **D44**, 817 (1991);
P. Langacker and N. Polonsky, Phys. Rev. **D47**, 4028 (1993);
M. Carena *et al.*, Nucl. Phys. **B406**, 59 (1993);
see also the review by S. Dimopoulos *et al.*, Physics Today, 25–33, October (1991).
20. S. Dimopoulos *et al.*, Phys. Rev. **D24**, 1681 (1981);
S. Dimopoulos and H. Georgi, Nucl. Phys. **B193**, 150 (1981);

L. Ibanez and G.G. Ross, Phys. Lett. **105B**, 439 (1981);
N. Sakai, Z. Phys. **C11**, 153 (1981);
M.B. Einhorn and D.R.T. Jones, Nucl. Phys. **B196**, 475 (1982);
W.J. Marciano and G. Senjanovic, Phys. Rev. **D25**, 3092 (1982).

21. D.I. Kazakov, Lectures given at the European School on High Energy Physics, Aug.-Sept. 2000, Caramulo, Portugal [hep-ph/0012288v2].
22. V. Lucas and S. Raby, Phys. Rev. **D54**, 2261 (1996) [hep-ph/9601303];
T. Blazek *et al.*, Phys. Rev. **D56**, 6919 (1997) [hep-ph/9611217];
G. Altarelli *et al.*, JHEP **0011**, 040 (2000) [hep-ph/0007254].
23. R. Dermíšek *et al.*, Phys. Rev. **D63**, 035001 (2001);
K.S. Babu *et al.*, Nucl. Phys. **B566**, 33 (2000).
24. M. L. Alciati *et al.*, JHEP **0503**, 054 (2005) [hep-ph/0501086].
25. G. Anderson *et al.*, eConf **C960625**, SUP107 (1996) [hep-ph/9609457].
26. R.N. Mohapatra and M.K. Parida, Phys. Rev. **D47**, 264 (1993).
27. D.G. Lee *et al.*, Phys. Rev. **D51**, 229 (1995).
28. S. Eidelman *et al.* [Particle Data Group], Phys. Lett. **B592**, 1 (2004).
29. K.S. Babu and S.M. Barr, Phys. Rev. **D48**, 5354 (1993);
V. Lucas and S. Raby, Phys. Rev. **D54**, 2261 (1996);
S.M. Barr and S. Raby, Phys. Rev. Lett. **79**, 4748 (1997) and references therein.
30. G. Altarelli *et al.*, JHEP **0011**, 040 (2000) See also earlier papers by A. Masiero *et al.*, Phys. Lett. **B115**, 380 (1982);
B. Grinstein, Nucl. Phys. **B206**, 387 (1982).
31. K.R. Dienes *et al.*, Phys. Rev. Lett. **91**, 061601 (2003).
32. L. J. Hall *et al.*, Phys. Rev. **D64**, 055003 (2001).
33. N. Arkani-Hamed and S. Dimopoulos, JHEP **0506**, 073 (2005) [hep-th/0405159].
34. G. F. Giudice and A. Romanino, Nucl. Phys. **B6999**, 65 (2004) [Erratum: *ibid.*, Nucl. Phys. **B706**, 65 (2005)] [hep-ph/0406088].
35. M. Chanowitz *et al.*, Nucl. Phys. **B135**, 66 (1978);
For the corresponding SUSY analysis, see M. Einhorn and D.R.T. Jones, Nucl. Phys. **B196**, 475 (1982);
K. Inoue *et al.*, Prog. Theor. Phys. **67**, 1889 (1982);
L. E. Ibanez and C. Lopez, Phys. Lett. **B126**, 54 (1983);
ibid., Nucl. Phys. **B233**, 511 (1984).
36. K. Agashe *et al.*, [hep-ph/0502222].
37. M. Gell-Mann *et al.*, in *Supergravity*, eds. P. van Nieuwenhuizen and D.Z. Freedman, North-Holland, Amsterdam, 1979, p. 315.
38. S. Weinberg, Phys. Rev. **D26**, 287 (1982);
N. Sakai and T Yanagida, Nucl. Phys. **B197**, 533 (1982).
39. G. Farrar and P. Fayet, Phys. Lett. **B76**, 575 (1978).
40. S. Dimopoulos *et al.*, Phys. Lett. **112B**, 133 (1982);
J. Ellis *et al.*, Nucl. Phys. **B202**, 43 (1982).
41. L.E. Ibanez and G.G. Ross, Nucl. Phys. **B368**, 3 (1992).
42. For a recent discussion, see C.S. Aulakh *et al.*, Nucl. Phys. **B597**, 89 (2001).

43. See talks by Matthew Earl, *NNN workshop*, Irvine, February (2000);
Y. Totsuka, *SUSY2K*, CERN, June (2000);
Y. Suzuki, *International Workshop on Neutrino Oscillations and their Origins*, Tokyo, Japan, December (2000), and *Baksan School, Baksan Valley*, Russia, April (2001), hep-ex/0110005;
K. Kobayashi [Super-Kamiokande Collaboration], "Search for nucleon decay from Super-Kamiokande," *Prepared for 27th International Cosmic Ray Conference (ICRC 2001), Hamburg, Germany, 7-15 Aug 2001.* For published results see : Y. Hayato *et al.* (Super-Kamiokande Collab.), Phys. Rev. Lett. **83**, 1529 (1999);
K. Kobayashi *et al.* [Super-Kamiokande Collaboration], [hep-ex/0502026].
44. T. Goto and T. Nihei, Phys. Rev. **D59**, 115009 (1999) [hep-ph/9808255];
H. Murayama and A. Pierce, Phys. Rev. **D65**, 055009 (2002) hep-ph/0108104.
45. B. Bajc *et al.*, Phys. Rev. **D66**, 075005 (2002) [hep-ph/0204311].
46. J. L. Chkareuli and I. G. Gogoladze, Phys. Rev. **D58**, 055011 (1998) [hep-ph/9803335].
47. L. J. Hall and Y. Nomura, Phys. Rev. **D66**, 075004 (2002);
H. D. Kim *et al.*, JHEP **0505**, 036 (2005).
48. M. Baldoceolin et al., Z. Phys. **C63**, 409 (1994).
49. K. S. Babu and R. N. Mohapatra, Phys. Lett. **B518**, 269 (2001) [hep-ph/0108089].
50. H. Georgi and D.V. Nanopoulos, Nucl. Phys. **B159**, 16 (1979);
J. Harvey *et al.*, Phys. Lett. **B92**, 309 (1980);
ibid., Nucl. Phys. **B199**, 223 (1982).
51. T. Banks, Nucl. Phys. **B303**, 172 (1988);
M. Olechowski and S. Pokorski, Phys. Lett. **B214**, 393 (1988);
S. Pokorski, Nucl. Phys. (Proc. Supp.) **B13**, 606 (1990);
B. Ananthanarayan *et al.*, Phys. Rev. **D44**, 1613 (1991);
Q. Shafi and B. Ananthanarayan, ICTP Summer School lectures (1991);
S. Dimopoulos *et al.*, Phys. Rev. Lett. **68**, 1984 (1992);
ibid., Phys. Rev. **D45**, 4192 (1992);
G. Anderson *et al.*, Phys. Rev. **D47**, 3702 (1993);
B. Ananthanarayan *et al.*, Phys. Lett. **B300**, 245 (1993);
G. Anderson *et al.*, Phys. Rev. **D49**, 3660 (1994);
B. Ananthanarayan *et al.*, Phys. Rev. **D50**, 5980 (1994).
52. LEP Higgs Working Group and ALEPH collaboration and DELPHI collaboration and L3 collaboration and OPAL Collaboration, Preliminary results, [hep-ex/0107030] (2001).
53. M. Carena and H. E. Haber, Prog. in Part. Nucl. Phys. **50**, 63 (2003) [hep-ph/0208209].
54. L.J. Hall *et al.*, Phys. Rev. **D50**, 7048 (1994);
M. Carena *et al.*, Nucl. Phys. **B419**, 213 (1994);
R. Rattazzi and U. Sarid, Nucl. Phys. **B501**, 297 (1997).
55. Blazek *et al.*, Phys. Rev. Lett. **88**, 111804 (2002) [hep-ph/0107097];
ibid., Phys. Rev. **D65**, 115004 (2002) [hep-ph/0201081];

K. Tobe and J. D. Wells, Nucl. Phys. **B663**, 123 (2003) [hep-ph/0301015];
D. Auto *et al.*, JHEP **0306**, 023 (2003) [hep-ph/0302155].
56. H. Georgi and C. Jarlskog, Phys. Lett. **B86**, 297 (1979).
57. K.S. Babu and R.N. Mohapatra, Phys. Rev. Lett. **74**, 2418 (1995);
V. Lucas and S. Raby, Phys. Rev. **D54**, 2261 (1996);
T. Blažek *et al.*, Phys. Rev. **D56**, 6919 (1997);
R. Barbieri *et al.*, Nucl. Phys. **B493**, 3 (1997);
T. Blazek *et al.*, Phys. Rev. **D60**, 113001 (1999);
ibid., Phys. Rev. **D62**, 055001 (2000);
Q. Shafi and Z. Tavartkiladze, Phys. Lett. **B487**, 145 (2000);
C.H. Albright and S.M. Barr, Phys. Rev. Lett. **85**, 244 (2000);
K.S. Babu *et al.*, Nucl. Phys. **B566**, 33 (2000);
G. Altarelli *et al.*, Ref. 30;
Z. Berezhiani and A. Rossi, Nucl. Phys. **B594**, 113 (2001);
C. H. Albright and S. M. Barr, Phys. Rev. **D64**, 073010 (2001) [hep-ph/0104294];
R. Dermisek and S. Raby, Phys. Lett. **B622**, 327 (2005) [hep-ph/0507045].
58. G. Lazarides *et al.*, Nucl. Phys. **B181**, 287 (1981);
T. E. Clark *et al.*, Phys. Lett. **B115**, 26 (1982);
K. S. Babu and R. N. Mohapatra, Phys. Rev. Lett. **70**, 2845 (1993) [hep-ph/9209215].
59. B. Bajc *et al.*, Phys. Rev. Lett. **90**, 051802 (2003) [hep-ph/0210207].
60. H. S. Goh *et al.*, Phys. Lett. **B570**, 215 (2003) [hep-ph/0303055];
ibid., Phys. Rev. **D68**, 115008 (2003) [hep-ph/0308197];
B. Dutta *et al.*, Phys. Rev. **D69**, 115014 (2004) [hep-ph/0402113];
S. Bertolini and M. Malinsky, [hep-ph/0504241];
K. S. Babu and C. Macesanu, [hep-ph/0505200].
61. P. Minkowski, Phys. Lett. **B67**, 421 (1977).
62. T. Yanagida, in *Proceedings of the Workshop on the Unified Theory and the Baryon Number of the Universe*, eds. O. Sawada and A. Sugamoto, KEK report No. 79-18, Tsukuba, Japan, 1979;
S. Glashow, Quarks and leptons, published in Proceedings of the Carg'ese Lectures, M. Levy (ed.), Plenum Press, New York, (1980);
M. Gell-Mann *et al.*, in Supergravity, ed. P. van Nieuwenhuizen *et al.*, North-Holland, Amsterdam, (1979), p. 315;
R.N. Mohapatra and G. Senjanovic, Phys. Rev. Lett. **44**, 912 (1980).
63. G. 't Hooft, Nucl. Phys. **B79**, 276 (1974);
A.M. Polyakov, Pis'ma Zh. Eksp. Teor. Fiz. **20**, 430 (1974) [JETP Lett. **20**, 194 (1974)];
For a pedagogical introduction, see S. Coleman, in *Aspects of Symmetry*, Selected Erice Lectures, Cambridge University Press, Cambridge, (1985), and P. Goddard and D. Olive, Rep. Prog. Phys. **41**, 1357 (1978).
64. I. De Mitri, (MACRO Collab.), Nucl. Phys. (Proc. Suppl.) **B95**, 82 (2001).
65. For a review, see A.D. Linde, *Particle Physics and Inflationary Cosmology*, Harwood Academic, Switzerland (1990).

66. G. R. Dvali *et al.*, Phys. Rev. Lett. **80**, 2281 (1998) [hep-ph/9710301].
67. P. Langacker and S. Y. Pi, Phys. Rev. Lett. **45**, 1 (1980).
68. G. R. Dvali *et al.*, Phys. Rev. Lett. **75**, 4559 (1995) [hep-ph/9507230].
69. V. Rubakov, Nucl. Phys. **B203**, 311 (1982), Institute of Nuclear Research Report No. P-0211, Moscow (1981), unpublished;
C. Callan, Phys. Rev. **D26**, 2058 (1982);
F. Wilczek, Phys. Rev. Lett. **48**, 1146 (1982);
See also, S. Dawson and A.N. Schellekens, Phys. Rev. **D27**, 2119 (1983).
70. K. Freese *et al.*, Phys. Rev. Lett. **51**, 1625 (1983).
71. V. A. Kuzmin *et al.*, Phys. Lett. **B155**, 36 (1985).
72. M. Fukugita and T. Yanagida, Phys. Lett. **B174**, 45 (1986);
See also the recent review by W. Buchmuller *et al.*, hep-ph/0502169 and references therein.
73. S. Dimopoulos and F. Wilczek, *Proceedings Erice Summer School*, ed. A. Zichichi (1981);
K.S. Babu and S.M. Barr, Phys. Rev. **D50**, 3529 (1994).
74. R. Barbieri *et al.*, Nucl. Phys. **B391**, 487 (1993);
Z. Berezhiani *et al.*, Nucl. Phys. **B444**, 61 (1995);
Q. Shafi and Z. Tavartkiladze, Phys. Lett. **B522**, 102 (2001).
75. G.F. Giudice and A. Masiero, Phys. Lett. **B206**, 480 (1988);
J.E. Kim and H.P. Nilles, Mod. Phys. Lett. **A9**, 3575 (1994).
76. L. Randall and C. Csaki, hep-ph/9508208.

NOTE: THE FIGURES IN THIS SECTION ARE INTENDED TO SHOW THE REPRESENTATIVE DATA. THEY ARE NOT MEANT TO BE COMPLETE COMPILATIONS OF ALL THE WORLD'S RELIABLE DATA.

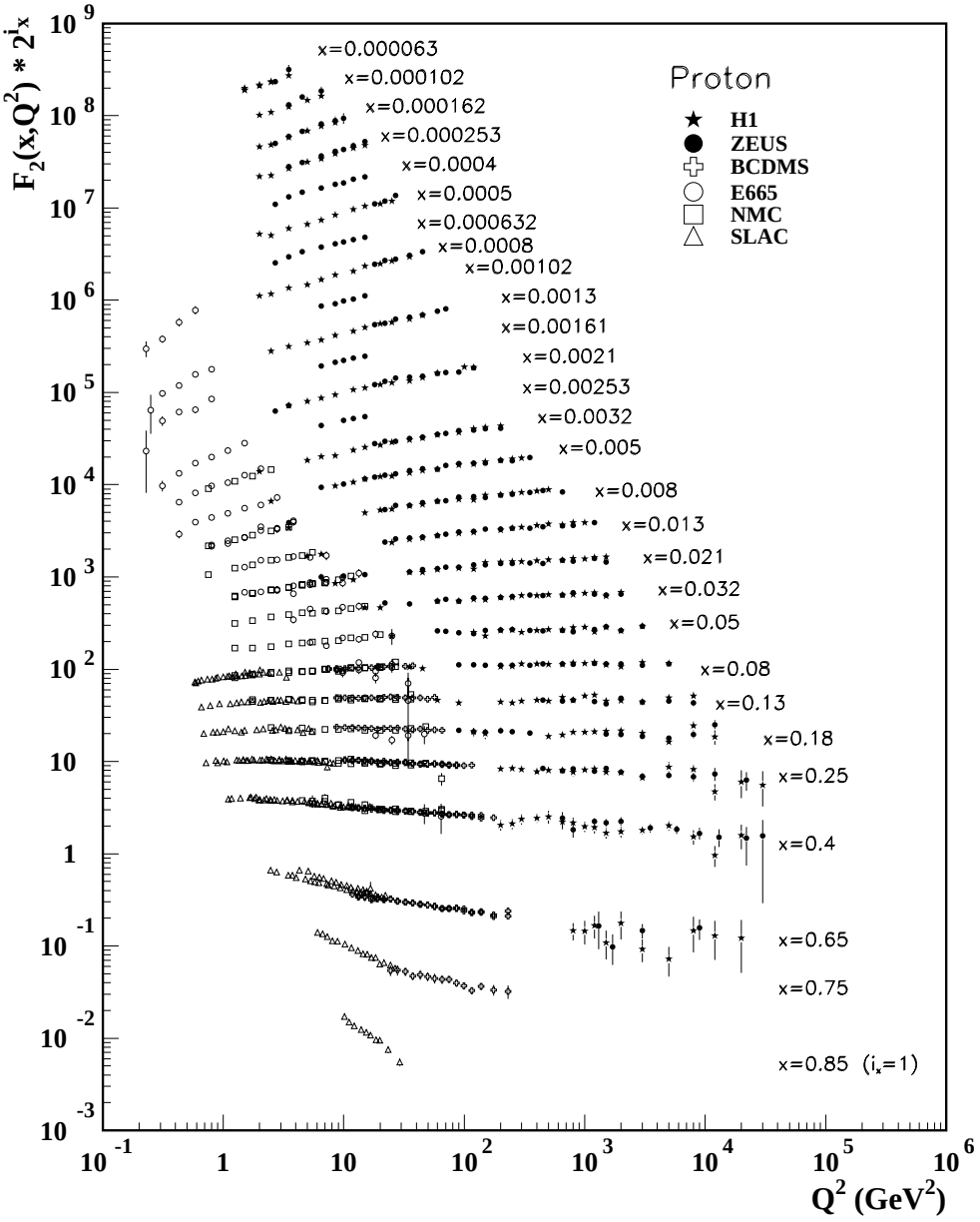


Figure 16.6: The proton structure function F_2^p measured in electromagnetic scattering of positrons on protons (collider experiments ZEUS and H1), in the kinematic domain of the HERA data, for $x > 0.00006$ (*cf.* Fig. 16.9 for data at smaller x and Q^2), and for electrons (SLAC) and muons (BCDMS, E665, NMC) on a fixed target. Statistical and systematic errors added in quadrature are shown. The data are plotted as a function of Q^2 in bins of fixed x. Some points have been slightly offset in Q^2 for clarity. The ZEUS binning in x is used in this plot; all other data are rebinned to the x values of the ZEUS data. For the purpose of plotting, F_2^p has been multiplied by 2^{i_x}, where i_x is the number of the x bin, ranging from $i_x = 1$ ($x = 0.85$) to $i_x = 28$ ($x = 0.000063$). References: **H1**—C. Adloff *et al.*, Eur. Phys. J. **C21**, 33 (2001); C. Adloff *et al.*, Eur. Phys. J. (accepted for publication) hep-ex/0304003; **ZEUS**—S. Chekanov *et al.*, Eur. Phys. J. **C21**, 443 (2001); **BCDMS**—A.C. Benvenuti *et al.*, Phys. Lett. **B223**, 485 (1989) (as given in [54]) ; **E665**—M.R. Adams *et al.*, Phys. Rev. **D54**, 3006 (1996); **NMC**—M. Arneodo *et al.*, Nucl. Phys. **B483**, 3 (97); **SLAC**—L.W. Whitlow *et al.*, Phys. Lett. **B282**, 475 (1992).

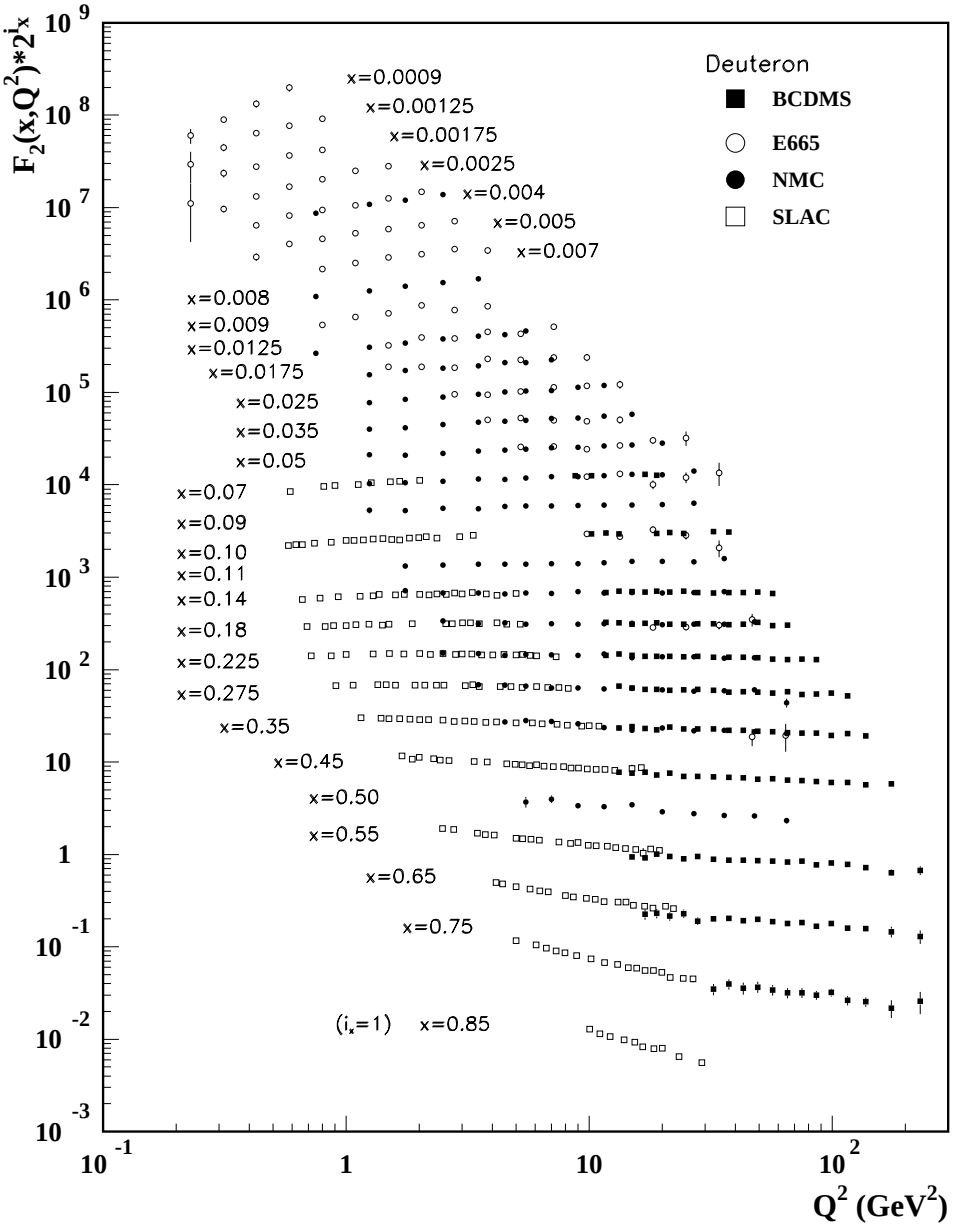


Figure 16.7: The deuteron structure function F_2^d measured in electromagnetic scattering of electrons (SLAC) and muons (BCDMS, E665, NMC) on a fixed target, shown as a function of Q^2 for bins of fixed x. Statistical and systematic errors added in quadrature are shown. For the purpose of plotting, F_2^p has been multiplied by 2^{ix}, where i_x is the number of the x bin, ranging from 1 ($x = 0.85$) to 29 ($x = 0.0009$). References: **BCDMS**—A.C. Benvenuti *et al.*, Phys. Lett. **B237**, 592 (1990). **E665, NMC, SLAC**—same references as Fig. 16.6.

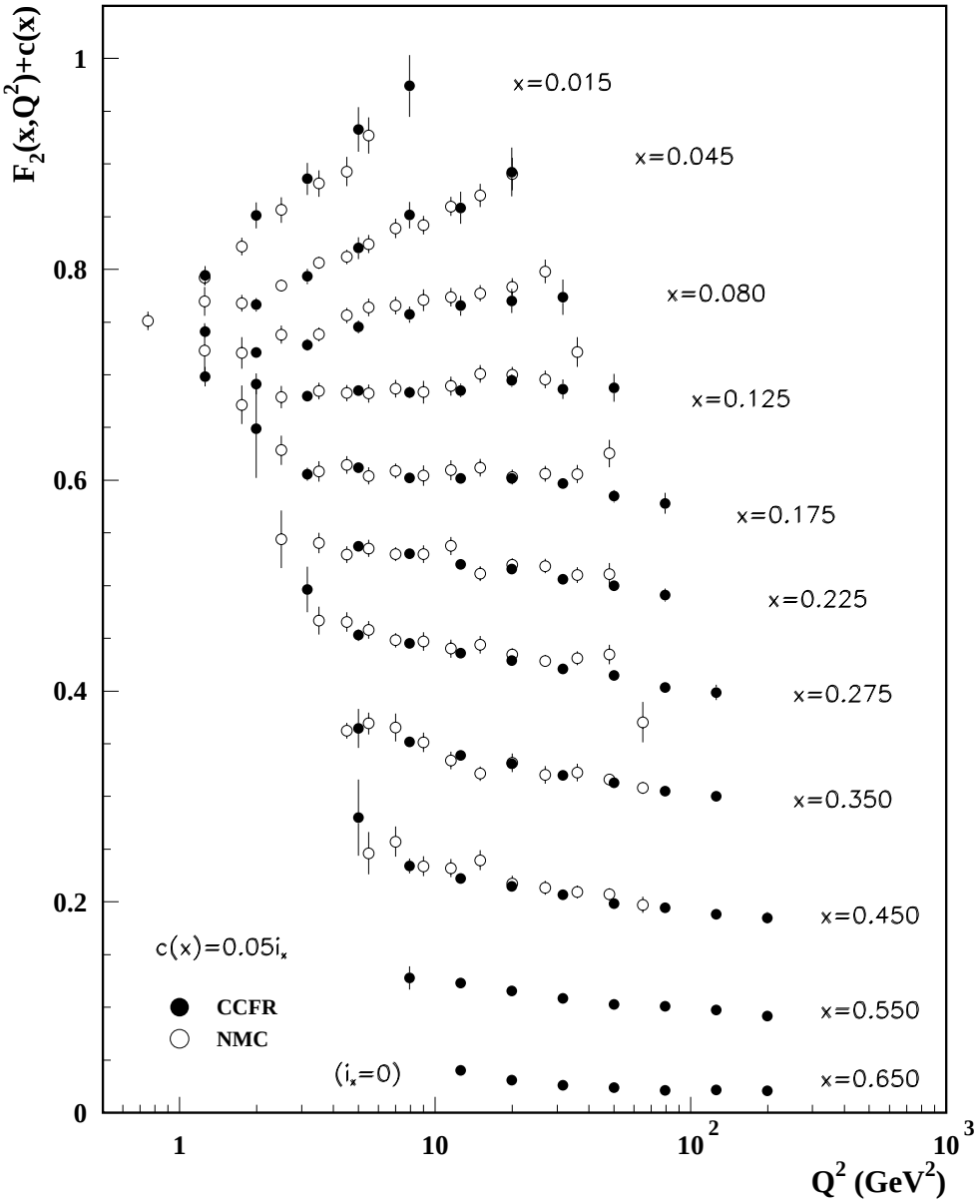


Figure 16.8: The deuteron structure function F_2 measured in deep inelastic scattering of muons on a fixed target (NMC) is compared to the structure function F_2 from neutrino-iron scattering (CCFR) using $F_2^\mu = (5/18)F_2^\nu - x(s+\overline{s})/6$, where heavy target effects have been taken into account. The data are shown versus Q^2, for bins of fixed x. The NMC data have been rebinned to CCFR x values. Statistical and systematic errors added in quadrature are shown. For the purpose of plotting, a constant $c(x) = 0.05i_x$ is added to F_2 where i_x is the number of the x bin, ranging from 0 ($x = 0.65$) to 10 ($x = 0.015$). References: **NMC**—M. Arneodo *et al.*, Nucl. Phys. **B483**, 3 (97); **CCFR/NuTeV**—U.K. Yang *et al.*, Phys. Rev. Lett. **86**, 2741 (2001).

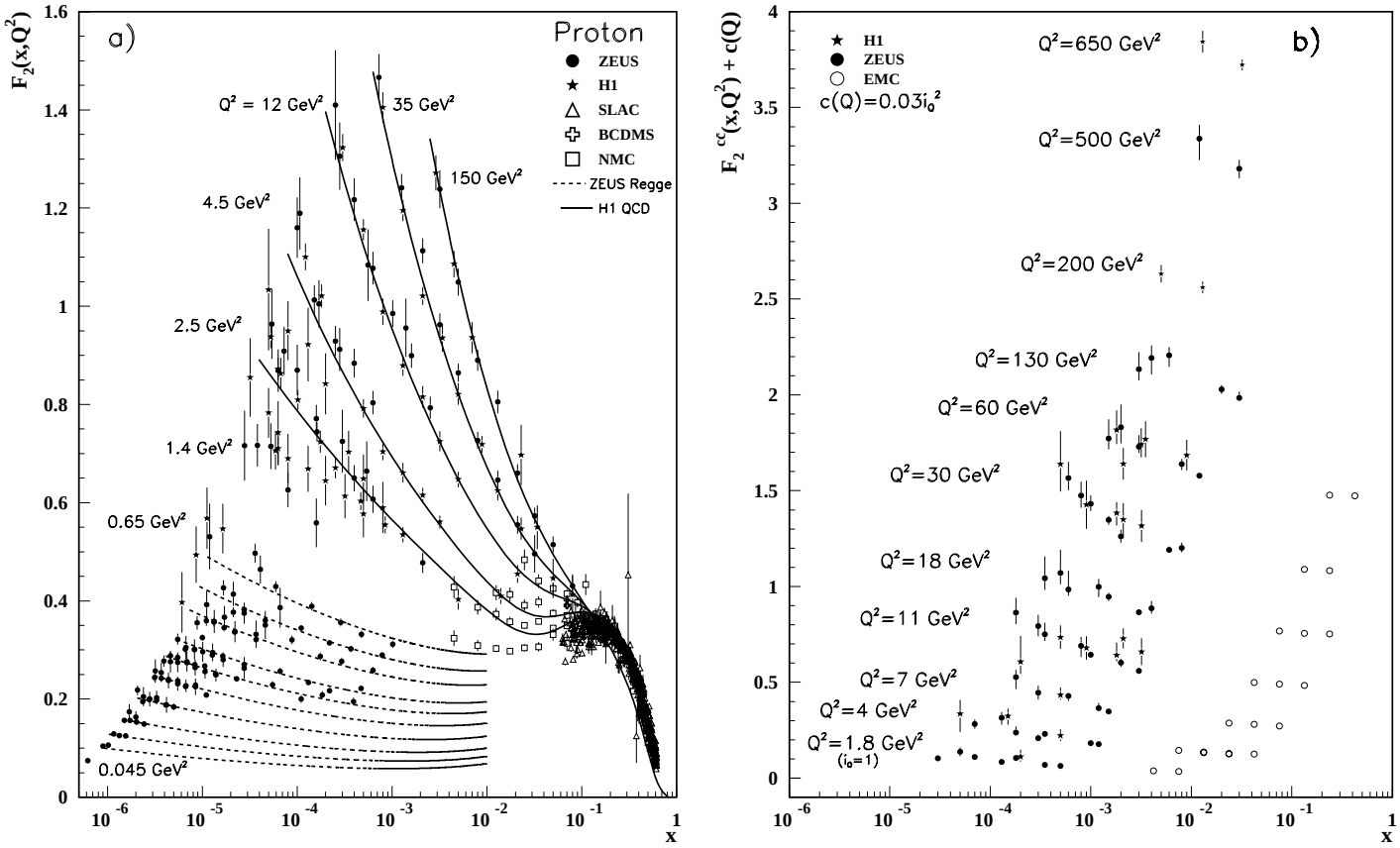


Figure 16.9: a) The proton structure function F_2^p mostly at small x and Q^2, measured in electromagnetic scattering of positrons (H1, ZEUS), electrons (SLAC), and muons (BCDMS, NMC) on protons. Lines are ZEUS and H1 parameterizations for lower (Regge) and higher (QCD) Q^2. Some points are only within 10% of the stated Q^2. Some points have been slightly offset in x for clarity. References: **ZEUS**—J. Breitweg *et al.*, Phys. Lett. **B407**, 432 (1997); J. Breitweg *et al.*, Eur. Phys. J. **C7**, 609 (1999); J. Breitweg *et al.*, Phys. Lett. **B487**, 53 (2000) (both data and ZEUS Regge parameterization); S. Chekanov *et al.*, Eur. Phys. J. **C21**, 443 (2001); **H1**—C. Adloff *et al.*, Nucl. Phys. **B497**, 3 (1997); C. Adloff *et al.*, Eur. Phys. J. **C21**, 33 (2001) (both data and H1 QCD parameterization) ; **BCDMS, NMC, SLAC**—same references as Fig. 16.6.

b) The charm structure function $F_2^{c\overline{c}}(x)$, i.e. that part of the inclusive structure function F_2^p arising from the production of charm quarks, measured in electromagnetic scattering of positrons on protons (H1, ZEUS) and muons on iron (EMC). The H1 points have been slightly offset in x for clarity. For the purpose of plotting, a constant $c(Q) = 0.03i_Q^2$ is added to $F_2^{c\overline{c}}$ where i_Q is the number of the Q^2 bin, ranging from 1 ($Q^2 = 1.8$ GeV^2) to 8 ($Q^2 = 130$ GeV^2). References: **ZEUS**—J. Breitweg *et al.*, Eur. Phys. J. **C12**, 35 (2000); **H1**—C. Adloff *et al.*, Z. Phys. **C72**, 593 (1996); C. Adloff *et al.*, Phys. Lett. **B528**, 199 (2002); **EMC**—J.J. Aubert *et al.*, Nucl. Phys. **B213**, 31 (1983).

Statistical and systematic errors added in quadrature are shown for both plots. The data are given as a function of x in bins of Q^2.

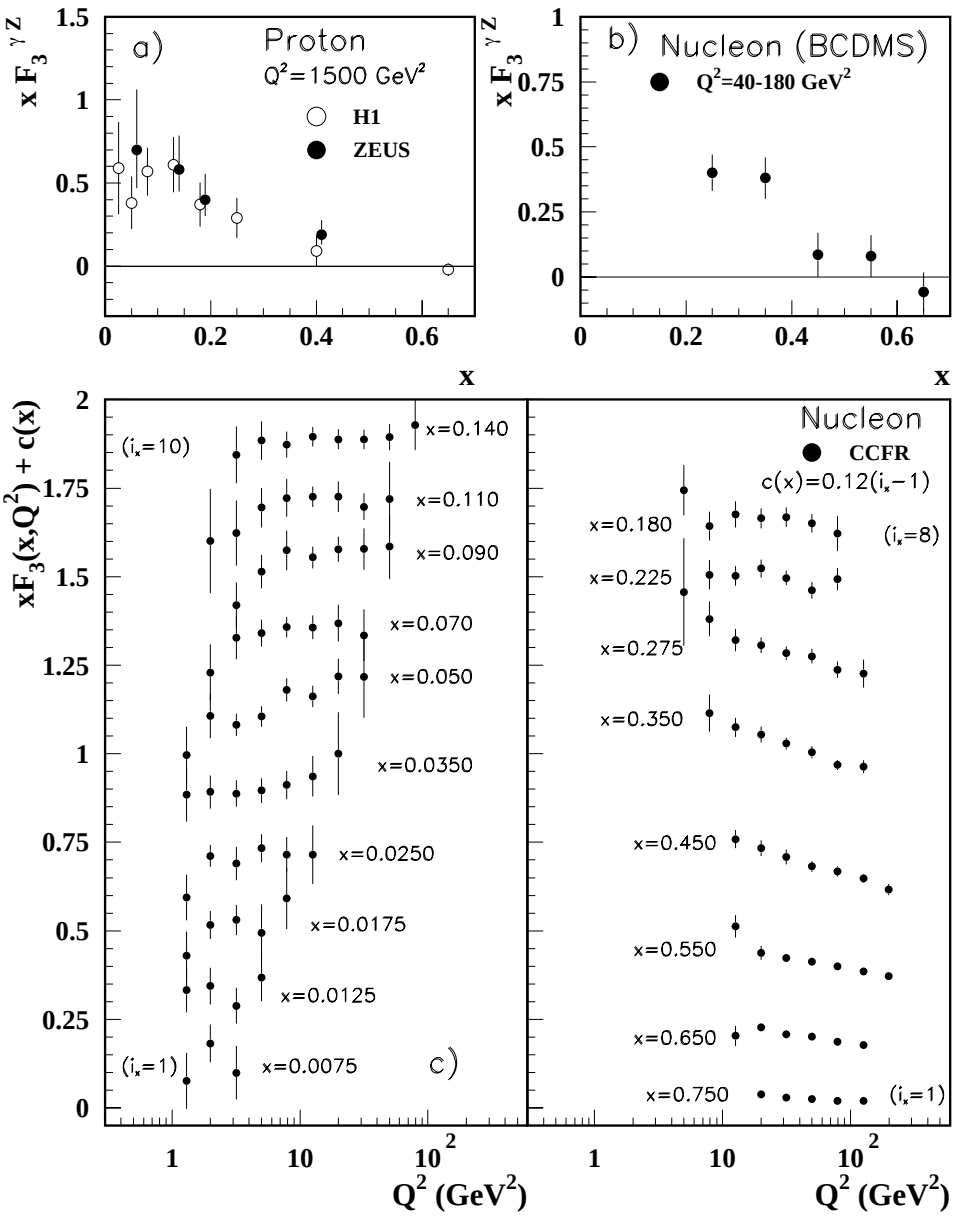


Figure 16.10: The structure function $xF_3^{\gamma Z}$ measured in electroweak scattering of **a)** electrons on protons (H1 and ZEUS) and **b)** muons on carbon (BCDMS). The ZEUS points have been slightly offset in x for clarity. References: **H1**—C. Adloff *et al.*, Eur. Phys. J. (accepted for publication) hep-ex/0304003; **ZEUS**–S. Chekanov *et al.*, Eur. Phys. J. **C28**, 175 (2003); **BCDMS**—A. Argento *et al.*, Phys. Lett. **B140**, 142 (1984).

c) The structure function xF_3 of the nucleon measured in ν-Fe scattering. The data are plotted as a function of Q^2 in bins of fixed x. For the purpose of plotting, a constant $c(x) = 0.12(i_x - 1)$ is added to xF_3, where i_x is the number of the x bin as shown in the plot. References: **CCFR**–W.G. Seligman *et al.*, Phys. Rev. Lett. **79**, 1213 (1997).

Statistical and systematic errors added in quadrature are shown for all plots.

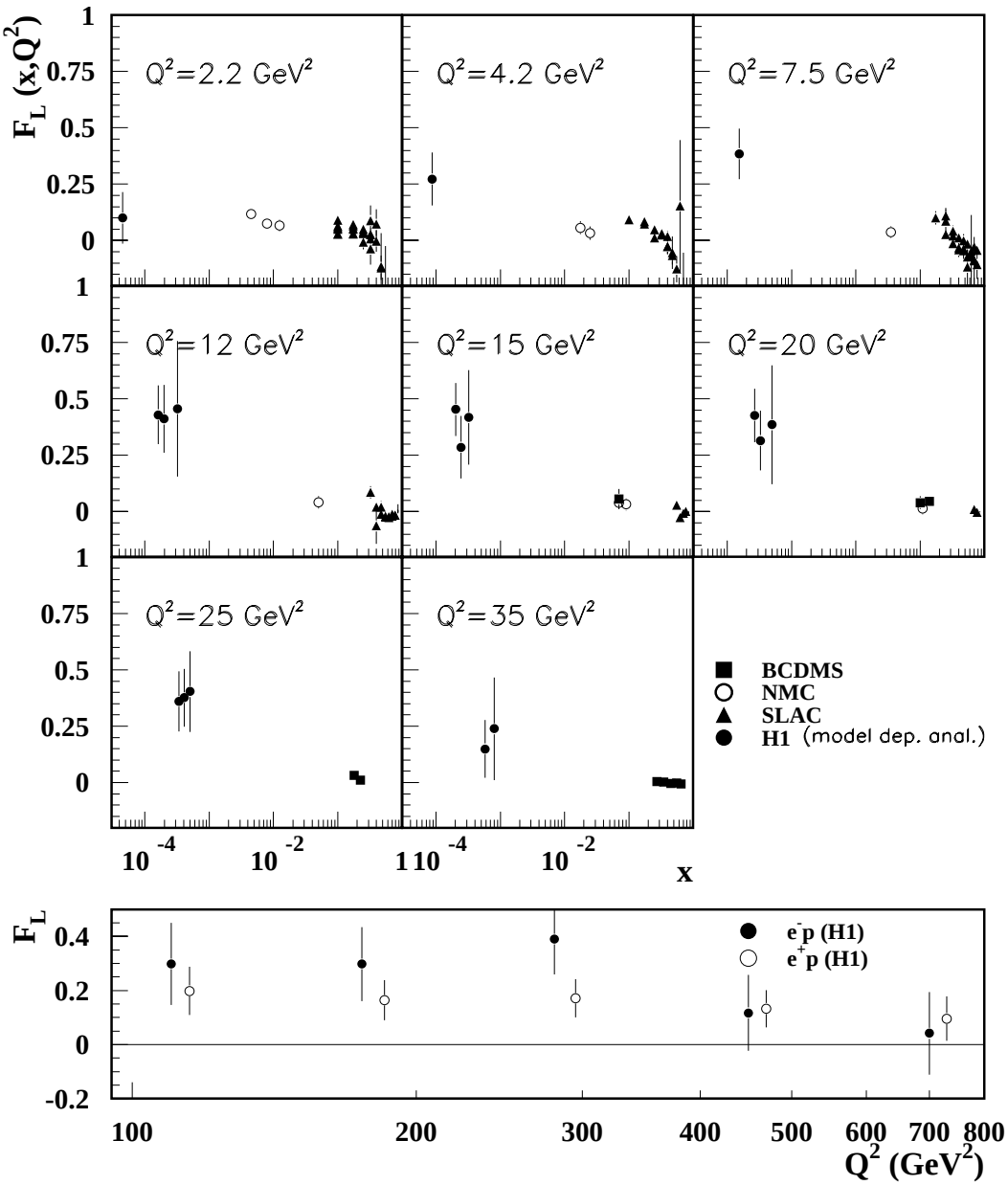


Figure 16.11: Top panel: The longitudinal structure function F_L as a function of x in bins of fixed Q^2 measured on the proton (except for the SLAC data which also contain deuterium data). BCDMS, NMC, and SLAC results are from measurements of R (the ratio of longitudinal to transverse photon absorption cross sections) which are converted to F_L by using the BDCMS parameterization of F_2 (A.C. Benvenuti *et al.*, Phys. Lett. **B223**, 485 (1989)). It is assumed that the Q^2 dependence of the fix-target data is small within a given Q^2 bin. References: **H1**—C. Adloff *et al.*, Eur. Phys. J. **C21**, 33 (2001); **BCDMS**—A. Benvenuti *et al.*, Phys. Lett. **B223**, 485 (1989); **NMC**—M. Arneodo *et al.*, Nucl. Phys. **B483**, 3 (1997); **SLAC**—L.W. Whitlow *et al.*, Phys. Lett. **B250**, 193 (1990) and numerical values from the thesis of L.W. Whitlow (SLAC-357).

Bottom panel: Higher Q^2 values of the longitudinal structure function F_L as a function of Q^2 given at the measured x for e^+/e^--proton scattering. Points have been slightly offset in Q^2 for clarity. References: **H1**—C. Adloff *et al.*, Eur. Phys. J. (accepted for publication) hep-ex/0304003.

The H1 results shown in both plots require the assumption of the validity of the QCD form for the F_2 structure function in order to extract F_L. Statistical and systematic errors added in quadrature are shown for both plots.

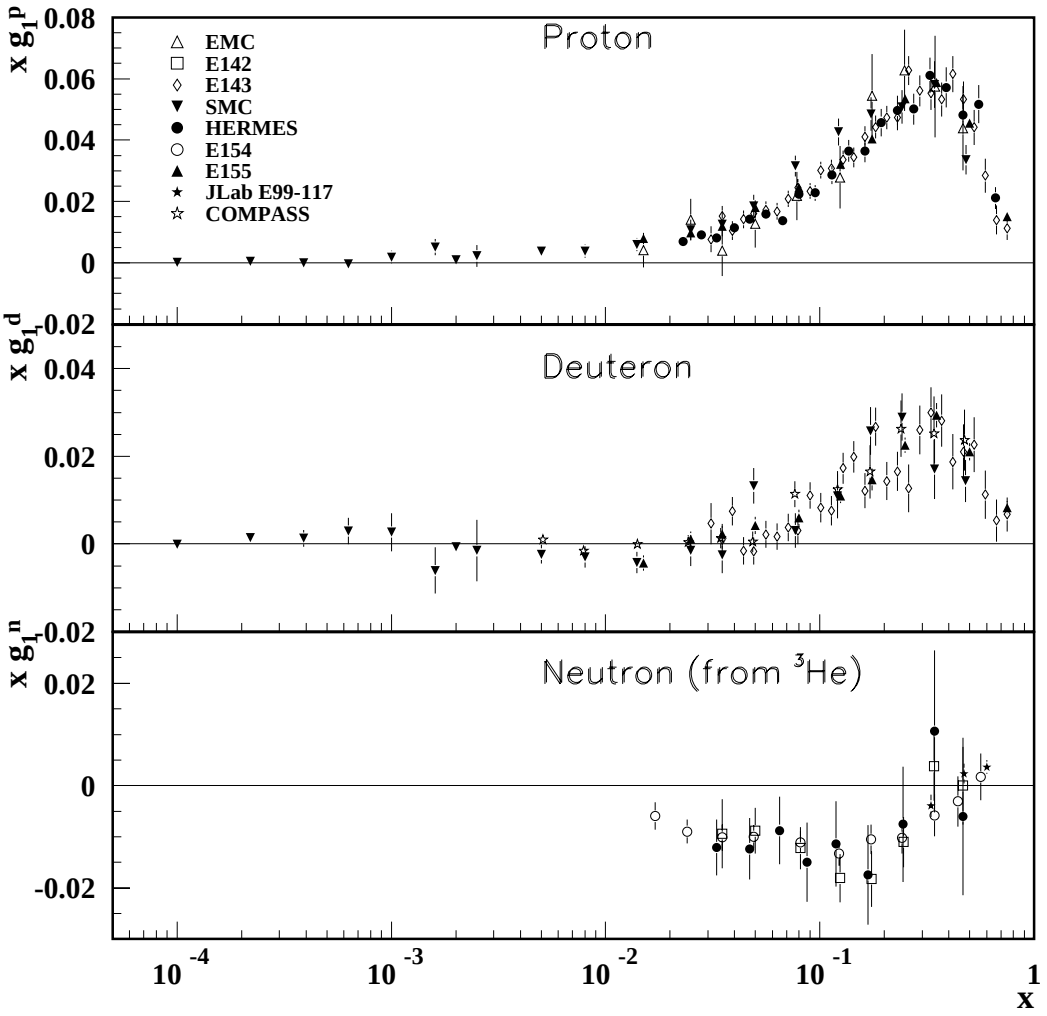


Figure 16.12: The spin-dependent structure function $xg_1(x)$ of the proton, deuteron, and neutron (from ^{3}He target) measured in deep inelastic scattering of polarized electrons/positrons: E142 ($Q^2 \sim 0.3-10$ GeV^2), E143 ($Q^2 \sim 0.3-10$ GeV^2), E154 ($Q^2 \sim 1-17$ GeV^2), E155 ($Q^2 \sim 1-40$ GeV^2), HERMES ($Q^2 \sim 0.8-20$ GeV^2) and muons: EMC ($Q^2 \sim 1.5-100$ GeV^2), SMC ($Q^2 \sim 0.01-100$ GeV^2), shown at the measured Q^2 (except for EMC data given $Q^2 = 10.7$ GeV^2 and E155 data given at $Q^2 = 5$ GeV^2). Note that $g_1^n(x)$ may also be extracted by taking the difference between $g_1^d(x)$ and $g_1^p(x)$, but these values have been omitted in the bottom plot for clarity. Statistical and systematic errors added in quadrature are shown. References: **EMC**—J. Ashman *et al.*, Nucl. Phys. **B328**, 1 (1989); **E142**—P.L. Anthony *et al.*, Phys. Rev. **D54**, 6620 (1996); **E143**—K. Abe *et al.*, Phys. Rev. **D58**, 112003 (1998); **SMC**—B. Adeva *et al.*, Phys. Rev. **D58**, 112001 (1998), B. Adeva *et al.*, Phys. Rev. **D60**, 072004 (1999) and Erratum-Phys. Rev. **D62**, 079902 (2000); **HERMES**—A. Airapetian *et al.*, Phys. Lett. **B442**, 484 (1998) and K. Ackerstaff *et al.*, Phys. Lett. **B404**, 383 (1997); **E154**—K. Abe *et al.*, Phys. Rev. Lett. **79**, 26 (1997); **E155**—P.L. Anthony *et al.*, Phys. Lett. **B463**, 339 (1999) and P.L. Anthony *et al.*, Phys. Lett. **B493**, 19 (2000).

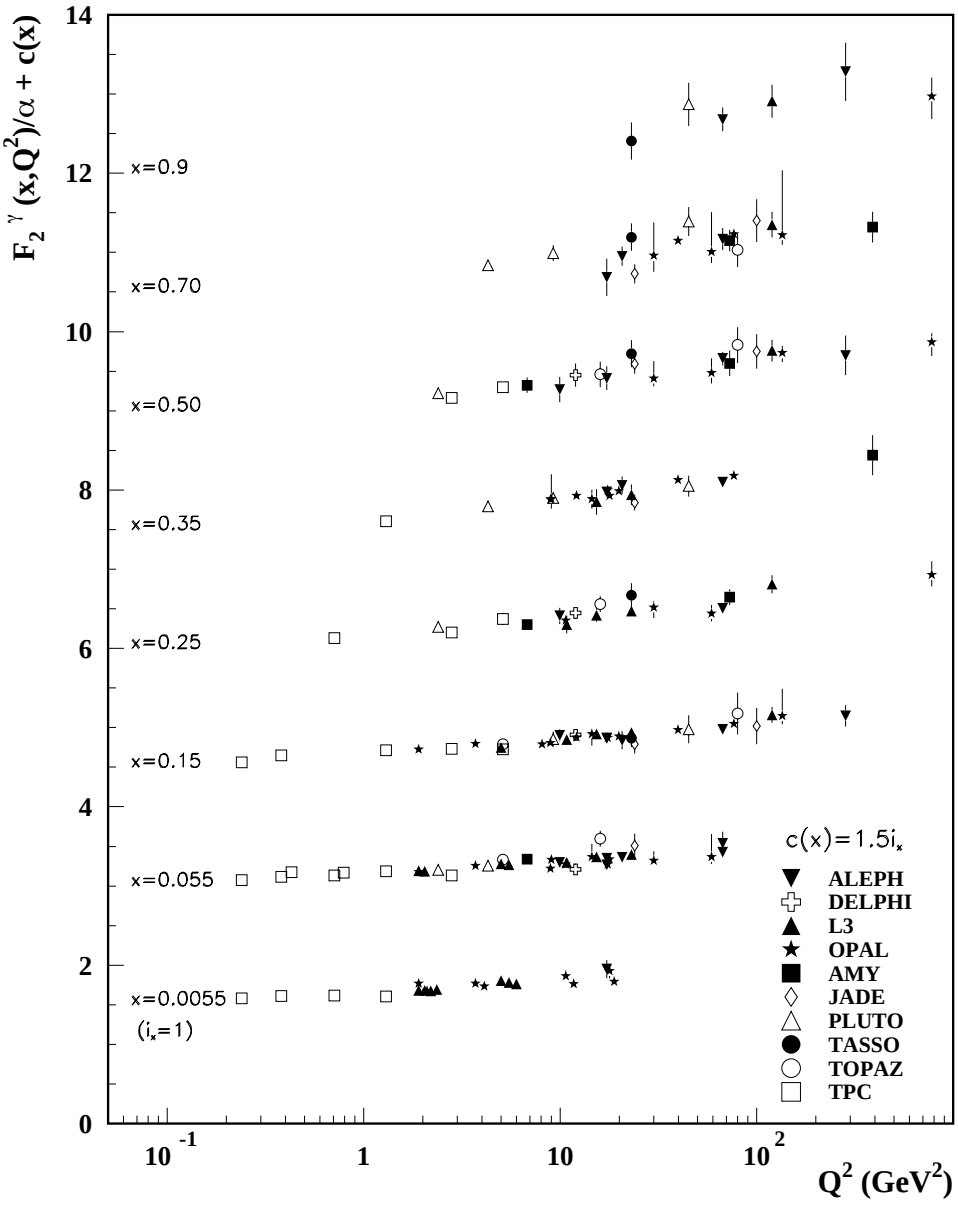


Figure 16.13: The hadronic structure function of the photon F_2^γ divided by the fine structure constant α measured in e^+e^- scattering, shown as a function of Q^2 for bins of x. Data points have been shifted to the nearest corresponding x bin as given in the plot. Some points have been offset in Q^2 for clarity. Statistical and systematic errors added in quadrature are shown. For the purpose of plotting, a constant $c(x) = 1.5i_x$ is added to F_2^γ/α where i_x is the number of the x bin, ranging from 1 ($x = 0.0055$) to 8 ($x = 0.9$). References: **ALEPH**–R. Barate *et al.*, Phys. Lett. **B458**, 152 (1999); **DELPHI**–P. Abreu *et al.*, Z. Phys. **C69**, 223 (1995); **L3**–M. Acciarri *et al.*, Phys. Lett. **B436**, 403 (1998); M. Acciarri *et al.*, Phys. Lett. **B447**, 147 (1999); M. Acciarri *et al.*, Phys. Lett. **B483**, 373 (2000); **OPAL**–A. Ackerstaff *et al.*, Phys. Lett. **B411**, 387 (1997); A. Ackerstaff *et al.*, Z. Phys. **C74**, 33 (1997); G. Abbiendi *et al.*, Eur. Phys. J. **C18**, 15 (2000); G. Abbiendi *et al.*, Phys. Lett. **B533**, 207 (2002) (note that there is some non-trivial statistical correlation between these last two papers); **AMY**–S.K. Sahu *et al.*, Phys. Lett. **B346**, 208 (1995); T. Kojima *et al.*, Phys. Lett. **B400**, 395 (1997); **JADE**–W. Bartel *et al.*, Z. Phys. **C24**, 231 (1984); **PLUTO**–C. Berger *et al.*, Phys. Lett. **142B**, 111 (1984); C. Berger *et al.*, Nucl. Phys. **B281**, 365 (1987); **TASSO**–M. Althoff *et al.*, Z. Phys. **C31**, 527 (1986); **TOPAZ**–K. Muramatsu *et al.*, Phys. Lett. **B332**, 477 (1994); **TPC/Two Gamma**–H. Aihara *et al.*, Z. Phys. **C34**, 1 (1987).

16. STRUCTURE FUNCTIONS

Written Summer 2001 by B. Foster (University of Oxford), A.D. Martin (University of Durham), M.G. Vincter (Carleton University). Updated Summer 2005.

16.1. Deep inelastic scattering

High energy lepton-nucleon scattering (deep inelastic scattering) plays a key role in determining the partonic structure of the proton. The process $\ell N \to \ell' X$ is illustrated in Fig. 16.1. The filled circle in this figure represents the internal structure of the proton which can be expressed in terms of structure functions.

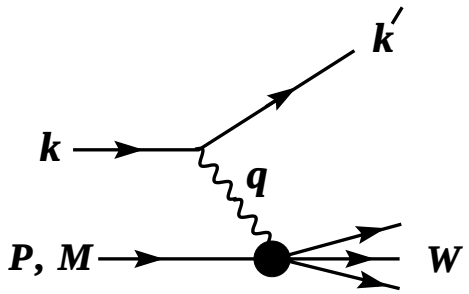


Figure 16.1: Kinematic quantities for the description of deep inelastic scattering. The quantities k and k' are the four-momenta of the incoming and outgoing leptons, P is the four-momentum of a nucleon with mass M, and W is the mass of the recoiling system X. The exchanged particle is a γ, $W^{\pm}$, or Z; it transfers four-momentum $q = k - k'$ to the nucleon.

Invariant quantities:

$\nu = \dfrac{q \cdot P}{M} = E - E'$ is the lepton's energy loss in the nucleon rest frame (in earlier literature sometimes $\nu = q \cdot P$). Here, E and E' are the initial and final lepton energies in the nucleon rest frame.

$Q^2 = -q^2 = 2(EE' - \vec{k} \cdot \vec{k}') - m_\ell^2 - m_{\ell'}^2$ where $m_\ell(m_{\ell'})$ is the initial (final) lepton mass. If $EE' \sin^2(\theta/2) \gg m_\ell^2,\ m_{\ell'}^2$, then

$\approx 4EE' \sin^2(\theta/2)$, where θ is the lepton's scattering angle with respect to the lepton beam direction.

$x = \dfrac{Q^2}{2M\nu}$ where, in the parton model, x is the fraction of the nucleon's momentum carried by the struck quark.

$y = \dfrac{q \cdot P}{k \cdot P} = \dfrac{\nu}{E}$ is the fraction of the lepton's energy lost in the nucleon rest frame.

$W^2 = (P+q)^2 = M^2 + 2M\nu - Q^2$ is the mass squared of the system X recoiling against the scattered lepton.

$s = (k+P)^2 = \dfrac{Q^2}{xy} + M^2 + m_\ell^2$ is the center-of-mass energy squared of the lepton-nucleon system.

The process in Fig. 16.1 is called deep ($Q^2 >> M^2$) inelastic ($W^2 >> M^2$) scattering (DIS). In what follows, the masses of the initial and scattered leptons, m_ℓ and m'_ℓ, are neglected.

16.1.1. ***DIS cross sections*** :

$$\frac{d^2\sigma}{dx\,dy} = x\,(s - M^2)\,\frac{d^2\sigma}{dx\,dQ^2} = \frac{2\pi\,M\nu}{E'}\frac{d^2\sigma}{d\Omega_{\rm Nrest}\,dE'}\ . \tag{16.1}$$

In lowest-order perturbation theory, the cross section for the scattering of polarized leptons on polarized nucleons can be expressed in terms of the products of leptonic and hadronic tensors associated with the coupling of the exchanged bosons at the upper and lower vertices in Fig. 16.1 (see Refs. 1–4)

$$\frac{d^2\sigma}{dxdy} \;=\; \frac{2\pi y\alpha^2}{Q^4}\,\sum_j\,\eta_j\,L_j^{\mu\nu}\,W^j_{\mu\nu}\ . \tag{16.2}$$

For neutral-current processes, the summation is over $j = \gamma, Z$ and γZ representing photon and Z exchange and the interference between them, whereas for charged-current interactions there is only W exchange, $j = W$. (For transverse nucleon polarization, there is a dependence on the azimuthal angle of the scattered lepton.) $L_{\mu\nu}$ is the lepton tensor associated with the coupling of the exchange boson to the leptons. For incoming leptons of charge $e = \pm 1$ and helicity $\lambda = \pm 1$,

$$\begin{aligned}
L^\gamma_{\mu\nu} =& 2\left(k_\mu k'_\nu + k'_\mu k_\nu - k\cdot k' g_{\mu\nu} - i\lambda\varepsilon_{\mu\nu\alpha\beta}k^\alpha k'^\beta\right),\\
L^{\gamma Z}_{\mu\nu} =& (g^e_V + e\lambda g^e_A)\,L^\gamma_{\mu\nu}, \qquad L^Z_{\mu\nu} = (g^e_V + e\lambda g^e_A)^2\,L^\gamma_{\mu\nu},\\
L^W_{\mu\nu} =& (1 + e\lambda)^2\,L^\gamma_{\mu\nu},
\end{aligned} \tag{16.3}$$

where $g^e_V = -\frac{1}{2} \,+\, 2\sin^2\theta_W, \qquad g^e_A \,=\, -\frac{1}{2}$.

Although here the helicity formalism is adopted, an alternative approach is to express the tensors in Eq. (16.3) in terms of the polarization of the lepton.

The factors η_j in Eq. (16.2) denote the ratios of the corresponding propagators and couplings to the photon propagator and coupling squared

$$\begin{aligned}
\eta_\gamma \;=\; 1 \quad ;\quad \eta_{\gamma Z} \;=\; & \left(\frac{G_F M_Z^2}{2\sqrt{2}\pi\alpha}\right)\,\left(\frac{Q^2}{Q^2 + M_Z^2}\right);\\
\eta_Z \;=\; \eta^2_{\gamma Z} \quad ;\quad \eta_W \;=\; & \tfrac{1}{2}\left(\frac{G_F M_W^2}{4\pi\alpha}\,\frac{Q^2}{Q^2 + M_W^2}\right)^2 .
\end{aligned} \tag{16.4}$$

The hadronic tensor, which describes the interaction of the appropriate electroweak currents with the target nucleon, is given by

$$W_{\mu\nu} \;=\; \frac{1}{4\pi}\,\int\,d^4z\,e^{iq\cdot z}\,\left\langle P,S\left|\left[J^\dagger_\mu(z), J_\nu(0)\right]\right|\,P,S\right\rangle, \tag{16.5}$$

where S denotes the nucleon-spin 4-vector, with $S^2 = -M^2$ and $S\cdot P = 0$.

16.2. Structure functions of the proton

The structure functions are defined in terms of the hadronic tensor (see Refs. 1–3)

$$\begin{aligned} W_{\mu\nu} = & \left(-g_{\mu\nu} \; + \; \frac{q_\mu q_\nu}{q^2}\right) \; F_1(x,Q^2) + \frac{\hat{P}_\mu \hat{P}_\nu}{P\cdot q} \; F_2(x,Q^2) \\ & - \; i\varepsilon_{\mu\nu\alpha\beta} \; \frac{q^\alpha P^\beta}{2P\cdot q} \; F_3(x,Q^2) \\ & + \; i\varepsilon_{\mu\nu\alpha\beta} \; \frac{q^\alpha}{P\cdot q} \; \left[S^\beta g_1(x,Q^2) + \left(S^\beta \; - \; \frac{S\cdot q}{P\cdot q} \; P^\beta\right) \; g_2(x,Q^2)\right] \\ & + \; \frac{1}{P\cdot q} \; \left[\tfrac{1}{2}\left(\hat{P}_\mu \hat{S}_\nu \; + \; \hat{S}_\mu \hat{P}_\nu\right) - \frac{S\cdot q}{P\cdot q} \; \hat{P}_\mu \hat{P}_\nu\right] \; g_3(x,Q^2) \\ & + \; \frac{S\cdot q}{P\cdot q} \; \left[\frac{\hat{P}_\mu \hat{P}_\nu}{P\cdot q} \; g_4(x,Q^2) + \left(-g_{\mu\nu} \; + \; \frac{q_\mu q_\nu}{q^2}\right) \; g_5(x,Q^2)\right] \end{aligned} \tag{16.6}$$

where

$$\hat{P}_\mu \; = \; P_\mu \; - \; \frac{P\cdot q}{q^2} \; q_\mu, \qquad \hat{S}_\mu \; = \; S_\mu \; - \; \frac{S\cdot q}{q^2} \; q_\mu \; . \tag{16.7}$$

In Ref. [2], the definition of $W_{\mu\nu}$ with $\mu \leftrightarrow \nu$ is adopted, which changes the sign of the $\varepsilon_{\mu\nu\alpha\beta}$ terms in Eq. (16.6), although the formulae given here below are unchanged. Ref. [1] tabulates the relation between the structure functions defined in Eq. (16.6) and other choices available in the literature.

The cross sections for neutral- and charged-current deep inelastic scattering on unpolarized nucleons can be written in terms of the structure functions in the generic form

$$\begin{aligned} \frac{d^2\sigma^i}{dxdy} \; = \; & \frac{4\pi\alpha^2}{xyQ^2} \; \eta^i \left\{\left(1 \; - \; y \; - \; \frac{x^2y^2M^2}{Q^2}\right) \; F_2^i \right. \\ & \left. + \, y^2 x F_1^i \; \mp \; \left(y \; - \; \frac{y^2}{2}\right) \; xF_3^i\right\}, \end{aligned} \tag{16.8}$$

where i = NC, CC corresponds to neutral-current ($eN \to eX$) or charged-current ($eN \to \nu X$ or $\nu N \to eX$) processes, respectively. For incoming neutrinos, $L^W_{\mu\nu}$ of Eq. (16.3) is still true, but with e, λ corresponding to the outgoing charged lepton. In the last term of Eq. (16.8), the $-$ sign is taken for an incoming e^+ or $\overline{\nu}$ and the $+$ sign for an incoming e^- or ν. The factor $\eta^{\rm NC} = 1$ for unpolarized $e^\pm$ beams, whereas*

$$\eta^{\rm CC} \; = \; (1 \pm \lambda)^2 \eta_W \tag{16.9}$$

with $\pm$ for $\ell^\pm$; and where λ is the helicity of the incoming lepton and η_W is defined in Eq. (16.4); for incoming neutrinos $\eta^{\rm CC} = 4\eta_W$. The CC structure functions, which derive exclusively from W exchange, are

$$F_1^{\rm CC} = F_1^W, \; F_2^{\rm CC} = F_2^W, \; xF_3^{\rm CC} = xF_3^W . \tag{16.10}$$

The NC structure functions $F_2^\gamma, F_2^{\gamma Z}, F_2^Z$ are, for $e^\pm N \to e^\pm X$, given by Ref. [5],

$$F_2^{\rm NC} \;=\; F_2^\gamma \;-\; (g_V^e \pm \lambda g_A^e)\eta_{\gamma Z} F_2^{\gamma Z} \;+\; ({g_V^e}^2 + {g_A^e}^2 \pm 2\lambda g_V^e g_A^e)\,\eta_Z F_2^Z \tag{16.11}$$

and similarly for $F_1^{\rm NC}$, whereas

$$xF_3^{\rm NC} = \; -(g_A^e \pm \lambda g_V^e)\eta_{\gamma Z} x F_3^{\gamma Z} + [2g_V^e g_A^e \pm \lambda({g_V^e}^2 + {g_A^e}^2)]\eta_Z x F_3^Z \;. \tag{16.12}$$

The polarized cross-section difference

$$\Delta\sigma \;=\; \sigma(\lambda_n = -1, \lambda_\ell) \;-\; \sigma(\lambda_n = 1, \lambda_\ell) \;, \tag{16.13}$$

where λ_ℓ, λ_n are the helicities (± 1) of the incoming lepton and nucleon, respectively, may be expressed in terms of the five structure functions $g_{1,\ldots 5}(x, Q^2)$ of Eq. (16.6). Thus,

$$\begin{aligned}\frac{d^2\Delta\sigma^i}{dxdy} \;=\; & \frac{8\pi\alpha^2}{xyQ^2}\,\eta^i \left\{ -\lambda_\ell y \left(2 - y - 2x^2y^2\frac{M^2}{Q^2}\right)\, x g_1^i + \lambda_\ell 4x^3y^2\,\frac{M^2}{Q^2}\, g_2^i \right. \\ & +\; 2x^2 y \frac{M^2}{Q^2}\left(1 - y - x^2y^2\frac{M^2}{Q^2}\right)\, g_3^i \\ & \left. -\; \left(1 + 2x^2y\frac{M^2}{Q^2}\right)\left[\left(1 - y - x^2y^2\frac{M^2}{Q^2}\right)\, g_4^i + xy^2 g_5^i\right]\right\}\end{aligned} \tag{16.14}$$

with $i =$ NC or CC as before. The Eq. (16.13) corresponds to the difference of antiparallel minus parallel spins of the incoming particles for e^- or ν initiated reactions, but parallel minus antiparallel for e^+ or $\overline{\nu}$ initiated processes. For longitudinal nucleon polarization, the contributions of g_2 and g_3 are suppressed by powers of M^2/Q^2. These structure functions give an unsuppressed contribution to the cross section for transverse polarization [1], but in this case the cross-section difference vanishes as $M/Q \to 0$.

Because the same tensor structure occurs in the spin-dependent and spin-independent parts of the hadronic tensor of Eq. (16.6) in the $M^2/Q^2 \to 0$ limit, the differential cross-section difference of Eq. (16.14) may be obtained from the differential cross section Eq. (16.8) by replacing

$$F_1 \;\to\; -g_5 \;, \quad F_2 \;\to\; -g_4 \;, \quad F_3 \;\to\; 2g_1 \;, \tag{16.15}$$

and multiplying by two, since the total cross section is the average over the initial-state polarizations. In this limit, Eq. (16.8) and Eq. (16.14) may be written in the form

$$\begin{aligned}\frac{d^2\sigma^i}{dxdy} &= \frac{2\pi\alpha^2}{xyQ^2}\,\eta^i\left[Y_+ F_2^i \mp Y_- x F_3^i - y^2 F_L^i\right], \\ \frac{d^2\Delta\sigma^i}{dxdy} &= \frac{4\pi\alpha^2}{xyQ^2}\,\eta^i\left[-Y_+ g_4^i \mp Y_- 2x g_1^i + y^2 g_L^i\right],\end{aligned} \tag{16.16}$$

with $i =$ NC or CC, where $Y_\pm = 1 \pm (1-y)^2$ and

$$F_L^i \;=\; F_2^i \;-\; 2xF_1^i \;, \qquad g_L^i \;=\; g_4^i \;-\; 2xg_5^i \;. \tag{16.17}$$

In the naive quark-parton model, the analogy with the Callan-Gross relations [6] $F_L^i = 0$, are the Dicus relations [7] $g_L^i = 0$. Therefore, there are only two independent polarized structure functions: g_1 (parity conserving) and g_5 (parity violating), in analogy with the unpolarized structure functions F_1 and F_3.

16.2.1. ***Structure functions in the quark-parton model*** :

In the quark-parton model [8,9], contributions to the structure functions F^i and g^i can be expressed in terms of the quark distribution functions $q(x, Q^2)$ of the proton, where $q = u, \overline{u}, d, \overline{d}$ etc. The quantity $q(x, Q^2)dx$ is the number of quarks (or antiquarks) of designated flavor that carry a momentum fraction between x and $x + dx$ of the proton's momentum in a frame in which the proton momentum is large.

For the neutral-current processes $ep \to eX$,

$$\begin{aligned}
\left[F_2^\gamma,\ F_2^{\gamma Z},\ F_2^Z\right] &= x\sum_q \left[e_q^2,\ 2e_q g_V^q,\ {g_V^q}^2 + {g_A^q}^2\right]\ (q + \overline{q})\ , \\
\left[F_3^\gamma,\ F_3^{\gamma Z},\ F_3^Z\right] &= \sum_q \left[0,\ 2e_q g_A^q,\ 2g_V^q g_A^q\right]\ (q - \overline{q})\ , \\
\left[g_1^\gamma,\ g_1^{\gamma Z},\ g_1^Z\right] &= \tfrac{1}{2}\sum_q \left[e_q^2,\ 2e_q g_V^q,\ {g_V^q}^2 + {g_A^q}^2\right]\ (\Delta q + \Delta\overline{q})\ , \\
\left[g_5^\gamma,\ g_5^{\gamma Z},\ g_5^Z\right] &= \sum_q \left[0,\ e_q g_A^q,\ g_V^q g_A^q\right]\ (\Delta q - \Delta\overline{q})\ ,
\end{aligned} \tag{16.18}$$

where $g_V^q = \pm\frac{1}{2} - 2e_q \sin^2\theta_W$ and $g_A^q = \pm\frac{1}{2}$, with $\pm$ according to whether q is a $u-$ or $d-$type quark respectively. The quantity Δq is the difference $q\uparrow - q\downarrow$ of the distributions with the quark spin parallel and antiparallel to the proton spin.

For the charged-current processes $e^- p \to \nu X$ and $\overline{\nu} p \to e^+ X$, the structure functions are:

$$\begin{aligned}
F_2^{W^-} &= 2x(u + \overline{d} + \overline{s} + c\ldots)\ , \\
F_3^{W^-} &= 2(u - \overline{d} - \overline{s} + c\ldots)\ , \\
g_1^{W^-} &= (\Delta u + \Delta\overline{d} + \Delta\overline{s} + \Delta c\ldots)\ , \\
g_5^{W^-} &= (-\Delta u + \Delta\overline{d} + \Delta\overline{s} - \Delta c\ldots)\ ,
\end{aligned} \tag{16.19}$$

where only the active flavors are to be kept and where CKM mixing has been neglected. For $e^+ p \to \overline{\nu} X$ and $\nu p \to e^- X$, the structure functions F^{W^+}, g^{W^+} are obtained by the flavor interchanges $d \leftrightarrow u, s \leftrightarrow c$ in the expressions for F^{W^-}, g^{W^-}. The structure functions for scattering on a neutron are obtained from those of the proton by the interchange $u \leftrightarrow d$. For both the neutral- and charged-current processes, the quark-parton model predicts $2xF_1^i = F_2^i$ and $g_4^i = 2xg_5^i$.

Neglecting masses, the structure functions g_2 and g_3 contribute only to scattering from transversely polarized nucleons (for which $S \cdot q = 0$), and have no simple interpretation in terms of the quark-parton model. They arise from off-diagonal matrix elements $\langle P, \lambda'|[J_\mu^\dagger(z), J_\nu(0)]|P, \lambda\rangle$, where the proton helicities satisfy $\lambda' \neq \lambda$. In fact, the leading-twist contributions to both g_2 and g_3 are both twist-2 and twist-3, which contribute at the same order of Q^2. The Wandzura-Wilczek relation [10] expresses the twist-2 part of

g_2 in terms of g_1 as

$$g_2^i(x) \;=\; -g_1^i(x) \;+\; \int_x^1 \frac{dy}{y}\, g_1^i(y)\;. \tag{16.20}$$

However, the twist-3 component of g_2 is unknown. Similarly, there is a relation expressing the twist-2 part of g_3 in terms of g_4. A complete set of relations, including M^2/Q^2 effects, can be found in Ref. [11].

16.2.2. *Structure functions and QCD* :

One of the most striking predictions of the quark-parton model is that the structure functions F_i, g_i *scale*, *i.e.*, $F_i(x, Q^2) \to F_i(x)$ in the Bjorken limit that Q^2 and $\nu \to \infty$ with x fixed [12]. This property is related to the assumption that the transverse momentum of the partons in the infinite-momentum frame of the proton is small. In QCD, however, the radiation of hard gluons from the quarks violates this assumption, leading to logarithmic scaling violations, which are particularly large at small x, see Fig. 16.2. The radiation of gluons produces the evolution of the structure functions. As Q^2 increases, more and more gluons are radiated, which in turn split into $q\overline{q}$ pairs. This process leads both to the softening of the initial quark momentum distributions and to the growth of the gluon density and the $q\overline{q}$ sea as x decreases.

In QCD, the above process is described in terms of scale-dependent parton distributions $f_a(x, \mu^2)$, where $a = g$ or q and, typically, μ is the scale of the probe Q. For $Q^2 \gg M^2$, the structure functions are of the form

$$F_i = \sum_a C_i^a \otimes f_a, \tag{16.21}$$

where $\otimes$ denotes the convolution integral

$$C \;\otimes\; f \;=\; \int_x^1 \frac{dy}{y}\, C(y)\, f\left(\frac{x}{y}\right)\;, \tag{16.22}$$

and where the coefficient functions C_i^a are given as a power series in α_s. The parton distribution f_a corresponds, at a given x, to the density of parton a in the proton integrated over transverse momentum k_t up to μ. Its evolution in μ is described in QCD by a DGLAP equation (see Refs. 14–17) which has the schematic form

$$\frac{\partial f_a}{\partial \ln \mu^2} \sim \frac{\alpha_s(\mu^2)}{2\pi} \sum_b \left(P_{ab} \otimes f_b\right)\;, \tag{16.23}$$

where the P_{ab}, which describe the parton splitting $b \to a$, are also given as a power series in α_s. Although perturbative QCD can predict, via Eq. (16.23), the evolution of the parton distribution functions from a particular scale, μ_0, these DGLAP equations cannot predict them *a priori* at any particular μ_0. Thus they must be measured at a starting point μ_0 before the predictions of QCD can be compared to the data at other scales, μ. In general, all observables involving a hard hadronic interaction (such as structure

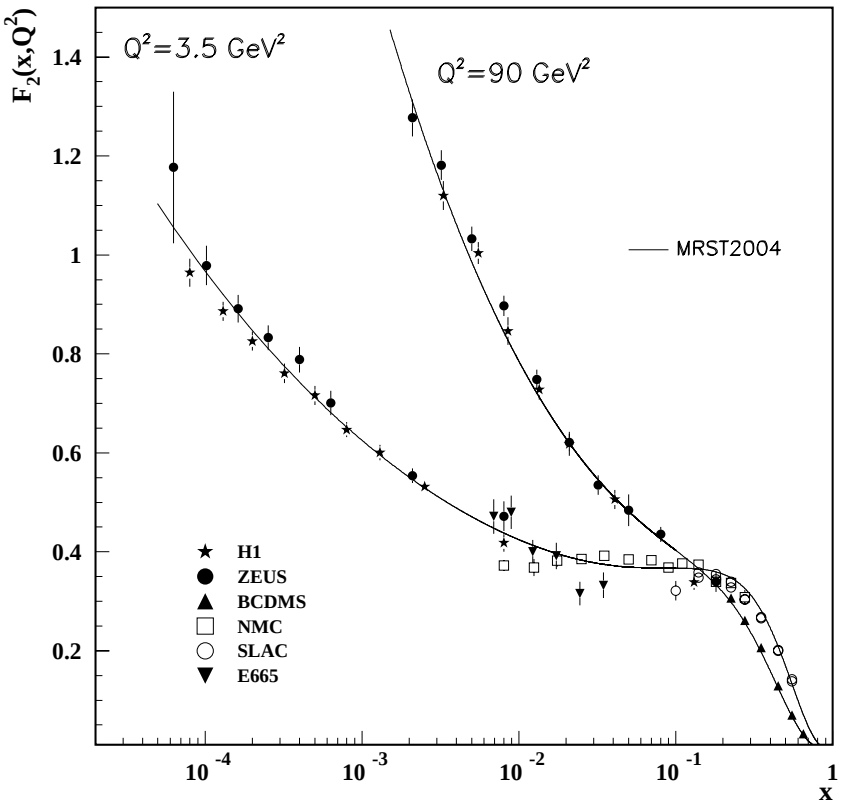


Figure 16.2: The proton structure function F_2^p given at two Q^2 values (3.5 GeV^2 and 90 GeV^2), which exhibit scaling at the 'pivot' point $x \sim 0.14$. See the caption in Fig. 16.6 for the references of the data. Also shown is the MRST2004 parameterization [13] given at the same scales.

functions) can be expressed as a convolution of calculable, process-dependent coefficient functions and these universal parton distributions, e.g. Eq. (16.21).

It is often convenient to write the evolution equations in terms of the gluon, non-singlet (q^{NS}) and singlet (q^S) quark distributions, such that

$$q^{NS} \;=\; q_i - \overline{q}_i \quad (\text{or } \; q_i - q_j), \qquad q^S \;=\; \sum_i (q_i + \overline{q}_i) \; . \tag{16.24}$$

The non-singlet distributions have non-zero values of flavor quantum numbers, such as isospin and baryon number. The DGLAP evolution equations then take the form

$$\frac{\partial q^{NS}}{\partial \ln \mu^2} = \frac{\alpha_s(\mu^2)}{2\pi} \; P_{qq} \; \otimes \; q^{NS} \; ,$$

$$\frac{\partial}{\partial \ln \mu^2} \begin{pmatrix} q^S \\ g \end{pmatrix} = \frac{\alpha_s(\mu^2)}{2\pi} \begin{pmatrix} P_{qq} & 2n_f \, P_{qg} \\ P_{gq} & P_{gg} \end{pmatrix} \otimes \begin{pmatrix} q^S \\ g \end{pmatrix} \; , \tag{16.25}$$

where P are splitting functions that describe the probability of a given parton splitting into two others, and n_f is the number of (active) quark flavors. The leading-order Altarelli-Parisi [16] splitting functions are

$$P_{qq} = \frac{4}{3}\left[\frac{1+x^2}{(1-x)}\right]_+ = \frac{4}{3}\left[\frac{1+x^2}{(1-x)_+}\right] + 2\delta(1-x)\ , \tag{16.26}$$

$$P_{qg} = \frac{1}{2}\left[x^2+(1-x)^2\right]\ , \tag{16.27}$$

$$P_{gq} = \frac{4}{3}\left[\frac{1+(1-x)^2}{x}\right]\ , \tag{16.28}$$

$$P_{gg} = 6\left[\frac{1-x}{x} + x(1-x) + \frac{x}{(1-x)_+}\right] + \left[\frac{11}{2} - \frac{n_f}{3}\right]\delta(1-x), \tag{16.29}$$

where the notation $[F(x)]_+$ defines a distribution such that for any sufficiently regular test function, $f(x)$,

$$\int_0^1 dx f(x)[F(x)]_+ = \int_0^1 dx\ (f(x)-f(1))F(x)\ . \tag{16.30}$$

In general, the splitting functions can be expressed as a power series in α_s. The series contains both terms proportional to $\ln\mu^2$ and to $\ln\ 1/x$. The leading-order DGLAP evolution sums up the $(\alpha_s \ln\mu^2)^n$ contributions, while at next-to-leading order (NLO) the sum over the $\alpha_s(\alpha_s\ln\mu^2)^{n-1}$ terms is included [18,19]. In fact, the NNLO contributions to the splitting functions and the DIS coefficient functions are now also all known [20,21,22].

In the kinematic region of very small x, it is essential to sum leading terms in $\ln\ 1/x$, independent of the value of $\ln\mu^2$. At leading order, LLx, this is done by the BFKL equation for the unintegrated distributions (see Refs. [23,24]) . The leading-order $(\alpha_s\ln(1/x))^n$ terms result in a power-like growth, $x^{-\omega}$ with $\omega = (12\alpha_s\ln 2)/\pi$, at asymptotic values of $\ln\ 1/x$. More recently, the next-to-leading $\ln\ 1/x$ (NLLx) contributions have become available [25,26]. They are so large (and negative) that the result appears to be perturbatively unstable. Methods, based on a combination of collinear and small x resummations, have been developed which reorganize the perturbative series into a more stable hierarchy [27,28,29]. These studies show that the asymptotic properties of the small x resummations are not significant at today's energies (which sample $x \gtrsim 10^{-4}$), and that in this domain NNLO DGLAP is a good approximation. Indeed, as yet, there is no firm evidence for any deviation from standard DGLAP evolution in the data for $Q^2 \gtrsim 2$ GeV^2. Nor is there any convincing indication that we have entered the 'non-linear' regime where the gluon density is so high that gluon-gluon recombination effects become significant.

The precision of the contemporary experimental data demands that NNLO (or at least NLO) DGLAP evolution be used in comparisons between QCD theory and

experiment. At higher orders, it is necessary to specify, and to use consistently, both a renormalization and a factorization scheme. Whereas the renormalization scheme used is almost universally the modified minimal subtraction ($\overline{\rm MS}$) scheme [30,31], there are two popular choices for factorization scheme, in which the form of the correction for each structure function is different. The two most-used factorization schemes are: DIS [32], in which there are no higher-order corrections to the F_2 structure function, and $\overline{\rm MS}$ [33]. They differ in how the non-divergent pieces are assimilated in the parton distribution functions.

It is usually assumed that the quarks are massless. The effects of the c and b-quark masses have been studied up to NNLO, for example, in Refs. 34–39. An approach using a variable flavor number is now generally adopted, in which evolution with $n_f = 3$ is matched to that with $n_f = 4$ at the charm threshold, with an analogous matching at the bottom threshold.

The discussion above relates to the Q^2 behavior of leading-twist (twist-2) contributions to the structure functions. Higher-twist terms, which involve their own non-perturbative input, exist. These die off as powers of Q; specifically twist-n terms are damped by $1/Q^{n-2}$. The higher-twist terms appear to be numerically unimportant for Q^2 above a few GeV^2, except for x close to 1.

Table 16.1: Lepton-nucleon and related hard-scattering processes and their primary sensitivity to the parton distributions that are probed.

Process	Main Subprocess	PDFs Probed
$\ell^\pm N \to \ell^\pm X$	$\gamma^* q \to q$	$g(x \lesssim 0.01), q, \overline{q}$
$\ell^+(\ell^-)N \to \overline{\nu}(\nu)X$	$W^* q \to q'$	
$\nu(\overline{\nu})N \to \ell^-(\ell^+)X$	$W^* q \to q'$	
$\nu\ N \to \mu^+\mu^- X$	$W^* s \to c \to \mu^+$	s
$\ell N \to \ell Q X$	$\gamma^* Q \to Q$	$Q = c, b$
	$\gamma^* g \to Q\overline{Q}$	$g(x \lesssim 0.01)$
$pp \to \gamma X$	$qg \to \gamma q$	g
$pN \to \mu^+\mu^- X$	$q\overline{q} \to \gamma^*$	$\overline{q}$
$pp, pn \to \mu^+\mu^- X$	$u\overline{u}, d\overline{d} \to \gamma^*$	$\overline{u} - \overline{d}$
	$u\overline{d}, d\overline{u} \to \gamma^*$	
$ep, en \to e\pi X$	$\gamma^* q \to q$	
$p\overline{p} \to W \to \ell^\pm X$	$ud \to W$	$u, d, u/d$
$p\overline{p} \to$ jet $+X$	$gg, qg, qq \to 2j$	$q, g(0.01 \lesssim x \lesssim 0.5)$

16.3. Determination of parton distributions

The parton distribution functions (PDFs) can be determined from data for deep inelastic lepton-nucleon scattering and for related hard-scattering processes initiated by nucleons. Table 16.1 given below (based on Ref. [40]) highlights some processes and their primary sensitivity to PDFs.

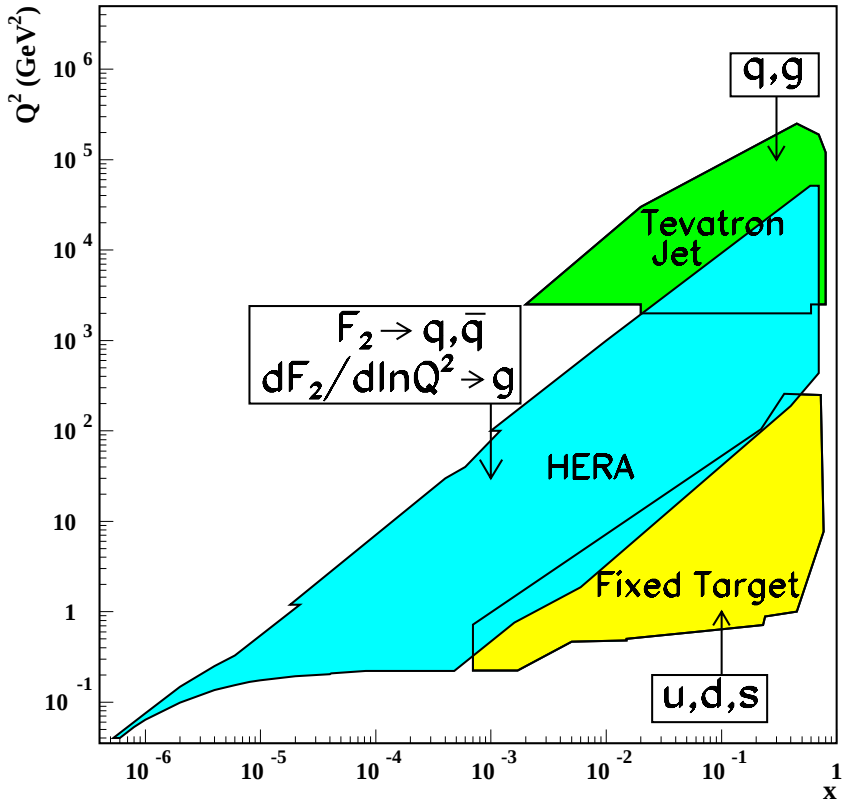


Figure 16.3: Kinematic domains in x and Q^2 probed by fixed-target and collider experiments, shown together with the important constraints they make on the various parton distributions. See full-color version on color pages at end of book.

The kinematic ranges of fixed-target and collider experiments are complementary (as is shown in Fig. 16.3), which enables the determination of PDFs over a wide range in x and Q^2. Recent determinations of the unpolarized PDF's from NLO global analyses are given in Ref. [13,41], and at NNLO in Ref. [13] (see also Ref. [42]). Recent studies of the uncertainties in the PDFs and observables can be found in Refs. [43,44] and Refs. [45,46] (see also Ref. [47]). The result of one analysis is shown in Fig. 16.4 at a scale $\mu^2 = 10$ GeV^2. The polarized PDFs are obtained through NLO global analyses of measurements of the g_1 structure function in inclusive polarized deep inelastic scattering (for recent examples see Refs. 48–50). The inclusive data do not provide

enough observables to determine all polarized PDFs. These polarized PDFs may be fully accessed via flavor tagging in semi-inclusive deep inelastic scattering. Fig. 16.5 shows several global analyses at a scale of 2.5 GeV^2 along with the data from semi-inclusive DIS.

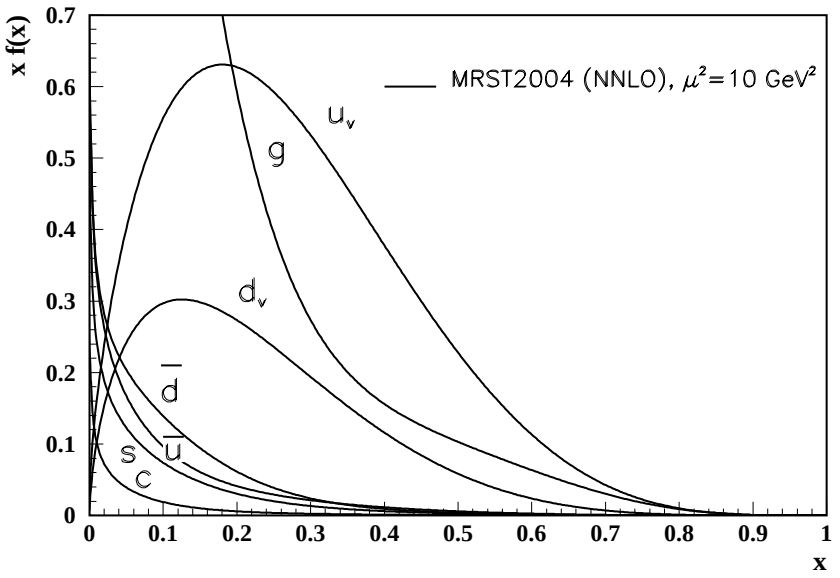


Figure 16.4: Distributions of x times the unpolarized parton distributions $f(x)$ (where $f = u_v, d_v, \overline{u}, \overline{d}, s, c, g$) using the NNLO MRST2004 parameterization [13] at a scale $\mu^2 = 10$ GeV^2.

Comprehensive sets of PDFs available as program-callable functions can be obtained from several sources *e.g.*, Refs. [53,54]. As a result of a Les Houches Accord, a PDF package (LHAPDF) exists [55] which facilitates the inclusion of recent PDFs in Monte Carlo/Matrix Element programs in a very compact and efficient format.

16.4. DIS determinations of α_s

Table 16.2 shows the values of $\alpha_S(M_Z^2)$ found in recent fits to DIS and related data in which the coupling is left as a free parameter.

There have been several other studies of α_s at NNLO, and beyond, using subsets of DIS data (see, for example, Refs. 59–61). Moreover, there exist global NLO analyses of polarized DIS data which give $\alpha_s(M_Z^2) = 0.120 \pm 0.009$ [62] and 0.114 ± 0.009 [50].

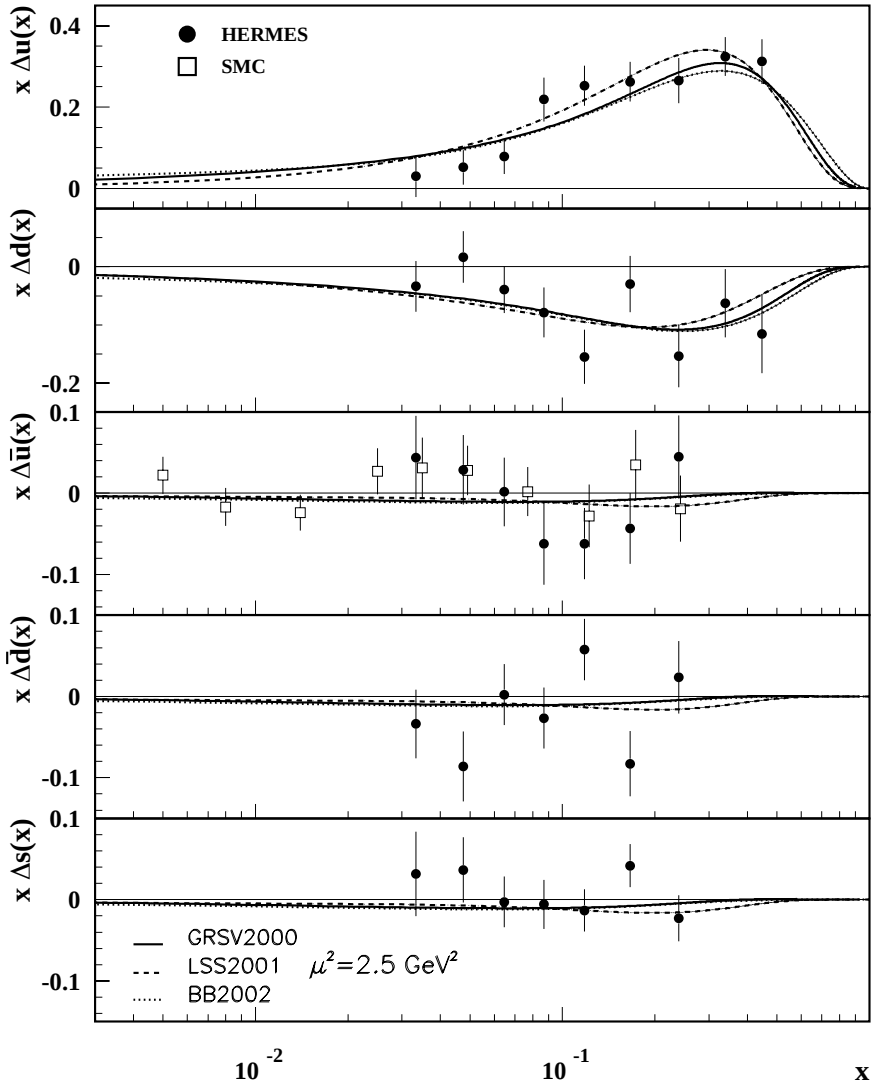


Figure 16.5: Distributions of x times the polarized parton distributions $\Delta q(x)$ (where $q = u, d, \overline{u}, \overline{d}, s$) using the GRSV2000 [48], LSS2001 [49], and BB2002 [50] parameterizations at a scale $\mu^2 = 2.5$ GeV^2. Points represent data from semi-inclusive positron (HERMES [51]) and muon (SMC [52]) deep inelastic scattering given at $Q^2 = 2.5$ GeV^2. SMC results are extracted under the assumption that $\Delta\overline{u}(x) = \Delta\overline{d}(x)$.

16.5. The hadronic structure of the photon

Besides the *direct* interactions of the photon, it is possible for it to fluctuate into a hadronic state via the process $\gamma \to q\overline{q}$. While in this state, the partonic content of the photon may be *resolved*, for example, through the process $e^+e^- \to e^+e^-\gamma^*\gamma \to e^+e^-X$ where the virtual photon emitted by the DIS lepton probes the hadronic structure of the quasi-real photon emitted by the other lepton. The perturbative LO contributions,

Table 16.2: The values of $\alpha_S(M_Z^2)$ found in NLO and NNLO fits to DIS and related data. CTEQ [56] and MRST04 [13] are global fits. H1 [57] fit only a subset of the F_2^{ep} data, while Alekhin [42] also includes F_2^{ed} and ZEUS [58] in addition include their charged current and jet data. The experimental errors quoted correspond to different choices of the effective increase $\Delta\chi^2$ from the best fit value of χ^2.

	$\Delta\chi^2$	$\alpha_S(M_Z^2)$ $\pm$ expt $\pm$ theory $\pm$ model	
NLO			
CTEQ	37	0.1169 ± 0.0045	
ZEUS	50	0.1183 ± 0.0028	± 0.0008
MRST04	20	$0.1205 \pm 0.002 \quad \pm 0.003$	
H1	1	$0.115 \quad \pm 0.0017 \pm 0.005$	$^{+0.0009}_{-0.0005}$
Alekhin	1	$0.1171 \pm 0.0015 \pm 0.0033$	
NNLO			
MRST04	20	$0.1167 \pm 0.002 \quad \pm 0.003$	
Alekhin	1	$0.1143 \pm 0.0014 \pm 0.0009$	

$\gamma \to q\overline{q}$ followed by $\gamma^* q \to q$, are subject to QCD corrections due to the coupling of quarks to gluons.

Often the equivalent-photon approximation is used to express the differential cross section for deep inelastic electron–photon scattering in terms of the structure functions of the transverse quasi-real photon times a flux factor N_γ^T (for these incoming quasi-real photons of transverse polarization)

$$\frac{d^2\sigma}{dxdQ^2} = N_\gamma^T \frac{2\pi\alpha^2}{xQ^4}\left[\left(1+(1-y)^2\right)F_2^\gamma(x,Q^2) - y^2 F_L^\gamma(x,Q^2)\right],$$

where we have used $F_2^\gamma = 2xF_T^\gamma + F_L^\gamma$. Complete formulae are given, for example, in the comprehensive review of Ref. [63].

The hadronic photon structure function, F_2^γ, evolves with increasing Q^2 from the 'hadron-like' behavior, calculable via the vector-meson-dominance model, to the dominating 'point-like' behaviour, calculable in perturbative QCD. Due to the point-like coupling, the logarithmic evolution of F_2^γ with Q^2 has a *positive* slope for all values of x, see Fig. 16.13. The 'loss' of quarks at large x due to gluon radiation is over-compensated by the 'creation' of quarks via the point-like $\gamma \to q\bar{q}$ coupling. The logarithmic evolution was first predicted in the quark–parton model ($\gamma^*\gamma \to q\bar{q}$) [64,65] and then in QCD in the limit of large Q^2 [66]. The evolution is now known to NLO [67,68,69]. Recent NLO data analyses to determine the parton densities of the photon can be found in Refs. [70,71,72].

References:

1. J. Blümlein and N. Kochelev, Nucl. Phys. **B498**, 285 (1997).
2. S. Forte *et al.*, Nucl. Phys. **B602**, 585 (2001).
3. M. Anselmino *et al.*, Z. Phys. **C64**, 267 (1994).
4. M. Anselmino *et al.*, Phys. Rep. **261**, 1 (1995).
5. M. Klein and T. Riemann, Z. Phys. **C24**, 151 (1984).
6. C.G. Callan and D.J. Gross, Phys. Rev. Lett. **22**, 156 (1969).
7. D.A. Dicus, Phys. Rev. **D5**, 1367 (1972).
8. J.D. Bjorken and E.A. Paschos, Phys. Rev. **185**, 1975 (1969).
9. R.P. Feynman, Photon Hadron Interactions (Benjamin, New York, 1972).
10. S. Wandzura and F. Wilczek, Phys. Rev. **B72**, 195 (1977).
11. J. Blümlein and A. Tkabladze, Nucl. Phys. **B553**, 427 (1999).
12. J.D. Bjorken, Phys. Rev. **179**, 1547 (1969).
13. A.D. Martin *et al.*, Phys. Lett. **B604**, 61 (2004).
14. V.N. Gribov and L.N. Lipatov, Sov. J. Nucl. Phys. **15**, 438 (1972).
15. L.N. Lipatov, Sov. J. Nucl. Phys. **20**, 95 (1975).
16. G. Altarelli and G. Parisi, Nucl. Phys. **B126**, 298 (1977).
17. Yu.L. Dokshitzer, Sov. Phys. JETP **46**, 641 (1977).
18. G. Curci *et al.*, Nucl. Phys. **B175**, 27 (1980); W. Furmanski, and R. Petronzio, Phys. Lett. **B97**, 437 (1980).
19. R.K. Ellis *et al.*, QCD and Collider Physics (Cambridge UP, 1996).
20. E.B. Zijlstra and W.L. van Neerven, Phys. Lett. **B272**, 127 (1991); Phys. Lett. **B273**, 476 (1991); Phys. Lett. **B297**, 377 (1992); Nucl. Phys. **B383**, 525 (1992).
21. S. Moch and J.A.M. Vermaseren, Nucl. Phys. **B573**, 853 (2000).
22. S. Moch *et al.*, Nucl. Phys. **B688**, 101 (2004); Nucl. Phys. **B691**, 129 (2004); Phys. Lett. **B606**, 123 (2005); Nucl. Phys. **B724**, 3 (2005).
23. E.A. Kuraev *et al.*, Phys. Lett. **B60**, 50 (1975); Sov. Phys. JETP **44**, 443 (1976); Sov. Phys. JETP **45**, 199 (1977).
24. Ya.Ya. Balitsky and L.N. Lipatov, Sov. J. Nucl. Phys. **28**, 822 (1978).
25. V.S. Fadin, and L.N. Lipatov, Phys. Lett. **B429**, 127 (1998).
26. G. Camici and M. Ciafaloni, Phys. Lett. **B412**, 396 (1997), erratum-Phys. Lett. **B147**, 390 (1997); Phys. Lett. **B430**, 349 (1998).
27. M. Ciafaloni *et al.*, Phys. Rev. **D60**, 114036 (1999); JHEP **0007** 054 (2000).
28. M. Ciafaloni *et al.*, Phys. Lett. **B576**, 143 (2003); Phys. Rev. **D68**, 114003 (2003).
29. G. Altarelli *et al.*, Nucl. Phys. **B621**, 359 (2002); Nucl. Phys. **B674**, 459 (2003).
30. G. 't Hooft and M. Veltman, Nucl. Phys. **B44**, 189 (1972).
31. G. 't Hooft, Nucl. Phys. **B61**, 455 (1973).
32. G. Altarelli *et al.*, Nucl. Phys. **B143**, 521 (1978) and erratum: Nucl. Phys. **B146**, 544 (1978).
33. W.A. Bardeen *et al.*, Phys. Rev. **D18**, 3998 (1978).
34. M.A.G. Aivazis *et al.*, Phys. Rev. **D50**, 3102 (1994).
35. J.C. Collins, Phys. Rev. **D58**, 094002 (1998).
36. A. Chuvakin *et al.*, Phys. Rev. **D61**, 096004 (2000).

37. R.S. Thorne and R.G. Roberts, Phys. Rev. **D57**, 6871 (1998); Phys. Lett. **B421**, 303 (1998); Eur. Phys. J. **C19**, 339 (2001).
38. W.-K. Tung, *et al.*, J. Phys. **G28**, 983 (2002); S. Kretzer *et al.*, Phys. Rev. **D69**, 114005 (2004).
39. R.S. Thorne, Phys. Rev. **D73**, 054019 (2006).
40. A.D. Martin *et al.*, Eur. Phys. J. **C4**, 463 (1998).
41. CTEQ, J. Pumplin *et al.*, JHEP **0207**, 012 (2002).
42. S. Alekhin, JHEP **0302**, 015 (2003).
43. CTEQ, D. Stump *et al.*, Phys. Rev. **D65**, 014012 (2001).
44. CTEQ, J. Pumplin *et al.*, Phys. Rev. **D65**, 014013 (2001).
45. A.D. Martin *et al.*, Eur. Phys. J. **C28**, 455 (2003).
46. A.D. Martin *et al.*, Eur. Phys. J. **C35**, 325 (2004).
47. W.T. Giele *et al.*, `hep-ph/0104052`.
48. M. Glück *et al.*, Phys. Rev. **D63**, 094005 (2001).
49. E. Leader *et al.*, Eur. Phys. J. **C23**, 479 (2002).
50. J. Blümlein and H. Böttcher, Nucl. Phys. **B636**, 225 (2002).
51. HERMES, A. Airpetian *et al.*, Phys. Rev. Lett. **92**, 012005 (2004).
52. SMC, B. Adeva *et al.*, Phys. Lett. **B420**, 180 (1998).
53. H. Plothow-Besch, CERN PDFLIB, W5051 (2000).
54. `http://durpdg.dur.ac.uk/HEPDATA/PDF`.
55. `http://durpdg.dur.ac.uk/lhapdf/index.html`.
56. J. Huston *et al.*, JHEP **0506**, 080 (2005).
57. H1, C. Adloff *et al.*, Eur. Phys. J. **C21**, 33 (2001).
58. ZEUS, S. Chekanov *et al.*, Eur. Phys. J. **C42**, 1 (2005).
59. W.L. van Neerven and A. Vogt, Nucl. Phys. **B603**, 42 (2001).
60. J. Santiago and F.J. Yndurain, Nucl. Phys. **B611**, 447 (2001).
61. A.L. Kataev *et al.*, Phys. Part. Nucl. **34**, 20 (2003).
62. G. Altarelli *et al.*, Nucl. Phys. **B496**, 337 (1997).
63. R. Nisius, Phys. Reports **332**, 165 (2000).
64. T.F. Walsh and P.M. Zerwas, Phys. Lett. **B44**, 195 (1973).
65. R.L. Kingsley, Nucl. Phys. **B60**, 45 (1973).
66. E. Witten, Nucl. Phys. **B120**, 189 (1977).
67. W.A. Bardeen and A.J. Buras, Phys. Rev. **D20**, 166 (1979), erratum Phys. Rev. **D21**, 2041 (1980).
68. M. Fontannaz and E. Pilon, Phys. Rev. **D45**, 382 (1992), erratum Phys. Rev. **D46**, 484 (1992).
69. M. Glück *et al.*, Phys. Rev. **D45**, 3986 (1992).
70. F. Cornet *et al.*, Phys. Rev. **D70**, 093004 (2004).
71. P. Aurenche, *et al.*, Eur. Phys. J. **C44**, 395 (2005).
72. W. Slominski *et al.*, Eur. Phys. J. **C45**, 633 (2006).

17. FRAGMENTATION FUNCTIONS IN e^+e^- ANNIHILATION AND LEPTON-NUCLEON DIS

Revised August 2005 by O. Biebel (Ludwig-Maximilians-Universität, Munich, Germany), D. Milstead (Fysikum, Stockholms Universitet, Sweden), P. Nason (INFN, Sez. di Milano, Milan, Italy), and B.R. Webber (Cavendish Laboratory, Cambridge, UK). An extended version of the 2001 review can be found in Ref. [1]

17.1. Concept of fragmentation

17.1.1. *Introduction* :

Fragmentation functions are dimensionless functions that describe the final-state single-particle energy distributions in hard scattering processes, like e^+e^- annihilation or deep inelastic lepton-nucleon scattering, or high transverse momentum hadrons in photon-hadron and hadron-hadron collisions. The total e^+e^- fragmentation function for hadrons of type h in annihilation at c.m. energy $\sqrt{s}$, via an intermediate vector boson $V = \gamma/Z^0$, is defined as

$$F^h(x,s) = \frac{1}{\sigma_{\rm tot}}\frac{d\sigma}{dx}(e^+e^- \to V \to hX) \tag{17.1}$$

where $x = 2E_h/\sqrt{s} \leq 1$ is the scaled hadron energy (in practice, the approximation $x = x_p = 2p_h/\sqrt{s}$ is often used). Its integral with respect to x gives the average multiplicity of those hadrons:

$$\langle n_h(s)\rangle = \int_0^1 dx\, F^h(x,s)\ . \tag{17.2}$$

Neglecting contributions suppressed by inverse powers of s, the fragmentation function (17.1) can be represented as a sum of contributions from the different parton types $i = u, \overline{u}, d, \overline{d}, \ldots, g$:

$$F^h(x,s) = \sum_i \int_x^1 \frac{dz}{z} C_i(s;z,\alpha_{\rm S}) D_i^h(x/z,s)\ . \tag{17.3}$$

where D_i^h are the parton fragmentation functions. At lowest order in $\alpha_{\rm S}$ the coefficient function C_g for gluons is zero, while for quarks $C_i = g_i(s)\delta(1-z)$ where $g_i(s)$ is the appropriate electroweak coupling. In particular, $g_i(s)$ is proportional to the charge-squared of parton i at $s \ll M_Z^2$, when weak effects can be neglected. In higher orders the coefficient functions and parton fragmentation functions are factorization-scheme dependent.

Parton fragmentation functions are analogous to the parton distributions in deep inelastic scattering (see sections on QCD and Structure Functions, 9 and 16 of this *Review*). In both cases, the simplest parton-model approach would predict a scale-independent x distribution. Furthermore we obtain similar violations of this scaling behaviour when QCD corrections are taken into account.

Fragmentation functions in lepton-hadron scattering and e^+e^- annihilation are complementary. Since e^+e^- annihilation results in a neutral off mass-shell photon or Z^0, fragmentation arising from a pure quark-antiquark system can be studied. Lepton-hadron scattering is a more complicated environment with which it is possible to study the influence on fragmentation functions from initial state QCD radiation, the partonic and spin structure of the hadron target, and the target remnant system.[1]

In lepton-hadron scattering, calling p the four momentum of the incoming hadron, and q the four momentum of the exchanged virtual boson, one can construct two independent kinematic invariants. One usually introduces $Q^2 = -q^2$ and defines $x_{\rm Bj} = Q^2/(2p \cdot q)$. Thus, there is a freedom in the choice of scale used to define a fragmentation function. For e^+e^-, the c.m. energy provides a natural choice of scale as twice the energy of each produced quark. For lepton-hadron interactions, fragmentation scales such as $Q = \sqrt{-q^2}$, or the invariant mass of the exchanged boson and target nucleon system $W = \sqrt{(p+q)^2} = \sqrt{m_h^2 + 2p \cdot q - Q^2}$ (where m_h is the incoming hadron mass) are typically used. Both W and Q can vary by several orders of magnitudes for a given c.m. energy thus allowing the study of fragmentation in different environments by a single experiment, *e.g.* in photoproduction the exchanged photon is quasi-real ($Q^2 \sim 0$) leading to processes akin to hadron-hadron scattering. In deep inelastic scattering (DIS)($Q^2 \gg 1$ GeV^2), using the Quark Parton Model (QPM), the hadronic fragments of the struck quark can be directly compared with quark fragmentation in e^+e^-. Results from lepton-hadron experiments quoted in this report primarily concern fragmentation in the DIS regime. Studies made by lepton-hadron experiments of fragmentation with photoproduction data containing high transverse momentum jets or particles are also reported, when these are directly comparable to DIS and e^+e^- results.

After the lepton-hadron interaction, the transverse momentum of the scattered lepton is balanced by the hadronic system. To remove the transverse momentum imbalance many fragmentation studies have been performed in frames in which the target hadron and the exchanged boson are collinear. Two frames of reference are typically used which fulfill this condition.

The so-called hadronic c.m. frame (HCMS) is defined as the rest system of the exchanged boson and incoming hadron, with the z^*-axis defined along the direction of the exchanged boson. The $+z^*$ direction defines the so-called current region. Fragmentation measurements performed in the HCMS often use the Feynman-x variable $x_F = 2p_z^*/W$, where p_z^* is the longitudinal momentum of the particle in the HCMS. Since W is the invariant mass of the hadronic final state, x_F ranges between -1 and 1.

The Breit system [3] is connected to the HCMS by a longitudinal boost such that the time component of q becomes 0, so that $q = (0, 0, 0, -Q)$. In this frame the target has three momentum $\vec{p} = (0, 0, Q/(2x_{\rm Bj}))$, as can be easily verified using the definition of $x_{\rm Bj}$. In the quark parton model, the struck parton has momentum $x_{\rm Bj} \cdot p$, and thus its longitudinal momentum is $Q/2$, which becomes $-Q/2$ after the collision. As

[1] For a comprehensive review of the measurements and models of fragmentation in lepton-hadron scattering see [2].

compared with the HCMS, the current region of the Breit frame is more closely matched to the partonic scattering process, and is thus appropriate for direct comparisons of fragmentation functions in DIS with those from e^+e^- annihilation. The variable $x_p = 2p^*/Q$ is used at HERA for measurements of fragmentation functions in the Breit frame, ensuring directly comparable DIS and e^+e^- results.

17.2. Scaling violation

The evolution of the parton fragmentation function $D_i(x,t)$ with increasing scale $t = s$, like that of the parton distribution function $f_i(x,t)$ with $t = s$ (see Sec. 39 of this *Review*), is governed by the DGLAP equation [4]

$$t\frac{\partial}{\partial t}D_i(x,t) = \sum_j \int_x^1 \frac{dz}{z}\frac{\alpha_S}{2\pi}P_{ji}(z,\alpha_S)D_j(x/z,t)\ . \tag{17.4}$$

In analogy to DIS, in some cases an evolution equation for the fragmentation function F itself (Eq. (17.3)) can be derived from Eq. (17.4) [5]. Notice that the splitting function is now P_{ji} rather than P_{ij} since here D_j represents the fragmentation of the final parton. The splitting functions again have perturbative expansions of the form

$$P_{ji}(z,\alpha_S) = P_{ji}^{(0)}(z) + \frac{\alpha_S}{2\pi}P_{ji}^{(1)}(z) + \cdots \tag{17.5}$$

where the lowest-order functions $P_{ji}^{(0)}(z)$ are the same as those in deep inelastic scattering but the higher-order terms [6,7] [2] are different. The effect of evolution is, however, the same in both cases: as the scale increases, one observes a scaling violation in which the x distribution is shifted towards lower values. This can be seen from Fig. 17.1.

The coefficient functions C_i in Eq. (17.3) and the splitting functions P_{ji} contain singularities at $z = 0$ and 1, which have important effects on fragmentation at small and large values of x, respectively. For details see *e.g.* Ref. [1].

Quantitative results of studies of scaling violation in e^+e^- fragmentation are reported in [26–30] The values of α_S obtained are consistent with the world average (see review on QCD in Sec. 9 of this *Review*).

Scaling violations in DIS are shown in Fig. 17.2 for both HCMS and Breit frame. In Fig. 17.2(a) the distribution in terms of $x_F = 2p_z/W$ shows for values of $x_F > 0.15$ a steeper slope in ep data than for the μp data, indicating the scaling violations. At smaller values of x_F in the current jet region, the multiplicity of particles substantially increases with W owing to the increased phase space available for the fragmentation process. The EMC data access both the current region and the region of the fragmenting target remnant system. At higher values of $|x_F|$, owing to the extended nature of the remnant, the multiplicity in the target region far exceeds that in the current region. Owing to

[2] There are misprints in the formulae in the published article [6]. The correct expressions can be found in the preprint version or in Ref. [8].

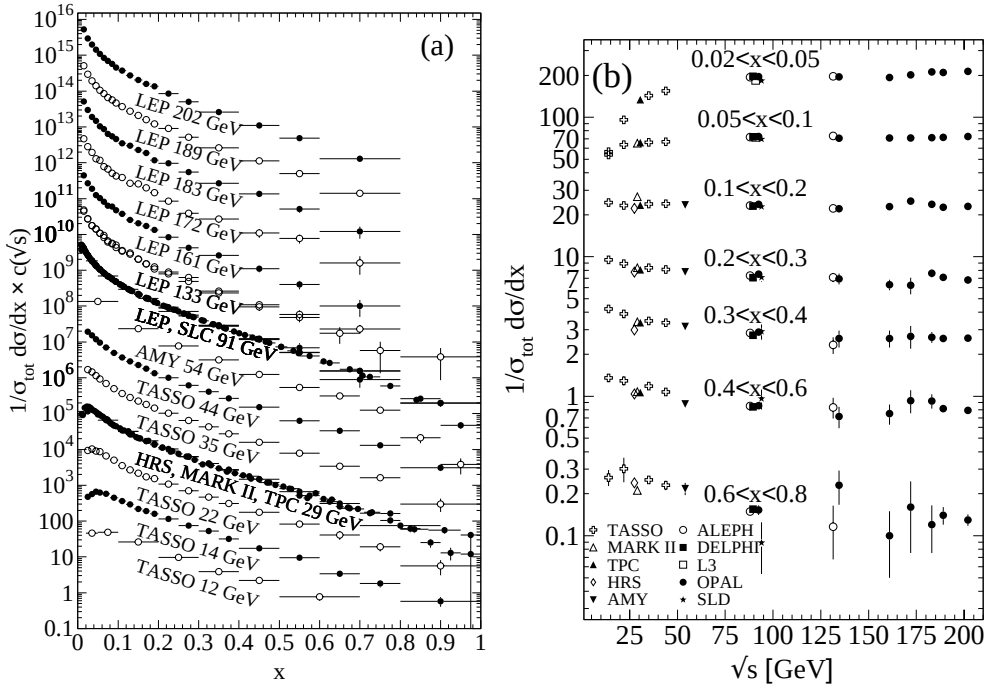


Figure 17.1: The e^+e^- fragmentation function for all charged particles is shown [9–25] (a) for different c.m. energies, $\sqrt{s}$, versus x and (b) for various ranges of x versus $\sqrt{s}$. For the purpose of plotting (a), the distributions were scaled by $c(\sqrt{s}) = 10^i$ where i is ranging from $i = 0$ ($\sqrt{s} = 12$ GeV) to $i = 13$ ($\sqrt{s} = 202$ GeV).

acceptance reasons, the remnant hemisphere of the HCMS is only accessible by the lower energy fixed target experiments

Using hadrons from the current hemisphere in the Breit frame, measurements of fragmentation functions and the production properties of particles in ep scattering have been made by [35–39]. Fig. 17.2(b) compares results from ep scattering and e^+e^- experiments, the latter results are halved as they cover both event hemispheres. The agreement between the DIS and e^+e^- results is fairly good. However, processes in DIS which are not present in e^+e^- annihilation, such as boson-gluon fusion and initial state QCD radiation, can depopulate the current region. These effects become most prominent at low values of Q and x_p. When compared with e^+e^- annihilation data at $\sqrt{s} = 5.2$, 6.5 GeV [40], which are not shown here, the DIS particle rates tend to lie below those from e^+e^- annihilation. NLO QCD calculations [41], convoluted with fragmentation functions derived from e^+e^- data, have been tested against the HERA scaling violations data and provide a good description of the data in the kinematic regions in which the calculations are predictive [35,39,42].

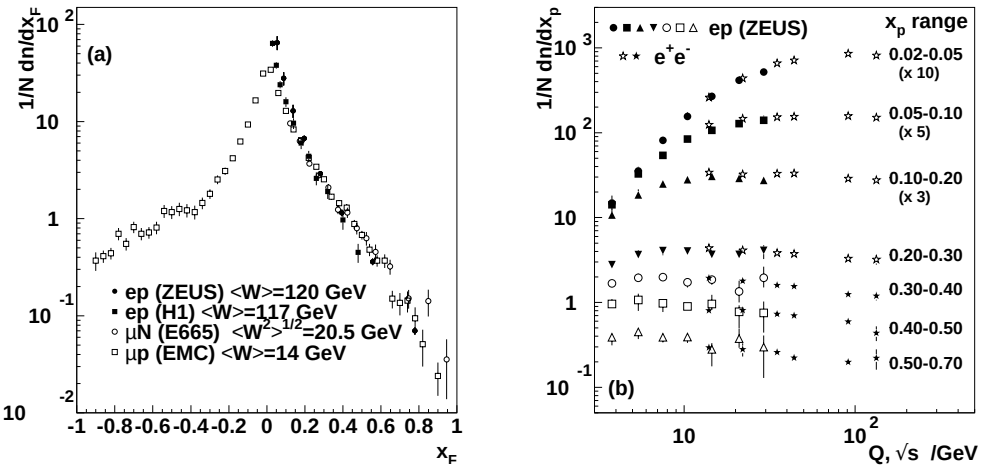


Figure 17.2: (a) The distribution $1/N \cdot dN/dx_F$ for all charged particles in DIS lepton-hadron experiments at different values of W, and measured in the HCMS [31–34]. (b) Scaling violations of the fragmentation function for all charged particles in the current region of the Breit frame of DIS [35] and in e^+e^- interactions [17,24,27]. The data are shown as a function of $\sqrt{s}$ for e^+e^- results, and as a function of Q for the DIS results, each within the same indicated intervals of the scaled momentum x_p. The data for the three lowest intervals of x_p are multiplied by factors 10, 5, and 3, respectively for clarity.

17.3. Fragmentation functions for small particle momenta

As in the case of the parton distribution functions, the most common strategy for solving the evolution equations Eq. (17.4) is to take moments (Mellin transforms) with respect to x:

$$\tilde{D}(j,s) = \int_0^1 dx\; x^{j-1}\; D(x,s)\;. \tag{17.6}$$

The behaviour of $\tilde{D}(j,s)$ away from $j=1$ determines the form of small-x fragmentation functions. Keeping the first three terms in a Taylor expansion around $j=1$ gives a simple Gaussian function of j which transforms by inverse Mellin transformation into a Gaussian in the variable $\xi \equiv \ln(1/x)$:

$$xD(x,s) \propto \exp\left[-\frac{1}{2\sigma^2}(\xi-\xi_p)^2\right]\;, \tag{17.7}$$

where the peak position is

$$\xi_p = \frac{1}{4b\alpha_{\rm S}(s)} \simeq \frac{1}{4}\ln\left(\frac{s}{\Lambda^2}\right) \tag{17.8}$$

with $b=(33-2n_f)/12\pi$ for n_f quark flavours and the width of the distribution of ξ is ($C_A=3$)

$$\sigma = \left(\frac{1}{24b}\sqrt{\frac{2\pi}{C_A\alpha_{\rm S}^3(s)}}\right)^{\frac{1}{2}} \propto \left[\ln\left(\frac{s}{\Lambda^2}\right)\right]^{\frac{3}{4}}\;. \tag{17.9}$$

Again, one can compute next-to-leading corrections to these predictions. In the method of [43], the corrections are included in an analytical form known as the 'modified leading logarithmic approximation' (MLLA). Alternatively they can be used to compute the higher moment corrections to the Gaussian form Eq. (17.7) [44]. Fig. 17.3[3] shows the ξ distribution for charged particles produced in the current region of the Breit frame in DIS interactions, and in e^+e^- annihilation. As expected from Eq. (17.7), Eq. (17.8), and Eq. (17.9), the distributions have a Gaussian shape, with the peak position and area becoming progressively larger with c.m. energy (e^+e^-) and Q^2 (DIS).

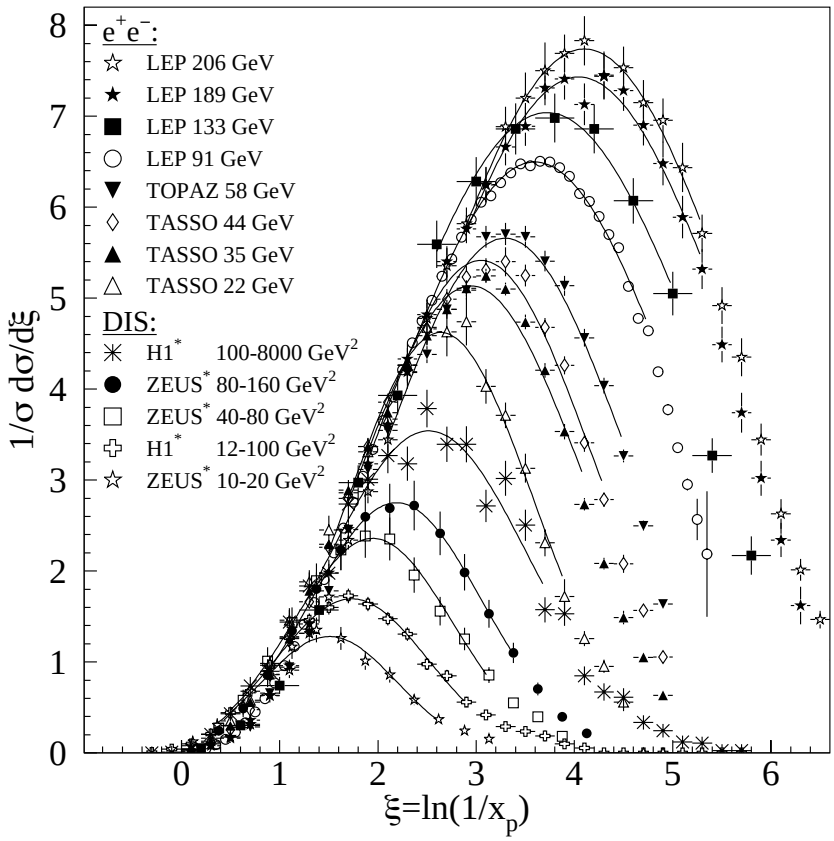


Figure 17.3: Distribution of $\xi = \ln(1/x_p)$ at several c.m. energies (e^+e^-) [9,10,14,17–20,24,45–48] and intervals of Q^2 (DIS) [37–39]. [3] At each energy only one representative measurement is shown. For clarity some measurements at intermediate c.m. energies (e^+e^-) or Q^2 ranges (DIS) are not shown. The DIS measurements marked with * have been scaled by a factor of 2 for direct comparability with the e^+e^- results. Overlaid are fits of a simple Gaussian function for illustration.

The predicted energy dependence Eq. (17.8) of the peak in the ξ distribution is a striking illustration of soft gluon coherence, which is the origin of the suppression of

hadron production at small x. Of course, a decrease at very small x is expected on purely kinematical grounds, but this would occur at particle energies proportional to their masses, i.e. at $x \propto m/\sqrt{s}$ and hence $\xi \sim \frac{1}{2}\ln s$. Thus if the suppression were purely kinematic the peak position ξ_p would vary twice as rapidly with energy, which is ruled out by the data (see Fig. 17.4). The e^+e^- and DIS data agree well with each other, demonstrating the universality of hadronization. The MLLA prediction describes the data. Measurements of the higher moments of the ξ distribution in e^+e^- [24,48–50] and DIS [39] have also been made and show consistency with each other.

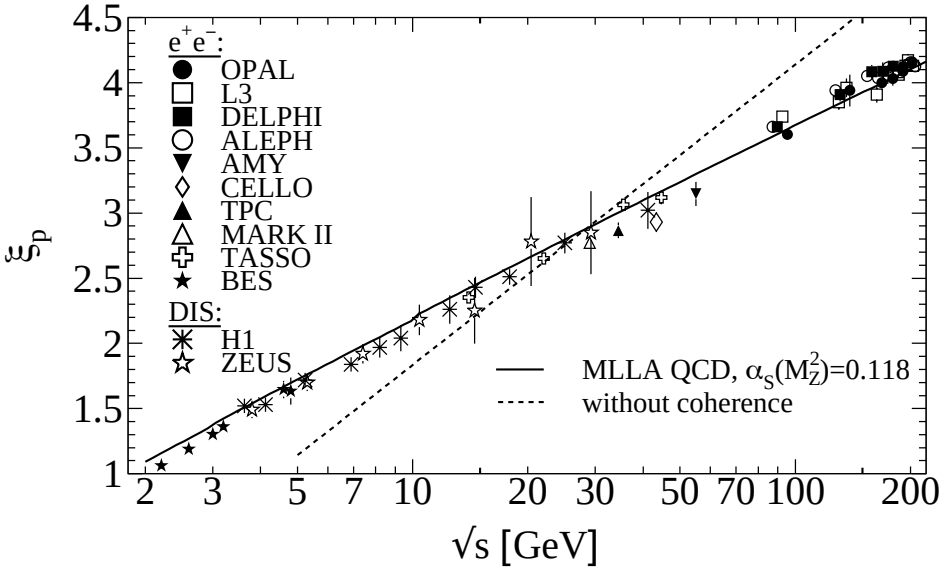


Figure 17.4: Evolution of the peak position, ξ_p, of the ξ distribution with the c.m. energy $\sqrt{s}$. The MLLA QCD prediction (solid) using $\alpha_S(s = M_Z^2) = 0.118$ and the expectation without gluon coherence (dashed) are superimposed to the data [9,11,14,16–18,20,24,37,38,46,47,49,51–58].

17.3.1. *Longitudinal Fragmentation* :

In the process $e^+e^- \to V \to hX$, the joint distribution in the energy fraction x and the angle θ between the observed hadron h and the incoming electron beam has the general form

$$\frac{1}{\sigma_{\rm tot}}\frac{d^2\sigma}{dx\, d\cos\theta} = \frac{3}{8}(1+\cos^2\theta)\, F_T(x) + \frac{3}{4}\sin^2\theta\, F_L(x) + \frac{3}{4}\cos\theta\, F_A(x)\ , \qquad (17.10)$$

where F_T, F_L and F_A are respectively the transverse, longitudinal and asymmetric fragmentation functions. All these functions also depend on the c.m. energy $\sqrt{s}$.

[3] Dominant systematic errors used for the results of [39] are according to this reference 10% for $x_p < 0.3$ and increase up to 30% for large x_p.

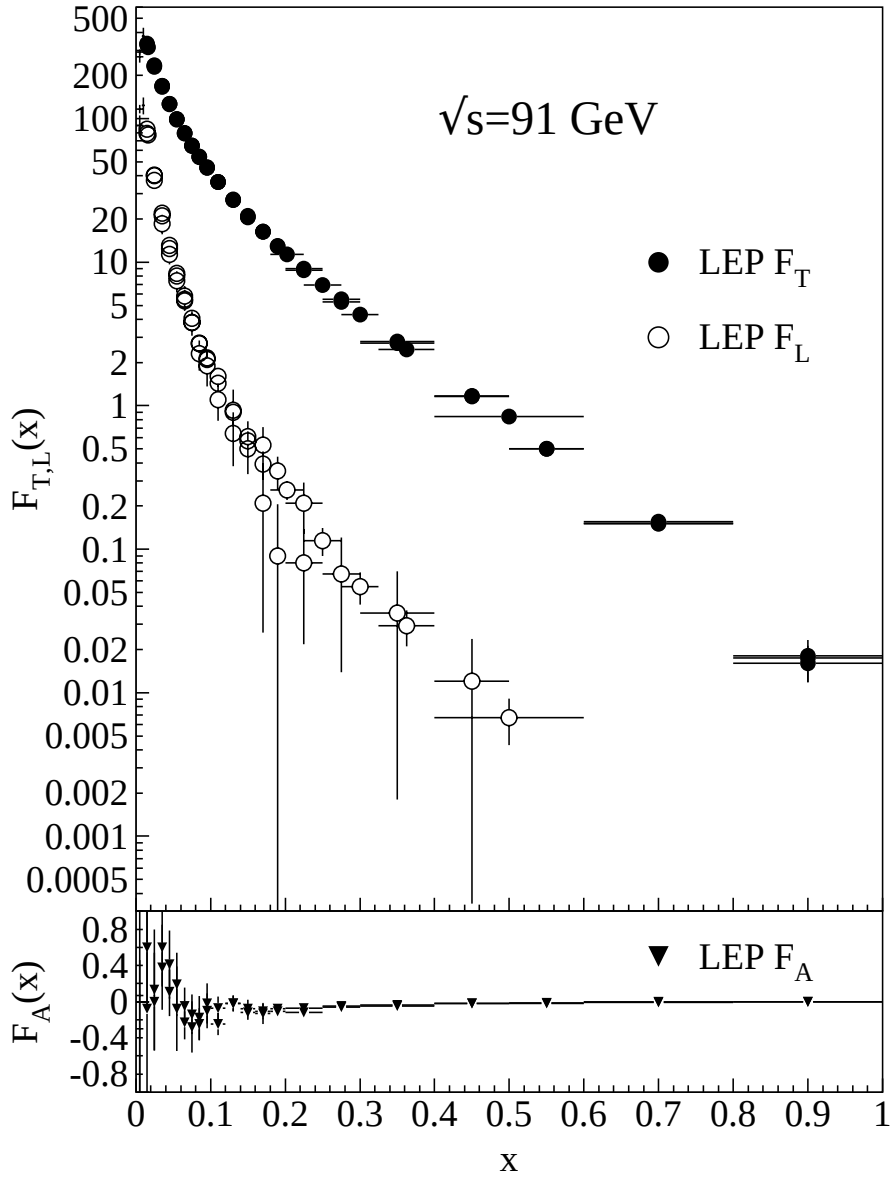


Figure 17.5: Transverse (F_T), longitudinal (F_L), and asymmetric (F_A) fragmentation functions are shown [12,26,59]. Data points with relative errors greater than 100% are omitted.

Eq. (17.10) is the most general form of the inclusive single particle production from the decay of a massive vector boson [5]. As their names imply, F_T and F_L represent the contributions from virtual bosons polarized transversely or longitudinally with respect to the direction of motion of the hadron h. F_A is a parity-violating contribution which comes from the interference between vector and axial vector contributions. Integrating over all angles, we obtain the total fragmentation function, $F = F_T + F_L$. Each of these

functions can be represented as a convolution of the parton fragmentation functions D_i with appropriate coefficient functions $C_i^{\mathrm{T,L,A}}$ as in Eq. (17.3). This representation works in the high energy limit. As $x \cdot \sqrt{s}/2$ approaches hadronic scales $\simeq m_\rho$, power suppressed effects can no longer be neglected, and the fragmentation function formalism no longer accounts correctly for the separation of F_T, F_L, and F_A. In Fig. 17.5, F_T, F_L, and F_A measured at $\sqrt{s} = 91$ GeV are shown.

17.4. Gluon fragmentation

The gluon fragmentation function $D_g(x)$ can be extracted from the longitudinal fragmentation function defined in Eq. (17.10). Since the coefficient functions C_i^{L} for quarks and gluons are comparable in $\mathcal{O}(\alpha_{\mathrm{S}})$, F_L can be expressed in terms of F_T and D_g which allows one to obtain D_g from the measured F_L and F_T, see *e.g.* [1] for details. At NLO, i.e. $\mathcal{O}(\alpha_S^2)$ coefficient functions C_i, quark fragmentation is dominant in F_L over a large part of the kinematic range, see *e.g.* [60]. This leaves some sensitivity of F_L to D_g but further constraints will be needed, for instance from hadro-production in deep-inelastic scattering.

D_g can also be deduced from the fragmentation of three-jet events in which the gluon jet is identified, for example by tagging the other two jets with heavy quark decays. To leading order the measured distributions of $x = E_{\mathrm{had}}/E_{\mathrm{jet}}$ for particles in gluon jets can be identified directly with the gluon fragmentation functions $D_g(x)$. The experimentally measured gluon fragmentation functions are shown in Fig. 17.6.

17.5. Spin-dependent fragmentation

Measurements of hadron production in polarized lepton-hadron scattering are used principally in the determination of polarized parton densities from DIS interactions with longitudinally polarized targets [62–64]. Since flavour-dependent fragmentation functions derived from the $e^+e^- \to hX$ can give significantly different flavour contributions [65–69] experiments use string fragmentation in JETSET [70], which is tuned to describe their own identified particle data, and those measured by other low energy DIS experiments.

Polarized scattering presents the possibility to measure the spin transfer from the struck quark to the final hadron, and thus develop spin-dependent fragmentation functions [71,72]. These are useful in the study of the quark transversity distribution [73], which describes the probability of finding a transversely polarized quark with its spin aligned or anti-aligned with the spin of a transversely polarized nucleon. The transversity function is chiral-odd and therefore not accessible through measurements of inclusive lepton-hadron scattering. Semi-inclusive DIS in which another chiral-odd observable may be involved provides a valuable tool to probe transversity. The Collins fragmentation function [74] relates the transverse polarization of the quark to that of the final hadron. It is chiral-odd and naive T-odd, leading to a characteristic single spin asymmetry in the azimuthal angular distribution of the produced hadron in the hadron scattering plane. A number of experiments have measured this asymmetry [75,76]. However, these studies were unable to distinguish between processes due to transversity in conjunction with the Collins fragmentation with other processes requiring non-polarized fragmentation

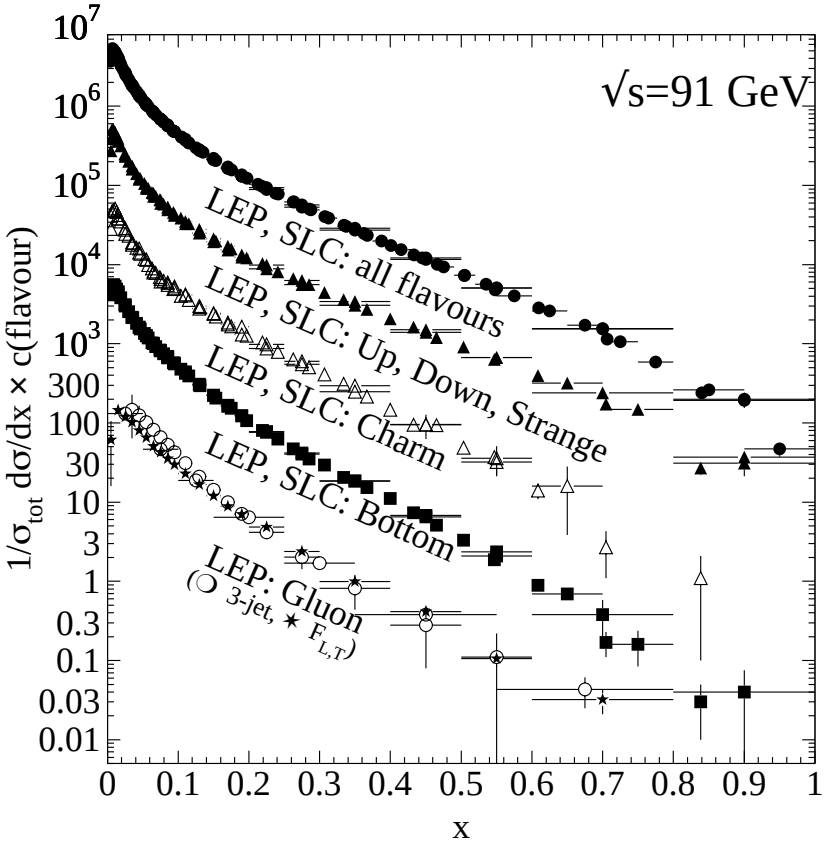


Figure 17.6: Comparison of the charged-particle and the flavour-dependent e^+e^- fragmentation functions obtained at $\sqrt{s} = 91$ GeV. The data [10,12,14,15,19,21] [22,28,59,61] are shown for the inclusive, light (up, down, strange) quarks, charm quark, bottom quark, and the gluon versus x. For the purpose of plotting, the distributions were scaled by $c(\text{flavour}) = 10^i$ where i is ranging from $i = 0$ (Gluon) to $i = 4$ (all flavours).

functions, such as the Sivers mechanism [77]. However, the HERMES and COMPASS collaborations have now made first measurements of the Collins and Sivers asymmetries using a transversely polarized target [76,78].

17.6. Fragmentation models

Although the scaling violation can be calculated perturbatively, the actual form of the parton fragmentation functions is non-perturbative. Perturbative evolution gives rise to a shower of quarks and gluons (partons). Phenomenological schemes are then used to model the carry-over of parton momenta and flavour to the hadrons. Two of the very popular models are the *string fragmentation* [79,80], implemented in the JETSET [70] and UCLA [81] Monte Carlo event generation programs, and the *cluster fragmentation* of the HERWIG Monte Carlo event generator [82].

17.6.1. ***String fragmentation*** : The string-fragmentation scheme considers the colour field between the partons, *i.e.*, quarks and gluons, to be the fragmenting entity rather than the partons themselves. The string can be viewed as a colour flux tube formed by gluon self-interaction as two coloured partons move apart. Energetic gluon emission is regarded as energy-momentum carrying "kinks" on the string. When the energy stored in the string is sufficient, a $q\overline{q}$ pair may be created from the vacuum. Thus the string breaks up repeatedly into colour singlet systems as long as the invariant mass of the string pieces exceeds the on-shell mass of a hadron. The $q\overline{q}$ pairs are created according to the probability of a tunnelling process $\exp(-\pi m^2_{q,\perp}/\kappa)$ which depends on the transverse mass squared $m^2_{q,\perp} \equiv m^2_q + p^2_{q,\perp}$ and the string tension $\kappa \approx 1$ GeV/fm. The transverse momentum $p_{q,\perp}$ is locally compensated between quark and antiquark. Due to the dependence on the parton mass m_q and/or hadron mass, m_h, the production of strange and, in particular, heavy-quark hadrons is suppressed. The light-cone momentum fraction $z = (E + p_\parallel)_h/(E + p)_q$, where $p_\parallel$ is the momentum of the formed hadron h along the direction of the quark q, is given by the string-fragmentation function

$$f(z) \sim \frac{1}{z}(1-z)^a \exp\left(-\frac{bm^2_{h,\perp}}{z}\right) \tag{17.11}$$

where a and b are free parameters. These parameters need to be adjusted to bring the fragmentation into accordance with measured data, *e.g.*, $a = 0.11$ and $b = 0.52$ GeV^{-2} as determined in Ref. [83] (for an overview on tuned parameters see Ref. [84]) .

17.6.2. ***Cluster fragmentation*** : Assuming a local compensation of colour based on the *pre-confinement* property of perturbative QCD [85], the remaining gluons at the end of the parton shower evolution are split non-perturbatively into quark-antiquark pairs. Colour singlet clusters of typical mass of a couple of GeV are then formed from quark and antiquark of colour-connected splittings. These clusters decay directly into two hadrons unless they are either too heavy (relative to an adjustable parameter `CLMAX`, default value 3.35 GeV), when they decay into two clusters, or too light, in which case a cluster decays into a single hadron, requiring a small rearrangement of energy and momentum with neighbouring clusters. The decay of a cluster into two hadrons is assumed to be isotropic in the rest frame of the cluster except if a perturbative-formed quark is involved. A decay channel is chosen based on the phase-space probability, the density of states, and the spin degeneracy of the hadrons. Cluster fragmentation has a compact description with few parameters, due to the phase-space dominance in the hadron formation.

17.7. Fragmentation into identified particles

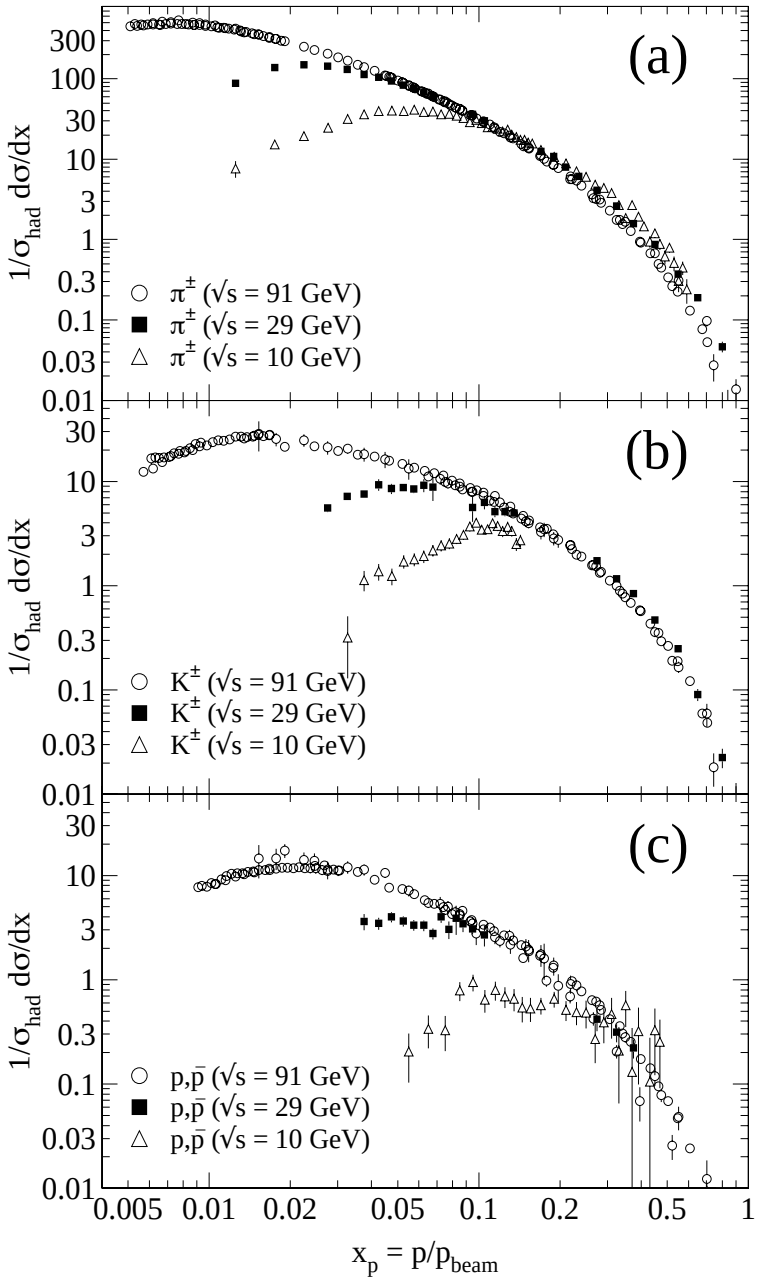


Figure 17.7: Scaled momentum spectra of (a) $\pi^\pm$, (b) $K^\pm$, and (c) $p/\overline{p}$ at $\sqrt{s} = 10$, 29, and 91 GeV are shown [22,25,55,86].

A great wealth of measurements of e^+e^- fragmentation into identified particles exists. A collection of references to find data on the fragmentation into identified particles is given for Table 40.1. As representatives of all the data, Fig. 17.7 shows fragmentation functions as the scaled momentum spectra of charged particles at several c.m. energies. Heavy flavour particles are dealt with separately in Sect. 17.8.

The measured fragmentation functions are solutions to the DGLAP equation (17.4) but need to be parametrized at some initial scale t_0 (usually 2 GeV^2 for light quarks and gluons). A general parametrization is [87]

$$D_{p\to h}(x,t_0) = Nx^{\alpha}(1-x)^{\beta}\left(1+\frac{\gamma}{x}\right) \tag{17.12}$$

where the normalization N, and the parameters α, β, and γ in general depend on the energy scale t_0 and also on the type of the parton, p, and the hadron, h. Frequently the term involving γ is left out [65–68]. The parameters of Eq. (17.12), listed in [65–68], were obtained by fitting data on various hadron types for different combinations of partons and hadrons in $p \to h$ in the range $\sqrt{s} \approx 5$ - 200 GeV.

Many studies have been made of identified particles produced in lepton-hadron scattering, although fewer particle species have been measured than in e^+e^- collisions. References [88–93] and [94–97] are representative of the data from fixed target and ep collider experiments.

Fig. 17.8(a) compares lower energy fixed target and HERA data on strangeness production showing that the HERA spectra have substantially increased multiplicities, albeit with insufficient statistical precision to study scaling violations. The fixed target data show that the Λ rate substantially exceeds the $\overline{\Lambda}$ rate in the remnant region, owing to the conserved baryon number from the baryon target. Fig. 17.8(b) shows neutral and charged pion fragmentation functions $1/N \cdot dn/dz$, where z is defined as the ratio of the pion energy to that of the exchanged boson, both measured in the laboratory frame. Results are shown from HERMES and the EMC experiments, where HERMES data have been evolved with NLO QCD to $\langle Q^2 \rangle = 25$ GeV^2 in order to be consistent with the EMC. Each of the experiments uses various kinematic cuts to ensure that the measured particles lie in the region which is expected to be associated with the struck quark. In the DIS kinematic regime accessed at these experiments, and over the range in z shown in Fig. 17.8, the z and x_F variables have similar values [31]. The precision data on identified particles can be used in the study of the quark flavour content of the proton.

Data on identified particle production are also useful in studying the universality of jet fragmentation in e^+e^- and DIS. The strangeness suppression factor γ_s, as derived principally from tuning the Lund string model [80] within JETSET [70] is typically found to be around 0.3 in e^+e^- experiments [45], although values closer to 0.2 [98] have also been obtained. The converse is true for DIS experiments, with the tendency from HERA [94,96] and recent fixed target measurements [88] of light strange particle production, to support a stronger suppression ($\gamma_s \approx 0.2$), although values close to 0.3 have also been obtained [99].

However, when comparing the description of QCD-based models for lepton-hadron interactions and e^+e^- collisions, it is important to note that the overall description by event generators of inclusively produced hadronic final states is more precise in e^+e^- collisions than lepton-hadron interactions [100]. Predictions of particle rates in lepton-hadron scattering are affected by uncertainties in the modelling of the parton composition of the proton and photon, the extended target remnant, and initial and final

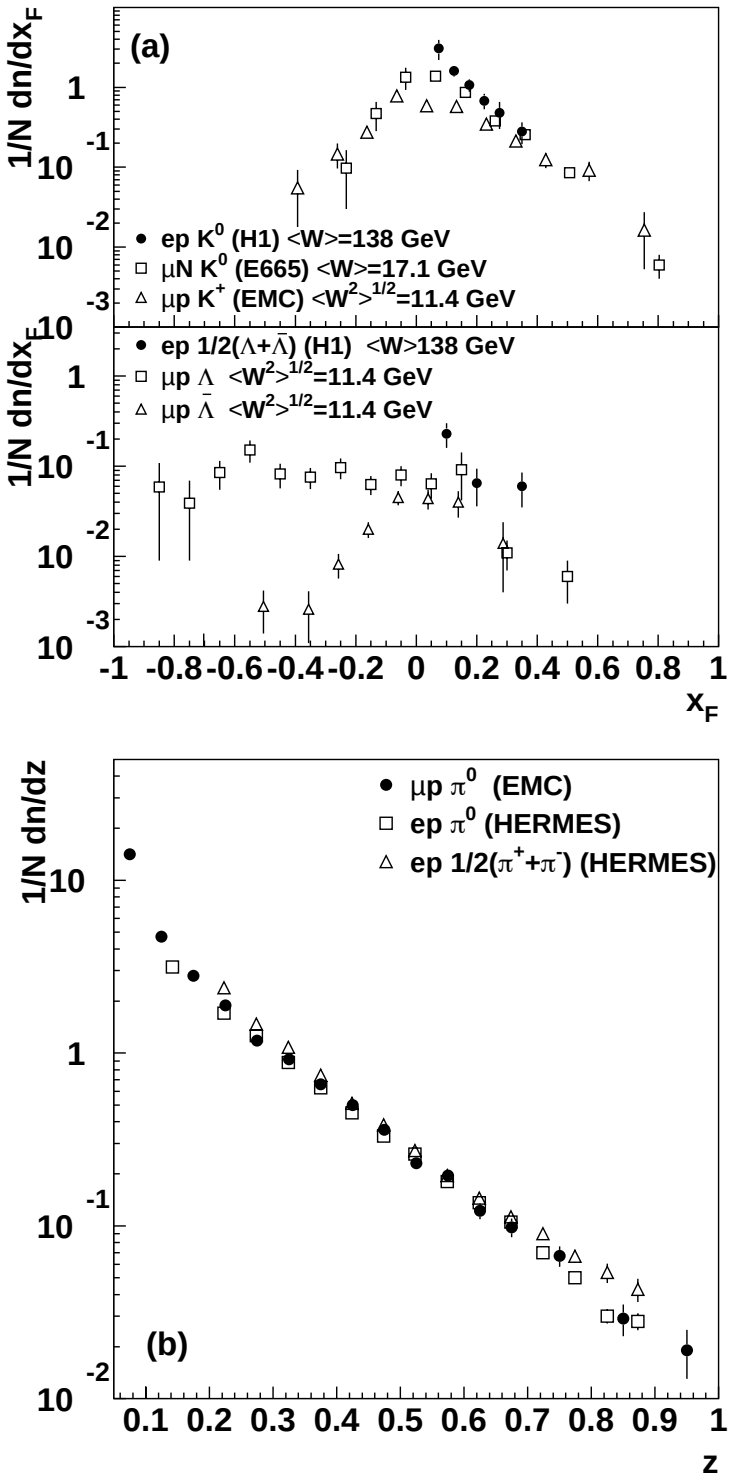


Figure 17.8: (a) $1/N{\cdot}\mathrm{d}n/\mathrm{d}x_F$ for identified strange particles in DIS at various values of W [88,91,94]. (b) $1/N{\cdot}\mathrm{d}n/\mathrm{d}z$ for measurements of pions from fixed target DIS experiment [89] [92,93].

state QCD radiation. Furthermore, the tuning of event generators for e^+e^- collisions is typically based on a larger set of parameters and use more observables [45] than are used when optimising models for lepton-hadron data [101].

17.8. Heavy quark fragmentation

It was recognized very early [102] that a heavy flavoured meson should retain a large fraction of the momentum of the primordial heavy quark, and therefore its fragmentation function should be much harder than that of a light hadron. In the limit of a very heavy quark, one expects the fragmentation function for a heavy quark to go into any heavy hadron to be peaked near 1.

When the heavy quark is produced at a momentum much larger than its mass, one expects important perturbative effects, enhanced by powers of the logarithm of the transverse momentum over the heavy quark mass, to intervene and modify the shape of the fragmentation function. In leading logarithmic order (*i.e.*, including all powers of $\alpha_S \log m_Q/p_T$) the total (*i.e.*, summed over all hadron types) perturbative fragmentation function is simply obtained by solving the leading evolution equation for fragmentation functions, Eq. (17.4), with the initial condition at a scale $\mu^2 = m_Q^2$ given by $D_Q(z, m_Q^2) = \delta(1-z)$ and $D_i(z, m_Q^2) = 0$ for $i \neq Q$ (the notation $D_i(z)$ stands for the probability to produce a heavy quark Q from parton i with a fraction z of the parton momentum).

Several extensions of the leading logarithmic result have appeared in the literature. Next-to-leading-log (NLL) order results for the perturbative heavy quark fragmentation function have been obtained in Ref. [103]. At large z, phase space for gluon radiation is suppressed. This exposes large perturbative corrections due to the incomplete cancellation of real gluon radiation and virtual gluon exchange (Sudakov effects), which should be resummed in order to get accurate results. A leading-log (LL) resummation formula has been obtained in [103,104]. Next-to-leading-log resummation has been performed in Ref. [105]. Fixed-order calculations of the fragmentation function at order α_S^2 in e^+e^- annihilation have appeared in Ref. [106]. This result does not include terms of order $(\alpha_S \log s/m^2)^k$ and $\alpha_S(\alpha_S \log s/m^2)^k$, but it does include correctly all terms up to the order α_S^2, including terms without any logarithmic enhancements. The result of Ref. [103] for the perturbative initial condition of the heavy quark fragmentation function has been extended to NNLO (next-to-next-to-leading order) in Ref. [107]. Other ingredients (*i.e.*, NNLO single inclusive production cross sections for light quarks, and NNLO evolution for fragmentation function) are however still missing for a full NNLO analysis of heavy flavour fragmentation functions.

Inclusion of non-perturbative effects in the calculation of the heavy quark fragmentation function is done in practice by convolving the perturbative result with a phenomenological non-perturbative form. Among the most popular parametrizations we have the following:

$$\text{Peterson}\textit{et al.}\ [108]: \quad D_{\rm np}(z) \propto \frac{1}{z}\left(1 - \frac{1}{z} - \frac{\epsilon}{1-z}\right)^{-2}, \tag{17.13}$$

$$\text{Kartvelishvili}\textit{et al.}\ [109]: \quad D_{\rm np}(z) \propto z^{\alpha}(1-z), \tag{17.14}$$

$$\text{Collins\&Spiller [110]}: \quad D_{\rm np}(z) \propto \left(\frac{1-z}{z} + \frac{(2-z)\epsilon_C}{1-z}\right) \times (1+z^2)\left(1 - \frac{1}{z} - \frac{\epsilon_C}{1-z}\right)^{-2} \tag{17.15}$$

$$\text{Colangelo\&Nason [111]}: \quad D_{\rm np}(z) \propto (1-z)^{\alpha} z^{\beta} \tag{17.16}$$

$$\text{Bowler [112]}: \quad D_{\rm np}(z) \propto z^{-(1+bm^2_{h,\perp})} (1-z)^{a} \exp\left(-\frac{bm^2_{h,\perp}}{z}\right) \tag{17.17}$$

$$\text{Braaten}\textit{et al.}\text{ [113]}: \quad \text{(see Eq. (31), (32) in [113])} \tag{17.18}$$

where ϵ, ϵ_C, a, $bm^2_{h,\perp}$, α, and β are non-perturbative parameters, depending upon the heavy hadron considered. In general, the non-perturbative parameters entering the non-perturbative forms do not have an absolute meaning. They are fitted together with some model of hard radiation, which can be either a shower Monte Carlo, a leading-log or NLL calculation (which may or may not include Sudakov resummation), or a fixed order calculation. In [106], for example, the Peterson [108] ϵ parameter for charm and bottom production is fitted from the measured distributions of refs. [114,115] for charm, and of [116] for bottom. If the leading-logarithmic approximation (LLA) is used for the perturbative part, one finds $\epsilon_c \approx 0.05$ and $\epsilon_b \approx 0.006$; if a second order calculation is used one finds $\epsilon_c \approx 0.035$ and $\epsilon_b \approx 0.0033$; if a NLLO calculation is used instead one finds $\epsilon_c \approx 0.022$ and $\epsilon_b \approx 0.0023$. The larger values found in the LL approximation are consistent with what is obtained in the context of parton shower models [117], as expected. The ϵ parameter for charm and bottom scales roughly with the inverse square of the heavy flavour mass. This behaviour can be justified by several arguments [102,118,119]. It can be used to relate the non-perturbative parts of the fragmentation functions of charm and bottom quarks [106,111,120].

17.8.1. ***Charm quark fragmentation*** : High statistics data for charmed mesons production near the Υ resonance (excluding decay products of B mesons) have been published [121,122]. They include results for D and D^*, D_s (see also [123,124]) and Λ_c. Shown in Fig. 17.9(a) are the CLEO and BELLE inclusive cross-sections times branching ratio $\mathcal{B}$, $s \cdot \mathcal{B} d\sigma/dx_p$, for the production of D^0 and D^{*+}. The variable x_p approximates the light-cone momentum fraction z in Eq. (17.13), but is not identical to it. The two measurements are fairly consistent with each other.

The branching ratio $\mathcal{B}$ represents $D^0 \to K^-\pi^+$ for the D^0 results and for the D^{*+} the product branching fraction: $D^{*+} \to D^0\pi^+$, $D^0 \to K^-\pi^+$. Older studies are reported in refs. [115,126]. Charmed meson spectra on the Z peak have been published by OPAL and ALEPH [83,127]. The relative production fractions of the various hadron species should be process independent at high energies according to QCD factorization. Combining results near the $\Upsilon(4S)$ from refs. [121,122,124,126,128], neglecting the gluon splitting contribution, which is negligible at these energies, we obtain $f(c \to D^0) = 0.565 \pm 0.032$, $f(c \to D^+) = 0.246 \pm 0.020$, $f(c \to D_s^+) = 0.080 \pm 0.017$,

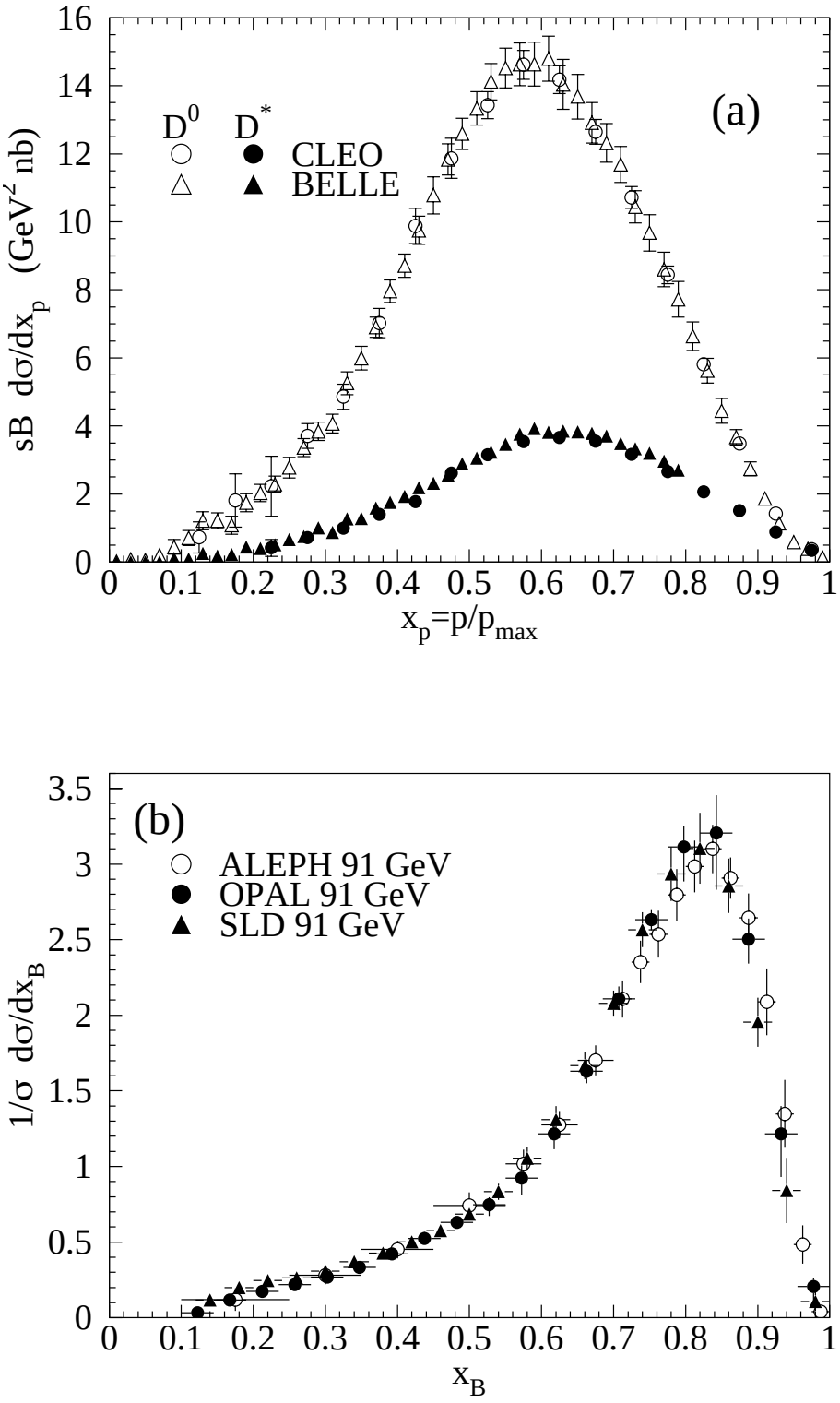


Figure 17.9: (a) Efficiency-corrected inclusive cross-section measurements for the production of D^0 and D^{*+} in e^+e^- measurements at $\sqrt{s} \approx 10.6$ GeV, excluding B decay products [121,122]. (b) Measured e^+e^- fragmentation function of b quarks into B hadrons at $\sqrt{s} \approx 91$ GeV [125].

$f(c \to \Lambda_c^+) = 0.094 \pm 0.035$, $f(c \to D^{*0}) = 0.213 \pm 0.024$, $f(c \to D^{*+}) = 0.224 \pm 0.028$, and $f(c \to D_s^{*+}) = 0.061 \pm 0.018$ in good agreement with those reported by LEP and SLD (see App. B of Ref. [129]) . Here, f is the fraction of produced charm quarks that hadronize into the respective hadron, eventually including decays of short-lived resonances.

As is well known, the large (isospin violating) difference in D^+ and D^0 production is well understood as a consequence of the fact that D^* mesons have a mass that is accidentally very near the sum of the D and the pion mass. The small isospin violating mass differences between states of different charge are such that both the D^{*+} and the D^{*0} can decay into the D^0 with large branching fractions, while only the D^{*+} can decay to a D^+, with a relatively small branching. If we can assume that similar accidents do not happen with higher resonances, we can conclude that D^{*+} and D^{*-} are produced with the same rate, and that D^0 and D^+ not coming from D^* decays are also produced with the same rate. The data are at present consistent with this view within errors, although it would seem to favour a $f(c \to D^{*+})$ larger than $f(c \to D^{*0})$. BELLE [122] publishes a ratio $(D^{*+} + D^{*0})/(D^+ + D^0) = 0.527 \pm 0.013 \pm 0.024$, so that the ratio of primary to total D is $0.473 \pm 0.013 \pm 0.024$.

Given the high precision of CLEO's and BELLE's data, it is difficult to obtain good fits of the inclusive cross-sections with the simple parametrizations generally used [108–112], see Ref. [122]. It is however still possible to obtain good fits to the data using relatively simple forms. In the context of a QCD calculation of the fragmentation function, including NLO initial condition, evolution and coefficient functions, and NLL resummation of Sudakov effects in the initial condition and in the coefficient functions, it was shown in Ref. [130] that a superposition of the form proposed in Ref. [111] and a delta function for the description of D^*'s and primary D mesons yields very good fits to CLEO and BELLE data. In the same work it is shown, however, that one cannot fit simultaneously CLEO/BELLE and ALEPH data in a pure perturbative QCD framework. This fact could be interpreted in terms of power corrections to the coefficient functions, with a suppression $1/Q^2$ (compatible with the result of Ref. [131]) or $1/Q$ (similar to the form proposed in Ref. [5]) .

Charm quark production has also been extensively studied at HERA by the H1 and ZEUS collaborations. Measurements have been made of $D^{*\pm}$, $D^{\pm}$ and $D_s^{\pm}$ mesons [132–135] and the Λ_c baryon [135]. Various fragmentation quantities have been extracted, some of which are shown in Table 17.1 as measured by H1 and ZEUS, along with averages of these quantites as obtained from e^+e^- data.

17.8.2. ***Bottom quark fragmentation*** : Experimental studies of the fragmentation function for b quarks, shown in Fig. 17.9(b), have been performed at LEP and SLD [116,125,137]. Commonly used methods identify the B meson through its semileptonic decay or based upon tracks emerging from the B secondary vertex. The most recent studies [125] fit the B spectrum using a Monte Carlo shower model supplemented with non-perturbative fragmentation functions yielding consistent results.

The experiments measure primarily the spectrum of B mesons. This defines a fragmentation function which includes the effect of the decay of higher mass excitations,

Table 17.1: Measurements of fragmentation ratios from ep [134,135] and e^+e^- experiments [136]. $R_{u/d}$ is ratio of neutral to charged D-mesons, γ_s is the strangeness suppression factor in charm fragmentation, and P_v^d is the fraction of charged D-mesons produced in a vector state.

	$R_{u/d}$
H1	1.26 ± 0.20 (stat.)± 0.11 (syst.)± 0.04 (br. $\oplus$ theo.)
ZEUS	1.100 ± 0.078 (stat.)$^{+0.038}_{-0.061}$ (syst.) $^{+0.047}_{-0.049}$ (br.)
e^+e^- av.	1.020 ± 0.069 (stat. $\oplus$ sys.) $^{+0.045}_{-0.047}$ (br.)
	γ_s
H1	0.36 ± 0.10 (stat.) ± 0.01 (syst.) ± 0.08 (br. $\oplus$ theo.)
ZEUS	0.257 ± 0.024 (stat.) $^{+0.013}_{-0.016}$ (syst.) $^{+0.078}_{-0.049}$ (br.)
e^+e^- av.	0.259 ± 0.023 (stat. $\oplus$ syst.) $^{+0.087}_{-0.052}$ (br.)
	P_v^d
H1	0.693 ± 0.045 (stat.)± 0.004 (syst.) ± 0.009 (br.$\oplus$ theo.)
ZEUS	0.566 ± 0.025 (stat.) $^{+0.007}_{-0.022}$ (syst.)$^{+0.022}_{-0.023}$ (br.)
e^+e^- av.	0.614 ± 0.019 (stat.$\oplus$syst.)$^{+0.023}_{-0.025}$ (br.)

like the B^* and B^{**}. In the literature there is sometimes ambiguity in what is defined to be the bottom fragmentation function. Instead of using what is directly measured (*i.e.*, the B meson spectrum) corrections are applied to account for B^* or B^{**} production in some cases. For a more detailed discussion see [1].

Heavy flavour production in e^+e^- collisions is the primary source of information for the role of fragmentation effects in heavy flavour production in hadron-hadron and lepton-hadron collisions. The QCD calculations tend to underestimate the data in certain regions of phase space. Recently, it was also pointed out [138] that the long-standing discrepancy between theoretical calculations and the measured B meson spectrum at the hadron colliders [139] was substantially reduced if a correct use of available information on heavy flavour production from e^+e^- data was made.

Both bottomed and charmed mesons spectra have been measured recently at the TEVATRON with unprecedented accuracy [140]. The measured spectra are in good agreement with QCD calculations (including non-perturbative fragmentation effects inferred from e^+e^- data [141]) , no longer supporting the previously reported discrepancies [139].

The HERA collaborations have produced a number of measurements of beauty production [133,142]. Compared with LEP data there is at present insufficient statistical

precision to make detailed measurements of the fragmentation properties of b-quarks in lepton-hadron scattering. The data which do exist have been tested against QCD-based models implementing Peterson *et al.* [108] fragmentation.

17.8.3. ***Gluon splitting into heavy quarks*** : Besides degrading the fragmentation function by gluon radiation, QCD evolution can also generate soft heavy quarks, increasing in the small x region as s increases. Several theoretical studies are available on the issue of how often $b\bar{b}$ or $c\bar{c}$ pairs are produced indirectly, via a gluon splitting mechanism [143–145]. Experimental results from studies on charm production via gluon splitting and measurements of $g \to b\bar{b}$ are given in Table 17.2.

Table 17.2: Measured fraction of events containing $g \to c\overline{c}$ and $g \to b\overline{b}$ subprocesses in Z decays, compared with theoretical predictions. The central/lower/upper values for the theoretical predictions are obtained with $m_c = (1.5 \pm 0.3)$ and $m_b = (4.75 \pm 0.25)\,\mathrm{GeV}$.

	$\overline{n}_{g\to c\overline{c}}$ (%)	$\overline{n}_{g\to b\overline{b}}$ (%)
ALEPH	[127] $3.26 \pm 0.23 \pm 0.42$	[146] $0.2777 \pm 0.042 \pm 0.057$
DELPHI		[147] $0.21 \pm 0.11 \pm 0.09$
L3	[148] $2.45 \pm 0.29 \pm 0.53$	
OPAL	[149] $3.20 \pm 0.21 \pm 0.38$	
SLD		[150] $0.307 \pm 0.071 \pm 0.066$
Theory [144]		
$\Lambda^{(5)}_{\overline{\mathrm{MS}}} = 150\,\mathrm{MeV}$	$1.35^{+0.48}_{-0.30}$	0.20 ± 0.02
$\Lambda^{(5)}_{\overline{\mathrm{MS}}} = 300\,\mathrm{MeV}$	$1.85^{+0.69}_{-0.44}$	0.26 ± 0.03

In Ref. [144] an explicit calculation of these quantities has been performed. Using these results, charm and bottom multiplicities as reported in Table 17.2 for different values of the masses and of $\Lambda^{(5)}_{\overline{\mathrm{MS}}}$ were computed in Ref. [151]. The averaged experimental result for charm, taking correlations into account, $(2.96 \pm 0.38)\%$ [152] is 1–2 standard deviations above the theoretical prediction, preferring lower values of the quark mass and/or a larger value of $\Lambda^{(5)}_{\overline{\mathrm{MS}}}$. However, higher-order corrections may well be substantial at the charm quark mass scale. Better agreement is achieved for bottom.

As reported in Ref. [144], Monte Carlo models are in qualitative agreement with these results, although the spread of the values they obtain is somewhat larger than the theoretical error estimated by the direct calculation. In particular, for charm one finds that while HERWIG [82] and JETSET [70] agree quite well with the theoretical calculation, ARIADNE [153] is higher by roughly a factor of 2, and thus is in better

agreement with data. For bottom, agreement between theory, models and data is adequate. For a detailed discussion see Ref. [154].

The discrepancy with the charm prediction may be due to experimental cuts forcing the final state configuration to be more 3-jet like, which increases the charm multiplicity. Calculations that take this possibility into account are given in Ref. [145].

References:

1. O. Biebel, P. Nason, and B.R. Webber, Bicocca-FT-01-20, Cavendish-HEP-01/12, MPI-PhE/2001-14, `hep-ph/0109282`.
2. W. Kittel, and E.A. De Wolf, *Soft Multihadron Dynamics*, World Scientific, Singapore (2005).
3. H.F. Jones, Nuovo Cimento **40A**, 1018 (1965);
   R.P. Feynman, *Photon Hadron Interacations*, W.A. Benjamin, New York (1972);
   Yu.L. Dokshitzer *et al.*, *Perturbative Quantum Chromodynamics*, ed. A.H. Mueller, World Scientific, Singapore (1989);
   K.H. Streng, Z. Phys. **C2**, 377 (1994).
4. V.N. Gribov and L.N. Lipatov, Sov. J. Nucl. Phys. **15**, 438 (1972);
   L.N. Lipatov, Sov. J. Nucl. Phys. **20**, 95 (1975);
   G. Altarelli and G. Parisi, Nucl. Phys. **B126**, 298 (1977);
   Yu.L. Dokshitzer, Sov. Phys. JETP **46**, 641 (1977).
5. P. Nason and B.R. Webber, Nucl. Phys. **B421**, 473 (1994),;
   Erratum *ibid.*, **B480**, 755 (1996).
6. W. Furmanski and R. Petronzio, preprint TH.2933–CERN (1980), Phys. Lett. **97B**, 437 (1980).
7. G. Curci *et al.*,Nucl. Phys. **B175**, 27 (1980);
   E.G. Floratos *et al.*,Nucl. Phys. **B192**, 417 (1981);
   J. Kalinowski *et al.*, Nucl. Phys. **B181**, 221 (1981);
   J. Kalinowski *et al.*, Nucl. Phys. **B181**, 253 (1981).
8. R.K. Ellis, J. Stirling, and B.R. Webber: *QCD and Collider Physics*, Cambridge University Press, Cambridge (1996).
9. ALEPH Collab.: D. Buskulic *et al.*, Z. Phys. **C73**, 409 (1997).
10. ALEPH Collab.: E. Barate *et al.*, Phys. Rev. **294**, 1 (1998).
11. AMY Collab.: Y.K. Li *et al.*, Phys. Rev. **D41**, 2675 (1990).
12. DELPHI Collab.: P. Abreu *et al.*, Eur. Phys. J. **C6**, 19 (1999).
13. HRS Collab.: D.Bender *et al.*, Phys. Rev. **D31**, 1 (1984).
14. L3 Collab.: B. Adeva *et al.*, Phys. Lett. **B259**, 199 (1991).
15. MARK II Collab.: G.S. Abrams *et al.*, Phys. Rev. Lett. **64**, 1334 (1990).
16. MARK II Collab.: A. Petersen *et al.*, Phys. Rev. **D37**, 1 (1988).
17. OPAL Collab.: R. Akers *et al.*, Z. Phys. **C72**, 191 (1996).
18. OPAL Collab.: K. Ackerstaff *et al.*, Z. Phys. **C75**, 193 (1997).
19. OPAL Collab.: K. Ackerstaff *et al.*, Eur. Phys. J. **C7**, 369 (1998).
20. OPAL Collab.: G. Abbiendi *et al.*, Eur. Phys. J. **C16**, 185 (2000);
    OPAL Collab.: G. Abbiendi *et al.*, Eur. Phys. J. **C27**, 467 (2003).
21. OPAL Collab.: G. Abbiendi *et al.*, Eur. Phys. J. **C37**, 25 (2004).
22. SLD Collab.: K. Abe *et al.*, Phys. Rev. **D69**, 072003 (2004).

23. TASSO Collab.: R. Brandelik *et al.*, Phys. Lett. **B114**, 65 (1982).
24. TASSO Collab.: W. Braunschweig *et al.*, Z. Phys. **C47**, 187 (1990).
25. TPC Collab.: H. Aihara *et al.*, Phys. Rev. Lett. **61**, 1263 (1988).
26. ALEPH Collab.: D. Barate *et al.*, Phys. Lett. **B357**, 487 (1995); Erratum *ibid.*, **B364**, 247 (1995).
27. DELPHI Collab.: P. Abreu *et al.*, Phys. Lett. **B311**, 408 (1993).
28. DELPHI Collab.: P. Abreu *et al.*, Phys. Lett. **B398**, 194 (1997).
29. DELPHI Collab.: P. Abreu *et al.*, Eur. Phys. J. **C13**, 573 (2000); W. de Boer and T. Kußmaul, IEKP-KA/93-8, `hep-ph/9309280`.
30. B.A. Kniehl, G. Kramer, and B. Pötter, Phys. Rev. Lett. **85**, 5288 (2001).
31. E665 Collab.: M.R. Adams *et al.*, Phys. Lett. **B272**, 163 (1991).
32. EMC Collab.: M. Arneodo *et al.*, Z. Phys. **C35**, 417 (1987).
33. H1 Collab.: I. Abt *et al.*, Z. Phys. **C63**, 377 (1994).
34. ZEUS Collab.: M. Derrick *et al.*, Z. Phys. **C70**, 1 (1996).
35. ZEUS Collab.: J. Breitweg *et al.*, Phys. Lett. **B414**, 428 (1997).
36. H1 Collab.: S. Aid *et al.*, Nucl. Phys. **B445**, 3 (1995).
37. H1 Collab.: C. Adloff *et al.*, Nucl. Phys. **B504**, 3 (1997).
38. ZEUS Collab.: M. Derrick *et al.*, Z. Phys. **C67**, 93 (1995).
39. ZEUS Collab.: J. Breitweg *et al.*, Eur. Phys. J. **C11**, 251 (1999).
40. MARK II Collab.: J.F. Patrick *et al.*, Phys. Rev. Lett. **49**, 1232, (1982).
41. D. Graudenz, CERN-TH/96-52.
42. P. Dixon *et al.*, J. Phys. **G25**, 1453 (1999).
43. Yu.L. Dokshitzer *et al.*, *Basics of Perturbative QCD*, Editions Frontières, Paris (1991).
44. C.P. Fong and B.R. Webber, Nucl. Phys. **B355**, 54 (1992).
45. DELPHI Collab.: P. Abreu *et al.*, Z. Phys. **C73**, 11 (1996).
46. DELPHI Collab.: P. Abreu *et al.*, Z. Phys. **C73**, 229 (1997).
47. L3 Collab.: P. Achard *et al.*, Phys. Reports **399**, 71 (2004).
48. TOPAZ Collab.: R. Itoh *et al.*, Phys. Lett. **B345**, 335 (1995).
49. OPAL Collab.: M.Z. Akrawy *et al.*, Phys. Lett. **B247**, 617 (1990).
50. TASSO Collab.: W. Braunschweig *et al.*, Z. Phys. **C22**, 307 (1990).
51. BES Collab.: J.Z. Bai *et al.*, Phys. Rev. **D69**, 072002 (2004).
52. ALEPH Collab.: D. Buskulic *et al.*, Z. Phys. **C55**, 209 (1992).
53. ALEPH Collab.: A. Heister *et al.*, Eur. Phys. J. **C35**, 457 (2004).
54. DELPHI Collab.: P. Abreu *et al.*, Phys. Lett. **B275**, 231 (1992).
55. DELPHI Collab.: P. Abreu *et al.*, Eur. Phys. J. **C5**, 585 (1998).
56. DELPHI Collab.: P. Abreu *et al.*, Phys. Lett. **B459**, 397 (1999).
57. L3 Collab.: M. Acciarri *et al.*, Phys. Lett. **B444**, 569 (1998).
58. TPC/TWO-GAMMA Collab.: H. Aihara *et al.*, "*Charged Hadron Production in e^+e^- Annihilation at $\sqrt{s} = 29$ GeV*", LBL 23737.
59. OPAL Collab.: R. Akers *et al.*, Z. Phys. **C86**, 203 (1995).
60. J. Binnewies, Hamburg University PhD Thesis, DESY 97-128, `hep-ph/9707269`.

61. ALEPH Collab.: R. Barate *et al.*, Eur. Phys. J. **C17**, 1 (2000);
OPAL Collab.: R. Akers *et al.*, Z. Phys. **C68**, 179 (1995);
OPAL Collab.: G. Abbiendi *et al.*, Eur. Phys. J. **C11**, 217 (1999).
62. COMPASS Collab.: G. Baum *et al.*, CERN-SPSLC-96-14;
CERN-SPSLC-P-297.
63. HERMES Collab.: A. Airapetian *et al.*, Phys. Rev. **D71**, 012003 (2005).
64. SMC Collab.: B. Adeva *et al.*, Phys. Lett. **B420**, 180, (1998).
65. J. Binnewies *et al.*, Phys. Rev. **D52**, 4947 (1995);
Z. Phys. **C65**, 471 (1995).
66. B.A. Kniehl *et al.*, Nucl. Phys. **B582**, 514 (2000).
67. S. Kretzer, Phys. Rev. **D62**, 054001 (2000).
68. L. Bourhis *et al.*, Eur. Phys. J. **C19**, 89 (2001).
69. S. Kretzer *et al.*, Eur. Phys. J. **C22**, 269 (2001).
70. T. Sjöstrand and M. Bengtsson, Comp. Phys. Comm. **43**, 367 (1987);
T. Sjöstrand, Comput. Phys. Commun. **82**, 74 (1994).
71. P.J. Mulders and R.D. Tangerman, Nucl. Phys. **B461**, 197 (1996);
Erratum *ibid.*, **B484**, 538 (1997).
72. R. Jacob, Nucl. Phys. **A711**, 35 (2002).
73. J.P. Ralston and D.E. Soper, Nucl. Phys. **B152**, 109 (1979).
74. J. Collins, Nucl. Phys. **B396**, 161 (1993).
75. CLAS Collab.: H. Avakian *et al.*, Phys. Rev. **D69**, 112004 (2004);
HERMES Collab.: A. Airapetian *et al.*, Phys. Rev. Lett. **84**, 4047 (2000);
HERMES Collab.: A. Airapetian *et al.*, Phys. Rev. **D64**, 097101 (2001).
76. HERMES Collab.: A. Airapetian *et al.*, Phys. Rev. Lett. **94**, 012002 (2005).
77. D. Sivers, Phys. Rev. **D43**, 261 (1991).
78. COMPASS Collab.: V.Y Alexakhin *et al.*, Phys. Rev. Lett. **94**, 202002 (2005).
79. X. Artru and G. Mennessier, Nucl. Phys. **B70**, 93 (1974).
80. B. Andersson *et al.*, Phys. Reports **97**, 31 (1983).
81. S. Chun and C. Buchanan, Phys. Reports **292**, 239 (1998).
82. G. Marchesini *et al.*, Comput. Phys. Commun. **67**, 465 (1992);
G. Corcella *et al.*, JHEP **0101**, 010 (2001).
83. OPAL Collab.: G. Alexander *et al.*, Z. Phys. **C69**, 543 (1996).
84. M. Schmelling, Phys. Script. **51**, 683 (1995).
85. D. Amati and G. Veneziano, Phys. Lett. **B83**, 87 (1979).
86. ALEPH Collab.: D. Buskulic *et al.*, Z. Phys. **C66**, 355 (1995);
ARGUS Collab.: H. Albrecht *et al.*, Z. Phys. **C44**, 547 (1989);
OPAL Collab.: R. Akers *et al.*, Z. Phys. **C63**, 181 (1994);
SLD Collab.: K. Abe *et al.*, Phys. Rev. **D59**, 052001 (1999).
87. B.A. Kniehl *et al.*, Nucl. Phys. **B597**, 337 (2001).
88. E665 Collab.: M.R. Adams *et al.*, Z. Phys. **C61**, 539 (1994).
89. EMC Collab.: J.J. Aubert *et al.*, Z. Phys. **C18**, 189 (1983);
EMC Collab.: M. Arneodo *et al.*, Phys. Lett. **B150**, 458 (1985).
90. EMC Collab.: M. Arneodo *et al.*, Z. Phys. **C33**, 167 (1986).
91. EMC Collab.: M. Arneodo *et al.*, Z. Phys. **C34**, 283 (1987).

92. HERMES Collab.: A. Airapetian *et al.*, Eur. Phys. J. **C21**, 599 (2001).
93. T.P. M$^{\rm c}$Pharlin *et al.*, Phys. Lett. **B90**, 479 (1980).
94. H1 Collab.: S. Aid *et al.*, Nucl. Phys. **B480**, 3 (1996).
95. H1 Collab.: C. Adloff *et al.*, Eur. Phys. J. **C18**, 293 (2000);
H1 Collab.: A. Aktas *et al.*, Eur. Phys. J. **C36**, 413 (2004).
96. ZEUS Collab.: M. Derrick *et al.*, Z. Phys. **C68**, 29 (1995);
ZEUS Collab.: J. Breitweg *et al.*, Eur. Phys. J. **C2**, 77 (1998).
97. ZEUS Collab.: S. Chekanov *et al.*, Phys. Lett. **B553**, 141 (2003).
98. OPAL Collab.: P.D. Acton *et al.*, Phys. Lett. **B305**, 407 (1993).
99. E632 Collab.: D. DeProspo *et al.*, Phys. Rev. **D50**, 6691 (1994).
100. G. Grindhammer *et al.*, in: Proceedings of the *Workshop on Monte Carlo Generators for HERA Physics*, Hamburg, Germany, 1998/1999.
101. N. Brook *et al.*, in: Proceedings of the *Workshop for Future HERA Physics at HERA*, Hamburg, Germany, 1996.
102. V.A. Khoze *et al.*, *Proceedings, Conference on High energy physics, Tbilisi 1976*;
J.D. Bjorken, Phys. Rev. **D17**, 171 (1978).
103. B. Mele and P. Nason, Phys. Lett. **B245**, 635 (1990);
Nucl. Phys. **B361**, 626 (1991).
104. Yu.L. Dokshitzer *et al.*, Phys. Rev. **D53**, 89 (1996).
105. M. Cacciari and S. Catani, CERN-TH/2001-174, UPRF-2001-11, `hep-ph/0107138`.
106. P. Nason and C. Oleari, Phys. Lett. **B418**, 199 (1998);
ibid., **B447**, 327 (1999);
Nucl. Phys. **B565**, 245 (2000).
107. K. Melnikov and A. Mitov, Phys. Rev. **D70**, 034027 (2004).
108. C. Peterson *et al.*, Phys. Rev. **D27**, 105 (1983).
109. V.G. Kartvelishvili *et al.*, Phys. Lett. **B78**, 615 (1978).
110. P. Collins and T. Spiller, J. Phys. **G11**, 1289 (1985).
111. G. Colangelo and P. Nason, Phys. Lett. **B285**, 167 (1992).
112. M.G. Bowler, Z. Phys. **C11**, 169 (1981).
113. E. Braaten *et al.*, Phys. Rev. **D51**, 4819 (1995).
114. OPAL Collab.: R. Akers *et al.*, Z. Phys. **C67**, 27 (1995).
115. ARGUS Collab.: H. Albrecht *et al.*, Z. Phys. **C52**, 353 (1991).
116. ALEPH Collab.: D. Buskulic *et al.*, Phys. Lett. **B357**, 699 (1995).
117. J. Chrin, Z. Phys. **C36**, 163 (1987).
118. R.L. Jaffe and L. Randall, Nucl. Phys. **B412**, 79 (1994);
P. Nason and B. Webber, Phys. Lett. **B395**, 355 (1997).
119. M. Cacciari and E. Gardi, Nucl. Phys. **B664**, 299 (2003).
120. L. Randall and N. Rius, Nucl. Phys. **B441**, 167 (1995).
121. CLEO Collab.: M. Artuso *et al.*, Phys. Rev. **D70**, 112001 (2004).
122. BELLE Collab.: R. Seuster *et al.*, Phys. Rev. **D73**, 032002 (206).
123. CLEO Collab.: R.A. Briere *et al.*, Phys. Rev. **D62**, 112003 (2000).
124. BABAR Collab.: B. Aubert *et al.*, Phys. Rev. **D65**, 0911004 (2002).
125. ALEPH Collab.: A. Heister *et al.*, Phys. Lett. **B512**, 30 (2001);
OPAL Collab.: G. Abbiendi *et al.*, Eur. Phys. J. **C29**, 463 (2003);

SLD Collab.: K. Abe *et al.*, Phys. Rev. **D65**, 092006 (2002);
Erratum *ibid.*, **D66**, 079905 (2002).
126. CLEO Collab.: D. Bortoletto *et al.*, Phys. Rev. **D37**, 1719 (1988).
127. ALEPH Collab.: R. Barate *et al.*, Phys. Lett. **B561**, 213 (2003).
128. ARGUS Collab.: H. Albrecht *et al.*, Z. Phys. **C54**, 1 (1992).
129. LEP Collab. and the LEP and SLD electroweak and heavy flavour working groups, `hep-ex/0412015`.
130. M. Cacciari *et al.*, `hep-ph/0510032`.
131. M. Dasgupta and B.R. Webber, Nucl. Phys. **B484**, 247 (1997).
132. H1 Collab.: C. Adloff *et al.*, Nucl. Phys. **B520**, 191 (2001);
H1 Collab.: C. Adloff *et al.*, Phys. Lett. **B528**, 199 (2002);
H1 Collab.: A. Aktas *et al.*, Eur. Phys. J. **C40**, 56 (2005);
H1 Collab.: A. Aktas *et al.*, `hep-ex/0507081`;
ZEUS Collab.: J. Breitweg *et al.*, Phys. Lett. **B481**, 213 (2000);
ZEUS Collab.: S. Chekanov *et al.*, Phys. Lett. **B545**, 244 (2002);
ZEUS Collab.: S. Chekanov *et al.*, Nucl. Phys. **B672**, 3 (2003);
ZEUS Collab.: S. Chekanov *et al.*, Phys. Lett. **B590**, 143 (2004);
ZEUS Collab.: S. Chekanov *et al.*, Phys. Rev. **D69**, 012004 (2004).
133. H1 Collab.: A. Aktas *et al.*, Phys. Lett. **B621**, 56 (2005).
134. H1 Collab.: A. Aktas *et al.*, Eur. Phys. J. **C38**, 447 (2005).
135. ZEUS Collab.: S. Chekanov *et al.*, DESY-05-147, `hep-ex/0508019`.
136. L. Gladillin, `hep-ex/9912064`.
137. L3 Collab.: B. Adeva *et al.*, Phys. Lett. **B261**, 177 (1991).
138. M. Cacciari and P. Nason, Phys. Rev. Lett. **89**, 122003 (2002).
139. CDF Collab.: F. Abe *et al.*, Phys. Rev. Lett. **71**, 500 (1993);
CDF Collab.: F. Abe *et al.*, Phys. Rev. Lett. **71**, 2396 (1993);
CDF Collab.: F. Abe *et al.*, Phys. Rev. **D50**, 4252 (1994);
CDF Collab.: F. Abe *et al.*, Phys. Rev. Lett. **75**, 1451 (1995);
CDF Collab.: D. Acosta *et al.*, Phys. Rev. **D66**, 032002 (2002);
CDF Collab.: D. Acosta *et al.*, Phys. Rev. **D65**, 052005 (2002);
D0 Collab.: S. Abachi *et al.*, Phys. Rev. Lett. **74**, 3548 (1995);
UA1 Collab.: C. Albajar *et al.*, Phys. Lett. **B186**, 237 (1987);
UA1 Collab.: C. Albajar *et al.*, Phys. Lett. **B256**, 121 (1991);
Erratum *ibid.*, **B272**, 497 (1991).
140. CDF Collab.: D. Acosta *et al.*, Phys. Rev. Lett. **91**, 241804 (2003);
CDF Collab.: D. Acosta *et al.*, Phys. Rev. **D71**, 032001 (2005).
141. M. Cacciari and P. Nason, JHEP **0309**, 006 (2003);
M. Cacciari *et al.*, JHEP **0407**, 033 (2004);
B.A. Kniehl *et al.*, DESY-05-146, MZ-TH-05-16, `hep-ph/0508129`.
142. H1 Collab.: C. Adloff *et al.*, Phys. Lett. **B467**, 156 (1999);
H1 Collab.: A. Aktas *et al.*, Eur. Phys. J. **C41**, 453 (2005);
ZEUS Collab.: J. Breitweg *et al.*, Eur. Phys. J. **C18**, 625 (2001);
ZEUS Collab.: S. Chekanov *et al.*, Phys. Lett. **B599**, 173 (2004).

143. A.H. Mueller and P. Nason, Nucl. Phys. **B266**, 265 (1986);
M.L. Mangano and P. Nason, Phys. Lett. **B285**, 160 (1992).
144. M.H. Seymour, Nucl. Phys. **B436**, 163 (1995).
145. D.J. Miller and M.H. Seymour, Phys. Lett. **B435**, 213 (1998).
146. ALEPH Collab.: R. Barate *et al.*, Phys. Lett. **B434**, 437 (1998).
147. DELPHI Collab.: P. Abreu *et al.*, Phys. Lett. **B405**, 202 (1997).
148. L3 Collab.: M. Acciarri *et al.*, Phys. Lett. **B476**, 243 (2000).
149. OPAL Collab.: G. Abbiendi *et al.*, Eur. Phys. J. **C13**, 1 (2000).
150. SLD Collab.: K. Abe *et al.*, SLAC-PUB-8157, `hep-ex/9908028`.
151. S. Frixione *et al.*, Heavy Quark Production, in A.J. Buras and M. Lindner (eds.), *Heavy flavours II*, World Scientific, Singapore (1998), `hep-ex/9702287`.
152. A. Giammanco, in: Proceedings of the *XII International Workshop on Deep Inelastic Scattering* DIS'2004, Štrbské Pleso, Slovakia.
153. L. Lönnblad, Comp. Phys. Comm. **71**, 15 (1992).
154. A. Ballestrero *et al.*, CERN-2000-09-B, `hep-ph/0006259`.

18. EXPERIMENTAL TESTS OF GRAVITATIONAL THEORY

Revised August 2005 by T. Damour (IHES, Bures-sur-Yvette, France).

Einstein's General Relativity, the current "standard" theory of gravitation, describes gravity as a universal deformation of the Minkowski metric:

$$g_{\mu\nu}(x^\lambda) = \eta_{\mu\nu} + h_{\mu\nu}(x^\lambda) \ , \text{ where } \eta_{\mu\nu} = \mathrm{diag}(-1,+1,+1,+1) \ . \tag{18.1}$$

Alternatively, it can be defined as the unique, consistent, local theory of a massless spin-2 field $h_{\mu\nu}$, whose source must then be the total, conserved energy-momentum tensor [1]. General Relativity is classically defined by two postulates. One postulate states that the Lagrangian density describing the propagation and self-interaction of the gravitational field is

$$\mathcal{L}_{\rm Ein}[g_{\mu\nu}] = \frac{c^4}{16\pi G_N}\sqrt{g}g^{\mu\nu}R_{\mu\nu}(g) \ , \tag{18.2}$$

$$R_{\mu\nu}(g) = \partial_\alpha\Gamma^\alpha_{\mu\nu} - \partial_\nu\Gamma^\alpha_{\mu\alpha} + \Gamma^\beta_{\alpha\beta}\Gamma^\alpha_{\mu\nu} - \Gamma^\beta_{\alpha\nu}\Gamma^\alpha_{\mu\beta} \ , \tag{18.3}$$

$$\Gamma^\lambda_{\mu\nu} = \tfrac{1}{2}g^{\lambda\sigma}(\partial_\mu g_{\nu\sigma} + \partial_\nu g_{\mu\sigma} - \partial_\sigma g_{\mu\nu}) \ , \tag{18.4}$$

where G_N is Newton's constant, $g = -\det(g_{\mu\nu})$, and $g^{\mu\nu}$ is the matrix inverse of $g_{\mu\nu}$. A second postulate states that $g_{\mu\nu}$ couples universally, and minimally, to all the fields of the Standard Model by replacing everywhere the Minkowski metric $\eta_{\mu\nu}$. Schematically (suppressing matrix indices and labels for the various gauge fields and fermions and for the Higgs doublet),

$$\begin{aligned}\mathcal{L}_{\rm SM}[\psi, A_\mu, H, g_{\mu\nu}] = &-\tfrac{1}{4}\sum\sqrt{g}g^{\mu\alpha}g^{\nu\beta}F^a_{\mu\nu}F^a_{\alpha\beta} \\ &-\sum\sqrt{g}\,\overline{\psi}\,\gamma^\mu D_\mu\psi \\ &-\tfrac{1}{2}\sqrt{g}g^{\mu\nu}\overline{D_\mu H}D_\nu H - \sqrt{g}\,V(H) \\ &-\sum\lambda\sqrt{g}\,\overline{\psi}\,H\psi \ ,\end{aligned} \tag{18.5}$$

where $\gamma^\mu\gamma^\nu + \gamma^\nu\gamma^\mu = 2g^{\mu\nu}$, and where the covariant derivative D_μ contains, besides the usual gauge field terms, a (spin dependent) gravitational contribution $\Gamma_\mu(x)$ [2]. From the total action $S_{\rm tot}[g_{\mu\nu}, \psi, A_\mu, H] = c^{-1}\int d^4x(\mathcal{L}_{\rm Ein} + \mathcal{L}_{\rm SM})$ follow Einstein's field equations,

$$R_{\mu\nu} - \tfrac{1}{2}Rg_{\mu\nu} = \frac{8\pi G_N}{c^4}T_{\mu\nu} \ . \tag{18.6}$$

Here $R = g^{\mu\nu}R_{\mu\nu}$, $T_{\mu\nu} = g_{\mu\alpha}g_{\nu\beta}T^{\alpha\beta}$, and $T^{\mu\nu} = (2/\sqrt{g})\delta\mathcal{L}_{\rm SM}/\delta g_{\mu\nu}$ is the (symmetric) energy-momentum tensor of the Standard Model matter. The theory is invariant under arbitrary coordinate transformations: $x'^\mu = f^\mu(x^\nu)$. To solve the field equations Eq. (18.6) one needs to fix this coordinate gauge freedom. E.g. the "harmonic gauge" (which is the analogue of the Lorentz gauge, $\partial_\mu A^\mu = 0$, in electromagnetism) corresponds to imposing the condition $\partial_\nu(\sqrt{g}g^{\mu\nu}) = 0$.

In this *Review*, we only consider the classical limit of gravitation (*i.e.* classical matter and classical gravity). Considering quantum matter in a classical gravitational

background already poses interesting challenges, notably the possibility that the zero-point fluctuations of the matter fields generate a nonvanishing vacuum energy density $\rho_{\rm vac}$, corresponding to a term $-\sqrt{g}\ \rho_{\rm vac}$ in $\mathcal{L}_{\rm SM}$ [3]. This is equivalent to adding a "cosmological constant" term $+\Lambda\, g_{\mu\nu}$ on the left-hand side of Einstein's equations Eq. (18.6), with $\Lambda = 8\pi G_N\, \rho_{\rm vac}/c^4$. Recent cosmological observations (see the following *Reviews*) suggest a positive value of Λ corresponding to $\rho_{\rm vac} \approx (2.3 \times 10^{-3}{\rm eV})^4$. Such a small value has a negligible effect on the tests discussed below. Quantizing the gravitational field itself poses a very difficult challenge because of the perturbative non-renormalizability of Einstein's Lagrangian. Superstring theory offers a promising avenue toward solving this challenge.

18.1. Experimental tests of the coupling between matter and gravity

The universality of the coupling between $g_{\mu\nu}$ and the Standard Model matter postulated in Eq. (18.5) ("Equivalence Principle") has many observable consequences. First, it predicts that the outcome of a local non-gravitational experiment, referred to local standards, does not depend on where, when, and in which locally inertial frame, the experiment is performed. This means, for instance, that local experiments should neither feel the cosmological evolution of the universe (constancy of the "constants"), nor exhibit preferred directions in spacetime (isotropy of space, local Lorentz invariance). These predictions are consistent with many experiments and observations. The best limit on a possible time variation of the basic coupling constants concerns the fine-structure constant $\alpha_{\rm em}$ and has been obtained by analyzing a natural fission reactor phenomenon which took place at Oklo, Gabon, two billion years ago [4]. A conservative estimate of the (95% C.L.) Oklo limit on the variability of $\alpha_{\rm em}$ is (see second reference in [4])

$$-0.9 \times 10^{-7} < \frac{\alpha_{\rm em}^{\rm Oklo} - \alpha_{\rm em}^{\rm now}}{\alpha_{\rm em}} < 1.2 \times 10^{-7}\,, \tag{18.7}$$

which corresponds to the following limit on the average time derivative of $\alpha_{\rm em}$

$$-6.7 \times 10^{-17}{\rm yr}^{-1} < \dot{\alpha}_{\rm em}/\alpha_{\rm em} < 5.0 \times 10^{-17}{\rm yr}^{-1}\ . \tag{18.8}$$

Direct laboratory limits on the time variation of $\alpha_{\rm em}$ [5] are less stringent than Eq. (18.8). Recently, Ref. 13 claimed to have measured a cosmological change of $\alpha_{\rm em}$ of magnitude $\Delta\alpha_{\rm em}/\alpha_{\rm em} = (-0.72 \pm 0.18) \times 10^{-5}$ around redshifts $z \simeq 0.5 - 3.5$. When analyzed within various dilaton-like theoretical models of $\alpha_{\rm em}$ variability [7,8,9] such a large cosmological variation of $\alpha_{\rm em}$ appears incompatible with the combined set of other experimental limits, notably the Oklo bound quoted above, a comparable "rhenium" geological bound [10], and the limits on possible violations of the universality of free fall quoted below. See [11,12] for discussions of systematic errors in astronomical measurements of $\alpha_{\rm em}$ and [13] for a general review of the issue of "variable constants".

The highest precision tests of the isotropy of space have been performed by looking for possible quadrupolar shifts of nuclear energy levels [14]. The (null) results can be

interpreted as testing the fact that the various pieces in the matter Lagrangian Eq. (18.5) are indeed coupled to one and the same external metric $g_{\mu\nu}$ to the 10^{-27} level. For astrophysical constraints on possible Planck-scale violations of Lorentz invariance see [15,16].

The universal coupling to $g_{\mu\nu}$ postulated in Eq. (18.5) implies that two (electrically neutral) test bodies dropped at the same location and with the same velocity in an external gravitational field fall in the same way, independently of their masses and compositions. The universality of the acceleration of free fall has been verified at the 10^{-12} level both for laboratory bodies [17],

$$(\Delta a/a)_{\rm BeCu} = (-1.9 \pm 2.5) \times 10^{-12} \ , \tag{18.9}$$

and for the gravitational accelerations of the Moon and the Earth toward the Sun [18],

$$(\Delta a/a)_{\rm MoonEarth} = (-3.2 \pm 4.6) \times 10^{-13} \ . \tag{18.10}$$

See also Ref. 19 for a $\pm 0.94 \times 10^{-12}$ limit on the fractional difference in free-fall acceleration toward the Sun of earth-core-like and moon-mantle-like laboratory bodies and Ref. 20 for *short-range* tests of the universality of free-fall.

Finally, Eq. (18.5) also implies that two identically constructed clocks located at two different positions in a static external Newtonian potential $U(\boldsymbol{x}) = \sum G_N m/r$ exhibit, when intercompared by means of electromagnetic signals, the (apparent) difference in clock rate,

$$\frac{\tau_1}{\tau_2} = \frac{\nu_2}{\nu_1} = 1 + \frac{1}{c^2}[U(\boldsymbol{x}_1) - U(\boldsymbol{x}_2)] + O\left(\frac{1}{c^4}\right) \ , \tag{18.11}$$

independently of their nature and constitution. This universal gravitational redshift of clock rates has been verified at the 10^{-4} level by comparing a hydrogen-maser clock flying on a rocket up to an altitude $\sim 10,000$ km to a similar clock on the ground [21]. For more details and references on experimental gravity see, *e.g.*, Refs. 22 and 23.

18.2. Tests of the dynamics of the gravitational field in the weak field regime

The effect on matter of one-graviton exchange, *i.e.* the interaction Lagrangian obtained when solving Einstein's field equations Eq. (18.6) written in, say, the harmonic gauge at first order in $h_{\mu\nu}$,

$$\Box h_{\mu\nu} = -\frac{16\pi G_N}{c^4}(T_{\mu\nu} - \tfrac{1}{2}T\eta_{\mu\nu}) + O(h^2) + O(hT) \ , \tag{18.12}$$

reads $-(8\pi G_N/c^4)T^{\mu\nu}\Box^{-1}(T_{\mu\nu} - \frac{1}{2}T\eta_{\mu\nu})$. For a system of N moving point masses, with free Lagrangian $L^{(1)} = \sum_{A=1}^{N} - m_A c^2\sqrt{1 - \boldsymbol{v}_A^2/c^2}$, this interaction, expanded to order

v^2/c^2, reads (with $r_{AB} \equiv |\boldsymbol{x}_A - \boldsymbol{x}_B|$, $\boldsymbol{n}_{AB} \equiv (\boldsymbol{x}_A - \boldsymbol{x}_B)/r_{AB}$)

$$L^{(2)} = \frac{1}{2} \sum_{A \neq B} \frac{G_N \, m_A \, m_B}{r_{AB}} \Bigg[1 + \frac{3}{2c^2}(\boldsymbol{v}_A^2 + \boldsymbol{v}_B^2) - \frac{7}{2c^2}(\boldsymbol{v}_A \cdot \boldsymbol{v}_B) - \frac{1}{2c^2}(\boldsymbol{n}_{AB} \cdot \boldsymbol{v}_A)(\boldsymbol{n}_{AB} \cdot \boldsymbol{v}_B) + O\left(\frac{1}{c^4}\right) \Bigg] \, . \tag{18.13}$$

The two-body interactions Eq. (18.13) exhibit v^2/c^2 corrections to Newton's $1/r$ potential induced by spin-2 exchange. Consistency at the "post-Newtonian" level $v^2/c^2 \sim G_N \, m/rc^2$ requires that one also considers the three-body interactions induced by some of the three-graviton vertices and other nonlinearities (terms $O(h^2)$ and $O(hT)$ in Eq. (18.12)),

$$L^{(3)} = -\frac{1}{2} \sum_{B \neq A \neq C} \frac{G_N^2 \, m_A \, m_B \, m_C}{r_{AB} \, r_{AC} \, c^2} + O\left(\frac{1}{c^4}\right) \, . \tag{18.14}$$

All currently performed gravitational experiments in the solar system, including perihelion advances of planetary orbits, the bending and delay of electromagnetic signals passing near the Sun, and very accurate ranging data to the Moon obtained by laser echoes, are compatible with the post-Newtonian results Eqs. (18.12)–(18.14).

Similarly to what is done in discussions of precision electroweak experiments (see Section 10 in this *Review*), it is useful to quantify the significance of precision gravitational experiments by parameterizing plausible deviations from General Relativity. The addition of a mass-term in Einstein's field equations leads to a score of theoretical difficulties [24] which have not yet received any consensual solution. We shall therefore not consider here the ill-defined "mass of the graviton" as a possible deviation parameter from General Relativity (see, however, the phenomenological limits quoted in the Section "Gauge and Higgs Bosons" of this *Review*). Deviations from Einstein's pure spin-2 theory are then defined by adding new, bosonic light or massless, macroscopically coupled fields. The possibility of new gravitational-strength couplings leading (on small, and possibly large, scales) to deviations from Einsteinian (and Newtonian) gravity is suggested by String Theory [25], and by Brane World ideas [26]. For compilations of experimental constraints on Yukawa-type additional interactions see Refs. [17,27,28] and the Section "Extra Dimensions" in this *Review*. Recent experiments have set limits on non-Newtonian forces below 0.1 mm [29].

Here, we shall focus on the parametrization of long-range deviations from relativistic gravity obtained by adding a massless scalar field φ coupled to the trace of the energy-momentum tensor $T = g_{\mu\nu} T^{\mu\nu}$ [30]. The most general such theory contains an arbitrary function $a(\varphi)$ of the scalar field, and can be defined by the Lagrangian

$$\mathcal{L}_{\rm tot}[g_{\mu\nu}, \varphi, \psi, A_\mu, H] = \frac{c^4}{16\pi G} \sqrt{g} (R(g) - 2 g^{\mu\nu} \partial_\mu \varphi \partial_\nu \varphi) + \mathcal{L}_{\rm SM}[\psi, A_\mu, H, \widetilde{g}_{\mu\nu}] \, , \tag{18.15}$$

where G is a "bare" Newton constant, and where the Standard Model matter is coupled not to the "Einstein" (pure spin-2) metric $g_{\mu\nu}$, but to the conformally related ("Jordan-Fierz") metric $\widetilde{g}_{\mu\nu} = \exp(2a(\varphi))g_{\mu\nu}$. The scalar field equation $\Box_g\varphi = -(4\pi G/c^4)\alpha(\varphi)T$ displays $\alpha(\varphi) \equiv \partial a(\varphi)/\partial\varphi$ as the basic (field-dependent) coupling between φ and matter [31]. The one-parameter (ω) Jordan-Fierz-Brans-Dicke theory [30] is the special case $a(\varphi) = \alpha_0\varphi$ leading to a field-independent coupling $\alpha(\varphi) = \alpha_0$ (with ${\alpha_0}^2 = 1/(2\omega+3)$).

In the weak field, slow motion, limit appropriate to describing gravitational experiments in the solar system, the addition of φ modifies Einstein's predictions only through the appearance of two "post-Einstein" dimensionless parameters: $\overline{\gamma} = -2\alpha_0^2/(1+\alpha_0^2)$ and $\overline{\beta} = +\frac{1}{2}\beta_0\alpha_0^2/(1+\alpha_0^2)^2$, where $\alpha_0 \equiv \alpha(\varphi_0)$, $\beta_0 \equiv \partial\alpha(\varphi_0)/\partial\varphi_0$, φ_0 denoting the vacuum expectation value of φ. These parameters show up also naturally (in the form $\gamma_{\rm PPN} = 1+\overline{\gamma}$, $\beta_{\rm PPN} = 1+\overline{\beta}$) in phenomenological discussions of possible deviations from General Relativity [32,22]. The parameter $\overline{\gamma}$ measures the admixture of spin 0 to Einstein's graviton, and contributes an extra term $+\overline{\gamma}(\boldsymbol{v}_A - \boldsymbol{v}_B)^2/c^2$ in the square brackets of the two-body Lagrangian Eq. (18.13). The parameter $\overline{\beta}$ modifies the three-body interaction Eq. (18.14) by a factor $1+2\overline{\beta}$. Moreover, the combination $\eta \equiv 4\overline{\beta} - \overline{\gamma}$ parameterizes the lowest order effect of the self-gravity of orbiting masses by modifying the Newtonian interaction energy terms in Eq. (18.13) into $G_{AB}m_Am_B/r_{AB}$, with a body-dependent gravitational "constant" $G_{AB} = G_N[1+\eta(E_A^{\rm grav}/m_Ac^2 + E_B^{\rm grav}/m_Bc^2) + O(1/c^4)]$, where $G_N = G\exp[2a(\varphi_0)](1+\alpha_0^2)$ and where $E_A^{\rm grav}$ denotes the gravitational binding energy of body A.

The best current limits on the post-Einstein parameters $\overline{\gamma}$ and $\overline{\beta}$ are (at the 68% confidence level): (i) $\overline{\gamma} = (2.1 \pm 2.3) \times 10^{-5}$ deduced from the additional Doppler shift experienced by radio-wave beams connecting the Earth to the Cassini spacecraft when they passed near the Sun [33], and (ii) $4\overline{\beta} - \overline{\gamma} = -0.0007 \pm 0.0010$ [18] from Lunar Laser Ranging measurements of a possible polarization of the Moon toward the Sun [34]. More stringent limits on $\overline{\gamma}$ are obtained in models (*e.g.*, string-inspired ones [35]) where scalar couplings violate the Equivalence Principle.

18.3. Tests of the dynamics of the gravitational field in the radiative and/or strong field regimes

The discovery of pulsars (*i.e.* rotating neutron stars emitting a beam of radio noise) in gravitationally bound orbits [36,37] has opened up an entirely new testing ground for relativistic gravity, giving us an experimental handle on the regime of radiative and/or strong gravitational fields. In these systems, the finite velocity of propagation of the gravitational interaction between the pulsar and its companion generates damping-like terms at order $(v/c)^5$ in the equations of motion [38]. These damping forces are the local counterparts of the gravitational radiation emitted at infinity by the system ("gravitational radiation reaction"). They cause the binary orbit to shrink and its orbital period P_b to decrease. The remarkable stability of the pulsar clock has allowed Taylor and collaborators to measure the corresponding very small orbital period decay $\dot{P}_b \equiv dP_b/dt \sim -(v/c)^5 \sim -10^{-12}$ [37,39,40], thereby giving us a direct experimental

confirmation of the propagation properties of the gravitational field, and, in particular, an experimental confirmation that the speed of propagation of gravity is equal to the velocity of light to better than a part in a thousand [41]. In addition, the surface gravitational potential of a neutron star $h_{00}(R) \simeq 2Gm/c^2R \simeq 0.4$ being a factor $\sim 10^8$ higher than the surface potential of the Earth, and a mere factor 2.5 below the black hole limit ($h_{00} = 1$), pulsar data are sensitive probes of the strong-gravitational-field regime.

Binary pulsar timing data record the times of arrival of successive electromagnetic pulses emitted by a pulsar orbiting around the center of mass of a binary system. After correcting for the Earth motion around the Sun and for the dispersion due to propagation in the interstellar plasma, the time of arrival of the Nth pulse t_N can be described by a generic, parameterized "timing formula" [42] whose functional form is common to the whole class of tensor-scalar gravitation theories:

$$t_N - t_0 = F[T_N(\nu_p, \dot{\nu}_p, \ddot{\nu}_p)\,;\,\{p^K\}\,;\,\{p^{PK}\}]\ . \tag{18.16}$$

Here, T_N is the pulsar proper time corresponding to the Nth turn given by $N/2\pi = \nu_p T_N + \frac{1}{2}\dot{\nu}_p T_N^2 + \frac{1}{6}\ddot{\nu}_p T_N^3$ (with $\nu_p \equiv 1/P_p$ the spin frequency of the pulsar, etc.), $\{p^K\} = \{P_b, T_0, e, \omega_0, x\}$ is the set of "Keplerian" parameters (notably, orbital period P_b, eccentricity e and projected semi-major axis $x = a \sin i/c$), and $\{p^{PK}\} = \{k, \gamma_{\rm timing}, \dot{P}_b, r, s, \delta_\theta, \dot{e}, \dot{x}\}$ denotes the set of (separately measurable) "post-Keplerian" parameters. Most important among these are: the fractional periastron advance per orbit $k \equiv \dot{\omega} P_b/2\pi$, a dimensionful time-dilation parameter $\gamma_{\rm timing}$, the orbital period derivative $\dot{P}_b$, and the "range" and "shape" parameters of the gravitational time delay caused by the companion, r and s.

Without assuming any specific theory of gravity, one can phenomenologically analyze the data from any binary pulsar by least-squares fitting the observed sequence of pulse arrival times to the timing formula Eq. (18.16). This fit yields the "measured" values of the parameters $\{\nu_p, \dot{\nu}_p, \ddot{\nu}_p\}$, $\{p^K\}$, $\{p^{PK}\}$. Now, each specific relativistic theory of gravity predicts that, for instance, k, $\gamma_{\rm timing}$, $\dot{P}_b$, r and s (to quote parameters that have been successfully measured from some binary pulsar data) are some theory-dependent functions of the Keplerian parameters and of the (unknown) masses m_1, m_2 of the pulsar and its companion. For instance, in General Relativity, one finds (with $M \equiv m_1 + m_2$, $n \equiv 2\pi/P_b$)

$$\begin{aligned}
k^{\rm GR}(m_1, m_2) =& 3(1-e^2)^{-1}(G_N M n/c^3)^{2/3}\ ,\\
\gamma^{\rm GR}_{\rm timing}(m_1, m_2) =& e n^{-1}(G_N M n/c^3)^{2/3} m_2(m_1 + 2m_2)/M^2\ ,\\
\dot{P}_b^{\rm GR}(m_1, m_2) =& -(192\pi/5)(1-e^2)^{-7/2}\left(1 + \tfrac{73}{24}e^2 + \tfrac{37}{96}e^4\right)\\
&\times (G_N M n/c^3)^{5/3} m_1 m_2/M^2\ ,\\
r(m_1, m_2) =& G_N m_2/c^3\ ,\\
s(m_1, m_2) =& n x (G_N M n/c^3)^{-1/3} M/m_2\ .
\end{aligned} \tag{18.17}$$

In tensor-scalar theories, each of the functions $k^{\rm theory}(m_1,m_2)$, $\gamma^{\rm theory}_{\rm timing}(m_1,m_2)$, $\dot{P}^{\rm theory}_b(m_1,m_2)$, etc is modified by quasi-static strong field effects (associated with the self-gravities of the pulsar and its companion), while the particular function $\dot{P}^{\rm theory}_b(m_1,m_2)$ is further modified by radiative effects (associated with the spin 0 propagator) [31,43].

Let us summarize the current experimental situation (see [44] for a more extensive review). In the first discovered binary pulsar PSR1913 + 16 [36,37], it has been possible to measure with accuracy the three post-Keplerian parameters k, $\gamma_{\rm timing}$ and $\dot{P}_b$. The three equations $k^{\rm measured} = k^{\rm theory}(m_1,m_2)$, $\gamma^{\rm measured}_{\rm timing} = \gamma^{\rm theory}_{\rm timing}(m_1,m_2)$, $\dot{P}^{\rm measured}_b = \dot{P}^{\rm theory}_b(m_1,m_2)$ determine, for each given theory, three curves in the two-dimensional mass plane. This yields *one* (combined radiative/strong-field) test of the specified theory, according to whether the three curves meet at one point, as they should. After subtracting a small ($\sim 10^{-14}$ level in $\dot{P}^{\rm obs}_b = (-2.4211 \pm 0.0014) \times 10^{-12}$), but significant, Newtonian perturbing effect caused by the Galaxy [45], one finds that General Relativity passes this $(k - \gamma_{\rm timing} - \dot{P}_b)_{1913+16}$ test with complete success at the 10^{-3} level [37,39,40]

$$\left[\frac{\dot{P}^{\rm obs}_b - \dot{P}^{\rm galactic}_b}{\dot{P}^{\rm GR}_b[k^{\rm obs},\gamma^{\rm obs}_{\rm timing}]}\right]_{1913+16} = 1.0026 \pm 0.0006({\rm obs}) \pm 0.0021({\rm galactic})$$
$$= 1.0026 \pm 0.0022\ . \qquad (18.18)$$

Here $\dot{P}^{\rm GR}_b[k^{\rm obs},\gamma^{\rm obs}_{\rm timing}]$ is the result of inserting in $\dot{P}^{\rm GR}_b(m_1,m_2)$ the values of the masses predicted by the two equations $k^{\rm obs} = k^{\rm GR}(m_1,m_2)$, $\gamma^{\rm obs}_{\rm timing} = \gamma^{\rm GR}_{\rm timing}(m_1,m_2)$. This experimental evidence for the reality of gravitational radiation damping forces at the 0.3% level is illustrated in Fig. 18.1, which shows actual orbital phase data (after subtraction of a linear drift).

The discovery of the binary pulsar PSR1534 + 12 [46] has allowed one to measure the four post-Keplerian parameters k, $\gamma_{\rm timing}$, r and s, and thereby to obtain *two* (four observables minus two masses) tests of strong field gravity, without mixing of radiative effects [47]. General Relativity passes these tests within the measurement accuracy [47,37]. The most precise of these new, pure, strong-field tests is the one obtained by combining the measurements of k, γ, and s. Using the most recent data [48], one finds agreement at the 1% level:

$$\left[\frac{s^{\rm obs}}{s^{\rm GR}[k^{\rm obs},\gamma^{\rm obs}_{\rm timing}]}\right]_{1534+12} = 1.000 \pm 0.007\ . \qquad (18.19)$$

It has also been possible to measure the orbital period change of PSR1534 + 12. General Relativity passes the corresponding $(k - \gamma_{\rm timing} - \dot{P}_b)_{1534+12}$ test with success at the 15% level [48].

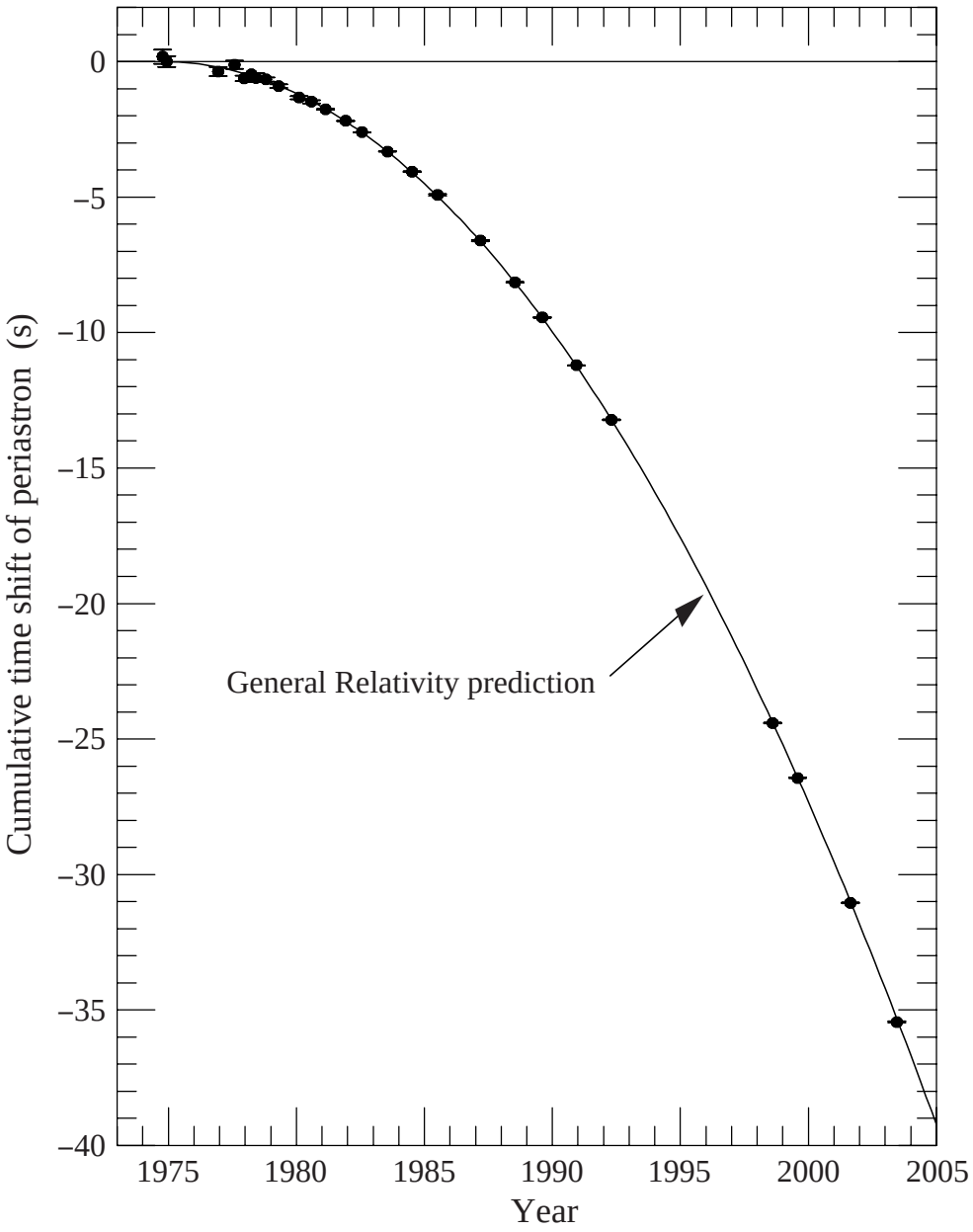


Figure 18.1: Accumulated shift of the times of periastron passage in the PSR 1913+16 system, relative to an assumed orbit with a constant period. The parabolic curve represents the general relativistic prediction, modified by Galactic effects, for orbital period decay from gravitational radiation damping forces. (Figure obtained with permission from Ref. 40.)

The discovery of the binary pulsar PSR J1141 − 6545 [49] (whose companion is probably a white dwarf) has recently led to the measurement of the three post-Keplerian parameters k, $\gamma_{\rm timing}$ and $\dot{P}_b$ [50]. As in the PSR $1913+16$ system this yields *one* combined radiative/strong-field test of relativistic gravity. One finds that General Relativity passes this $(k - \gamma_{\rm timing} - \dot{P}_b)_{1141-6545}$ test with success at the 25% level [50].

Several other binary pulsar systems, of a nonsymmetric type (nearly circular systems made of a neutron star and a white dwarf), can also be used to test relativistic gravity [51,52,53,54]. The constraints on tensor-scalar theories provided by three binary-pulsar "experiments" have been analyzed in [43] and shown to exclude a large portion of the parameter space allowed by solar-system tests. Measurements of the pulse shape of PSR1913 + 16 [55] have detected a time variation of the pulse shape compatible with the prediction [56] that the general relativistic spin-orbit coupling should cause a secular change in the orientation of the pulsar beam with respect to the line of sight ("geodetic precession").

The tests considered above have examined the gravitational interaction on scales between a fraction of a millimeter and a few astronomical units. On the other hand, the general relativistic action on light and matter of an external gravitational field on a length scale ~ 100 kpc has been verified to $\sim 30\%$ in some gravitational lensing systems (see, *e.g.*, Ref. 57). Some tests on cosmological scales are also available. In particular, Big Bang Nucleosynthesis (see Section 20 of this *Review*) has been used to set significant constraints on the variability of the gravitational "constant" [58]. For other cosmological tests of the "constancy of constants" see the review [13].

18.4. Conclusions

All present experimental tests are compatible with the predictions of the current "standard" theory of gravitation: Einstein's General Relativity. The universality of the coupling between matter and gravity (Equivalence Principle) has been verified at the 10^{-12} level. Solar system experiments have tested all the weak-field predictions of Einstein's theory at better than the 10^{-3} level (and down to the 2×10^{-5} level for the post-Einstein parameter $\overline{\gamma}$). The propagation properties of relativistic gravity, as well as several of its strong-field aspects, have been verified at the 10^{-3} level in binary pulsar experiments. Recent laboratory experiments have set strong constraints on sub-millimeter modifications of Newtonian gravity.

Several important new developments in experimental gravitation are expected in the near future. The approved NASA Gravity Probe B mission [59](a space gyroscope experiment; due for launch in December 2003) will directly measure the gravitational spin-orbit and spin-spin couplings, thereby measuring the parameter $\overline{\gamma}$ to better than the 10^{-5} level. The universality of free-fall acceleration should soon be tested to much better than the 10^{-12} level by some satellite experiments: the approved CNES MICROSCOPE [60] mission (10^{-15} level), and the planned (cryogenic) NASA-ESA STEP [61] mission (10^{-18} level). The recently constructed kilometer-size laser interferometers (notably LIGO [62] in the USA and VIRGO [63] and GEO600 [64] in Europe) should soon directly detect gravitational waves arriving on Earth. As the sources of these waves are expected to be extremely relativistic objects with strong internal gravitational fields (*e.g.*, coalescing binary black holes), their detection will allow one to experimentally probe gravity in highly dynamical circumstances. Note finally that arrays of millisecond pulsars are sensitive detectors of (very low frequency) gravitational waves [65,66].

References:

1. S.N. Gupta, Phys. Rev. **96**, 1683 (1954);
   R.H. Kraichnan, Phys. Rev. **98**, 1118 (1955);
   R.P. Feynman, F.B. Morinigo, and W.G. Wagner, *Feynman Lectures on Gravitation*, edited by Brian Hatfield (Addison-Wesley, Reading, 1995);
   S. Weinberg, Phys. Rev. **138**, B988 (1965);
   V.I. Ogievetsky and I.V. Polubarinov, Ann. Phys. (NY) **35**, 167 (1965);
   W. Wyss, Helv. Phys. Acta **38**, 469 (1965);
   S. Deser, Gen. Rel. Grav. **1**, 9 (1970);
   D.G. Boulware and S. Deser, Ann. Phys. (NY) **89**, 193 (1975);
   J. Fang and C. Fronsdal, J. Math. Phys. **20**, 2264 (1979);
   R.M. Wald, Phys. Rev. **D33**, 3613 (1986);
   C. Cutler and R.M. Wald, Class. Quantum Grav. **4**, 1267 (1987);
   R.M. Wald, Class. Quantum Grav. **4**, 1279 (1987);
   N. Boulanger *et al.*, Nucl. Phys. **B597**, 127 (2001).
2. S. Weinberg, *Gravitation and Cosmology* (John Wiley, New York, 1972).
3. S. Weinberg, Rev. Mod. Phys. **61**, 1 (1989).
4. A.I. Shlyakhter, Nature **264**, 340 (1976);
   T. Damour and F. Dyson, Nucl. Phys. **B480**, 37 (1996);
   Y. Fujii *et al.*, Nucl. Phys. **B573**, 377 (2000).
5. J.D. Prestage, R.L. Tjoelker, and L. Maleki, Phys. Rev. Lett. **74**, 3511 (1995);
   Y. Sortais *et al.*, Physica Scripta **95**, 50 (2001).
6. J.K. Webb *et al.*, Phys. Rev. Lett. **87**, 091301 (2001);
   M.T. Murphy *et al.*, Mon. Not. Roy. Astron. Soc. **327**, 1208 (2001).
7. K.A. Olive and M. Pospelov, Phys. Rev. **D65**, 085044 (2002).
8. H.B. Sandvik, J.D. Barrow, and J. Magueijo, Phys. Rev. Lett. **88**, 031302 (2002).
9. T. Damour, F. Piazza, and G. Veneziano, Phys. Rev. Lett. **89**, 081601 (2002); and Phys. Rev. **D66**, 046007 (2002).
10. K.A. Olive *et al.*, Phys. Rev. **D66**, 045022 (2002).
11. M.T. Murphy *et al.*, Astrophys. Space Sci. **283**, 577 (2003).
12. J.N. Bahcall, C.L. Steinhardt, and D. Schlegel, `astro-ph/0301507`.
13. J.P. Uzan, Rev. Mod. Phys. **75**, 403 (2003).
14. J.D. Prestage *et al.*, Phys. Rev. Lett. **54**, 2387 (1985);
    S.K. Lamoreaux *et al.*, Phys. Rev. Lett. **57**, 3125 (1986);
    T.E. Chupp *et al.*, Phys. Rev. Lett. **63**, 1541 (1989).
15. T. Jacobson, S. Liberati and D. Mattingly, Nature **424**, 1019 (2003).
16. R.C. Myers and M. Pospelov, Phys. Rev. Lett. **90**, 211601 (2003).
17. Y. Su *et al.*, Phys. Rev. **D50**, 3614 (1994).
18. J.O. Dickey *et al.*, Science **265**, 482 (1994);
    J.G. Williams, X.X. Newhall, and J.O. Dickey, Phys. Rev. **D53**, 6730 (1996).
19. S. Baessler *et al.*, Phys. Rev. Lett. **83**, 3585 (1999).
20. G.L. Smith *et al.*, Phys. Rev. **D61**, 022001 (1999).
21. R.F.C. Vessot and M.W. Levine, Gen. Rel. Grav. **10**, 181 (1978);
    R.F.C. Vessot *et al.*, Phys. Rev. Lett. **45**, 2081 (1980).

22. C.M. Will, *Theory and Experiment in Gravitational Physics* (Cambridge University Press, Cambridge, 1993); and Living Rev. Rel. **4**, 4 (2001).
23. T. Damour, in *Gravitation and Quantizations*, ed. B. Julia and J. Zinn-Justin, Les Houches, Session LVII (Elsevier, Amsterdam, 1995), pp. 1–61.
24. H. van Dam and M.J. Veltman, Nucl. Phys. **B22**, 397 (1970);
V.I. Zakharov, Sov. Phys. JETP Lett. **12**, 312 (1970);
D.G. Boulware and S. Deser, Phys. Rev. **D6**, 3368 (1972);
C. Aragone and J. Chela-Flores, Nuovo Cim. **10A**, 818 (1972);
A.I. Vainshtein, Phys. Lett. **B39**, 393 (1972);
C. Deffayet *et al.*, Phys. Rev. **D65**, 044026 (2002);
M. Porrati, Phys. Lett. **B534**, 209 (2002);
N. Arkani-Hamed, H. Georgi, and M.D. Schwartz, Annals Phys. **305**, 96 (2003);
T. Damour, I.I. Kogan, and A. Papazoglou, Phys. Rev. **D67**, 064009 (2003);
G. Dvali, A. Gruzinov, and M. Zaldarriaga, Phys. Rev. **D68**, 024012 (2003).
25. T.R. Taylor and G. Veneziano, Phys. Lett. **B213**, 450 (1988);
S. Dimopoulos and G. Giudice, Phys. Lett. **B379**, 105 (1996);
I. Antoniadis, S. Dimopoulos, and G. Dvali, Nucl. Phys. **B516**, 70 (1998).
26. V.A. Rubakov, Phys. Usp **44**, 871 (2001);
I.I. Kogan, in *Proceedings of the XXXVIth Rencontres de Moriond, Electro-Weak Interactions and Unified Theories* (March 2001); `astro-ph/0108220`.
27. E. Fischbach and C.L. Talmadge, *The search for non-Newtonian gravity* (Springer-Verlag, New York, 1999).
28. J.C. Long, H.W. Chan, and J.C. Price, Nucl. Phys. **B539**, 23 (1999).
29. C.D. Hoyle *et al.*, Phys. Rev. Lett. **86**, 1418 (2001);
J. Chiaverini *et al.*, Phys. Rev. Lett. **90**, 151101 (2003);
J.C. Long *et al.*, Nature **421**, 922 (2003).
30. P. Jordan, *Schwerkraft und Weltall* (Vieweg, Braunschweig, 1955);
M. Fierz, Helv. Phys. Acta **29**, 128 (1956);
C. Brans and R.H. Dicke, Phys. Rev. **124**, 925 (1961).
31. T. Damour and G. Esposito-Farèse, Class. Quantum Grav. **9**, 2093 (1992).
32. A.S. Eddington, *The Mathematical Theory of Relativity* (Cambridge University Press, Cambridge, 1923);
K. Nordtvedt, Phys. Rev. **169**, 1017 (1968);
C.M. Will, Astrophys. J. **163**, 611 (1971).
33. B. Bertotti, L. Iess and P. Tortora, Nature, **425**, 374 (2003).
34. K. Nordtvedt, Phys. Rev. **170**, 1186 (1968).
35. T.R. Taylor and G. Veneziano, Phys. Lett. **B213**, 450 (1988);
T. Damour and A.M. Polyakov, Nucl. Phys. **B423**, 532 (1994).
36. R.A. Hulse, Rev. Mod. Phys. **66**, 699 (1994).
37. J.H. Taylor, Rev. Mod. Phys. **66**, 711 (1994).
38. T. Damour and N. Deruelle, Phys. Lett. **A87**, 81 (1981);
T. Damour, C.R. Acad. Sci. Paris **294**, 1335 (1982).
39. J.H. Taylor, Class. Quantum Grav. **10**, S167 (Supplement 1993).

40. J. Weisberg and J.H. Taylor, in *Radio Pulsars, ASP Conference Series* **302**, 93 (2003); astro-ph/0211217.
41. By contrast to the binary pulsar case, which does involve the gauge-invariant, helicity-two propagating degrees of freedom of the gravitational field, the recent measurement of light deflection from Jupiter (E.B. Fomalont and S.M. Kopeikin, astro-ph/0302294) does not depend (in spite of contrary claims: S.M. Kopeikin, Astrophys. J. **556**, L1 (2001)), at the considered precision level, on the propagation speed of gravity (see C.M. Will, Astrophys. J. **590**, 683 (2003) and S. Samuel, Phys. Rev. Lett. **90**, 231101 (2003)).
42. T. Damour and J.H. Taylor, Phys. Rev. **D45**, 1840 (1992).
43. T. Damour and G. Esposito-Farèse, Phys. Rev. **D54**, 1474 (1996); and Phys. Rev. **D58**, 042001 (1998).
44. I.H. Stairs, Living Rev. Rel. **6** 5 (2003).
45. T. Damour and J.H. Taylor, Astrophys. J. **366**, 501 (1991).
46. A. Wolszczan, Nature **350**, 688 (1991).
47. J.H. Taylor *et al.*, Nature **355**, 132 (1992).
48. I.H. Stairs *et al.*, Astrophys. J. **505**, 352 (1998);
    I.H. Stairs *et al.*, Astrophys. J. **581**, 501 (2002).
49. V.M. Kaspi *et al.*, Astrophys. J. **528**, 445 (2000).
50. M. Bailes *et al.*, Astrophys. J. **595**, L49 (2003).
51. C.M. Will and H.W. Zaglauer, Astrophys. J. **346**, 366 (1989).
52. T. Damour and G. Schäfer, Phys. Rev. Lett. **66**, 2549 (1991).
53. Ch. Lange *et al.*, Mon. Not. Roy. Astron. Soc. **326**, 274 (2001).
54. Z. Arzoumanian, in *Radio Pulsars, ASP Conference Series* **302**, 69 (2003); astro-ph/0212180.
55. M. Kramer, Astrophys. J. **509**, 856 (1998);
    J.M. Weisberg and J.H. Taylor, Astrophys. J. **576**, 942 (2002).
56. T. Damour and R. Ruffini, C. R. Acad. Sc. Paris **279**, série A, 971 (1974);
    B.M. Barker and R.F. O'Connell, Phys. Rev. D 12, 329 (1975).
57. A. Dar, Nucl. Phys. (Proc. Supp.) **B28**, 321 (1992).
58. J. Yang *et al.*, Astrophys. J. **227**, 697 (1979);
    T. Rothman and R. Matzner, Astrophys. J. **257**, 450 (1982);
    F.S. Accetta, L.M. Krauss, and P. Romanelli, Phys. Lett. **B248**, 146 (1990).
59. http://einstein.stanford.edu.
60. P. Touboul *et al.*, C.R.Acad. Sci. Paris **2** (série IV) 1271 (2001).
61. P.W. Worden, in *Proc. 7th Marcel Grossmann Meeting on General Relativity*, edited by R.J. Jantzen and G. MacKeiser, (World Scientific, Singapore, 1995), pp. 1569-1573.
62. http://www.ligo.caltech.edu .
63. http://www.virgo.infn.it .
64. http://www.geo600.uni-hannover.de .
65. V.M. Kaspi, J.H. Taylor and M.F. Ryba, Astrophys. J. **428**, 713 (1994).
66. A.N. Lommen and D.C. Backer, Bulletin of the American Astronomical Society **33**, 1347 (2001); and Astrophys. J. **562**, 297 (2001).

19. BIG-BANG COSMOLOGY

Written July 2001 by K.A. Olive (University of Minnesota) and J.A. Peacock (University of Edinburgh). Revised September 2005.

19.1. Introduction to Standard Big-Bang Model

The observed expansion of the Universe [1,2,3] is a natural (almost inevitable) result of any homogeneous and isotropic cosmological model based on general relativity. However, by itself, the Hubble expansion does not provide sufficient evidence for what we generally refer to as the Big-Bang model of cosmology. While general relativity is in principle capable of describing the cosmology of any given distribution of matter, it is extremely fortunate that our Universe appears to be homogeneous and isotropic on large scales. Together, homogeneity and isotropy allow us to extend the Copernican Principle to the Cosmological Principle, stating that all spatial positions in the Universe are essentially equivalent.

The formulation of the Big-Bang model began in the 1940s with the work of George Gamow and his collaborators, Alpher and Herman. In order to account for the possibility that the abundances of the elements had a cosmological origin, they proposed that the early Universe which was once very hot and dense (enough so as to allow for the nucleosynthetic processing of hydrogen), and has expanded and cooled to its present state [4,5]. In 1948, Alpher and Herman predicted that a direct consequence of this model is the presence of a relic background radiation with a temperature of order a few K [6,7]. Of course this radiation was observed 16 years later as the microwave background radiation [8]. Indeed, it was the observation of the 3 K background radiation that singled out the Big-Bang model as the prime candidate to describe our Universe. Subsequent work on Big-Bang nucleosynthesis further confirmed the necessity of our hot and dense past. (See the review on BBN—Sec. 20 of this *Review* for a detailed discussion of BBN.) These relativistic cosmological models face severe problems with their initial conditions, to which the best modern solution is inflationary cosmology, discussed in Sec. 19.3.5. If correct, these ideas would strictly render the term 'Big Bang' redundant, since it was first coined by Hoyle to represent a criticism of the lack of understanding of the initial conditions.

19.1.1. *The Robertson-Walker Universe* :

The observed homogeneity and isotropy enable us to describe the overall geometry and evolution of the Universe in terms of two cosmological parameters accounting for the spatial curvature and the overall expansion (or contraction) of the Universe. These two quantities appear in the most general expression for a space-time metric which has a (3D) maximally symmetric subspace of a 4D space-time, known as the Robertson-Walker metric:

$$ds^2 = dt^2 - R^2(t)\left[\frac{dr^2}{1-kr^2} + r^2\left(d\theta^2 + \sin^2\theta\, d\phi^2\right)\right] . \tag{19.1}$$

Note that we adopt $c = 1$ throughout. By rescaling the radial coordinate, we can choose the curvature constant k to take only the discrete values $+1$, -1, or 0 corresponding to closed, open, or spatially flat geometries. In this case, it is often more convenient to re-express the metric as

$$ds^2 = dt^2 - R^2(t)\left[d\chi^2 + S_k^2(\chi)\,(d\theta^2 + \sin^2\theta\, d\phi^2)\right] \ , \tag{19.2}$$

where the function $S_k(\chi)$ is $(\sin\chi, \chi, \sinh\chi)$ for $k = (+1, 0, -1)$. The coordinate r (in Eq. (19.1)) and the 'angle' χ (in Eq. (19.2)) are both dimensionless; the dimensions are carried by $R(t)$, which is the cosmological scale factor which determines proper distances in terms of the comoving coordinates. A common alternative is to define a dimensionless scale factor, $a(t) = R(t)/R_0$, where $R_0 \equiv R(t_0)$ is R at the present epoch. It is also sometimes convenient to define a dimensionless or conformal time coordinate, η, by $d\eta = dt/R(t)$. Along constant spatial sections, the proper time is defined by the time coordinate, t. Similarly, for $dt = d\theta = d\phi = 0$, the proper distance is given by $R(t)\chi$. For standard texts on cosmological models see *e.g.*, Refs. [9–14].

19.1.2. *The redshift* :

The cosmological redshift is a direct consequence of the Hubble expansion, determined by $R(t)$. A local observer detecting light from a distant emitter sees a redshift in frequency. We can define the redshift as

$$z \equiv \frac{\nu_1 - \nu_2}{\nu_2} \simeq \frac{v_{12}}{c} \ , \tag{19.3}$$

where ν_1 is the frequency of the emitted light, ν_2 is the observed frequency and v_{12} is the relative velocity between the emitter and the observer. While the definition, $z = (\nu_1 - \nu_2)/\nu_2$ is valid on all distance scales, relating the redshift to the relative velocity in this simple way is only true on small scales (*i.e.*, less than cosmological scales) such that the expansion velocity is non-relativistic. For light signals, we can use the metric given by Eq. (19.1) and $ds^2 = 0$ to write

$$\frac{v_{12}}{c} = \dot{R}\,\delta r = \frac{\dot{R}}{R}\,\delta t = \frac{\delta R}{R} = \frac{R_2 - R_1}{R_1} \ , \tag{19.4}$$

where $\delta r(\delta t)$ is the radial coordinate (temporal) separation between the emitter and observer. Thus, we obtain the simple relation between the redshift and the scale factor

$$1 + z = \frac{\nu_1}{\nu_2} = \frac{R_2}{R_1} \ . \tag{19.5}$$

This result does not depend on the non-relativistic approximation.

19.1.3. ***The Friedmann-Lemaître equations of motion*** :

The cosmological equations of motion are derived from Einstein's equations

$$\mathcal{R}_{\mu\nu} - \tfrac{1}{2} g_{\mu\nu}\mathcal{R} = 8\pi G_{\rm N} T_{\mu\nu} + \Lambda g_{\mu\nu} \ . \tag{19.6}$$

Gliner [15] and Zeldovich [16] seem to have pioneered the modern view, in which the Λ term is taken to the rhs and interpreted as particle-physics processes yielding an effective energy–momentum tensor $T_{\mu\nu}$ for the vacuum of $\Lambda g_{\mu\nu}/8\pi G_{\rm N}$. It is common to assume that the matter content of the Universe is a perfect fluid, for which

$$T_{\mu\nu} = -p g_{\mu\nu} + (p+\rho)\, u_\mu u_\nu \ , \tag{19.7}$$

where $g_{\mu\nu}$ is the space-time metric described by Eq. (19.1), p is the isotropic pressure, ρ is the energy density and $u = (1,0,0,0)$ is the velocity vector for the isotropic fluid in co-moving coordinates. With the perfect fluid source, Einstein's equations lead to the Friedmann-Lemaître equations

$$H^2 \equiv \left(\frac{\dot R}{R}\right)^2 = \frac{8\pi\ G_{\rm N}\ \rho}{3} \ - \ \frac{k}{R^2} + \frac{\Lambda}{3} \ , \tag{19.8}$$

and

$$\frac{\ddot R}{R} = \frac{\Lambda}{3} - \frac{4\pi G_{\rm N}}{3}\ \ (\rho + 3p) \ , \tag{19.9}$$

where $H(t)$ is the Hubble parameter and Λ is the cosmological constant. The first of these is sometimes called the Friedmann equation. Energy conservation via $T^{\mu\nu}_{\ \ ;\mu} = 0$, leads to a third useful equation [which can also be derived from Eq. (19.8) and Eq. (19.9)]

$$\dot\rho = -3H\,(\rho + p) \ . \tag{19.10}$$

Eq. (19.10) can also be simply derived as a consequence of the first law of thermodynamics.

Eq. (19.8) has a simple classical mechanical analog if we neglect (for the moment) the cosmological term Λ. By interpreting $-k/R^2$ as a "total energy", then we see that the evolution of the Universe is governed by a competition between the potential energy, $8\pi G_{\rm N}\rho/3$ and the kinetic term $(\dot R/R)^2$. For $\Lambda = 0$, it is clear that the Universe must be expanding or contracting (except at the turning point prior to collapse in a closed Universe). The ultimate fate of the Universe is determined by the curvature constant k. For $k = +1$, the Universe will recollapse in a finite time, whereas for $k = 0, -1$, the Universe will expand indefinitely. These simple conclusions can be altered when $\Lambda \neq 0$ or more generally with some component with $(\rho + 3p) < 0$.

19.1.4. ***Definition of cosmological parameters*** :

In addition to the Hubble parameter, it is useful to define several other measurable cosmological parameters. The Friedmann equation can be used to define a critical density such that $k = 0$ when $\Lambda = 0$,

$$\begin{aligned} \rho_c \equiv \frac{3H^2}{8\pi G_{\rm N}} &= 1.88 \times 10^{-26}\, h^2 \text{ kg m}^{-3} \\ &= 1.05 \times 10^{-5}\, h^2 \text{ GeV cm}^{-3} \ , \end{aligned} \tag{19.11}$$

where the scaled Hubble parameter, h, is defined by

$$\begin{aligned} H &\equiv 100\, h \text{ km s}^{-1} \text{ Mpc}^{-1} \\ \Rightarrow H^{-1} &= 9.78\, h^{-1} \text{ Gyr} \\ &= 2998\, h^{-1} \text{ Mpc} \ . \end{aligned} \tag{19.12}$$

The cosmological density parameter $\Omega_{\rm tot}$ is defined as the energy density relative to the critical density,

$$\Omega_{\rm tot} = \rho/\rho_c \ . \tag{19.13}$$

Note that one can now rewrite the Friedmann equation as

$$k/R^2 = H^2(\Omega_{\rm tot} - 1) \ , \tag{19.14}$$

From Eq. (19.14), one can see that when $\Omega_{\rm tot} > 1$, $k = +1$ and the Universe is closed, when $\Omega_{\rm tot} < 1$, $k = -1$ and the Universe is open, and when $\Omega_{\rm tot} = 1$, $k = 0$, and the Universe is spatially flat.

It is often necessary to distinguish different contributions to the density. It is therefore convenient to define present-day density parameters for pressureless matter (Ω_m) and relativistic particles (Ω_r), plus the quantity $\Omega_\Lambda = \Lambda/3H^2$. In more general models, we may wish to drop the assumption that the vacuum energy density is constant, and we therefore denote the present-day density parameter of the vacuum by Ω_v. The Friedmann equation then becomes

$$k/R_0^2 = H_0^2(\Omega_m + \Omega_r + \Omega_v - 1) \ , \tag{19.15}$$

where the subscript 0 indicates present-day values. Thus, it is the sum of the densities in matter, relativistic particles and vacuum that determines the overall sign of the curvature. Note that the quantity $-k/R_0^2H_0^2$ is sometimes referred to as Ω_k. This usage is unfortunate: it encourages one to think of curvature as a contribution to the energy density of the Universe, which is not correct.

19.1.5. ***Standard Model solutions*** :

Much of the history of the Universe in the standard Big-Bang model can be easily described by assuming that either matter or radiation dominates the total energy density. During inflation or perhaps even today if we are living in an accelerating Universe, domination by a cosmological constant or some other form of dark energy should be considered. In the following, we shall delineate the solutions to the Friedmann equation when a single component dominates the energy density. Each component is distinguished by an equation of state parameter $w = p/\rho$.

19.1.5.1. *Solutions for a general equation of state:*

Let us first assume a general equation of state parameter for a single component, w which is constant. In this case, Eq. (19.10) can be written as $\dot{\rho} = -3(1+w)\rho\dot{R}/R$ and is easily integrated to yield

$$\rho \propto R^{-3(1+w)} \ . \tag{19.16}$$

Note that at early times when R is small, the less singular curvature term k/R^2 in the Friedmann equation can be neglected so long as $w > -1/3$. Curvature domination occurs at rather late times (if a cosmological constant term does not dominate sooner). For $w \neq -1$, one can insert this result into the Friedmann equation Eq. (19.8) and if one neglects the curvature and cosmological constant terms, it is easy to integrate the equation to obtain,

$$R(t) \propto t^{2/[3(1+w)]} \ . \tag{19.17}$$

19.1.5.2. *A Radiation-dominated Universe:*

In the early hot and dense Universe, it is appropriate to assume an equation of state corresponding to a gas of radiation (or relativistic particles) for which $w = 1/3$. In this case, Eq. (19.16) becomes $\rho \propto R^{-4}$. The "extra" factor of $1/R$ is due to the cosmological redshift; not only is the number density of particles in the radiation background decreasing as R^{-3} since volume scales as R^3, but in addition, each particle's energy is decreasing as $E \propto \nu \propto R^{-1}$. Similarly, one can substitute $w = 1/3$ into Eq. (19.17) to obtain

$$R(t) \propto t^{1/2} \ ; \qquad H = 1/2t \ . \tag{19.18}$$

19.1.5.3. *A Matter-dominated Universe:*

At relatively late times, non-relativistic matter eventually dominates the energy density over radiation (see Sec. 19.3.8). A pressureless gas ($w = 0$) leads to the expected dependence $\rho \propto R^{-3}$ from Eq. (19.16) and, if $k = 0$, we get

$$R(t) \propto t^{2/3} \ ; \qquad H = 2/3t \ . \tag{19.19}$$

19.1.5.4. *A Universe dominated by vacuum energy:*

If there is a dominant source of vacuum energy, V_0, it would act as a cosmological constant with $\Lambda = 8\pi G_{\rm N} V_0$ and equation of state $w = -1$. In this case, the solution to the Friedmann equation is particularly simple and leads to an exponential expansion of the Universe

$$R(t) \propto e^{\sqrt{\Lambda/3}t} \ . \tag{19.20}$$

A key parameter is the equation of state of the vacuum, $w \equiv p/\rho$: this need not be the $w = -1$ of Λ, and may not even be constant [17,18,19]. It is now common to use w to stand for this vacuum equation of state, rather than of any other constituent of the Universe, and we use the symbol in this sense hereafter. We generally assume w to be independent of time, and where results relating to the vacuum are quoted without an explicit w dependence, we have adopted $w = -1$.

The presence of vacuum energy can dramatically alter the fate of the Universe. For example, if $\Lambda < 0$, the Universe will eventually recollapse independent of the sign of k. For large values of Λ (larger than the Einstein static value needed to halt any cosmological expansion or contraction), even a closed Universe will expand forever. One way to quantify this is the deceleration parameter, q_0, defined as

$$q_0 = -\left.\frac{R\ddot{R}}{\dot{R}^2}\right|_0 = \frac{1}{2}\Omega_m + \Omega_r + \frac{(1+3w)}{2}\Omega_v \ . \tag{19.21}$$

This equation shows us that $w < -1/3$ for the vacuum may lead to an accelerating expansion. Astonishingly, it appears that such an effect has been observed in the Supernova Hubble diagram [20–23] (see Fig. 19.1 below); current data indicate that vacuum energy is indeed the largest contributor to the cosmological density budget, with $\Omega_v = 0.72 \pm 0.05$ and $\Omega_m = 0.28 \pm 0.05$ if $k = 0$ is assumed [23].

The nature of this dominant term is presently uncertain, but much effort is being invested in dynamical models (*e.g.*, rolling scalar fields), under the catch-all heading of "quintessence."

19.2. Introduction to Observational Cosmology

19.2.1. ***Fluxes, luminosities, and distances*** :

The key quantities for observational cosmology can be deduced quite directly from the metric.

(1) The *proper* transverse size of an object seen by us to subtend an angle $d\psi$ is its comoving size $d\psi\, S_k(\chi)$ times the scale factor at the time of emission:

$$d\ell = d\psi\ R_0 S_k(\chi)/(1+z) \ . \tag{19.22}$$

(2) The apparent flux density of an object is deduced by allowing its photons to flow through a sphere of current radius $R_0 S_k(\chi)$; but photon energies and arrival rates are

redshifted, and the bandwidth $d\nu$ is reduced. The observed photons at frequency ν_0 were emitted at frequency $\nu_0(1+z)$, so the flux density is the luminosity at this frequency, divided by the total area, divided by $1+z$:

$$S_\nu(\nu_0) = \frac{L_\nu([1+z]\nu_0)}{4\pi R_0^2 S_k^2(\chi)(1+z)} \ . \tag{19.23}$$

These relations lead to the following common definitions:

$$\begin{aligned} \text{angular-diameter distance:} \quad & D_{\rm A} = (1+z)^{-1} R_0 S_k(\chi) \\ \text{luminosity distance:} \quad & D_{\rm L} = (1+z)\, R_0 S_k(\chi) \end{aligned} \tag{19.24}$$

These distance-redshift relations are expressed in terms of observables by using the equation of a null radial geodesic ($R(t)d\chi = dt$) plus the Friedmann equation:

$$\begin{aligned} R_0 d\chi =& \frac{1}{H(z)}\, dz = \frac{1}{H_0} \Big[(1-\Omega_m-\Omega_v-\Omega_r)(1+z)^2 \\ &+ \ \Omega_v(1+z)^{3+3w} \ + \ \Omega_m(1+z)^3 + \Omega_r(1+z)^4\Big]^{-1/2} \ dz \ . \end{aligned} \tag{19.25}$$

The main scale for the distance here is the Hubble length, $1/H_0$.

The flux density is the product of the specific intensity I_ν and the solid angle $d\Omega$ subtended by the source: $S_\nu = I_\nu \ d\Omega$. Combining the angular size and flux-density relations thus gives the relativistic version of surface-brightness conservation:

$$I_\nu(\nu_0) = \frac{B_\nu([1+z]\nu_0)}{(1+z)^3} \ , \tag{19.26}$$

where B_ν is surface brightness (luminosity emitted into unit solid angle per unit area of source). We can integrate over ν_0 to obtain the corresponding total or bolometric formula:

$$I_{\rm tot} = \frac{B_{\rm tot}}{(1+z)^4} \ . \tag{19.27}$$

This cosmology-independent form expresses Liouville's Theorem: photon phase-space density is conserved along rays.

19.2.2. *Distance data and geometrical tests of cosmology* :

In order to confront these theoretical predictions with data, we have to bridge the divide between two extremes. Nearby objects may have their distances measured quite easily, but their radial velocities are dominated by deviations from the ideal Hubble flow, which typically have a magnitude of several hundred $\mathrm{km\,s^{-1}}$. On the other hand, objects at redshifts $z \gtrsim 0.01$ will have observed recessional velocities that differ from their ideal values by $\lesssim 10\%$, but absolute distances are much harder to supply in this case. The traditional solution to this problem is the construction of the distance ladder: an interlocking set of methods for obtaining relative distances between various classes of

object, which begins with absolute distances at the 10 to 100 pc level and terminates with galaxies at significant redshifts. This is reviewed in the review on Global cosmological parameters—Sec. 21 of this *Review*.

By far the most exciting development in this area has been the use of type Ia Supernovae (SNe), which now allow measurement of relative distances with 5% precision. In combination with Cepheid data from the HST key project on the distance scale, SNe results are the dominant contributor to the best modern value for H_0: $72\,\mathrm{km\,s^{-1}Mpc^{-1}} \pm 10\%$ [24]. Better still, the analysis of high-z SNe has allowed the first meaningful test of cosmological geometry to be carried out: as shown in Fig. 19.1 and Fig. 19.2, a combination of supernova data and measurements of microwave-background anisotropies strongly favors a $k = 0$ model dominated by vacuum energy. (See the review on Global cosmological parameters—Sec. 21 of this *Review* for a more comprehensive review of Hubble parameter determinations.)

19.2.3. *Age of the Universe* :

The most striking conclusion of relativistic cosmology is that the Universe has not existed forever. The dynamical result for the age of the Universe may be written as

$$\begin{aligned} H_0 t_0 &= \int_0^\infty \frac{dz}{(1+z)H(z)} \\ &= \int_0^\infty \frac{dz}{(1+z)\,\left[(1+z)^2(1+\Omega_m z) - z(2+z)\Omega_v\right]^{1/2}} \ , \end{aligned} \tag{19.28}$$

where we have neglected Ω_r and chosen $w = -1$. Over the range of interest ($0.1 \lesssim \Omega_m \lesssim 1$, $|\Omega_v| \lesssim 1$), this exact answer may be approximated to a few % accuracy by

$$H_0 t_0 \simeq \tfrac{2}{3}\,(0.7\Omega_m + 0.3 - 0.3\Omega_v)^{-0.3} \ . \tag{19.29}$$

For the special case that $\Omega_m + \Omega_v = 1$, the integral in Eq. (19.28) can be expressed analytically as

$$H_0 t_0 = \frac{2}{3\sqrt{\Omega_v}}\,\ln\frac{1+\sqrt{\Omega_v}}{\sqrt{1-\Omega_v}} \quad (\Omega_m < 1) \ . \tag{19.30}$$

The most accurate means of obtaining ages for astronomical objects is based on the natural clocks provided by radioactive decay. The use of these clocks is complicated by a lack of knowledge of the initial conditions of the decay. In the Solar System, chemical fractionation of different elements helps pin down a precise age for the pre-Solar nebula of 4.6 Gyr, but for stars it is necessary to attempt an a priori calculation of the relative abundances of nuclei that result from supernova explosions. In this way, a lower limit for the age of stars in the local part of the Milky Way of about 11 Gyr is obtained [25].

The other major means of obtaining cosmological age estimates is based on the theory of stellar evolution. In principle, the main-sequence turnoff point in the color-magnitude diagram of a globular cluster should yield a reliable age. However, these have been

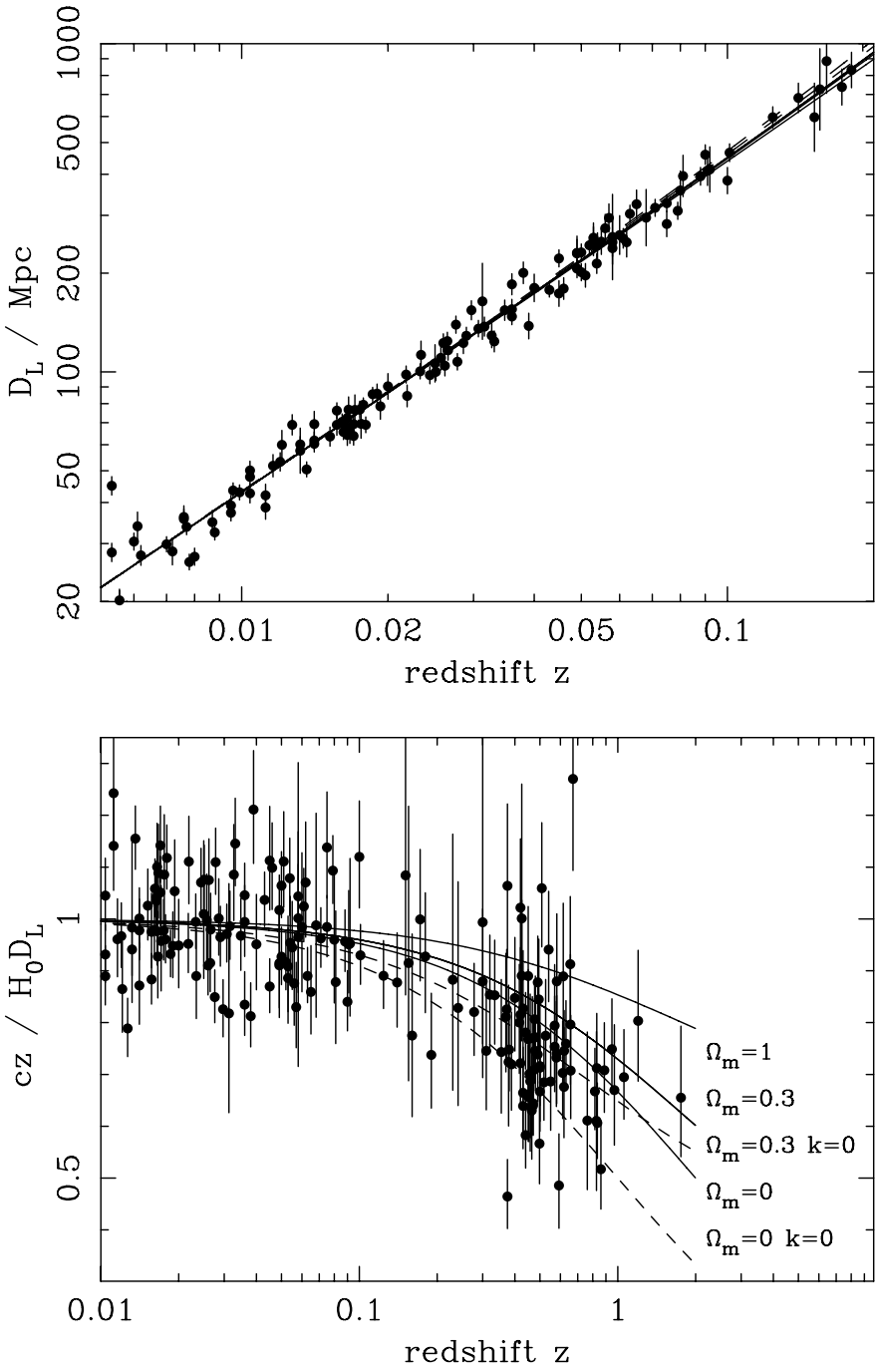


Figure 19.1: The type Ia supernova Hubble diagram [20–22]. The first panel shows that for $z \ll 1$ the large-scale Hubble flow is indeed linear and uniform; the second panel shows an expanded scale, with the linear trend divided out, and with the redshift range extended to show how the Hubble law becomes nonlinear. ($\Omega_r = 0$ is assumed.) Comparison with the prediction of Friedmann-Lemaître models appears to favor a vacuum-dominated Universe.

controversial owing to theoretical uncertainties in the evolution model, as well as

observational uncertainties in the distance, dust extinction and metallicity of clusters. The present consensus favors ages for the oldest clusters of about 12 Gyr [26,27].

These methods are all consistent with the age deduced from studies of structure formation, using the microwave background and large-scale structure: $t_0 = 13.7 \pm 0.2$ Gyr [28], where the extra accuracy comes at the price of assuming the Cold Dark Matter model to be true.

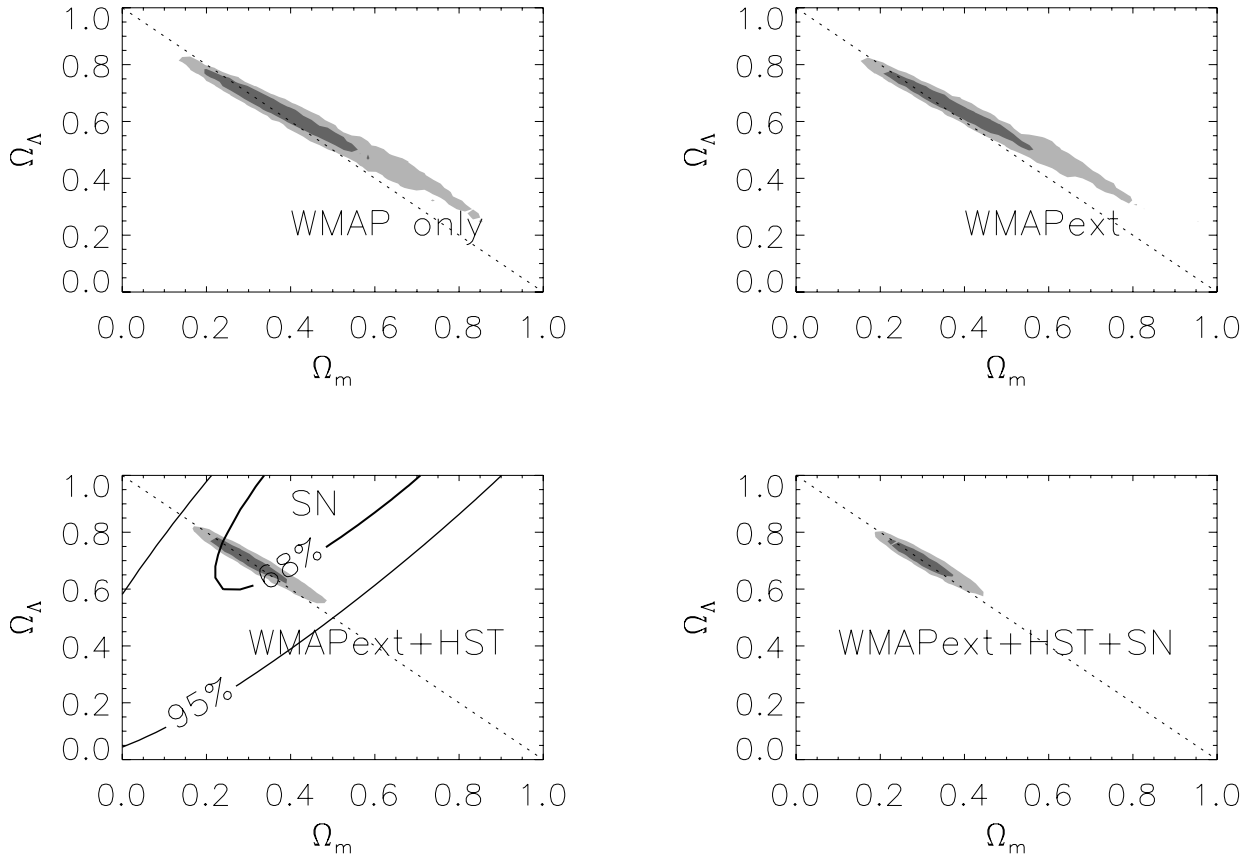


Figure 19.2: Likelihood-based confidence contours [28] over the plane Ω_Λ (*i.e.* Ω_v assuming $w = -1$) vs Ω_m. The SNe Ia results very nearly constrain $\Omega_v - \Omega_m$, whereas the results of CMB anisotropies (from the first-year WMAP data) favor a flat model with $\Omega_v + \Omega_m \simeq 1$. The intersection of these constraints is the most direct (but far from the only) piece of evidence favoring a flat model with $\Omega_m \simeq 0.3$.

19.2.4. *Horizon, isotropy, flatness problems* :

For photons, the radial equation of motion is just $c\,dt = R\,d\chi$. How far can a photon get in a given time? The answer is clearly

$$\Delta\chi = \int_{t_1}^{t_2} \frac{dt}{R(t)} \equiv \Delta\eta \ , \tag{19.31}$$

i.e., just the interval of conformal time. We can replace dt by $dR/\dot{R}$, which the Friedmann equation says is $\propto dR/\sqrt{\rho R^2}$ at early times. Thus, this integral converges if $\rho R^2 \to \infty$ as

$t_1 \to 0$, otherwise it diverges. Provided the equation of state is such that ρ changes faster than R^{-2}, light signals can only propagate a finite distance between the Big Bang and the present; there is then said to be a particle horizon. Such a horizon therefore exists in conventional Big-Bang models, which are dominated by radiation ($\rho \propto R^{-4}$) at early times.

At late times, the integral for the horizon is largely determined by the matter-dominated phase, for which

$$D_{\rm H} = R_0\,\chi_{\rm H} \equiv R_0 \int_0^{t(z)} \frac{dt}{R(t)} \simeq \frac{6000}{\sqrt{\Omega z}}\; h^{-1}\,{\rm Mpc} \quad (z \gg 1)\ . \tag{19.32}$$

The horizon at the time of formation of the microwave background ('last scattering': $z \simeq 1100$) was thus of order 100 Mpc in size, subtending an angle of about 1°. Why then are the large number of causally disconnected regions we see on the microwave sky all at the same temperature? The Universe is very nearly isotropic and homogeneous, even though the initial conditions appear not to permit such a state to be constructed.

A related problem is that the $\Omega = 1$ Universe is unstable:

$$\Omega(a) - 1 = \frac{\Omega - 1}{1 - \Omega + \Omega_v a^2 + \Omega_m a^{-1} + \Omega_r a^{-2}}\ , \tag{19.33}$$

where Ω with no subscript is the total density parameter, and $a(t) = R(t)/R_0$. This requires $\Omega(t)$ to be unity to arbitrary precision as the initial time tends to zero; a universe of non-zero curvature today requires very finely tuned initial conditions.

19.3. The Hot Thermal Universe

19.3.1. *Thermodynamics of the early Universe* :

As alluded to above, we expect that much of the early Universe can be described by a radiation-dominated equation of state. In addition, through much of the radiation-dominated period, thermal equilibrium is established by the rapid rate of particle interactions relative to the expansion rate of the Universe (see Sec. 19.3.3 below). In equilibrium, it is straightforward to compute the thermodynamic quantities, ρ, p, and the entropy density, s. In general, the energy density for a given particle type i can be written as

$$\rho_i = \int E_i\; dn_{q_i}\ , \tag{19.34}$$

with the density of states given by

$$dn_{q_i} = \frac{g_i}{2\pi^2}\left(\exp[(E_{q_i} - \mu_i)/T_i] \pm 1\right)^{-1}\; q_i^2\, dq_i\ , \tag{19.35}$$

where g_i counts the number of degrees of freedom for particle type i, $E_{q_i}^2 = m_i^2 + q_i^2$, μ_i is the chemical potential, and the $\pm$ corresponds to either Fermi or Bose statistics. Similarly, we can define the pressure of a perfect gas as

$$p_i = \frac{1}{3}\int \frac{q_i^2}{E_i}\; dn_{q_i}\ . \tag{19.36}$$

The number density of species i is simply

$$n_i = \int dn_{q_i} \ , \tag{19.37}$$

and the entropy density is

$$s_i = \frac{\rho_i + p_i - \mu_i n_i}{T_i} \ , \tag{19.38}$$

In the Standard Model, a chemical potential is often associated with baryon number, and since the net baryon density relative to the photon density is known to be very small (of order 10^{-10}), we can neglect any such chemical potential when computing total thermodynamic quantities.

For photons, we can compute all of the thermodynamic quantities rather easily. Taking $g_i = 2$ for the 2 photon polarization states, we have

$$\rho_\gamma = \frac{\pi^2}{15} T^4 \ ; \quad p_\gamma = \frac{1}{3}\rho_\gamma \ ; \quad s_\gamma = \frac{4\rho_\gamma}{3T} \ ; \quad n_\gamma = \frac{2\zeta(3)}{\pi^2} T^3 \ , \tag{19.39}$$

with $2\zeta(3)/\pi^2 \simeq 0.2436$. Note that Eq. (19.10) can be converted into an equation for entropy conservation. Recognizing that $\dot{p} = s\dot{T}$, Eq. (19.10) becomes

$$d(sR^3)/dt = 0 \ . \tag{19.40}$$

For radiation, this corresponds to the relationship between expansion and cooling, $T \propto R^{-1}$ in an adiabatically expanding Universe. Note also that both s and n_γ scale as T^3.

19.3.2. *Radiation content of the Early Universe* :

At the very high temperatures associated with the early Universe, massive particles are pair produced, and are part of the thermal bath. If for a given particle species i we have $T \gg m_i$, then we can neglect the mass in Eq. (19.34) to Eq. (19.38), and the thermodynamic quantities are easily computed as in Eq. (19.39). In general, we can approximate the energy density (at high temperatures) by including only those particles with $m_i \ll T$. In this case, we have

$$\rho = \left(\sum_B g_B + \frac{7}{8} \sum_F g_F \right) \frac{\pi^2}{30} T^4 \equiv \frac{\pi^2}{30} \, N(T) \, T^4 \ , \tag{19.41}$$

where $g_{B(F)}$ is the number of degrees of freedom of each boson (fermion) and the sum runs over all boson and fermion states with $m \ll T$. The factor of 7/8 is due to the difference between the Fermi and Bose integrals. Eq. (19.41) defines the effective number of degrees of freedom, $N(T)$, by taking into account new particle degrees of freedom as the temperature is raised. This quantity is plotted in Fig. 19.3 [29].

Temperature	New Particles	$4N(T)$
$T < m_e$	γ's + ν's	29
$m_e < T < m_\mu$	$e^\pm$	43
$m_\mu < T < m_\pi$	$\mu^\pm$	57
$m_\pi < T < T_c$ [†]	π's	69
$T_c < T < m_{\rm strange}$	π's + $u, \bar{u}, d, \bar{d}$ + gluons	205
$m_s < T < m_{\rm charm}$	$s, \bar{s}$	247
$m_c < T < m_\tau$	$c, \bar{c}$	289
$m_\tau < T < m_{\rm bottom}$	$\tau^\pm$	303
$m_b < T < m_{\rm W,Z}$	$b, \bar{b}$	345
$m_{W,Z} < T < m_{\rm Higgs}$	$W^\pm, Z$	381
$m_H < T < m_{\rm top}$	H^0	385
$m_t < T$	$t, \bar{t}$	427

[†] T_c corresponds to the confinement-deconfinement transition between quarks and hadrons.

The value of $N(T)$ at any given temperature depends on the particle physics model. In the standard SU(3) × SU(2) × U(1) model, we can specify $N(T)$ up to temperatures of O(100) GeV. The change in N (ignoring mass effects) can be seen in the above table.

At higher temperatures, $N(T)$ will be model dependent. For example, in the minimal SU(5) model, one needs to add 24 states to $N(T)$ for the X and Y gauge bosons, another 24 from the adjoint Higgs, and another 6 (in addition to the 4 already counted in $W^\pm, Z$, and H) from the $\overline{5}$ of Higgs. Hence for $T > m_{\rm X}$ in minimal SU(5), $N(T) = 160.75$. In a supersymmetric model this would at least double, with some changes possibly necessary in the table if the lightest supersymmetric particle has a mass below m_t.

In the radiation-dominated epoch, Eq. (19.10) can be integrated (neglecting the T-dependence of N) giving us a relationship between the age of the Universe and its temperature

$$t = \left(\frac{90}{32\pi^3 G_{\rm N} N(T)}\right)^{1/2} T^{-2} \ , \tag{19.42}$$

Put into a more convenient form

$$t\, T_{\rm MeV}^2 = 2.4[N(T)]^{-1/2} \ , \tag{19.43}$$

where t is measured in seconds and $T_{\rm MeV}$ in units of MeV.

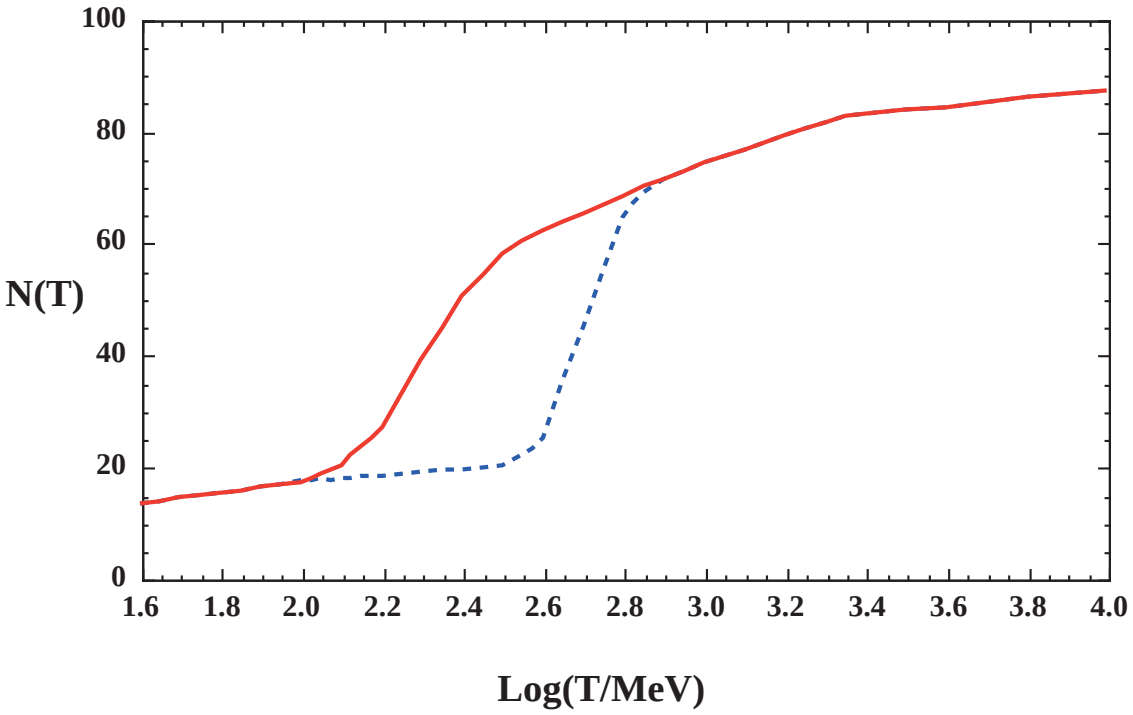


Figure 19.3: The effective numbers of relativistic degrees of freedom as a function of temperature. The sharp drop corresponds to the quark-hadron transition. The solid curve assume a QCD scale of 150 MeV, while the dashed curve assumes 450 MeV.

19.3.3. ***Neutrinos and equilibrium*** : Due to the expansion of the Universe, certain rates may be too slow to either establish or maintain equilibrium. Quantitatively, for each particle i, as a minimal condition for equilibrium, we will require that some rate Γ_i involving that type be larger than the expansion rate of the Universe or

$$\Gamma_i > H \ . \tag{19.44}$$

Recalling that the age of the Universe is determined by H^{-1}, this condition is equivalent to requiring that on average, at least one interaction has occurred over the lifetime of the Universe.

A good example for a process which goes in and out of equilibrium is the weak interactions of neutrinos. On dimensional grounds, one can estimate the thermally averaged scattering cross section

$$\langle \sigma v \rangle \sim O(10^{-2}) T^2 / m_{\rm W}^4 \tag{19.45}$$

for $T \lesssim m_{\rm W}$. Recalling that the number density of leptons is $n \propto T^3$, we can compare the weak interaction rate, $\Gamma \sim n \langle \sigma v \rangle$, with the expansion rate,

$$\begin{aligned} H = \left(\frac{8\pi G_{\rm N} \rho}{3} \right)^{1/2} &= \left(\frac{8\pi^3}{90} N(T) \right)^{1/2} T^2 / M_{\rm P} \\ &\sim 1.66 N(T)^{1/2} T^2 / M_{\rm P} . \end{aligned} \tag{19.46}$$

The Planck mass $M_{\rm P} = G_{\rm N}^{-1/2} = 1.22 \times 10^{19}$ GeV.

Neutrinos will be in equilibrium when $\Gamma_{\rm wk} > H$ or

$$T > (500\, m_{\rm W}^4/M_{\rm P})^{1/3} \sim 1 \ \text{MeV} \ . \tag{19.47}$$

The temperature at which these rates are equal is commonly referred to as the neutrino decoupling or freeze-out temperature and is defined by $\Gamma(T_d) = H(T_d)$.

At very high temperatures, the Universe is too young for equilibrium to have been established. For $T \gg m_{\rm W}$, we should write $\langle \sigma v \rangle \sim O(10^{-2})/T^2$, so that $\Gamma \sim 10^{-2} T$. Thus at temperatures $T \gtrsim 10^{-2} M_{\rm P}/\sqrt{N}$, equilibrium will not have been established.

For $T < T_d$, neutrinos drop out of equilibrium. The Universe becomes transparent to neutrinos and their momenta simply redshift with the cosmic expansion. The effective neutrino temperature will simply fall with $T \sim 1/R$.

Soon after decoupling, $e^\pm$ pairs in the thermal background begin to annihilate (when $T \lesssim m_e$). Because the neutrinos are decoupled, the energy released due to annihilation heats up the photon background relative to the neutrinos. The change in the photon temperature can be easily computed from entropy conservation. The neutrino entropy must be conserved separately from the entropy of interacting particles. A straightforward computation yields

$$T_\nu = (4/11)^{1/3}\, T_\gamma \simeq 1.9 \ \text{K} \ . \tag{19.48}$$

Today, the total entropy density is therefore given by

$$s = \frac{4}{3}\frac{\pi^2}{30}\left(2 + \frac{21}{4}(T_\nu/T_\gamma)^3\right) T_\gamma^3 = \frac{4}{3}\frac{\pi^2}{30}\left(2 + \frac{21}{11}\right) T_\gamma^3 = 7.04\, n_\gamma \ . \tag{19.49}$$

Similarly, the total relativistic energy density today is given by

$$\rho_r = \frac{\pi^2}{30}\left[2 + \frac{21}{4}(T_\nu/T_\gamma)^4\right] T_\gamma^4 \simeq 1.68 \rho_\gamma \ . \tag{19.50}$$

In practice, a small correction is needed to this, since neutrinos are not totally decoupled at $e^\pm$ annihilation: the effective number of massless neutrino species is 3.04, rather than 3 [30].

This expression ignores neutrino rest masses, but current oscillation data require at least one neutrino eigenstate to have a mass exceeding 0.05 eV. In this minimal case, $\Omega_\nu h^2 = 5 \times 10^{-4}$, so the neutrino contribution to the matter budget would be negligibly small (which is our normal assumption). However, a nearly degenerate pattern of mass eigenstates could allow larger densities, since oscillation experiments only measure differences in m^2 values. Note that a 0.05-eV neutrino has $kT_\nu = m_\nu$ at $z \simeq 297$, so the above expression for the total present relativistic density is really only an extrapolation. However, neutrinos are almost certainly relativistic at all epochs where the radiation content of the universe is dynamically significant.

19.3.4. *Field Theory and Phase transitions* :

It is very likely that the Universe has undergone one or more phase transitions during the course of its evolution [31–34]. Our current vacuum state is described by $SU(3)_c \times U(1)_{em}$ which in the Standard Model is a remnant of an unbroken $SU(3)_c \times SU(2)_L \times U(1)_Y$ gauge symmetry. Symmetry breaking occurs when a non-singlet gauge field (the Higgs field in the Standard Model) picks up a non-vanishing vacuum expectation value, determined by a scalar potential. For example, a simple (non-gauged) potential describing symmetry breaking is $V(\phi) = \frac{1}{4}\lambda\phi^4 - \frac{1}{2}\mu^2\phi^2 + V(0)$. The resulting expectation value is simply $\langle\phi\rangle = \mu/\sqrt{\lambda}$.

In the early Universe, finite temperature radiative corrections typically add terms to the potential of the form $\phi^2 T^2$. Thus, at very high temperatures, the symmetry is restored and $\langle\phi\rangle = 0$. As the Universe cools, depending on the details of the potential, symmetry breaking will occur via a first order phase transition in which the field tunnels through a potential barrier, or via a second order transition in which the field evolves smoothly from one state to another (as would be the case for the above example potential).

The evolution of scalar fields can have a profound impact on the early Universe. The equation of motion for a scalar field ϕ can be derived from the energy-momentum tensor

$$T_{\mu\nu} = \partial_\mu\phi\partial_\nu\phi - \frac{1}{2}g_{\mu\nu}\partial_\rho\phi\partial^\rho\phi - g_{\mu\nu}V(\phi) \ . \tag{19.51}$$

By associating $\rho = T_{00}$ and $p = R^{-2}(t)T_{ii}$ we have

$$\begin{aligned} \rho &= \frac{1}{2}\dot{\phi}^2 + \frac{1}{2}R^{-2}(t)(\nabla\phi)^2 + V(\phi) \\ p &= \frac{1}{2}\dot{\phi}^2 - \frac{1}{6}R^{-2}(t)(\nabla\phi)^2 - V(\phi) \end{aligned} \tag{19.52}$$

and from Eq. (19.10) we can write the equation of motion (by considering a homogeneous region, we can ignore the gradient terms)

$$\ddot{\phi} + 3H\dot{\phi} = -\partial V/\partial\phi \ . \tag{19.53}$$

19.3.5. *Inflation* :

In Sec. 19.2.4, we discussed some of the problems associated with the standard Big-Bang model. However, during a phase transition, our assumptions of an adiabatically expanding universe are generally not valid. If, for example, a phase transition occurred in the early Universe such that the field evolved slowly from the symmetric state to the global minimum, the Universe may have been dominated by the vacuum energy density associated with the potential near $\phi \approx 0$. During this period of slow evolution, the energy density due to radiation will fall below the vacuum energy density, $\rho \ll V(0)$. When this happens, the expansion rate will be dominated by the constant V(0) and we obtain the exponentially expanding solution given in Eq. (19.20). When the field evolves towards

the global minimum it will begin to oscillate about the minimum, energy will be released during its decay and a hot thermal universe will be restored. If released fast enough, it will produce radiation at a temperature $NT_R{}^4 \lesssim V(0)$. In this reheating process entropy has been created and the final value of RT is greater than the initial value of RT. Thus, we see that, during a phase transition, the relation $RT \sim$ constant need not hold true. This is the basis of the inflationary Universe scenario [35–37].

If during the phase transition the value of RT changed by a factor of $O(10^{29})$, the cosmological problems discussed above would be solved. The observed isotropy would be generated by the immense expansion; one small causal region could get blown up and hence our entire visible Universe would have been in thermal contact some time in the past. In addition, the density parameter Ω would have been driven to 1 (with exponential precision). Density perturbations will be stretched by the expansion, $\lambda \sim R(t)$. Thus it will appear that $\lambda \gg H^{-1}$ or that the perturbations have left the horizon, where in fact the size of the causally connected region is now no longer simply H^{-1}. However, not only does inflation offer an explanation for large scale perturbations, it also offers a source for the perturbations themselves through quantum fluctuations.

Early models of inflation were based on a first order phase transition of a Grand Unified theory [38]. Although these models led to sufficient exponential expansion, completion of the transition through bubble percolation did not occur. Later models of inflation [39,40], also based on Grand Unified symmetry breaking, through second order transitions were also doomed. While they successfully inflated and reheated, and in fact produced density perturbations due to quantum fluctuations during the evolution of the scalar field, they predicted density perturbations many orders of magnitude too large. Most models today are based on an unknown symmetry breaking involving a new scalar field, the inflaton, ϕ.

19.3.6. *Baryogenesis* :

The Universe appears to be populated exclusively with matter rather than antimatter. Indeed antimatter is only detected in accelerators or in cosmic rays. However, the presence of antimatter in the latter is understood to be the result of collisions of primary particles in the interstellar medium. There is in fact strong evidence against primary forms of antimatter in the Universe. Furthermore, the density of baryons compared to the density of photons is extremely small, $\eta \sim 10^{-10}$.

The production of a net baryon asymmetry requires baryon number violating interactions, C and CP violation and a departure from thermal equilibrium [41]. The first two of these ingredients are expected to be contained in grand unified theories as well as in the non-perturbative sector of the standard model, the third can be realized in an expanding universe where as we have seen interactions come in and out of equilibrium.

There are several interesting and viable mechanisms for the production of the baryon asymmetry. While, we can not review any of them here in any detail, we mention some of the important scenarios. In all cases, all three ingredients listed above are incorporated. One of the first mechanisms was based on the out of equilibrium decay of a massive particle such as a superheavy GUT gauge of Higgs boson [42,43]. A novel mechanism involving the decay of flat directions in supersymmetric models is known as the

Affleck-Dine scenario [44]. Recently, much attention has been focused on the possibility of generating the baryon asymmetry at the electro-weak scale using the non-perturbative interactions of sphalerons [45]. Because these interactions conserve the sum of baryon and lepton number, $B + L$, it is possible to first generate a lepton asymmetry (*e.g.*, by the out-of-equilibrium decay of a superheavy right-handed neutrino), which is converted to a baryon asymmetry at the electro-weak scale [46]. This mechanism is known as lepto-baryogenesis.

19.3.7. *Nucleosynthesis* :

An essential element of the standard cosmological model is Big-Bang nucleosynthesis (BBN), the theory which predicts the abundances of the light element isotopes D, ^{3}He, ^{4}He, and ^{7}Li. Nucleosynthesis takes place at a temperature scale of order 1 MeV. The nuclear processes lead primarily to ^{4}He, with a primordial mass fraction of about 24%. Lesser amounts of the other light elements are produced: about 10^{-5} of D and ^{3}He and about 10^{-10} of ^{7}Li by number relative to H. The abundances of the light elements depend almost solely on one key parameter, the baryon-to-photon ratio, η. The nucleosynthesis predictions can be compared with observational determinations of the abundances of the light elements. Consistency between theory and observations leads to a conservative range of

$$3.4 \times 10^{-10} < \eta < 6.9 \times 10^{-10} \ . \tag{19.54}$$

η is related to the fraction of Ω contained in baryons, Ω_b

$$\Omega_b = 3.66 \times 10^7 \eta \, h^{-2} \ , \tag{19.55}$$

or $10^{10}\eta = 274\Omega_b h^2$. The WMAP result [28] for $\Omega_b h^2$ of 0.0224 ± 0.0009 translates into a value of $\eta = 6.15 \pm 0.25$. This result can be used to 'predict' the light element abundance which can in turn be compared with observation [47]. The resulting D/H abundance is in excellent agreement with that found in quasar absorption systems. It is in reasonable agreement with the helium abundance observed in extra-galactic HII regions (once systematic uncertainties are accounted for) but is in poor agreement with the Li abundance observed in the atmospheres of halo dwarf stars. (See the review on BBN—Sec. 20 of this *Review* for a detailed discussion of BBN or references [48,49].)

19.3.8. *The transition to a matter-dominated Universe* :

In the Standard Model, the temperature (or redshift) at which the Universe undergoes a transition from a radiation dominated to a matter dominated Universe is determined by the amount of dark matter. Assuming three nearly massless neutrinos, the energy density in radiation at temperatures $T \ll 1$ MeV, is given by

$$\rho_r = \frac{\pi^2}{30} \left[2 + \frac{21}{4} \left(\frac{4}{11} \right)^{4/3} \right] T^4 \ . \tag{19.56}$$

In the absence of non-baryonic dark matter, the matter density can be written as

$$\rho_m = m_{\rm N} \eta \, n_\gamma \ , \tag{19.57}$$

where $m_{\rm N}$ is the nucleon mass. Recalling that $n_\gamma \propto T^3$ [cf. Eq. (19.39)], we can solve for the temperature or redshift at the matter-radiation equality when $\rho_r = \rho_m$,

$$T_{eq} = 0.22\, m_{\rm N}\, \eta \qquad \text{or} \qquad (1 + z_{\rm eq}) = 0.22\, \eta \frac{m_{\rm N}}{T_0}\ , \tag{19.58}$$

where T_0 is the present temperature of the microwave background. For $\eta = 5 \times 10^{-10}$, this corresponds to a temperature $T_{\rm eq} \simeq 0.1$ eV or $(1 + z_{\rm eq}) \simeq 425$. A transition this late is very problematic for structure formation (see Sec. 19.4.5).

The redshift of matter domination can be pushed back significantly if non-baryonic dark matter is present. If instead of Eq. (19.57), we write

$$\rho_m = \Omega_m \rho_c \left(\frac{T}{T_0}\right)^3\ , \tag{19.59}$$

we find that

$$T_{eq} = 0.9 \frac{\Omega_m \rho_c}{T_0^3} \qquad \text{or} \qquad (1 + z_{eq}) = 2.4 \times 10^4 \Omega_m h^2\ . \tag{19.60}$$

19.4. The Universe at late times

19.4.1. *The CMB* :

One form of the infamous Olbers' paradox says that, in Euclidean space, surface brightness is independent of distance. Every line of sight will terminate on matter that is hot enough to be ionized and so scatter photons: $T \gtrsim 10^3$ K; the sky should therefore shine as brightly as the surface of the Sun. The reason the night sky is dark is entirely due to the expansion, which cools the radiation temperature to 2.73 K. This gives a Planck function peaking at around 1 mm to produce the microwave background (CMB).

The CMB spectrum is a very accurate match to a Planck function [50]. (See the review on CBR–Sec. 23 of this *Review*.) The COBE estimate of the temperature is [51]

$$T = 2.725 \pm 0.002\,{\rm K}\ . \tag{19.61}$$

The lack of any distortion of the Planck spectrum is a strong physical constraint. It is very difficult to account for in any expanding universe other than one that passes through a hot stage. Alternative schemes for generating the radiation, such as thermalization of starlight by dust grains, inevitably generate a superposition of temperatures. What is required in addition to thermal equilibrium is that $T \propto 1/R$, so that radiation from different parts of space appears identical.

Although it is common to speak of the CMB as originating at "recombination," a more accurate terminology is the era of "last scattering," In practice, this takes place at $z \simeq 1100$, almost independently of the main cosmological parameters, at which time the fractional ionization is very small. This occurred when the age of the Universe was a few hundred thousand years. (See the review on CBR–Sec. 23 of this *Review* for a full discussion of the CMB.)

19.4.2. ***Matter in the Universe*** :

One of the main tasks of cosmology is to measure the density of the Universe, and how this is divided between dark matter and baryons. The baryons consist partly of stars, with $0.002 \lesssim \Omega_* \lesssim 0.003$ [52] but mainly inhabit the IGM. One powerful way in which this can be studied is via the absorption of light from distant luminous objects such as quasars. Even very small amounts of neutral hydrogen can absorb rest-frame UV photons (the Gunn-Peterson effect), and should suppress the continuum by a factor $\exp(-\tau)$, where

$$\tau \simeq \left[\frac{n_{\rm HI}(z)}{(1+z)\sqrt{1+\Omega_m z}}\right] \, / \, 10^{-4.62} \, h\,{\rm m}^{-3}, \tag{19.62}$$

and this expression applies while the Universe is matter dominated ($z \gtrsim 1$ in the $\Omega_m = 0.3$ $\Omega_v = 0.7$ model). It is possible that this general absorption has now been seen at $z = 6.2$ [53]. In any case, the dominant effect on the spectrum is a 'forest' of narrow absorption lines, which produce a mean $\tau = 1$ in the Lyα forest at about $z = 3$, and so we have $\Omega_{\rm HI} \simeq 10^{-5.5} h^{-1}$. This is such a small number that clearly the IGM is very highly ionized at these redshifts.

The Lyα forest is of great importance in pinning down the abundance of deuterium. Because electrons in deuterium differ in reduced mass by about 1 part in 4000 compared to Hydrogen, each absorption system in the Lyα forest is accompanied by an offset deuterium line. By careful selection of systems with an optimal HI column density, a measurement of the D/H ratio can be made. This has now been done in 5 quasars, with relatively consistent results [49]. Combining these determinations with the theory of primordial nucleosynthesis yields a baryon density of $\Omega_b h^2 = 0.021 \pm 0.004$ (95% confidence). (See also the review on BBN—Sec. 20 of this *Review*.)

Ionized IGM can also be detected in emission when it is densely clumped, via bremsstrahlung radiation. This generates the spectacular X-ray emission from rich clusters of galaxies. Studies of this phenomenon allow us to achieve an accounting of the total baryonic material in clusters. Within the central $\simeq 1$ Mpc, the masses in stars, X-ray emitting gas and total dark matter can be determined with reasonable accuracy (perhaps 20% rms), and this allows a minimum baryon fraction to be determined [54,55]:

$$\frac{M_{\rm baryons}}{M_{\rm total}} \gtrsim 0.009 + (0.066 \pm 0.003)\, h^{-3/2} \, . \tag{19.63}$$

Because clusters are the largest collapsed structures, it is reasonable to take this as applying to the Universe as a whole. This equation implies a minimum baryon fraction of perhaps 12% (for reasonable h), which is too high for $\Omega_m = 1$ if we take $\Omega_b h^2 \simeq 0.02$ from nucleosynthesis. This is therefore one of the more robust arguments in favor of $\Omega_m \simeq 0.3$. (See the review on Global cosmological parameters—Sec. 21 of this *Review*.) This argument is also consistent with the inference on Ω_m that can be made from Fig. 19.2.

This method is much more robust than the older classical technique for weighing the Universe: '$L \times M/L$'. The overall light density of the Universe is reasonably well determined from redshift surveys of galaxies, so that a good determination of mass M

and luminosity L for a single object suffices to determine Ω_m *if* the mass-to-light ratio is universal.

Galaxy redshift surveys allow us to deduce the galaxy luminosity function, ϕ, which is the comoving number density of galaxies; this may be described by the Schechter function, which is a power law with an exponential cutoff:

$$\phi = \phi^* \left(\frac{L}{L^*}\right)^{-\alpha} e^{-L/L^*} \frac{dL}{L^*} \tag{19.64}$$

The total luminosity density produced by integrating over the distribution is

$$\rho_L = \phi^* L^* \Gamma(2-\alpha) \ , \tag{19.65}$$

and this tells us the average mass-to-light ratio needed to close the Universe. Answers vary (principally owing to uncertainties in ϕ^*). In blue light, the total luminosity density is $\rho_{\rm L} = 2 \pm 0.2 \times 10^8 \ hL_\odot {\rm Mpc}^{-3}$ [56,57]. The critical density is $2.78 \times 10^{11} \ \Omega h^2 M_\odot {\rm Mpc}^{-3}$, so the critical M/L for closure is

$$(M/L)_{\rm crit,\,B} = 1390\, h \pm 10\% \ . \tag{19.66}$$

Dynamical determinations of mass on the largest accessible scales consistently yield blue M/L values of at least $300\,h$, but normally fall short of the closure value [58]. This was a long-standing argument against the $\Omega_m = 1$ model, but it was never conclusive because the stellar populations in objects such as rich clusters (where the masses can be determined) differ systematically from those in other regions.

19.4.3. *Gravitational lensing* :

A robust method for determining masses in cosmology is to use gravitational light deflection. Most systems can be treated as a geometrically thin gravitational lens, where the light bending is assumed to take place only at a single distance. Simple geometry then determines a mapping between the coordinates in the intrinsic source plane and the observed image plane:

$$\alpha(D_{\rm L}\theta_{\rm I}) = \frac{D_{\rm S}}{D_{\rm LS}}(\theta_{\rm I} - \theta_{\rm S}) \ , \tag{19.67}$$

where the angles $\theta_{\rm I}, \theta_{\rm S}$ and α are in general two-dimensional vectors on the sky. The distances $D_{\rm LS}$ etc. are given by an extension of the usual distance-redshift formula:

$$D_{\rm LS} = \frac{R_0 S_k(\chi_{\rm S} - \chi_{\rm L})}{1 + z_{\rm S}} \ . \tag{19.68}$$

This is the angular-diameter distance for objects on the source plane as perceived by an observer on the lens.

Solutions of this equation divide into weak lensing, where the mapping between source plane and image plane is one-to-one, and strong lensing, in which multiple imaging is

possible. For circularly-symmetric lenses, an on-axis source is multiply imaged into a 'caustic' ring, whose radius is the Einstein radius:

$$\begin{aligned}\theta_{\rm E} &= \left(4GM\,\frac{D_{\rm LS}}{D_{\rm L}D_{\rm S}}\right)^{1/2} \\ &= \left(\frac{M}{10^{11.09}M_\odot}\right)^{1/2}\,\left(\frac{D_{\rm L}D_{\rm S}/D_{\rm LS}}{\rm Gpc}\right)^{-1/2}\quad {\rm arcsec}\end{aligned} \tag{19.69}$$

The observation of 'arcs' (segments of near-perfect Einstein rings) in rich clusters of galaxies has thus given very accurate masses for the central parts of clusters—generally in good agreement with other indicators, such as analysis of X-ray emission from the cluster IGM [59].

19.4.4. *Density Fluctuations* :

The overall properties of the Universe are very close to being homogeneous; and yet telescopes reveal a wealth of detail on scales varying from single galaxies to large-scale structures of size exceeding 100 Mpc. The existence of these structures must be telling us something important about the initial conditions of the Big Bang, and about the physical processes that have operated subsequently. This motivates the study of the density perturbation field, defined as

$$\delta(\mathbf{x}) \equiv \frac{\rho(\mathbf{x}) - \langle\rho\rangle}{\langle\rho\rangle}\ . \tag{19.70}$$

A critical feature of the δ field is that it inhabits a universe that is isotropic and homogeneous in its large-scale properties. This suggests that the statistical properties of δ should also be statistically homogeneous—*i.e.*, it is a stationary random process.

It is often convenient to describe δ as a Fourier superposition:

$$\delta(\mathbf{x}) = \sum \delta_{\mathbf{k}} e^{-i\mathbf{k}\cdot\mathbf{x}}\ . \tag{19.71}$$

We avoid difficulties with an infinite universe by applying periodic boundary conditions in a cube of some large volume V. The cross-terms vanish when we compute the variance in the field, which is just a sum over modes of the power spectrum

$$\langle\delta^2\rangle = \sum |\delta_{\mathbf{k}}|^2 \equiv \sum P(k)\ . \tag{19.72}$$

Note that the statistical nature of the fluctuations must be isotropic, so we write $P(k)$ rather than $P(\mathbf{k})$. The $\langle\ldots\rangle$ average here is a volume average. Cosmological density fields are an example of an ergodic process, in which the average over a large volume tends to the same answer as the average over a statistical ensemble.

The statistical properties of discrete objects sampled from the density field are often described in terms of N-point correlation functions, which represent the excess probability over random for finding one particle in each of N boxes in a given configuration.

For the 2-point case, the correlation function is readily shown to be identical to the autocorrelation function of the δ field: $\xi(r) = \langle \delta(x)\delta(x+r)\rangle$.

The power spectrum and correlation function are Fourier conjugates, and thus are equivalent descriptions of the density field (similarly, k-space equivalents exist for the higher-order correlations). It is convenient to take the limit $V \to \infty$ and use k-space integrals, defining a dimensionless power spectrum as $\Delta^2(k) = d\langle\delta^2\rangle/d\ln k = Vk^3P(k)/2\pi^2$:

$$\xi(r) = \int \Delta^2(k)\,\frac{\sin kr}{kr}\,d\ln k; \quad \Delta^2(k) = \frac{2}{\pi}k^3\int_0^\infty \xi(r)\,\frac{\sin kr}{kr}\,r^2\,dr\ . \tag{19.73}$$

For many years, an adequate approximation to observational data on galaxies was $\xi = (r/r_0)^{-\gamma}$, with $\gamma \simeq 1.8$ and $r_0 \simeq 5\,h^{-1}\,$Mpc. Modern surveys are now able to probe into the large-scale linear regime where traces of the curved primordial spectrum can be detected [60,61].

19.4.5. *Formation of cosmological structure* :

The simplest model for the generation of cosmological structure is gravitational instability acting on some small initial fluctuations (for the origin of which a theory such as inflation is required). If the perturbations are adiabatic (*i.e.*, fractionally perturb number densities of photons and matter equally), the linear growth law for matter perturbations is simple:

$$\delta \propto \begin{cases} a(t)^2 & \text{(radiation domination; } \Omega_{\rm r} = 1) \\ a(t) & \text{(matter domination; } \Omega_{\rm m} = 1) \end{cases} \tag{19.74}$$

For low density universes, the present-day amplitude is suppressed by a factor $g(\Omega)$, where

$$g(\Omega) \simeq \frac{5}{2}\Omega_m\left[\Omega_m^{4/7} - \Omega_v + (1+\Omega_m/2)(1+\frac{1}{70}\Omega_v)\right]^{-1}\ , \tag{19.75}$$

is an accurate fit for models with matter plus cosmological constant. The alternative perturbation mode is isocurvature: only the equation of state changes, and the total density is initially unperturbed. These modes perturb the total entropy density, and thus induce additional large-scale CMB anisotropies [62]. Although the character of perturbations in the simplest inflationary theories are purely adiabatic, correlated adiabatic and isocurvature modes are predicted in many models; the simplest example is the curvaton, which is a scalar field that decays to yield a perturbed radiation density. If the matter content already exists at this time, the overall perturbation field will have a significant isocurvature component. Such a prediction is inconsistent with current CMB data [63], and most analyses of CMB and LSS data assume the adiabatic case to hold exactly.

Linear evolution preserves the shape of the power spectrum. However, a variety of processes mean that growth actually depends on the matter content:

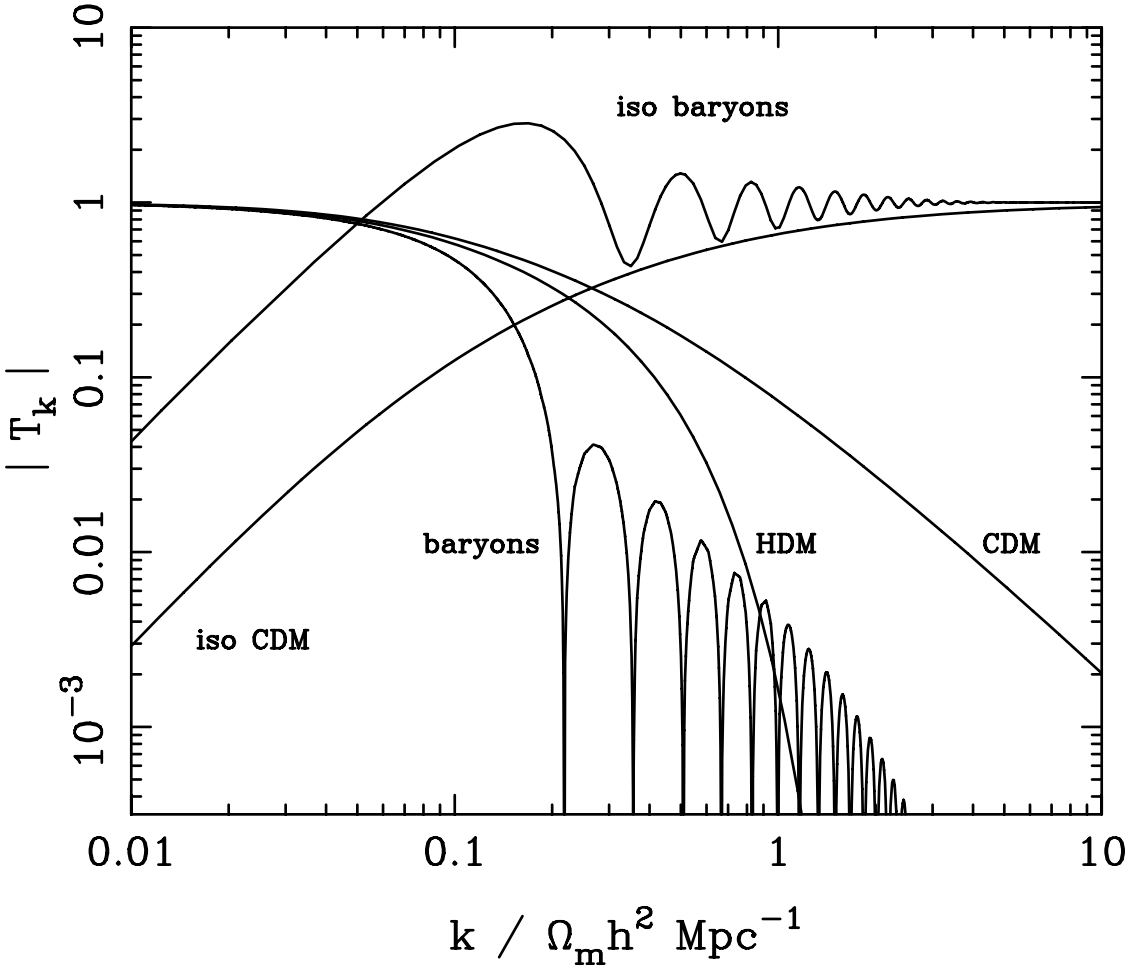


Figure 19.4: A plot of transfer functions for various models. For adiabatic models, $T_k \to 1$ at small k, whereas the opposite is true for isocurvature models. For dark-matter models, the characteristic wavenumber scales proportional to $\Omega_m h^2$. The scaling for baryonic models does not obey this exactly; the plotted cases correspond to $\Omega_m = 1$, $h = 0.5$.

(1) Pressure opposes gravity effectively for wavelengths below the horizon length while the Universe is radiation dominated. The *comoving* horizon size at $z_{\rm eq}$ is therefore an important scale:

$$D_{\rm H}(z_{\rm eq}) = \frac{2(\sqrt{2}-1)}{(\Omega_m z_{\rm eq})^{1/2} H_0} = \frac{16.0}{\Omega_m h^2}\,{\rm Mpc} \tag{19.76}$$

(2) At early times, dark matter particles will undergo free streaming at the speed of light, and so erase all scales up to the horizon—a process that only ceases when the particles go nonrelativistic. For light massive neutrinos, this happens at $z_{\rm eq}$; all structure up to the horizon-scale power-spectrum break is in fact erased. Hot(cold) dark matter models are thus sometimes dubbed large(small)-scale damping models.

(3) A further important scale arises where photon diffusion can erase perturbations in the matter–radiation fluid; this process is named Silk damping.

The overall effect is encapsulated in the transfer function, which gives the ratio of the late-time amplitude of a mode to its initial value (see Fig. 19.4). The overall power spectrum is thus the primordial power-law, times the square of the transfer function:

$$P(k) \propto k^n \, T_k^2 \ . \tag{19.77}$$

The most generic power-law index is $n = 1$: the 'Zeldovich' or 'scale-invariant' spectrum. Inflationary models tend to predict a small 'tilt': $|n - 1| \lesssim 0.03$ [13,14]. On the assumption that the dark matter is cold, the power spectrum then depends on 5 parameters: n, h, Ω_b, $\Omega_{\rm cdm}$ ($\equiv \Omega_m - \Omega_b$) and an overall amplitude. The latter is often specified as σ_8, the linear-theory fractional rms in density when a spherical filter of radius $8\,h^{-1}\,$Mpc is applied in linear theory. This scale can be probed directly via weak gravitational lensing, and also via its effect on the abundance of rich galaxy clusters. The favored value is approximately [64,65]

$$\sigma_8 = (0.7 \pm 15\%)\,(\Omega_m/0.3)^{-0.5} \ . \tag{19.78}$$

A direct measure of mass inhomogeneity is valuable, since the galaxies inevitably are biased with respect to the mass. This means that the fractional fluctuations in galaxy number, $\delta n/n$ may differ from the mass fluctuations, $\delta\rho/\rho$. It is commonly assumed that the two fields obey some proportionality on large scales where the fluctuations are small, $\delta n/n = b\delta\rho/\rho$, but even this is not guaranteed [66].

The main shape of the transfer function is a break around the horizon scale at $z_{\rm eq}$, which depends just on $\Omega_m h$ when wavenumbers are measured in observable units ($h\,$Mpc^{-1}). In principle, accurate data over a wide range of k could determine both Ωh and n, but in practice there is a strong degeneracy between these. For reasonable baryon content, weak oscillations in the transfer function may be visible, giving an alternative means of fixing the baryon content. Current data [60,61] favor $\Omega_m h \simeq 0.20$ and a baryon fraction of about 0.15 for $n = 1$ (see Fig. 19.5). In order to constrain n itself, it is necessary to examine data on anisotropies in the CMB.

19.4.6. *CMB anisotropies* :

The CMB has a clear dipole anisotropy, of magnitude 1.23×10^{-3}. This is interpreted as being due to the Earth's motion, which is equivalent to a peculiar velocity for the Milky Way of

$$v_{\rm MW} \simeq 600\,{\rm km\,s^{-1}} \quad \text{towards} \quad (\ell, b) \simeq (270^\circ, 30^\circ) \ . \tag{19.79}$$

All higher-order multipole moments of the CMB are however much smaller (of order 10^{-5}), and interpreted as signatures of density fluctuations at last scattering ($\simeq 1100$). To analyze these, the sky is expanded in spherical harmonics as explained in the review on CBR–Sec. 23 of this *Review*. The dimensionless power per ln k or 'bandpower' for the CMB is defined as

$$\mathcal{T}^2(\ell) = \frac{\ell(\ell+1)}{2\pi}\, C_\ell \ . \tag{19.80}$$

This function encodes information from the three distinct mechanisms that cause CMB anisotropies:

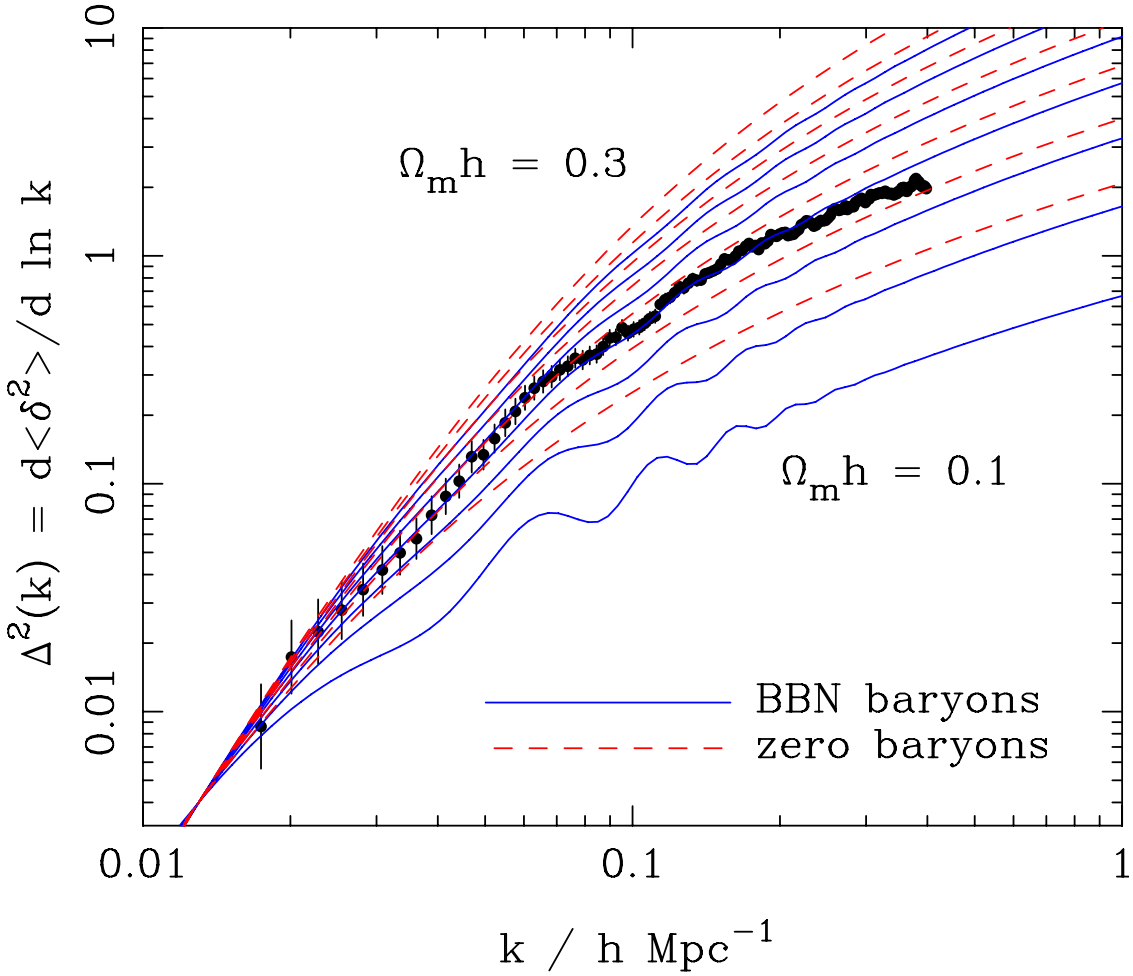


Figure 19.5: The galaxy power spectrum from the 2dFGRS, shown in dimensionless form, $\Delta^2(k) \propto k^3 P(k)$. The solid points with error bars show the power estimate. The window function correlates the results at different k values, and also distorts the large-scale shape of the power spectrum An approximate correction for the latter effect has been applied. The solid and dashed lines show various CDM models, all assuming $n = 1$. For the case with non-negligible baryon content, a big-bang nucleosynthesis value of $\Omega_b h^2 = 0.02$ is assumed, together with $h = 0.7$. A good fit is clearly obtained for $\Omega_m h \simeq 0.2$. See full-color version on color pages at end of book.

(1) Gravitational (Sachs–Wolfe) perturbations. Photons from high-density regions at last scattering have to climb out of potential wells, and are thus redshifted.

(2) Intrinsic (adiabatic) perturbations. In high-density regions, the coupling of matter and radiation can compress the radiation also, giving a higher temperature.

(3) Velocity (Doppler) perturbations. The plasma has a non-zero velocity at recombination, which leads to Doppler shifts in frequency and hence shifts in brightness temperature.

Because the potential fluctuations obey Poisson's equation, $\nabla^2\Phi = 4\pi G\rho\delta$, and the velocity field satisfies the continuity equation $\nabla \cdot \mathbf{u} = -\dot{\delta}$, the resulting different powers

of k ensure that the Sachs-Wolfe effect dominates on large scales and adiabatic effects on small scales.

The relation between angle and comoving distance on the last-scattering sphere requires the comoving angular-diameter distance to the last-scattering sphere; because of its high redshift, this is effectively identical to the horizon size at the present epoch, $D_{\rm H}$:

$$\begin{aligned} D_{\rm H} &= \frac{2}{\Omega_m H_0} \quad (\Omega_v = 0) \\ D_{\rm H} &\simeq \frac{2}{\Omega_m^{0.4} H_0} \quad (\text{flat} : \Omega_m + \Omega_v = 1) \end{aligned} \tag{19.81}$$

These relations show how the CMB is strongly sensitive to curvature: the horizon length at last scattering is $\propto 1/\sqrt{\Omega_m}$, so that this subtends an angle that is virtually independent of Ω_m for a flat model. Observations of a peak in the CMB power spectrum at relatively large scales ($\ell \simeq 225$) are thus strongly inconsistent with zero-Λ models with low density: current CMB data require $\Omega_m + \Omega_v \simeq 1 \pm 0.05$. (See *e.g.*, Fig. 19.2).

In addition to curvature, the CMB encodes information about several other key cosmological parameters. Within the compass of simple adiabatic CDM models, there are 9 of these:

$$\omega_c,\ \omega_b,\ \Omega_t,\ h,\ \tau,\ n_s,\ n_t,\ r,\ Q\ . \tag{19.82}$$

The symbol ω denotes the physical density, Ωh^2: the transfer function depends only on the densities of CDM (ω_c) and baryons (ω_b). Transcribing the power spectrum at last scattering into an angular power spectrum brings in the total density parameter ($\Omega_t \equiv \Omega_m + \Omega_v = \Omega_c + \Omega_b + \Omega_v$) and h: there is an exact geometrical degeneracy [67] between these that keeps the angular-diameter distance to last scattering invariant, so that models with substantial spatial curvature and large vacuum energy cannot be ruled out without prior knowledge of the Hubble parameter. Alternatively, the CMB alone cannot measure the Hubble parameter.

The other main parameter degeneracy involves the tensor contribution to the CMB anisotropies. These are important at large scales (up to the horizon scales); for smaller scales, only scalar fluctuations (density perturbations) are important. Each of these components is characterized by a spectral index, n, and a ratio between the power spectra of tensors and scalars (r). Finally, the overall amplitude of the spectrum must be specified (Q), together with the optical depth to Compton scattering owing to recent reionization (τ). The tensor degeneracy operates as follows: the main effect of adding a large tensor contribution is to reduce the contrast between low ℓ and the peak at $\ell \simeq 225$ (because the tensor spectrum has no acoustic component). The required height of the peak can be recovered by increasing n_s to increase the small-scale power in the scalar component; this in turn over-predicts the power at $\ell \sim 1000$, but this effect can be counteracted by raising the baryon density [68]. In order to break this degeneracy, additional data are required. For example, an excellent fit to the CMB data is obtained with a scalar-only model with zero curvature and $\omega_b = 0.023$, $\omega_c = 0.134$, $h = 0.73$, $n_s = 0.97$ [28], but this is indistinguishable from a model where tensors dominate at $\ell \lesssim 100$, but we raise ω_b to

0.03 and n_s to 1.2. This baryon density is too high for nucleosynthesis, which disfavors the high-tensor solution [69].

The reason the tensor component is introduced, and why it is so important, is that it is the only non-generic prediction of inflation. Slow-roll models of inflation involve two dimensionless parameters:

$$\epsilon \equiv \frac{M_{\rm P}^2}{16\pi}\,(V'/V)^2 \qquad \eta \equiv \frac{M_{\rm P}^2}{8\pi}\,(V''/V)\ , \tag{19.83}$$

where V is the inflaton potential, and dashes denote derivatives with respect to the inflation field. In terms of these, the tensor-to-scalar ratio is $r \simeq 12\epsilon$, and the spectral indices are $n_s = 1-6\epsilon+2\eta$ and $n_t = -2\epsilon$. The natural expectation of inflation is that the quasi-exponential phase ends once the slow-roll parameters become significantly non-zero, so that both $n_s \neq 1$ and a significant tensor component are expected. These prediction can be avoided in some models, but it is undeniable that observation of such features would be a great triumph for inflation. Much future effort in cosmology will therefore be directed towards the question of whether the Universe contains anything other than scale-invariant scalar fluctuations.

References:

1. V.M. Slipher, Pop. Astr. **23**, 21 (1915).
2. K. Lundmark, MNRAS **84**, 747 (1924).
3. E. Hubble and M.L. Humason, Ap. J. **74**, 43 (1931).
4. G. Gamow, Phys. Rev. **70**, 572 (1946).
5. R.A. Alpher *et al.*, Phys. Rev. **73**, 803 (1948).
6. R.A. Alpher and R.C. Herman, Phys. Rev. **74**, 1737 (1948).
7. R.A. Alpher and R.C. Herman, Phys. Rev. **75**, 1089 (1949).
8. A.A. Penzias and R.W. Wilson, Ap. J. **142**, 419 (1965).
9. S. Weinberg, *Gravitation and Cosmology*, John Wiley and Sons (1972).
10. P.J.E. Peebles, *Principles of Physical Cosmology* Princeton University Press (1993).
11. G. Börner, *The Early Universe: Facts and Fiction*, Springer-Verlag (1988).
12. E.W. Kolb and M.S. Turner, *The Early Universe*, Addison-Wesley (1990).
13. J.A. Peacock, *Cosmological Physics*, Cambridge Univ. Press (1999).
14. A.R. Liddle and D. Lyth, *Cosmological Inflation and Large-Scale Structure*, Cambridge University Press (2000).
15. E.B. Gliner, Sov. Phys. JETP 22, 378 (1966).
16. Y.B. Zeldovich, (1967), Sov. Phys. Uspekhi, **11**, 381 (1968).
17. P.M. Garnavich *et al.*, Ap. J. **507**, 74 (1998).
18. S. Perlmutter *et al.*, Phys. Rev. Lett. **83**, 670 (1999).
19. I. Maor *et al.*, Phys. Rev. **D65**, 123003 (2002).
20. A.G. Riess *et al.*, A. J. **116**, 1009 (1998).
21. S. Perlmutter *et al.*, Ap. J. **517**, 565 (1999).
22. A.G. Riess, PASP, **112**, 1284 (2000).
23. J.L. Tonry *et al.*, Ap. J. **594**, 1 (2003).
24. W.L. Freedman *et al.*, ApJ **553**, 47 (2001).

25. J.A. Johnson and M. Bolte, ApJ **554**, 888 (2001).
26. R. Jimenez and P. Padoan, ApJ **498**, 704 (1998).
27. E. Carretta *et al.*, ApJ **533**, 215 (2000).
28. D.N. Spergel *et al.*, Ap. J. Supp. **148**, 175 (2003).
29. M. Srednicki *et al.*, Nucl. Phys. B **310**, 693 (1988).
30. G. Mangano *et al.*, Phys. Lett. **B534**, 8 (2002).
31. A. Linde, Phys. Rev. **D14**, 3345 (1976).
32. A. Linde, Rep. Prog. Phys. **42**, 389 (1979).
33. C.E. Vayonakis, Surveys High Energ. Phys. **5**, 87 (1986).
34. S.A. Bonometto and A. Masiero, Riv. Nuovo Cim. **9N5**, 1 (1986).
35. A. Linde, A.D. Linde, *Particle Physics And Inflationary Cosmology*, Harwood (1990).
36. K.A. Olive, Phys. Rep. **190**, 3345 (1990).
37. D. Lyth and A. Riotto, Phys. Rep. **314**, 1 (1999).
38. A.H. Guth, Phys. Rev. **D23**, 347 (1981).
39. A.D. Linde, Phys. Lett. **108B**, 389 (1982).
40. A. Albrecht and P.J. Steinhardt, Phys. Rev. Lett. **48**, 1220 (1982).
41. A.D. Sakharov, Sov. Phys. JETP Lett. **5**, 24 (1967).
42. S. Weinberg, Phys. Rev. Lett. **42**, 850 (1979).
43. D. Toussaint *et al.*, Phys. Rev. **D19**, 1036 (1979).
44. I. Affleck and M. Dine, Nucl. Phys. **B249**, 361 (1985).
45. V. Kuzmin *et al.*, Phys. Lett. **B155**, 36 (1985).
46. M. Fukugita and T. Yanagida, Phys. Lett. **B174**, 45 (1986).
47. R.H. Cyburt *et al.*, Phys. Lett. **B567**, 227 (2003).
48. K.A. Olive *et al.*, Phys. Rept. **333**, 389 (2000).
49. D. Kirkman *et al.*, Ap. J. Supp. **149**, 1 (2003).
50. D.J. Fixsen *et al.*, ApJ **473**, 576 (1996).
51. J.C. Mather *et al.*, ApJ **512**, 511 (1999).
52. S.M. Cole *et al.*, MNRAS **326**, 255 (2001).
53. R.H. Becker *et al.*, Astr. J. **122**, 2850 (2001).
54. S.D.M. White *et al.*, Nature **366**, 429 (1993).
55. S.W. Allen *et al.*, MNRAS, **334**, L11 (2002).
56. S. Folkes *et al.*, MNRAS **308**, 459 (1999).
57. M.L. Blanton *et al.*, Astr. J. **121**, 2358 (2001).
58. H. Hoekstra *et al.*, ApJ, **548**, L5 (2001).
59. S.W. Allen, MNRAS **296**, 392 (1998).
60. W.J. Percival *et al.*, MNRAS **327**, 1297 (2001).
61. D.J. Eisenstein *et al.*, astro-ph/0501171.
62. G. Efstathiou and J.R. Bond, MNRAS **218**, 103 (1986).
63. C. Gordon and A. Lewis, Phys. Rev. **D67**, 123513 (2003).
64. M.L. Brown *et al.*, MNRAS, **341**, 1 (2003).
65. P.T.P. Viana, R.C. Nichol, A.R. Liddle, ApJ, **569**, L75 (2002).
66. A. Dekel and O. Lahav, ApJ **520**, 24 (1999).
67. G. Efstathiou and J.R. Bond, MNRAS, **304**, 75 (1999).

68. G.P. Efstathiou *et al.*, MNRAS **330**, L29 (2002).
69. G.P. Efstathiou, MNRAS **332**, 193 (2002).

20. BIG-BANG NUCLEOSYNTHESIS

Revised October 2005 by B.D. Fields (Univ. of Illinois) and S. Sarkar (Univ. of Oxford).

Big-bang nucleosynthesis (BBN) offers the deepest reliable probe of the early universe, being based on well-understood Standard Model physics [1–4]. Predictions of the abundances of the light elements, D, ^{3}He, ^{4}He, and ^{7}Li, synthesized at the end of the "first three minutes" are in good overall agreement with the primordial abundances inferred from observational data, thus validating the standard hot big-bang cosmology (see [5] for a review). This is particularly impressive given that these abundances span nine orders of magnitude — from $^4\mathrm{He/H} \sim 0.08$ down to $^7\mathrm{Li/H} \sim 10^{-10}$ (ratios by number). Thus BBN provides powerful constraints on possible deviations from the standard cosmology [2], and on new physics beyond the Standard Model [3].

20.1. Theory

The synthesis of the light elements is sensitive to physical conditions in the early radiation-dominated era at temperatures $T \lesssim 1$ MeV, corresponding to an age $t \gtrsim 1$ s. At higher temperatures, weak interactions were in thermal equilibrium, thus fixing the ratio of the neutron and proton number densities to be $n/p = \mathrm{e}^{-Q/T}$, where $Q = 1.293$ MeV is the neutron-proton mass difference. As the temperature dropped, the neutron-proton inter-conversion rate, $\Gamma_{n\leftrightarrow p} \sim G_{\mathrm{F}}^2 T^5$, fell faster than the Hubble expansion rate, $H \sim \sqrt{g_* G_{\mathrm{N}}}\, T^2$, where g_* counts the number of relativistic particle species determining the energy density in radiation. This resulted in departure from chemical equilibrium ("freeze-out") at $T_{\mathrm{fr}} \sim (g_* G_{\mathrm{N}}/G_F^4)^{1/6} \simeq 1$ MeV. The neutron fraction at this time, $n/p = \mathrm{e}^{-Q/T_{\mathrm{fr}}} \simeq 1/6$, is thus sensitive to every known physical interaction, since Q is determined by both strong and electromagnetic interactions while T_{fr} depends on the weak as well as gravitational interactions. Moreover the sensitivity to the Hubble expansion rate affords a probe of e.g. the number of relativistic neutrino species [6]. After freeze-out the neutrons were free to β-decay so the neutron fraction dropped to $\simeq 1/7$ by the time nuclear reactions began. A simplified analytic model of freeze-out yields the n/p ratio to an accuracy of $\sim 1\%$ [7,8].

The rates of these reactions depend on the density of baryons (strictly speaking, nucleons), which is usually expressed normalized to the relic blackbody photon density as $\eta \equiv n_{\mathrm{B}}/n_\gamma$. As we shall see, all the light-element abundances can be explained with $\eta_{10} \equiv \eta \times 10^{10}$ in the range 4.7–6.5 (95% CL). With n_γ fixed by the present CMB temperature 2.725 K (see Cosmic Microwave Background review), this can be stated as the allowed range for the baryon mass density today, $\rho_{\mathrm{B}} = (3.2\text{–}4.5) \times 10^{-31}\ \mathrm{g\,cm^{-3}}$, or as the baryonic fraction of the critical density, $\Omega_{\mathrm{B}} = \rho_{\mathrm{B}}/\rho_{\mathrm{crit}} \simeq \eta_{10} h^{-2}/274 = (0.017\text{–}0.024) h^{-2}$, where $h \equiv H_0/100\ \mathrm{km\,s^{-1}\,Mpc^{-1}} = 0.72 \pm 0.08$ is the present Hubble parameter (see Cosmological Parameters review).

The nucleosynthesis chain begins with the formation of deuterium in the process $p(n,\gamma)\mathrm{D}$. However, photo-dissociation by the high number density of photons delays production of deuterium (and other complex nuclei) well after T drops below the binding energy of deuterium, $\Delta_{\mathrm{D}} = 2.23$ MeV. The quantity $\eta^{-1}\mathrm{e}^{-\Delta_{\mathrm{D}}/T}$, i.e. the number of

photons per baryon above the deuterium photo-dissociation threshold, falls below unity at $T \simeq 0.1$ MeV; nuclei can then begin to form without being immediately photo-dissociated again. Only 2-body reactions such as $\mathrm{D}(p,\gamma)^3\mathrm{He}$, $^3\mathrm{He}(\mathrm{D},p)^4\mathrm{He}$, are important because the density has become rather low by this time.

Nearly all the surviving neutrons when nucleosynthesis begins end up bound in the most stable light element ^{4}He. Heavier nuclei do not form in any significant quantity both because of the absence of stable nuclei with mass number 5 or 8 (which impedes nucleosynthesis via n^4He, p^4He or ^{4}He^{4}He reactions) and the large Coulomb barriers for reactions such as $\mathrm{T}(^4\mathrm{He},\gamma)^7\mathrm{Li}$ and $^3\mathrm{He}(^4\mathrm{He},\gamma)^7\mathrm{Be}$. Hence the primordial mass fraction of ^{4}He, conventionally referred to as Y_p, can be estimated by the simple counting argument

$$Y_\mathrm{p} = \frac{2(n/p)}{1+n/p} \simeq 0.25 \ . \tag{20.1}$$

There is little sensitivity here to the actual nuclear reaction rates, which are however important in determining the other "left-over" abundances: D and ^{3}He at the level of a few times 10^{-5} by number relative to H, and ^{7}Li/H at the level of about 10^{-10} (when η_{10} is in the range 1–10). These values can be understood in terms of approximate analytic arguments [8,9]. The experimental parameter most important in determining Y_p is the neutron lifetime, τ_n, which normalizes (the inverse of) $\Gamma_{n\leftrightarrow p}$. (This is not fully determined by G_F alone since neutrons and protons also have strong interactions, the effects of which cannot be calculated very precisely.) The experimental uncertainty in τ_n used to be a source of concern but has recently been reduced substantially: $\tau_n = 885.7 \pm 0.8$ s (see N Baryons listing).

The elemental abundances, calculated using the (publicly available [10]) Wagoner code [1,11], are shown in Fig. 20.1 as a function of η_{10}. The ^{4}He curve includes small corrections due to radiative processes at zero and finite temperature [12], non-equilibrium neutrino heating during $e^\pm$ annihilation [13], and finite nucleon mass effects [14]; the range reflects primarily the 1σ uncertainty in the neutron lifetime. The spread in the curves for D, ^{3}He and ^{7}Li corresponds to the 1σ uncertainties in nuclear cross sections estimated by Monte Carlo methods [15–16]. Recently the input nuclear data have been carefully reassessed [17–21], leading to improved precision in the abundance predictions. Polynomial fits to the predicted abundances and the error correlation matrix have been given [16,22]. The boxes in Fig. 20.1 show the observationally inferred primordial abundances with their associated statistical and systematic uncertainties, as discussed below.

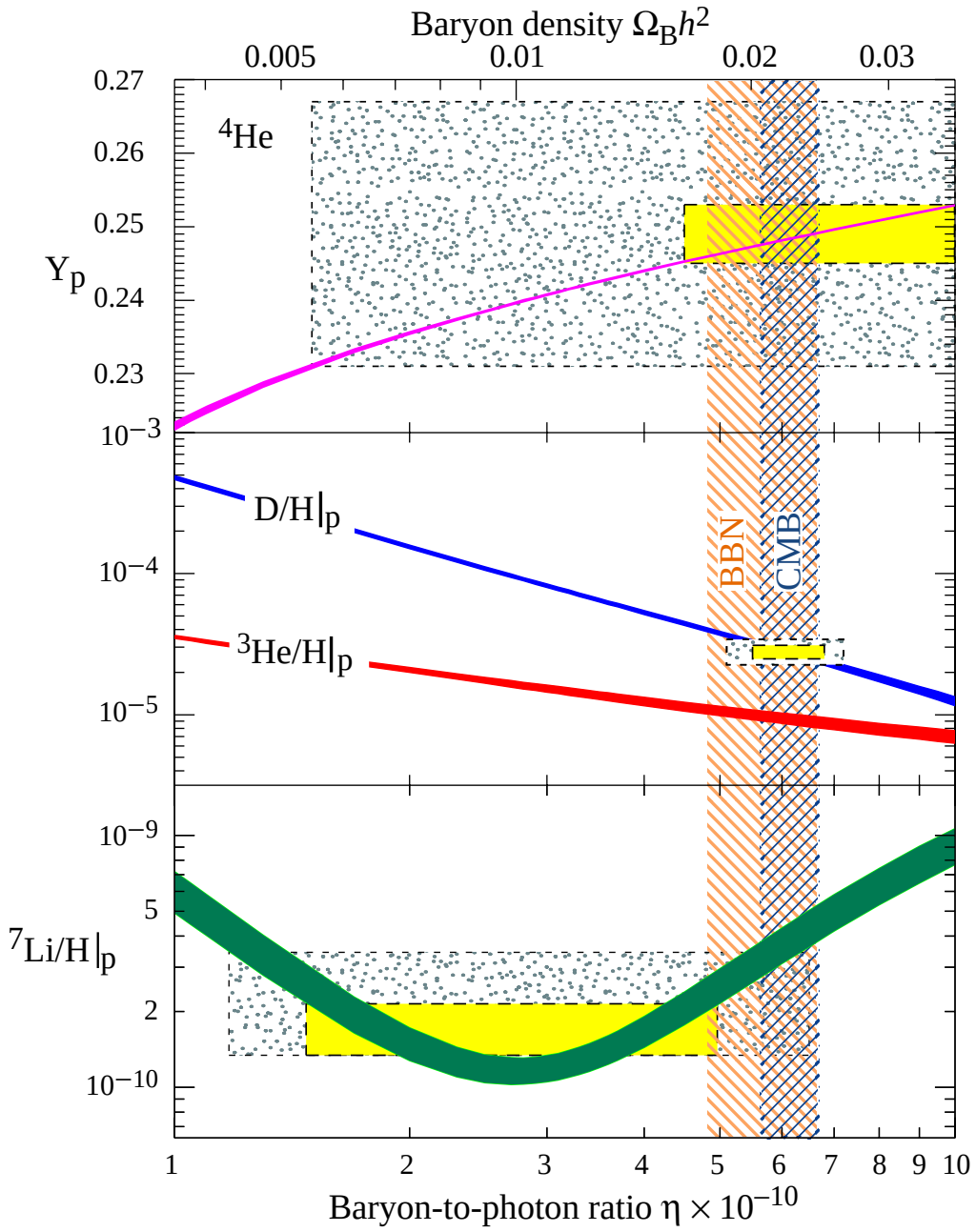


Figure 20.1: The abundances of ^{4}He, D, ^{3}He and ^{7}Li as predicted by the standard model of big-bang nucleosynthesis. Boxes indicate the observed light element abundances (smaller boxes: 2σ statistical errors; larger boxes: $\pm 2\sigma$ statistical and systematic errors). The narrow vertical band indicates the CMB measure of the cosmic baryon density. See full-color version on color pages at end of book.

20.2. Light Element Abundances

BBN theory predicts the universal abundances of D, ^{3}He, ^{4}He, and ^{7}Li, which are essentially determined by $t \sim 180$ s. Abundances are however observed at much later

epochs, after stellar nucleosynthesis has commenced. The ejected remains of this stellar processing can alter the light element abundances from their primordial values, but also produce heavy elements such as C, N, O, and Fe ("metals"). Thus one seeks astrophysical sites with low metal abundances, in order to measure light element abundances which are closer to primordial. For all of the light elements, systematic errors are an important and often dominant limitation to the precision with which primordial abundances can be inferred.

In recent years, high-resolution spectra have revealed the presence of D in high-redshift, low-metallicity quasar absorption systems (QAS), via its isotope-shifted Lyman-α absorption [23–27]. It is believed that there are no astrophysical sources of deuterium [28], so any detection provides a lower limit to primordial D/H and thus an upper limit on η; for example, the local interstellar value of $\mathrm{D/H}|_{\mathrm{p}} = (1.5 \pm 0.1) \times 10^{-5}$ [29] requires $\eta_{10} \leq 9$. Recent observations find an unexpected scatter of a factor of ~ 2 [30] so interstellar D may have been depleted by stellar processing. However, for the high-redshift systems, conventional models of galactic nucleosynthesis (chemical evolution) do not predict D/H depletion [31].

The observed extragalactic D values are bracketed by the non-detection of D in a high-redshift system, $\mathrm{D/H}|_{\mathrm{p}} < 6.7 \times 10^{-5}$ at 1σ [32], and low values in some (damped Lyman-α) systems [24,25]. Averaging the 5 most precise observations of deuterium in QAS gives $\mathrm{D/H} = (2.78 \pm 0.15) \times 10^{-5}$, where the error is statistical only [23,26]. However there remains concern over systematic errors, the dispersion between the values being much larger than is expected from the individual measurement errors ($\chi^2 = 16.5$ for $\nu = 4$ d.o.f.). Increasing the error by a factor $\sqrt{\chi^2/\nu}$ gives, as shown on Fig. 20.1:

$$\mathrm{D/H}|_{\mathrm{p}} = (2.78 \pm 0.29) \times 10^{-5}. \tag{20.2}$$

We observe ^{4}He in clouds of ionized hydrogen (H II regions), the most metal-poor of which are in dwarf galaxies. There is now a large body of data on ^{4}He and CNO in these systems [33]. These data confirm that the small stellar contribution to helium is positively correlated with metal production. Extrapolating to zero metallicity gives the primordial ^{4}He abundance [34]

$$Y_{\mathrm{p}} = 0.249 \pm 0.009. \tag{20.3}$$

Here the latter error is a careful (and significantly enlarged) estimate of the systematic uncertainties; these dominate, and is based on the scatter in different analyses of the physical properties of the H II regions [33,34]. Other extrapolations to zero metallicity give $Y_{\mathrm{p}} = 0.2443 \pm 0.0015$ [33], and $Y_{\mathrm{p}} = 0.2391 \pm 0.0020$ [35]. These are consistent (given the systematic errors) with the above estimate [34], which appears in Fig. 20.1.

The systems best suited for Li observations are metal-poor stars in the spheroid (Pop II) of our Galaxy, which have metallicities going down to at least 10^{-4} and perhaps 10^{-5} of the Solar value [36]. Observations have long shown [37–41] that Li does not vary significantly in Pop II stars with metallicities $\lesssim 1/30$ of Solar — the "Spite plateau" [37]. Recent precision data suggest a small but significant correlation between

Li and Fe [38] which can be understood as the result of Li production from Galactic cosmic rays [39]. Extrapolating to zero metallicity one arrives at a primordial value $\mathrm{Li/H}|_{\rm p} = (1.23 \pm 0.06) \times 10^{-10}$ [40]. One systematic error stems from the differences in techniques to determine the physical parameters (*e.g.*, the temperature) of the stellar atmosphere in which the Li absorption line is formed. Alternative analyses, using methods that give systematically higher temperatures, yield $\mathrm{Li/H}|_{\rm p} = (2.19 \pm 0.28) \times 10^{-10}$ [41] and $\mathrm{Li/H}|_{\rm p} = (2.34 \pm 0.32) \times 10^{-10}$ [42]; the difference with [40] indicates a systematic uncertainty of a factor of ~ 2. Moreover it is possible that the Li in Pop II stars has been partially destroyed, due to mixing of the outer layers with the hotter interior [43]. Such processes can be constrained by the absence of significant scatter in Li-Fe [38], and by observations of the fragile isotope ^{6}Li [39]. Nevertheless, depletion by a factor as large as ~ 1.6 remains allowed [40,43]. Including these systematics, we estimate a primordial Li range, as shown in Fig. 20.1:

$$\mathrm{Li/H}|_{\rm p} = (1.7 \pm 0.02^{+1.1}_{-0}) \times 10^{-10}. \tag{20.4}$$

Finally, we turn to ^{3}He. Here, the only observations available are in the Solar system and (high-metallicity) H II regions in our Galaxy [44]. This makes inference of the primordial abundance difficult, a problem compounded by the fact that stellar nucleosynthesis models for ^{3}He are in conflict with observations [45]. Consequently, it is no longer appropriate to use ^{3}He as a cosmological probe; instead, one might hope to turn the problem around and constrain stellar astrophysics using the predicted primordial ^{3}He abundance [46].

20.3. Concordance, Dark Matter, and the CMB

We now use the observed light element abundances to assess the theory. We first consider standard BBN, which is based on Standard Model physics alone, so $N_\nu = 3$ and the only free parameter is the baryon-to-photon ratio η. (The implications of BBN for physics beyond the Standard Model will be considered below, §4). Thus, any abundance measurement determines η, while additional measurements overconstrain the theory and thereby provide a consistency check.

First we note that the overlap in the η ranges spanned by the larger boxes in Fig. 20.1 indicates overall concordance. More quantitatively, when we account for theoretical uncertainties as well as the statistical and systematic errors in observations, there is acceptable agreement among the abundances when

$$4.7 \leq \eta_{10} \leq 6.5 \ (95\%\ \mathrm{CL}). \tag{20.5}$$

However the agreement is much less satisfactory if we use only the quoted statistical errors in the observations. In particular, as seen in Fig. 20.1, D and ^{4}He are consistent with each other but favour a value of η which is higher by $\sim 2\sigma$ from that indicated by the ^{7}Li abundance. Additional studies are required to clarify if this discrepancy is real.

Even so, the overall concordance is remarkable: using well-established microphysics we have extrapolated back to an age of ~ 1 s to correctly predict light element abundances

spanning 9 orders of magnitude. This is a major success for the standard cosmology, and inspires confidence in extrapolation back to still earlier times.

This concordance provides a measure of the baryon content

$$0.017 \leq \Omega_{\rm B} h^2 \leq 0.024 \ (95\% \ {\rm CL}) \ , \tag{20.6}$$

a result that plays a key role in our understanding of the matter budget of the universe. First we note that $\Omega_{\rm B} \ll 1$, i.e., baryons cannot close the universe [47]. Furthermore, the cosmic density of (optically) luminous matter is $\Omega_{\rm lum} \simeq 0.0024 h^{-1}$ [48], so that $\Omega_{\rm B} \gg \Omega_{\rm lum}$: most baryons are optically dark, probably in the form of a $\sim 10^6$ K X-ray emitting intergalactic medium [49]. Finally, given that $\Omega_{\rm M} \sim 0.3$ (see Dark Matter and Cosmological Parameters reviews), we infer that most matter in the universe is not only dark but also takes some non-baryonic (more precisely, non-nucleonic) form.

The BBN prediction for the cosmic baryon density can be tested through precision observations of CMB temperature fluctuations (see Cosmic Microwave Background review). One can determine η from the amplitudes of the acoustic peaks in the CMB angular power spectrum [50], making it possible to compare two measures of η using very different physics, at two widely separated epochs. In the standard cosmology, there is no change in η between BBN and CMB decoupling, thus, a comparison of $\eta_{\rm BBN}$ and $\eta_{\rm CMB}$ is a key test. Agreement would endorse the standard picture while disagreement could point to new physics during/between the BBN and CMB epochs.

The release of the first-year WMAP results was a landmark event in this test of BBN. As with other cosmological parameter determinations from CMB data, the derived $\eta_{\rm CMB}$ depends on the adopted priors [51], in particular the form assumed for the power spectrum of primordial density fluctuations. If this is taken to be a scale-free power-law, the WMAP data implies $\Omega_{\rm B} h^2 = 0.024 \pm 0.001$ or $\eta_{10} = 6.58 \pm 0.27$, while allowing for a "running" spectral index lowers the value to $\Omega_{\rm B} h^2 = 0.0224 \pm 0.0009$ or $\eta_{10} = 6.14 \pm 0.25$ [52]; this latter range appears in Fig. 20.1. Other assumptions for the shape of the power spectrum can lead to baryon densities as low as $\Omega_{\rm B} h^2 = 0.019$ [53]. Thus outstanding uncertainties regarding priors are a source of systematic error which presently exceeds the statistical error in the prediction for η.

It is remarkable that the CMB estimate of the baryon density is consistent with the BBN range quoted in Eq. (20.6), and is in fact in very good agreement with the value inferred from recent high-redshift D/H measurements [26] and ^{4}He determinations; together these observations span diverse environments from redshifts $z = 1000$ to the present. However ^{7}Li is at best marginally consistent with the CMB (and with D) given the error budgets we have quoted. The question then becomes more pressing as to whether this mismatch comes from systematic errors in the observed abundances, and/or uncertainties in stellar astrophysics, or whether there might be new physics at work. Inhomogeneous nucleosynthesis can alter abundances for a given $\eta_{\rm BBN}$ but will overproduce ^{7}Li [54]. Note that entropy generation by some non-standard process could have decreased η between the BBN era and CMB decoupling, however the lack of spectral distortions in the CMB rules out any significant energy injection upto a redshift $z \sim 10^7$ [55]. Interestingly, the CMB itself offers the promise of measuring ^{4}He [56] and possibly ^{7}Li [57] directly at $z \sim 300 - 1000$.

Bearing in mind the importance of priors, the promise of precision determinations of the baryon density using the CMB motivates using this value as an input to BBN calculations. Within the context of the Standard Model, BBN then becomes a zero-parameter theory, and the light element abundances are completely determined to within the uncertainties in $\eta_{\rm CMB}$ and the BBN theoretical errors. Comparison with the observed abundances then can be used to test the astrophysics of post-BBN light element evolution [58]. Alternatively, one can consider possible physics beyond the Standard Model (e.g., which might change the expansion rate during BBN) and then use all of the abundances to test such models; this is the subject of our final section.

20.4. Beyond the Standard Model

Given the simple physics underlying BBN, it is remarkable that it still provides the most effective test for the cosmological viability of ideas concerning physics beyond the Standard Model. Although baryogenesis and inflation must have occurred at higher temperatures in the early universe, we do not as yet have 'standard models' for these so BBN still marks the boundary between the established and the speculative in big bang cosmology. It might appear possible to push the boundary back to the quark-hadron transition at $T \sim \Lambda_{\rm QCD}$ or electroweak symmetry breaking at $T \sim 1/\sqrt{G_{\rm F}}$; however so far no observable relics of these epochs have been identified, either theoretically or observationally. Thus although the Standard Model provides a precise description of physics up to the Fermi scale, cosmology cannot be traced in detail before the BBN era.

Limits on particle physics beyond the Standard Model come mainly from the observational bounds on the ^{4}He abundance. This is proportional to the n/p ratio which is determined when the weak-interaction rates fall behind the Hubble expansion rate at $T_{\rm fr} \sim 1$ MeV. The presence of additional neutrino flavors (or of any other relativistic species) at this time increases g_*, hence the expansion rate, leading to a larger value of $T_{\rm fr}$, n/p, and therefore $Y_{\rm p}$ [6,59]. In the Standard Model, the number of relativistic particle species at 1 MeV is $g_* = 5.5 + \frac{7}{4}N_\nu$, where 5.5 accounts for photons and $e^\pm$, and N_ν is the number of (nearly massless) neutrino flavors (see Big Bang Cosmology review). The helium curves in Fig. 20.1 were computed taking $N_\nu = 3$; the computed abundance scales as $\Delta Y_{\rm BBN} \simeq 0.013 \Delta N_\nu$ [7]. Clearly the central value for N_ν from BBN will depend on η, which is independently determined (with weaker sensitivity to N_ν) by the adopted D or ^{7}Li abundance. For example, if the best value for the observed primordial ^{4}He abundance is 0.249, then, for $\eta_{10} \sim 6$, the central value for N_ν is very close to 3. This limit depends sensitively on the adopted light element abundances, particularly $Y_{\rm BBN}$. A maximum likelihood analysis on η and N_ν based on the above ^{4}He and D abundances finds the (correlated) 95% CL ranges to be $4.9 < \eta_{10} < 7.1$ and $1.8 < N_\nu < 4.5$ [60]. On the other hand, using the 'low' ^{4}He and ^{7}Li gives [61] $\eta_{10} \leq 4.3$, and $1.4 \leq N_\nu \leq 4.9$. Similar results were obtained in another study [62] which presented a simpler method FASTBBN [10] to extract such bounds based on χ^2 statistics, given a set of input abundances. Using the CMB determination of η improves the constraints: with a 'low' ^{4}He, $N_\nu = 3$ is barely allowed at 2σ [63] but using the ^{4}He (and D) abundance quoted above gives $5.66 < \eta_{10} < 6.58$ ($\Omega_{\rm B}h^2 = 0.0226 \pm 0.0017$) and $N_\nu = 3.24 \pm 1.2$ at 95% CL [60].

Just as one can use the measured helium abundance to place limits on g_* [59], any changes in the strong, weak, electromagnetic, or gravitational coupling constants, arising e.g. from the dynamics of new dimensions, can be similarly constrained [64].

The limits on N_ν can be translated into limits on other types of particles or particle masses that would affect the expansion rate of the Universe during nucleosynthesis. For example consider 'sterile' neutrinos with only right-handed interactions of strength $G_{\rm R} < G_{\rm F}$. Such particles would decouple at higher temperature than (left-handed) neutrinos, so their number density ($\propto\ T^3$) relative to neutrinos would be reduced by any subsequent entropy release, e.g. due to annihilations of massive particles that become non-relativistic in between the two decoupling temperatures. Thus (relativistic) particles with less than full strength weak interactions contribute less to the energy density than particles that remain in equilibrium up to the time of nucleosynthesis [65]. If we impose $N_\nu < 4$ as an illustrative constraint, then the three right-handed neutrinos must have a temperature $3(T_{\nu_{\rm R}}/T_{\nu_{\rm L}})^4 < 1$. Since the temperature of the decoupled $\nu_{\rm R}$'s is determined by entropy conservation (see Big Bang Cosmology review), $T_{\nu_{\rm R}}/T_{\nu_{\rm L}} = [(43/4)/g_*(T_{\rm d})]^{1/3} < 0.76$, where $T_{\rm d}$ is the decoupling temperature of the $\nu_{\rm R}$'s. This requires $g_*(T_{\rm d}) > 24$ so decoupling must have occurred at $T_{\rm d} > 140$ MeV. The decoupling temperature is related to $G_{\rm R}$ through $(G_{\rm R}/G_{\rm F})^2 \sim (T_{\rm d}/3\,{\rm MeV})^{-3}$, where 3 MeV is the decoupling temperature for $\nu_{\rm L}$s. This yields a limit $G_{\rm R} \lesssim 10^{-2} G_{\rm F}$. The above argument sets lower limits on the masses of new Z' gauge bosons in superstring models [66] or in extended technicolor models [67] to which such right-handed neutrinos would be coupled. Similarly a Dirac magnetic moment for neutrinos, which would allow the right-handed states to be produced through scattering and thus increase g_*, can be significantly constrained [68], as can any new interactions for neutrinos which have a similar effect [69]. Right-handed states can be populated directly by helicity-flip scattering if the neutrino mass is large enough and this can be used to used to infer e.g. a bound of $m_{\nu_\tau} \lesssim 1$ MeV taking $N_\nu < 4$ [70]. If there is mixing between active and sterile neutrinos then the effect on BBN is more complicated [71].

The limit on the expansion rate during BBN can also be translated into bounds on the mass/lifetime of particles which are non-relativistic during BBN resulting in an even faster speed-up rate; the subsequent decays of such particles will typically also change the entropy, leading to further constraints [72]. Even more stringent constraints come from requiring that the synthesized light element abundances are not excessively altered through photodissociations by the electromagnetic cascades triggered by the decays [73–75], or by the effects of hadrons in the cascades [76]. Such arguments have been applied to e.g. rule out a MeV mass ν_τ which decays during nucleosynthesis [77]; even if the decays are to non-interacting particles (and light neutrinos), bounds can be derived from considering their effects on BBN [78].

Such arguments have proved very effective in constraining supersymmetry. For example if the gravitino is light and contributes to g_*, the illustrative BBN limit $N_\nu < 4$ requires its mass to exceed ~ 1 eV [79]. In models where supersymmetry breaking is gravity mediated, the gravitino mass is usually much higher, of order the electroweak scale; such gravitinos would be unstable and decay after BBN. The constraints on unstable particles discussed above imply stringent bounds on the allowed abundance of such particles,

which in turn impose powerful constraints on supersymmetric inflationary cosmology [74–76]. These can be evaded only if the gravitino is massive enough to decay before BBN, i.e. $m_{3/2} \gtrsim 50$ TeV [80] which would be unnatural, or if it is in fact the LSP and thus stable [74,81]. Similar constraints apply to moduli — very weakly coupled fields in string theory which obtain an electroweak-scale mass from supersymmetry breaking [82].

Finally we mention that BBN places powerful constraints on the recently suggested possibility that there are new large dimensions in nature, perhaps enabling the scale of quantum gravity to be as low as the electroweak scale [83]. Thus Standard Model fields may be localized on a 'brane' while gravity alone propagates in the 'bulk'. It has been further noted that the new dimensions may be non-compact, even infinite [84] and the cosmology of such models has attracted considerable attention. The expansion rate in the early universe can be significantly modified so BBN is able to set interesting constraints on such possibilities [85].

References:

1. R.V. Wagoner *et al.*, Astrophys. J. **148**, 3 (1967).
2. R.A. Malaney and G.J. Mathews, Phys. Reports **229**, 145 (1993).
3. S. Sarkar, Rept. on Prog. in Phys. **59**, 1493 (1996).
4. D.N. Schramm and M.S. Turner, Rev. Mod. Phys. **70**, 303 (1998).
5. K.A. Olive *et al.*, Phys. Reports **333**, 389 (2000).
6. P.J.E. Peebles, Phys. Rev. Lett. **16**, 411 (1966).
7. J. Bernstein *et al.*, Rev. Mod. Phys. **61**, 25 (1989).
8. S. Mukhanov, Int. J. Theor. Phys. **143**, 669 (2004).
9. R. Esmailzadeh *et al.*, Astrophys. J. **378**, 504 (1991).
10. `http://www-thphys.physics.ox.ac.uk/users/SubirSarkar/bbn.html`.
11. L. Kawano, FERMILAB-PUB-92/04-A.
12. S. Esposito *et al.*, Nucl. Phys. **B568**, 421 (2000).
13. S. Dodelson and M.S. Turner, Phys. Rev. **D46**, 3372 (1992).
14. D. Seckel, hep-ph/9305311;
    R. Lopez and M.S. Turner, Phys. Rev. **D59**, 103502 (1999).
15. M.S. Smith *et al.*, Astrophys. J. Supp. **85**, 219 (1993).
16. G. Fiorentini *et al.*, Phys. Rev. **D58**, 063506 (1998).
17. K.M. Nollett and S. Burles, Phys. Rev. **D61**, 123505 (2000).
18. R.H. Cyburt *et al.*, New Astron. **6**, 215 (2001).
19. A. Coc *et al.*, Astrophys. J. **600**, 544 (2004).
20. R.H. Cyburt, Phys. Rev. **D70**, 023505 (2004).
21. P.D. Serpico *et al.*, J. Cosmo. Astropart. Phys. **12**, 010 (2004).
22. K.M. Nollett *et al.*, Astrophys. J. Lett. **552**, L1 (2001).
23. J.M. O'Meara *et al.*, Astrophys. J. **552**, 718 (2001).
24. S. D'Odorico *et al.*, Astron. & Astrophys. **368**, L21 (2001).
25. M. Pettini and D. Bowen, Astrophys. J. **560**, 41 (2001).
26. D. Kirkman *et al.*, Astrophys. J. Supp. **149**, 1 (2003).
27. N.H.M. Crighton *et al.*, MNRAS **355**, 1042 (2004).

28. R.I. Epstein *et al.*, Nature **263**, 198 (1976).
29. J. Linsky, Space Sci. Rev., **106**, 49 (2003).
30. B.E. Wood *et al.*,Astrophys. J. **609**, 838 (2004).
31. B.D. Fields, Astrophys. J. **456**, 678 (1996).
32. D. Kirkman *et al.*, Astrophys. J. **529**, 655 (2000).
33. Y.I. Izotov *et al.*, Astrophys. J. **527**, 757 (1999).
34. K.A. Olive and E. Skillman, Astrophys. J. **617**, 29 (2004).
35. V. Luridiana *et al.*, Astrophys. J. **592**, 846 (2003).
36. N. Christlieb *et al.*, Nature **419**, 904 (2002).
37. M. Spite and F. Spite, Nature **297**, 483 (1982).
38. S.G. Ryan *et al.*, Astrophys. J. **523**, 654 (1999).
39. E. Vangioni-Flam *et al.*, New Astron. **4**, 245 (1999).
40. S.G. Ryan *et al.*, Astrophys. J. Lett. **530**, L57 (2000).
41. P. Bonifacio *et al.*, Astron. & Astrophys. **390**, 91 (2002).
42. J. Meléndez and I. Ramirez, Astrophys. J. Lett. **615**, 33 (2004).
43. M.H. Pinsonneault *et al.*, Astrophys. J. **574**, 389 (2002).
44. D.S. Balser *et al.*, Astrophys. J. **510**, 759 (1999).
45. K.A. Olive *et al.*, Astrophys. J. **479**, 752 (1997).
46. E. Vangioni-Flam *et al.*, Astrophys. J. **585**, 611 (2003).
47. H. Reeves *et al.*, Astrophys. J. **179**, 909 (1973).
48. M. Fukugita *et al.*, Astrophys. J. **503**, 518 (1998).
49. R. Cen and J.P. Ostriker, Astrophys. J. **514**, 1 (1999).
50. G. Jungman *et al.*, Phys. Rev. **D54**, 1332 (1996).
51. M. Tegmark *et al.*, Phys. Rev. **D63**, 043007 (2001).
52. D.N. Spergel *et al.*, Astrophys. J. Supp. **148**, 175 (2003).
53. A. Blanchard *et al.*, Astron. & Astrophys. **412**, 35 (2003).
54. K. Jedamzik and J.B. Rehm, Phys. Rev. **D64**, 023510 (2001).
55. D.J. Fixsen *et al.*, Astrophys. J. **473**, 576 (1996).
56. R. Trotta and S.H. Hansen, Phys. Rev. **D69**, 023509 (2004).
57. P.C. Stancil *et al.*, Astrophys. J. **580**, 29 (2002).
58. R.H. Cyburt *et al.*, Phys. Lett. **B567**, 227 (2003).
59. G. Steigman *et al.*, Phys. Lett. **B66**, 202 (1977).
60. R. H. Cyburt *et al.*, Astropart. Phys. **23**, 313 (2005).
61. K.A. Olive and D. Thomas, Astropart. Phys. **11**, 403 (1999).
62. E. Lisi *et al.*, Phys. Rev. **D59**, 123520 (1999).
63. V. Barger *et al.*, Phys. Lett. **B566**, 8 (2003).
64. E.W. Kolb *et al.*, Phys. Rev. **D33**, 869 (1986);
F.S. Accetta *et al.*, Phys. Lett. **B248**, 146 (1990);
B.A. Campbell and K.A. Olive, Phys. Lett. **B345**, 429 (1995);
K.M. Nollett and R. Lopez, Phys. Rev. **D66**, 063507 (2002);
C. Bambi *et al.*, Phys. Rev. **D71**, 123524 (2005).
65. K.A. Olive *et al.*, Nucl. Phys. **B180**, 497 (1981).
66. J. Ellis *et al.*, Phys. Lett. **B167**, 457 (1986).
67. L.M. Krauss *et al.*, Phys. Rev. Lett. **71**, 823 (1993).

68. J.A. Morgan, Phys. Lett. **B102**, 247 (1981).
69. E.W. Kolb *et al.*, Phys. Rev. **D34**, 2197 (1986);
    J.A. Grifols and E. Massó, Mod. Phys. Lett. **A2**, 205 (1987);
    K.S. Babu *et al.*, Phys. Rev. Lett. **67**, 545 (1991).
70. A.D. Dolgov *et al.*, Nucl. Phys. **B524**, 621 (1998).
71. K. Enqvist *et al.*, Nucl. Phys. **B373**, 498 (1992);
    A.D. Dolgov, Phys. Reports **370**, 333 (2002).
72. K. Sato and M. Kobayashi, Prog. Theor. Phys. **58**, 1775 (1977);
    D.A. Dicus *et al.*, Phys. Rev. **D17**, 1529 (1978);
    R.J. Scherrer and M.S. Turner, Astrophys. J. **331**, 19 (1988).
73. D. Lindley, MNRAS **188**, 15 (1979), Astrophys. J. **294**, 1 (1985).
74. J. Ellis *et al.*, Nucl. Phys. **B259**, 175 (1985).
75. J. Ellis *et al.*, Nucl. Phys. **B373**, 399 (1992);
    R.H. Cyburt *et al.*, Phys. Rev. **D67**, 103521 (2003).
76. M.H. Reno and D. Seckel, Phys. Rev. **D37**, 3441 (1988);
    S. Dimopoulos *et al.*, Nucl. Phys. **B311**, 699 (1989);
    K. Kohri *et al.*, Phys. Rev. **D71**, 083502 (2005).
77. S. Sarkar and A.M. Cooper, Phys. Lett. **B148**, 347 (1984).
78. S. Dodelson *et al.*, Phys. Rev. **D49**, 5068 (1994).
79. J.A. Grifols *et al.*, Phys. Lett. **B400**, 124 (1997).
80. S. Weinberg, Phys. Rev. Lett. **48**, 1303 (1979).
81. M. Bolz *et al.*, Nucl. Phys. **B606**, 518 (2001).
82. G. Coughlan *et al.*, Phys. Lett. **B131**, 59 (1983).
83. N. Arkani-Hamed *et al.*, Phys. Rev. **D59**, 086004 (1999).
84. L. Randall and R. Sundrum, Phys. Rev. Lett. **83**, 3370 (1999).
85. J.M. Cline *et al.*, Phys. Rev. Lett. **83**, 4245 (1999);
    P. Binetruy *et al.*, Phys. Lett. **B477**, 285 (2000).

21. THE COSMOLOGICAL PARAMETERS

Written August 2003, updated May 2006, by O. Lahav (University College London) and A.R. Liddle (University of Sussex).

21.1. Parametrizing the Universe

Rapid advances in observational cosmology are leading to the establishment of the first precision cosmological model, with many of the key cosmological parameters determined to one or two significant figure accuracy. Particularly prominent are measurements of cosmic microwave anisotropies, led by the three-year results from the Wilkinson Microwave Anisotropy Probe (WMAP) [1,2]. However the most accurate model of the Universe requires consideration of a wide range of different types of observation, with complementary probes providing consistency checks, lifting parameter degeneracies, and enabling the strongest constraints to be placed.

The term 'cosmological parameters' is forever increasing in its scope, and nowadays includes the parametrization of some functions, as well as simple numbers describing properties of the Universe. The original usage referred to the parameters describing the global dynamics of the Universe, such as its expansion rate and curvature. Also now of great interest is how the matter budget of the Universe is built up from its constituents: baryons, photons, neutrinos, dark matter, and dark energy. We need to describe the nature of perturbations in the Universe, through global statistical descriptions such as the matter and radiation power spectra. There may also be parameters describing the physical state of the Universe, such as the ionization fraction as a function of time during the era since decoupling. Typical comparisons of cosmological models with observational data now feature between five and ten parameters.

21.1.1. *The global description of the Universe* :

Ordinarily, the Universe is taken to be a perturbed Robertson–Walker space-time with dynamics governed by Einstein's equations. This is described in detail by Olive and Peacock in this volume. Using the density parameters Ω_i for the various matter species and Ω_Λ for the cosmological constant, the Friedmann equation can be written

$$\sum_i \Omega_i + \Omega_\Lambda = \frac{k}{R^2 H^2}\,, \tag{21.1}$$

where the sum is over all the different species of matter in the Universe. This equation applies at any epoch, but later in this article we will use the symbols Ω_i and Ω_Λ to refer to the present values. A typical collection would be baryons, photons, neutrinos, and dark matter (given charge neutrality, the electron density is guaranteed to be too small to be worth considering separately).

The complete present state of the homogeneous Universe can be described by giving the present values of all the density parameters and the present Hubble parameter h. These also allow us to track the history of the Universe back in time, at least until an epoch where interactions allow interchanges between the densities of the different species, which is believed to have last happened at neutrino decoupling shortly before nucleosynthesis. To probe further back into the Universe's history requires assumptions about particle interactions, and perhaps about the nature of physical laws themselves.

21.1.2. ***Neutrinos*** :

The standard neutrino sector has three flavors. For neutrinos of mass in the range $5 \times 10^{-4}\,\mathrm{eV}$ to $1\,\mathrm{MeV}$, the density parameter in neutrinos is predicted to be

$$\Omega_\nu h^2 = \frac{\sum m_\nu}{93\,\mathrm{eV}}\,, \tag{21.2}$$

where the sum is over all families with mass in that range (higher masses need a more sophisticated calculation). We use units with $c = 1$ throughout. Results on atmospheric and solar neutrino oscillations [3] imply non-zero mass-squared differences between the three neutrino flavors. These oscillation experiments cannot tell us the absolute neutrino masses, but within the simple assumption of a mass hierarchy suggest a lower limit of $\Omega_\nu \approx 0.001$ on the neutrino mass density parameter.

For a total mass as small as $0.1\,\mathrm{eV}$, this could have a potentially observable effect on the formation of structure, as neutrino free-streaming damps the growth of perturbations. Present cosmological observations have shown no convincing evidence of any effects from either neutrino masses or an otherwise non-standard neutrino sector, and impose quite stringent limits, which we summarize in Section 21.3.4. Consequently, the standard assumption at present is that the masses are too small to have a significant cosmological impact, but this may change in the near future.

The cosmological effect of neutrinos can also be modified if the neutrinos have decay channels, or if there is a large asymmetry in the lepton sector manifested as a different number density of neutrinos versus anti-neutrinos. This latter effect would need to be of order unity to be significant, rather than the 10^{-9} seen in the baryon sector, which may be in conflict with nucleosynthesis [4].

21.1.3. ***Inflation and perturbations*** :

A complete model of the Universe should include a description of deviations from homogeneity, at least in a statistical way. Indeed, some of the most powerful probes of the parameters described above come from the evolution of perturbations, so their study is naturally intertwined in the determination of cosmological parameters.

There are many different notations used to describe the perturbations, both in terms of the quantity used to describe the perturbations and the definition of the statistical measure. We use the dimensionless power spectrum Δ^2 as defined in Olive and Peacock (also denoted $\mathcal{P}$ in some of the literature). If the perturbations obey Gaussian statistics, the power spectrum provides a complete description of their properties.

From a theoretical perspective, a useful quantity to describe the perturbations is the curvature perturbation $\mathcal{R}$, which measures the spatial curvature of a comoving slicing of the space-time. A case of particular interest is the Harrison–Zel'dovich spectrum, which corresponds to a constant spectrum $\Delta^2_{\mathcal{R}}$. More generally, one can approximate the spectrum by a power-law, writing

$$\Delta^2_{\mathcal{R}}(k) = \Delta^2_{\mathcal{R}}(k_*) \left[\frac{k}{k_*}\right]^{n-1}\,, \tag{21.3}$$

where n is known as the spectral index, always defined so that $n = 1$ for the Harrison–Zel'dovich spectrum, and k_* is an arbitrarily chosen scale. The initial spectrum, defined at some early epoch of the Universe's history, is usually taken to have a simple form such as this power-law, and we will see that observations require n close to one, which corresponds to the perturbations in the curvature being independent of scale. Subsequent evolution will modify the spectrum from its initial form.

The simplest viable mechanism for generating the observed perturbations is the inflationary cosmology, which posits a period of accelerated expansion in the Universe's early stages [5]. It is a useful working hypothesis that this is the sole mechanism for generating perturbations. Commonly, it is further assumed to be the simplest class of inflationary model, where the dynamics are equivalent to that of a single scalar field ϕ slowly rolling on a potential $V(\phi)$. One aim of cosmology is to verify that this simple picture can match observations, and to determine the properties of $V(\phi)$ from the observational data.

Inflation generates perturbations through the amplification of quantum fluctuations, which are stretched to astrophysical scales by the rapid expansion. The simplest models generate two types, density perturbations which come from fluctuations in the scalar field and its corresponding scalar metric perturbation, and gravitational waves which are tensor metric fluctuations. The former experience gravitational instability and lead to structure formation, while the latter can influence the cosmic microwave background anisotropies. Defining slow-roll parameters, with primes indicating derivatives with respect to the scalar field, as

$$\epsilon = \frac{m_{\rm Pl}^2}{16\pi}\left(\frac{V'}{V}\right)^2 \quad ; \quad \eta = \frac{m_{\rm Pl}^2}{8\pi}\frac{V''}{V}\,, \tag{21.4}$$

which should satisfy $\epsilon, |\eta| \ll 1$, the spectra can be computed using the slow-roll approximation as

$$\begin{aligned} \Delta^2_{\mathcal{R}}(k) &\simeq \frac{8}{3m_{\rm Pl}^4}\frac{V}{\epsilon}\bigg|_{k=aH}\,; \\ \Delta^2_{\rm grav}(k) &\simeq \frac{128}{3m_{\rm Pl}^4}V\bigg|_{k=aH}\,. \end{aligned} \tag{21.5}$$

In each case, the expressions on the right-hand side are to be evaluated when the scale k is equal to the Hubble radius during inflation. The symbol '$\simeq$' indicates use of the slow-roll approximation, which is expected to be accurate to a few percent or better.

From these expressions, we can compute the spectral indices

$$n \simeq 1 - 6\epsilon + 2\eta \quad ; \quad n_{\rm grav} \simeq -2\epsilon\,. \tag{21.6}$$

Another useful quantity is the ratio of the two spectra, defined by

$$r \equiv \frac{\Delta^2_{\rm grav}(k_*)}{\Delta^2_{\mathcal{R}}(k_*)}\,. \tag{21.7}$$

The literature contains a number of definitions of r; this convention matches that of recent versions of CMBFAST [6] and that used by WMAP [7], while definitions based on the relative effect on the microwave background anisotropies typically differ by tens of percent. We have

$$r \simeq 16\epsilon \simeq -8n_{\rm grav}\,, \tag{21.8}$$

which is known as the consistency equation.

In general one could consider corrections to the power-law approximation, which we discuss later. However for now we make the working assumption that the spectra can be approximated by power laws. The consistency equation shows that r and $n_{\rm grav}$ are not independent parameters, and so the simplest inflation models give initial conditions described by three parameters, usually taken as $\Delta^2_{\mathcal{R}}$, n, and r, all to be evaluated at some scale k_*, usually the 'statistical centre' of the range explored by the data. Alternatively, one could use the parametrization V, ϵ, and η, all evaluated at a point on the putative inflationary potential.

After the perturbations are created in the early Universe, they undergo a complex evolution up until the time they are observed in the present Universe. While the perturbations are small, this can be accurately followed using a linear theory numerical code such as CMBFAST [6]. This works right up to the present for the cosmic microwave background, but for density perturbations on small scales non-linear evolution is important and can be addressed by a variety of semi-analytical and numerical techniques. However the analysis is made, the outcome of the evolution is in principle determined by the cosmological model, and by the parameters describing the initial perturbations, and hence can be used to determine them.

Of particular interest are cosmic microwave background anisotropies. Both the total intensity and two independent polarization modes are predicted to have anisotropies. These can be described by the radiation angular power spectra C_ℓ as defined in the article of Scott and Smoot in this volume, and again provide a complete description if the density perturbations are Gaussian.

21.1.4. *The standard cosmological model* :

We now have most of the ingredients in place to describe the cosmological model. Beyond those of the previous subsections, there are two parameters which are essential — a measure of the ionization state of the Universe and the galaxy bias parameter. The Universe is known to be highly ionized at low redshifts (otherwise radiation from distant quasars would be heavily absorbed in the ultra-violet), and the ionized electrons can scatter microwave photons altering the pattern of observed anisotropies. The most convenient parameter to describe this is the optical depth to scattering τ (i.e. the probability that a given photon scatters once); in the approximation of instantaneous and complete re-ionization, this could equivalently be described by the redshift of re-ionization $z_{\rm ion}$. The bias parameter, described fully later, is needed to relate the observed galaxy power spectrum to the predicted dark matter power spectrum.

The basic set of cosmological parameters is therefore as shown in Table 21.1. The spatial curvature does not appear in the list, because it can be determined from the other parameters using Eq. (21.1). The total present matter density $\Omega_{\rm m} = \Omega_{\rm dm} + \Omega_{\rm b}$ is usually used in place of the dark matter density.

Table 21.1: The basic set of cosmological parameters. We give values as obtained using a fit of a ΛCDM cosmology with a power-law initial spectrum to WMAP3 data alone [2]. Tensors are assumed zero except in quoting a limit on them. We cannot stress too much that the exact values and uncertainties depend on both the precise datasets used and the choice of parameters allowed to vary, and the effects of varying some assumptions will be shown later in Table 21.2. Limits on the cosmological constant depend on whether the Universe is assumed flat, while there is no established convention for specifying the density perturbation amplitude. Uncertainties are one-sigma/68% confidence unless otherwise stated.

Parameter	Symbol	Value
Hubble parameter	h	$0.73^{+0.03}_{-0.04}$
Total matter density	$\Omega_{\rm m}$	$\Omega_{\rm m}h^2 = 0.127^{+0.007}_{-0.009}$
Baryon density	$\Omega_{\rm b}$	$\Omega_{\rm b}h^2 = 0.0223^{+0.0007}_{-0.0009}$
Cosmological constant	Ω_Λ	See Ref. 2
Radiation density	$\Omega_{\rm r}$	$\Omega_{\rm r}h^2 = 2.47 \times 10^{-5}$
Neutrino density	Ω_ν	See Sec. 21.1.2
Density perturbation amplitude	$\Delta^2_{\mathcal{R}}(k_*)$	See Ref. 2
Density perturbation spectral index	n	$n = 0.951^{+0.015}_{-0.019}$
Tensor to scalar ratio	r	$r < 0.55$ (95% conf)
Ionization optical depth	τ	$\tau = 0.088^{+0.028}_{-0.034}$
Bias parameter	b	See Sec. 21.3.4

Most attention to date has been on parameter estimation, where a set of parameters is chosen by hand and the aim is to constrain them. Interest has been growing towards the higher-level inference problem of model selection, which compares different choices of parameter sets. Bayesian inference offers an attractive framework for cosmological model selection, setting a tension between model complexity and ability to fit the data.

As described in Sec. 21.4, models based on these eleven parameters are able to give a good fit to the complete set of high-quality data available at present, and indeed some simplification is possible. Observations are consistent with spatial flatness, and indeed the inflation models so far described automatically generate spatial flatness, so we can set $k = 0$; the density parameters then must sum to one, and so one can be eliminated. The neutrino energy density is often not taken as an independent parameter.

Provided the neutrino sector has the standard interactions the neutrino energy density while relativistic can be related to the photon density using thermal physics arguments, and it is currently difficult to see the effect of the neutrino mass although observations of large-scale structure have already placed interesting upper limits. This reduces the standard parameter set to nine. In addition, there is no observational evidence for the existence of tensor perturbations (though the upper limits are quite weak), and so r could be set to zero. Presently n is in a somewhat controversial position regarding whether it needs to be varied in a fit, or can be set to the Harrison–Zel'dovich value $n = 1$. Parameter estimation [2] suggests $n = 1$ is ruled out at reasonable significance, but Bayesian model selection techniques [8] suggest the data is not conclusive. With n set to one, this leaves seven parameters, which is the smallest set that can usefully be compared to the present cosmological data set. This model (usually with n kept as a parameter) is referred to by various names, including ΛCDM, the concordance cosmology, and the standard cosmological model.

Of these parameters, only $\Omega_{\rm r}$ is accurately measured directly. The radiation density is dominated by the energy in the cosmic microwave background, and the COBE FIRAS experiment has determined its temperature to be $T = 2.725 \pm 0.001$ Kelvin [9], corresponding to $\Omega_{\rm r} = 2.47 \times 10^{-5} h^{-2}$. It typically does not need to be varied in fitting other data. If galaxy clustering data is not included in a fit, then the bias parameter is also unnecessary.

In addition to this minimal set, there is a range of other parameters which might prove important in future as the dataset further improves, but for which there is so far no direct evidence, allowing them to be set to a specific value. We discuss various speculative options in the next section. For completeness at this point, we mention one other interesting parameter, the helium fraction, which is a non-zero parameter that can affect the microwave anisotropies at a subtle level. Presently, big-bang nucleosynthesis provides the best measurement of this parameter, and it is usually fixed in microwave anisotropy studies, but the data are just reaching a level where allowing its variation may become mandatory.

21.1.5. *Derived parameters* :

The parameter list of the previous subsection is sufficient to give a complete description of cosmological models which agree with observational data. However, it is not a unique parametrization, and one could instead use parameters derived from that basic set. Parameters which can be obtained from the set given above include the age of the Universe, the present horizon distance, the present microwave background and neutrino background temperatures, the epoch of matter–radiation equality, the epochs of recombination and decoupling, the epoch of transition to an accelerating Universe, the baryon-to-photon ratio, and the baryon to dark matter density ratio. The physical densities of the matter components, $\Omega_i h^2$, are often more useful than the density parameters. The density perturbation amplitude can be specified in many different ways other than the large-scale primordial amplitude, for instance, in terms of its effect on the cosmic microwave background, or by specifying a short-scale quantity, a common choice being the present linear-theory mass dispersion on a scale of $8\,h^{-1}$Mpc, known as σ_8.

Different types of observation are sensitive to different subsets of the full cosmological parameter set, and some are more naturally interpreted in terms of some of the derived parameters of this subsection than on the original base parameter set. In particular, most types of observation feature degeneracies whereby they are unable to separate the effects of simultaneously varying several of the base parameters.

21.2. Extensions to the standard model

This section discusses some ways in which the standard model could be extended. At present, there is no positive evidence in favor of any of these possibilities, which are becoming increasingly constrained by the data, though there always remains the possibility of trace effects at a level below present observational capability.

21.2.1. *More general perturbations* :

The standard cosmology assumes adiabatic, Gaussian perturbations. Adiabaticity means that all types of material in the Universe share a common perturbation, so that if the space-time is foliated by constant-density hypersurfaces, then all fluids and fields are homogeneous on those slices, with the perturbations completely described by the variation of the spatial curvature of the slices. Gaussianity means that the initial perturbations obey Gaussian statistics, with the amplitudes of waves of different wavenumbers being randomly drawn from a Gaussian distribution of width given by the power spectrum. Note that gravitational instability generates non-Gaussianity; in this context, Gaussianity refers to a property of the initial perturbations before they evolve significantly.

The simplest inflation models, based on one dynamical field, predict adiabatic fluctuations and a level of non-Gaussianity which is too small to be detected by any experiment so far conceived. For present data, the primordial spectra are usually assumed to be power laws.

21.2.1.1. *Non-power-law spectra:*

For typical inflation models, it is an approximation to take the spectra as power laws, albeit usually a good one. As data quality improves, one might expect this approximation to come under pressure, requiring a more accurate description of the initial spectra, particularly for the density perturbations. In general, one can write a Taylor expansion of $\ln \Delta^2_{\mathcal{R}}$ as

$$\ln \Delta^2_{\mathcal{R}}(k) = \ln \Delta^2_{\mathcal{R}}(k_*) + (n_* - 1) \ln \frac{k}{k_*} + \frac{1}{2} \left. \frac{dn}{d \ln k} \right|_* \ln^2 \frac{k}{k_*} + \cdots , \qquad (21.9)$$

where the coefficients are all evaluated at some scale k_*. The term $dn/d\ln k|_*$ is often called the running of the spectral index [10]. Once non-power-law spectra are allowed, it is necessary to specify the scale k_* at which the spectral index is defined.

21.2.1.2. *Isocurvature perturbations:*

An isocurvature perturbation is one which leaves the total density unperturbed, while perturbing the relative amounts of different materials. If the Universe contains N fluids, there is one growing adiabatic mode and $N-1$ growing isocurvature modes. These can be excited, for example, in inflationary models where there are two or more fields which acquire dynamically-important perturbations. If one field decays to form normal matter, while the second survives to become the dark matter, this will generate a cold dark matter isocurvature perturbation.

In general there are also correlations between the different modes, and so the full set of perturbations is described by a matrix giving the spectra and their correlations. Constraining such a general construct is challenging, though constraints on individual modes are beginning to become meaningful, with no evidence that any other than the adiabatic mode must be non-zero.

21.2.1.3. *Non-Gaussianity:*

Multi-field inflation models can also generate primordial non-Gaussianity. The extra fields can either be in the same sector of the underlying theory as the inflation, or completely separate, an interesting example of the latter being the curvaton model [11]. Current upper limits on non-Gaussianity are becoming stringent, but there remains much scope to push down those limits and perhaps reveal trace non-Gaussianity in the data. If non-Gaussianity is observed, its nature may favor an inflationary origin, or a different one such as topological defects. A plausible possibility is non-Gaussianity caused by defects forming in a phase transition which ended inflation.

21.2.2. ***Dark matter properties*** :

Dark matter properties are discussed in the article by Drees and Gerbier in this volume. The simplest assumption concerning the dark matter is that it has no significant interactions with other matter, and that its particles have a negligible velocity. Such dark matter is described as 'cold,' and candidates include the lightest supersymmetric particle, the axion, and primordial black holes. As far as astrophysicists are concerned, a complete specification of the relevant cold dark matter properties is given by the density parameter $\Omega_{\rm cdm}$, though those seeking to directly detect it are as interested in its interaction properties.

Cold dark matter is the standard assumption and gives an excellent fit to observations, except possibly on the shortest scales where there remains some controversy concerning the structure of dwarf galaxies and possible substructure in galaxy halos. For all the dark matter to have a large velocity dispersion, so-called hot dark matter, has long been excluded as it does not permit galaxies to form; for thermal relics the mass must be above about 1 keV to satisfy this constraint, though relics produced non-thermally, such as the axion, need not obey this limit. However, there remains the possibility that further parameters might need to be introduced to describe dark matter properties relevant to astrophysical observations. Suggestions which have been made include a modest velocity dispersion (warm dark matter) and dark matter self-interactions. There remains the possibility that the dark matter comprises two separate components, *e.g.*, a cold one and a hot one, an example being if massive neutrinos have a non-negligible effect.

21.2.3. ***Dark energy*** :

While the standard cosmological model given above features a cosmological constant, in order to explain observations indicating that the Universe is presently accelerating, further possibilities exist under the general heading dark energy.† A particularly attractive possibility (usually called quintessence, though that word is used with various different meanings in the literature) is that a scalar field is responsible, with the mechanism mimicking that of early Universe inflation [12]. As described by Olive and Peacock, a fairly model-independent description of dark energy can be given just using the equation of state parameter w, with $w = -1$ corresponding to a cosmological constant. In general, the function w could itself vary with redshift, though practical experiments devised so far would be sensitive primarily to some average value weighted over recent epochs. For high-precision predictions of microwave background anisotropies, it is better to use a scalar-field description in order to have a self-consistent evolution of the 'sound speed' associated with the dark energy perturbations.

Present observations are consistent with a cosmological constant, but it is quite common to see w kept as a free parameter to be added to the set described in the previous section. Most, but not all, researchers assume the weak energy condition $w \geq -1$. In the future it may be necessary to use a more sophisticated parametrization of the dark energy.

21.2.4. ***Complex ionization history*** :

The full ionization history of the Universe is given by specifying the ionization fraction as a function of redshift z. The simplest scenario takes the ionization to be zero from recombination up to some redshift $z_{\rm ion}$, at which point the Universe instantaneously re-ionizes completely. In that case, there is a one-to-one correspondence between τ and $z_{\rm ion}$ (that relation, however, also depending on other cosmological parameters).

While simple models of the re-ionization process suggest that rapid ionization is a good approximation, observational evidence is mixed, with indications of a high optical depth inferred from the microwave background difficult to reconcile with absorption seen in some high-redshift quasar systems, and also perhaps with the temperature of the intergalactic medium at $z \simeq 3$. Accordingly, a more complex ionization history may need to be considered, and perhaps separate histories for hydrogen and helium, which will necessitate new parameters. Additionally, high-precision microwave anisotropy experiments may require consideration of the level of residual ionization left after recombination, which in principle is computable from the other cosmological parameters.

† Unfortunately this is rather a misnomer, as it is the negative pressure of this material, rather than its energy, that is responsible for giving the acceleration. Furthermore, while generally in physics matter and energy are interchangeable terms, dark matter and dark energy are quite distinct concepts.

21.2.5. ***Varying 'constants'*** :

Variation of the fundamental constants of nature over cosmological times is another possible enhancement of the standard cosmology. There is a long history of study of variation of the gravitational constant G, and more recently attention has been drawn to the possibility of small fractional variations in the fine-structure constant. There is presently no observational evidence for the former, which is tightly constrained by a variety of measurements. Evidence for the latter has been claimed from studies of spectral line shifts in quasar spectra at redshifts of order two [13], but this is presently controversial and in need of further observational study.

More broadly, one can ask whether general relativity is valid at all epochs under consideration.

21.2.6. ***Cosmic topology*** :

The usual hypothesis is that the Universe has the simplest topology consistent with its geometry, for example that a flat Universe extends forever. Observations cannot tell us whether that is true, but they can test the possibility of a non-trivial topology on scales up to roughly the present Hubble scale. Extra parameters would be needed to specify both the type and scale of the topology, for example, a cuboidal topology would need specification of the three principal axis lengths. At present, there is no direct evidence for cosmic topology, though the low values of the observed cosmic microwave quadrupole and octupole have been cited as a possible signature.

21.3. Probes

The goal of the observational cosmologist is to utilize astronomical objects to derive cosmological parameters. The transformation from the observables to the key parameters usually involves many assumptions about the nature of the objects, as well as about the nature of the dark matter. Below we outline the physical processes involved in each probe, and the main recent results. The first two subsections concern probes of the homogeneous Universe, while the remainder consider constraints from perturbations.

We note three types of uncertainties that enter into any errors on the cosmological parameters of interest: (i) due to the assumptions on the cosmological model and its priors (i.e. the number of assumed cosmological parameters and their allowed range); (ii) due to the uncertainty in the astrophysics of the objects (e.g. the mass–temperature relation of galaxy clusters); and (iii) due to instrumental and observational limitations (e.g. the effect of 'seeing' on weak gravitational lensing measurements).

21.3.1. ***Direct measures of the Hubble constant*** :

In 1929 Edwin Hubble discovered the law of expansion of the Universe by measuring distances to nearby galaxies. The slope of the relation between the distance and recession velocity is defined to be the Hubble constant H_0. Astronomers argued for decades on the systematic uncertainties in various methods and derived values over the wide range, $40\,\mathrm{km\,s^{-1}\,Mpc^{-1}} \lesssim H_0 \lesssim 100\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$.

One of the most reliable results on the Hubble constant comes from the Hubble Space Telescope Key Project [14]. The group has used the empirical period–luminosity

relations for Cepheid variable stars to obtain distances to 31 galaxies, and calibrated a number of secondary distance indicators (Type Ia Supernovae, Tully-Fisher, surface brightness fluctuations and Type II Supernovae) measured over distances of 400 to 600 Mpc. They estimated $H_0 = 72 \pm 3$ (statistical) ± 7 (systematic) $\mathrm{km\,s^{-1}\,Mpc^{-1}}$.[‡] The major sources of uncertainty in this result are due to the metallicity of the Cepheids and the distance to the fiducial nearby galaxy (called the Large Magellanic Cloud) relative to which all Cepheid distances are measured. Nevertheless, it is remarkable that this result is in such good agreement with the result derived from the WMAP CMB measurements (see Table 21.2).

21.3.2. *Supernovae as cosmological probes* :

The relation between observed flux and the intrinsic luminosity of an object depends on the luminosity distance d_L, which in turn depends on cosmological parameters. More specifically

$$d_L = (1+z) r_e(z)\,, \tag{21.10}$$

where $r_e(z)$ is the coordinate distance. For example, in a flat Universe

$$r_e(z) = \int_0^z dz'/H(z')\,. \tag{21.11}$$

For a general dark energy equation of state $w(z) = p_Q(z)/\rho_Q(z)$, the Hubble parameter is, still considering only the flat case,

$$H^2(z)/H_0^2 = (1+z)^3\Omega_{\mathrm{m}} + \Omega_Q \exp[3X(z)]\,, \tag{21.12}$$

where

$$X(z) = \int_0^z [1+w(z')](1+z')^{-1}dz'\,, \tag{21.13}$$

and Ω_{m} and Ω_Q are the present density parameters of matter and dark energy components. If a general equation of state is allowed, then one has to solve for $w(z)$ (parametrized, for example, as $w(z) = w = \text{const.}$, or $w(z) = w_0 + w_1 z$) as well as for Ω_Q.

Empirically, the peak luminosity of supernova of Type Ia (SNe Ia) can be used as an efficient distance indicator (*e.g.*, Ref. 15). The favorite theoretical explanation for SNe Ia is the thermonuclear disruption of carbon-oxygen white dwarfs. Although not perfect 'standard candles', it has been demonstrated that by correcting for a relation between the light curve shape and the luminosity at maximum brightness, the dispersion of the measured luminosities can be greatly reduced. There are several possible systematic effects which may affect the accuracy of the SNe Ia as distance indicators, for example, evolution with redshift and interstellar extinction in the host galaxy and in the Milky Way, but there is no indication that any of these effects are significant for the cosmological constraints.

[‡] Unless stated otherwise, all quoted uncertainties in this article are one-sigma/68% confidence. It is common for cosmological parameters to have significantly non-Gaussian error distributions.

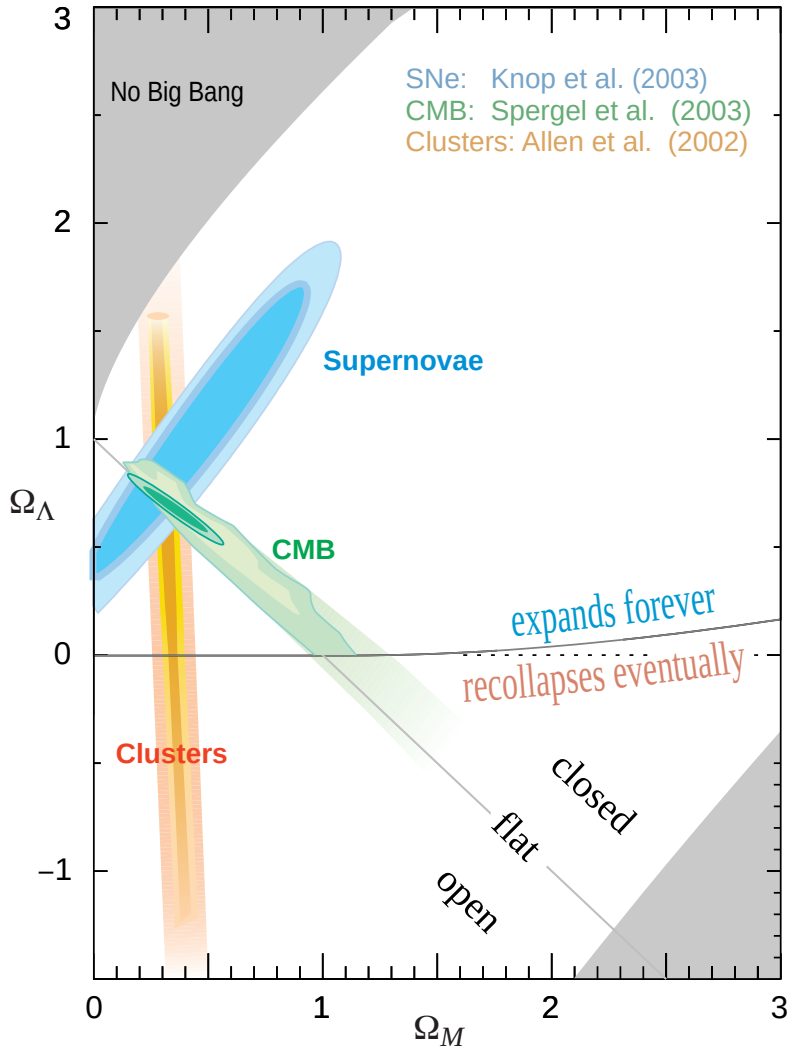


Figure 21.1: This shows the preferred region in the $\Omega_{\rm m}$–Ω_Λ plane from the compilation of supernovae data in Ref. 17, and also the complementary results coming from some other observations. See full-color version on color pages at end of book. [Courtesy of the Supernova Cosmology Project.]

Two major studies, the 'Supernova Cosmology Project' and the 'High-z Supernova Search Team', found evidence for an accelerating Universe [16], interpreted as due to a cosmological constant, or to a more general 'dark energy' component. Current results from the Supernova Cosmology Project [17] are shown in Fig. 21.1 (see also Ref. 18). The SNe Ia data alone can only constrain a combination of $\Omega_{\rm m}$ and Ω_Λ. When combined with the CMB data (which indicates flatness, *i.e.*, $\Omega_{\rm m} + \Omega_\Lambda \approx 1$), the best-fit values are $\Omega_{\rm m} \approx 0.3$ and $\Omega_\Lambda \approx 0.7$. Future experiments will aim to set constraints on the cosmic equation of state $w(z)$. However, given the integral relation between the luminosity distance and $w(z)$, it is not straightforward to recover $w(z)$ (*e.g.*, Ref. 19).

21.3.3. ***Cosmic microwave background*** :

The physics of the cosmic microwave background (CMB) is described in detail by Scott and Smoot in this volume. Before recombination, the baryons and photons are tightly coupled, and the perturbations oscillate in the potential wells generated primarily by the dark matter perturbations. After decoupling, the baryons are free to collapse into those potential wells. The CMB carries a record of conditions at the time of decoupling, often called primary anisotropies. In addition, it is affected by various processes as it propagates towards us, including the effect of a time-varying gravitational potential (the integrated Sachs-Wolfe effect), gravitational lensing, and scattering from ionized gas at low redshift.

The primary anisotropies, the integrated Sachs-Wolfe effect, and scattering from a homogeneous distribution of ionized gas, can all be calculated using linear perturbation theory, a widely-used implementation being the CMBFAST code of Seljak and Zaldarriaga [6]. Gravitational lensing is also calculated in this code. Secondary effects such as inhomogeneities in the re-ionization process, and scattering from gravitationally-collapsed gas (the Sunyaev–Zel'dovich effect), require more complicated, and more uncertain, calculations.

The upshot is that the detailed pattern of anisotropies, quantified, for instance, by the angular power spectrum C_ℓ, depends on all of the cosmological parameters. In a typical cosmology, the anisotropy power spectrum [usually plotted as $\ell(\ell+1)C_\ell$] features a flat plateau at large angular scales (small ℓ), followed by a series of oscillatory features at higher angular scales, the first and most prominent being at around one degree ($\ell \simeq 200$). These features, known as acoustic peaks, represent the oscillations of the photon-baryon fluid around the time of decoupling. Some features can be closely related to specific parameters—for instance, the location of the first peak probes the spatial geometry, while the relative heights of the peaks probes the baryon density—but many other parameters combine to determine the overall shape.

The three-year data release from the WMAP satellite [1], henceforth WMAP3, has provided the most accurate results to date on the spectrum of CMB fluctuations, with a precision determination of the temperature power spectrum up to $\ell \simeq 900$, shown in Fig. 21.2, and the best measurements of the spectrum of E-polarization anisotropies and the correlation spectrum between temperature and polarization (those spectra having first been detected by DASI [20]) . These are consistent with models based on the parameters we have described, and provide quite accurate determinations of many of them [2]. In this subsection, we will refer to results from WMAP alone, with the following section studying some combinations with other observations. We note that as the parameter fitting is done in a multi-parameter space, one has to assume a 'prior' range for each of the parameters (*e.g.*, Hubble constant $0.5 < h < 1$), and there may be some dependence on these assumed priors.

WMAP3 provides an exquisite measurement of the location of the first acoustic peak, which directly probes the spatial geometry and yields a total density $\Omega_{\rm tot} \equiv \sum \Omega_i + \Omega_\Lambda$ of

$$\Omega_{\rm tot} = 1.003^{+0.013}_{-0.017}\,, \tag{21.14}$$

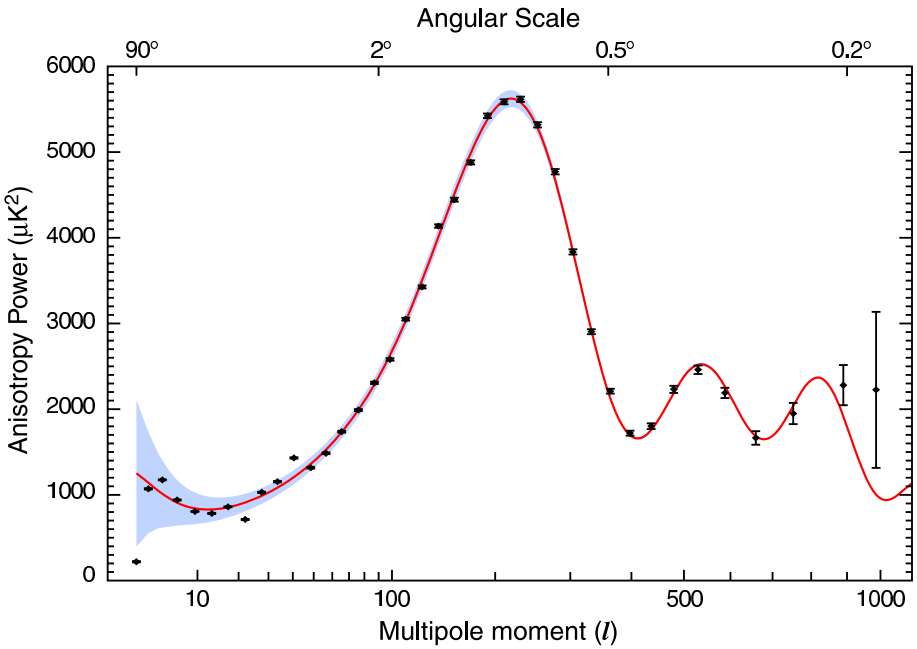


Figure 21.2: The angular power spectrum of the cosmic microwave background temperature from WMAP3. The solid line shows the prediction from the best-fitting ΛCDM model [2]. The error bars on the data points (which are tiny for most of them) indicate the observational errors, while the shaded region indicates the statistical uncertainty from being able to observe only one microwave sky, known as cosmic variance, which is the dominant uncertainty on large angular scales. See full-color version on color pages at end of book. [Figure courtesy NASA/WMAP Science Team.]

consistent with spatial flatness and completely excluding significantly curved Universes. (This result does however assume a prior on the Hubble parameter from other measurements; WMAP3 alone constrains $\Omega_{\rm tot}$ only weakly, and allows significantly closed Universes if h is small. This result also assumes that the dark energy is a cosmological constant.) WMAP3 also gives a precision measurement of the age of the Universe. It gives a baryon density consistent with, and at much higher precision than, that coming from nucleosynthesis. It affirms the need for both dark matter and dark energy if the data are to be explained. It shows no evidence for dynamics of the dark energy, being consistent with a pure cosmological constant ($w = -1$).

The density perturbations are consistent with a power-law primordial spectrum. There are indications that the spectral slope is less than the Harrison–Zel'dovich value $n = 1$ [2], though the result appears less strong using Bayesian techniques [8]. There is no indication of tensor perturbations, but the upper limit is quite weak.

WMAP3 gives a much lower result for the reionization optical depth τ than did their first year results [21]. The current best-fit value $\tau = 0.088$ is in reasonable agreement with models of how early structure formation induces reionization.

WMAP3 is consistent with other experiments and its dynamic range can be enhanced

by including information from small-angle CMB experiments including ACBAR, CBI and VSA. However the WMAP3 dataset on its own is so powerful that these add little constraining power.

21.3.4. *Galaxy clustering* :

The power spectrum of density perturbations depends on the nature of the dark matter. Within the Cold Dark Matter model, the shape of the power spectrum depends primarily on the primordial power spectrum and on the combination $\Omega_{\rm m}h$ which determines the horizon scale at matter–radiation equality, with a subdominant dependence on the baryon density. The matter distribution is most easily probed by observing the galaxy distribution, but this must be done with care as the galaxies do not perfectly trace the dark matter distribution. Rather, they are a 'biased' tracer of the dark matter. The need to allow for such bias is emphasized by the observation that different types of galaxies show bias with respect to each other. Further, the observed 3D galaxy distribution is in redshift space, *i.e.*, the observed redshift is the sum of the Hubble expansion and the line-of-sight peculiar velocity, leading to linear and non-linear dynamical effects which also depend on the cosmological parameters. On the largest length scales, the galaxies are expected to trace the location of the dark matter, except for a constant multiplier b to the power spectrum, known as the linear bias parameter. On scales smaller than 20 h^{-1} Mpc or so, the clustering pattern is 'squashed' in the radial direction due to coherent infall, which depends on the parameter $\beta \equiv \Omega_{\rm m}^{0.6}/b$ (on these shorter scales, more complicated forms of biasing are not excluded by the data). On scales of a few h^{-1} Mpc, there is an effect of elongation along the line of sight (colloquially known as the 'finger of God' effect) which depends on the galaxy velocity dispersion $\sigma_{\rm p}$.

21.3.4.1. *The galaxy power spectrum:*

The 2-degree Field (2dF) Galaxy Redshift Survey is now complete and publicly available.** The power-spectrum analysis of the final 2dFGRS data set of approximately 220,000 galaxies was fitted to a CDM model [22]. It shows evidence for baryon acoustic oscillations, with baryon fraction $\Omega_{\rm b}/\Omega_{\rm m} = 0.185 \pm 0.046$ (1-σ errors). The shape of the power spectrum is characterized by $\Omega_{\rm m}h = 0.168 \pm 0.016$, and in combination with WMAP data gives $\Omega_{\rm m} = 0.231 \pm 0.021$ (see also Ref. 23). The 2dF power spectrum is compared with the Sloan Digital Sky Survey (SDSS) †† power spectrum [24] in Fig. 21.3. We see agreement in the gross features, but also some discrepancies. Eisenstein et al. [25] reported on detection of baryon acoustic peak in the large-scale correlation function of the SDSS sample of nearly 47,000 Luminous Red Galaxies. By using the baryon acoustic peak as a 'standard ruler' they found, independent of WMAP, that $\Omega_{\rm m} = 0.273 \pm 0.025$ for a flat ΛCDM model.

Combination of the 2dF data with the CMB indicates a 'biasing' parameter $b \sim 1$, in agreement with a 2dF-alone analysis of higher-order clustering statistics. However, results for biasing also depend on the length scale over which a fit is done, and the selection of the objects by luminosity, spectral type, or color. In particular, on scales smaller than

** See `http://www.mso.anu.edu.au/2dFGRS`

†† See `http://www.sdss.org`

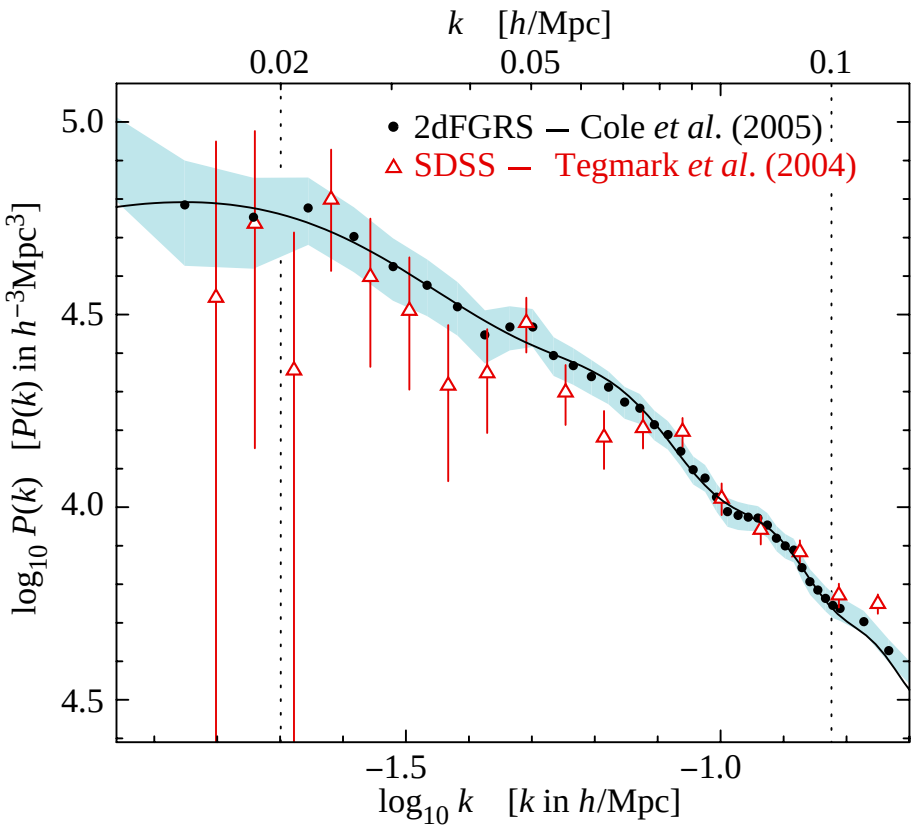


Figure 21.3: The galaxy power spectrum from the 2dF galaxy redshift survey [22] compared with that from SDSS [24], each corrected for its survey geometry. The 2dFGRS power spectrum (with distances measured in redshift space) is shown by solid circles with one-sigma errors shown by the shaded area. The triangles and error bars show the SDSS power spectrum. The solid curve shows a linear-theory ΛCDM model with $\Omega_{\rm m}h = 0.168$, $\Omega_{\rm b}/\Omega_{\rm m} = 0.17$, $h = 0.72$, $n = 1$ and normalization matched to the 2dFGRS power spectrum. The dotted vertical lines indicate the range over which the best-fit model was evaluated. See full-color version on color pages at end of book. [Figure provided by Shaun Cole and Will Percival; see Ref. 22.]

10 h^{-1}Mpc, different galaxy types are clustered differently. This 'biasing' introduces a systematic effect on the determination of cosmological parameters from redshift surveys. Prior knowledge from simulations of galaxy formation could help, but is model-dependent. We also note that the present-epoch power spectrum is not sensitive to dark energy, so it is mainly a probe of the matter density.

21.3.4.2. *Limits on neutrino mass from galaxy surveys and other probes:*

Large-scale structure data can put an upper limit on the ratio $\Omega_\nu/\Omega_{\rm m}$ due to the neutrino 'free streaming' effect [26]. By comparing the 2dF galaxy power spectrum with a four-component model (baryons, cold dark matter, a cosmological constant, and massive neutrinos), it was estimated that $\Omega_\nu/\Omega_{\rm m} < 0.13$ (95% confidence limit), giving $\Omega_\nu < 0.04$ if a concordance prior of $\Omega_{\rm m} = 0.3$ is imposed. The latter corresponds to an upper limit of about 2 eV on the total neutrino mass, assuming a prior of $h \approx 0.7$ [27]. Potential systematic effects include biasing of the galaxy distribution and non-linearities of the power spectrum. A similar upper limit of 2 eV was derived from CMB anisotropies alone [28]. The above analyses assume that the primordial power spectrum is adiabatic, scale-invariant and Gaussian.

Additional cosmological data sets bring down this upper limit [29,30]. An upper limit on the total neutrino mass of 0.17 eV was reported by combining a large number of cosmological probes [31].

Laboratory limits on absolute neutrino masses from tritium beta decay and especially from neutrinoless double-beta decay should, within the next decade, push down towards (or perhaps even beyond) the 0.1 eV level that has cosmological significance.

21.3.5. ***Clusters of galaxies*** :

A cluster of galaxies is a large collection of galaxies held together by their mutual gravitational attraction. The largest ones are around 10^{15} solar masses, and are the largest gravitationally-collapsed structures in the Universe. Even at the present epoch they are relatively rare, with only a few percent of galaxies being in clusters. They provide various ways to study the cosmological parameters; here we discuss constraints from the measurements of the cluster number density and the baryon fraction in clusters.

21.3.5.1. *Cluster number density:*

The first objects of a given kind form at the rare high peaks of the density distribution, and if the primordial density perturbations are Gaussian-distributed, their number density is exponentially sensitive to the size of the perturbations, and hence can strongly constrain it. Clusters are an ideal application in the present Universe. They are usually used to constrain the amplitude σ_8, as a box of side $8\,h^{-1}\,$Mpc contains about the right amount of material to form a cluster. The most useful observations at present are of X-ray emission from hot gas lying within the cluster, whose temperature is typically a few keV, and which can be used to estimate the mass of the cluster. A theoretical prediction for the mass function of clusters can come either from semi-analytic arguments or from numerical simulations. At present, the main uncertainty is the relation between the observed gas temperature and the cluster mass, despite extensive study using simulations. A recent analysis [32] gives

$$\sigma_8 = 0.78^{+0.30}_{-0.06} \qquad (95\%\ \text{confidence}) \tag{21.15}$$

for $\Omega_{\rm m} = 0.35$, with highly non-Gaussian error bars, but different authors still find a spread of values. Scaling to lower $\Omega_{\rm m}$ increases σ_8. This result is somewhat above the values predicted in cosmologies compatible with WMAP3.

The same approach can be adopted at high redshift (which for clusters means redshifts approaching one) to attempt to measure σ_8 at an earlier epoch. The evolution of σ_8 is primarily driven by the value of the matter density $\Omega_{\rm m}$, with a sub-dominant dependence on the dark energy density. It is generally recognized that such analyses favor a low matter density, though there is not complete consensus on this, and at present this technique for constraining the density is not competitive with the CMB.

21.3.5.2. *Cluster baryon fraction:*

If clusters are representative of the mass distribution in the Universe, the fraction of the mass in baryons to the overall mass distribution would be $f_{\rm b} = \Omega_{\rm b}/\Omega_{\rm m}$. If $\Omega_{\rm b}$, the baryon density parameter, can be inferred from the primordial nucleosynthesis abundance of the light elements, the cluster baryon fraction $f_{\rm b}$ can then be used to constrain $\Omega_{\rm m}$ and h (*e.g.*, Ref. 33). The baryons in clusters are primarily in the form of X-ray-emitting gas that falls into the cluster, and secondarily in the form of stellar baryonic mass. Hence, the baryon fraction in clusters is estimated to be

$$f_{\rm b} = \frac{\Omega_{\rm b}}{\Omega_{\rm m}} \simeq f_{\rm gas} + f_{\rm gal}\,, \tag{21.16}$$

where $f_{\rm b} = M_{\rm b}/M_{\rm grav}$, $f_{\rm gas} = M_{\rm gas}/M_{\rm grav}$, $f_{\rm gal} = M_{\rm gal}/M_{\rm grav}$, and $M_{\rm grav}$ is the total gravitating mass.

This can be used to obtain an approximate relation between $\Omega_{\rm m}$ and h:

$$\Omega_{\rm m} = \frac{\Omega_{\rm b}}{f_{\rm gas} + f_{\rm gal}} \simeq \frac{\Omega_{\rm b}}{0.08h^{-1.5} + 0.01h^{-1}}\,. \tag{21.17}$$

Big Bang Nucleosynthesis gives $\Omega_{\rm b}h^2 \approx 0.02$, allowing the above relation to be approximated as $\Omega_{\rm m}h^{0.5} \approx 0.25$ (*e.g.*, Ref. 34). For example, Allen *et al.* [35] derived a density parameter consistent with $\Omega_{\rm m} = 0.3$ from Chandra observations.

21.3.6. ***Clustering in the inter-galactic medium*** :

It is commonly assumed, based on hydrodynamic simulations, that the neutral hydrogen in the inter-galactic medium (IGM) can be related to the underlying mass distribution. It is then possible to estimate the matter power spectrum on scales of a few megaparsecs from the absorption observed in quasar spectra, the so-called Lyman-alpha forest. The usual procedure is to measure the power spectrum of the transmitted flux, and then to infer the mass power spectrum. Photo-ionization heating by the ultraviolet background radiation and adiabatic cooling by the expansion of the Universe combine to give a simple power-law relation between the gas temperature and the baryon density. It also follows that there is a power-law relation between the optical depth τ and $\rho_{\rm b}$. Therefore, the observed flux $F = \exp(-\tau)$ is strongly correlated with $\rho_{\rm b}$, which itself traces the mass density. The matter and flux power spectra can be related by

$$P_{\rm m}(k) = b^2(k)\, P_F(k)\,, \tag{21.18}$$

where $b(k)$ is a bias function which is calibrated from simulations. Croft *et al.* [36] derived cosmological parameters from Keck Telescope observations of the Lyman-alpha forest at redshifts $z = 2 - 4$. Their derived power spectrum corresponds to that of a CDM model, which is in good agreement with the 2dF galaxy power spectrum. A recent study using VLT spectra [37] agrees with the flux power spectrum of Ref. 36. This method depends on various assumptions. Seljak *et al.* [38] pointed out that errors are sensitive to the range of cosmological parameters explored in the simulations, and the treatment of the mean transmitted flux.

21.3.7. *Gravitational lensing* :

Images of background galaxies get distorted due to the gravitational effect of mass fluctuations along the line of sight. Deep gravitational potential wells such as galaxy clusters generate 'strong lensing', *i.e.*, arcs and arclets, while more moderate fluctuations give rise to 'weak lensing'. Weak lensing is now widely used to measure the mass power spectrum in random regions of the sky (see Ref. 39 for recent reviews). As the signal is weak, the CCD frame of deformed galaxy shapes ('shear map') is analyzed statistically to measure the power spectrum, higher moments, and cosmological parameters.

The shear measurements are mainly sensitive to the combination of $\Omega_{\rm m}$ and the amplitude σ_8. There are various systematic effects in the interpretation of weak lensing, *e.g.*, due to atmospheric distortions during observations, the redshift distribution of the background galaxies, intrinsic correlation of galaxy shapes, and non-linear modeling uncertainties. Hoekstra *et al.* [40] derived the result $\sigma_8\Omega_{\rm m}^{0.52} = 0.46^{+0.05}_{-0.07}$ (95% confidence level), assuming a ΛCDM model. Other recent results are summarized in Ref. 39. For a $\Omega_{\rm m} = 0.3, \Omega_\Lambda = 0.7$ cosmology, different groups derived normalizations σ_8 over a wide range, indicating that the systematic errors are still larger than some of the quoted error bars.

21.3.8. *Peculiar velocities* :

Deviations from the Hubble flow directly probe the mass fluctuations in the Universe, and hence provide a powerful probe of the dark matter. Peculiar velocities are deduced from the difference between the redshift and the distance of a galaxy. The observational difficulty is in accurately measuring distances to galaxies. Even the best distance indicators (*e.g.*, the Tully–Fisher relation) give an error of 15% per galaxy, hence limiting the application of the method at large distances. Peculiar velocities are mainly sensitive to $\Omega_{\rm m}$, not to Ω_Λ or quintessence. Extensive analyses in the early 1990s (*e.g.*, Ref. 41) suggested a value of $\Omega_{\rm m}$ close to unity. A more recent analysis [42], which takes into account non-linear corrections, gives $\sigma_8\Omega_{\rm m}^{0.6} = 0.49 \pm 0.06$ and $\sigma_8\Omega_{\rm m}^{0.6} = 0.63 \pm 0.08$ (90% errors) for two independent data sets. While at present cosmological parameters derived from peculiar velocities are strongly affected by random and systematic errors, a new generation of surveys may improve their accuracy. Three promising approaches are the 6dF near-infrared survey of 15,000 peculiar velocities[‡‡], Supernovae Type Ia and the kinematic Sunyaev–Zel'dovich effect.

[‡‡] See `http://www.mso.anu.edu.au/6dFGS/`

21.4. Bringing observations together

Although it contains two ingredients—dark matter and dark energy—which have not yet been verified by laboratory experiments, the ΛCDM model is almost universally accepted by cosmologists as the best description of present data. The basic ingredients are given by the parameters listed in Sec. 21.1.4, with approximate values of some of the key parameters being $\Omega_{\rm b} \approx 0.04$, $\Omega_{\rm dm} \approx 0.20$, $\Omega_\Lambda \approx 0.76$, and a Hubble constant $h \approx 0.73$. The spatial geometry is very close to flat (and often assumed to be precisely flat), and the initial perturbations Gaussian, adiabatic, and nearly scale-invariant.

Table 21.2: Parameter constraints reproduced from Spergel *et al.* [2]. All columns assume the ΛCDM cosmology with a power-law initial spectrum, no tensors, spatial flatness, and a cosmological constant as dark energy. Three different data combinations are shown to highlight the extent to which this choice matters. The first column is WMAP3 alone, the second combines this with 2dF, and the third column shows a combination of all datasets considered in Ref. [2]. This last data is not in their paper but can be found online at `http://lambda.gsfc.nasa.gov`. The parameter A is a measure of the perturbation amplitude; see Ref. 2 for details. Uncertainties are shown at one sigma, and caution is needed in extrapolating them to higher significance levels due to non-Gaussian likelihoods and assumed priors.

	WMAP alone	WMAP + 2dF	WMAP + all
$\Omega_{\rm m}h^2$	$0.127^{+0.007}_{-0.009}$	$0.126^{+0.004}_{-0.006}$	$0.131^{+0.004}_{-0.010}$
$\Omega_{\rm b}h^2$	$0.0223^{+0.0007}_{-0.0009}$	$0.0222^{+0.0007}_{-0.0008}$	$0.0220^{+0.0006}_{-0.0008}$
h	$0.73^{+0.03}_{-0.04}$	$0.73^{+0.02}_{-0.03}$	$0.71^{+0.01}_{-0.02}$
n	$0.951^{+0.015}_{-0.019}$	$0.948^{+0.013}_{-0.018}$	$0.938^{+0.013}_{-0.018}$
τ	$0.088^{+0.028}_{-0.034}$	$0.083^{+0.027}_{-0.031}$	$0.069^{+0.026}_{-0.029}$
A	$0.68^{+0.04}_{-0.06}$	$0.68^{+0.04}_{-0.05}$	$0.67^{+0.04}_{-0.05}$

The most powerful single experiment is WMAP3, which on its own supports all these main tenets. Values for some parameters, as given in Spergel *et al.* [2], are reproduced in Table 21.2. This model presumes a flat Universe, and so Ω_Λ is a derived quantity in this analysis, with best-fit value $\Omega_\Lambda = 0.76$.

These constraints can be somewhat strengthened by adding additional datasets, as shown in the Table. However WMAP3 on its own is sufficiently powerful that inclusion of other datasets only changes things at quite a detailed level. In our view, the most robust present constraints are those from WMAP3 alone.

The baryon density Ω_b is now measured with quite high accuracy from the CMB and large-scale structure, and is consistent with the determination from big bang nucleosynthesis; Fields and Sarkar in this volume quote the range $0.017 \leq \Omega_{\rm b} h^2 \leq 0.024$. Given the sensitivity of the measurement, it is important to note that it has some dependence on the datasets chosen, as seen in Table 21.2.

While Ω_Λ is measured to be non-zero with very high confidence, there is no evidence of evolution of the dark energy density. The WMAP team find the limit $w < -0.78$ at 95% confidence from a compilation of data including SNe Ia data, with the cosmological constant case $w = -1$ giving an excellent fit to the data.

The data provides strong support for the main predictions of the simplest inflation models: spatial flatness and adiabatic, Gaussian, nearly scale-invariant density perturbations. But it is disappointing that there is no sign of primordial gravitational waves, with WMAP3 providing only a weak upper limit $r < 0.55$ at 95% confidence [2] (this assumes no running, and weakens significantly if running is allowed). The spectral index n is placed in an interesting position by WMAP3, with indications that $n < 1$ is required by the data. However the conclusion that $n = 1$ is ruled out that is suggested by parameter estimation [2] receives much less compelling support in Bayesian model selection analyses [8] and in our view it is premature to conclude that $n = 1$ is no longer viable.

Tests have been made for various types of non-Gaussianity, a particular example being a parameter $f_{\rm NL}$ which measures a quadratic contribution to the perturbations and is constrained to $-54 < f_{\rm NL} < 114$ at 95% confidence [2] (this looks weak, but prominent non-Gaussianity requires the product $f_{\rm NL}\Delta_{\mathcal{R}}$ to be large, and $\Delta_{\mathcal{R}}$ is of order 10^{-5}).

Two parameters which are still uncertain are $\Omega_{\rm m}$ and σ_8, both of which were revised downwards significantly by WMAP3 to a level where they do not sit well against local measures of σ_8, particularly those using weak gravitational lensing. However an analysis including Lyman-alpha data with WMAP3 has found that this brings σ_8 up again [31]. It is clear that we have yet to reach the last word on these parameters. It is also worth noting that WMAP3 only probes larger length scales, and the constraint comes from using WMAP to estimate all the parameters of the model needed to determine σ_8. As such, their constraint depends strongly on the assumed set of cosmological parameters being sufficient.

One parameter which is surprisingly robust is the age of the Universe. There is a useful coincidence that for a flat Universe the position of the first peak is strongly correlated with the age of the Universe. The WMAP3 result is $13.7^{+0.1}_{-0.2}$ Gyr (assuming a flat Universe). This is in good agreement with the ages of the oldest globular clusters [43] and radioactive dating [44].

21.5. Outlook for the future

The concordance model is now well established, and there seems little room left for any dramatic revision of this paradigm. A measure of the strength of that statement is how difficult it has proven to formulate convincing alternatives. For example, one corner of parameter space that has been explored is the possibility of abandoning the dark energy, and instead considering a mixed dark matter model with $\Omega_{\rm m} = 1$ and $\Omega_\nu = 0.2$. Such a model fits both the 2dF and WMAP data reasonably well, but only for a Hubble constant $h < 0.5$ [27,45]. However, this model is inconsistent with the HST key project value of h, the results from SNe Ia, cluster number density evolution, and baryon fraction in clusters.

Should there indeed be no major revision of the current paradigm, we can expect future developments to take one of two directions. Either the existing parameter set will continue to prove sufficient to explain the data, with the parameters subject to ever-tightening constraints, or it will become necessary to deploy new parameters. The latter outcome would be very much the more interesting, offering a route towards understanding new physical processes relevant to the cosmological evolution. There are many possibilities on offer for striking discoveries, for example:

- The cosmological effects of a neutrino mass may be unambiguously detected, shedding light on fundamental neutrino properties;
- Compelling detection of deviations from scale-invariance in the initial perturbations would indicate dynamical processes during perturbation generation by, for instance, inflation;
- Detection of primordial non-Gaussianities would indicate that non-linear processes influence the perturbation generation mechanism;
- Detection of variation in the dark energy density (*i.e.*, $w \neq -1$) would provide much-needed experimental input into the question of the properties of the dark energy.

These provide more than enough motivation for continued efforts to test the cosmological model and improve its precision.

Over the coming years, there are a wide range of new observations, which will bring further precision to cosmological studies. Indeed, there are far too many for us to be able to mention them all here, and so we will just highlight a few areas.

The cosmic microwave background observations will improve in several directions. The new frontier is the study of polarization, first detected in 2002 by DASI and for which power spectrum measurements have now been made by WMAP and Boomerang [46]. Dedicated ground-based polarization experiments, such as CBI, QUaD, and Clover promise powerful measures of the polarization spectrum in the next few years, and may be able to separately detect the two modes of polarization. Another area of development is pushing accurate power spectrum measurements to smaller angular scales, typically achieved by interferometry, which should allow measurements of secondary anisotropy effects, such as the Sunyaev–Zel'dovich effect, whose detection has already been tentatively claimed by CBI. Finally, we mention the Planck satellite, due to launch in 2008, which will make high-precision all-sky maps of temperature and polarization, utilizing a very wide

frequency range for observations to improve understanding of foreground contaminants, and to compile a large sample of clusters via the Sunyaev–Zel'dovich effect.

On the supernova side, the most ambitious initiative at present is a proposed satellite mission JDEM (Joint Dark Energy Mission) funded by NASA and DoE. There are several candidates for this mission, including the much-publicized SNAP satellite, but the funding has yet to be secured. An impressive array of ground-based dark energy surveys are also already operational or proposed, including the ESSENCE project, the Dark Energy Survey, LSST and WFMOS. With large samples, it may be possible to detect evolution of the dark energy density, thus measuring its equation of state and perhaps even its variation.

The above future surveys will address fundamental questions of physics well beyond just testing the 'concordance' ΛCDM model and minor variations. It would be important to distinguish the imprint of dark energy and dark matter on the geometry from the growth of perturbations and to test theories of modified gravity as alternatives for fitting the observations to a Dark Energy component.

The development of the first precision cosmological model is a major achievement. However, it is important not to lose sight of the motivation for developing such a model, which is to understand the underlying physical processes at work governing the Universe's evolution. On that side, progress has been much less dramatic. For instance, there are many proposals for the nature of the dark matter, but no consensus as to which is correct. The nature of the dark energy remains a mystery. Even the baryon density, now measured to an accuracy of a few percent, lacks an underlying theory able to predict it even within orders of magnitude. Precision cosmology may have arrived, but at present many key questions remain unanswered.

Acknowledgements:

ARL was supported by the Leverhulme Trust, and both authors acknowledge PPARC Senior Research Fellowships. We thank Sarah Bridle and Jochen Weller for useful comments on this article, and OL thanks members of the Cambridge Leverhulme Quantitative Cosmology and 2dFGRS Teams for helpful discussions.

References:

1. G. Hinshaw *et al.*, `astro-ph/0603451`;
   L. Page *et al.*, `astro-ph/0603450`;
   N. Jarosik *et al.*, `astro-ph/0603452`.
2. D. N. Spergel *et al.*, `astro-ph/0603449`.
3. S. Fukuda *et al.*, Phys. Rev. Lett. **85**, 3999 (2000);
   Q.R. Ahmad *et al.*, Phys. Rev. Lett. **87**, 071301 (2001).
4. A.D. Dolgov *et al.*, Nucl. Phys. **B632**, 363 (2002).
5. For detailed accounts of inflation see E.W. Kolb and M.S. Turner, *The Early Universe*, Addison–Wesley (Redwood City, 1990);
   A.R. Liddle and D.H. Lyth, *Cosmological Inflation and Large-Scale Structure*, (Cambridge University Press, 2000).

6. U. Seljak and M. Zaldarriaga, Astrophys. J. **469**, 1 (1996).
7. H.V. Peiris *et al.*, Astrophys. J. Supp. **148**, 213 (2003).
8. D. Parkinson *et al.*, `astro-ph/0605003`.
9. J.C. Mather *et al.*, Astrophys. J. **512**, 511 (1999).
10. A. Kosowsky and M.S. Turner, Phys. Rev. **D52**, 1739 (1995).
11. D.H. Lyth and D. Wands, Phys. Lett. **B524**, 5 (2002);
    K. Enqvist and M.S. Sloth, Nucl. Phys. **B626**, 395 (2002);
    T. Moroi and T. Takahashi, Phys. Lett. **B522**, 215 (2001).
12. B. Ratra and P.J.E. Peebles, Phys. Rev. **D37**, 3406 (1988);
    C. Wetterich, Nucl. Phys. **B302**, 668 (1988);
    T. Padmanabhan, Phys. Rept. **380**, 235 (2003);
    V. Sahni and A. Starobinsky, Int. J. Mod. Phys. **D9**, 373 (2000).
13. J.K. Webb *et al.*, Phys. Rev. Lett. **82**, 884 (1999);
    J.K. Webb *et al.*, Phys. Rev. Lett. **87**, 091301 (2001);
    J.K. Webb *et al.*, Astrophys. Sp. Sci. **283**, 565 (2003);
    H. Chand *et al.*, Astron. Astrophys. **417**, 853 (2004);
    R. Srianand *et al.*, Phys. Rev. Lett. **92**, 121302 (2004).
14. W.L. Freedman *et al.*, Astrophys. J. **553**, 47 (2001).
15. A. Filippenko, `astro-ph/0307139`.
16. A.G. Riess *et al.*, Astron. J. **116**, 1009 (1998);
    P. Garnavich *et al.*, Astrophys. J. **509**, 74 (1998);
    S. Perlmutter *et al.*, Astrophys. J. **517**, 565 (1999).
17. R.A. Knop *et al.*, Astrophys. J. **598**, 102 (2003).
18. J.L. Tonry *et al.*, Astrophys. J. **594**, 1 (2003);
    A.G. Riess *et al.*, Astrophys. J. **607**, 665 (2004).
19. I. Maor *et al.*, Phys. Rev. **D65**, 123003 (2002).
20. J. Kovac *et al.*, Nature **420**, 772 (2002).
21. D.N. Spergel *et al.*, Astrophys. J. Supp. **148**, 175 (2003).
22. S. Cole *et al.*, Mon. Not. Roy. Astr. Soc. **362**, 505 (2005).
23. A. Sanchez *et al.*, Mon. Not. Roy. Astr. Soc. **366**, 189 (2006).
24. M. Tegmark *et al.*, Astrophys. J. **606**, 702 (2004).
25. D. Eisenstein *et al.*, Astrophys. J. **633**, 560 (2005).
26. W. Hu *et al.*, Phys. Rev. Lett. **80**, 5255 (1998).
27. O. Elgaroy and O. Lahav, JCAP **0304**, 004 (2003).
28. K. Ichikawa *et al.*, Phys. Rev. D**71**, 043001 (2005).
29. S. Hannestad, JCAP **0305**, 004 (2003).
30. O. Elgaroy and O. Lahav, New J. Physics, **7**, 61 (2005).
31. U. Seljak *et al.*, `astro-ph/0604335`.
32. P.T.P. Viana *et al.*, Mon. Not. Roy. Astr. Soc., **346**, 319 (2003).
33. S.D.M. White *et al.*, Nature **366**, 429 (1993).
34. P. Erdogdu *et al.*, Mon. Not. Roy. Astr. Soc. **340**, 573 (2003).
35. S.W. Allen *et al.*, Mon. Not. Roy. Astr. Soc. **342**, 287 (2003).
36. R.A.C. Croft *et al.*, Astrophys. J. **581**, 20 (2002).
37. S. Kim *et al.*, Mon. Not. Roy. Astr. Soc. **347**, 355 (2004).

38. U. Seljak *et al.*, Mon. Not. Roy. Astr. Soc. **342**, L79 (2003);
U. Seljak *et al.*, Phys. Rev. D**71**, 103515 (2005).
39. P. Schneider, `astro-ph/0306465`;
A. Refregier, Ann. Rev. Astron. Astrophys, **41**, 645 (2003).
40. H. Hoekstra *et al.*, Astrophys. J. **577**, 595 (2002).
41. A. Dekel, Ann. Rev. Astron. Astrophys. **32**, 371 (1994).
42. L. Silberman *et al.*, Astrophys. J. **557**, 102 (2001).
43. B. Chaboyer and L.M. Krauss, Science **299**, 65 (2003).
44. R. Cayrel *et al.*, Nature **409**, 691 (2001).
45. A. Blanchard *et al.*, Astron. Astrophys. **412**, 35 (2003).
46. C.J. MacTavish *et al.*, `astro-ph/0507503`.

22. DARK MATTER

Written September 2003 by M. Drees (Technical University, Munich) and G. Gerbier (Saclay, CEA). Revised September 2005.

22.1. Theory

22.1.1. *Evidence for Dark Matter* :

The existence of Dark (*i.e.*, non-luminous and non-absorbing) Matter (DM) is by now well established. The earliest [1], and perhaps still most convincing, evidence for DM came from the observation that various luminous objects (stars, gas clouds, globular clusters, or entire galaxies) move faster than one would expect if they only felt the gravitational attraction of other visible objects. An important example is the measurement of galactic rotation curves. The rotational velocity v of an object on a stable Keplerian orbit with radius r around a galaxy scales like $v(r) \propto \sqrt{M(r)/r}$, where $M(r)$ is the mass inside the orbit. If r lies outside the visible part of the galaxy and mass tracks light, one would expect $v(r) \propto 1/\sqrt{r}$. Instead, in most galaxies one finds that v becomes approximately constant out to the largest values of r where the rotation curve can be measured; in our own galaxy, $v \simeq 220$ km/s at the location of our solar system, with little change out to the largest observable radius. This implies the existence of a *dark halo*, with mass density $\rho(r) \propto 1/r^2$, *i.e.*, $M(r) \propto r$; at some point ρ will have to fall off faster (in order to keep the total mass of the galaxy finite), but we do not know at what radius this will happen. This leads to a lower bound on the DM mass density, $\Omega_{\rm DM} \gtrsim 0.1$, where $\Omega_X \equiv \rho_X/\rho_{\rm crit}$, $\rho_{\rm crit}$ being the critical mass density (*i.e.*, $\Omega_{\rm tot} = 1$ corresponds to a flat Universe).

The observation of clusters of galaxies tends to give somewhat larger values, $\Omega_{\rm DM} \simeq 0.2$ to 0.3. These observations include measurements of the peculiar velocities of galaxies in the cluster, which are a measure of their potential energy if the cluster is virialized; measurements of the *X-ray* temperature of hot gas in the cluster, which again correlates with the gravitational potential felt by the gas; and—most directly—studies of (weak) gravitational lensing of background galaxies on the cluster.

The currently most accurate, if somewhat indirect, determination of $\Omega_{\rm DM}$ comes from global fits of cosmological parameters to a variety of observations; see the Section on Cosmological Parameters for details. For example, using measurements of the anisotropy of the cosmic microwave background (CMB) and of the spatial distribution of galaxies, Ref. 2 finds a density of cold, non–baryonic matter

$$\Omega_{\rm nbm} h^2 = 0.111 \pm 0.006 \ , \tag{22.1}$$

where h is the Hubble constant in units of 100 km/(s·Mpc). Some part of the baryonic matter density [2],

$$\Omega_{\rm b} h^2 = 0.023 \pm 0.001 \ , \tag{22.2}$$

may well contribute to (baryonic) DM, *e.g.*, MACHOs [3] or cold molecular gas clouds [4].

The DM density in the "neighborhood" of our solar system is also of considerable interest. This was first estimated as early as 1922 by J.H. Jeans, who analyzed the motion of nearby stars transverse to the galactic plane [1]. He concluded that in our galactic neighborhood the average density of DM must be roughly equal to that of luminous

matter (stars, gas, dust). Remarkably enough, the most recent estimates, based on a detailed model of our galaxy, find quite similar results [5]:

$$\rho_{\rm DM}^{\rm local} \simeq 0.3 \frac{\rm GeV}{\rm cm^3} \ ; \tag{22.3}$$

this value is known to within a factor of two or so.

22.1.2. *Candidates for Dark Matter* :

Analyses of structure formation in the Universe [6] indicate that most DM should be "cold", *i.e.*, should have been non-relativistic at the onset of galaxy formation (when there was a galactic mass inside the causal horizon). This agrees well with the upper bound [2] on the contribution of light neutrinos to Eq. (22.1),

$$\Omega_\nu h^2 \leq 0.0076 \quad 95\% \ {\rm CL} \tag{22.4}$$

Candidates for non-baryonic DM in Eq. (22.1) must satisfy several conditions: they must be stable on cosmological time scales (otherwise they would have decayed by now), they must interact very weakly with electromagnetic radiation (otherwise they wouldn't qualify as *dark* matter), and they must have the right relic density. Candidates include primordial black holes, axions, and weakly interacting massive particles (WIMPs).

Primordial black holes must have formed before the era of Big-Bang nucleosynthesis, since otherwise they would have been counted in Eq. (22.2) rather than Eq. (22.1). Such an early creation of a large number of black holes is possible only in certain somewhat contrived cosmological models [7].

The existence of axions [8] was first postulated to solve the strong *CP* problem of QCD; they also occur naturally in superstring theories. They are pseudo Nambu-Goldstone bosons associated with the (mostly) spontaneous breaking of a new global "Peccei-Quinn" (PQ) U(1) symmetry at scale f_a; see the Section on Axions in this *Review* for further details. Although very light, axions would constitute cold DM, since they were produced non-thermally. At temperatures well above the QCD phase transition, the axion is massless, and the axion field can take any value, parameterized by the "misalignment angle" θ_i. At $T \lesssim 1$ GeV the axion develops a mass m_a due to instanton effects. Unless the axion field happens to find itself at the minimum of its potential ($\theta_i = 0$), it will begin to oscillate once m_a becomes comparable to the Hubble parameter H. These coherent oscillations transform the energy originally stored in the axion field into physical axion quanta. The contribution of this mechanism to the present axion relic density is [8]

$$\Omega_a h^2 = \kappa_a \left(f_a / 10^{12} \ {\rm GeV} \right)^{1.175} \theta_i^2 \ , \tag{22.5}$$

where the numerical factor κ_a lies roughly between 0.5 and a few. If $\theta_i \sim \mathcal{O}(1)$, Eq. (22.5) will saturate Eq. (22.1) for $f_a \sim 10^{11}$ GeV, comfortably above laboratory and astrophysical constraints [8]; this would correspond to an axion mass around 0.1 meV. However, if the post-inflationary reheat temperature $T_R > f_a$, cosmic strings will

form during the PQ phase transition at $T \simeq f_a$. Their decay will give an additional contribution to Ω_a, which is often bigger than that in Eq. (22.5) [9], leading to a smaller preferred value of f_a, *i.e.*, larger m_a. On the other hand, values of f_a near the Planck scale become possible if θ_i is for some reason very small.

Weakly interacting massive particles (WIMPs) χ are particles with mass roughly between 10 GeV and a few TeV, and with cross sections of approximately weak strength. Their present relic density can be calculated reliably if the WIMPs were in thermal and chemical equilibrium with the hot "soup" of Standard Model (SM) particles after inflation. In this case their density would become exponentially (Boltzmann) suppressed at $T < m_\chi$. The WIMPs therefore drop out of thermal equilibrium ("freeze out") once the rate of reactions that change SM particles into WIMPs or vice versa, which is proportional to the product of the WIMP number density and the WIMP pair annihilation cross section into SM particles σ_A times velocity, becomes smaller than the Hubble expansion rate of the Universe. After freeze out, the co-moving WIMP density remains essentially constant. Their present relic density is then approximately given by (ignoring logarithmic corrections) [10]

$$\Omega_\chi h^2 \simeq const. \cdot \frac{T_0^3}{M_{\rm Pl}^3 \langle \sigma_A v \rangle} \simeq \frac{0.1 \text{ pb} \cdot c}{\langle \sigma_A v \rangle} \ . \tag{22.6}$$

Here T_0 is the current CMB temperature, $M_{\rm Pl}$ is the Planck mass, c is the speed of light, σ_A is the total annihilation cross section of a pair of WIMPs into SM particles, v is the relative velocity between the two WIMPs in their cms system, and $\langle \ldots \rangle$ denotes thermal averaging. Freeze out happens at temperature $T_F \simeq m_\chi/20$ almost independently of the properties of the WIMP. This means that WIMPs are already non-relativistic when they decouple from the thermal plasma; it also implies that Eq. (22.6) is applicable if $T_R \gtrsim m_\chi/10$. Notice that the 0.1 pb in Eq. (22.6) contains factors of T_0 and $M_{\rm Pl}$; it is therefore quite intriguing that it "happens" to come out near the typical size of weak interaction cross sections.

The seemingly most obvious WIMP candidate is a heavy neutrino. However, an SU(2) doublet neutrino will have too small a relic density if its mass exceeds $M_Z/2$, as required by LEP data. One can suppress the annihilation cross section, and hence increase the relic density, by postulating mixing between a heavy SU(2) doublet and some "sterile" SU(2) $\times$ U(1)$_Y$ singlet neutrino. However, one also has to require the neutrino to be stable; it is not obvious why a massive neutrino should not be allowed to decay.

The currently best motivated WIMP candidate is therefore the lightest superparticle (LSP) in supersymmetric models [11] with exact R-parity (which guarantees the stability of the LSP). Searches for exotic isotopes [12] imply that a stable LSP has to be neutral. This leaves basically two candidates among the superpartners of ordinary particles, a sneutrino, and a neutralino. Sneutrinos again have quite large annihilation cross sections; their masses would have to exceed several hundred GeV for them to make good DM candidates. This is uncomfortably heavy for the lightest sparticle, in view of naturalness arguments. Moreover, the negative outcome of various WIMP searches (see below) rules out "ordinary" sneutrinos as primary component of the DM halo of our galaxy. (In models with gauge-mediated SUSY breaking the lightest "messenger sneutrino" could

make a good WIMP [13].) The most widely studied WIMP is therefore the lightest neutralino. Detailed calculations [14] show that the lightest neutralino will have the desired thermal relic density Eq. (22.1) in at least four distinct regions of parameter space. χ could be (mostly) a bino or photino (the superpartner of the $U(1)_Y$ gauge boson and photon, respectively), if both χ and some sleptons have mass below ~ 150 GeV, or if m_χ is close to the mass of some sfermion (so that its relic density is reduced through co-annihilation with this sfermion), or if $2m_\chi$ is close to the mass of the *CP*-odd Higgs boson present in supersymmetric models [15]. Finally, Eq. (22.1) can also be satisfied if χ has a large higgsino component.

Although WIMPs are attractive DM candidates because their thermal relic density naturally has at least the right order of magnitude, non-thermal production mechanisms have also been suggested, *e.g.*, LSP production from the decay of some moduli fields [16], from the decay of the inflaton [17], or from the decay of "$Q-$balls" (non-topological solitons) formed in the wake of Affleck-Dine baryogenesis [18]. Although LSPs from these sources are typically highly relativistic when produced, they quickly achieve kinetic (but not chemical) equilibrium if T_R exceeds a few MeV [19](but stays below $m_\chi/20$). They therefore also contribute to cold DM.

Primary black holes (as MACHOs), axions, and WIMPs are all (in principle) detectable with present or near-future technology (see below). There are also particle physics DM candidates which currently seem almost impossible to detect. These include the gravitino (the spin-3/2 superpartner of the graviton) [20], states from the "hidden sector" thought responsible for supersymmetry breaking [13], and the axino (the spin-1/2 superpartner of the axion) [21].

22.2. Experimental detection of Dark Matter

22.2.1. *The case of baryonic matter in our galaxy* :

The search for hidden galactic baryonic matter in the form of MAssive Compact Halo Objects (MACHOs) has been initiated following the suggestion that they may represent a large part of the galactic DM and could be detected through the microlensing effect [3]. The MACHO, EROS, and OGLE collaborations have performed a program of observation of such objects by monitoring the luminosity of millions of stars in the Large and Small Magellanic Clouds for several years. EROS concluded that MACHOs cannot contribute more than 20% to the mass of the galactic halo [22], while MACHO observed a signal at 0.4 solar mass and put an upper limit of 40%. Overall, this strengthens the need for non-baryonic DM, also supported by the arguments developed above.

22.2.2. *Axion searches* :

Axions can be detected by looking for $a \to \gamma$ conversion in a strong magnetic field [23]. Such a conversion proceeds through the loop-induced $a\gamma\gamma$ coupling, whose strength $g_{a\gamma\gamma}$ is an important parameter of axion models. Currently two experiments searching for axionic DM are taking data. They both employ high quality cavities. The cavity "Q factor" enhances the conversion rate on resonance, *i.e.*, for $m_a c^2 = \hbar\omega_{\rm res}$. One then needs to scan the resonance frequency in order to cover a significant range in m_a or, equivalently,

f_a. The bigger of the two experiments, situated at the LLNL in California [24], started taking data in the first half of 1996. It uses very sophisticated "conventional" electronic amplifiers with very low noise temperature to enhance the conversion signal. Their published results [25] exclude axions with mass between 1.9 and 3.3 μeV, corresponding to $f_a \simeq 4 \cdot 10^{13}$ GeV, as a major component of the dark halo of our galaxy, if $g_{a\gamma\gamma}$ is near the upper end of the theoretically expected range.

The smaller "CARRACK" experiment now under way in Kyoto, Japan [26] uses Rydberg atoms (atoms excited to a very high state, $n \simeq 230$) to detect the microwave photons that would result from axion conversion. This allows almost noise-free detection of single photons. Preliminary results of the CARRACK I experiment [27] exclude axions with mass in a narrow range around 10 μeV as major component of the galactic dark halo for some plausible range of $g_{a\gamma\gamma}$ values. This experiment is being upgraded to CARRACK II, which intends to probe the range between 2 and 50 μeV with sensitivity to all plausible axion models, if axions form most of DM [27].

22.2.3. *Basics of direct WIMP search* :

As stated above, WIMPs should be gravitationally trapped inside galaxies and should have the adequate density profile to account for the observed rotational curves. These two constraints determine the main features of experimental detection of WIMPs, which have been detailed in the reviews [28].

Their mean velocity inside our galaxy is expected to be similar to that of stars around the center of the galaxy, *i.e.*, a few hundred kilometers per second at the location of our solar system. For these velocities, WIMPs interact with ordinary matter through elastic scattering on nuclei. With expected WIMP masses in the range 10 GeV to 10 TeV, typical nuclear recoil energies are of order of 1 to 100 keV.

The shape of the nuclear recoil spectrum results from a convolution of the WIMP velocity distribution, usually taken as a shifted Maxwellian distribution, with the angular scattering distribution, which is isotropic to first approximation but forward-peaked for high nuclear mass (typically higher than Ge mass) due to the nuclear form factor. Overall, this results in a roughly exponential spectrum. The higher the WIMP mass, the higher the mean value of the exponential. This points to the need for low nuclear energy threshold detectors.

On the other hand, expected interaction rates depend on the product of the WIMP local flux and the interaction cross section. The first term is fixed by the local density of dark matter, taken as 0.3 GeV/cm^3 (see above), the mean WIMP velocity, typically 220 km/s, and the mass of the WIMP. The expected interaction rate then mainly depends on two unknowns, the mass and cross section of WIMP (with some uncertainty [5] due to the halo model). This is why the experimental observable, which is basically the scattering rate as a function of energy, is usually expressed as a contour in the WIMP mass—cross section plane.

The cross section depends on the nature of the couplings. For non-relativistic WIMPs one in general has to distinguish spin–independent and spin–dependent couplings. The former can involve scalar and vector WIMP and nucleon currents (vector currents are absent for Majorana WIMPs, e.g. the neutralino), while the latter involve axial

vector currents (and obviously only exist if χ carries spin). Due to coherence effects the spin–independent cross section scales approximately as the square of the mass of the nucleus, so higher mass nuclei, from Ge to Xe, are preferred for this search. For spin-dependent coupling, the cross section depends on the nuclear spin factor; the useful target nuclei are ^{19}F and ^{127}I.

Cross sections calculated in MSSM models induce rates of at most 1 evt day^{-1} kg^{-1} of detector, much lower than the usual radioactive backgrounds. This indicates the need for underground laboratories to protect against cosmic ray induced backgrounds, and for the selection of extremely radiopure materials.

The typical shape of exclusion contours can be anticipated from this discussion: at low WIMP mass, the sensitivity drops because of the detector energy threshold, whereas at high masses, the sensitivity also decreases because, for a fixed mass density, the WIMP flux decreases $\propto 1/m_\chi$. The sensitivity is best for WIMP masses near the mass of the recoiling nucleus.

22.2.4. *Status and prospects of direct WIMP searches* :

The first searches have been performed with ultra-pure semiconductors installed in pure lead and copper shields in underground environments [29]. Combining a priori excellent energy resolutions and very pure detector material, they produced the first limits on WIMP searches and until recently had the best performance (Heidelberg-Moscow, IGEX, COSME-II, HDMS) [29]. Without positive identification of nuclear recoil events, however, these experiments could only set limits, *e.g.*, excluding sneutrinos as major component of the galactic halo. Still, planned experiments using several tens of kgs to a ton of Germanium (many of which were designed for double-beta decay search)—GENIUS TF, GEDEON, MAJORANA—are based on only passive reduction of the external and internal electromagnetic, and neutron background by using segmented detectors, minimal detector housing, close electronics, and large liquid nitrogen shields.

To make further progress, active background rejection and signal identification questions have to be addressed. This has been the focus of many recent investigations and improvements. Active background rejection in detectors relies on the relatively small ionization in nuclear recoils due to their low velocity. This induces a reduction—quenching—of the ionization/scintillation signal for nuclear recoil signal events relative to e or γ induced backgrounds. Energies calibrated with gamma sources are then called "electron equivalent energies" (eee). This effects has been calculated and measured [29]. It is exploited in cryogenic detectors described later. In scintillation detectors, it induces in addition a difference in decay times of pulses induced by e/γ events vs nuclear recoils. Due to the limited resolution and discrimination power of this technique at low energies, this effect allows only a statistical background rejection. It has been used in NaI(Tl) (DAMA, NAIAD, Saclay NaI), in CsI(Tl)(KIMS), Xe (ZEPLIN) [29]. No observation of nuclear recoils has been reported by these experiments.

There are two experimental signatures of WIMP detection that would prove its astrophysical origin. One is the measurement of the strong daily forward/backward asymmetry of the nuclear recoil direction, due to the alternate sweeping of the WIMP cloud by the rotating Earth. Detection of this effect requires gaseous detectors or

anisotropic response scintillators (stilbene). The second is the few percent annual modulation of the recoil rate due to the Earth speed adding to or subtracting from the speed of the Sun. This tiny effect can only be detected with large masses; nuclear recoil identification should also be performed, as the much larger background may also be subject to modulation.

The DAMA experiment operating 100 kg of NaI(Tl) in Gran Sasso has observed, with a statistical significance of 6.3 σ, an annually modulated signal with the expected phase, over a period of 7 years with a total exposure of around 100 000 kg·d , in the 2 to 6 keV (eee) energy interval [30]. This effect is attributed to a WIMP signal by the authors. If interpreted within the standard halo model described above, it would require a WIMP with $m_\chi \simeq 50$ GeV and $\sigma_{\chi p} \simeq 7 \cdot 10^{-6}$ pb (central values). This interpretation has however several unaddressed implications. In particular, the expected nuclear recoil rate from WIMPs should be of the order of 50% of the total measured rate in the 2–3 keV (eee) bin and 7% in the 4–6 keV (eee) bin. The rather large WIMP signal should be detectable by the pulse shape analysis. Moreover, the remaining, presumably e/γ induced, background would have to rise with energy; no explanation for this is given by the authors.

Annual modulation has also been searched for by NaI-32 (Zaragoza), DEMOS (Ge), ELEGANTS (NaI) [29]. No signal has been seen in these experiments, but their sensitivity is too low to contradict DAMA.

New limits on the spin independent coupling of WIMPs were obtained by the CDMS collaboration which has operated Ge cryogenics detectors at the Soudan mine [31]. They supercede the earlier results of EDELWEISS also obtained with Ge cryogenics detectors in the deep underground Fréjus lab [32]. The simultaneous measurement of the phonon signal and the ionization signal in such semiconductor detectors permits event by event discrimination between nuclear and electronic recoils down to 5 to 10 keV recoil energy. In addition, the advanced rejection technique allowed CDMS to reject surface detector interactions, which mimic nuclear recoils. Assuming conventional WIMP halo parameters described above and spin independent coupling WIMP interactions, the CDMS limit and DAMA signal are clearly incompatible. Varying the halo parameters, and/or including spin–dependent interactions compatible with the neutrino flux limit from the Sun, does not allow reconciliation of both results without finetuning [33]. The obtained sensitivity of $\sigma_{\chi p} \simeq few \times 10^{-7}$ pb tests most optimistic cross sections that can be accommodated in the MSSM [34].

CDMS also provides the best sensitivity for spin dependent WIMP-neutron interactions thanks to the ^{73}Ge (^{29}Si) content of natural Germanium (Silicon) [35].

Other cryogenic experiments like CRESST and ROSEBUD [36] use the scintillation of $CaWO_4$ as second variable for background discrimination, while CUORICINO will use TeO_2 in the purely thermal mode. The cryogenic experimental programs of CDMS II, EDELWEISS II, CRESST II, CUORICINO, and ROSEBUD [36] intend to increase their sensitivity by a factor 100, by operating from a few to 40 kg of detectors.

Liquid or two-phase Xenon detectors are coming on line. ZEPLIN has been operating 6 kg in the Boulby laboratory for about 1 year and announced a sensitivity close to that of

EDELWEISS [37]. With only 1.5 photoelectron/keV (eee), and a three-fold coincidence, searching for the WIMP signal in the 2–10 keV (eee) region is quite challenging. In particular, the neutron recoil discrimination calibration results do not appear to be convincing enough to consider the limits set on the WIMP signal to be as reliable as the ones set by the cryogenic experiments cited above. With masses of 7 to 30 kg, ZEPLIN II and III aim at sensitivities down to 10^{-8} pb, while XMASS in Japan has operated a 100 kg detector at the SuperKamiokande site, and demonstrated the self-shielding effect to lower the background [29]. They intend to ultimately operate 800 kg while XENON in US has approximately the same program.

On the other hand, the extended version of DAMA, LIBRA, has started operating 250 kg of NaI(Tl) in Gran Sasso, ANAIS will operate 107 kg of NaI(Tl) in Canfranc laboratory, KIMS, 80 kg of CsI(Tl)in Yang Yang lab in Korea, and ELEGANTS VI the large shield of 750 kg of NaI(Tl) [29].

There is also continuous work in the development of a low pressure Time Projection Chamber, the only convincing technique to measure the direction of nuclear recoils [38]. DRIFT, a $1\mathrm{m}^3$ volume detector has been operated underground but did not obtain competitive results. The sub-keV energy threshold gaseous detector MICROMEGAS is being investigated for WIMP searches [38]. Other exotic techniques include the superheated droplet detectors SIMPLE, PICASSO, which has started to obtain interesting limits, and an ultra cold pure ^{3}He detector (MacHe3) has been operated with a very small sensitive mass [29].

Sensitivities down to $\sigma_{\chi p}$ of 10^{-10} pb, as needed to probe large regions of MSSM parameter space [34], can be reached with detectors of typical masses of 1 ton [36], assuming nearly perfect background discrimination capabilities. Note that the expected WIMP rate is then 5 evts/ton/year for Ge. The ultimate neutron background will only be identified by its multiple interactions in a finely segmented or multiple interaction sensitive detector, and/or by operating detectors containing different target materials within the same set–up. Information on various neutron backgrounds calculations and measurements can be found in [39]. With an intermediate mass of 10 to 30 kg and less efficient multiple interaction detection, a muon veto seems mandatory.

22.2.5. *Status and prospects of indirect WIMP searches* :

WIMPs can annihilate and their annihilation products can be detected; these include neutrinos, gamma rays, positrons, antiprotons, and antinuclei [40]. These methods are complementary to direct detection and can explore higher masses and different coupling scenarios. "Smoking gun" signals for indirect detection are neutrinos coming from the center of the Sun or Earth, and monoenergetic photons from the halo.

WIMPs can be slowed down, captured, and trapped in celestial objects like the Earth or the Sun, thus enhancing their density and their probability of annihilation. This is a source of muon neutrinos which can interact in the Earth. Upward going muons can then be detected in large neutrino telescopes such as MACRO, BAKSAN, SuperKamiokande, Baikal, AMANDA, ANTARES, NESTOR, and the future large sensitive area IceCube [40]. The best upper limits, of $\simeq$ 1000 muons/km2/year, have been set recently by SuperKamiokande [41]. However, at least in the framework of

the MSSM and with standard halo velocity profiles, only the limits from the Sun are competitive with direct WIMP search limits. ANTARES (IceCube) will increase this sensitivity respectively by $\simeq$ one (two) order(s) of magnitude.

WIMP annihilation in the halo can give a continuous spectrum of gamma rays and (at one-loop level) also monoenergetic photon contributions from the $\gamma\gamma$ and γZ channels. The size of this signal depends very strongly on the halo model, but is expected to be most prominent towards the galactic center. Existing limits come from the EGRET satellite below 10 GeV, and from the WHIPPLE ground based telescope above 100 GeV [42]. However, only the planned space mission GLAST will be able to provide competitive SUSY sensitivities in both the continuous and γ line channels. Also, Atmospheric Cherenkov Telescopes like MAGIC, VERITAS, and H.E.S.S. should be able to test some SUSY models, at large WIMP mass, for halo models showing a significant WIMP enhancement at the galactic center [40]. H.E.S.S. has actually recently observed TeV gamma rays from the galactic center. The original data [43] would have been compatible with a WIMP signal if its mass exceeds 12 TeV. Newer data presented at the 2005 ICRC in Pune, India, show a featureless spectrum, well described by a power law, extending beyond 20 TeV. Similarly, at the other end of the mass spectrum, a WIMP mass below 20 MeV would be required to explain the excess of 511 keV gamma rays from galactic center observed by INTEGRAL [44]. Astrophysical sources are likely to be the explanations of these excesses.

Diffuse continuum gammas could also give a signature due to their isotropic halo origin. The excess of GeV gamma-rays observed by EGRET [42] and attributed to a possible WIMP signal could however be due to classical sources.

Antiprotons arise as another WIMP annihilation product in the halo. The signal is expected to be detectable above background only at very low energies. The BESS balloon-borne experiment indeed observed antiprotons below 1 GeV [45]. However, the uncertainties in the calculation of the expected signal and background energy spectra are too large to reach a firm conclusion. Precision measurements by the future experiments BESS, AMS2, and PAMELA may allow one to disentangle signal and background [40].

A cosmic-ray positron flux excess at around 8 GeV measured by HEAT [46] has given rise to numerous calculations and conjectures concerning a possible SUSY interpretation. The need for an ad-hoc "boost" of expected flux to match the observed one and the failure to reproduce the energy shape by including a component from WIMP annihilation are illustrative of the difficulty to assign a Dark Matter origin to such measurements.

Last but not least, an antideuteron signal [47], as potentially observable by AMS2 or PAMELA, could constitute a signal for WIMP annihilation in the halo.

An interesting comparison of respective sensitivities to MSSM parameter space of future direct and various indirect searches has been performed with the DARKSUSY tool [48]. Searching for neutrinos from the Sun tests the spin-dependent WIMP couplings to matter, whereas direct searches are mostly sensitive to spin-independent couplings. Moreover, γ line searches are sensitive to higgsino-like neutralinos, whereas direct detection methods are more sensitive to gaugino-like neutralinos. However, it should be kept in mind that interpretations of excess in the halo or in the galactic center as being

due to signals for WIMP annihilation strongly depend on details of the halo model.

Numerous new experiments are in line to bring accurate measurements to constrain or discover Dark Matter.

References:

1. For a brief but delightful history of DM, see V. Trimble, in Proceedings of the *First International Symposium on Sources of Dark Matter in the Universe*, Bel Air, California, 1994; published by World Scientific, Singapore (ed. D.B. Cline), see also a recent review : G. Bertone, D. Hooper and J. Silk,.
2. See *Global cosmological parameters* in this *Review*.
3. B. Paczynski, Astrophys. J. **304**, 1 (1986);
   K. Griest, Astrophys. J. **366**, 412 (1991).
4. F. De Paolis *et al.*, Phys. Rev. Lett. **74**, 14 (1995).
5. M. Kamionkowski and A. Kinkhabwala, Phys. Rev. **D57**, 3256 (1998).
6. See *e.g.*, J.R. Primack, in the Proceedings of *Midrasha Mathematicae in Jerusalem: Winter School in Dynamical Systems*, Jerusalem, Israel, January 1997, `astro-ph/9707285`. There is currently some debate whether cold DM models correctly reproduce the DM density profile near the center of galactic haloes. See *e.g.*, R.A. Swaters *et al.*, Astrophys. J. **583**, 732 (2003).
7. B.J. Carr and S.W. Hawking, MNRAS **168**, 399 (1974).
8. See *Axions and other Very Light Bosons* in this *Review*.
9. R.A. Battye and E.P.S. Shellard, Phys. Rev. Lett. **73**, 2954 (1994);
   Erratum-ibid. **76**, 2203 (1996).
10. E.W. Kolb and M.E. Turner, *The Early Universe*, Addison-Wesley (1990).
11. For a review, see G. Jungman, M. Kamionkowski, and K. Griest, Phys. Reports **267**, 195 (1996).
12. See *Searches for WIMPs and other Particles* in this *Review*.
13. S. Dimopoulos, G.F. Giudice, and A. Pomarol, Phys. Lett. **B389**, 37 (1996).
14. See *e.g.*, J.R. Ellis *et al.*, Nucl. Phys. **B652**, 259 (2003);
    J.R. Ellis *et al.*, Phys. Lett. **B565**, 176 (2003);
    H. Baer *et al.*, JHEP **0306**, 054 (2003);
    A. Bottino *et al.*, Phys. Rev. **D68**, 043506 (2003).
15. G. Griest and D. Seckel, Phys. Rev. **D43**, 3191 (1991).
16. T. Moroi and L. Randall, Nucl. Phys. **B570**, 455 (2000).
17. R. Allahverdi and M. Drees, Phys. Rev. Lett. **89**, 091302 (2002).
18. M. Fujii and T. Yanagida, Phys. Lett. **B542**, 80 (2002).
19. J. Hisano, K. Kohri, and M.M. Nojiri, Phys. Lett. **B505**, 169 (2001);
    X. Chen, M. Kamionkowski, and X. Zhang, Phys. Rev. **D64**, 021302 (2001).
20. M. Bolz, W. Buchmüller, and M. Plümacher, Phys. Lett. **B443**, 209 (1998).
21. L. Covi *et al.*, JHEP **0105**, 033 (2001).
22. MACHO Collab., C. Alcock *et al.*, Astrophys. J. **542**, 257 (2000);
    EROS Collab., AA 355, 39 (2000);
    OGLE Collab., AA 343, 10 (1999).
23. P. Sikivie, Phys. Rev. Lett. **51**, 1415 (1983), Erratum-ibid. **52**, 695 (1984).
24. H. Peng *et al.*, Nucl. Instrum. Methods **A444**, 569 (2000).

25. S. Asztalos *et al.*, Phys. Rev. **D69**, 011101 (2004).
26. M. Tada *et al.*, `physics/0101028` .
27. K. Yamamoto *et al.*, in *Heidelberg 2000, Dark matter in astro— and particle—physics*, `hep-ph/0101200` .
28. P.F. Smith and J.D. Lewin, Phys. Reports **187**, 203 (1990);
J.R. Primack, D. Seckel, and B. Sadoulet, Ann. Rev. Nucl. Part. Sci. **38** 751 (1988).
29. For recent reviews on non cryogenic detectors, see *e.g.*, A. Morales, Nucl. Phys. (Proc. Suppl.) **B138** 135 (2005).;
Proceedings of Topics in Astroparticles and Underground Physics TAUP 2003, Nucl. Phys B (Proc. Suppl.) vol. 138 (2005);
Proceedings of *Identification of Dark Matter*. IDM 2002, World Scientific, ed. N. Spooner and V. Kudryavtsev, (York, UK, 2002).
30. DAMA Collab., R. Bernabei *et al.*, Phys. Lett. **B480**, 23 (2000), and Riv. Nuovo Cimento **26**, 1 (2003).
31. CDMS Collab., D.S. Akerib *et al.*, Phys. Rev. Lett. **93**, 211301 (2004).
32. EDELWEISS Collab., A. Benoit *et al.*, Phys. Rev. **D71**, 122002 (2005).
33. C.J. Copi and L.M. Krauss, Phys. Rev. **67**, 103507 (2003); G. Gelmini and P. Gondolo, Phys. Rev. **D71**, 123520 (2005);
A. Kurylov and M. Kamionkowski, Phys. Rev. **D69**, 063503 (2004);
C.J. Copi and L.M. Krauss, New Astron. Rev. **49**, 185 (2005).
34. For a general introduction to SUSY, see the section devoted in this *Review of Particle Physics*. For a review on cross sections for direct detection, see J. Ellis *et al.*, Phys. Rev. **D67**, 123502 (2003) and for a recent update Phys. Rev. **D71**, 095007 (2005).
35. C. Savage, P. Gondolo and K. Freese, Phys. Rev. **D70**, 123513 (2004).
36. For a recent review on cryogenic detectors, see *e.g.*, W. Seidel, Nucl. Phys. (Proc. Suppl.) **B138** 130 (2005). In addition to the TAUP and IDM Conference Proceedings, see also the Proceedings of *Int. Workshop on Low Temperature Detectors*, LTD10, NIM A,(2003).
37. UKDMC collab, G.J. Alner *et al.*, Astropart. Phys. **23** 444 (2005).
38. Workshop on large TPC for low energy rare event detection, Paris, December 2004, `http://www.unine.ch/phys/tpc.html` .
39. These sites gather informations on neutrons from various underground labs:
`http://www.physics.ucla.edu/wimps/nBG/nBG.html`;
`http://ilias-darkmatter.uni-tuebingen.de/BSNS_WG.html`.
40. L. Bergstrom, Rept. on Prog. in Phys. **63**, 793 (2000);
L. Bergstrom *et al.*, Phys. Rev. **D59**, 043506 (1999);
C. Tao, Phys. Scripta **T93**, 82 (2001);
Y.Mambrini and C. Muoz, Journ. of Cosm. And Astrop. Phys., **10**, 3(2004).
41. Super K. collaboration, S. Desai *et al.*, Phys. Rev. **D70**, 109901 (2004).
42. EGRET Collab., D. Dixon *et al.*, New Astron. **3**, 539 (1998).
43. D. Horns, Phys. Lett. **B607**, 225 (2005).
44. C. Bohem *et al.*, Phys. Rev. Lett. **92**, 101301 (2004).
45. BESS Collab., S. Orito *et al.*, Phys. Rev. Lett. **84**, 1078 (2000).

46. HEAT Collab., S. W. Barwick *et al.*, Astrophys. J. **482**, L191 (2000).
47. F. Donato, N. Fornengo and P. Salati, Phys. Rev. **D62**, 043003 (2000).
48. DARKSUSY site: `http://www.physto.se/ edsjo/darksusy/`.

23. COSMIC MICROWAVE BACKGROUND

Revised April 2006 by D. Scott (University of British Columbia) and G.F. Smoot (UCB/LBNL).

23.1. Introduction

The energy content in radiation from beyond our Galaxy is dominated by the Cosmic Microwave Background (CMB), discovered in 1965 [1]. The spectrum of the CMB is well described by a blackbody function with $T = 2.725\,\mathrm{K}$. This spectral form is one of the main pillars of the hot Big Bang model for the early Universe. The lack of any observed deviations from a blackbody spectrum constrains physical processes over cosmic history at redshifts $z \lesssim 10^7$ (see previous versions of this mini-review). However, at the moment, all viable cosmological models predict a very nearly Planckian spectrum, and so are not stringently limited.

Another observable quantity inherent in the CMB is the variation in temperature (or intensity) from one part of the microwave sky to another [2]. Since the first detection of these anisotropies by the *COBE* satellite [3], there has been intense activity to map the sky at increasing levels of sensitivity and angular resolution by ground-based and balloon-borne measurements. These were joined in 2003 by the first results from NASA's Wilkinson Microwave Anisotropy Probe (*WMAP*) [4], which were improved upon by analysis of the 3 year *WMAP* data [5]. These observations have led to a stunning confirmation of the 'Standard Model of Cosmology.' In combination with other astrophysical data, the CMB anisotropy measurements place quite precise constraints on a number of cosmological parameters, and have launched us into an era of precision cosmology.

23.2. Description of CMB Anisotropies

Observations show that the CMB contains anisotropies at the 10^{-5} level, over a wide range of angular scales. These anisotropies are usually expressed by using a spherical harmonic expansion of the CMB sky:

$$T(\theta,\phi) = \sum_{\ell m} a_{\ell m} Y_{\ell m}(\theta,\phi).$$

The vast majority of the cosmological information is contained in the temperature 2 point function, *i.e.*, the variance as a function of separation θ. Equivalently, the power per unit $\ln \ell$ is $\ell \sum_m |a_{\ell m}|^2 / 4\pi$.

23.2.1. *The Monopole* :

The CMB has a mean temperature of $T_\gamma = 2.725 \pm 0.001\,\mathrm{K}$ (1σ) [6], which can be considered as the monopole component of CMB maps, a_{00}. Since all mapping experiments involve difference measurements, they are insensitive to this average level. Monopole measurements can only be made with absolute temperature devices, such as the FIRAS instrument on the *COBE* satellite [6]. Such measurements of the spectrum are consistent with a blackbody distribution over more than three decades in frequency. A blackbody of the measured temperature corresponds to $n_\gamma = (2\zeta(3)/\pi^2)\,T_\gamma^3 \simeq 411\,\mathrm{cm}^{-3}$ and $\rho_\gamma = (\pi^2/15)\,T_\gamma^4 \simeq 4.64 \times 10^{-34}\,\mathrm{g\,cm}^{-3} \simeq 0.260\,\mathrm{eV\,cm}^{-3}$.

23.2.2. *The Dipole* :

The largest anisotropy is in the $\ell = 1$ (dipole) first spherical harmonic, with amplitude $3.358 \pm 0.017\,\mathrm{mK}$ [7]. The dipole is interpreted to be the result of the Doppler shift caused by the solar system motion relative to the nearly isotropic blackbody field, as confirmed by measurements of the radial velocities of local galaxies [8]. The motion of an observer with velocity $\beta = v/c$ relative to an isotropic Planckian radiation field of temperature T_0 produces a Doppler-shifted temperature pattern

$$\begin{aligned} T(\theta) &= T_0(1-\beta^2)^{1/2}/(1-\beta\cos\theta) \\ &\approx T_0\,\left(1+\beta\cos\theta+(\beta^2/2)\cos 2\theta + O(\beta^3)\right). \end{aligned}$$

At every point in the sky one observes a blackbody spectrum, with temperature $T(\theta)$. The spectrum of the dipole is the differential of a blackbody spectrum, as confirmed by Ref. 9.

The implied velocity [10,7] for the solar system barycenter is $v = 369 \pm 2\,\mathrm{km\,s}^{-1}$, assuming a value $T_0 = T_\gamma$, towards $(\ell, b) = (263.86^\circ \pm 0.04^\circ, 48.24^\circ \pm 0.10^\circ)$. Such a solar system motion implies a velocity for the Galaxy and the Local Group of galaxies relative to the CMB. The derived value is $v_{\mathrm{LG}} = 627 \pm 22\,\mathrm{km\,s}^{-1}$ towards $(\ell, b) = (276^\circ \pm 3^\circ, 30^\circ \pm 3^\circ)$, where most of the error comes from uncertainty in the velocity of the solar system relative to the Local Group.

The dipole is a frame dependent quantity, and one can thus determine the 'absolute rest frame' of the Universe as that in which the CMB dipole would be zero. Our velocity relative to the Local Group, as well as the velocity of the Earth around the Sun, and any velocity of the receiver relative to the Earth, is normally removed for the purposes of CMB anisotropy study.

23.2.3. *Higher-Order Multipoles* :

Excess variance in CMB maps at higher multipoles ($\ell \geq 2$) is interpreted as being the result of perturbations in the density of the early Universe, manifesting themselves at the epoch of the last scattering of the CMB photons. In the hot Big Bang picture, the expansion of the Universe cools the plasma so that by a redshift $z \simeq 1100$ (with little dependence on the details of the model) the hydrogen and helium nuclei can bind electrons into neutral atoms, a process usually referred to as recombination [11]. Before this epoch, the CMB photons are tightly coupled to the baryons, while afterwards they can freely stream towards us.

Theoretical models generally predict that the $a_{\ell m}$ modes are Gaussian random fields to high precision, e.g. standard slow-roll inflation's non-Gaussion contribution is expected to be one or two orders of magnitude below current observational limits [12]. Although non-Gaussianity of various forms is possible in early Universe models, the signatures found in existing *WMAP* data are generally considered to be subtle foreground or instrumental artefacts [13,14]. Tests show that Gaussianity is an extremely good simplifying approximation [15,16], with only some relatively weak indications of non-Gaussianity or statistical anisotropy at large scales.

With the assumption of Gaussian statistics, and if there is no preferred axis, then it is the variance of the temperature field which carries the cosmological information, rather than the values of the individual $a_{\ell m}$s; in other words the power spectrum in ℓ fully characterizes the anisotropies. The power at each ℓ is $(2\ell+1)C_\ell/(4\pi)$, where $C_\ell \equiv \langle |a_{\ell m}|^2 \rangle$, and a statistically isotropic sky means that all ms are equivalent. We use our estimators of the C_ℓs to constrain their expectation values, which are the quantities predicted by a theoretical model. For an idealized full-sky observation, the variance of each measured C_ℓ (*i.e.*, the variance of the variance) is $[2/(2\ell+1)]C_\ell^2$. This sampling uncertainty (known as 'cosmic variance') comes about because each C_ℓ is χ^2 distributed with $(2\ell+1)$ degrees of freedom for our observable volume of the Universe. For fractional sky coverage, $f_{\rm sky}$, this variance is increased by $1/f_{\rm sky}$ and the modes become partially correlated.

It is important to understand that theories predict the expectation value of the power spectrum, whereas our sky is a single realization. Hence the cosmic variance is an unavoidable source of uncertainty when constraining models; it dominates the scatter at lower ℓs, while the effects of instrumental noise and resolution dominate at higher ℓs [17].

23.2.4. *Angular Resolution and Binning* :

There is no one-to-one conversion between multipole ℓ and the angle subtended by a particular wavevector projected onto the sky. However, a single spherical harmonic $Y_{\ell m}$ corresponds to angular variations of $\theta \sim \pi/\ell$. CMB maps contain anisotropy information from the size of the map (or in practice some fraction of that size) down to the beam-size of the instrument, σ. One can think of the effect of a Gaussian beam as rolling off the power spectrum with the function $\mathrm{e}^{-\ell(\ell+1)\sigma^2}$.

For less than full sky coverage, the ℓ modes become correlated. Hence, experimental results are usually quoted as a series of 'band powers', defined as estimators of $\ell(\ell+1)C_\ell/2\pi$ over different ranges of ℓ. Because of the strong foreground signals in the Galactic Plane, even 'all-sky' surveys, such as *COBE* and *WMAP* involve a cut sky. The amount of binning required to obtain uncorrelated estimates of power also depends on the map size.

23.3. Cosmological Parameters

The current 'Standard Model' of cosmology contains around 10 free parameters (see The Cosmological Parameters—Sec. 21 of this *Review*). The basic framework is the Friedmann-Robertson-Walker metric (*i.e.*, a universe that is approximately homogeneous and isotropic on large scales), with density perturbations laid down at early times and evolving into today's structures (see Big-Bang cosmology—Sec. 19 of this *Review*). These perturbations can be either 'adiabatic' (meaning that there is no change to the entropy per particle for each species, *i.e.*, $\delta\rho/\rho$ for matter is $(3/4)\delta\rho/\rho$ for radiation) or 'isocurvature' (meaning that, for example, matter perturbations compensate radiation perturbations so that the total energy density remains unperturbed, *i.e.*, $\delta\rho$ for matter is $-\delta\rho$ for radiation). These different modes give rise to distinct phases during growth, with those of the adiabatic scenario being strongly preferred by the data. Models that generate mainly isocurvature type perturbations (such as most topological defect scenarios) are no longer considered to be viable.

Within the adiabatic family of models, there is, in principle, a free function describing how the comoving curvature perturbations, $\mathcal{R}$, vary with scale. In inflationary models [18], the Taylor series expansion of $\ln\mathcal{R}(\ln k)$ has terms of steadily decreasing size. For the simplest models, there are thus 2 parameters describing the initial conditions for density perturbations: the amplitude and slope of the power spectrum, $\left\langle|\mathcal{R}|^2\right\rangle \propto k^n$. This can be explicitly defined, for example, through:

$$\Delta^2_{\mathcal{R}} \equiv (k^3/2\pi^2)\left\langle|\mathcal{R}|^2\right\rangle,$$

and using $A \equiv \Delta^2_{\mathcal{R}}(k_0)$ with $k_0 = 0.002\,\mathrm{Mpc}^{-1}$, say. There are many other equally valid definitions of the amplitude parameter (see also Sec. 19 and Sec. 21 of this *Review*), and we caution that the relationships between some of them can be cosmology dependent. In 'slow roll' inflationary models this normalization is proportional to the combination $V^3/(V')^2$, for the inflationary potential $V(\phi)$. The slope n also involves V'', and so the combination of A and n can, in principle, constrain potentials.

Inflation generates tensor (gravity wave) modes as well as scalar (density perturbation) modes. This fact introduces another parameter, measuring the amplitude of a possible tensor component, or equivalently the ratio of the tensor to scalar contributions. The tensor amplitude $A_{\mathrm{T}} \propto V$, and thus one expects a larger gravity wave contribution in models where inflation happens at higher energies. The tensor power spectrum also has a slope, often denoted n_{T}, but since this seems likely to be extremely hard to measure, it is sufficient for now to focus only on the amplitude of the gravity wave component. It is most common to define the tensor contribution through r, the ratio of tensor to scalar perturbation spectra at large scales (say $k = 0.002\,\mathrm{Mpc}^{-1}$); however, there are other definitions, for example in terms of the ratio of contributions to C_2. Different inflationary potentials will lead to different predictions, e.g. for $\lambda\phi^4$ inflation with 50 e-folds, $r = 0.32$, while other models can have arbitrarily small values of r. In any case, whatever the specific definition, and whether they come from inflation or something else, the 'initial conditions' give rise to a minimum of 3 parameters: A, n, and r.

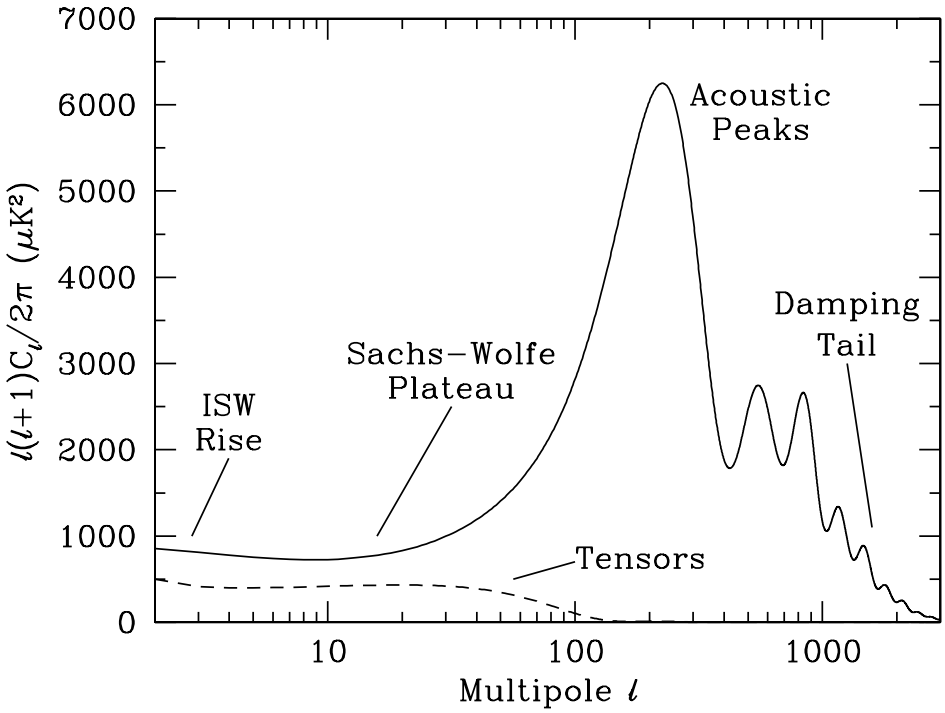


Figure 23.1: The theoretical CMB anisotropy power spectrum, using a standard ΛCDM model from `CMBFAST`. The x-axis is logarithmic here. The regions, each covering roughly a decade in ℓ, are labeled as in the text: the ISW Rise; Sachs-Wolfe Plateau; Acoustic Peaks; and Damping Tail. Also shown is the shape of the tensor (gravity wave) contribution, with an arbitrary normalization.

The background cosmology requires an expansion parameter (the Hubble Constant, H_0, often represented through $H_0 = 100\,h\,\mathrm{km\,s^{-1}Mpc^{-1}}$) and several parameters to describe the matter and energy content of the Universe. These are usually given in terms of the critical density, *i.e.*, for species 'x', $\Omega_\mathrm{x} = \rho_\mathrm{x}/\rho_\mathrm{crit}$, where $\rho_\mathrm{crit} = 3H_0^2/8\pi G$. Since physical densities $\rho_\mathrm{x} \propto \Omega_\mathrm{x}h^2 \equiv \omega_\mathrm{x}$ are what govern the physics of the CMB anisotropies, it is these ωs that are best constrained by CMB data. In particular CMB observations constrain $\Omega_\mathrm{B}h^2$ for baryons and $\Omega_\mathrm{M}h^2$ for baryons plus Cold Dark Matter.

The contribution of a cosmological constant Λ (or other form of Dark Energy) is usually included through a parameter which quantifies the curvature, $\Omega_\mathrm{K} \equiv 1 - \Omega_\mathrm{tot}$, where $\Omega_\mathrm{tot} = \Omega_\mathrm{M} + \Omega_\Lambda$. The radiation content, while in principle a free parameter, is precisely enough determined by the measurement of T_γ, and makes a negligible contribution to Ω_tot today.

The main effect of astrophysical processes on the C_ℓs comes through reionization. The Universe became reionized at some redshift z_i, long after recombination, affecting the CMB through the integrated Thomson scattering optical depth:

$$\tau = \int_0^{z_\mathrm{i}} \sigma_\mathrm{T} n_\mathrm{e}(z) \frac{dt}{dz} dz,$$

where $\sigma_{\rm T}$ is the Thomson cross-section, $n_{\rm e}(z)$ is the number density of free electrons (which depends on astrophysics) and dt/dz is fixed by the background cosmology. In principle, τ can be determined from the small scale matter power spectrum together with the physics of structure formation and feedback processes. However, this is a sufficiently complicated calculation that τ needs to be considered as a free parameter.

Thus we have 8 basic cosmological parameters: A, n, r, h, $\Omega_{\rm B}h^2$, $\Omega_{\rm M}h^2$, $\Omega_{\rm tot}$, and τ. One can add additional parameters to this list, particularly when using the CMB in combination with other data sets. The next most relevant ones might be: $\Omega_\nu h^2$, the massive neutrino contribution; w ($\equiv p/\rho$), the equation of state parameter for the Dark Energy; and $dn/d\ln k$, measuring deviations from a constant spectral index. To these 11 one could of course add further parameters describing additional physics, such as details of the reionization process, features in the initial power spectrum, a sub-dominant contribution of isocurvature modes, *etc.*

As well as these underlying parameters, there are other quantities that can be obtained from them. Such derived parameters include the actual Ωs of the various components (*e.g.*, $\Omega_{\rm M}$), the variance of density perturbations at particular scales (*e.g.*, σ_8), the age of the Universe today (t_0), the age of the Universe at recombination, reionization, *etc.*

23.4. Physics of Anisotropies

The cosmological parameters affect the anisotropies through the well understood physics of the evolution of linear perturbations within a background FRW cosmology. There are very effective, fast, and publicly-available software codes for computing the CMB anisotropy, polarization, and matter power spectra, *e.g.*, `CMBFAST` [19] and `CAMB` [20]. `CMBFAST` is the most extensively used code; it has been tested over a wide range of cosmological parameters and is considered to be accurate to better than the 1% level [21].

A description of the physics underlying the C_ℓs can be separated into 3 main regions, as shown in Fig. 23.1.

23.4.1. *The Sachs-Wolfe plateau:* $\ell \lesssim 100$:

The horizon scale (or more precisely, the angle subtended by the Hubble radius) at last scattering corresponds to $\ell \simeq 100$. Anisotropies at larger scales have not evolved significantly, and hence directly reflect the 'initial conditions.' The combination of gravitational redshift and intrinsic temperature fluctuations leads to $\delta T/T \simeq (1/3)\delta\phi/c^2$, where $\delta\phi$ is the perturbation to the gravitational potential. This is usually referred to as the 'Sachs-Wolfe' effect [22].

Assuming that a nearly scale-invariant spectrum of density perturbations was laid down at early times (*i.e.*, $n \simeq 1$, meaning equal power per decade in k), then $\ell(\ell+1)C_\ell \simeq$ constant at low ℓs. This effect is hard to see unless the multipole axis is plotted logarithmically (as in Fig. 23.1, but not Fig. 23.2).

Time variation of the potentials (*i.e.*, time-dependent metric perturbations) leads to an upturn in the C_ℓs in the lowest several multipoles; any deviation from a total equation of state $w = 0$ has such an effect. So the dominance of the Dark Energy at low redshift makes the lowest ℓs rise above the plateau. This is sometimes called the

'integrated Sachs-Wolfe effect' (or ISW Rise), since it comes from the line integral of $\dot{\phi}$; it has been confirmed through correlations between the large-angle anisotropies and large-scale structure [23]. Specific models can also give additional contributions at low ℓ (*e.g.*, perturbations in the Dark Energy component itself [24]) , but typically these are buried in the cosmic variance.

In principle, the mechanism that produces primordial perturbations could generate scalar, vector, and tensor modes. However, the vector (vorticity) modes decay with the expansion of the Universe. Tensors also decay when they enter the horizon, and so they contribute only to angular scales above about 1° (see Fig. 23.1). Hence some fraction of the low ℓ signal could be due to a gravity wave contribution, although small amounts of tensors are essentially impossible to discriminate from other effects that might raise the level of the plateau. However, the tensors *can* be distinguished using polarization information (see Sec. 23.6).

23.4.2. *The acoustic peaks:* $100 \lesssim \ell \lesssim 1000$:

On sub-degree scales, the rich structure in the anisotropy spectrum is the consequence of gravity-driven acoustic oscillations occurring before the atoms in the Universe became neutral. Perturbations inside the horizon at last scattering have been able to evolve causally and produce anisotropy at the last scattering epoch, which reflects that evolution. The frozen-in phases of these sound waves imprint a dependence on the cosmological parameters, which gives CMB anisotropies their great constraining power.

The underlying physics can be understood as follows. Before the Universe became neutral the proton-electron plasma was tightly coupled to the photons, and these components behaved as a single 'photon-baryon fluid'. Perturbations in the gravitational potential, dominated by the dark matter component, were steadily evolving. They drove oscillations in the photon-baryon fluid, with photon pressure providing most of the restoring force and baryons giving some additional inertia. The perturbations were quite small in amplitude, $\mathrm{O}(10^{-5})$, and so evolved linearly. That means each Fourier mode evolved independently and hence can be described by a driven harmonic oscillator, with frequency determined by the sound speed in the fluid. Thus the fluid density oscillated, which gives time variations in temperature, together with a velocity effect which is $\pi/2$ out of phase and has its amplitude reduced by the sound speed.

After the Universe recombined the baryons and radiation decoupled, and the radiation could travel freely towards us. At that point the phases of the oscillations were frozen-in, and projected on the sky as a harmonic series of peaks. The main peak is the mode that went through 1/4 of a period, reaching maximal compression. The even peaks are maximal *under*-densities, which are generally of smaller amplitude because the rebound has to fight against the baryon inertia. The troughs, which do not extend to zero power, are partially filled by the Doppler effect because they are at the velocity maxima.

An additional ingredient comes from geometrical projection. The scale associated with the peaks is the sound horizon at last scattering, which can be straightforwardly calculated as a physical length scale. This length is projected onto the sky, leading to an angular scale that depends on the background cosmology. Hence the angular position of the peaks is a sensitive probe of the spatial curvature of the Universe (*i.e.*, $\Omega_{\rm tot}$), with the peaks lying at higher ℓ in open universes and lower ℓ in closed geometry.

One last effect arises from reionization at redshift $z_{\rm i}$. A fraction of photons (τ) will be isotropically scattered at $z < z_{\rm i}$, partially erasing the anisotropies at angular scales smaller than those subtended by the Hubble radius at $z_{\rm i}$, which corresponds typically to ℓs above about a few 10s, depending on the specific reionization model. The acoustic peaks are therefore reduced by a factor $e^{-2\tau}$ relative to the plateau.

These peaks were a clear theoretical prediction going back to about 1970 [25]. One can think of them as a snapshot of stochastic standing waves. Since the physics governing them is simple and their structure rich, then one can see how they encode extractable information about the cosmological parameters. Their empirical existence started to become clear around 1994 [26], and the emergence, over the following decade, of a coherent series of acoustic peaks and troughs is a triumph of modern cosmology. This picture has received further confirmation with the recent detection in the power spectrum of galaxies (at redshifts close to zero) of the imprint of the acoustic oscillations in the baryon component [27].

23.4.3. *The damping tail:* $\ell \gtrsim 1000$:

The recombination process is not instantaneous, giving a thickness to the last scattering surface. This leads to a damping of the anisotropies at the highest ℓs, corresponding to scales smaller than that subtended by this thickness. One can also think of the photon-baryon fluid as having imperfect coupling, so that there is diffusion between the two components, and hence the amplitudes of the oscillations decrease with time. These effects lead to a damping of the C_ℓs, sometimes called Silk damping [28], which cuts off the anisotropies at multipoles above about 2000.

An extra effect at high ℓs comes from gravitational lensing, caused mainly by non-linear structures at low redshift. The C_ℓs are convolved with a smoothing function in a calculable way, partially flattening the peaks, generating a power-law tail at the highest multipoles, and complicating the polarization signal [29]. This is an example of a 'secondary effect', *i.e.*, the processing of anisotropies due to relatively nearby structures (see Sec. 23.7.2). Galaxies and clusters of galaxies give several such effects; all are expected to be of low amplitude and typically affect only the highest ℓs, but they will be increasingly important as experiments push to higher sensitivity and angular resolution.

23.5. Current Anisotropy Data

There has been a steady improvement in the quality of CMB data that has led to the development of the present-day cosmological model. Probably the most robust constraints currently available come from the combination of the *WMAP* three year data [16] with smaller scale results from the ACBAR [30], BOOMERANG [31], CBI [32] and VSA [33] experiments (together with constraints from other cosmological data-sets). We plot power spectrum estimates from these five experiments in Fig. 23.2. Other recent experiments, such as ARCHEOPS [34], DASI [35] and MAXIMA [36] also give powerful constraints, which are quite consistent with what we describe below. There have been some comparisons among data-sets [37], which indicate very good agreement, both in maps and in derived power spectra (up to systematic uncertainties in the overall calibration for some experiments). This makes it clear that systematic effects are largely under control. However, a fully self-consistent joint analysis of all the current data sets

has not been attempted, one of the reasons being that it requires a careful treatment of the overlapping sky coverage.

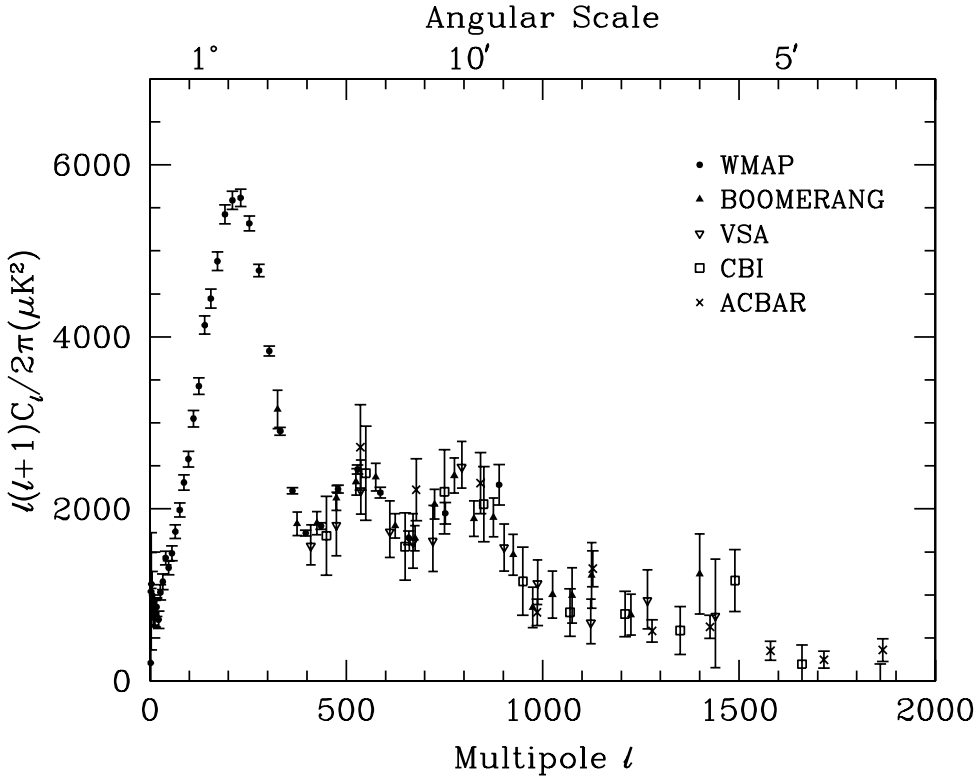


Figure 23.2: Band-power estimates from the *WMAP*, BOOMERANG, VSA, CBI, and ACBAR experiments. We have suppressed some of the low-ℓ and high-ℓ band-powers which have large error bars. Note also that the widths of the ℓ-bands varies between experiments. This plot represent only a selection of available experimental results, with some other data-sets being of similar quality. The multipole axis here is linear, so the Sachs-Wolfe plateau is hard to see. However, the acoustic peaks and damping region are very clearly observed, with no need for a theoretical curve to guide the eye.

Fig. 23.2 shows band-powers from the three year *WMAP* data [7], together with BOOMERANG [31], VSA [33], CBI [32] and ACBAR [30] data at higher ℓ. The points are in very good agreement with a 'ΛCDM' type model, as described earlier, with several of the peaks and troughs quite apparent. For details of how these estimates were arrived at, the strength of any correlations between band-powers and other information required to properly interpret them, turn to the original papers.

23.6. CMB Polarization

Since Thomson scattering of an anisotropic radiation field also generates linear polarization, the CMB is predicted to be polarized at the roughly 5% level [38]. Polarization is a spin 2 field on the sky, and the algebra of the modes in ℓ-space is strongly analogous to spin-orbit coupling in quantum mechanics [39]. The linear polarization pattern can be decomposed in a number of ways, with two quantities required for each pixel in a map, often given as the Q and U Stokes parameters. However, the most intuitive and physical decomposition is a geometrical one, splitting the polarization pattern into a part that comes from a divergence (often referred to as the 'E-mode') and a part with a curl (called the 'B-mode') [40]. More explicitly, the modes are defined in terms of second derivatives of the polarization amplitude, with the Hessian for the E-modes having principle axes in the same sense as the polarization, while the B-mode pattern can be thought of simply as a 45° rotation of the E-mode pattern. Globally one sees that the E-modes have $(-1)^\ell$ parity (like the spherical harmonics), while the B-modes have $(-1)^{\ell+1}$ parity.

The existence of this linear polarization allows for 6 different cross power spectra to be determined from data that measure the full temperature and polarization anisotropy information. Parity considerations make 2 of these zero, and we are left with 4 potential observables: $C_\ell^{\rm TT}$, $C_\ell^{\rm TE}$, $C_\ell^{\rm EE}$, and $C_\ell^{\rm BB}$. Since scalar perturbations have no handedness, the B-mode power spectrum can only be generated by vectors or tensors. Hence, in the context of inflationary models, the determination of a non-zero B-mode signal is a way to measure the gravity wave contribution (and thus potentially derive the energy scale of inflation), even if it is rather weak. However, one must first eliminate the foreground contributions and other systematic effects down to very low levels.

The oscillating photon-baryon fluid also results in a series of acoustic peaks in the polarization C_ℓs. The main 'EE' power spectrum has peaks that are out of phase with those in the 'TT' spectrum, because the polarization anisotropies are sourced by the fluid velocity. The 'TE' part of the polarization and temperature patterns comes from correlations between density and velocity perturbations on the last scattering surface, which can be both positive and negative, and is of larger amplitude than the EE signal. There is no polarization 'Sachs-Wolfe' effect, and hence no large-angle plateau. However, scattering during a recent period of reionization can create a polarization 'bump' at large angular scales.

The strongest upper limits on polarization are at the roughly $10\,\mu$K level from the POLAR [41] experiment at large angular scales and the PIQUE [42], COMPASS [43] and CBI [44] experiments at smaller scales. The first measurement of a polarization signal came in 2002 from the DASI experiment [45], which provided a convincing detection, confirming the general paradigm, but of low enough significance that it lent little constraint to models. As well as the E-mode signal, DASI also made a statistical detection of the TE correlation.

In 2003 the *WMAP* experiment demonstrated that is was able to measure the TE cross-correlation power spectrum with high precision [46], and this was improved upon in the 3 year results, which also included EE measurements [47]. Other recent experimental results include a weak detection of the EE signal from CAPMAP [48] and more significant detections from CBI [49], DASI [50] and BOOMERANG [51]. In addition the TE signal

has been detected in several multipole bands by BOOMERANG [52], and there are statistical detections by CBI [49] and DASI [50]. Some upper limits on $C_\ell^{\rm BB}$ also exist, but are currently not very constraining.

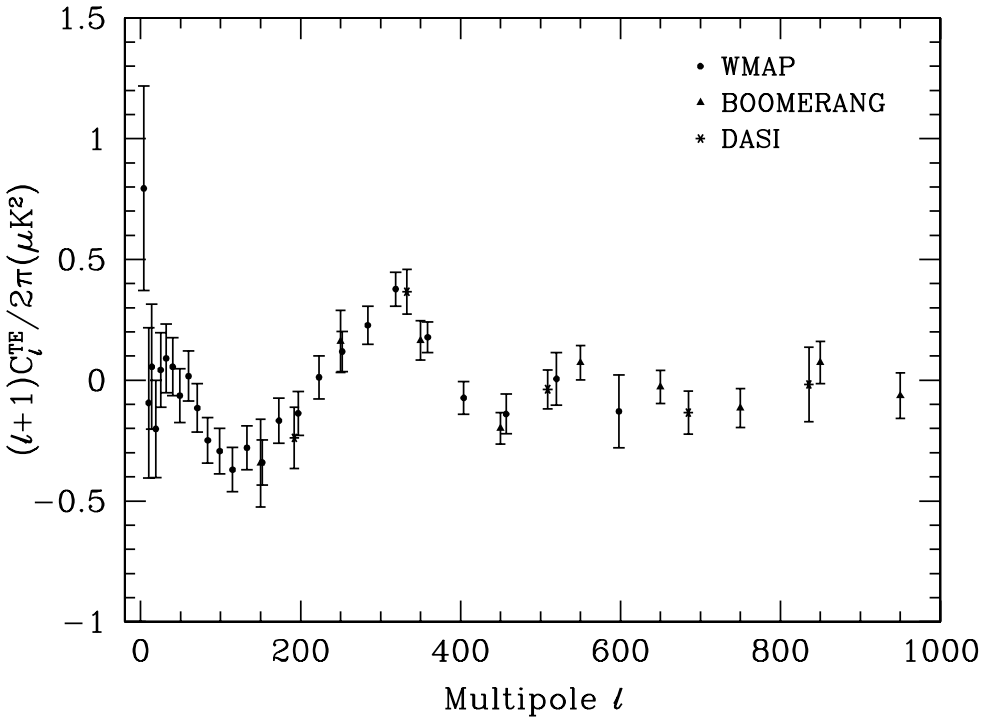


Figure 23.3: Cross power spectrum of the temperature anisotropies and E-mode polarization signal from *WMAP* [47], together with estimates from DASI and BOOMERANG which extend to higher ℓ. Note that the BOOMERANG bands are wider in ℓ than those of *WMAP*, while those of DASI are almost as wide as the features in the power spectrum. Also note that the y-axis here is not multiplied by the additional ℓ, which helps to show both the large and small angular scale features.

The results of *WMAP* $C_\ell^{\rm TE}$ [47] are shown in Fig. 23.3, along with estimates from the DASI and BOOMERANG experiments. The measured shape of the cross-correlation power spectrum provides supporting evidence of the adiabatic nature of the perturbations, as well as directly constraining the thickness of the last scattering surface. Since the polarization anisotropies are generated in this scattering surface, the existence of correlations at angles above about a degree demonstrates that there were super-Hubble fluctuations at the recombination epoch.

Perhaps the most intriguing result from the polarization measurements is at the largest angular scales ($\ell < 10$), where there is an excess signal compared to that expected from the temperature power spectrum alone. This is precisely the signal expected from an early period of reionization, arising from Doppler shifts during the partial scattering at $z < z_{\rm i}$. This signal is also confirmed in the *WMAP* $C_\ell^{\rm EE}$ results at $\ell = 2$–6. The

amplitude of the signal indicates that the first stars, presumably the source of the ionizing radiation, formed around $z \approx 10$ (somewhat lower than the value suggested by the first year *WMAP* results, although the uncertainty is still quite large).

23.7. Complications

There are a number of issues which complicate the interpretation of CMB anisotropy data, some of which we sketch out below.

23.7.1. *Foregrounds* :

The microwave sky contains significant emission from our Galaxy and from extra-galactic sources [53]. Fortunately, the frequency dependence of these various sources is in general substantially different from that of the CMB anisotropy signals. The combination of Galactic synchrotron, bremsstrahlung and dust emission reaches a minimum at a wavelength of roughly 3 mm (or about 100 GHz). As one moves to greater angular resolution, the minimum moves to slightly higher frequencies, but becomes more sensitive to unresolved (point-like) sources.

At frequencies around 100 GHz and for portions of the sky away from the Galactic Plane the foregrounds are typically 1 to 10% of the CMB anisotropies. By making observations at multiple frequencies, it is relatively straightforward to separate the various components and determine the CMB signal to the few per cent level. For greater sensitivity it is necessary to use the spatial information and statistical properties of the foregrounds to separate them from the CMB.

The foregrounds for CMB polarization are expected to follow a similar pattern, but are less well studied, and are intrinsically more complicated. The three year *WMAP* data have shown that the polarized foregrounds dominate at large angular scales, and that they must be well characterized in order to be discriminated [47]. Whether it is possible to achieve sufficient separation to detect B-mode CMB polarization is still an open question. However, for the time being, foreground contamination is not a 'show-stopper' for CMB experiments.

23.7.2. *Secondary Anisotropies* :

With increasingly precise measurements of the primary anisotropies, there is growing theoretical and experimental interest in 'secondary anisotropies.' Effects which happen at $z \ll 1000$ become more important as experiments push to higher angular resolution and sensitivity.

These secondary effects include gravitational lensing, patchy reionization and the Sunyaev-Zel'dovich (SZ) effect [54]. This is Compton scattering ($\gamma e \to \gamma' e'$) of the CMB photons by a hot electron gas, which creates spectral distortions by transferring energy from the electrons to the photons. The effect is particularly important for clusters of galaxies, through which one observes a partially Comptonized spectrum, resulting in a decrement at radio wavelengths and an increment in the submillimeter. This can be used to find and study individual clusters and to obtain estimates of the Hubble constant. There is also the potential to constrain the equation of state of the Dark Energy through counts of detected clusters as a function of redshift [55].

23.7.3. ***Higher-order Statistics*** :

Although most of the CMB anisotropy information is contained in the power spectra, there will also be weak signals present in higher-order statistics. These statistics can measure any primordial non-Gaussianity in the perturbations, as well as non-linear growth of the fluctuations on small scales and other secondary effects (plus residual foreground contamination). Although there are an infinite variety of ways in which the CMB could be non-Gaussian, there is a generic form to consider for the initial conditions, where a quadratic contribution to the curvature perturbations is parameterized through a dimensionless number $f_{\rm NL}$. This weakly non-linear component can be constrained through measurements of the bispectrum or Minkowski functionals for example, and the result from *WMAP* is $-54 < f_{\rm NL} < 114$ (95% confidence region) [16].

23.8. Constraints on Cosmologies

The clearest outcome of the newer experimental results is that the standard cosmological paradigm is in good shape. A large amount of high precision data on the power spectrum is adequately fit with fewer than 10 free parameters. The framework is that of Friedmann-Robertson-Walker models, which have nearly flat geometry, containing Dark Matter and Dark Energy, and with adiabatic perturbations having close to scale invariant initial conditions.

Within this framework, bounds can be placed on the values of the cosmological parameters. Of course, much more stringent constraints can be placed on models which cover a restricted number of parameters, e.g. assuming that $\Omega_{\rm tot} = 1$, $n = 1$ or $r = 0$. More generally, the constraints depend upon the adopted priors, even if they are implicit, for example by restricting the parameter freedom or the ranges of parameters (particularly where likelihoods peak near the boundaries), or by using different choices of other data in combination with the CMB. When the data become even more precise, these considerations will be less important, but for now we caution that restrictions on model space and choice of priors need to be kept in mind when adopting specific parameter values and uncertainties.

There are some combinations of parameters that fit the CMB anisotropies almost equivalently. For example, there is a nearly exact geometric degeneracy, where any combination of $\Omega_{\rm M}$ and Ω_Λ that gives the same angular diameter distance to last scattering will give nearly identical C_ℓs. There are also other less exact degeneracies among the parameters. Such degeneracies can be broken when using the CMB data in combination with other cosmological data sets. Particularly useful are complementary constraints from galaxy clustering, the abundance of galaxy clusters, weak gravitational lensing measurements, Type Ia supernova distances and the distribution of Lyman α forest clouds. For an overview of some of these other cosmological constraints, see The Cosmological Parameters—Sec. 21 of this *Review*.

The 3 year *WMAP* alone, together with weak priors (on h and $\Omega_{\rm B}h^2$ for example), and within the context of a 6 parameter family of models (which fixes $\Omega_{\rm tot} = 1$ and $r = 0$), yields the following results [16]: $A = 2.26^{+0.13}_{-0.16} \times 10^{-9}$, $n = 0.951^{+0.015}_{-0.019}$, $h = 0.73^{+0.03}_{-0.04}$, $\Omega_{\rm B}h^2 = 0.0223^{+0.0007}_{-0.0009}$, $\Omega_{\rm M}h^2 = 0.127^{+0.007}_{-0.010}$ and $\tau = 0.09 \pm 0.03$. The main changes of the 3 year data compared with the first year results are: a lowering of $\Omega_{\rm M}$ (from constraints

on the third acoustic peak); a tightening of the confidence interval and decrease in the estimate for τ (driven by the large-angle EE measurements); and a subsequent breaking of the degeneracy between A and τ, which leads to a lowering of A and some evidence (at the roughly 3σ level) for $n < 1$. The *WMAP* data alone therefore now seem to require a 6 parameter model space. Other combinations of data, e.g. including BOOMERANG and other CMB measurements, or using large-scale structure data or supernova constraints, lead to consistent results to those given above, sometimes with smaller error bars, and with the precise values depending on data selection [56,57,16]. Note that for h, the CMB data alone provide only a very weak constraint, unless spatial flatness or some other cosmological data are used. For $\Omega_{\rm B}h^2$ the precise value depends sensitively on how much freedom is allowed in the shape of the primordial power spectrum (see Big-Bang nucleosynthesis—Sec. 20 of this *Review*). The addition of other cosmological data-sets allows for constraints to be placed on further parameters.

The best constraint on $\Omega_{\rm tot}$ is $1.003^{+0.013}_{-0.017}$. This comes from including a constraint on h [14], although similar results come from using Supernova Legacy Survey data [59] or large-scale structure data.

The 95% confidence upper limit on r is 0.55 using *WMAP* alone, tightening to $r < 0.28$ with the addition of the Sloan Digital Sky Survey data for example [60]. This limits depends on how we restrict the slope n and whether we allow $dn/d\ln k \neq 0$.

There are also constraints on parameters over and above the basic 8 that we have described. But for such constraints it is necessary to include additional data in order to break the degeneracies. For example the addition of the Dark Energy equation of state, w adds the partial degeneracy of being able to fit a ridge in (w, h) space, extending to low values of both parameters. This degeneracy is broken when the CMB is used in combination with independent H_0 limits, or other data. *WMAP* plus supernova and large-scale structure data yield $w = -1.06^{+0.13}_{-0.08}$, with stronger constraints for flat models.

For the optical depth τ, the best-fit corresponds to a reionization redshift centered on 11 in the best-fit cosmology and assuming instantaneous reionization. This redshift is not much higher that that suggested from studies of absorption in high-z quasar spectra [61]. The excitement here is that we have direct information from CMB polarization which can be combined with other astrophysical measurements to understand when the first stars formed and brought about the end of the cosmic dark ages.

23.9. Particle Physics Constraints

CMB data are beginning to put limits on parameters which are directly relevant for particle physics models. For example there is a limit on the neutrino contribution $\Omega_\nu h^2 < 0.0072$ (95% confidence) from a combination of *WMAP*, galaxy clustering and supernovae data [16]. This directly implies a limit on neutrino mass, $\sum m_\nu < 0.68\,{\rm eV}$, assuming the usual number density of fermions which decoupled when they were relativistic.

A combination of the *WMAP* data with other data-sets is better fit with models which have a running spectral index, *i.e.*, $dn/d\ln k \neq 0$ [16]. However, the improvement is not significant at this time. The indication that $n < 1$, if borne out, would be quite constraining for inflationary models. Moreover, this gives a real target that the value of r in simple models may be in the range of detectability.

One other hint of new physics lies in the fact that the quadrupole and possibly some of the other low ℓ modes seem anomalously low compared with the best-fit ΛCDM model [62,7]. This is what might be expected in a universe which has a large scale cut-off to the power spectrum, or is topologically non-trivial. However, because of cosmic variance, possible foregrounds, apparent correlations between modes (as mentioned in Sec. 23.2), *etc.*, the significance of this feature is still a matter of debate [13,63].

In addition it is also possible to put limits on other pieces of physics [64], for example the neutrino chemical potentials, contribution of warm dark matter, decaying particles, time variation of the fine-structure constant, or physics beyond general relativity. Further particle physics constraints will follow as the anisotropy measurements increase in precision.

Careful measurement of the CMB power spectra and non-Gaussianity can in principle put constraints on physics at the highest energies, including ideas of string theory, extra dimensions, colliding branes, *etc.* At the moment any calculation of predictions appears to be far from definitive. However, there is a great deal of activity on implications of string theory for the early Universe, and hence a very real chance that there might be observational implications for specific scenarios.

23.10. Fundamental Lessons

More important than the precise values of parameters is what we have learned about the general features which describe our observable Universe. Beyond the basic hot Big Bang picture, the CMB has taught us that:

- The Universe recombined at $z \simeq 1100$ and started to become ionized again at $z \simeq 10$.
- The geometry of the Universe is close to flat.
- Both Dark Matter and Dark Energy are required.
- Gravitational instability is sufficient to grow all of the observed large structures in the Universe.
- Topological defects were not important for structure formation.
- There are 'synchronized' super-Hubble modes generated in the early Universe.
- The initial perturbations were adiabatic in nature.
- The perturbations had close to Gaussian (*i.e.*, maximally random) initial conditions.

It is very tempting to make an analogy between the status of the cosmological 'Standard Model' and that of particle physics. In cosmology there are about 10 free parameters, each of which is becoming well determined, and with a great deal of consistency between different measurements. However, none of these parameters can be calculated from a fundamental theory, and so hints of the bigger picture, 'physics beyond the Standard Model' are being searched for with ever more ambitious experiments.

Despite this analogy, there are some basic differences. For one thing, many of the cosmological parameters change with cosmic epoch, and so the measured values are simply the ones determined today, and hence they are not 'constants', like particle masses for example (although they *are* deterministic, so that if one knows their values at one epoch, they can be calculated at another). Moreover, the number of parameters is not as fixed as

it is in the particle physics Standard Model; different researchers will not necessarily agree on what the free parameters are, and new ones can be added as the quality of the data improves. In addition, parameters like τ, which come from astrophysics, are in principle calculable from known physical processes, although this is currently impractical. On top of all this, other parameters might be 'stochastic' in that they may be fixed only in our observable patch of the Universe or among certain vacuum states in the 'Landscape' [65].

In a more general sense the cosmological 'Standard Model' is much further from the underlying 'fundamental theory' which will ultimately provide the values of the parameters from first principles. Nevertheless, any genuinely complete 'theory of everything' must include an explanation for the values of these cosmological parameters as well as the parameters of the Standard Model of particle physics.

23.11. Future Directions

With all the observational progress in the CMB and the tying down of cosmological parameters, what can we anticipate for the future? Of course there will be a steady improvement in the precision and confidence with which we can determine the appropriate cosmological model and its parameters. We can anticipate that the addition of three more years of *WMAP* data (6 years total) will bring improvements from the increased statistical accuracy and from the more detailed treatment of calibration and systematic effects. Ground-based experiments operating at smaller angular scales will also improve over the next few years, providing significantly tighter constraints on the damping tail. In addition, the next CMB satellite mission, *Planck*, is scheduled for launch in 2008, and there are even more ambitious projects currently being discussed.

Despite the increasing improvement in the results, it is also true that the addition of the latest experiments has not significantly changed the cosmological model. It is therefore appropriate to ask: what should we expect to come from *Planck* and from other more grandiose future experiments, including those being discussed as part of the 'Beyond Einstein' initiative? *Planck* certainly has the advantage of high sensitivity and a full sky survey. A precise measurement of the third acoustic peak provides a good determination of the matter density; this can only be done by measurements which are accurate relative to the first two peaks (which themselves constrain the curvature and the baryon density). A detailed measurement of the damping tail region will also significantly improve the determination of n and any running of the slope. *Planck* should also be capable of measuring $C_\ell^{\rm EE}$ quite well, providing both a strong check on the cosmological Standard Model and extra constraints that will improve parameter estimation.

A set of cosmological parameters are now known to roughly 10% accuracy, and that may seem sufficient for many people. However, we should certainly demand more of measurements which describe *the entire observable Universe*! Hence a lot of activity in the coming years will continue to focus on determining those parameters with increasing precision. This necessarily includes testing for consistency among different predictions of the cosmological Standard Model, and searching for signals which might require additional physics.

A second area of focus will be the smaller scale anisotropies and 'secondary effects.' There is a great deal of information about structure formation at $z \ll 1000$ encoded in

the CMB sky. This may involve higher-order statistics as well as spectral signatures. Such investigations can also provide constraints on the Dark Energy equation of state, for example. *Planck*, as well as experiments aimed at the highest ℓs, should be able to make a lot of progress in this arena.

A third direction is increasingly sensitive searches for specific signatures of physics at the highest energies. The most promising of these may be the primordial gravitational wave signals in $C_\ell^{\rm BB}$, which could be a probe of the $\sim 10^{16}$ GeV energy range. Whether the amplitude of the effect coming from inflation will be detectable is unclear, but the prize makes the effort worthwhile, and the indications that $n \approx 0.95$ give some genuine optimism that r may be of order 0.1, and hence reachable.

Anisotropies in the CMB have proven to be the premier probe of cosmology and the early Universe. Theoretically the CMB involves well-understood physics in the linear regime, and is under very good calculational control. A substantial and improving set of observational data now exists. Systematics appear to be well understood and not a limiting factor. And so for the next few years we can expect an increasing amount of cosmological information to be gleaned from CMB anisotropies, with the prospect also of some genuine surprises.

References:

1. A.A. Penzias and R. Wilson, Astrophys. J. **142**, 419 (1965);
   R.H. Dicke *et al.*, Astrophys. J. **142**, 414 (1965).
2. M. White, D. Scott, and J. Silk, Ann. Rev. Astron. & Astrophys. **32**, 329 (1994);
   W. Hu and S. Dodelson, Ann. Rev. Astron. & Astrophys. **40**, 171 (2002).
3. G.F. Smoot *et al.*, Astrophys. J. **396**, L1 (1992).
4. C.L. Bennett *et al.*, Astrophys. J. Supp. **148**, 1 (2003).
5. N. Jarosik *et al.*, `astro-ph/0603452`.
6. J.C. Mather *et al.*, Astrophys. J. **512**, 511 (1999).
7. G. Hinshaw *et al.*, `astro-ph/0603451`.
8. S. Courteau *et al.*, Astrophys. J. **544**, 636 (2000).
9. D.J. Fixsen *et al.*, Astrophys. J. **420**, 445 (1994).
10. D.J. Fixsen *et al.*, Astrophys. J. **473**, 576 (1996);
    A. Kogut *et al.*, Astrophys. J. **419**, 1 (1993).
11. S. Seager, D.D. Sasselov, and D. Scott, Astrophys. J. Supp. **128**, 407 (2000).
12. N. Bartolo *et al.*, Phys. Rep. **402**, 103 (2004).
13. A. de Oliveira-Costa *et al.*, Phys. Rev. **D69**, 063516 (2004).
14. Eriksen H.K. *et al.*, Astrophys. J. **605**, 14 (2004);
    P. Vielva *et al.*, Astrophys. J. **609**, 22 (2004);
    D.L. Larson, B.D. Wandelt, Astrophys. J. **613**, L85 (2004);
    P. Bielewicz, K.M. Gorski, and A.J. Banday, Monthly Not. Royal Astron. Soc. **355**, 1283 (2004);
    S. Prunet *et al.*, Phys. Rev. **D71**, 083508 (2005);
    Eriksen H.K. *et al.*, Astrophys. J. **622**, 58 (2005).
15. E. Komatsu *et al.*, Astrophys. J. Supp. **148**, 119 (2003).
16. D.N. Spergel *et al.*, `astro-ph/0603449`.

17. L. Knox, Phys. Rev. **D52**, 4307 (1995).
18. A.R. Liddle and D.H. Lyth, Cosmological Inflation and Large-Scale Structure, Cambridge University Press (2000).
19. U. Seljak and M. Zaldarriaga, Astrophys. J. **469**, 437 (1996).
20. A. Lewis, A. Challinor, A. Lasenby, Astrophys. J. **538**, 473 (2000).
21. U. Seljak *et al.*, Phys. Rev. **D68**, 083507 (2003).
22. R.K. Sachs and A.M. Wolfe, Astrophys. J. **147**, 73 (1967).
23. M.R. Nolta *et al.*, Astrophys. J. **608**, 10 (2004);
S. Boughn and R. Crittenden, Nature, **427**, 45 (2004);
P. Fosalba, E. Gaztañaga, and F. Castander, Astrophys. J. **597**, L89 (2003);
N. Afshordi, Y.-S. Loh, and M.A. Strauss, Phys. Rev. **D69**, 083524 (2004);
N. Padmanabhan *et al.*, Phys. Rev. **D72**, 043525 (2005).
24. W. Hu and D.J. Eisenstein, Phys. Rev. **D59**, 083509 (1999);
W. Hu *et al.*, Phys. Rev. **D59**, 023512 (1999).
25. P.J.E. Peebles and J.T. Yu, Astrophys. J. **162**, 815 (1970);
R.A. Sunyaev and Ya.B. Zel'dovich, Astrophysics & Space Science **7**, 3 (1970).
26. D. Scott, J. Silk, and M. White, Science **268**, 829 (1995).
27. D.J. Eisenstein *et al.*, Astrophys. J. in press `astro-ph/0501171`;
S. Cole *et al.*, Monthly Not. Royal Astron. Soc. **362**, 505 (2005).
28. J. Silk, Astrophys. J. **151**, 459 (1968).
29. M. Zaldarriaga and U. Seljak, Phys. Rev. **D58**, 023003 (1998).
30. M.C. Runyan *et al.*, Astrophys. J. Supp. **149**, 265 (2003);
C.L. Kuo *et al.*, Astrophys. J. **600**, 32 (2004).
31. J.E. Ruhl *et al.*, Astrophys. J. **599**, 786 (2003);
S. Masi *et al.*, Astrophys. J. in press `astro-ph/0507509`;
W.C. Jones *et al.*, Astrophys. J. in press `astro-ph/0507494`.
32. T.J. Pearson *et al.*, Astrophys. J. **591**, 556 (2003);
A.C.S. Readhead *et al.*, Astrophys. J. **609**, 498 (2004).
33. P.F. Scott *et al.*, Monthly Not. Royal Astron. Soc. **341**, 1076 (2003);
Grainge K. *et al.*, Monthly Not. Royal Astron. Soc. **341**, L23 (2003);
C. Dickinson *et al.*, Monthly Not. Royal Astron. Soc. **353**, 732 (2004).
34. A. Benoit *et al.*, Astronomy & Astrophysics **399**, L19 (2003);
M. Tristram *et al.*, Astronomy & Astrophysics **436**, 785 (2005).
35. N.W. Halverson *et al.*, Astrophys. J. **568**, 38 (2002).
36. A.T. Lee *et al.*, Astrophys. J. **561**, L1 (2001).
37. M.E. Abroe *et al.*, Astrophys. J. **605**, 607 (2004);
N. Rajguru *et al.*, Monthly Not. Royal Astron. Soc. **363**, 1125.
38. W. Hu, M. White, New Astron. **2**, 323 (1997).
39. W. Hu, M. White, Phys. Rev. **D56**, 596 (1997).
40. M. Zaldarriaga and U. Seljak, Phys. Rev. **D55**, 1830 (1997);
M. Kamionkowski, A. Kosowsky, and A. Stebbins, Phys. Rev. **D55**, 7368 (1997).
41. B.G. Keating *et al.*, Astrophys. J. **560**, L1 (2001).
42. M.M. Hedman *et al.*, Astrophys. J. **548**, L111 (2001).
43. P.C. Farese *et al.*, Astrophys. J. **610**, 625 (2004).

44. J.K. Cartwright *et al.*, Astrophys. J. **623**, 11 (2005).
45. J. Kovac *et al.*, Nature, 420, 772 (2002).
46. A. Kogut *et al.*, Astrophys. J. Supp. **148**, 161 (2003).
47. L. Page *et al.*, `astro-ph/0603450`.
48. D. Barkats *et al.*, Astrophys. J. **619**, L127 (2005).
49. A.C.S. Readhead *et al.*, Science, **306**, 836 (2004).
50. E.M. Leitch *et al.*, Astrophys. J. **624**, 10 (2005).
51. T.E. Montroy *et al.*, Astrophys. J. in press `astro-ph/0507514`.
52. F. Piacentini *et al.*Astrophys. J. in press `astro-ph/0507507`.
53. A. de Oliveira-Costa, M. Tegmark (eds.), Microwave Foregrounds, Astron. Soc. of the Pacific, San Francisco (1999).
54. R.A. Sunyaev and Ya.B. Zel'dovich, Ann. Rev. Astron. Astrophys. **18**, 537 (1980); M. Birkinshaw, Phys. Rep. **310**, 98 (1999).
55. J.E. Carlstrom, G.P. Holder, and E.D. Reese, Ann. Rev. Astron. & Astrophys. **40**, 643 (2002).
56. M. Tegmark *et al.*, Phys. Rev. **D69**, 103501 (2004);
    C.J. MacTavish *et al.*, Astrophys. J. in press `astro-ph/0507503`;
    A.G. Sánchez *et al.*, Monthly Not. Royal Astron. Soc. **366**, 189 (2005);
    M. Viel, M.G. Haehnelt, A. Lewis, `astro-ph/0604310`;
    U. Seljak, A. Slosar, P. McDonald, `astro-ph/0604335`.
57. D.N. Spergel *et al.*, Astrophys. J. **148**, 175 (2003).
58. W.L. Freedman *et al.*, Astrophys. J. **553**, 47 (2001).
59. P. Astier *et al.*, Astronomy & Astrophysics **447**, 31 (2006).
60. M. Tegmark *et al.*, Astrophys. J. **606**, 702 (2004).
61. X. Fan *et al.*, Astrophys. J. **123**, 1247 (2002).
62. G. Hinshaw *et al.*, Astrophys. J. Supp. **148**, 135 (2003).
63. G. Efstathiou, Monthly Notices of the Royal Astronomical Society, **346**, L26 (2003).
64. M. Kamionkowski and A. Kosowsky, Ann. Rev. Nucl. Part. Sci. **49**, 77 (1999).
65. R. Bousso and J. Polchinski, JHEP 0006, 006 (2000);
    L. Susskind, The Davis Meeting On Cosmic Inflation (2003), `hep-th/0302219`.

24. COSMIC RAYS

Revised March 2002 by T.K. Gaisser and T. Stanev (Bartol Research Inst., Univ. of Delaware); revised September 2005 by P.V. Sokolsky (Univ. of Utah) and R.E. Streitmatter (NASA)

24.1. Primary spectra

The cosmic radiation incident at the top of the terrestrial atmosphere includes all stable charged particles and nuclei with lifetimes of order 10^6 years or longer. Technically, "primary" cosmic rays are those particles accelerated at astrophysical sources and "secondaries" are those particles produced in interaction of the primaries with interstellar gas. Thus electrons, protons and helium, as well as carbon, oxygen, iron, and other nuclei synthesized in stars, are primaries. Nuclei such as lithium, beryllium, and boron (which are not abundant end-products of stellar nucleosynthesis) are secondaries. Antiprotons and positrons are also in large part secondary. Whether a small fraction of these particles may be primary is a question of current interest.

Apart from particles associated with solar flares, the cosmic radiation comes from outside the solar system. The incoming charged particles are "modulated" by the solar wind, the expanding magnetized plasma generated by the Sun, which decelerates and partially excludes the lower energy galactic cosmic rays from the inner solar system. There is a significant anticorrelation between solar activity (which has an alternating eleven-year cycle) and the intensity of the cosmic rays with energies below about 10 GeV. In addition, the lower-energy cosmic rays are affected by the geomagnetic field, which they must penetrate to reach the top of the atmosphere. Thus the intensity of any component of the cosmic radiation in the GeV range depends both on the location and time.

There are four different ways to describe the spectra of the components of the cosmic radiation: (1) By particles per unit rigidity. Propagation (and probably also acceleration) through cosmic magnetic fields depends on gyroradius or *magnetic rigidity,* R, which is gyroradius multiplied by the magnetic field strength:

$$R = \frac{p\,c}{Z\,e} = r_{_L} B\ . \tag{24.1}$$

(2) By particles per energy-per-nucleon. Fragmentation of nuclei propagating through the interstellar gas depends on energy per nucleon, since that quantity is approximately conserved when a nucleus breaks up on interaction with the gas. (3) By nucleons per energy-per-nucleon. Production of secondary cosmic rays in the atmosphere depends on the intensity of nucleons per energy-per-nucleon, approximately independently of whether the incident nucleons are free protons or bound in nuclei. (4) By particles per energy-per-nucleus. Air shower experiments that use the atmosphere as a calorimeter generally measure a quantity that is related to total energy per particle.

The units of differential intensity I are $[\mathrm{cm}^{-2}\mathrm{s}^{-1}\mathrm{sr}^{-1}\mathcal{E}^{-1}]$, where $\mathcal{E}$ represents the units of one of the four variables listed above.

The intensity of primary nucleons in the energy range from several GeV to somewhat beyond 100 TeV is given approximately by

$$I_N(E) \approx 1.8\, E^{-\alpha}\, \frac{\text{nucleons}}{\text{cm}^2 \text{ s sr GeV}}\ , \tag{24.2}$$

where E is the energy-per-nucleon (including rest mass energy) and $\alpha\ (\equiv \gamma + 1) = 2.7$ is the differential spectral index of the cosmic ray flux and γ is the integral spectral index. About 79% of the primary nucleons are free protons and about 70% of the rest are nucleons bound in helium nuclei. The fractions of the primary nuclei are nearly constant over this energy range (possibly with small but interesting variations). Fractions of both primary and secondary incident nuclei are listed in Table 24.1. Figure 24.1 [1] shows the major components as a function of energy at a particular epoch of the solar cycle. There has been a series of more precise measurements of the primary spectrum of protons and helium in the past decade [2–7].

Table 24.1: Relative abundances F of cosmic-ray nuclei at 10.6 GeV/nucleon normalized to oxygen ($\equiv$ 1) [8]. The oxygen flux at kinetic energy of 10.6 GeV/nucleon is 3.26×10^{-6} cm^{-2} s^{-1} sr^{-1} (GeV/nucleon)$^{-1}$. Abundances of hydrogen and helium are from Ref. [5,6].

Z	Element	F	Z	Element	F
1	H	540	13–14	Al-Si	0.19
2	He	26	15–16	P-S	0.03
3–5	Li-B	0.40	17–18	Cl-Ar	0.01
6–8	C-O	2.20	19–20	K-Ca	0.02
9–10	F-Ne	0.30	21–25	Sc-Mn	0.05
11–12	Na-Mg	0.22	26–28	Fe-Ni	0.12

Up to energies of at least 10^{15} eV, the composition and energy spectra of nuclei are typically interpreted in the context of "diffusion" or "leaky box" models, in which the sources of the primary cosmic radiation are located within the galaxy [9]. The ratio of secondary to primary nuclei is observed to decrease approximately as $E^{-0.5}$ with increasing energy, a fact interpreted to mean that the lifetime of cosmic rays in the galaxy decreases with energy. Measurements of radioactive "clock" isotopes [10,11] in the low energy cosmic radiation are consistent with a lifetime in the galaxy of about 15 Myr.

The spectrum of electrons and positrons incident at the top of the atmosphere is steeper than the spectra of protons and nuclei, as shown in Fig. 24.2. The positron fraction decreases from ~ 0.2 below 1 GeV [12–14] to ~ 0.1 around 2 GeV and to ~ 0.05 in at the highest energies for which it is measured ($5 - 20$ GeV) [15]. This behavior

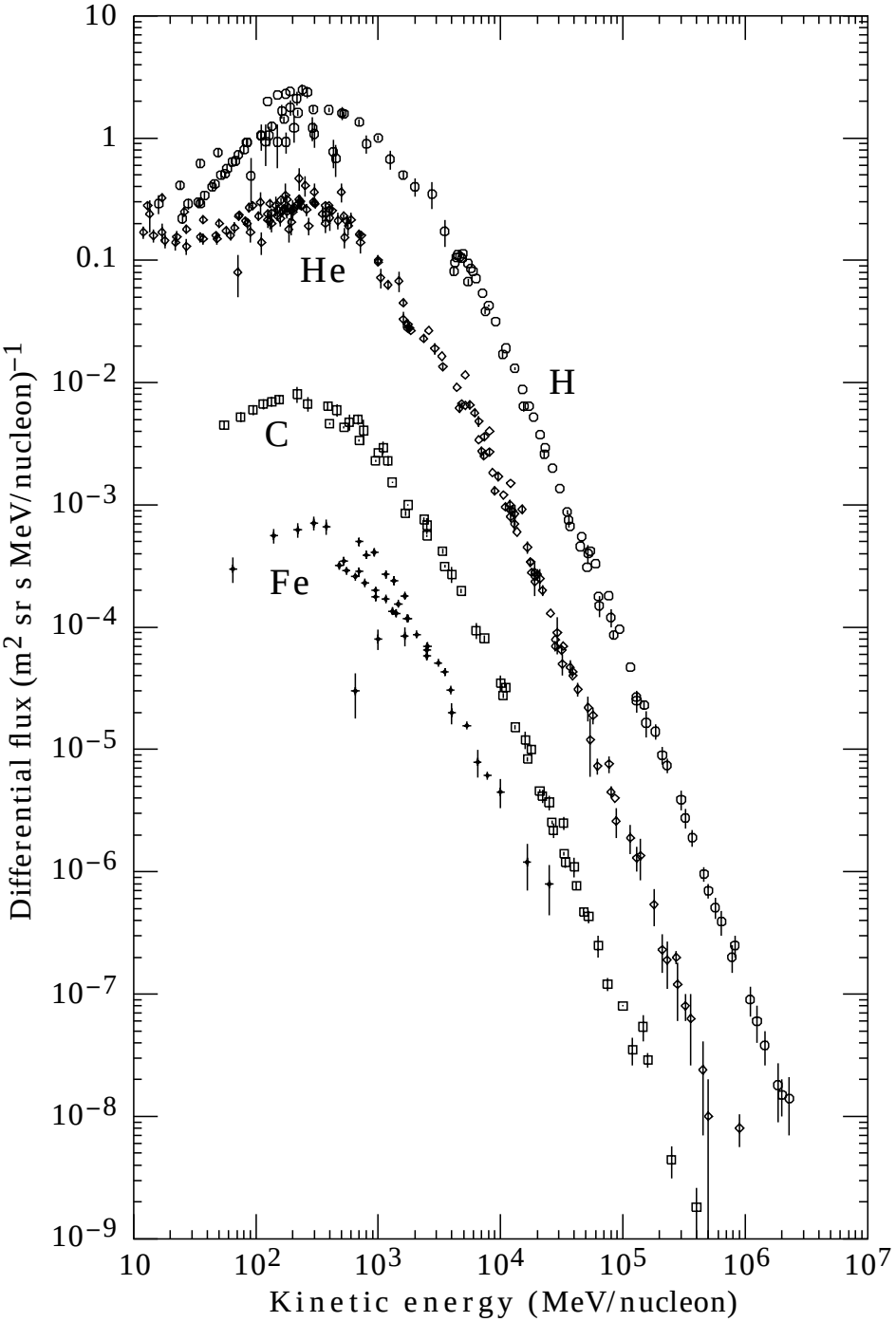


Figure 24.1: Major components of the primary cosmic radiation (from Ref. 1).

refers to measurements made during solar cycles of positive magnetic polarity and at high geomagnetic latitude. Ref. 14 discusses the dependence of the positron fraction on solar cycle and Ref. 5 studies the geomagnetic effects.

The ratio of antiprotons to protons is $\sim 2 \times 10^{-4}$ [16,17] at around 10–20 GeV, and

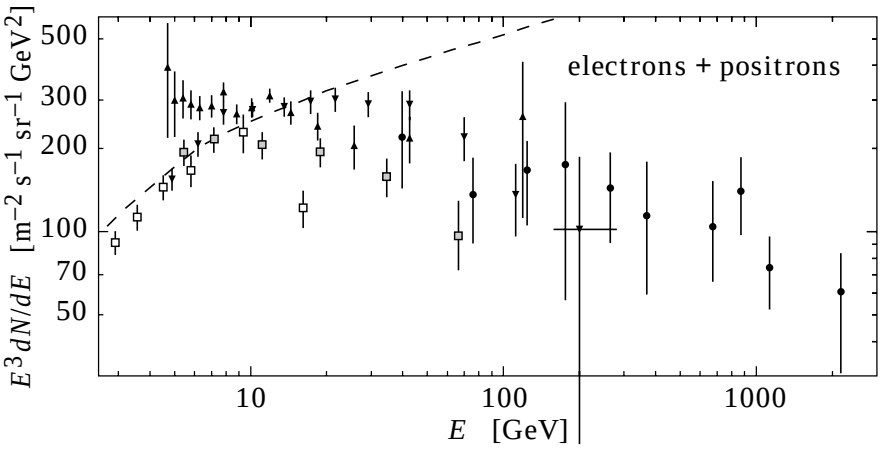


Figure 24.2: Differential spectrum of electrons plus positrons multiplied by E^3 (data summary from Ref. 9). The dashed line shows the proton spectrum multiplied by 0.01.

there is clear evidence [18–20] for the kinematic suppression at lower energy that is the signature of secondary antiprotons. The $\overline{p}/p$ ratio also shows a strong dependence on the phase and polarity of the solar cycle [21] in the opposite sense to that of the positron fraction. There is at this time no evidence for a significant primary component either of positrons or of antiprotons. No antihelium or antideuteron has been found in the cosmic radiation. The best current measured upper limit on the ratio antihelium/helium is approximately 7×10^{-4} [22]. The upper limit on the flux of antideuterons around 1 GeV/nuleon is approximately $2 \times 10^{-4}\,\mathrm{m}^2$ s sr GeV/nucleon [23].

24.2. Cosmic rays in the atmosphere

Figure 24.3 shows the vertical fluxes of the major cosmic ray components in the atmosphere in the energy region where the particles are most numerous (except for electrons, which are most numerous near their critical energy, which is about 81 MeV in air). Except for protons and electrons near the top of the atmosphere, all particles are produced in interactions of the primary cosmic rays in the air. Muons and neutrinos are products of the decay of charged mesons, while electrons and photons originate in decays of neutral mesons.

Most measurements are made at ground level or near the top of the atmosphere, but there are also measurements of muons and electrons from airplanes and balloons. Fig. 24.3 includes recent measurements of negative muons [4,24–26]. Since $\mu^+(\mu^-)$ are produced in association with $\nu_\mu(\overline{\nu}_\mu)$, the measurement of muons near the maximum of the intensity curve for the parent pions serves to calibrate the atmospheric ν_μ beam [27]. Because muons typically lose almost two GeV in passing through the atmosphere, the comparison near the production altitude is important for the sub-GeV range of $\nu_\mu(\overline{\nu}_\mu)$ energies.

The flux of cosmic rays through the atmosphere is described by a set of coupled cascade equations with boundary conditions at the top of the atmosphere to match the primary

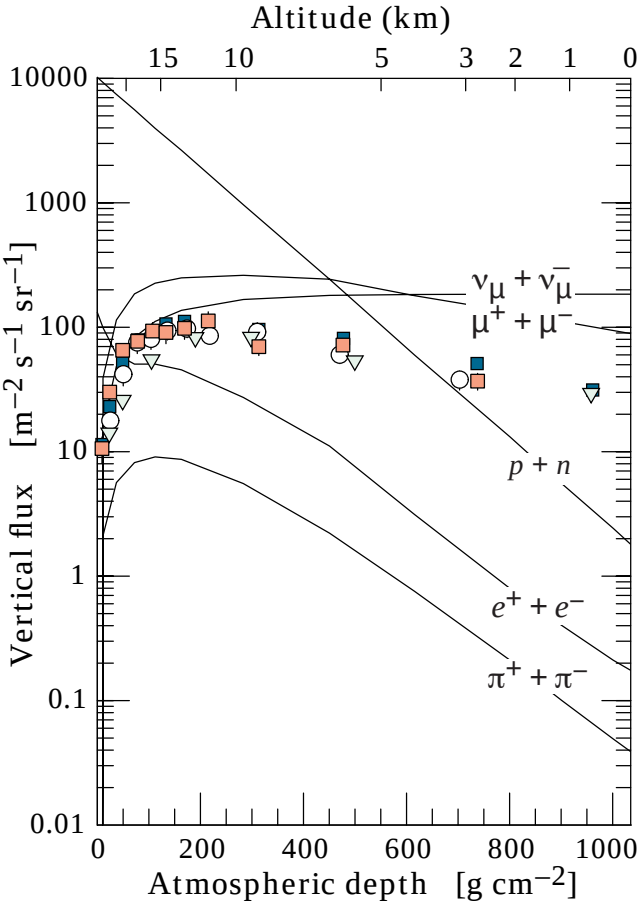


Figure 24.3: Vertical fluxes of cosmic rays in the atmosphere with $E > 1$ GeV estimated from the nucleon flux of Eq. (24.2). The points show measurements of negative muons with $E_\mu > 1$ GeV [4,24–26]. See full-color version on color pages at end of book.

spectrum. Numerical or Monte Carlo calculations are needed to account accurately for decay and energy-loss processes, and for the energy-dependences of the cross sections and of the primary spectral index γ. Approximate analytic solutions are, however, useful in limited regions of energy [28]. For example, the vertical intensity of nucleons at depth X (g cm^{-2}) in the atmosphere is given by

$$I_N(E, X) \;\approx\; I_N(E, 0)\, e^{-X/\Lambda} \;, \tag{24.3}$$

where Λ is the attenuation length of nucleons in air.

The corresponding expression for the vertical intensity of charged pions with energy $E_\pi \ll \epsilon_\pi = 115$ GeV is

$$I_\pi(E_\pi, X) \;\approx\; \frac{Z_{N\pi}}{\lambda_N}\, I_N(E_\pi, 0)\, e^{-X/\Lambda}\, \frac{X\, E_\pi}{\epsilon_\pi} \;. \tag{24.4}$$

This expression has a maximum at $t = \Lambda \approx 120$ g cm^{-2}, which corresponds to an altitude of 15 kilometers. The quantity $Z_{N\pi}$ is the spectrum-weighted moment of the inclusive

distribution of charged pions in interactions of nucleons with nuclei of the atmosphere. The intensity of low-energy pions is much less than that of nucleons because $Z_{N\pi} \approx 0.079$ is small and because most pions with energy much less than the critical energy ϵ_π decay rather than interact.

24.3. Cosmic rays at the surface

24.3.1. ***Muons*** : Muons are the most numerous charged particles at sea level (see Fig. 24.3). Most muons are produced high in the atmosphere (typically 15 km) and lose about 2 GeV to ionization before reaching the ground. Their energy and angular distribution reflect a convolution of production spectrum, energy loss in the atmosphere, and decay. For example, 2.4 GeV muons have a decay length of 15 km, which is reduced to 8.7 km by energy loss. The mean energy of muons at the ground is ≈ 4 GeV. The energy spectrum is almost flat below 1 GeV, steepens gradually to reflect the primary spectrum in the 10–100 GeV range, and steepens further at higher energies because pions with $E_\pi > \epsilon_\pi \approx 115$ GeV tend to interact in the atmosphere before they decay. Asymptotically ($E_\mu \gg 1$ TeV), the energy spectrum of atmospheric muons is one power steeper than the primary spectrum. The integral intensity of vertical muons above 1 GeV/c at sea level is ≈ 70 m^{-2}s^{-1}sr^{-1} [29,30], with recent measurements [31–33] tending to give lower normalization by 10-15%. Experimentalists are familiar with this number in the form $I \approx 1$ cm^{-2} min^{-1} for horizontal detectors.

The overall angular distribution of muons at the ground is $\propto \cos^2\theta$, which is characteristic of muons with $E_\mu \sim 3$ GeV. At lower energy the angular distribution becomes increasingly steep, while at higher energy it flattens, approaching a $\sec\theta$ distribution for $E_\mu \gg \epsilon_\pi$ and $\theta < 70°$.

Figure 24.4 shows the muon energy spectrum at sea level for two angles. At large angles low energy muons decay before reaching the surface and high energy pions decay before they interact, thus the average muon energy increases. An approximate extrapolation formula valid when muon decay is negligible ($E_\mu > 100/\cos\theta$ GeV) and the curvature of the Earth can be neglected ($\theta < 70°$) is

$$\frac{dN_\mu}{dE_\mu} \approx \frac{0.14\, E_\mu^{-2.7}}{\text{cm}^2 \text{ s sr GeV}} \times \left\{ \frac{1}{1 + \dfrac{1.1 E_\mu \cos\theta}{115\,\text{GeV}}} + \frac{0.054}{1 + \dfrac{1.1 E_\mu \cos\theta}{850\,\text{GeV}}} \right\} , \tag{24.5}$$

where the two terms give the contribution of pions and charged kaons. Eq. (24.5) neglects a small contribution from charm and heavier flavors which is negligible except at very high energy [38].

The muon charge ratio reflects the excess of π^+ over π^- in the forward fragmentation region of proton initiated interactions together with the fact that there are more protons

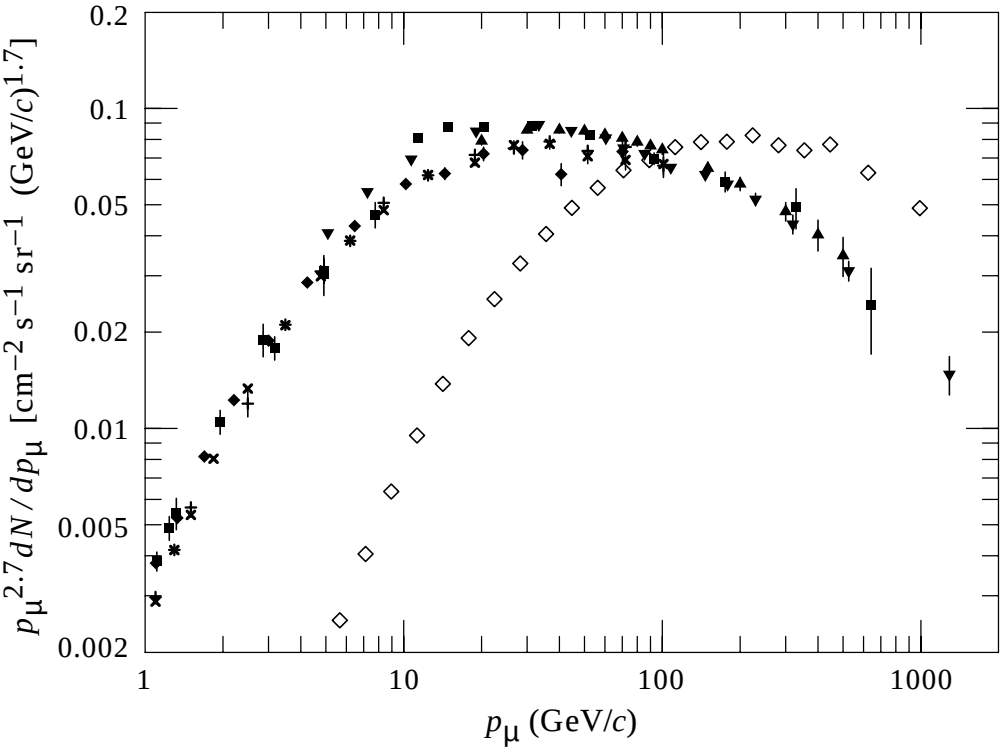


Figure 24.4: Spectrum of muons at $\theta = 0°$ (⧫ [29], ■ [34], ▼ [35], ▲ [36], × and + [31], and $\theta = 75°$ ◇ [37]) .

than neutrons in the primary spectrum. The charge ratio is between 1.1 and 1.4 from 1 GeV to 100 GeV [29,31,32]. Below 1 GeV there is a systematic dependence on location due to geomagnetic effects [31,32].

24.3.2. ***Electromagnetic component*** : At the ground, this component consists of electrons, positrons, and photons primarily from electromagnetic cascades initiated by decay of neutral and charged mesons. Muon decay is the dominant source of low-energy electrons at sea level. Decay of neutral pions is more important at high altitude or when the energy threshold is high. Knock-on electrons also make a small contribution at low energy [39]. The integral vertical intensity of electrons plus positrons is very approximately 30, 6, and 0.2 $m^{-2}s^{-1}sr^{-1}$ above 10, 100, and 1000 MeV respectively [30,40], but the exact numbers depend sensitively on altitude, and the angular dependence is complex because of the different altitude dependence of the different sources of electrons [39–41]. The ratio of photons to electrons plus positrons is approximately 1.3 above a GeV and 1.7 below the critical energy [41].

24.3.3. ***Protons*** : Nucleons above 1 GeV/c at ground level are degraded remnants of the primary cosmic radiation. The intensity is approximately represented by Eq. (24.3) with the replacement $t \to t/\cos\theta$ for $\theta < 70^\circ$ and an attenuation length $\Lambda = 123$ g cm^{-2}. At sea level, about 1/3 of the nucleons in the vertical direction are neutrons (up from $\approx 10\%$ at the top of the atmosphere as the n/p ratio approaches equilibrium). The integral intensity of vertical protons above 1 GeV/c at sea level is ≈ 0.9 m^{-2}s^{-1}sr^{-1} [30,42].

24.4. Cosmic rays underground

Only muons and neutrinos penetrate to significant depths underground. The muons produce tertiary fluxes of photons, electrons, and hadrons.

24.4.1. ***Muons*** : As discussed in Section 27.6 of this *Review*, muons lose energy by ionization and by radiative processes: bremsstrahlung, direct production of e^+e^- pairs, and photonuclear interactions. The total muon energy loss may be expressed as a function of the amount of matter traversed as

$$-\frac{dE_\mu}{dX} = a + b\,E_\mu \ , \tag{24.6}$$

where a is the ionization loss and b is the fractional energy loss by the three radiation processes. Both are slowly varying functions of energy. The quantity $\epsilon \equiv a/b$ (≈ 500 GeV in standard rock) defines a critical energy below which continuous ionization loss is more important than radiative losses. Table 24.2 shows a and b values for standard rock as a function of muon energy. The second column of Table 24.2 shows the muon range in standard rock ($A = 22$, $Z = 11$, $\rho = 2.65$ g cm^{-3}). These parameters are quite sensitive to the chemical composition of the rock, which must be evaluated for each experimental location.

Table 24.2: Average muon range R and energy loss parameters calculated for standard rock [43]. Range is given in km-water-equivalent, or 10^5 g cm^{-2}.

E_μ	R	a	$b_{\rm brems}$	$b_{\rm pair}$	$b_{\rm nucl}$	$\sum b_i$
GeV	km.w.e.	MeV g^{-1} cm^2	——	10^{-6} g^{-1} cm^2		——
10	0.05	2.17	0.70	0.70	0.50	1.90
100	0.41	2.44	1.10	1.53	0.41	3.04
1000	2.45	2.68	1.44	2.07	0.41	3.92
10000	6.09	2.93	1.62	2.27	0.46	4.35

The intensity of muons underground can be estimated from the muon intensity in the atmosphere and their rate of energy loss. To the extent that the mild energy dependence

of a and b can be neglected, Eq. (24.6) can be integrated to provide the following relation between the energy $E_{\mu,0}$ of a muon at production in the atmosphere and its average energy E_μ after traversing a thickness X of rock (or ice or water):

$$E_\mu \;=\; (E_{\mu,0} + \epsilon)\, e^{-bX} \;-\; \epsilon \;. \tag{24.7}$$

Especially at high energy, however, fluctuations are important and an accurate calculation requires a simulation that accounts for stochastic energy-loss processes [44].

There are two depth regimes for Eq. (24.7). For $X \ll b^{-1} \approx 2.5$ km water equivalent, $E_{\mu,0} \approx E_\mu(X) + aX$, while for $X \gg b^{-1}$ $E_{\mu,0} \approx (\epsilon + E_\mu(X)) \exp(bX)$. Thus at shallow depths the differential muon energy spectrum is approximately constant for $E_\mu < aX$ and steepens to reflect the surface muon spectrum for $E_\mu > aX$, whereas for $X > 2.5$ km.w.e. the differential spectrum underground is again constant for small muon energies but steepens to reflect the surface muon spectrum for $E_\mu > \epsilon \approx 0.5$ TeV. In the deep regime the shape is independent of depth although the intensity decreases exponentially with depth. In general the muon spectrum at slant depth X is

$$\frac{dN_\mu(X)}{dE_\mu} \;=\; \frac{dN_\mu}{dE_{\mu,0}}\,\frac{dE_{\mu,0}}{dE_\mu} \;=\; \frac{dN_\mu}{dE_{\mu,0}}\, e^{bX} \;, \tag{24.8}$$

where $E_{\mu,0}$ is the solution of Eq. (24.7) in the approximation neglecting fluctuations.

Fig. 24.5 shows the vertical muon intensity versus depth. In constructing this "depth-intensity curve," each group has taken account of the angular distribution of the muons in the atmosphere, the map of the overburden at each detector, and the properties of the local medium in connecting measurements at various slant depths and zenith angles to the vertical intensity. Use of data from a range of angles allows a fixed detector to cover a wide range of depths. The flat portion of the curve is due to muons produced locally by charged-current interactions of ν_μ.

24.4.2. ***Neutrinos*** : Because neutrinos have small interaction cross sections, measurements of atmospheric neutrinos require a deep detector to avoid backgrounds. There are two types of measurements: contained (or semi-contained) events, in which the vertex is determined to originate inside the detector, and neutrino-induced muons. The latter are muons that enter the detector from zenith angles so large (*e.g.*, nearly horizontal or upward) that they cannot be muons produced in the atmosphere. In neither case is the neutrino flux measured directly. What is measured is a convolution of the neutrino flux and cross section with the properties of the detector (which includes the surrounding medium in the case of entering muons).

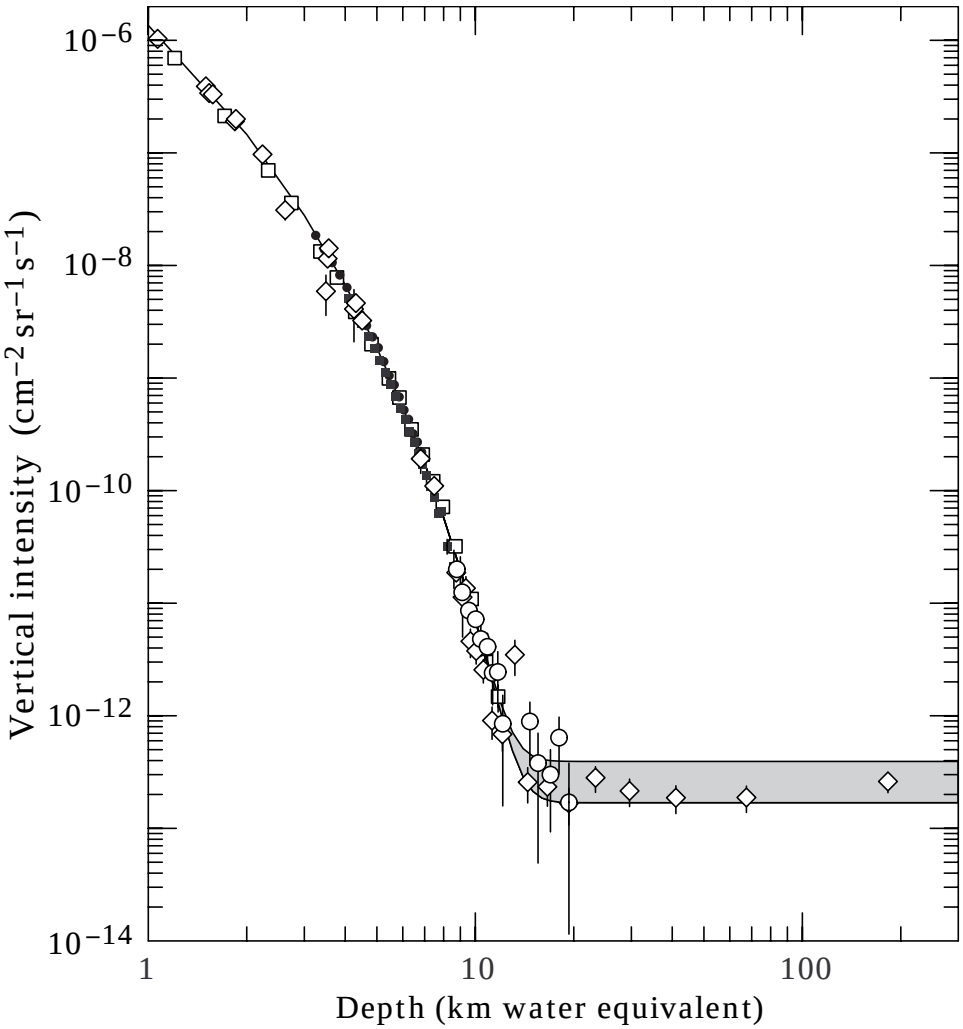


Figure 24.5: Vertical muon intensity vs depth (1 km.w.e. = 10^5 g cm^{-2} of standard rock). The experimental data are from: ♢: the compilations of Crouch [45], □: Baksan [46], ○: LVD [47], ●: MACRO [48], ■: Frejus [49]. The shaded area at large depths represents neutrino-induced muons of energy above 2 GeV. The upper line is for horizontal neutrino-induced muons, the lower one for vertically upward muons.

Contained and semi-contained events reflect neutrinos in the sub-GeV to multi-GeV region where the product of increasing cross section and decreasing flux is maximum. In the GeV region the neutrino flux and its angular distribution depend on the geomagnetic location of the detector and, to a lesser extent, on the phase of the solar cycle. Naively, we expect $\nu_\mu/\nu_e = 2$ from counting neutrinos of the two flavors coming from the chain of pion and muon decay. This ratio is only slightly modified by the details of the decay kinematics, but the fraction of electron neutrinos gradually decreases above a GeV as parent muons begin to reach the ground before decaying. Experimental measurements have to account for the ratio of $\overline{\nu}/\nu$, which have cross sections different by a factor of 3 in this energy range. In addition, detectors generally have different efficiencies for detecting

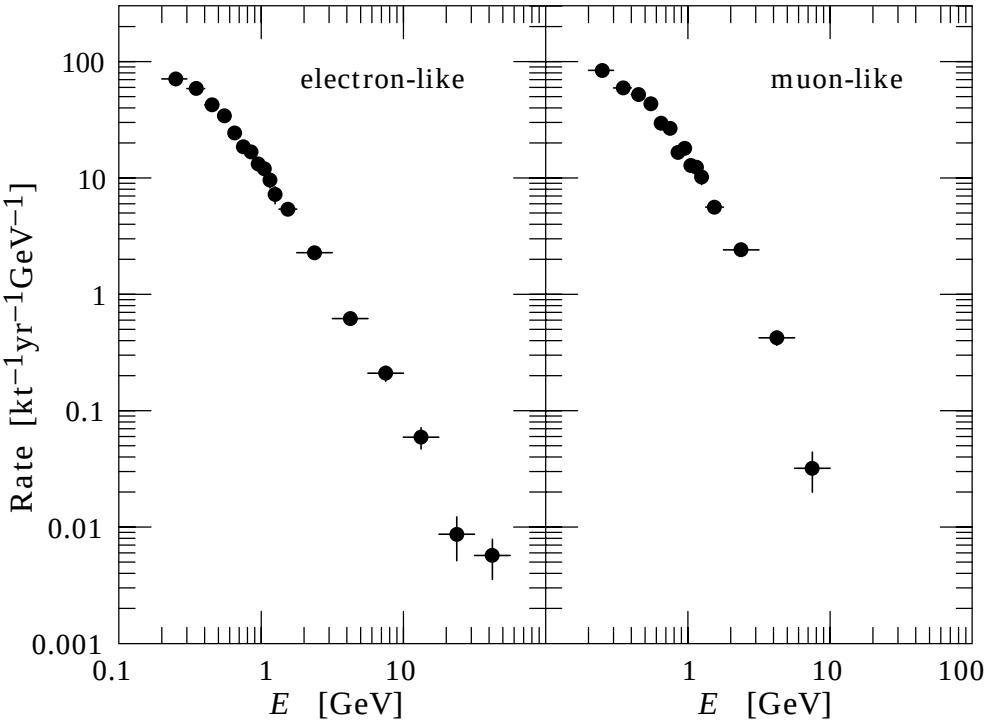


Figure 24.6: Sub-GeV and multi-GeV neutrino interactions from SuperKamiokande [50]. The plot shows the spectra of visible energy in the detector.

muon neutrinos and electron neutrinos which need to be accounted for in comparing measurements with expectation. Fig. 24.6 shows the distributions of the visible energy in the Super-Kamiokande detector [50] for electron-like and muon-like charged current neutrino interactions. Contrary to expectation, the numbers of the two classes of events are similar rather than different by a factor of two. The exposure for the data sample shown here is 50 kiloton-years. The falloff of the muon-like events at high energy is a consequence of the poor containment for high energy muons. Corrections for detection efficiencies and backgrounds are, however, insufficient to account for the large difference from the expectation.

Two well-understood properties of atmospheric cosmic rays provide a standard for comparison of the measurements of atmospheric neutrinos. These are the "$\sec\theta$ effect" and the "east-west effect". The former refers originally to the enhancement of the flux of > 10 GeV muons (and neutrinos) at large zenith angles because the parent pions propagate more in the low density upper atmosphere where decay is enhanced relative to interaction. For neutrinos from muon decay, the enhancement near the horizontal becomes important for $E_\nu > 1$ GeV and arises mainly from the increased pathlength through the atmosphere for muon decay in flight. Fig. 24.7 from Ref. 51 shows a comparison between measurement and expectation for the zenith angle dependence of multi-GeV electron-like (mostly ν_e) and muon-like (mostly ν_μ) events separately. The ν_e show an enhancement near the horizontal and approximate equality for nearly upward ($\cos\theta \approx -1$) and nearly downward ($\cos\theta \approx 1$) events. There is, however, a very significant deficit of upward ($\cos\theta < 0$) ν_μ events, which have long pathlengths comparable to the radius of the

Earth. This pattern has been interpreted as evidence for oscillations involving muon neutrinos [50]. (See the article on neutrino properties in this *Review.*) Including three dimensional effects in the calculation of atmospheric neutrinos may change somewhat the expected angular distributions of neutrinos at low energy [52], but it does not change the fundamental expectation of up-down symmetry, which is the basis of the evidence for oscillations.

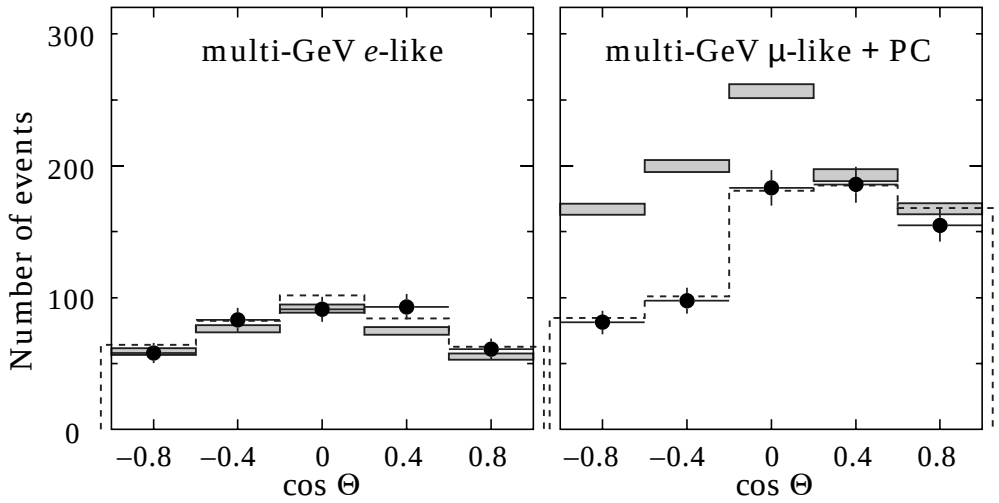


Figure 24.7: Zenith-angle dependence of multi-GeV neutrino interactions from SuperKamiokande [51]. The shaded boxes show the expectation in the absence of any oscillations. The lines show fits with some assumed oscillation parameters, as described in Ref. 51.

The east-west effect [53,54] is the enhancement, especially at low geomagnetic latitudes, of cosmic rays incident on the atmosphere from the west as compared to those from the east. This is a consequence of the fact that the cosmic rays are positively charged nuclei which are bent systematically in one sense in the geomagnetic field. Not all trajectories can reach the atmosphere from outside the geomagnetic field. The standard procedure to see which trajectories are allowed is to inject antiprotons outward from near the top of the atmosphere in various directions and see if they escape from the geomagnetic field without becoming trapped indefinitely or intersecting the surface of the Earth. Any direction in which an antiproton of a given momentum can escape is an allowed direction from which a proton of the opposite momentum can arrive. Since the geomagnetic field is oriented from south to north in the equatorial region, antiprotons injected toward the east are bent back towards the Earth. Thus there is a range of momenta and zenith angles for which positive particles cannot arrive from the east but can arrive from the west. This east-west asymmetry of the incident cosmic rays induces a similar asymmetry on the secondaries, including neutrinos. Since this is an azimuthal effect, the resulting asymmetry is independent of possible oscillations, which depend on pathlength (equivalently zenith angle), but not on azimuth. Fig. 24.8 (from Ref. 55) is a comparison of data and expectation for this effect, which serves as a consistency check of the measurement and analysis.

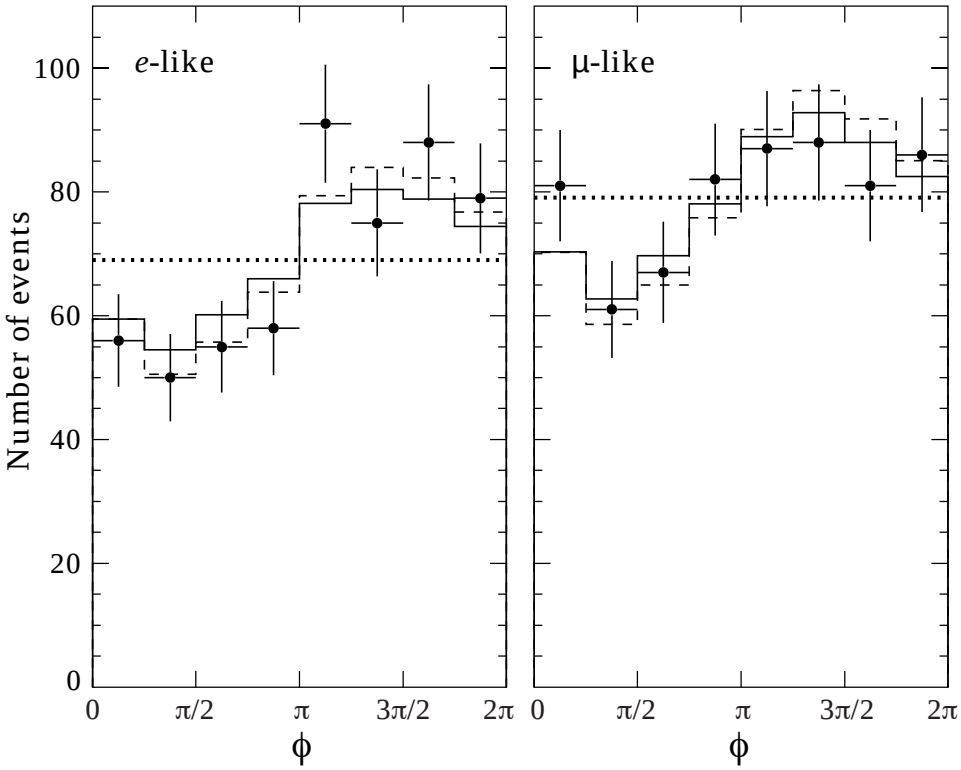


Figure 24.8: Azimuthal dependence of ∼GeV neutrino interactions from SuperKamiokande [55]. The cardinal points of the compass are S, E, N, W starting at 0. These are the direction from which the particles arrive. The lines show the expectation based on two different calculations, as described in Ref. 55.

Muons that enter the detector from outside after production in charged-current interactions of neutrinos naturally reflect a higher energy portion of the neutrino spectrum than contained events because the muon range increases with energy as well as the cross section. The relevant energy range is $\sim 10 < E_\nu < 1000$ GeV, depending somewhat on angle. Neutrinos in this energy range show a $\sec\theta$ effect similar to muons (see Eq. (24.5)). This causes the flux of horizontal neutrino-induced muons to be approximately a factor two higher than the vertically upward flux. The upper and lower edges of the horizontal shaded region in Fig. 24.5 correspond to horizontal and vertical intensities of neutrino-induced muons. Table 24.3 gives the measured fluxes of upward-moving neutrino-induced muons averaged over the lower hemisphere. Generally the definition of minimum muon energy depends on where it passes through the detector. The tabulated effective minimum energy estimates the average over various accepted trajectories.

Table 24.3: Measured fluxes (10^{-13} cm^{-2} s^{-1} sr^{-1}) of neutrino-induced muons as a function of the effective minimum muon energy E_μ.

$E_\mu >$	1 GeV	1 GeV	1 GeV	2 GeV	3 GeV	3 GeV
Ref.	CWI [56]	Baksan [57]	MACRO [58]	IMB [59]	Kam [60]	SuperK [61]
F_μ	2.17±0.21	2.77±0.17	2.29 ± 0.15	2.26±0.11	1.94±0.12	1.74±0.07

24.5. Air showers

So far we have discussed inclusive or uncorrelated fluxes of various components of the cosmic radiation. An air shower is caused by a single cosmic ray with energy high enough for its cascade to be detectable at the ground. The shower has a hadronic core, which acts as a collimated source of electromagnetic subshowers, generated mostly from $\pi^0 \to \gamma\,\gamma$. The resulting electrons and positrons are the most numerous particles in the shower. The number of muons, produced by decays of charged mesons, is an order of magnitude lower.

Air showers spread over a large area on the ground, and arrays of detectors operated for long times are useful for studying cosmic rays with primary energy $E_0 > 100$ TeV, where the low flux makes measurements with small detectors in balloons and satellites difficult.

Greisen [62] gives the following approximate expressions for the numbers and lateral distributions of particles in showers at ground level. The total number of muons N_μ with energies above 1 GeV is

$$N_\mu(> 1 \text{ GeV}) \;\approx\; 0.95 \times 10^5 \left(N_e/10^6\right)^{3/4} \;, \tag{24.9}$$

where N_e is the total number of charged particles in the shower (not just $e^\pm$). The number of muons per square meter, ρ_μ, as a function of the lateral distance r (in meters) from the center of the shower is

$$\rho_\mu = \frac{1.25\,N_\mu}{2\pi\,\Gamma(1.25)} \left(\frac{1}{320}\right)^{1.25} r^{-0.75} \left(1 + \frac{r}{320}\right)^{-2.5} \;, \tag{24.10}$$

where Γ is the gamma function. The number density of charged particles is

$$\rho_e = C_1(s, d, C_2)\, x^{(s-2)}(1+x)^{(s-4.5)}(1 + C_2 x^d) \;. \tag{24.11}$$

Here s, d, and C_2 are parameters in terms of which the overall normalization constant $C_1(s, d, C_2)$ is given by

$$C_1(s, d, C_2) = \frac{N_e}{2\pi r_1^2}\,[\,B(s, 4.5 - 2s)$$

$$+ C_2\, B(s+d,\, 4.5-d-2s)]^{-1}\ , \tag{24.12}$$

where $B(m,n)$ is the beta function. The values of the parameters depend on shower size (N_e), depth in the atmosphere, identity of the primary nucleus, etc. For showers with $N_e \approx 10^6$ at sea level, Greisen uses $s = 1.25$, $d = 1$, and $C_2 = 0.088$. Finally, x is r/r_1, where r_1 is the Molière radius, which depends on the density of the atmosphere and hence on the altitude at which showers are detected. At sea level $r_1 \approx 78$ m. It increases with altitude.

The lateral spread of a shower is determined largely by Coulomb scattering of the many low-energy electrons and is characterized by the Molière radius. The lateral spread of the muons (ρ_μ) is larger and depends on the transverse momenta of the muons at production as well as multiple scattering.

There are large fluctuations in development from shower to shower, even for showers of the same energy and primary mass—especially for small showers, which are usually well past maximum development when observed at the ground. Thus the shower size N_e and primary energy E_0 are only related in an average sense, and even this relation depends on depth in the atmosphere. One estimate of the relation is [63]

$$E_0 \ \sim\ 3.9 \times 10^6 \text{ GeV } (N_e/10^6)^{0.9} \tag{24.13}$$

for vertical showers with $10^{14} < E < 10^{17}$ eV at 920 g cm^{-2} (965 m above sea level). Because of fluctuations, N_e as a function of E_0 is not the inverse of Eq. (24.13). As E_0 increases the shower maximum (on average) moves down into the atmosphere and the relation between N_e and E_0 changes. At the maximum of shower development, there are approximately 2/3 particles per GeV of primary energy.

Detailed simulations and cross-calibrations between different types of detectors are necessary to establish the primary energy spectrum from air-shower experiments [63,64]. Figure 24.9 shows the "all-particle" spectrum. The differential energy spectrum has been multiplied by $E^{2.5}$ in order to display the features of the steep spectrum that are otherwise difficult to discern. The steepening that occurs between 10^{15} and 10^{16} eV is known as the *knee* of the spectrum. The feature around 10^{19} eV is called the *ankle* of the spectrum.

Measurements with small air shower experiments in the knee region differ by as much as a factor of two, indicative of systematic uncertainties in interpretation of the data. (For a recent review see Ref. 66.) In establishing the spectrum shown in Fig. 24.9, efforts have been made to minimize the dependence of the analysis on the primary composition. Ref. 67 uses an unfolding procedure to obtain the spectra of the individual components, giving a result for the all-particle spectrum between 10^{15} and 10^{17} eV that lies toward the upper range of the data shown in Fig. 24.9. In the energy range above 10^{17} eV, the Fly's Eye technique [68] is particularly useful because it can establish the primary energy in a model-independent way by observing most of the longitudinal development of each shower, from which E_0 is obtained by integrating the energy deposition in the atmosphere.

If the cosmic ray spectrum below 10^{18} eV is of galactic origin, the *knee* could reflect the fact that some (but not all) cosmic accelerators have reached their maximum energy.

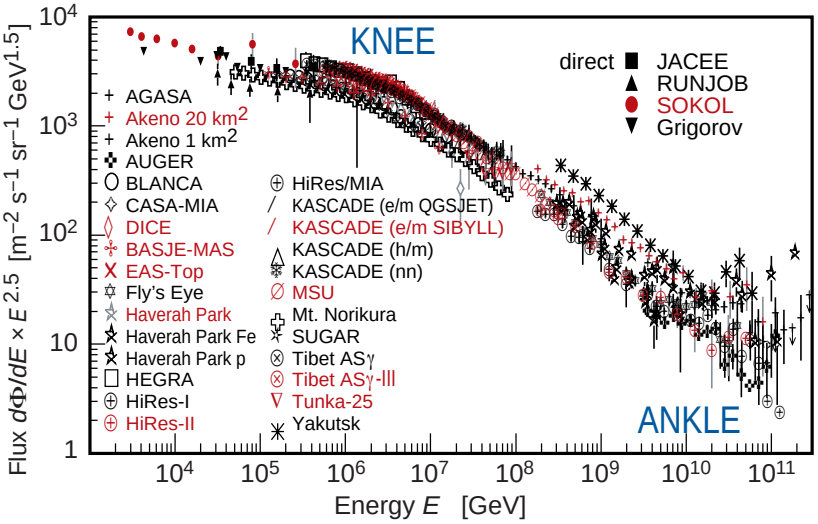


Figure 24.9: The all-particle spectrum: for references see [65]. Figure used by permission of author. See full-color version on color pages at end of book.

Some types of expanding supernova remnants, for example, are estimated not to be able to accelerate particles above energies in the range of 10^{15} eV total energy per particle. Effects of propagation and confinement in the galaxy [69] also need to be considered.

It was previously thought that the most likely explanation for the dip in the spectrum near 3×10^{18} eV called the *ankle* is a higher energy population of particles overtaking a lower energy population, for example an extragalactic flux beginning to dominate over the galactic flux. The situation now seems more complicated because of better evidence for the existence of a break in the spectrum near 5×10^{17} eV called the *second knee* [70], just below the ankle structure. There are clear predictions of a dip structure in the region of the observed ankle produced by e^+/e^- energy losses of extragalactic protons on the 2.7 K cosmic microwave radiation (CMB) [71]. This dip structure has been claimed to be a more robust signature of both the protonic and extragalactic nature of the highest energy cosmic rays than the GZK cutoff (see below) itself. If this interpretation is correct, then the most likely explanation of the second knee would be the termination of the galactic cosmic ray flux. Composition changes across this energy region would be important correlative evidence for this view.

If the cosmic ray flux above the second knee is cosmological in origin, then there should be a rapid change in the spectral index (called the GZK cutoff) around 5×10^{19} eV, resulting from the onset of inelastic interactions with the (CMB) [72,73]. While several experiments have reported events that have been assigned energies above 10^{20} eV [74–77], more recent experiments such as HiRes [78] have failed to confirm this; results are consistent with the expected cutoff. An implication of a continued spectrum would be that some sources of the highest energy particles must be relatively nearby. For example, the attenuation length for protons at $2 \cdot 10^{20}$ eV is 30 Mpc [73]. Both cosmic accelerators (bottom up) and massive cosmological relics (top down) have been suggested [74–77].

Figure 24.10 gives an expanded view of the high energy end of the spectrum, showing

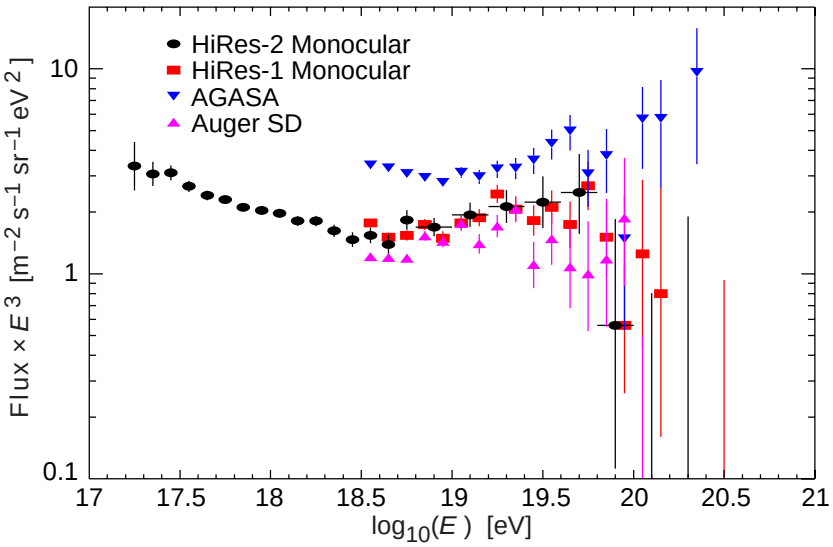


Figure 24.10: Expanded view of the highest energy portion of the cosmic-ray spectrum. ■ [78](HiRes1 monocular), • [78](HiRes2 monocular), ▲ [79](Auger) ▼ [76](AGASA) See full-color version on color pages at end of book.

only the more recent experiments. This figure and the previous one have shown the differential flux multiplied by a power of the energy, a procedure that enables one to see structure in the spectrum more clearly but amplifies small systematic differences in energy assignments into sizable normalization differences. All existing experiments are actually consistent in normalization if one takes quoted systematic errors in the energy scales into account. However, the continued power law type of flux beyond the GZK cutoff claimed by the AGASA experiment is contradicted by the HiRes data. New data from the Pierre Auger experiment [79] are still too preliminary in energy scale and statistics to impact this debate.

References:

1. J.A. Simpson, Ann. Rev. Nucl. and Part. Sci. **33**, 323 (1983).
2. M. Boezio *et al.*, Astrophys. J. **532**, 653 (1999).
3. W. Menn *et al.*, Astrophys. J. **533**, 281 (2000).
4. R. Bellotti *et al.*, Phys. Rev. **D60**, 052002 (1999).
5. AMS Collaboration, Phys. Lett. **B490**, 27 (2000).
6. T. Sanuki *et al.*, Astrophys. J. **545**, 1135 (2000).
7. S. Haino *et al.*, Phys. Lett. **B594**, 35 (2004).
8. J.J. Engelmann *et al.*, Astron. & Astrophys. **233**, 96 (1990);
   See also *Cosmic Abundances of Matter* (ed. C. Jake Waddington) A.I.P. Conf. Proc. No. 183 (1988), p. 111.
9. S.W. Barwick *et al.*, Astrophys. J. **498**, 779 (1998).
10. N.E. Yanasak *et al.*, Astrophys. J. 563, **768** (2001).
11. T. Hams *et al.*, Astrophys. J. 611, 892 (2004).
12. M. Boezio *et al.*, Astrophys. J. **552**, 635 (2000).
13. J. Alcaraz *et al.*, Phys. Lett. **B484**, 10 (2000).
14. J.M. Clem & P.A. Evenson, Astrophys. J. **568**, 216 (2002).

15. M.A. DuVernois *et al.*, Astrophys. J. **559**, 296 (2000).
16. M. Hof *et al.*, Astrophys. J. **467**, L33 (1996);
    See also G. Basini *et al.*, *Proc. 26th Int. Cosmic Ray Conf.*, Salt Lake City, **3**, 77 (1999).
17. A.S. Beach *et al.*, Phys. Rev. Lett. **87**, 271101 (2001).
18. J.W. Mitchell *et al.*, Phys. Rev. Lett. **76**, 3057 (1996).
19. M. Boezio *et al.*, Astrophys. J. **487**, 415 (1997).
20. S. Orito *et al.*, Phys. Rev. Lett. **84**, 1078 (2000).
21. Y. Asaoka *et al.*, Phys. Rev. Lett. **88**, 51101 (2002).
22. M. Sasaki *et al.*, Nucl. Phys. **B113**, 202 (2002).
23. H. Fuke *et al.*, Phys. Rev. Lett. **95**, 081101 (2005).
24. R. Bellotti *et al.*, Phys. Rev. **D53**, 35 (1996).
25. M. Boezio *et al.*, Phys. Rev. **D62**, 032007 (2000).
26. S. Coutu *et al.*, Phys. Rev. **D62**, 032001 (2000).
27. D.H. Perkins, Astropart. Phys. **2**, 249 (1994).
28. T.K. Gaisser, *Cosmic Rays and Particle Physics*, Cambridge University Press (1990).
29. M.P. De Pascale *et al.*, J. Geophys. Res. **98**, 3501 (1993).
30. P.K.F. Grieder, *Cosmic Rays at Earth*, Elsevier Science (2001).
31. J. Kremer *et al.*, Phys. Rev. Lett. **83**, 4241 (1999).
32. M. Motoki *et al.*, Astropart. Phys. **19**, 113 (2003).
33. P. Le Coultre on behalf of the L3 Collaboration, *Proc. 27th Int. Cosmic Ray Conf.*, (Hamburg) **3**, 974 (2001)..
34. O.C. Allkofer, K. Carstensen, and W.D. Dau, Phys. Lett. **B36**, 425 (1971).
35. B.C. Rastin, J. Phys. **G10**, 1609 (1984).
36. C.A. Ayre *et al.*, J. Phys. **G1**, 584 (1975).
37. H. Jokisch *et al.*, Phys. Rev. **D19**, 1368 (1979).
38. C.G.S. Costa, Astropart. Phys. **16**, 193 (2001).
39. S. Hayakawa, *Cosmic Ray Physics*, Wiley, Interscience, New York (1969).
40. R.R. Daniel and S.A. Stephens, Revs. Geophysics & Space Sci. **12**, 233 (1974).
41. K.P. Beuermann and G. Wibberenz, Can. J. Phys. **46**, S1034 (1968).
42. I.S. Diggory *et al.*, J. Phys. **A7**, 741 (1974).
43. D.E. Groom, N.V. Mokhov, and S.I. Striganov, "Muon stopping-power and range tables," Atomic Data and Nuclear Data Tables, 78, 183 (2001).
44. P. Lipari and T. Stanev, Phys. Rev. **D44**, 3543 (1991).
45. M. Crouch, in *Proc. 20th Int. Cosmic Ray Conf.*, Moscow, **6**, 165 (1987).
46. Yu.M. Andreev, V.I. Gurentzov, and I.M. Kogai, in *Proc. 20th Int. Cosmic Ray Conf.*, Moscow, **6**, 200 (1987).
47. M. Aglietta *et al.*, (LVD Collaboration), Astropart. Phys. **3**, 311 (1995).
48. M. Ambrosio *et al.*, (MACRO Collaboration), Phys. Rev. **D52**, 3793 (1995).
49. Ch. Berger *et al.*, (Frejus Collaboration), Phys. Rev. **D40**, 2163 (1989).
50. Y. Fukuda *et al.*, Phys. Rev. Lett. **81**, 1562 (1998).
51. K. Scholberg *et al.*, in *Proc. Eighth Int. Workshop on Neutrino Telescopes* (ed. Milla Baldo Ceolin, edizioni papergraf, Venezia, 1999) p. 183.
52. G. Battistoni *et al.*, Astropart. Phys. **12**, 315 (2000).

53. T.H. Johnson, Phys. Rev. **43**, 834 (1933).;
See also T.H. Johnson & E.C. Street, Phys. Rev. **44**, 125 (1933).
54. L.W. Alvarez and A.H. Compton, Phys. Rev. **43**, 835 (1933).
55. T. Futagami *et al.*, Phys. Rev. Lett. **82**, 5194 (1999).
56. F. Reines *et al.*, Phys. Rev. Lett. **15**, 429 (1965).
57. M.M. Boliev *et al.*, in *Proc. 3rd Int. Workshop on Neutrino Telescopes*(ed. Milla Baldo Ceolin), 235 (1991).
58. M. Ambrosio *et al.*, (MACRO) Phys. Lett. **B434**, 451 (1998). The number quoted for MACRO is the average over 90% of the lower hemisphere, $\cos\theta < -0.1$; see F. Ronga *et al.*, `hep-ex/9905025`.
59. R. Becker-Szendy *et al.*, Phys. Rev. Lett. **69**, 1010 (1992);
Proc. 25th Int. Conf. High-Energy Physics, Singapore (ed. K.K. Phua and Y. Yamaguchi, World Scientific), 662 1991).
60. S. Hatakeyama *et al.*, Phys. Rev. Lett. **81**, 2016 (1998).
61. Y. Fukuda *et al.*, Phys. Rev. Lett. **82**, 2644 (1999).
62. K. Greisen, Ann. Rev. Nucl. Sci. **10**, 63 (1960).
63. M. Nagano *et al.*, J. Phys. **G10**, 1295 (1984).
64. M. Teshima *et al.*, J. Phys. **G12**, 1097 (1986).
65. J. Hoerandel, Astropart. Phys. **19**, 193 (2003).
66. S.P. Swordy *et al.*, Astropart. Phys. **18**, 129 (2002).
67. H. Ulrich *et al.*, *Proc. 27th Int. Cosmic Ray Conf.*, Hamburg, **1**, 97 (2001).
68. D.J. Bird *et al.*, Astrophys. J. **424**, 491 (1994).
69. V.S. Ptuskin *et al.*, Astron. & Astrophys. **268**, 726 (1993).
70. Z. Cao, *Very High Energy Phenomena in the Universe*, 2001 Rencontre de Moriond, Les Arcs, Gioi Publications (2003), p. 205.
71. V. Berezinskii *et al.*, `hep-ph/0204357`.
72. K. Greisen, Phys. Rev. Lett. **16**, 748 (1966).
73. G.T. Zatsepin and V.A. Kuz'min, Sov. Phys. JETP Lett. **4**, 78 (1966).
74. J. Linsley, Phys. Rev. Lett. **10**, 146 (1963).
75. D.J. Bird *et al.*, Astrophys. J. **441**, 144 (1995).
76. N. Hayashima *et al.*, Phys. Rev. Lett. **73**, 3941 (1994).
77. M. Nagano & A.A. Watson, Rev. Mod. Phys. **72**, 689 (2000).
78. R. Abbasi *et al.*, Phys. Lett. B619, 271 (2005).
79. P. Sommers *et al.*, `astro-ph/0507150`, *Proc. 29th Int. Cosmic Ray Conf.*, Pune, India (2005).

25. ACCELERATOR PHYSICS OF COLLIDERS

Revised August 2005 by K. Desler and D. A. Edwards (DESY).

25.1. Luminosity

The event rate R in a collider is proportional to the interaction cross section $\sigma_{\rm int}$ and the factor of proportionality is called the *luminosity*:

$$R = \mathscr{L}\sigma_{\rm int} \ . \tag{25.1}$$

If two bunches containing n_1 and n_2 particles collide with frequency f, the luminosity is

$$\mathscr{L} = f\frac{n_1 n_2}{4\pi\sigma_x\sigma_y} \tag{25.2}$$

where σ_x and σ_y characterize the Gaussian transverse beam profiles in the horizontal (bend) and vertical directions and to simplify the expression it is assumed that the bunches are identical in transverse profile, that the profiles are independent of position along the bunch, and the particle distributions are not altered during collision. Whatever the distribution at the source, by the time the beam reaches high energy, the normal form is a good approximation thanks to the central limit theorem of probability and the diminished importance of space charge effects.

The beam size can be expressed in terms of two quantities, one termed the *transverse emittance*, ϵ, and the other, the *amplitude function*, β. The transverse emittance is a beam quality concept reflecting the process of bunch preparation, extending all the way back to the source for hadrons and, in the case of electrons, mostly dependent on synchrotron radiation. The amplitude function is a beam optics quantity and is determined by the accelerator magnet configuration. When expressed in terms of σ and β the transverse emittance becomes

$$\epsilon = \pi\sigma^2/\beta \ . \tag{25.3}$$

Of particular significance is the value of the amplitude function at the interaction point, β^*. Clearly one wants β^* to be as small as possible; how small depends on the capability of the hardware to make a near-focus at the interaction point.

Eq. (25.2) can now be recast in terms of emittances and amplitude functions as

$$\mathscr{L} = f\frac{n_1 n_2}{4\sqrt{\epsilon_x\,\beta_x^*\,\epsilon_y\,\beta_y^*}} \ . \tag{25.4}$$

Thus, to achieve high luminosity, all one has to do is make high population bunches of low emittance to collide at high frequency at locations where the beam optics provides as low values of the amplitude functions as possible.

25.2. Beam dynamics

Today's operating HEP colliders are all synchrotrons, and the organization of this section reflects that circumstance.

A major concern of beam dynamics is stability: conservation of adequate beam properties over a sufficiently long time scale. Several time scales are involved, and the approximations used in writing the equations of motion reflect the time scale under consideration. For example, when, in Sec. 25.2.1 below, we write the equations for transverse stability no terms associated with phase stability or synchrotron radiation appear; the time scale associated with the last two processes is much longer than that demanded by the need for transverse stability.

25.2.1. *Betatron oscillations* :

Present-day high-energy accelerators employ alternating gradient focussing provided by quadrupole magnetic fields [1], [2]. The equations of motion of a particle undergoing oscillations with respect to the design trajectory are

$$x'' + K_x(s)\, x = 0 \ , \quad y'' + K_y(s)\, y = 0 \ , \tag{25.5}$$

with

$$x' \equiv dx/ds \ , \quad y' \equiv dy/ds \tag{25.6}$$

$$K_x \equiv B'/(B\rho) + \rho^{-2} \ , \quad K_y \equiv -B'/(B\rho) \tag{25.7}$$

$$B' \equiv \partial B_y/\partial x \ . \tag{25.8}$$

The independent variable s is path length along the design trajectory. This motion is called a *betatron* oscillation because it was initially studied in the context of that type of accelerator. The functions K_x and K_y reflect the transverse focussing—primarily due to quadrupole fields except for the radius of curvature, ρ, term in K_x for a synchrotron—so each equation of motion resembles that for a harmonic oscillator but with spring constants that are a function of position. No terms relating to synchrotron oscillations appear, because their time scale is much longer and in this approximation play no role.

These equations have the form of Hill's equation and so the solution in one plane may be written as

$$x(s) = A\sqrt{\beta(s)}\ \cos(\psi(s) + \delta), \tag{25.9}$$

where A and δ are constants of integration and the phase advances according to $d\psi/ds = 1/\beta$. The dimension of A is the square root of length, reflecting the fact that the oscillation amplitude is modulated by the square root of the amplitude function. In addition to describing the envelope of the oscillation, β also plays the role of an 'instantaneous' $\lambda\!\!\!^{-}$. The wavelength of a betatron oscillation may be some tens of meters, and so typically values of the amplitude function are of the order of meters rather than on the order of the beam size. The beam optics arrangement generally has some periodicity and the amplitude function is chosen to reflect that periodicity. As noted above, a small value of the amplitude function is desired at the interaction point, and so the focussing optics is tailored in its neighborhood to provide a suitable β^*.

The number of betatron oscillations per turn in a synchrotron is called the *tune* and is given by

$$\nu = \frac{1}{2\pi} \oint \frac{ds}{\beta} \ . \tag{25.10}$$

Expressing the integration constant A in the solution above in terms of $x,\ x'$ yields the *Courant-Snyder invariant*

$$A^2 = \gamma(s)\, x(s)^2 + 2\alpha(s)\, x(s)\, x'(s) + \beta(s)\, x'(s)^2$$

where

$$\alpha \equiv -\,\beta'/2, \ \ \gamma \equiv \frac{1+\alpha^2}{\beta} \ . \tag{25.11}$$

(The Courant-Snyder parameters α, β and γ employ three Greek letters which have other meanings and the significance at hand must often be recognized from context.) Because β is a function of position in the focussing structure, this ellipse changes orientation and aspect ratio from location to location but the area πA^2 remains the same.

As noted above the transverse emittance is a measure of the area in $x,\ x'$ (or $y,\ y'$) phase space occupied by an ensemble of particles. The definition used in Eq. (25.3) is the area that encloses 39% of a Gaussian beam.

For electron synchrotrons the equilibrium emittance results from the balance between synchrotron radiation damping and excitation from quantum fluctuations in the radiation rate. The equilibrium is reached in a time small compared with the storage time.

For present-day hadron synchrotrons, synchrotron radiation does not play a similar role in determining the transverse emittance. Rather the emittance during storage reflects the source properties and the abuse suffered by the particles throughout acceleration and storage. Nevertheless it is useful to argue as follows: Though x' and x can serve as canonically conjugate variables at constant energy this definition of the emittance would not be an adiabatic invariant when the energy changes during the acceleration cycle. However, $\gamma(v/c)x'$, where here γ is the Lorentz factor, is proportional to the transverse momentum and so qualifies as a variable conjugate to x. So often one sees a normalized emittance defined according to

$$\epsilon_N = \gamma\, \frac{v}{c}\, \epsilon, \tag{25.12}$$

which is an approximate adiabatic invariant, e.g. during acceleration.

25.2.2. ***Phase stability*** : The particles in a circular collider also undergo synchrotron oscillations. This is usually referred to as motion in the *longitudinal* degree-of-freedom because particles arrive at a particular position along the accelerator earlier or later than an ideal reference particle. This circumstance results in a finite bunch length, which is related to an energy spread.

For dynamical variables in longitudinal phase space, let us take ΔE and Δt, where these are the energy and time differences from that of the ideal particle. A positive Δt means a particle is behind the ideal particle. The equation of motion is the same as that

for a physical pendulum and therefore is nonlinear. But for small oscillations, it reduces to a simple harmonic oscillator:

$$\frac{d^2\Delta t}{dn^2} = -(2\pi\nu_s)^2\Delta t \tag{25.13}$$

where the independent variable n is the turn number and ν_s is the number of synchrotron oscillations per turn, analogous to the betatron oscillation tune defined earlier. Implicit in this equation is the approximation that n is a continuos variable. This approximation is valid provided $\nu_s \ll 1$, which is usually well satisfied in practice.

In the high-energy limit, where $v/c \approx 1$,

$$\nu_s = \left[\frac{h\eta\, eV\, \cos\phi_s}{2\pi E}\right]^{1/2} \quad . \tag{25.14}$$

There are four as yet undefined quantities in this expression: the harmonic number h, the slip factor η, the maximum energy eV gain per turn from the acceleration system, and the synchronous phase ϕ_s. The frequency of the RF system is normally a relatively high multiple, h, of the orbit frequency. The slip factor relates the fractional change in the orbit period τ to changes in energy according to

$$\frac{\Delta\tau}{\tau} = \eta\frac{\Delta E}{E} \quad . \tag{25.15}$$

At sufficiently high energy, the slip factor just reflects the relationship between path length and energy, since the speed is a constant; η is positive for all the synchrotrons in the "Tables of Collider Parameters" (Sec. 26).

The synchronous phase is a measure of how far up on the RF wave the average particle must ride in order to maintain constant energy to counteract of synchrotron radiation. That is, $\sin\phi_s$ is the ratio of the energy loss per turn to the maximum energy per turn that can be provided by the acceleration system. For hadron colliders built to date, $\sin\phi_s$ is effectively zero. This is not the case for electron storage rings; for example, the electron ring of HERA runs at a synchronous phase of 45°.

Now if one has a synchrotron oscillation with amplitudes $\widehat{\Delta t}$ and $\widehat{\Delta E}$,

$$\Delta t = \widehat{\Delta t}\,\sin(2\pi\nu_s n) \ , \qquad \Delta E = \widehat{\Delta E}\,\cos(2\pi\nu_s n) \tag{25.16}$$

then the amplitudes are related according to

$$\widehat{\Delta E} = \frac{2\pi\nu_s E}{\eta\tau}\widehat{\Delta t} \ . \tag{25.17}$$

The longitudinal emittance ϵ_ℓ may be defined as the phase space area bounded by particles with amplitudes $\widehat{\Delta t}$ and $\widehat{\Delta E}$. In general, the longitudinal emittance for a given

amplitude is found by numerical integration. For $\sin\phi_s = 0$, an analytical expression is as follows:

$$\epsilon_\ell = \left[\frac{2\pi^3 EeVh}{\tau^2\eta}\right]^{1/2} (\widehat{\Delta t})^2 \tag{25.18}$$

Again, a Gaussian is a reasonable representation of the longitudinal profile of a well-behaved beam bunch; if $\sigma_{\Delta t}$ is the standard deviation of the time distribution, then the bunch length can be characterized by

$$\ell = c\,\sigma_{\Delta t} \ . \tag{25.19}$$

In the electron case the longitudinal emittance is determined by the synchrotron radiation process just as in the transverse degrees of freedom. For the hadron case the history of acceleration plays a role and because energy and time are conjugate coordinates, the longitudinal emittance is a quasi-invariant.

For HEP bunch length is a significant quantity because if the bunch length becomes larger than β^* the luminosity is adversely affected. This is because β grows parabolically as one proceeds away from the IP and so the beam size increases thus lowering the contribution to the luminosity from such locations. Further discussion and modified expressions for the luminosity may be found in Furman and Zisman [3].

25.2.3. ***Synchrotron radiation*** [4] : A relativistic particle undergoing centripetal acceleration radiates at a rate given by the Larmor formula multiplied by the 4th power of the Lorentz factor:

$$P = \frac{1}{6\pi\epsilon_0}\frac{e^2a^2}{c^3}\gamma^4. \tag{25.20}$$

Here, $a = v^2/\rho$ is the centripetal acceleration of a particle with speed v undergoing deflection with radius of curvature ρ. In a synchrotron that has a constant radius of curvature within bending magnets, the energy lost due to synchrotron radiation per turn is the above multiplied by the time spent in bending magnets, $2\pi\rho/v$. Expressed in familiar units, this result may be written

$$W = 8.85\times10^{-5}E^4/\rho \ \text{ MeV per turn} \tag{25.21}$$

for electrons at sufficiently high energy that $v \approx c$. The energy E is in GeV and ρ is in kilometers. The radiation has a broad energy spectrum which falls off rapidly above the *critical energy*, $E_c = (3c/2\rho)\hbar\gamma^3$. Typically, E_c is in the hard x-ray region.

The characteristic time for synchrotron radiation processes is the time during which the energy must be replenished by the acceleration system. If f_0 is the orbit frequency, then the characteristic time is given by

$$\tau_0 = \frac{E}{f_0 W} \ . \tag{25.22}$$

Oscillations in each of the three degrees of freedom either damp or antidamp depending on the design of the accelerator. For a simple separated-function alternating gradient

synchrotron, all three modes damp. The damping time constants are related by Robinson's Theorem, which, expressed in terms of τ_0, is

$$\frac{1}{\tau_x} + \frac{1}{\tau_y} + \frac{1}{\tau_s} = 2\frac{1}{\tau_0} \ . \tag{25.23}$$

Even though all three modes may damp, the emittances do not tend toward zero. Statistical fluctuations in the radiation rate excite synchrotron oscillations and radial betatron oscillations. Thus there is an equilibrium emittance at which the damping and excitation are in balance. The vertical emittance is non-zero due to horizontal-vertical coupling.

Polarization can develop from an initially unpolarized beam as a result of synchrotron radiation. A small fraction $\approx E_c/E$ of the radiated power flips the electron spin. Because the lower energy state is that in which the particle magnetic moment points in the same direction as the magnetic bend field, the transition rate toward this alignment is larger than the rate toward the reverse orientation. An equilibrium polarization of 92% is predicted, and despite a variety of depolarizing processes, polarization above 80% has been observed at a number of facilities.

The radiation rate for protons is of course down by a factor of the fourth power of the mass ratio, and is given by

$$W = 7.8 \times 10^{-3} E^4/\rho \ \text{ keV per turn} \tag{25.24}$$

where E is now in TeV and ρ in km. For the LHC, synchrotron radiation presents a significant load to the cryogenic system, and impacts magnet design due to gas desorption and secondary electron emission from the wall of the cold beam tube. The critical energy for the LHC is 44 eV.

25.2.4. ***Beam-beam tune shift*** [5] : In a bunch-bunch collision the particles of one bunch see the other bunch as a nonlinear lens. Therefore the focussing properties of the ring are changed in a way that depends on the transverse oscillation amplitude. Hence there is a spread in the frequency of betatron oscillations.

There is an extensive literature on the subject of how large this tune spread can be. In practice, the limiting value is hard to predict. It is consistently larger for electrons because of the beneficial effects of damping from synchrotron radiation.

In order that contributions to the total tune spread arise only at the detector locations, the beams in a multibunch collider are kept apart elsewhere in the collider by a variety of techniques. For equal energy particles of opposite charge circulating in the same vacuum chamber, electrostatic separators may be used assisted by a crossing angle if appropriate. For particles of equal energy and of the same charge, a crossing angle is needed not only for tune spread reasons but to steer the particles into two separate beam pipes. In HERA, because of the large ratio of proton to electron energy, separation can be achieved by bending magnets.

25.2.5. ***Luminosity lifetime*** : In electron synchrotrons the luminosity degrades during the store primarily due to particles leaving the phase stable region in longitudinal phase space as a result of quantum fluctuations in the radiation rate and bremsstrahlung. For hadron colliders the luminosity deteriorates due to emittance dilution resulting from a variety of processes. In practice, stores are intentionally terminated when the luminosity drops to the point where a refill will improve the integrated luminosity, while synchrotron radiation facilitates a "topping-up" process in electron-positron rings to provide continuous luminosity.

25.2.6. ***Luminosity: Achieved and Desired*** :

Many years ago, the CERN Intersecting Storage Rings set an enviable luminosity record for proton-proton collisions at 1.3×10^{32} $\mathrm{cm}^{-2}\mathrm{s}^{-1}$. Not until the Summer of 2005 was this level matched in a proton-antiproton collider as the Tevatron approached an integrated luminosity of 1 fb^{-1} in its Run II. Thus far, electrons have proved more amenable; KEKB has reached 1.5×10^{34} $\mathrm{cm}^{-2}\mathrm{s}^{-1}$, with the implication of integrated luminosity in the 100 $\mathrm{fb}^{-1}\mathrm{yr}^{-1}$ range.

The luminosity specification for a yet-to-be-built system is a complicated process, involving considerations of organization, funding, politics and so forth outside the scope of this paper. In the electron world, the 1/s dependence of cross sections such as "Higgs-strahlung" plays an important role. In the complex hadron environment, the higher luminosity potential of the LHC helped offset the otherwise totally adverse consequences of the SSC project cancellation. As a historical observation, it is interesting that although improvement directions of present and past facilities could not spelled out in advance, such opportunities have proved to be of critical value to the advance of HEP.

25.3. Prospects

While this update is in preparation, it is interesting to recall that exactly 20 years ago, the Tevatron was entering its commissioning phase as a proton-antiproton collider to provide a c.m.s. energy at the 2 TeV level. Now, the next major step in discovery reach is approaching. The LHC is scheduled to begin operation in 2007, with its proton beams colliding at 14 TeV c.m.s., and at a luminosity two orders of magnitude above that reached in the Tevatron. LHC progress may be followed at the website provided by CERN [6].

A concerted international effort is underway aimed toward the construction of the electron-positron counterpart of the LHC – a linear collider to operate at the 0.5–1.0 TeV c.m.s. level. A recent decision of an International Technical Review Panel decided in favor of the superconducting RF approach for the main linacs [7]. A current goal is to develop the design report by the end of 2006.

References:

1. E. D. Courant and H. S. Snyder, Ann. Phys. **3**, 1 (1958). This is the classic article on the alternating gradient synchrotron.
2. A. W. Chao and M. Tigner (eds.), *Handbook of Accelerator Physics and Engineering*, World Science Publishing Co. (Singapore, 2nd printing, 2002.), Sec. 2.1.

3. M. A. Furman and M. S. Zisman, *Handbook of Accelerator Physics and Engineering*, *op cit*, Sec. 4.1.
4. H. Wiedemann, *Handbook of Accelerator Physics and Engineering*, *op cit*, Sec. 3.1.
5. K. Hirata, P. Chen, J. M. Jowett, *Handbook of Accelerator Physics and Engineering*, *op cit*, Secs. 2.6.1, 2.6.2, 2.6.3, respectively.
6. http://lhc-new-homepage.web.cern.ch/lhc-new-homepage/.
7. http://www.interactions.org/linearcollider/gde/.

HIGH-ENERGY COLLIDER PARAMETERS: e^+e^- Colliders (I)

Updated in early 2006 with numbers received from representatives of the colliders (contact J. Beringer, LBNL). For existing (future) colliders the latest achieved (design) values are given. Quantities are, where appropriate, r.m.s.; H and V indicate horizontal and vertical directions; s.c. stands for superconducting. Parameters for the defunct SPEAR, DORIS, PETRA, PEP, SLC, TRISTAN, and VEPP-2M colliders may be found in our 1996 edition (Phys. Rev. **D54**, 1 July 1996, Part I).

	VEPP-2000 (Novosibirsk)	VEPP-4M (Novosibirsk)	BEPC (China)	BEPC-II (China)	DAΦNE (Frascati)
Physics start date	2006	1994	1989	2007	1999
Physics end date	—	—	2005	—	2008
Maximum beam energy (GeV)	1.0	6	2.2	1.89 (2.1 max)	0.700
Luminosity (10^{30} cm^{-2}s^{-1})	100	20	12.6 at 1.843 GeV/beam 5 at 1.55 GeV/beam	1000	150 (300 achievable)
Time between collisions (μs)	0.04	0.6	0.8	0.008	0.0027
Crossing angle (μ rad)	0	0	0	1.1×10^4	$(2.5 \text{ to } 3.2)\times 10^4$
Energy spread (units 10^{-3})	0.64	1	0.58 at 2.2 GeV	0.52	0.40
Bunch length (cm)	4	5	≈ 5	1.3	low current: 1 high current: 3
Beam radius (10^{-6} m)	125 (round)	H: 1000 V: 30	H: 890 V: 37	H: 380 V: 5.7	H: 800 V: 4.8
Free space at interaction point (m)	± 1	± 2	± 2.15	± 0.63	± 0.40
Luminosity lifetime (hr)	continuous	2	7–12	1.5	0.7
Filling time (min)	continuous	15	30	26	0.8 (topping up)
Acceleration period (s)	—	150	120	—	on energy
Injection energy (GeV)	0.2–1.0	1.8	1.55	1.89	on energy
Transverse emittance ($10^{-9}\pi$ rad-m)	H: 250 V: 250	H: 200 V: 20	H: 660 V: 28	H: 144 V: 2.2	H: 300 V: 1
β^*, amplitude function at interaction point (m)	H: 0.06 – 0.11 V: 0.06 – 0.10	H: 0.75 V: 0.05	H: 1.2 V: 0.05	H: 1.0 V: 0.015	H: 1.7 V: 0.0017
Beam-beam tune shift per crossing (units 10^{-4})	H: 750 V: 750	500	350	400	250
RF frequency (MHz)	172	180	199.53	499.8	356
Particles per bunch (units 10^{10})	16	15	20 at 2 GeV 11 at 1.55 GeV	4.8	e^-: 3.3 e^+: 2.4
Bunches per ring per species	1	2	1	93	120 (incl. 10 bunch gap)
Average beam current per species (mA)	150	80	40 at 2 GeV 22 at 1.55 GeV	910	e^-: 1800 e^+: 1300
Circumference or length (km)	0.024	0.366	0.2404	0.23753	0.098
Interaction regions	2	1	2	1	2
Utility insertions	2	1	4	5	2
Magnetic length of dipole (m)	1.2	2	1.6	Outer ring: 1.6 Inner ring: 1.41	1
Length of standard cell (m)	12	7.2	6.6	Outer ring: 6.6 Inner ring: 6.2	12
Phase advance per cell (deg)	H: 738 V: 378	65	≈ 60	60–90 no standard cell	360
Dipoles in ring	8	78	40 + 4 weak	84 + 8 weak	8
Quadrupoles in ring	20	150	68	134+2 s.c.	48
Peak magnetic field (T)	2.4	0.6	0.9028 at 2.8 GeV	Outer ring: 0.67712 Inner ring: 0.76636	1.7

HIGH-ENERGY COLLIDER PARAMETERS: e^+e^- Colliders (II)

Updated in early 2006 with numbers received from representatives of the colliders (contact J. Beringer, LBNL). For existing (future) colliders the latest achieved (design) values are given. Quantities are, where appropriate, r.m.s.; H and V indicate horizontal and vertical directions; s.c. stands for superconducting.

	CESR (Cornell)	CESR-C (Cornell)	KEKB (KEK)	PEP-II (SLAC)	LEP (CERN)
Physics start date	1979	2002	1999	1999	1989
Physics end date	2002	—	—	—	2000
Maximum beam energy (GeV)	6	6	$e^- \times e^+$: 8×3.5	e^-: 7–12 (9.0 nominal) e^+: 2.5–4 (3.1 nominal) (nominal $E_{\rm cm} = 10.5$ GeV)	100 - 104.6
Luminosity (10^{30} cm^{-2}s^{-1})	1280 at 5.3 GeV/beam	60 at 1.9 GeV/beam	16270	10025	24 at Z^0 100 at > 90 GeV
Time between collisions (μs)	0.014 to 0.22	0.014 to 0.22	0.00590 or 0.00786	0.0042	22
Crossing angle (μ rad)	± 2000	± 3500	$\pm 11,000$	0	0
Energy spread (units 10^{-3})	0.6 at 5.3 GeV/beam	0.8 at 1.9 GeV/beam	0.7	e^-/e^+: 0.61/0.77	0.7→1.5
Bunch length (cm)	1.8	1.2	0.65	e^-/e^+: 1.1/1.0	1.0
Beam radius (μm)	H: 460 V: 4	H: 340 V: 5.7	H: 110 V: 2.4	H: 157 V: 4.7	H: 200 → 300 V: 2.5 → 8
Free space at interaction point (m)	± 2.2 (± 0.6 to REC quads)	± 2.2 (± 0.3 to PM quads)	$+0.75/-0.58$ ($+300/-500$) mrad cone	± 0.2, ± 300 mrad cone	± 3.5
Luminosity lifetime (hr)	2–3	2–3	continuous	continuous	20 at Z^0 10 at > 90 GeV
Filling time (min)	5 (topping up)	4 (topping up)	continuous	continuous	20 to setup 20 to accumulate
Acceleration period (s)	—	—	—	—	600
Injection energy (GeV)	1.8–6	1.5–6	e^-/e^+ : 8/3.5	2.5–12	22
Transverse emittance (π rad-nm)	H: 210 V: 1	H: 130 V: 1.0	e^-: 24 (H), 0.71 (V) e^+: 18 (H), 0.68 (V)	e^-: 48 (H), 1.5 (V) e^+: 24 (H), 1.5 (V)	H: 20–45 V: 0.25 → 1
β^*, amplitude function at interaction point (m)	H: 1.0 V: 0.018	H: 0.90 V: 0.013	e^-: 0.56 (H), 0.0062 (V) e^+: 0.59 (H), 0.0065 (V)	e^-: 0.50 (H), 0.012 (V) e^+: 0.50 (H), 0.012 (V)	H: 1.5 V: 0.05
Beam-beam tune shift per crossing (units 10^{-4})	H: 250 V: 620	e^-: 270 (H), 380 (V) e^+: 180 (H), 260 (V)	e^-: 730 (H), 550 (V) e^+: 1170 (H), 960 (V)	e^-: 400 (H), 400 (V) e^+: 990 (H), 800 (V)	830
RF frequency (MHz)	500	500	508.887	476	352.2
Particles per bunch (units 10^{10})	1.15	1.15	e^-/e^+: 6.0/8.0	e^-/e^+: 4.6/7.8	45 in collision 60 in single beam
Bunches per ring per species	9 trains of 5 bunches	8 trains of 5 bunches	1388	1732	4 trains of 1 or 2
Average beam current per species (mA)	340	62	e^-/e^+: 1350/1720	e^-/e^+: 1750/2950	4 at Z^0 4→6 at > 90 GeV
Beam polarization (%)	—	—	—	—	55 at 45 GeV 5 at 61 GeV
Circumference or length (km)	0.768	0.768	3.016	2.2	26.66
Interaction regions	1	1	1	1 (2 possible)	4
Utility insertions	3	3	3 per ring	5	4
Magnetic length of dipole (m)	1.6–6.6	1.6–6.6	e^-/e^+ : 5.86/0.915	e^-/e^+: 5.4/0.45	11.66/pair
Length of standard cell (m)	16	16	e^-/e^+ : 75.7/76.1	15.2	79
Phase advance per cell (deg)	45–90 (no standard cell)	45–90 (no standard cell)	450	e^-/e^+: 60/90	102/90
Dipoles in ring	86	84	e^-/e^+ : 116/112	e^-/e^+: 192/192	3280+24 inj. + 64 weak
Quadrupoles in ring	101 + 4 s.c.	101 + 4 s.c.	e^-/e^+ : 452/452	e^-/e^+: 290/326	520+288 + 8 s.c.
Peak magnetic field (T)	0.3 normal, 0.8 high field } at 8 GeV	0.3 normal, 0.8 high field } at 8 GeV 2.1 wigglers at 1.9 GeV	e^-/e^+ : 0.25/0.72	e^-/e^+: 0.18/0.75	0.135

HIGH-ENERGY COLLIDER PARAMETERS: ep, $\overline{p}p$, pp, and Heavy Ion Colliders

Updated in early 2006 with numbers received from representatives of the colliders (contact J. Beringer, LBNL). For existing (future) colliders the latest achieved (design) values are given. Quantities are, where appropriate, r.m.s.; H and V indicate horizontal and vertical directions; s.c. stands for superconducting.

	HERA (DESY)	TEVATRON (Fermilab)	RHIC (Brookhaven)				LHC (CERN)	
Physics start date	1992	1987	2001	2000	2004	2002	2007	2008
Physics end date	—	—	—				—	
Particles collided	ep	$p\overline{p}$	pp (pol.)	Au Au	Cu Cu	d Au	pp	Pb Pb
Maximum beam energy (TeV)	e: 0.030 p: 0.92	0.980	0.1 46% pol	0.1 TeV/n	0.1 TeV/n	0.1 TeV/n	7.0	2.76 TeV/n
Luminosity (10^{30} cm^{-2}s^{-1})	75	171	10	0.0015	0.020	0.07	1.0×10^4	0.001
Time between collisions (ns)	96	396	107	213	213	213	24.95	99.8
Crossing angle (μ rad)	0	0	0				300	≤ 100
Energy spread (units 10^{-3})	e: 0.91 p: 0.2	0.14	0.2	0.5	0.5	0.5	0.113	0.11
Bunch length (cm)	e: 0.83 p: 8.5	p: 50 $\bar{p}$: 45	40	20	20	20	7.55	7.94
Beam radius (10^{-6} m)	e: $280(H), 50(V)$ p: $265(H), 50(V)$	p: 29 $\bar{p}$: 21	175 (β^*=1 m)	150 (β^*=1m)	145 (β^*=0.9m)	215 (β^*=2m)	16.6	15.9
Free space at interaction point (m)	± 2	± 6.5	16				38	38
Luminosity lifetime (hr)	10	7 (average, start of store)	10	3	5	6	14.9	10.9 - 3.6 (initial, 1 - 3 exp.)
Filling time (min)	e: 60, p: 120 (includes magnet cycling, accel.)	30	30	15	15	15	8.6 (both beams)	
Acceleration period (s)	e: 200 p: 1500	86	220	280	290	230	1200	
Injection energy (TeV)	e: 0.012 p: 0.040	0.15	0.023	0.011 TeV/n	0.011 TeV/n	0.012 TeV/n	0.450	0.1774 TeV/n
Transverse emittance ($10^{-9}\pi$ rad-m)	e: $20(H), 3.5(V)$ p: $5(H), 5(V)$	p: 3 $\bar{p}$: 1.5	31	23	23	23	0.5	0.5
β^*, ampl. function at interaction point (m)	e: $0.6(H), 0.26(V)$ p: $2.45(H), 0.18(V)$	0.28	1–10	1–5	1–5	2–5	0.55	0.5
Beam-beam tune shift per crossing (units 10^{-4})	e: $190(H), 450(V)$ p: $12(H), 9(V)$	p: 50 $\bar{p}$: 100	45	17	29	d: 27 Au: 21	34	—
RF frequency (MHz)	e: 499.7 p: 208.2/52.05	53	accel: 28 store: 28	accel: 28 store: 197	accel: 28 store: 197	accel: 28 store: 197	400.8	400.8
Particles per bunch (units 10^{10})	e: 3 p: 7	p: 24 $\bar{p}$: 6	9	0.11	0.38	d: 11 Au: 0.07	11.5	0.007
Bunches per ring per species	e: 189 p: 180	36	106	45	37	55	2808	592
Average beam current per species (mA)	e: 40 p: 90	p: 66 $\bar{p}$: 16	119	49	60	d: 77 Au: 38	584	6.12
Circumference (km)	6.336	6.28	3.834				26.659	
Interaction regions	2 colliding beams 1 fixed target (e beam)	2 high $\mathscr{L}$	6 total, 2 high $\mathscr{L}$				2 high $\mathscr{L}$ +1	1 dedicated +2
Utility insertions	4	4	13/ring				4	
Magnetic length of dipole (m)	e: 9.185 p: 8.82	6.12	9.45				14.3	
Length of standard cell (m)	e: 23.5 p: 47	59.5	29.7				106.90	
Phase advance per cell (deg)	e: 60 p: 90	67.8	84				90	
Dipoles in ring	e: 396 p: 416	774	192 per ring + 12 common				1232 main dipoles	
Quadrupoles in ring	e: 580 p: 280	216	246 per ring				482 2-in-1 24 1-in-1	
Magnet type	e: C-shaped p: s.c., collared, cold iron	s.c. $\cos\theta$ warm iron	s.c. $\cos\theta$ cold iron				s.c. 2 in 1 cold iron	
Peak magnetic field (T)	e: 0.274 p: 5	4.4	3.5				8.3	
$\overline{p}$ source accum. rate (hr^{-1})	—	16×10^{10}	—				—	
Max. no. $\overline{p}$ in accum. ring	—	2.4×10^{12}	—				—	

27. PASSAGE OF PARTICLES THROUGH MATTER

Revised April 2006 by H. Bichsel (University of Washington), D.E. Groom (LBNL), and S.R. Klein (LBNL).

27.1. Notation

Table 27.1: Summary of variables used in this section. The kinematic variables β and γ have their usual meanings.

Symbol	Definition	Units or Value
α	Fine structure constant $(e^2/4\pi\epsilon_0\hbar c)$	$1/137.035\,999\,11(46)$
M	Incident particle mass	MeV/c^2
E	Incident particle energy γMc^2	MeV
T	Kinetic energy	MeV
$m_e c^2$	Electron mass $\times\ c^2$	$0.510\,998\,918(44)$ MeV
r_e	Classical electron radius $e^2/4\pi\epsilon_0 m_e c^2$	$2.817\,940\,325(28)$ fm
N_A	Avogadro's number	$6.022\,1415(10) \times 10^{23}$ mol^{-1}
ze	Charge of incident particle	
Z	Atomic number of absorber	
A	Atomic mass of absorber	g mol^{-1}
K/A	$4\pi N_A r_e^2 m_e c^2/A$	$0.307\,075$ MeV g^{-1} cm^2 for A = 1 g mol^{-1}
I	Mean excitation energy	eV (*Nota bene!*)
$\delta(\beta\gamma)$	Density effect correction to ionization energy loss	
$\hbar\omega_p$	Plasma energy $(\sqrt{4\pi N_e r_e^3}\ m_e c^2/\alpha)$	$28.816\sqrt{\rho\langle Z/A\rangle}$ eV$^{(a)}$
N_c	Electron density	(units of r_e)$^{-3}$
w_j	Weight fraction of the jth element in a compound or mixture	
n_j	$\propto$ number of jth kind of atoms in a compound or mixture	
—	$4\alpha r_e^2 N_A/A$	$(716.408$ g cm$^{-2})^{-1}$ for $A = 1$ g mol^{-1}
X_0	Radiation length	g cm^{-2}
E_c	Critical energy for electrons	MeV
$E_{\mu c}$	Critical energy for muons	GeV
E_s	Scale energy $\sqrt{4\pi/\alpha}\ m_e c^2$	21.2052 MeV
R_M	Molière radius	g cm^{-2}

$^{(a)}$ For ρ in g cm^{-3}.

27.2. Electronic energy loss by heavy particles [1–22, 24–30, 82]

Moderately relativistic charged particles other than electrons lose energy in matter primarily by ionization and atomic excitation. The mean rate of energy loss (or stopping power) is given by the Bethe-Bloch equation,

$$-\frac{dE}{dx} = Kz^2\frac{Z}{A}\frac{1}{\beta^2}\left[\frac{1}{2}\ln\frac{2m_ec^2\beta^2\gamma^2T_{\max}}{I^2} - \beta^2 - \frac{\delta(\beta\gamma)}{2}\right] . \qquad (27.1)$$

Here $T_{\max}$ is the maximum kinetic energy which can be imparted to a free electron in a single collision, and the other variables are defined in Table 27.1. With K as defined in Table 27.1 and A in g mol^{-1}, the units are MeV g^{-1}cm^2.

In this form, the Bethe-Bloch equation describes the energy loss of pions in a material such as copper to about 1% accuracy for energies between about 6 MeV and 6 GeV (momenta between about 40 MeV/c and 6 GeV/c). At lower energies various corrections

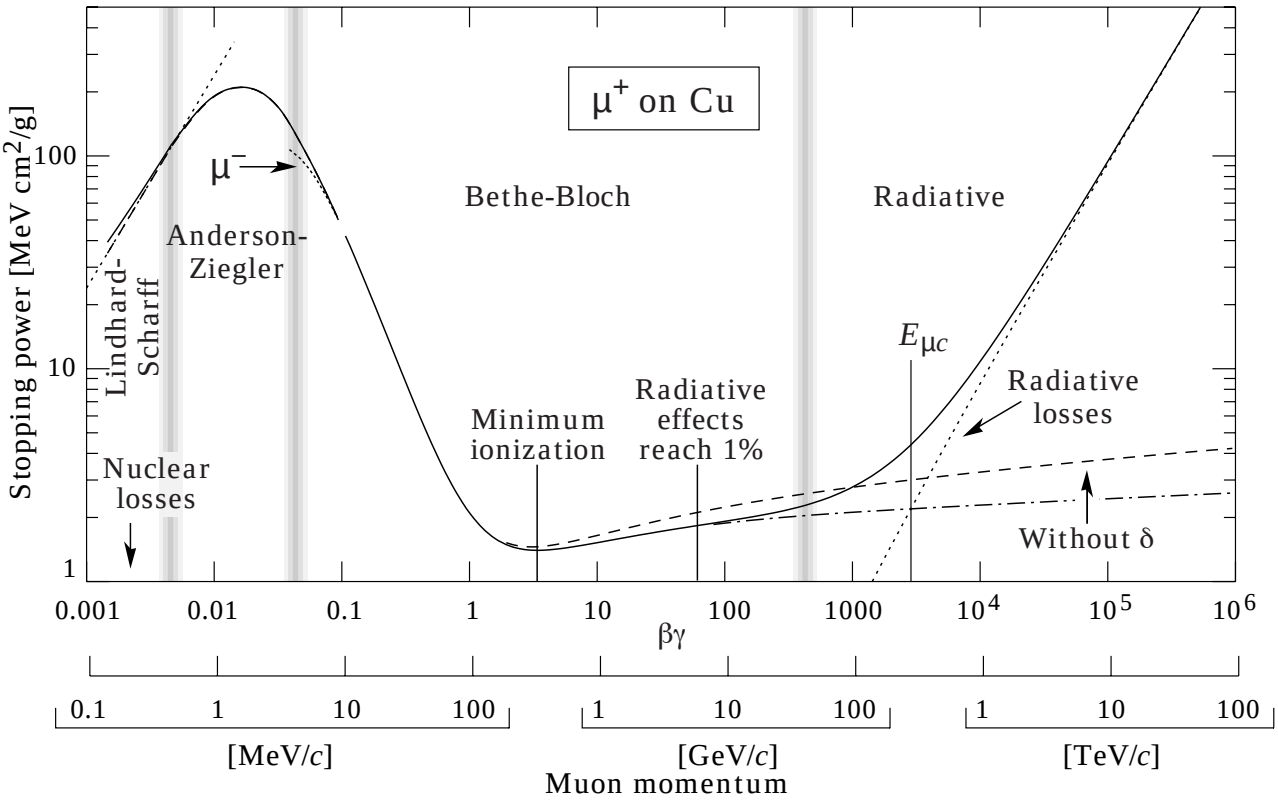


Fig. 27.1: Stopping power ($= \langle -dE/dx\rangle$) for positive muons in copper as a function of $\beta\gamma = p/Mc$ over nine orders of magnitude in momentum (12 orders of magnitude in kinetic energy). Solid curves indicate the total stopping power. Data below the break at $\beta\gamma \approx 0.1$ are taken from ICRU 49 [2], and data at higher energies are from Ref. 1. Vertical bands indicate boundaries between different approximations discussed in the text. The short dotted lines labeled "μ^-" illustrate the "Barkas effect," the dependence of stopping power on projectile charge at very low energies [3].

discussed in Sec. 27.2.1 must be made. At higher energies, radiative effects begin to be important. These limits of validity depend on both the effective atomic number of the absorber and the mass of the slowing particle.

The function as computed for muons on copper is shown by the solid curve in Fig. 27.1, and for pions on other materials in Fig. 27.3. A minor dependence on M at the highest energies is introduced through $T_{\rm max}$, but for all practical purposes in high-energy physics dE/dx in a given material is a function only of β. Except in hydrogen, particles of the same velocity have similar rates of energy loss in different materials; there is a slow decrease in the rate of energy loss with increasing Z. The qualitative difference in stopping power behavior at high energies between a gas (He) and the other materials shown in Fig. 27.3 is due to the density-effect correction, $\delta(\beta\gamma)$, discussed below. The stopping power functions are characterized by broad minima whose position drops from $\beta\gamma = 3.5$ to 3.0 as Z goes from 7 to 100. The values of minimum ionization as a function of atomic number are shown in Fig. 27.2.

In practical cases, most relativistic particles (*e.g.*, cosmic-ray muons) have mean energy loss rates close to the minimum, and are said to be minimum ionizing particles, or mip's.

As discussed below, the most probable energy loss in a detector is considerably below the mean given by the Bethe-Bloch equation.

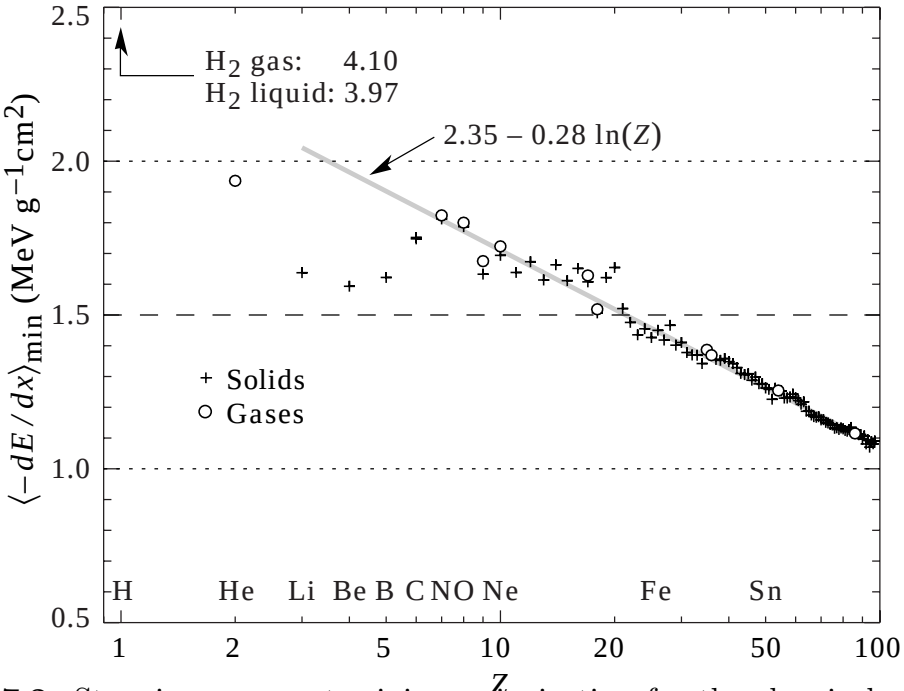


Figure 27.2: Stopping power at minimum ionization for the chemical elements. The straight line is fitted for $Z > 6$. A simple functional dependence on Z is not to be expected, since $\langle -dE/dx \rangle$ also depends on other variables.

Eq. (27.1) may be integrated to find the total (or partial) "continuous slowing-down approximation" (CSDA) range R for a particle which loses energy only through ionization

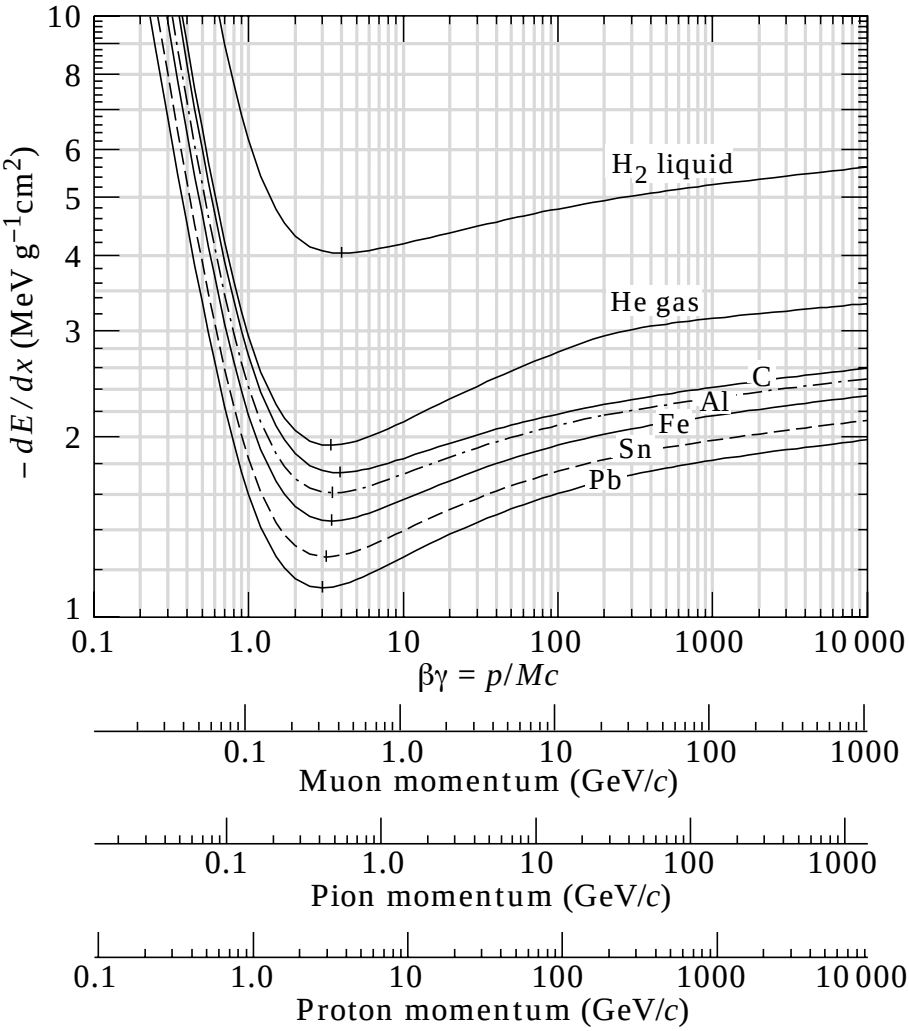


Figure 27.3: Mean energy loss rate in liquid (bubble chamber) hydrogen, gaseous helium, carbon, aluminum, iron, tin, and lead. Radiative effects, relevant for muons and pions, are not included. These become significant for muons in iron for $\beta\gamma \gtrsim 1000$, and at lower momenta for muons in higher-Z absorbers. See Fig. 27.21.

and atomic excitation. Since dE/dx depends only on β, R/M is a function of E/M or pc/M. In practice, range is a useful concept only for low-energy hadrons ($R \lesssim \lambda_I$, where λ_I is the nuclear interaction length), and for muons below a few hundred GeV (above which radiative effects dominate). R/M as a function of $\beta\gamma = p/Mc$ is shown for a variety of materials in Fig. 27.4.

The mass scaling of dE/dx and range is valid for the electronic losses described by the Bethe-Bloch equation, but not for radiative losses, relevant only for muons and pions.

For a particle with mass M and momentum $M\beta\gamma c$, $T_{\rm max}$ is given by

$$T_{\rm max} = \frac{2m_e c^2\,\beta^2\gamma^2}{1 + 2\gamma m_e/M + (m_e/M)^2}\ . \tag{27.2}$$

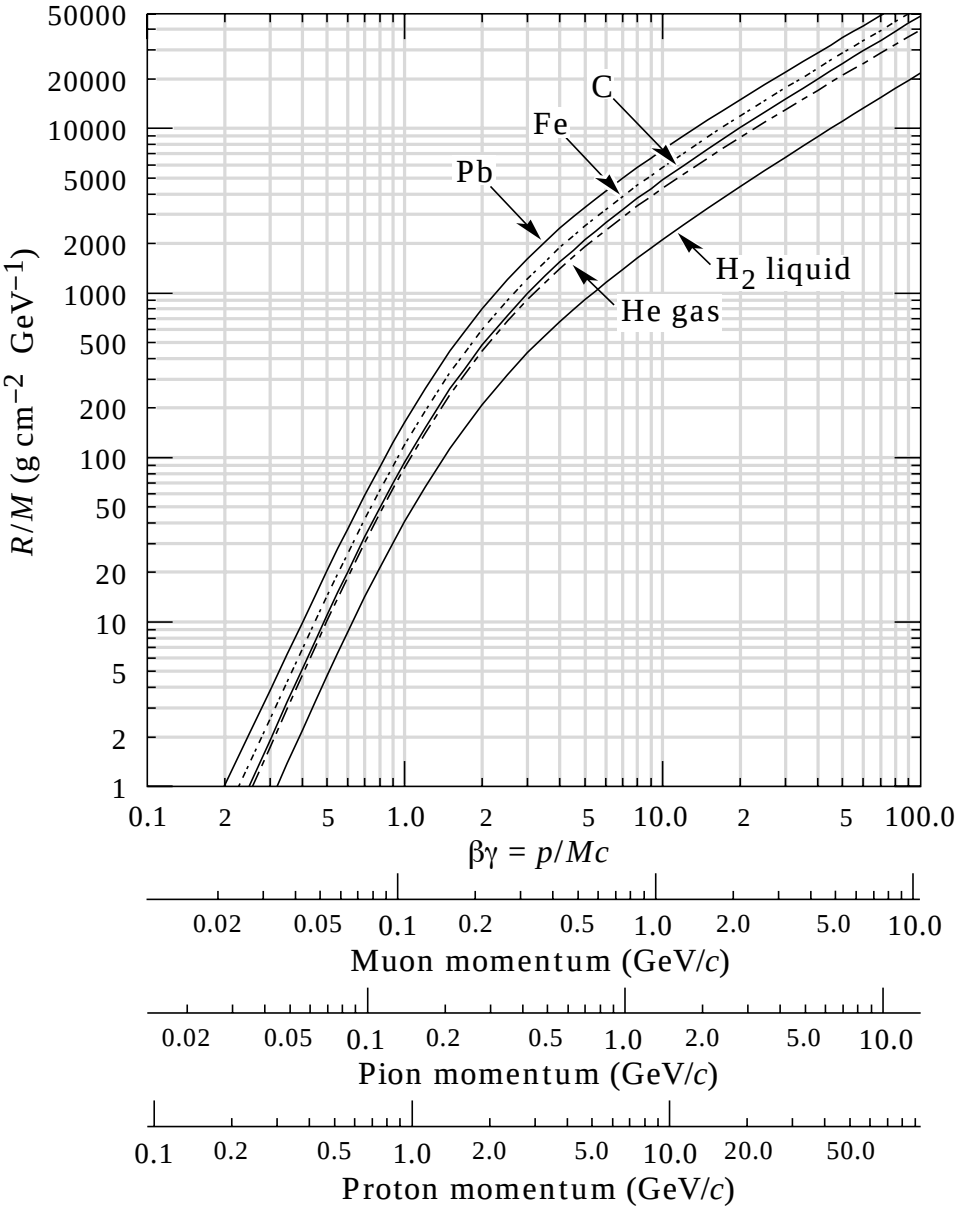


Figure 27.4: Range of heavy charged particles in liquid (bubble chamber) hydrogen, helium gas, carbon, iron, and lead. For example: For a K^+ whose momentum is 700 MeV/c, $\beta\gamma = 1.42$. For lead we read $R/M \approx 396$, and so the range is 195 g cm^{-2}.

In older references [4,5] the "low-energy" approximation $T_{\max} = 2m_e c^2 \beta^2\gamma^2$, valid for $2\gamma m_e/M \ll 1$, is often implicit. For a pion in copper, the error thus introduced into dE/dx is greater than 6% at 100 GeV.

At energies of order 100 GeV, the maximum 4-momentum transfer to the electron can exceed 1 GeV/c, where hadronic structure effects significantly modify the cross sections. This problem has been investigated by J.D. Jackson [6], who concluded that for hadrons (but not for large nuclei) corrections to dE/dx are negligible below energies where

radiative effects dominate. While the cross section for rare hard collisions is modified, the average stopping power, dominated by many softer collisions, is almost unchanged.

"The determination of the mean excitation energy is the principal non-trivial task in the evaluation of the Bethe stopping-power formula" [7]. Recommended values have varied substantially with time. Estimates based on experimental stopping-power measurements for protons, deuterons, and alpha particles and on oscillator-strength distributions and dielectric-response functions were given in ICRU 49 [2]. See also ICRU 37 [8]. These values, shown in Fig. 27.5, have since been widely used. Machine-readable versions can also be found [9]. These values are widely used.

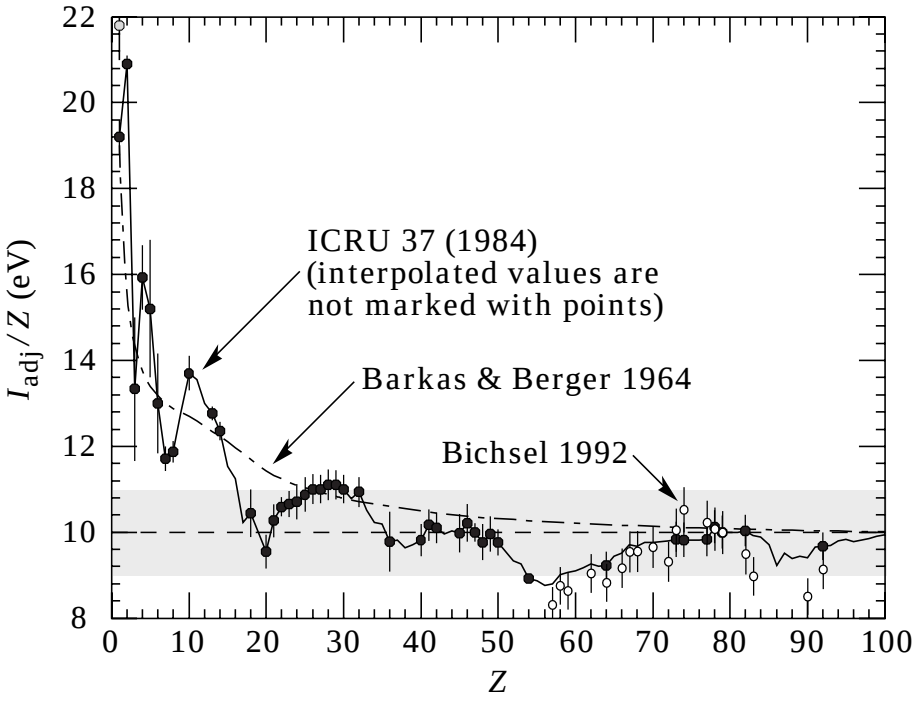


Figure 27.5: Mean excitation energies (divided by Z) as adopted by the ICRU [8]. Those based on experimental measurements are shown by symbols with error flags; the interpolated values are simply joined. The grey point is for liquid H_2; the black point at 19.2 eV is for H_2 gas. The open circles show more recent determinations by Bichsel [10]. The dotted curve is from the approximate formula of Barkas [11] used in early editions of this *Review*.

27.2.1. ***Energy loss at low energies*** : Shell corrections C/Z must be included in the square brackets of of Eq. (27.1) [2,8,10,11] to correct for atomic binding having been neglected in calculating some of the contributions to Eq. (27.1). The Barkas form [11] was used in generating Fig. 27.1. For copper it contributes about 1% at $\beta\gamma = 0.3$ (kinetic energy 6 MeV for a pion), and the correction decreases very rapidly with energy.

Eq. (27.1) is based on a first-order Born approximation. Higher-order corrections, again important only at lower energy, are normally included by adding the "Bloch correction" $z^2 L_2(\beta)$ inside the square brackets (Eq.(2.5) in [2]) .

An additional "Barkas correction" $zL_1(\beta)$ makes the stopping power for a negative particle somewhat smaller than for a positive particle with the same mass and velocity. In a 1956 paper, Barkas *et al.* noted that negative pions had a longer range than positive pions [3]. The effect has been measured for a number of negative/positive particle pairs, most recently for antiprotons at the CERN LEAR facility [12].

A detailed discussion of low-energy corrections to the Bethe formula is given in ICRU Report 49 [2]. When the corrections are properly included, the accuracy of the Bethe-Bloch treatment is accurate to about 1% down to $\beta \approx 0.05$, or about 1 MeV for protons.

For $0.01 < \beta < 0.05$, there is no satisfactory theory. For protons, one usually relies on the phenomenological fitting formulae developed by Andersen and Ziegler [2,13]. For particles moving more slowly than $\approx 0.01c$ (more or less the velocity of the outer atomic electrons), Lindhard has been quite successful in describing electronic stopping power, which is proportional to β [14]. Finally, we note that at low energies, *e.g.*, for protons of less than several hundred eV, non-ionizing nuclear recoil energy loss dominates the total energy loss [2,14,15].

As shown in ICRU 49 [2] (using data taken from Ref. 13), the nuclear plus electronic proton stopping power in copper is 113 MeV $\mathrm{cm}^2\,\mathrm{g}^{-1}$ at $T = 10$ keV, rises to a maximum of 210 MeV $\mathrm{cm}^2\,\mathrm{g}^{-1}$ at 100–150 keV, then falls to 120 MeV $\mathrm{cm}^2\,\mathrm{g}^{-1}$ at 1 MeV. Above 0.5–1.0 MeV the corrected Bethe-Bloch theory is adequate.

27.2.2. ***Density effect*** : As the particle energy increases, its electric field flattens and extends, so that the distant-collision contribution to Eq. (27.1) increases as $\ln\beta\gamma$. However, real media become polarized, limiting the field extension and effectively truncating this part of the logarithmic rise [4–5,16–18]. At very high energies,

$$\delta/2 \to \ln(\hbar\omega_p/I) + \ln\beta\gamma - 1/2\ , \tag{27.3}$$

where $\delta(\beta\gamma)/2$ is the density effect correction introduced in Eq. (27.1) and $\hbar\omega_p$ is the plasma energy defined in Table 27.1. A comparison with Eq. (27.1) shows that $|dE/dx|$ then grows as $\ln\beta\gamma$ rather than $\ln\beta^2\gamma^2$, and that the mean excitation energy I is replaced by the plasma energy $\hbar\omega_p$. The ionization stopping power as calculated with and without the density effect correction is shown in Fig. 27.1. Since the plasma frequency scales as the square root of the electron density, the correction is much larger for a liquid or solid than for a gas, as is illustrated by the examples in Fig. 27.3.

The density effect correction is usually computed using Sternheimer's parameterization [16]:

$$\delta(\beta\gamma) = \begin{cases} 2(\ln 10)x - \overline{C} & \text{if } x \geq x_1; \\ 2(\ln 10)x - \overline{C} + a(x_1 - x)^k & \text{if } x_0 \leq x < x_1; \\ 0 & \text{if } x < x_0 \text{ (nonconductors);} \\ \delta_0 10^{2(x-x_0)} & \text{if } x < x_0 \text{ (conductors)} \end{cases} \tag{27.4}$$

Here $x = \log_{10}\eta = \log_{10}(p/Mc)$. $\overline{C}$ (the negative of the C used in Ref. 16) is obtained by equating the high-energy case of Eq. (27.4) with the limit given in Eq. (27.3). The other parameters are adjusted to give a best fit to the results of detailed calculations for momenta below $Mc\exp(x_1)$. Parameters for elements and nearly 200 compounds and

mixtures of interest are published in a variety of places, notably in Ref. 18. A recipe for finding the coefficients for nontabulated materials is given by Sternheimer and Peierls [20], and is summarized in Ref. 1.

The remaining relativistic rise comes from the $\beta^2\gamma^2$ growth of $T_{\rm max}$, which in turn is due to (rare) large energy transfers to a few electrons. When these events are excluded, the energy deposit in an absorbing layer approaches a constant value, the Fermi plateau (see Sec. 27.2.4 below). At extreme energies (*e.g.*, > 332 GeV for muons in iron, and at a considerably higher energy for protons in iron), radiative effects are more important than ionization losses. These are especially relevant for high-energy muons, as discussed in Sec. 27.6.

27.2.3. ***Energetic knock-on electrons (δ rays)*** : The distribution of secondary electrons with kinetic energies $T \gg I$ is [4]

$$\frac{d^2N}{dTdx} = \frac{1}{2}\,Kz^2\frac{Z}{A}\,\frac{1}{\beta^2}\,\frac{F(T)}{T^2} \tag{27.5}$$

for $I \ll T \leq T_{\rm max}$, where $T_{\rm max}$ is given by Eq. (27.2). Here β is the velocity of the primary particle. The factor F is spin-dependent, but is about unity for $T \ll T_{\rm max}$. For spin-0 particles $F(T) = (1 - \beta^2 T/T_{\rm max})$; forms for spins 1/2 and 1 are also given by Rossi [4]. For incident electrons, the indistinguishability of projectile and target means that the range of T extends only to half the kinetic energy of the incident particle. Additional formulae are given in Ref. 19. Equation (27.5) is inaccurate for T close to I.

δ rays of even modest energy are rare. For $\beta \approx 1$ particle, for example, on average only one collision with $T_e > 1$ keV will occur along a path length of 90 cm of Ar gas [25].

A δ ray with kinetic energy T_e and corresponding momentum p_e is produced at an angle θ given by

$$\cos\theta = (T_e/p_e)(p_{\rm max}/T_{\rm max})\ , \tag{27.6}$$

where $p_{\rm max}$ is the momentum of an electron with the maximum possible energy transfer $T_{\rm max}$.

27.2.4. ***Restricted energy loss rates for relativistic ionizing particles*** : Further insight can be obtained by examining the mean energy deposit by an ionizing particle when energy transfers are restricted to $T \leq T_{\rm cut} \leq T_{\rm max}$. The restricted energy loss rate is

$$-\left.\frac{dE}{dx}\right|_{T<T_{\rm cut}} = Kz^2\frac{Z}{A}\frac{1}{\beta^2}\left[\frac{1}{2}\ln\frac{2m_ec^2\beta^2\gamma^2T_{\rm cut}}{I^2} - \frac{\beta^2}{2}\left(1+\frac{T_{\rm cut}}{T_{\rm max}}\right) - \frac{\delta}{2}\right]\ . \tag{27.7}$$

This form approaches the normal Bethe-Bloch function (Eq. (27.1)) as $T_{\rm cut} \to T_{\rm max}$. It can be verified that the difference between Eq. (27.1) and Eq. (27.7) is equal to $\int_{T_{\rm cut}}^{T_{\rm max}} T(d^2N/dTdx)dT$, where $d^2N/dTdx$ is given by Eq. (27.5).

Since $T_{\rm cut}$ replaces $T_{\rm max}$ in the argument of the logarithmic term of Eq. (27.1), the $\beta\gamma$ term producing the relativistic rise in the close-collision part of dE/dx is replaced by a constant, and $|dE/dx|_{T<T_{\rm cut}}$ approaches the constant "Fermi plateau." (The density effect correction δ eliminates the explicit $\beta\gamma$ dependence produced by the distant-collision contribution.) This behavior is illustrated in Fig. 27.6, where restricted loss rates for two examples of $T_{\rm cut}$ are shown in comparison with the full Bethe-Bloch dE/dx and the Landau-Vavilov most probable energy loss (to be discussed in Sec. 27.2.5 below).

27.2.5. ***Fluctuations in energy loss*** : For detectors of moderate thickness x (*e.g.* scintillators or LAr cells),* the energy loss probability distribution $f(\Delta;\beta\gamma,x)$ is adequately described by the highly-skewed Landau (or Landau-Vavilov) distribution [22,23]. The most probable energy loss is [24]

$$\Delta_p = \xi\left[\ln\frac{2mc^2\beta^2\gamma^2}{I} + \ln\frac{\xi}{I} + j - \beta^2 - \delta(\beta\gamma)\right] , \tag{27.8}$$

where $\xi = (K/2)\langle Z/A\rangle(x/\beta^2)$ MeV for a detector with a thickness x in g cm^{-2}, and $j = 0.200$ [24]. † While dE/dx is independent of thickness, Δ_p/x scales as $a\ln x + b$. The density correction $\delta(\beta\gamma)$ was not included in Landau's or Vavilov's work, but it was later included by Bichsel [24]. The high-energy behavior of $\delta(\beta\gamma)$ (Eq. (27.3)), is such that

$$\Delta_p \underset{\beta\gamma\gtrsim 100}{\longrightarrow} \xi\left[\ln\frac{2mc^2\xi}{(\hbar\omega_p)^2} + j\right] . \tag{27.9}$$

Thus the Landau-Vavilov most probable energy loss, like the restricted energy loss, reaches a Fermi plateau. The Bethe-Bloch dE/dx and Landau-Vavilov-Bichsel Δ_p/x in silicon are shown as a function of muon energy in Fig. 27.6. The case $x/\rho = 1600$ μm was chosen since it has about the same stopping power as does 3 mm of plastic scintillator. Folding in experimental resolution displaces the peak of the distribution, usually toward a higher value.

The mean of the energy-loss given by the Bethe-Bloch equation, Eq. (27.1), is ill-defined experimentally and is not useful for describing energy loss by single particles. (It finds its application in dosimetry, where only bulk deposit is of relevance.) It rises as $\ln\beta\gamma$ because $T_{\rm max}$ increases as $\beta^2\gamma^2$. The large single-collision energy transfers that increasingly extend the long tail are rare, making the mean of an experimental distribution consisting of a few hundred events subject to large fluctuations and sensitive to cuts as well as to background. The most probable energy loss should be used.

For very thick absorbers the distribution is less skewed but never approaches a Gaussian. In the case of Si illustrated in Fig. 27.6, the most probable energy loss per unit thickness for $x \approx 35$ g cm^{-2} is very close to the restricted energy loss with $T_{\rm cut} = 2\,dE/dx|_{\rm min}$.

The Landau distribution fails to describe energy loss in thin absorbers such as gas TPC cells [25] and Si detectors [24], as shown clearly in Fig. 1 of Ref. 25 for an

* $G \lesssim 0.05$–0.1, where G is given by Rossi [Ref. 4, Eq. 2.7.10]. It is Vavilov's κ [23].

† Rossi [4], Talman [26], and others give somewhat different values for j. The most probable loss is not sensitive to its value.

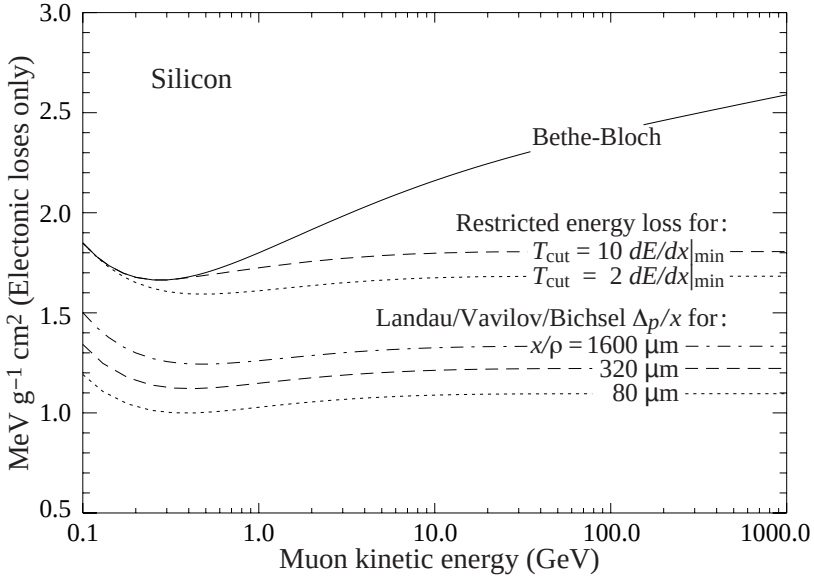


Figure 27.6: Bethe-Bloch dE/dx, two examples of restricted energy loss, and the Landau most probable energy per unit thickness in silicon. The change of Δ_p/x with thickness x illustrates its $a \ln x + b$ dependence. Minimum ionization ($dE/dx|_{\rm min}$) is 1.664 MeV g^{-1} cm^2. Radiative losses are excluded. The incident particles are muons.

argon-filled TPC cell. Also see Talman [26]. While Δ_p/x may be calculated adequately with Eq. (27.8), the distributions are significantly wider than the Landau width $w = 4\xi$ [Ref. 24, Fig. 15]. Examples for thin silicon detectors are shown in Fig. 27.7.

27.2.6. ***Energy loss in mixtures and compounds*** : A mixture or compound can be thought of as made up of thin layers of pure elements in the right proportion (Bragg additivity). In this case,

$$\frac{dE}{dx} = \sum w_j \left.\frac{dE}{dx}\right|_j , \tag{27.10}$$

where $dE/dx|_j$ is the mean rate of energy loss (in MeV g cm^{-2}) in the jth element. Eq. (27.1) can be inserted into Eq. (27.10) to find expressions for $\langle Z/A\rangle$, $\langle I\rangle$, and $\langle\delta\rangle$; for example, $\langle Z/A\rangle = \sum w_j Z_j/A_j = \sum n_j Z_j / \sum n_j A_j$. However, $\langle I\rangle$ as defined this way is an underestimate, because in a compound electrons are more tightly bound than in the free elements, and $\langle\delta\rangle$ as calculated this way has little relevance, because it is the electron density which matters. If possible, one uses the tables given in Refs. 18 and 27, which include effective excitation energies and interpolation coefficients for calculating the density effect correction for the chemical elements and nearly 200 mixtures and compounds. If a compound or mixture is not found, then one uses the recipe for δ given in Ref. 20 (repeated in Ref. 1), and calculates $\langle I\rangle$ according to the discussion in Ref. 7. (Note the "13%" rule!)

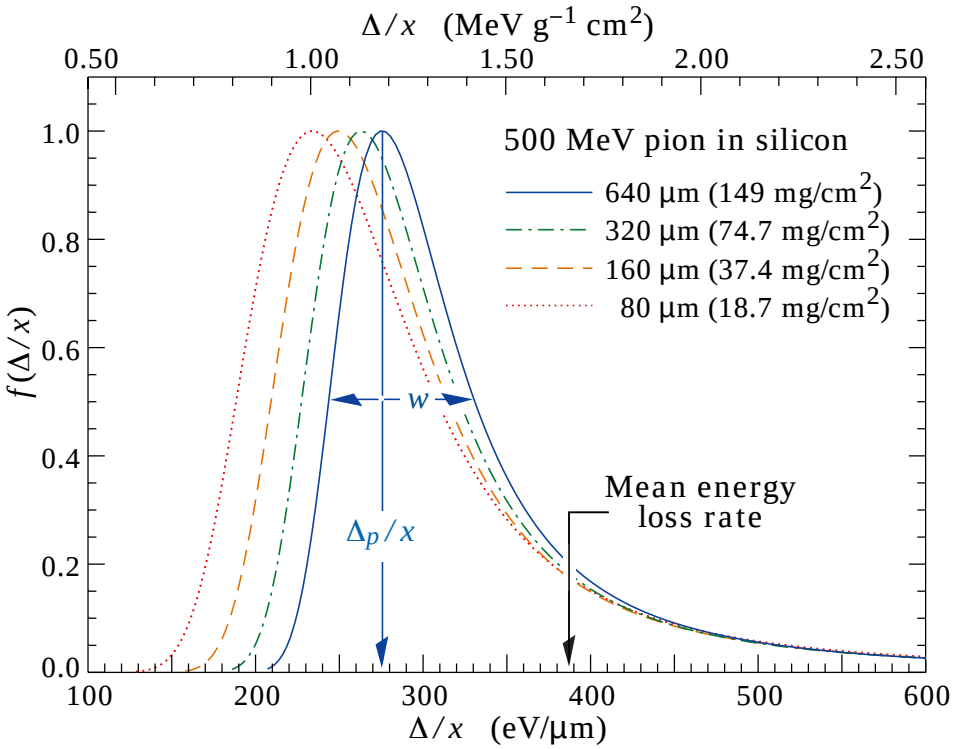


Figure 27.7: Straggling functions in silicon for 500 MeV pions, normalized to unity at the most probable value δ_p/x. The width w is the full width at half maximum. See full-color version on color pages at end of book.

27.2.7. ***Ionization yields*** : Physicists frequently relate total energy loss to the number of ion pairs produced near the particle's track. This relation becomes complicated for relativistic particles due to the wandering of energetic knock-on electrons whose ranges exceed the dimensions of the fiducial volume. For a qualitative appraisal of the nonlocality of energy deposition in various media by such modestly energetic knock-on electrons, see Ref. 28. The mean local energy dissipation per local ion pair produced, W, while essentially constant for relativistic particles, increases at slow particle speeds [29]. For gases, W can be surprisingly sensitive to trace amounts of various contaminants [29]. Furthermore, ionization yields in practical cases may be greatly influenced by such factors as subsequent recombination [30].

27.3. Multiple scattering through small angles

A charged particle traversing a medium is deflected by many small-angle scatters. Most of this deflection is due to Coulomb scattering from nuclei, and hence the effect is called multiple Coulomb scattering. (However, for hadronic projectiles, the strong interactions also contribute to multiple scattering.) The Coulomb scattering distribution is well represented by the theory of Molière [31]. It is roughly Gaussian for small deflection angles, but at larger angles (greater than a few θ_0, defined below) it behaves like Rutherford scattering, with larger tails than does a Gaussian distribution.

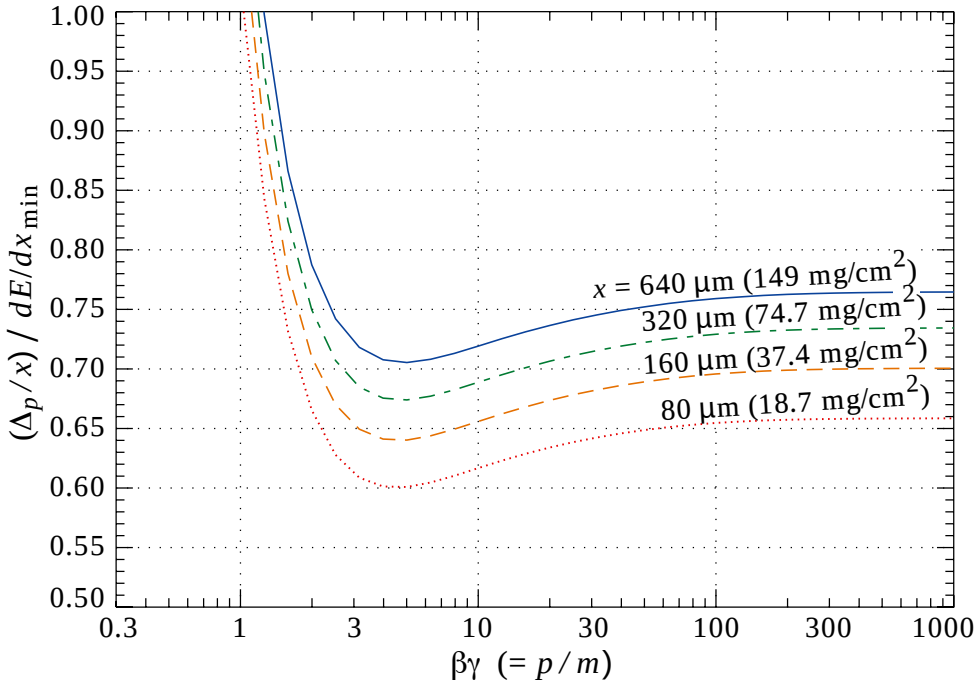


Figure 27.8: Most probable energy loss in silicon, scaled to the mean loss of a minimum ionizing particle, 388 eV/μm (1.66 MeV g^{-1}cm^2). See full-color version on color pages at end of book.

If we define

$$\theta_0 = \theta^{\rm rms}_{\rm plane} = \frac{1}{\sqrt{2}}\,\theta^{\rm rms}_{\rm space}\ . \tag{27.11}$$

then it is sufficient for many applications to use a Gaussian approximation for the central 98% of the projected angular distribution, with a width given by [32,33]

$$\theta_0 = \frac{13.6\ {\rm MeV}}{\beta cp}\ z\ \sqrt{x/X_0}\Big[1 + 0.038\ln(x/X_0)\Big]\ . \tag{27.12}$$

Here p, βc, and z are the momentum, velocity, and charge number of the incident particle, and x/X_0 is the thickness of the scattering medium in radiation lengths (defined below). This value of θ_0 is from a fit to Molière distribution [31] for singly charged particles with $\beta = 1$ for all Z, and is accurate to 11% or better for $10^{-3} < x/X_0 < 100$.

Eq. (27.12) describes scattering from a single material, while the usual problem involves the multiple scattering of a particle traversing many different layers and mixtures. Since it is from a fit to a Molière distribution, it is incorrect to add the individual θ_0 contributions in quadrature; the result is systematically too small. It is much more accurate to apply Eq. (27.12) once, after finding x and X_0 for the combined scatterer.

Lynch and Dahl have extended this phenomenological approach, fitting Gaussian distributions to a variable fraction of the Molière distribution for arbitrary scatterers [33], and achieve accuracies of 2% or better.

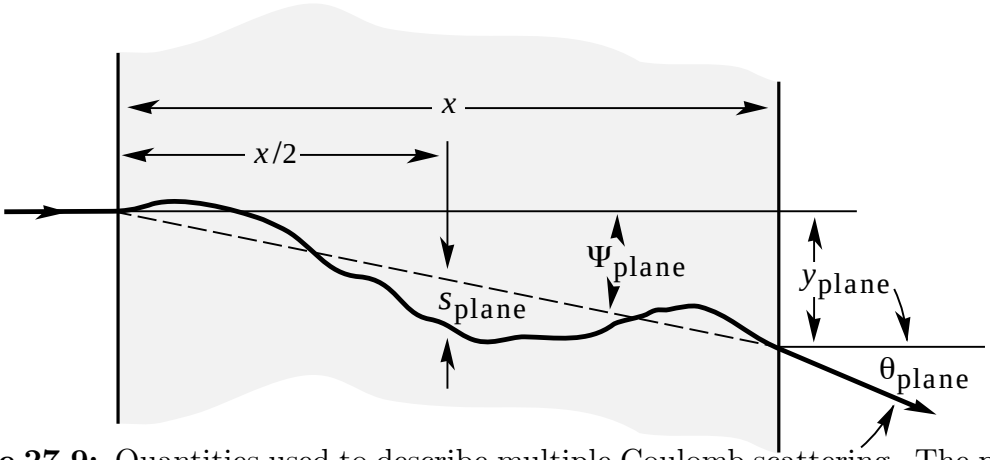


Figure 27.9: Quantities used to describe multiple Coulomb scattering. The particle is incident in the plane of the figure.

The nonprojected (space) and projected (plane) angular distributions are given approximately by [31]

$$\frac{1}{2\pi\,\theta_0^2}\,\exp\left(-\frac{\theta^2_{\rm space}}{2\theta_0^2}\right)\,d\Omega\ , \tag{27.13}$$

$$\frac{1}{\sqrt{2\pi}\,\theta_0}\,\exp\left(-\frac{\theta^2_{\rm plane}}{2\theta_0^2}\right)\,d\theta_{\rm plane}\ , \tag{27.14}$$

where θ is the deflection angle. In this approximation, $\theta^2_{\rm space} \approx (\theta^2_{{\rm plane},x} + \theta^2_{{\rm plane},y})$, where the x and y axes are orthogonal to the direction of motion, and $d\Omega \approx d\theta_{{\rm plane},x}\,d\theta_{{\rm plane},y}$. Deflections into $\theta_{{\rm plane},x}$ and $\theta_{{\rm plane},y}$ are independent and identically distributed.

Figure 27.9 shows these and other quantities sometimes used to describe multiple Coulomb scattering. They are

$$\psi^{\,\rm rms}_{\rm plane} = \frac{1}{\sqrt{3}}\,\theta^{\,\rm rms}_{\rm plane} = \frac{1}{\sqrt{3}}\,\theta_0\ , \tag{27.15}$$

$$y^{\,\rm rms}_{\rm plane} = \frac{1}{\sqrt{3}}\,x\,\theta^{\,\rm rms}_{\rm plane} = \frac{1}{\sqrt{3}}\,x\,\theta_0\ , \tag{27.16}$$

$$s^{\,\rm rms}_{\rm plane} = \frac{1}{4\sqrt{3}}\,x\,\theta^{\,\rm rms}_{\rm plane} = \frac{1}{4\sqrt{3}}\,x\,\theta_0\ . \tag{27.17}$$

All the quantitative estimates in this section apply only in the limit of small $\theta^{\,\rm rms}_{\rm plane}$ and in the absence of large-angle scatters. The random variables s, ψ, y, and θ in a given plane are distributed in a correlated fashion (see Sec. 31.1 of this *Review* for the definition of the correlation coefficient). Obviously, $y \approx x\psi$. In addition, y and θ have the correlation coefficient $\rho_{y\theta} = \sqrt{3}/2 \approx 0.87$. For Monte Carlo generation of a joint $(y_{\,\rm plane}, \theta_{\rm plane})$ distribution, or for other calculations, it may be most convenient to work with independent Gaussian random variables (z_1, z_2) with mean zero and variance one, and then set

$$y_{\rm plane} = z_1\,x\,\theta_0(1-\rho^2_{y\theta})^{1/2}/\sqrt{3} + z_2\,\rho_{y\theta}x\,\theta_0/\sqrt{3}$$

$$= z_1\, x\, \theta_0/\sqrt{12} + z_2\, x\, \theta_0/2 \ ; \tag{27.18}$$

$$\theta_{\rm plane} = z_2\, \theta_0 \ . \tag{27.19}$$

Note that the second term for $y_{\rm\ plane}$ equals $x\,\theta_{\rm plane}/2$ and represents the displacement that would have occurred had the deflection $\theta_{\rm plane}$ all occurred at the single point $x/2$.

For heavy ions the multiple Coulomb scattering has been measured and compared with various theoretical distributions [34].

27.4. Photon and electron interactions in matter

27.4.1. ***Radiation length*** : High-energy electrons predominantly lose energy in matter by bremsstrahlung, and high-energy photons by e^+e^- pair production. The characteristic amount of matter traversed for these related interactions is called the radiation length X_0, usually measured in g cm^{-2}. It is both (a) the mean distance over which a high-energy electron loses all but $1/e$ of its energy by bremsstrahlung, and (b) $\frac{7}{9}$ of the mean free path for pair production by a high-energy photon [35]. It is also the appropriate scale length for describing high-energy electromagnetic cascades. X_0 has been calculated and tabulated by Y.S. Tsai [36]:

$$\frac{1}{X_0} = 4\alpha r_e^2 \frac{N_A}{A}\Big\{ Z^2\big[L_{\rm rad} - f(Z)\big] + Z\, L'_{\rm rad}\Big\} \ . \tag{27.20}$$

For $A = 1$ g mol^{-1}, $4\alpha r_e^2 N_A/A = (716.408\ {\rm g\ cm^{-2}})^{-1}$. $L_{\rm rad}$ and $L'_{\rm rad}$ are given in Table 27.2. The function $f(Z)$ is an infinite sum, but for elements up to uranium can be represented to 4-place accuracy by

$$f(Z) = a^2\big[(1+a^2)^{-1} + 0.20206 \\ -0.0369\, a^2 + 0.0083\, a^4 - 0.002\, a^6\big] \ , \tag{27.21}$$

where $a = \alpha Z$ [37].

Table 27.2: Tsai's $L_{\rm rad}$ and $L'_{\rm rad}$, for use in calculating the radiation length in an element using Eq. (27.20).

Element	Z	$L_{\rm rad}$	$L'_{\rm rad}$
H	1	5.31	6.144
He	2	4.79	5.621
Li	3	4.74	5.805
Be	4	4.71	5.924
Others	> 4	$\ln(184.15\, Z^{-1/3})$	$\ln(1194\, Z^{-2/3})$

Although it is easy to use Eq. (27.20) to calculate X_0, the functional dependence on Z is somewhat hidden. Dahl provides a compact fit to the data [38]:

$$X_0 = \frac{716.4\ {\rm g\ cm^{-2}}\, A}{Z(Z+1)\ln(287/\sqrt{Z})} \ . \tag{27.22}$$

Results using this formula agree with Tsai's values to better than 2.5% for all elements except helium, where the result is about 5% low.

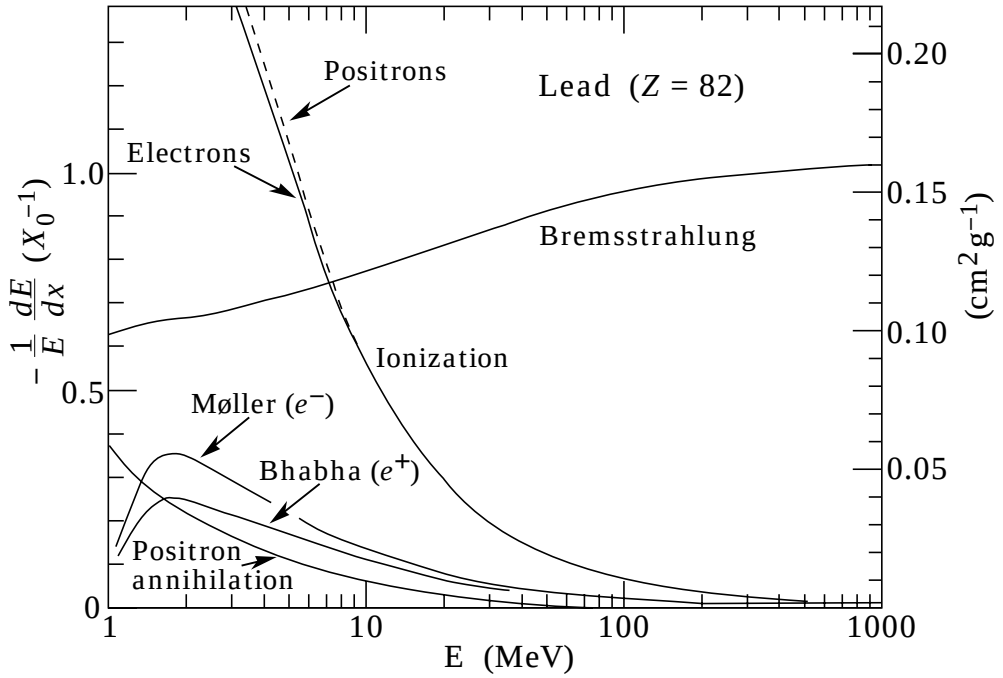


Figure 27.10: Fractional energy loss per radiation length in lead as a function of electron or positron energy. Electron (positron) scattering is considered as ionization when the energy loss per collision is below 0.255 MeV, and as Møller (Bhabha) scattering when it is above. Adapted from Fig. 3.2 from Messel and Crawford, *Electron-Photon Shower Distribution Function Tables for Lead, Copper, and Air Absorbers*, Pergamon Press, 1970. Messel and Crawford use $X_0(\mathrm{Pb}) = 5.82$ g/cm^2, but we have modified the figures to reflect the value given in the Table of Atomic and Nuclear Properties of Materials ($X_0(\mathrm{Pb}) = 6.37$ g/cm^2).

The radiation length in a mixture or compound may be approximated by

$$1/X_0 = \sum w_j/X_j \ , \tag{27.23}$$

where w_j and X_j are the fraction by weight and the radiation length for the jth element.

27.4.2. ***Energy loss by electrons*** : At low energies electrons and positrons primarily lose energy by ionization, although other processes (Møller scattering, Bhabha scattering, e^+ annihilation) contribute, as shown in Fig. 27.10. While ionization loss rates rise logarithmically with energy, bremsstrahlung losses rise nearly linearly (fractional loss is nearly independent of energy), and dominates above a few tens of MeV in most materials

Ionization loss by electrons and positrons differs from loss by heavy particles because of the kinematics, spin, and the identity of the incident electron with the electrons which it ionizes. Complete discussions and tables can be found in Refs. 7, 8, and 27.

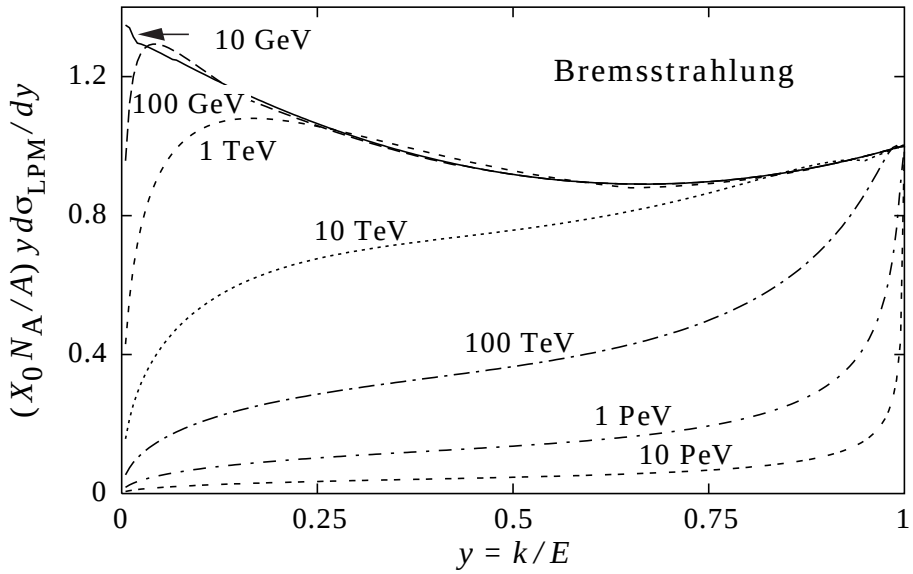


Figure 27.11: The normalized bremsstrahlung cross section $k\, d\sigma_{LPM}/dk$ in lead versus the fractional photon energy $y = k/E$. The vertical axis has units of photons per radiation length.

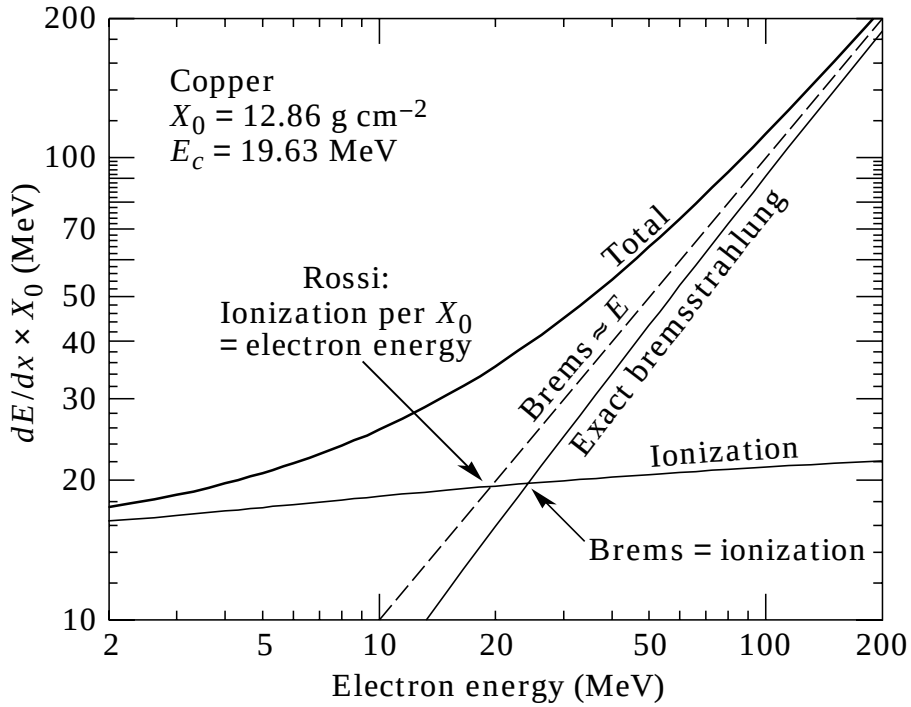


Figure 27.12: Two definitions of the critical energy E_c.

At very high energies and except at the high-energy tip of the bremsstrahlung

spectrum, the cross section can be approximated in the "complete screening case" as [36]

$$d\sigma/dk = (1/k)4\alpha r_e^2\{(\tfrac{4}{3} - \tfrac{4}{3}y + y^2)[Z^2(L_{\rm rad} - f(Z)) + Z\,L'_{\rm rad}] + \tfrac{1}{9}(1-y)(Z^2+Z)\}\ , \tag{27.24}$$

where $y = k/E$ is the fraction of the electron's energy transfered to the radiated photon. At small y (the "infrared limit") the term on the second line ranges from 1.7% (low Z) to 2.5% (high Z) of the total. If it is ignored and the first line simplified with the definition of X_0 given in Eq. (27.20), we have

$$\frac{d\sigma}{dk} = \frac{A}{X_0 N_A k}\left(\tfrac{4}{3} - \tfrac{4}{3}y + y^2\right)\ . \tag{27.25}$$

This cross section (times k) is shown by the top curve in Fig. 27.11.

This formula is accurate except in near $y = 1$, where screening may become incomplete, and near $y = 0$, where the infrared divergence is removed by the interference of bremsstrahlung amplitudes from nearby scattering centers (the LPM effect) [39,40] and dielectric suppression [41,42]. These and other suppression effects in bulk media are discussed in Sec. 27.4.5.

With decreasing energy ($E \lesssim 10$ GeV) the high-y cross section drops and the curves become rounded as $y \to 1$. Curves of this familar shape can be seen in Rossi [4] (Figs. 2.11.2,3); see also the review by Koch & Motz [43].

Except at these extremes, and still in the complete-screening approximation, the number of photons with energies between $k_{\rm min}$ and $k_{\rm max}$ emitted by an electron travelling a distance $d \ll X_0$ is

$$N_\gamma = \frac{d}{X_0}\left[\frac{4}{3}\ln\left(\frac{k_{\rm max}}{k_{\rm min}}\right) - \frac{4(k_{\rm max} - k_{\rm min})}{3E} + \frac{(k_{\rm max} - k_{\rm min})^2}{2E^2}\right]\ . \tag{27.26}$$

27.4.3. ***Critical energy*** : An electron loses energy by bremsstrahlung at a rate nearly proportional to its energy, while the ionization loss rate varies only logarithmically with the electron energy. The *critical energy* E_c is sometimes defined as the energy at which the two loss rates are equal [44]. Berger and Seltzer [44] also give the approximation $E_c = (800\ {\rm MeV})/(Z + 1.2)$. This formula has been widely quoted, and has been given in older editions of this *Review* [45]. Among alternate definitions is that of Rossi [4], who defines the critical energy as the energy at which the ionization loss per radiation length is equal to the electron energy. Equivalently, it is the same as the first definition with the approximation $|dE/dx|_{\rm brems} \approx E/X_0$. This form has been found to describe transverse electromagnetic shower development more accurately (see below). These definitions are illustrated in the case of copper in Fig. 27.12.

The accuracy of approximate forms for E_c has been limited by the failure to distinguish between gases and solid or liquids, where there is a substantial difference in ionization at the relevant energy because of the density effect. We distinguish these two cases in Fig. 27.13. Fits were also made with functions of the form $a/(Z + b)^\alpha$, but α was found to be essentially unity. Since E_c also depends on A, I, and other factors, such forms are at best approximate.

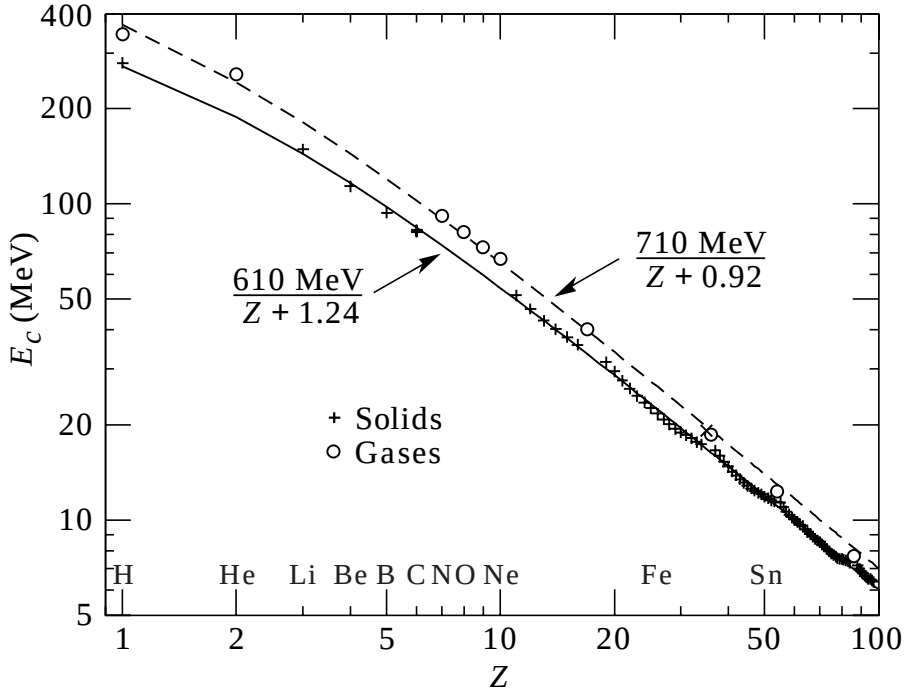


Figure 27.13: Electron critical energy for the chemical elements, using Rossi's definition [4]. The fits shown are for solids and liquids (solid line) and gases (dashed line). The rms deviation is 2.2% for the solids and 4.0% for the gases. (Computed with code supplied by A. Fassó.)

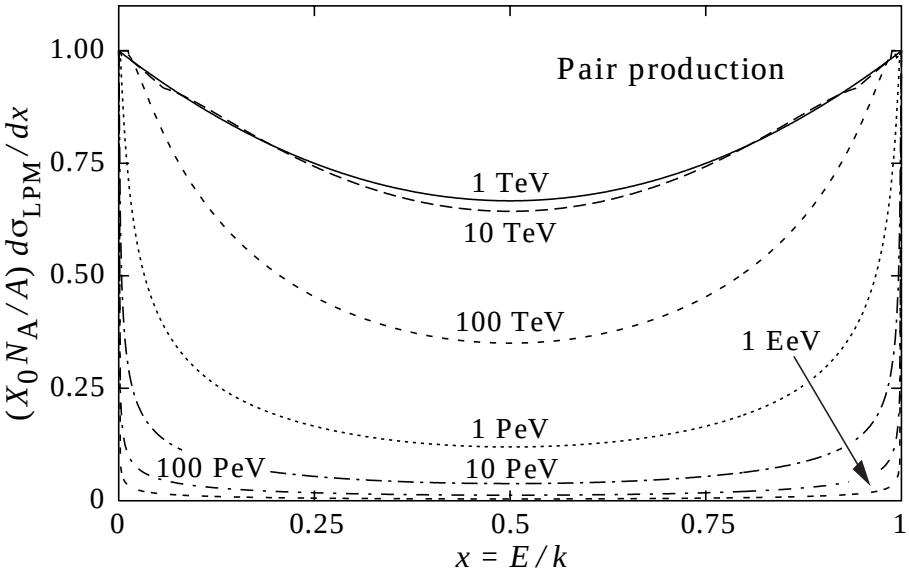


Figure 27.15: The normalized pair production cross section $d\sigma_{LPM}/dy$, versus fractional electron energy $x = E/k$.

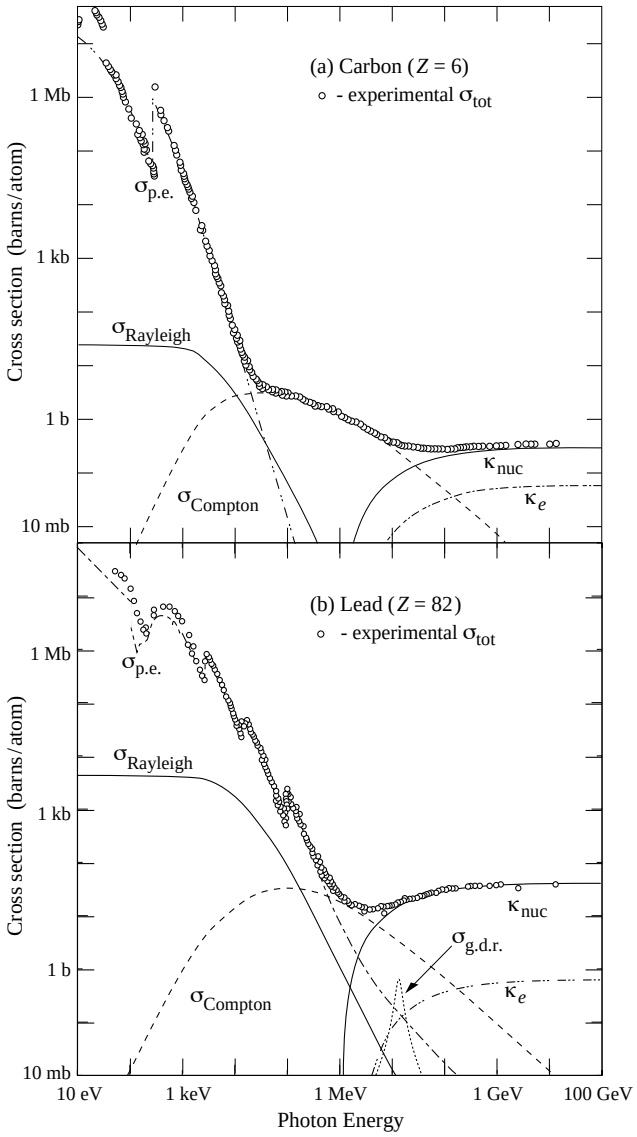


Figure 27.14: Photon total cross sections as a function of energy in carbon and lead, showing the contributions of different processes:

$\sigma_{\text{p.e.}}$ = Atomic photoelectric effect (electron ejection, photon absorption)

σ_{Rayleigh} = Rayleigh (coherent) scattering–atom neither ionized nor excited

σ_{Compton} = Incoherent scattering (Compton scattering off an electron)

κ_{nuc} = Pair production, nuclear field

κ_e = Pair production, electron field

$\sigma_{\text{g.d.r.}}$ = Photonuclear interactions, most notably the Giant Dipole Resonance [46]. In these interactions, the target nucleus is broken up.

Data from [47]; parameters for $\sigma_{\text{g.d.r.}}$ from [48]. Curves for these and other elements, compounds, and mixtures may be obtained from `http://physics.nist.gov/PhysRefData`. The photon total cross section is approximately flat for at least two decades beyond the energy range shown. Original figures courtesy J.H. Hubbell (NIST).

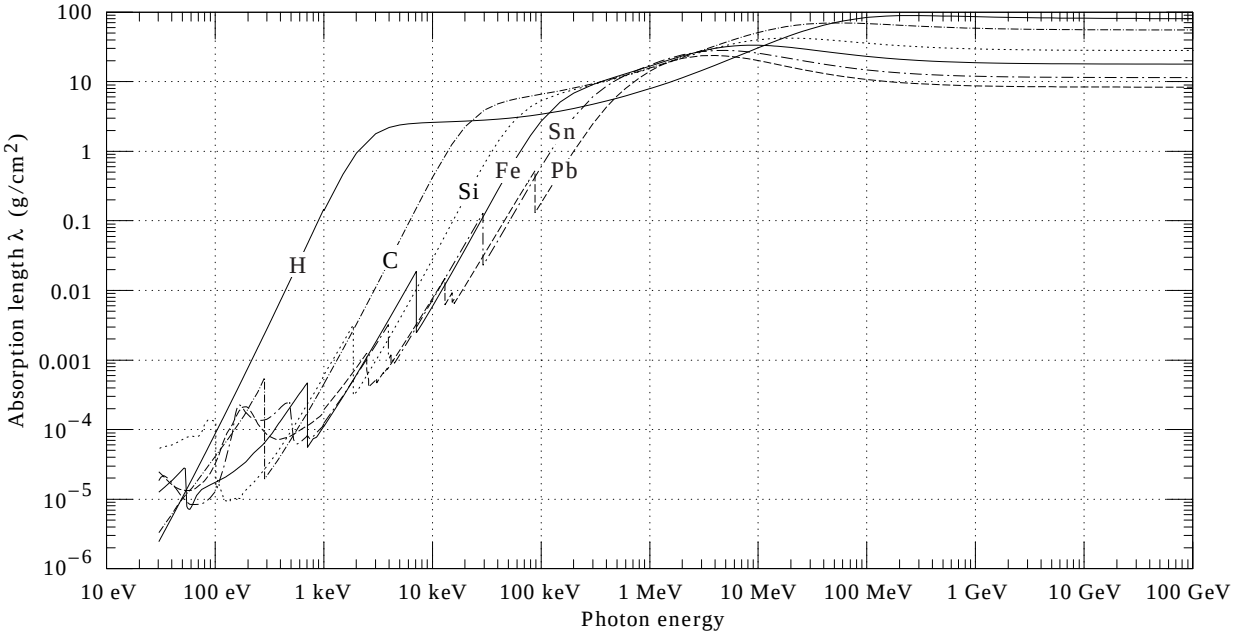


Figure 27.16: The photon mass attenuation length (or mean free path) $\lambda = 1/(\mu/\rho)$ for various elemental absorbers as a function of photon energy. The mass attenuation coefficient is μ/ρ, where ρ is the density. The intensity I remaining after traversal of thickness t (in mass/unit area) is given by $I = I_0 \exp(-t/\lambda)$. The accuracy is a few percent. For a chemical compound or mixture, $1/\lambda_{\rm eff} \approx \sum_{\rm elements} w_Z/\lambda_Z$, where w_Z is the proportion by weight of the element with atomic number Z. The processes responsible for attenuation are given in not Fig. 27.10. Since coherent processes are included, not all these processes result in energy deposition. The data for 30 eV $< E <$ 1 keV are obtained from `http://www-cxro.lbl.gov/optical_constants` (courtesy of Eric M. Gullikson, LBNL). The data for 1 keV $< E <$ 100 GeV are from `http://physics.nist.gov/PhysRefData`, through the courtesy of John H. Hubbell (NIST).

27.4.4. ***Energy loss by photons*** : Contributions to the photon cross section in a light element (carbon) and a heavy element (lead) are shown in Fig. 27.14. At low energies it is seen that the photoelectric effect dominates, although Compton scattering, Rayleigh scattering, and photonuclear absorption also contribute. The photoelectric cross section is characterized by discontinuities (absorption edges) as thresholds for photoionization of various atomic levels are reached. Photon attenuation lengths for a variety of elements are shown in Fig. 27.16, and data for 30 eV$< k <$100 GeV for all elements is available from the web pages given in the caption. Here k is the photon energy.

The increasing domination of pair production as the energy increases is shown in Fig. 27.17. Using approximations similar to those used to obtain Eq. (27.25), Tsai's

formula for the differential cross section [36] reduces to

$$\frac{d\sigma}{dx} = \frac{A}{X_0 N_A}\left[1 - \tfrac{4}{3}x(1-x)\right] \tag{27.27}$$

in the complete-screening limit valid at high energies. Here $x = E/k$ is the fractional energy transfer to the pair-produced electron (or positron), and k is the incident photon energy. The cross section is very closely related to that for bremsstrahlung, since the Feynman diagrams are variants of one another. The cross section is of necessity symmetric between x and $1-x$, as can be seen by the solid curve in Fig. 27.15. See the review by Motz, Olsen, & Koch for a more detailed treatment [49].

Eq. (27.27) may be integrated to find the high-energy limit for the total e^+e^- pair-production cross section:

$$\sigma = \tfrac{7}{9}(A/X_0 N_A)\ . \tag{27.28}$$

Equation Eq. (27.28) is accurate to within a few percent down to energies as low as 1 GeV, particularly for high-Z materials.

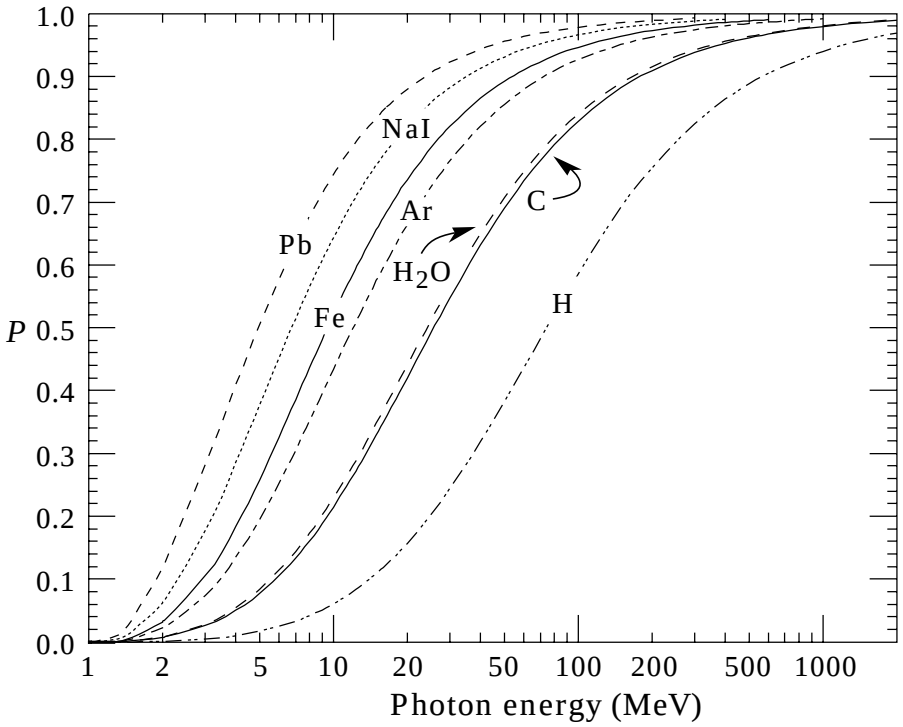


Figure 27.17: Probability P that a photon interaction will result in conversion to an e^+e^- pair. Except for a few-percent contribution from photonuclear absorption around 10 or 20 MeV, essentially all other interactions in this energy range result in Compton scattering off an atomic electron. For a photon attenuation length λ (Fig. 27.16), the probability that a given photon will produce an electron pair (without first Compton scattering) in thickness t of absorber is $P[1 - \exp(-t/\lambda)]$.

27.4.5. ***Bremsstrahlung and pair production at very high energies*** : At ultra-high energies, Eqns. 27.24–27.28 will fail because of quantum mechanical interference between amplitudes from different scattering centers. Since the longitudinal momentum transfer to a given center is small ($\propto k/E(E-k)$, in the case of bremsstrahlung), the interaction is spread over a comparatively long distance called the formation length ($\propto E(E-k)/k$) via the uncertainty principle. In alternate language, the formation length is the distance over which the highly relativistic electron and the photon "split apart." The interference is usually destructive. Calculations of the "Landau-Pomeranchuk-Migdal" (LPM) effect may be made semi-classically based on the average multiple scattering, or more rigorously using a quantum transport approach [39,40].

In amorphous media, bremsstrahlung is suppressed if the photon energy k is less than $E^2/(E+E_{LPM})$ [40], where*

$$E_{LPM} = \frac{(m_e c^2)^2 \alpha X_0}{4\pi\hbar c\rho} = (7.7 \text{ TeV/cm}) \times \frac{X_0}{\rho} \ . \tag{27.29}$$

Since physical distances are involved, X_0/ρ, in cm, appears. The energy-weighted bremsstrahlung spectrum for lead, $k\, d\sigma_{LPM}/dk$, is shown in Fig. 27.11. With appropriate scaling by X_0/ρ, other materials behave similarly.

For photons, pair production is reduced for $E(k-E) > k\, E_{LPM}$. The pair-production cross sections for different photon energies are shown in Fig. 27.15.

If $k \ll E$, several additional mechanisms can also produce suppression. When the formation length is long, even weak factors can perturb the interaction. For example, the emitted photon can coherently forward scatter off of the electrons in the media. Because of this, for $k < \omega_p E/m_e \sim 10^{-4}$, bremsstrahlung is suppressed by a factor $(k m_e/\omega_p E)^2$ [42]. Magnetic fields can also suppress bremsstrahlung.

In crystalline media, the situation is more complicated, with coherent enhancement or suppression possible. The cross section depends on the electron and photon energies and the angles between the particle direction and the crystalline axes [51].

27.5. Electromagnetic cascades

When a high-energy electron or photon is incident on a thick absorber, it initiates an electromagnetic cascade as pair production and bremsstrahlung generate more electrons and photons with lower energy. The longitudinal development is governed by the high-energy part of the cascade, and therefore scales as the radiation length in the material. Electron energies eventually fall below the critical energy, and then dissipate their energy by ionization and excitation rather than by the generation of more shower particles. In describing shower behavior, it is therefore convenient to introduce the scale variables

$$t = x/X_0 \ , \qquad\qquad y = E/E_c \ , \tag{27.30}$$

* This definition differs from that of Ref. 50 by a factor of two. E_{LPM} scales as the 4th power of the mass of the incident particle, so that $E_{LPM} = (1.4 \times 10^{10} \text{ TeV/cm}) \times X_0/\rho$ for a muon.

so that distance is measured in units of radiation length and energy in units of critical energy.

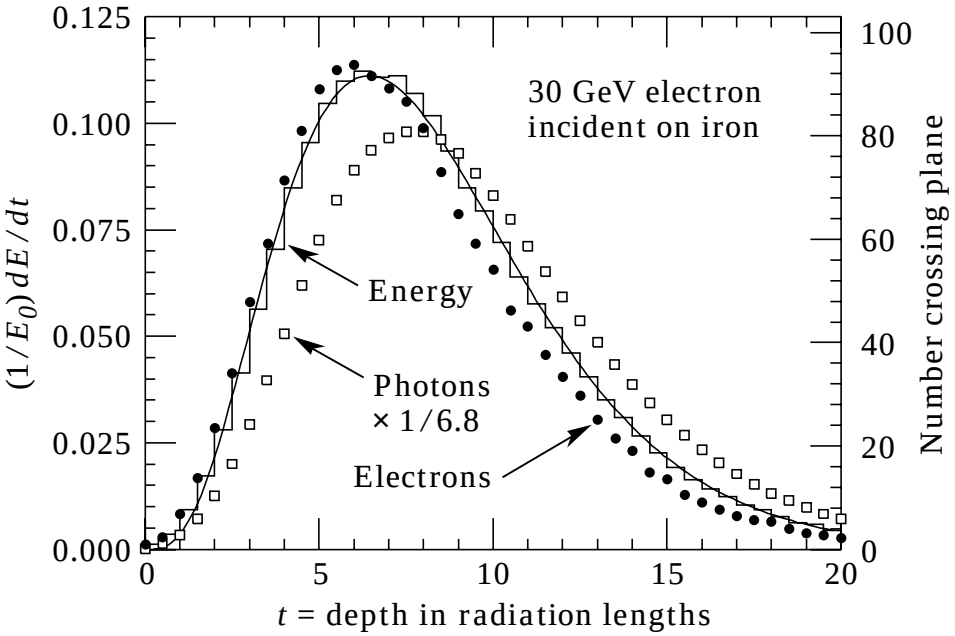


Figure 27.18: An EGS4 simulation of a 30 GeV electron-induced cascade in iron. The histogram shows fractional energy deposition per radiation length, and the curve is a gamma-function fit to the distribution. Circles indicate the number of electrons with total energy greater than 1.5 MeV crossing planes at $X_0/2$ intervals (scale on right) and the squares the number of photons with $E \geq 1.5$ MeV crossing the planes (scaled down to have same area as the electron distribution).

Longitudinal profiles from an EGS4 [52] simulation of a 30 GeV electron-induced cascade in iron are shown in Fig. 27.18. The number of particles crossing a plane (very close to Rossi's Π function [4]) is sensitive to the cutoff energy, here chosen as a total energy of 1.5 MeV for both electrons and photons. The electron number falls off more quickly than energy deposition. This is because, with increasing depth, a larger fraction of the cascade energy is carried by photons. Exactly what a calorimeter measures depends on the device, but it is not likely to be exactly any of the profiles shown. In gas counters it may be very close to the electron number, but in glass Cherenkov detectors and other devices with "thick" sensitive regions it is closer to the energy deposition (total track length). In such detectors the signal is proportional to the "detectable" track length T_d, which is in general less than the total track length T. Practical devices are sensitive to electrons with energy above some detection threshold E_d, and $T_d = T\,F(E_d/E_c)$. An analytic form for $F(E_d/E_c)$ obtained by Rossi [4] is given by Fabjan [53]; see also Amaldi [54].

The mean longitudinal profile of the energy deposition in an electromagnetic cascade

is reasonably well described by a gamma distribution [55]:

$$\frac{dE}{dt} = E_0\, b\, \frac{(bt)^{a-1} e^{-bt}}{\Gamma(a)} \tag{27.31}$$

The maximum $t_{\max}$ occurs at $(a-1)/b$. We have made fits to shower profiles in elements ranging from carbon to uranium, at energies from 1 GeV to 100 GeV. The energy deposition profiles are well described by Eq. (27.31) with

$$t_{\max} = (a-1)/b = 1.0 \times (\ln y + C_j)\ , \qquad j = e, \gamma\ , \tag{27.32}$$

where $C_e = -0.5$ for electron-induced cascades and $C_\gamma = +0.5$ for photon-induced cascades. To use Eq. (27.31), one finds $(a-1)/b$ from Eq. (27.32) and Eq. (27.30), then finds a either by assuming $b \approx 0.5$ or by finding a more accurate value from Fig. 27.19. The results are very similar for the electron number profiles, but there is some dependence on the atomic number of the medium. A similar form for the electron number maximum was obtained by Rossi in the context of his "Approximation B," [4] (see Fabjan's review in Ref. 53), but with $C_e = -1.0$ and $C_\gamma = -0.5$; we regard this as superseded by the EGS4 result.

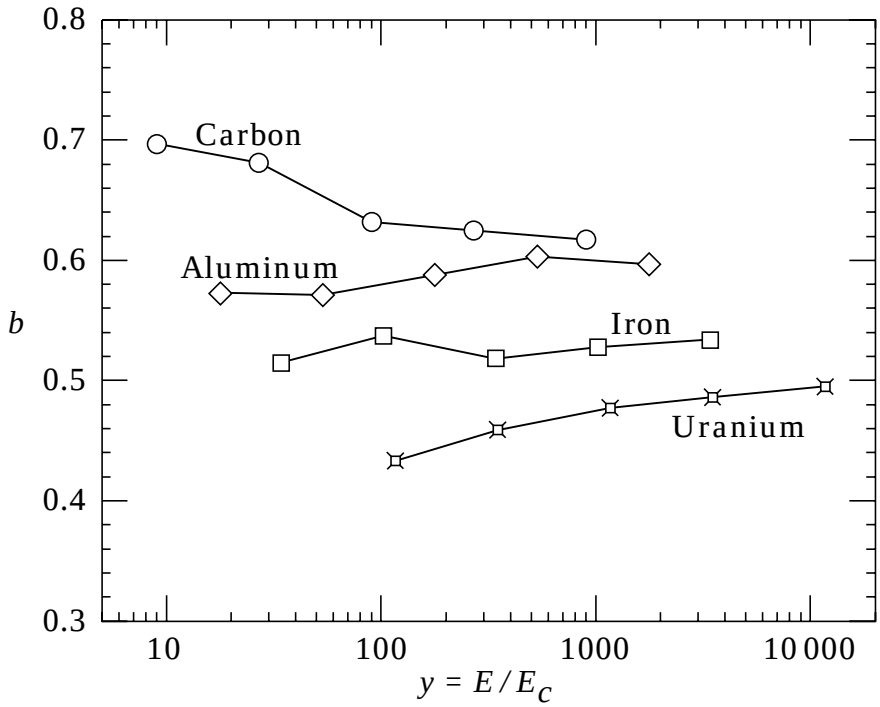


Figure 27.19: Fitted values of the scale factor b for energy deposition profiles obtained with EGS4 for a variety of elements for incident electrons with $1 \le E_0 \le 100$ GeV. Values obtained for incident photons are essentially the same.

The "shower length" $X_s = X_0/b$ is less conveniently parameterized, since b depends upon both Z and incident energy, as shown in Fig. 27.19. As a corollary of this Z dependence, the number of electrons crossing a plane near shower maximum is underestimated using Rossi's approximation for carbon and seriously overestimated for

uranium. Essentially the same b values are obtained for incident electrons and photons. For many purposes it is sufficient to take $b \approx 0.5$.

The gamma function distribution is very flat near the origin, while the EGS4 cascade (or a real cascade) increases more rapidly. As a result Eq. (27.31) fails badly for about the first two radiation lengths; it was necessary to exclude this region in making fits.

Because fluctuations are important, Eq. (27.31) should be used only in applications where average behavior is adequate. Grindhammer *et al.* have developed fast simulation algorithms in which the variance and correlation of a and b are obtained by fitting Eq. (27.31) to individually simulated cascades, then generating profiles for cascades using a and b chosen from the correlated distributions [56].

The transverse development of electromagnetic showers in different materials scales fairly accurately with the *Molière radius* R_M, given by [57,58]

$$R_M = X_0 \, E_s / E_c \ , \tag{27.33}$$

where $E_s \approx 21$ MeV (Table 27.1), and the Rossi definition of E_c is used.

In a material containing a weight fraction w_j of the element with critical energy E_{cj} and radiation length X_j, the Molière radius is given by

$$\frac{1}{R_M} = \frac{1}{E_s} \sum \frac{w_j \, E_{cj}}{X_j} \ . \tag{27.34}$$

Measurements of the lateral distribution in electromagnetic cascades are shown in Refs. 57 and 58. On the average, only 10% of the energy lies outside the cylinder with radius R_M. About 99% is contained inside of $3.5 R_M$, but at this radius and beyond composition effects become important and the scaling with R_M fails. The distributions are characterized by a narrow core, and broaden as the shower develops. They are often represented as the sum of two Gaussians, and Grindhammer [56] describes them with the function

$$f(r) = \frac{2r \, R^2}{(r^2 + R^2)^2} \ , \tag{27.35}$$

where R is a phenomenological function of x/X_0 and $\ln E$.

At high enough energies, the LPM effect (Sec. 27.4.5) reduces the cross sections for bremsstrahlung and pair production, and hence can cause significant elongation of electromagnetic cascades [40].

27.6. Muon energy loss at high energy

At sufficiently high energies, radiative processes become more important than ionization for all charged particles. For muons and pions in materials such as iron, this "critical energy" occurs at several hundred GeV. (There is no simple scaling with particle mass, but for protons the "critical energy" is much, much higher.) Radiative effects dominate the energy loss of energetic muons found in cosmic rays or produced at the newest accelerators. These processes are characterized by small cross sections, hard spectra, large energy fluctuations, and the associated generation of electromagnetic and (in the case of photonuclear interactions) hadronic showers [59–67]. As a consequence, at these

energies the treatment of energy loss as a uniform and continuous process is for many purposes inadequate.

It is convenient to write the average rate of muon energy loss as [68]

$$-dE/dx = a(E) + b(E)\,E\ . \tag{27.36}$$

Here $a(E)$ is the ionization energy loss given by Eq. (27.1), and $b(E)$ is the sum of e^+e^- pair production, bremsstrahlung, and photonuclear contributions. To the approximation that these slowly-varying functions are constant, the mean range x_0 of a muon with initial energy E_0 is given by

$$x_0 \approx (1/b)\ln(1 + E_0/E_{\mu c})\ , \tag{27.37}$$

where $E_{\mu c} = a/b$. Figure 27.20 shows contributions to $b(E)$ for iron. Since $a(E) \approx 0.002$ GeV g^{-1} cm^2, $b(E)E$ dominates the energy loss above several hundred GeV, where $b(E)$ is nearly constant. The rates of energy loss for muons in hydrogen, uranium, and iron are shown in Fig. 27.21 [1].

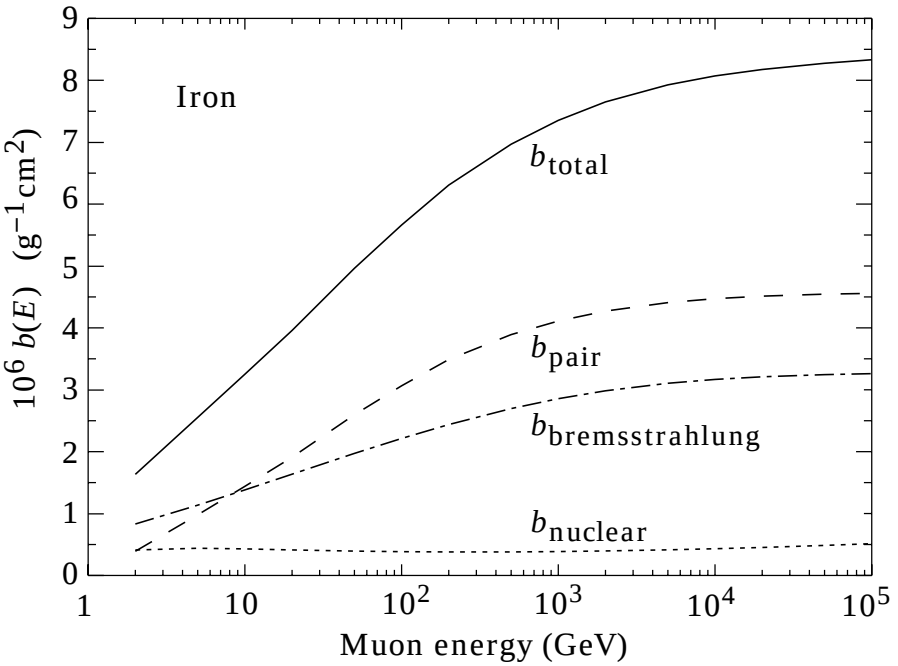


Figure 27.20: Contributions to the fractional energy loss by muons in iron due to e^+e^- pair production, bremsstrahlung, and photonuclear interactions, as obtained from Groom *et al.* [1] except for post-Born corrections to the cross section for direct pair production from atomic electrons.

The "muon critical energy" $E_{\mu c}$ can be defined more exactly as the energy at which radiative and ionization losses are equal, and can be found by solving $E_{\mu c} = a(E_{\mu c})/b(E_{\mu c})$. This definition corresponds to the solid-line intersection in Fig. 27.12, and is different from the Rossi definition we used for electrons. It serves the same function: below $E_{\mu c}$ ionization losses dominate, and above $E_{\mu c}$ radiative effects dominate. The dependence of $E_{\mu c}$ on atomic number Z is shown in Fig. 27.22.

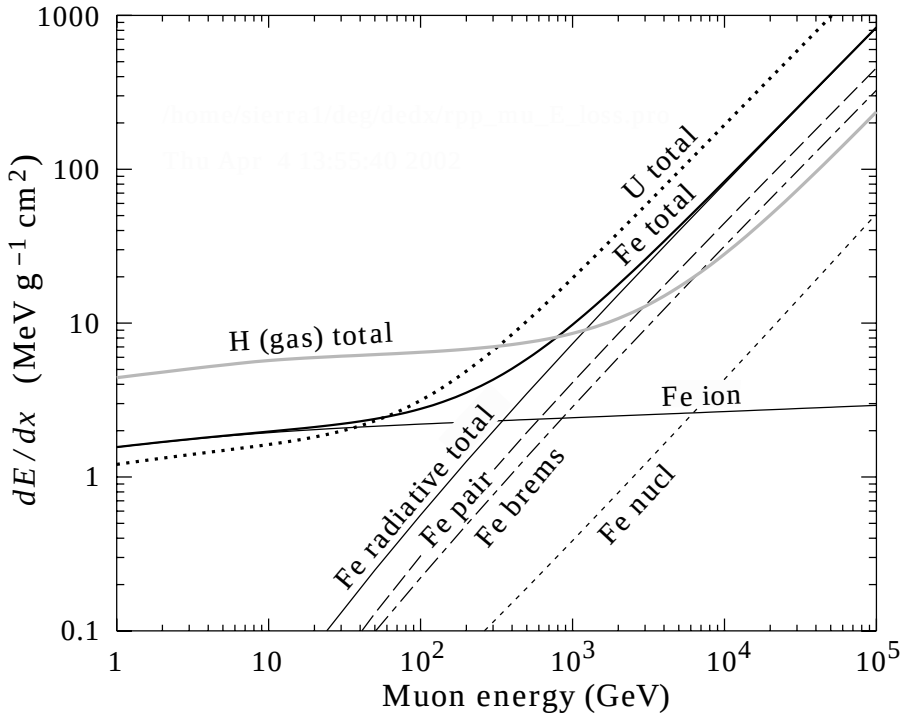


Figure 27.21: The average energy loss of a muon in hydrogen, iron, and uranium as a function of muon energy. Contributions to dE/dx in iron from ionization and the processes shown in Fig. 27.20 are also shown.

The radiative cross sections are expressed as functions of the fractional energy loss ν. The bremsstrahlung cross section goes roughly as $1/\nu$ over most of the range, while for the pair production case the distribution goes as ν^{-3} to ν^{-2} [69]. "Hard" losses are therefore more probable in bremsstrahlung, and in fact energy losses due to pair production may very nearly be treated as continuous. The simulated [67] momentum distribution of an incident 1 TeV/c muon beam after it crosses 3 m of iron is shown in Fig. 27.23. The most probable loss is 8 GeV, or 3.4 MeV g^{-1}cm^2. The full width at half maximum is 9 GeV/c, or 0.9%. The radiative tail is almost entirely due to bremsstrahlung, although most of the events in which more than 10% of the incident energy lost experienced relatively hard photonuclear interactions. The latter can exceed detector resolution [70], necessitating the reconstruction of lost energy. Tables [1] list the stopping power as 9.82 MeV g^{-1}cm^2 for a 1 TeV muon, so that the mean loss should be 23 GeV ($\approx$ 23 GeV/c), for a final momentum of 977 GeV/c, far below the peak. This agrees with the indicated mean calculated from the simulation. Electromagnetic and hadronic cascades in detector materials can obscure muon tracks in detector planes and reduce tracking efficiency [71].

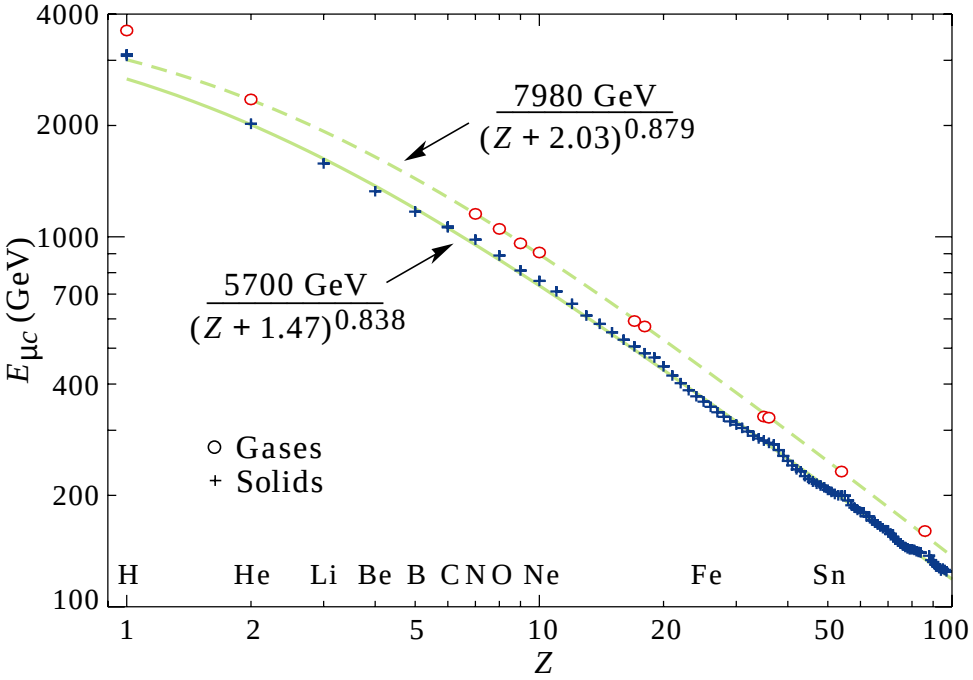


Figure 27.22: Muon critical energy for the chemical elements, defined as the energy at which radiative and ionization energy loss rates are equal [1]. The equality comes at a higher energy for gases than for solids or liquids with the same atomic number because of a smaller density effect reduction of the ionization losses. The fits shown in the figure exclude hydrogen. Alkali metals fall 3–4% above the fitted function, while most other solids are within 2% of the function. Among the gases the worst fit is for radon (2.7% high).

27.7. Cherenkov and transition radiation [72,73,82]

A charged particle radiates if its velocity is greater than the local phase velocity of light (Cherenkov radiation) or if it crosses suddenly from one medium to another with different optical properties (transition radiation). Neither process is important for energy loss, but both are used in high-energy physics detectors.

Cherenkov Radiation. The angle θ_c of Cherenkov radiation, relative to the particle's direction, for a particle with velocity βc in a medium with index of refraction n is

$$\begin{aligned} & \cos\theta_c = (1/n\beta) \\ \text{or} \quad & \tan\theta_c = \sqrt{\beta^2 n^2 - 1} \\ & \approx \sqrt{2(1 - 1/n\beta)} \quad \text{for small } \theta_c, \textit{e.g.} \text{ in gases.} \end{aligned} \tag{27.38}$$

The threshold velocity β_t is $1/n$, and $\gamma_t = 1/(1-\beta_t^2)^{1/2}$. Therefore, $\beta_t\gamma_t = 1/(2\delta+\delta^2)^{1/2}$, where $\delta = n - 1$. Values of δ for various commonly used gases are given as a function of pressure and wavelength in Ref. 74. For values at atmospheric pressure, see Table 6.1. Data for other commonly used materials are given in Ref. 75.

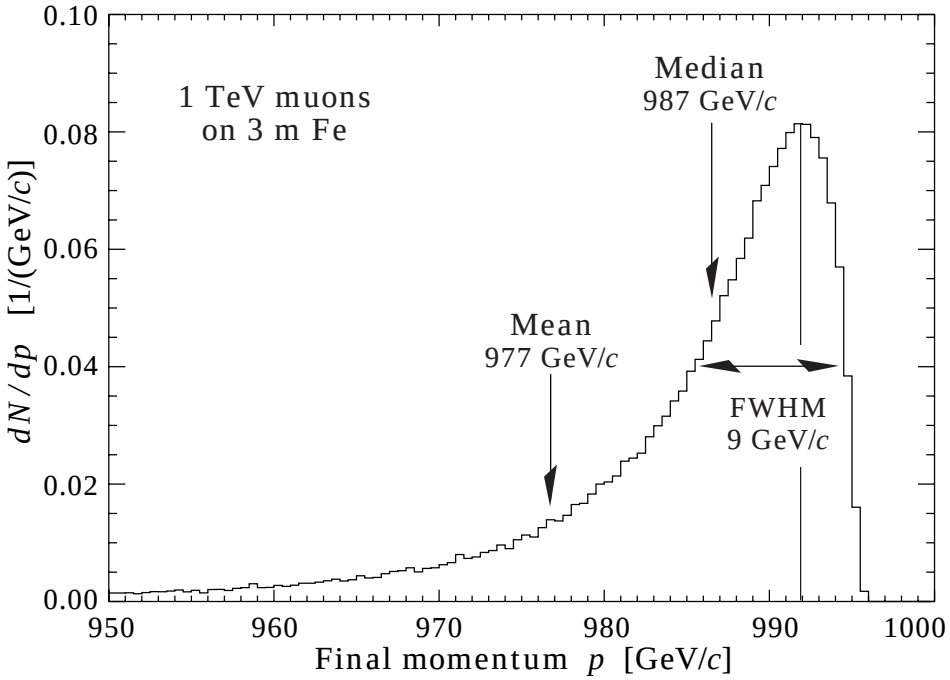


Figure 27.23: The momentum distribution of 1 TeV/c muons after traversing 3 m of iron as calculated with the MARS15 Monte Carlo code [67] by S.I. Striganov [1].

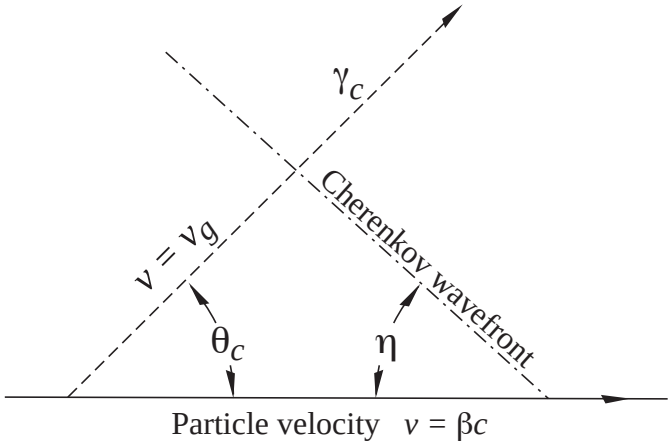


Figure 27.24: Cherenkov light emission and wavefront angles. In a dispersive medium, $\theta_c + \eta \neq 90^0$.

Practical Cherenkov radiator materials are dispersive. Let ω be the photon's frequency, and let $k = 2\pi/\lambda$ be its wavenumber. The photons propage at the group velocity $v_g = d\omega/dk = c/[n(\omega) + \omega(dn/d\omega)]$. In a non-dispersive medium, this simplies to $v_g = c/n$.

In his classical paper, Tamm [76] showed that for dispersive media the radiation is concentrated in a thin conical shell whose vertex is at the moving charge, and whose

opening half-angle η is given by

$$\cot\eta = \left[\frac{d}{d\omega}(\omega\tan\theta_c)\right]_{\omega_0}$$
$$= \left[\tan\theta_c + \beta^2\omega\, n(\omega)\,\frac{dn}{d\omega}\cot\theta_c\right]_{\omega_0} , \qquad (27.39)$$

where ω_0 is the central value of the small frequency range under consideration. (See Fig. 27.24.) This cone has a opening half-angle η, and, unless the medium is non-dispersive ($dn/d\omega = 0$), $\theta_c + \eta \neq 90^0$. The Cherenkov wavefront 'sideslips' along with the particle [77]. This effect may have timing implications for ring imaging Cherenkov counters [78], but it is probably unimportant for most applications.

The number of photons produced per unit path length of a particle with charge ze and per unit energy interval of the photons is

$$\frac{d^2N}{dEdx} = \frac{\alpha z^2}{\hbar c}\sin^2\theta_c = \frac{\alpha^2 z^2}{r_e\, m_e c^2}\left(1 - \frac{1}{\beta^2 n^2(E)}\right)$$
$$\approx 370\sin^2\theta_c(E)\ \mathrm{eV}^{-1}\mathrm{cm}^{-1} \qquad (z=1)\ , \qquad (27.40)$$

or, equivalently,

$$\frac{d^2N}{dxd\lambda} = \frac{2\pi\alpha z^2}{\lambda^2}\ \left(1 - \frac{1}{\beta^2 n^2(\lambda)}\right)\ . \qquad (27.41)$$

The index of refraction is a function of photon energy $E = \hbar\omega$, as is the sensitivity of the transducer used to detect the light. For practical use, Eq. (27.40) must be multiplied by the the transducer response function and integrated over the region for which $\beta\, n(\omega) > 1$. Further details are given in the discussion of Cherenkov detectors in the Particle Detectors section (Sec. 28 of this *Review*).

When two particles are close together (within $\lesssim 1$ wavelength), the electromagnetic fields from the particles may add coherently, affecting the Cherenkov radiation. The radiation from an e^+e^- pair at close separation is suppressed compared to two independent leptons [79].

Coherent radio Cherenkov radiation from electromagnetic showers (containing a net excess of e^- over e^+) is significant [80], and has been used to study cosmic ray air showers [81] and to search for ν_e induced showers.

Transition radiation. The energy radiated when a particle with charge ze crosses the boundary between vacuum and a medium with plasma frequency ω_p is

$$I = \alpha z^2\gamma\hbar\omega_p/3\ , \qquad (27.42)$$

where

$$\hbar\omega_p = \sqrt{4\pi N_e r_e^3}\ m_e c^2/\alpha = \sqrt{4\pi N_e a_\infty^3}\ 2\times 13.6\ \mathrm{eV}\ . \qquad (27.43)$$

Here N_e is the electron density in the medium, r_e is the classical electron radius, and a_∞ is the Bohr radius. For styrene and similar materials, $\sqrt{4\pi N_e a_\infty^3} \approx 0.8$, so that $\hbar\omega_p \approx 20$ eV. The typical emission angle is $1/\gamma$.

The radiation spectrum is logarithmically divergent at low energies and decreases rapidly for $\hbar\omega/\gamma\hbar\omega_p > 1$. About half the energy is emitted in the range $0.1 \leq \hbar\omega/\gamma\hbar\omega_p \leq 1$. For a particle with $\gamma = 10^3$, the radiated photons are in the soft x-ray range 2 to 20 keV. The γ dependence of the emitted energy thus comes from the hardening of the spectrum rather than from an increased quantum yield. For a typical radiated photon energy of $\gamma\hbar\omega_p/4$, the quantum yield is

$$N_\gamma \approx \frac{1}{2}\,\frac{\alpha z^2 \gamma\hbar\omega_p}{3}\Big/\frac{\gamma\hbar\omega_p}{4} \approx \tfrac{2}{3}\alpha z^2 \approx 0.5\% \times z^2\ . \tag{27.44}$$

More precisely, the number of photons with energy $\hbar\omega > \hbar\omega_0$ is given by [82]

$$N_\gamma(\hbar\omega > \hbar\omega_0) = \frac{\alpha z^2}{\pi}\left[\left(\ln\frac{\gamma\hbar\omega_p}{\hbar\omega_0} - 1\right)^2 + \frac{\pi^2}{12}\right]\ , \tag{27.45}$$

within corrections of order $(\hbar\omega_0/\gamma\hbar\omega_p)^2$. The number of photons above a fixed energy $\hbar\omega_0 \ll \gamma\hbar\omega_p$ thus grows as $(\ln\gamma)^2$, but the number above a fixed fraction of $\gamma\hbar\omega_p$ (as in the example above) is constant. For example, for $\hbar\omega > \gamma\hbar\omega_p/10$, $N_\gamma = 2.519\,\alpha z^2/\pi = 0.59\% \times z^2$.

The yield can be increased by using a stack of plastic foils with gaps between. However, interference can be important, and the soft x rays are readily absorbed in the foils. The first problem can be overcome by choosing thicknesses and spacings large compared to the "formation length" $D = \gamma c/\omega_p$, which in practical situations is tens of μm. Other practical problems are discussed in Sec. 28.

References:

1. D.E. Groom, N.V. Mokhov, and S.I. Striganov, "Muon stopping-power and range tables: 10 MeV–100 TeV" Atomic Data and Nuclear Data Tables **78**, 183–356 (2001). Since submission of this paper it has become likely that post-Born corrections to the direct pair production cross section should be made. Code used to make Figs. 27.20, 27.21, and 27.12 included these corrections [D.Yu. Ivanov *et al.*, Phys. Lett. **B442**, 453 (1998)]. The effect is negligible for except at high Z. (It is less than 1% for iron);
   More extensive printable and machine-readable tables are given at `http://pdg.lbl.gov/AtomicNuclearProperties/`.
2. "Stopping Powers and Ranges for Protons and Alpha Particles," ICRU Report No. 49 (1993);
   Tables and graphs of these data are available at `http://physics.nist.gov/PhysRefData/`.
3. W.H. Barkas, W. Birnbaum, and F.M. Smith, Phys. Rev. **101**, 778 (1956).
4. B. Rossi, *High Energy Particles*, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1952.
5. U. Fano, Ann. Rev. Nucl. Sci. **13**, 1 (1963).
6. J. D. Jackson, Phys. Rev. **D59**, 017301 (1999).
7. S.M. Seltzer and M.J. Berger, Int. J. of Applied Rad. **33**, 1189 (1982).
8. "Stopping Powers for Electrons and Positrons," ICRU Report No. 37 (1984).
9. `http://physics.nist.gov/PhysRefData/XrayMassCoef/tab1.html`.
10. H. Bichsel, Phys. Rev. **A46**, 5761 (1992).

11. W.H. Barkas and M.J. Berger, *Tables of Energy Losses and Ranges of Heavy Charged Particles*, NASA-SP-3013 (1964).
12. M. Agnello *et al.*, Phys. Rev. Lett. **74**, 371 (1995).
13. H.H. Andersen and J.F. Ziegler, *Hydrogen: Stopping Powers and Ranges in All Elements.* Vol. 3 of *The Stopping and Ranges of Ions in Matter* (Pergamon Press 1977).
14. J. Lindhard, Kgl. Danske Videnskab. Selskab, Mat.-Fys. Medd. **28**, No. 8 (1954); J. Lindhard, M. Scharff, and H.E. Schiøtt, Kgl. Danske Videnskab. Selskab, Mat.-Fys. Medd. **33**, No. 14 (1963).
15. J.F. Ziegler, J.F. Biersac, and U. Littmark, *The Stopping and Range of Ions in Solids*, Pergamon Press 1985.
16. R.M. Sternheimer, Phys. Rev. **88**, 851 (1952).
17. A. Crispin and G.N. Fowler, Rev. Mod. Phys. **42**, 290 (1970).
18. R.M. Sternheimer, S.M. Seltzer, and M.J. Berger, "The Density Effect for the Ionization Loss of Charged Particles in Various Substances," Atomic Data and Nuclear Data Tables **30**, 261 (1984). Minor errors are corrected in Ref. 1. Chemical composition for the tabulated materials is given in Ref. 7.
19. For unit-charge projectiles, see E.A. Uehling, Ann. Rev. Nucl. Sci. **4**, 315 (1954). For highly charged projectiles, see J.A. Doggett and L.V. Spencer, Phys. Rev. **103**, 1597 (1956). A Lorentz transformation is needed to convert these center-of-mass data to knock-on energy spectra.
20. R.M. Sternheimer and R.F. Peierls, Phys. Rev. **B3**, 3681 (1971).
21. N.F. Mott and H.S.W. Massey, *The Theory of Atomic Collisions*, Oxford Press, London, 1965.
22. L.D. Landau, J. Exp. Phys. (USSR) **8**, 201 (1944).
23. P.V. Vavilov, Sov. Phys. JETP **5**, 749 (1957).
24. H. Bichsel, Rev. Mod. Phys. **60**, 663 (1988).
25. H. Bichsel, "A method to improve tracking and particle identification in TPCs and silicon detectors," Nucl. Inst. & Meth. A (in press, 2006).
26. R. Talman, Nucl. Instrum. Methods **159**, 189 (1979).
27. S.M. Seltzer and M.J. Berger, Int. J. of Applied Rad. **35**, 665 (1984). This paper corrects and extends the results of Ref. 7.
28. L.V. Spencer "Energy Dissipation by Fast Electrons," Nat'l Bureau of Standards Monograph No. 1 (1959).
29. "Average Energy Required to Produce an Ion Pair," ICRU Report No. 31 (1979).
30. N. Hadley *et al.*, "List of Poisoning Times for Materials," Lawrence Berkeley Lab Report TPC-LBL-79-8 (1981).
31. H.A. Bethe, Phys. Rev. **89**, 1256 (1953). A thorough review of multiple scattering is given by W.T. Scott, Rev. Mod. Phys. **35**, 231 (1963). However, the data of Shen *et al.*, (Phys. Rev. **D20**, 1584 (1979)) show that Bethe's simpler method of including atomic electron effects agrees better with experiment that does Scott's treatment. For a thorough discussion of simple formulae for single scatters and methods of compounding these into multiple-scattering formulae, see W.T. Scott, Rev. Mod.

Phys. **35**, 231 (1963). Detailed summaries of formulae for computing single scatters are given in J.W. Motz, H. Olsen, and H.W. Koch, Rev. Mod. Phys. **36**, 881 (1964).
32. V.L. Highland, Nucl. Instrum. Methods **129**, 497 (1975), and Nucl. Instrum. Methods **161**, 171 (1979).
33. G.R. Lynch and O.I Dahl, Nucl. Instrum. Methods **B58**, 6 (1991).
34. M. Wong *et al.*, Med. Phys. **17**, 163 (1990).
35. E. Segré, *Nuclei and Particles*, New York, Benjamin (1964) p. 65 ff.
36. Y.S. Tsai, Rev. Mod. Phys. **46**, 815 (1974).
37. H. Davies, H.A. Bethe, and L.C. Maximon, Phys. Rev. **93**, 788 (1954).
38. O.I. Dahl, private communication.
39. L.D. Landau and I.J. Pomeranchuk, Dokl. Akad. Nauk. SSSR **92**, 535 (1953); **92**, 735 (1953). These papers are available in English in L. Landau, *The Collected Papers of L.D. Landau*, Pergamon Press, 1965; A.B. Migdal, Phys. Rev. **103**, 1811 (1956).
40. S. Klein, Rev. Mod. Phys. **71**, 1501 (1999).
41. M. L. Ter-Mikaelian, SSSR **94**, 1033 (1954);
M. L. Ter-Mikaelian, *High Energy Electromagnetic Processes in Condensed Media* (John Wiley & Sons, New York, 1972).
42. P. Anthony *et al.*, Phys. Rev. Lett. **76**, 3550 (1996).
43. H. W. Koch and J. W. Motz, Rev. Mod. Phys. **31**, 920 (1959).
44. M.J. Berger and S.M. Seltzer, "Tables of Energy Losses and Ranges of Electrons and Positrons," National Aeronautics and Space Administration Report NASA-SP-3012 (Washington DC 1964).
45. K. Hikasa *et al.*, *Review of Particle Properties*, Phys. Rev. **D46** (1992) S1.
46. B. L. Berman and S. C. Fultz, Rev. Mod. Phys. **47**, 713 (1975).
47. J.S. Hubbell, H. Gimm, and I Øverbø, J. Phys. Chem. Ref. Data **9**, 1023 (1980).
48. A. Veyssiere *et al.*, Nucl. Phys. **A159**, 561 (1970).
49. J. W. Motz, H. A. Olsen, and H. W. Koch, Rev. Mod. Phys. **41**, 581 (1969).
50. P. Anthony *et al.*, Phys. Rev. Lett. **75**, 1949 (1995).
51. U.I. Uggerhoj, Rev. Mod. Phys. **77**, 1131 (2005).
52. W.R. Nelson, H. Hirayama, and D.W.O. Rogers, "The EGS4 Code System," SLAC-265, Stanford Linear Accelerator Center (Dec. 1985).
53. *Experimental Techniques in High Energy Physics*, ed. by T. Ferbel (Addision-Wesley, Menlo Park CA 1987).
54. U. Amaldi, Phys. Scripta **23**, 409 (1981).
55. E. Longo and I. Sestili, Nucl. Instrum. Methods **128**, 283 (1975).
56. G. Grindhammer *et al.*, in *Proceedings of the Workshop on Calorimetry for the Supercollider*, Tuscaloosa, AL, March 13–17, 1989, edited by R. Donaldson and M.G.D. Gilchriese (World Scientific, Teaneck, NJ, 1989), p. 151.
57. W.R. Nelson, T.M. Jenkins, R.C. McCall, and J.K. Cobb, Phys. Rev. **149**, 201 (1966).
58. G. Bathow *et al.*, Nucl. Phys. **B20**, 592 (1970).
59. H.A. Bethe and W. Heitler, *Proc. Roy. Soc.* **A146**, 83 (1934);
H.A. Bethe, *Proc. Cambridge Phil. Soc.* **30**, 542 (1934).

60. A.A. Petrukhin and V.V. Shestakov, Can. J. Phys. **46**, S377 (1968).
61. V.M. Galitskii and S.R. Kel'ner, Sov. Phys. JETP **25**, 948 (1967).
62. S.R. Kel'ner and Yu.D. Kotov, Sov. J. Nucl. Phys. **7**, 237 (1968).
63. R.P. Kokoulin and A.A. Petrukhin, in *Proceedings of the International Conference on Cosmic Rays*, Hobart, Australia, August 16–25, 1971, Vol. **4**, p. 2436.
64. A.I. Nikishov, Sov. J. Nucl. Phys. **27**, 677 (1978).
65. Y.M. Andreev *et al.*, Phys. Atom. Nucl. **57**, 2066 (1994).
66. L.B. Bezrukov and E.V. Bugaev, Sov. J. Nucl. Phys. **33**, 635 (1981).
67. N.V. Mokhov, "The MARS Code System User's Guide," Fermilab-FN-628 (1995);
N. V. Mokhov *et al.*, Radiation Protection and Dosimetry, vol. 116, part 2, pp. 99 (2005);
Fermilab-Conf-04/053 (2004);
N. V. Mokhov *et al.*, in *Proc. of Intl. Conf. on Nuclear Data for Science and Tech*, (Santa Fe, NM, 2004), AIP C onf. Proc. 769, part 2, p. 1618;
Fermilab-Conf-04/269-AD (2004); `http://www-ap.fnal.gov/MARS/`.
68. P.H. Barrett, L.M. Bollinger, G. Cocconi, Y. Eisenberg, and K. Greisen, Rev. Mod. Phys. **24**, 133 (1952).
69. A. Van Ginneken, Nucl. Instrum. Methods **A251**, 21 (1986).
70. U. Becker *et al.*, Nucl. Instrum. Methods **A253**, 15 (1986).
71. J.J. Eastman and S.C. Loken, in *Proceedings of the Workshop on Experiments, Detectors, and Experimental Areas for the Supercollider*, Berkeley, CA, July 7–17, 1987, edited by R. Donaldson and M.G.D. Gilchriese (World Scientific, Singapore, 1988), p. 542.
72. *Methods of Experimental Physics*, L.C.L. Yuan and C.-S. Wu, editors, Academic Press, 1961, Vol. 5A, p. 163.
73. W.W.M. Allison and P.R.S. Wright, "The Physics of Charged Particle Identification: dE/dx, Cherenkov Radiation, and Transition Radiation," p. 371 in *Experimental Techniques in High Energy Physics*, T. Ferbel, editor, (Addison-Wesley 1987).
74. E.R. Hayes, R.A. Schluter, and A. Tamosaitis, "Index and Dispersion of Some Cherenkov Counter Gases," ANL-6916 (1964).
75. T. Ypsilantis, "Particle Identification at Hadron Colliders", CERN-EP/89-150 (1989), or ECFA 89-124, **2** 661 (1989).
76. I. Tamm, J. Phys. U.S.S.R., **1**, 439 (1939).
77. H. Motz and L. I. Schiff, Am. J. Phys. **21**, 258 (1953).
78. B. N. Ratcliff, Nucl. Instrum & Meth. **A502**, 211 (2003).
79. S. K. Mandal, S. R. Klein, and J. D. Jackson, Phys. Rev. **D72**, 093003 (2005).
80. E. Zas, F. Halzen and T. Stanev, Phys. Rev. **D45**, 362 (1991).
81. H. Falcke *et al*, Nature **435**, 313 (2005).
82. J.D. Jackson, *Classical Electrodynamics*, 3rd edition, (John Wiley & Sons, New York, 1998).

28. PARTICLE DETECTORS

Revised 2006 (see the various sections for authors).

In this section we give various parameters for common detector components. The quoted numbers are usually based on typical devices, and should be regarded only as rough approximations for new designs. More detailed discussions of detectors and their underlying physics can be found in books by Ferbel [1], Grupen [2], Kleinknecht [3], Knoll [4], and Green [5]. In Table 28.1 are given typical spatial and temporal resolutions of common detectors.

Table 28.1: Typical spatial and temporal resolutions of common detectors. Revised September 2003 by R. Kadel (LBNL).

Detector Type	Accuracy (rms)	Resolution Time	Dead Time
Bubble chamber	10–150 μm	1 ms	50 ms[a]
Streamer chamber	300 μm	2 μs	100 ms
Proportional chamber	50–300 μm[b,c,d]	2 ns	200 ns
Drift chamber	50–300 μm	2 ns[e]	100 ns
Scintillator	—	100 ps/n[f]	10 ns
Emulsion	1 μm	—	—
Liquid Argon Drift [Ref. 6]	$\sim$175–450 μm	$\sim$ 200 ns	$\sim$ 2 μs
Gas Micro Strip [Ref. 7]	30–40 μm	$<$ 10 ns	—
Resistive Plate chamber [Ref. 8]	$\lesssim$10 μm	1–2 ns	—
Silicon strip	pitch/(3 to 7)[g]	*h*	*h*
Silicon pixel	2 μm[i]	*h*	*h*

[a] Multiple pulsing time.

[b] 300 μm is for 1 mm pitch.

[c] Delay line cathode readout can give $\pm$150 μm parallel to anode wire.

[d] wirespacing/$\sqrt{12}$.

[e] For two chambers.

[f] n = index of refraction.

[g] The highest resolution ("7") is obtained for small-pitch detectors ($\lesssim$25 μm) with pulse-height-weighted center finding.

[h] Limited by the readout electronics [9]. (Time resolution of $\leq$ 25 ns is planned for the ATLAS SCT.)

[i] Analog readout of 34 μm pitch, monolithic pixel detectors.

28.1. Photon detectors

Written August 2005 by D. Chakraborty (Northern Illinois U) and T. Sumiyoshi (Tokyo Metropolitan U).

Most detectors in high-energy, nuclear, and astrophysics rely on the detection of photons in or near the visible range, $100\,\text{nm} \lesssim \lambda \lesssim 1000\,\text{nm}$, or $E \approx$ a few eV. This range covers scintillation and Cherenkov radiation as well as the light detected in many astronomical observations.

Generally, photodetection involves generating a detectable electrical signal proportional to the (usually very small) number of incident photons. The process involves three distinct steps:

1. Generation of a primary photoelectron or electron-hole (e-h) pair by an incident photon by the photoelectric or photoconductive effect,
2. Amplification of the p.e. signal to detectable levels by one or more multiplicative bombardment steps and/or an avalanche process (usually), and,
3. Collection of the secondary electrons to form the electrical signal.

The important characteristics of a photodetector include the following in statistical averages:

1. Quantum efficiency (QE or ϵ_Q): the number of primary photoelectrons generated per incident photon ($0 \le \epsilon_Q \le 1$; in silicon more than one e-h pair per incident photon can be generated for $\lambda \lesssim 165$ nm),
2. Collection efficiency (CE or ϵ_C): the overall acceptance factor other than the generation of photoelectrons ($0 \le \epsilon_C \le 1$),
3. Gain (G): the number of electrons collected for each photoelectron generated,
4. Dark current or dark noise: the electrical signal when there is no photon,
5. Energy resolution: electronic noise (ENC or N_e) and statistical fluctuations in the amplification process compound the Poisson distribution of n_γ photons from a given source:
$$\frac{\sigma(E)}{\langle E \rangle} = \sqrt{\frac{f_N}{n_\gamma \epsilon_Q \epsilon_C} + \left(\frac{N_e}{G n_\gamma \epsilon_Q \epsilon_C} \right)^2}, \tag{28.1}$$
where f_N, or the excess noise factor (ENF), is the contribution to the energy distribution variance due to amplification statistics [10],
6. Dynamic range: the maximum signal available from the detector (this is usually expressed in units of the response to noise-equivalent power, or NEP, which is the optical input power that produces a signal-to-noise ratio of 1),
7. Time dependence of the response: this includes the transit time, which is the time between the arrival of the photon and the electrical pulse, and the transit time spread, which contributes to the pulse rise time and width, and
8. Rate capability: inversely proportional to the time needed, after the arrival of one photon, to get ready to receive the next.

The QE is a strong function of the photon wavelength (λ), and is usually quoted at maximum, together with a range of λ where the QE is comparable to its maximum.

Spatial uniformity and linearity with respect to the number of photons are highly desirable in a photodetector's response.

Optimization of these factors involves many trade-offs and vary widely between applications. For example, while a large gain is desirable, attempts to increase the gain for a given device also increases the ENF and after-pulsing ("echos" of the main pulse). In solid-state devices, a higher QE often requires a compromise in the timing properties. In other types, coverage of large areas by focusing increases the transit time spread.

Other important considerations also are highly application-specific. These include the photon flux and wavelength range, the total area to be covered and the efficiency required, the volume available to accommodate the detectors, characteristics of the environment such as chemical composition, temperature, magnetic field, ambient background, as well as ambient radiation of different types and, mode of operation (continuous or triggered), bias (high-voltage) requirements, power consumption, calibration needs, aging, cost, and so on. Several technologies employing different phenomena for the three steps described above, and many variants within each, offer a wide range of solutions to choose from. The salient features of the main technologies and the common variants are described below. Some key characteristics are summarized in Table 28.2.

28.1.1. ***Vacuum photodetectors*** : Vacuum photodetectors can be broadly subdivided into three types: photomultiplier tubes, microchannel plates, and hybrid photodetectors.

28.1.1.1. *Photomultiplier tubes:* A versatile class of photon detectors, vacuum photomultiplier tubes (PMT) has been employed by a vast majority of all particle physics experiments to date [10]. Both "transmission-" and "reflection-type" PMT's are widely used. In the former, the photocathode material is deposited on the inside of a transparent window through which the photons enter, while in the latter, the photocathode material rests on a separate surface that the incident photons strike. The cathode material has a low work function, chosen for the wavelength band of interest. When a photon hits the cathode and liberates an electron (the photoelectric effect), the latter is accelerated and guided by electric fields to impinge on a secondary-emission electrode, or dynode, which then emits a few (~ 5) secondary electrons. The multiplication process is repeated typically 10 times in series to generate a sufficient number of electrons, which are collected at the anode for delivery to the external circuit. The total gain of a PMT depends on the applied high voltage V as $G = AV^{kn}$, where $k \approx 0.7$–0.8 (depending on the dynode material), n is the number of dynodes in the chain, and A a constant (which also depends on n). Typically, G is in the range of 10^5–10^6. Pulse risetimes are usually in the few nanosecond range. With *e.g.* two-level discrimination the effective time resolution can be much better.

A large variety of PMT's, including many just recently developed, covers a wide span of wavelength ranges from infrared (IR) to extreme ultraviolet (XUV) [11]. They are categorized by the window materials, photocathode materials, dynode structures, anode configurations, *etc.* Common window materials are borosilicate glass for IR to near-UV, fused quartz and sapphire (Al_2O_3) for UV, and MgF_2 or LiF for XUV. The choice of photocathode materials include a variety of mostly Cs- and/or Sb-based compounds such as CsI, CsTe, bi-alkali (SbRbCs, SbKCs), multi-alkali ($SbNa_2KCs$), GaAs(Cs),

Table 28.2: Representative characteristics of some photodetectors commonly used in particle physics. The time resolution of the devices listed here vary in the 10–2000 ps range.

Type	λ (nm)	$\epsilon_Q\,\epsilon_C$	Gain	Risetime (ns)	Area (mm^2)	1-p.e noise (Hz)	HV (V)	Price (USD)
PMT*	115–1100	0.15–0.25	10^3–10^7	0.7–10	10^2–10^5	10–10^4	500–3000	100–5000
MCP*	100–650	0.01–0.10	10^3–10^7	0.15–0.3	10^2–10^4	0.1–200	500–3500	10–6000
HPD*	115–850	0.1–0.3	10^3–10^4	7	10^2–10^5	10–10^3	$\sim 2\times10^4$	$\sim$600
GPM*	115–500	0.15–0.3	10^3–10^6	$O(0.1)$	$O(10)$	10–10^3	300–2000	$O(10)$
APD	300–1700	$\sim$0.7	10–10^8	$O(1)$	10–10^3	1–10^3	400–1400	$O(100)$
SiPM	400–550	0.15–0.3	10^5–10^6	~ 1	1–10	$O(10^6)$	30–60	$O(10)$
VLPC	500–600	$\sim$0.9	$\sim 5\times10^4$	~ 10	1	$O(10^4)$	$\sim$7	$\sim$1

∗These devices often come in multi-anode configurations. In such cases, area, noise, and price are to be considered on a "per readout-channel" basis.

GaAsP, *etc.* Sensitive wavelengths and peak quantum efficiencies for these materials are summarized in Table 28.3. Typical dynode structures used in PMT's are circular cage, line focusing, box and grid, venetian blind, and fine mesh. In some cases, limited spatial resolution can be obtained by using a mosaic of multiple anodes.

PMT's are vulnerable to magnetic fields—sometimes even the geomagnetic field causes large orientation-dependent gain changes. A high-permeability metal shield is often necessary. However, proximity-focused PMT's, *e.g.* the fine-mesh types, can be used even in a high magnetic field (≥ 1 T) if the electron drift direction is parallel to the field.

28.1.1.2. *Microchannel plates:* A typical Microchannel plate (MCP) photodetector consists of one or more $\sim$2 mm thick glass plates with densely packed $O(10\ \mu$m)-diameter cylindrical holes, or "channels", sitting between the transmission-type photocathode and anode planes, separated by $O(1$ mm) gaps. Instead of discrete dynodes, the inner surface of each cylindrical tube serves as a continuous dynode for the entire cascade of multiplicative bombardments initiated by a photoelectron. Gain fluctuations can be minimized by operating in a saturation mode, whence each channel is only capable of a binary output, but the sum of all channel outputs remains proportional to the number of photons received so long as the photon flux is low enough to ensure that the probability of a single channel receiving more than one photon during a single time gate is negligible. MCP's are thin, offer good spatial resolution, have excellent time resolution ($\sim$20 ps), and can tolerate random magnetic fields up to 0.1 T and axial fields up to ~ 1 T. However, they suffer from relatively long recovery time per channel and short lifetime. MCP's are widely employed as image-intensifiers, although not so much in HEP or astrophysics.

28.1.1.3. *Hybrid photon detectors:* Hybrid photon detectors (HPD) combine the sensitivity of a vacuum PMT with the excellent spatial and energy resolutions of a Si sensor [19]. A single photoelectron ejected from the photocathode is accelerated through a potential difference of ∼20 kV before it impinges on the silicon sensor/anode. The gain nearly equals the maximum number of e-h pairs that could be created from the entire kinetic energy of the accelerated electron: $G \approx eV/w$, where e is the electronic charge, V is the applied potential difference, and $w \approx 3.7$ eV is the mean energy required to create an e-h pair in Si at room temperature. Since the gain is achieved in a single step, one might expect to have the excellent resolution of a simple Poisson statistic with large mean, but in fact it is even better, thanks to the Fano effect discussed in Sec. 28.11.

Low-noise electronics must be used to read out HPD's if one intends to take advantage of the low fluctuations in gain, *e.g.* when counting small numbers of photons. HPD's can have the same $\epsilon_Q\,\epsilon_C$ and window geometries as PMT's and can be segmented down to ∼50 μm. However, they require rather high biases and will not function in a magnetic field. The exception is proximity-focused devices (⇒ no (de)magnification) in an axial field. Current applications of HPD's include the CMS hadronic calorimeter and the RICH detector in LHCb.

Table 28.3: Properties of photocathode and window materials commonly used in vacuum photodetectors [11].

Photocathode material	λ (nm)	Window material	Peak ϵ_Q (λ/nm)
CsI	115–200	MgF_2	0.15 (135)
CsTe	115–240	MgF_2	0.18 (210)
Bi-alkali	300–650	Borosilicate	0.27 (390)
	160-650	Quartz	0.27 (390)
Multi-alkali	300–850	Borosilicate	0.20 (360)
	160-850	Quartz	0.23 (280)
GaAs(Cs)*	160–930	Quartz	0.23 (280)
GaAsP(Cs)	300-750	Borosilicate	0.42 (560)

*Reflection type photocathode is used.

28.1.2. ***Gaseous photon detectors*** : In gaseous photomultipliers (GPM) a photoelectron in a suitable gas mixture initiates an avalanche in a high-field region, producing a large number of secondary impact-ionization electrons. In principle the charge multiplication and collection processes are identical to those employed in gaseous tracking detectors such as multiwire proportional chambers, micromesh gaseous detectors (Micromegas), or gas electron multipliers (GEM). These are discussed in Sections 28.7 and 28.8.

The devices can be divided into two types depending on the photocathode material. One type uses solid photocathode materials much in the same way as PMT's. Since it is resistant to gas mixtures typically used in tracking chambers, CsI is a common choice. In the other type, photoionization occurs on suitable molecules vaporized and mixed in the drift volume. Most gases have photoionization work functions in excess of 10 eV, which would limit their sensitivity to wavelengths far too short. However, vapors of TMAE (tetrakis dimethyl-amine ethylene) or TEA (tri-ethyl-amine), which have smaller work functions (5.3 eV for TMAE and 7.5 eV for TEA), are suited for XUV photon detection [12]. Since devices like GEM's offer sub-mm spatial resolution, GPM's are often used as position-sensitive photon detectors. They can be made into flat panels to cover large areas ($O(1\ \mathrm{m}^2)$), can operate in high magnetic fields, and are relatively inexpensive. Many of the ring imaging Cherenkov (RICH) detectors to date have used GPM's for the detection of Cherenkov light [13]. Special care must be taken to suppress the photon-feedback process in GPM's. It is also important to maintain high purity of the gas as minute traces of O_2 can significantly degrade the detection efficiency.

28.1.3. ***Solid-state photon detectors*** : In a phase of rapid development, solid-state photodetectors are competing with vacuum- or gas-based devices for many existing applications and making way for a multitude of new ones. Compared to traditional vacuum- and gaseous photodetectors, solid-state devices are more compact, lightweight, rugged, tolerant to magnetic fields, and often cheaper. They also allow fine pixelization, are easy to integrate into large systems, and can operate at low electric potentials, while matching or exceeding most performance criteria. They are particularly well suited for detection of γ- and X-rays. Except for applications where coverage of very large areas or dynamic range is required, solid-state detectors are proving to be the better choice. Some hybrid devices attempt to combine the best features of different technologies while applications of nanotechnology are opening up exciting new possibilities.

Silicon photodiodes (PD) are widely used in high-energy physics as particle detectors and in a great number of applications (including solar cells!) as light detectors. The structure is discussed in some detail in Sec. 28.11. In its simplest form, the PD is a reverse-biased *p*-*n* junction. Photons with energies above the indirect bandgap energy (wavelengths shorter than about 1050 nm, depending on the temperature) can create *e*-*h* pairs (the photoconductive effect), which are collected on the *p* and *n* sides, respectively. Often, as in the PD's used for crystal scintillator readout in CLEO, L3, Belle, BaBar, and GLAST, intrinsic silicon is doped to create a *p*-*i*-*n* structure. The reverse bias increases the thickness of the depleted region; in the case of these particular detectors, to full depletion at a depth of about 100 μm. Increasing the depletion depth decreases the capacitance (and hence electronic noise) and extends the red response. Quantum efficiency can exceed 90%, but falls toward the red because of the increasing absorption length of light in silicon. The absorption length reaches 100 μm at 985 nm. However, since $G = 1$, amplification is necessary. Optimal low-noise amplifiers are slow, but, even so, noise limits the minimum detectable signal in room-temperature devices to several hundred photons.

Very large arrays containing $O(10^7)$ of $O(10\ \mu\mathrm{m}^2)$-sized photodiodes pixelizing a plane are widely used to photograph all sorts of things from everyday subjects at visible

wavelengths to crystal structures with X-rays and astronomical objects from infrared to UV. To limit the number of readout channels, these are made into charge-coupled devices (CCD), where pixel-to-pixel signal transfer takes place over thousands of synchronous cycles with sequential output through shift registers [14]. Thus, high spatial resolution is achieved at the expense of speed and timing precision. Custom-made CCD's have virtually replaced photographic plates and other imagers for astronomy and in spacecraft. Typical QE's exceed 90% over much of the visible spectrum, and "thick" CCD's have useful QE up to $\lambda = 1$ μm. Active Pixel Sensor (APS) arrays with a preamplifier on each pixel and CMOS processing afford higher speeds, but are challenged at longer wavelengths. Much R&D is underway to overcome the limitations of both CCD and CMOS imagers.

In avalanche photodiodes (APD), an exponential cascade of impact ionizations initiated by the initial photogenerated e-h pair under a large reverse-bias voltage leads to an avalanche breakdown [15]. As a result, detectable electrical response can be obtained from low-intensity optical signals down to single photons. Excellent junction uniformity is critical, and a guard ring is generally used as a protection against edge breakdown. Well-designed APD's, such as those used in CMS' crystal-based electromagnetic calorimeter, have achieved $\epsilon_Q\,\epsilon_C \approx 0.7$ with sub-ns response time. The sensitive wavelength window and gain depend on the semiconductor used. The gain is typically 10–200 in linear and up to 10^8 in Geiger mode of operation. Stability and close monitoring of the operating temperature are important for linear-mode operation, and substantial cooling is often necessary. Position-sensitive APD's use time information at multiple anodes to calculate the hit position.

One of the most promising recent developments in the field is that of devices consisting of large arrays ($O(10^3)$) of tiny APD's packed over a small area ($O(1\ \mathrm{mm}^2)$) and operated in a limited Geiger mode [16]. Of the many names for this class of photodetectors, "SiPM" (for "Silicon PhotoMultiplier") seems to be the most common at the moment. Although each cell only offers a binary output, linearity with respect to the number of photons is achieved by summing the cell outputs in the same way as with a MCP in saturation mode (see above). While SiPM's have yet to go into commercial production, prototypes are being studied by the thousands for various purposes including medical imaging, *e.g.* positron emission tomography (PET). These compact, rugged, and economical devices allow auto-calibration through decent separation of photoelectron peaks and offer gains of $O(10^6)$ at a moderate bias voltage ($\sim$50 V). However, the single-photoelectron noise of a SiPM, being the logical "or" of $O(10^3)$ Geiger APD's, is rather large: $O(1\ \mathrm{MHz/mm}^2)$ at room temperature. Also, the recovery time for each cell can be several microseconds. SiPM's are particularly well-suited for applications where triggered pulses of 10 or more photoelectrons are expected over a small area, *e.g.* fiber-guided scintillation light. Intense R&D is expected to lower the noise level and improve the device-to-device uniformity, resulting in coverage of larger areas and wider applications. Attempts are being made to combine the fabrication of the sensors and the front-end electronics (ASIC) in the same process with the goal of making SiPMs and other finely pixelized solid-state photodetectors extremely easy to use.

Of late, much R&D has been directed to p-i-n diode arrays based on thin polycrys-

talline diamond films formed by chemical vapor deposition (CVD) on a hot substrate (∼1000 K) from a hydrocarbon-containing gas mixture under low pressure (∼100 mbar). These devices have maximum sensitivity in the extreme- to moderate-UV region [17]. Many desirable characteristics, including high tolerance to radiation and temperature fluctuations, low dark noise, blindness to most of the solar radiation spectrum, and relatively low cost make them ideal for space-based UV/XUV astronomy, measurement of synchrotron radiation, and luminosity monitoring at (future) lepton collider(s).

Visible-light photon counters (VLPC) utilize the formation of an impurity band only 50 meV below the conduction band in As-doped Si to generate strong ($G \approx 5 \times 10^4$) yet sharp response to single photons with $\epsilon_Q \approx 0.9$ [18]. The smallness of the band gap considerably reduces the gain dispersion. Only a very small bias (∼7 V) is needed, but high sensitivity to infrared photons requires cooling below 10 K. The dark noise increases sharply and exponentially with both temperature and bias. The Run 2 DØ detector uses 86000 VLPC's to read the optical signal from its scintillating-fiber tracker and scintillator-strip preshower detectors.

28.2. Organic scintillators

Revised September 2001 by K.F. Johnson (FSU).

Organic scintillators are broadly classed into three types, crystalline, liquid, and plastic, all of which utilize the ionization produced by charged particles (see Sec. 27.2) of this *Review*) to generate optical photons, usually in the blue to green wavelength regions [20]. Plastic scintillators are by far the most widely used. Crystal organic scintillators are practically unused in high-energy physics.

Densities range from 1.03 to 1.20 g cm^{-3}. Typical photon yields are about 1 photon per 100 eV of energy deposit [21]. A one-cm-thick scintillator traversed by a minimum-ionizing particle will therefore yield $\approx 2 \times 10^4$ photons. The resulting photoelectron signal will depend on the collection and transport efficiency of the optical package and the quantum efficiency of the photodetector.

Plastic scintillators do not respond linearly to the ionization density. Very dense ionization columns emit less light than expected on the basis of dE/dx for minimum-ionizing particles. A widely used semi-empirical model by Birks posits that recombination and quenching effects between the excited molecules reduce the light yield [22]. These effects are more pronounced the greater the density of the excited molecules. Birks' formula is

$$\frac{d\mathscr{L}}{dx} = \mathscr{L}_0 \frac{dE/dx}{1 + k_B\, dE/dx} \ ,$$

where $\mathscr{L}$ is the luminescence, $\mathscr{L}_0$ is the luminescence at low specific ionization density, and k_B is Birks' constant, which must be determined for each scintillator by measurement.

Decay times are in the ns range; rise times are much faster. The combination of high light yield and fast response time allows the possibility of sub-ns timing resolution [23]. The fraction of light emitted during the decay "tail" can depend on the exciting particle. This allows pulse shape discrimination as a technique to carry out particle identification. Because of the hydrogen content (carbon to hydrogen ratio ≈ 1) plastic scintillator is

sensitive to proton recoils from neutrons. Ease of fabrication into desired shapes and low cost has made plastic scintillators a common detector component. Recently, plastic scintillators in the form of scintillating fibers have found widespread use in tracking and calorimetry [24].

28.2.1. ***Scintillation mechanism*** :
Scintillation: A charged particle traversing matter leaves behind it a wake of excited molecules. Certain types of molecules, however, will release a small fraction ($\approx 3\%$) of this energy as optical photons. This process, scintillation, is especially marked in those organic substances which contain aromatic rings, such as polystyrene (PS) and polyvinyltoluene (PVT). Liquids which scintillate include toluene and xylene.

Fluorescence: In fluorescence, the initial excitation takes place via the absorption of a photon, and de-excitation by emission of a longer wavelength photon. Fluors are used as "waveshifters" to shift scintillation light to a more convenient wavelength. Occurring in complex molecules, the absorption and emission are spread out over a wide band of photon energies, and have some overlap, that is, there is some fraction of the emitted light which can be re-absorbed [25]. This "self-absorption" is undesirable for detector applications because it causes a shortened attenuation length. The wavelength difference between the major absorption and emission peaks is called the Stokes' shift. It is usually the case that the greater the Stokes' shift, the smaller the self absorption—thus, a large Stokes' shift is a desirable property for a fluor (aka the "Better red than dead" strategy).

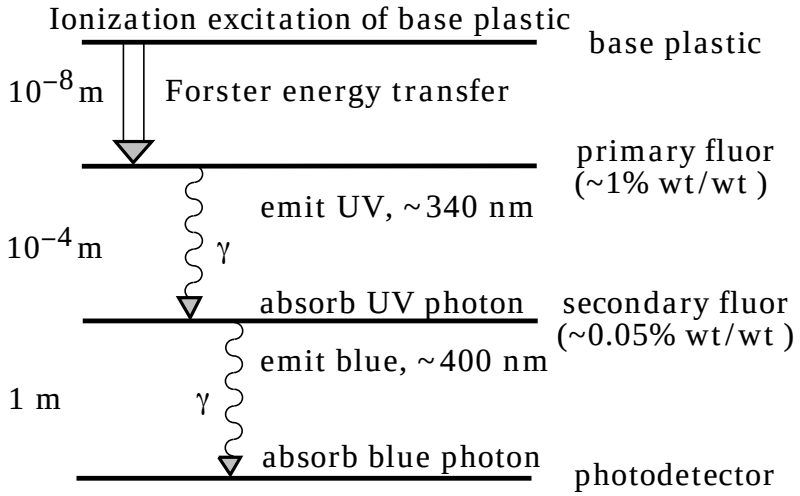


Figure 28.1: Cartoon of scintillation "ladder" depicting the operating mechanism of plastic scintillator. Approximate fluor concentrations and energy transfer distances for the separate sub-processes are shown.

Scintillators: The plastic scintillators used in high-energy physics are binary or ternary solutions of selected fluors in a plastic base containing aromatic rings. (See the appendix in Ref. 26 for a comprehensive list of components.) Virtually all plastic scintillators contain as a base either PVT or PS. PVT-based scintillator can be up to 50% brighter.

The fluors must satisfy additional conditions besides being fluorescent. They must be sufficiently stable, soluble, chemically inert, fast, radiation tolerant, and efficient.

Ionization in the plastic base produces UV photons with short attenuation length (several mm). Longer attenuation lengths are obtained by dissolving a "primary" fluor in high concentration (1% by weight) into the base, which is selected to efficiently re-radiate absorbed energy at wavelengths where the base is more transparent.

The primary fluor has a second important function. The decay time of the scintillator base material can be quite long—in pure polystyrene it is 16 ns, for example. The addition of the primary fluor in high concentration can shorten the decay time by an order of magnitude and increase the total light yield. At the concentrations used (1% and greater), the average distance between a fluor molecule and an excited base unit is around 100 Å, much less than a wavelength of light. At these distances the predominant mode of energy transfer from base to fluor is not the radiation of a photon, but a resonant dipole-dipole interaction, first described by Foerster, which strongly couples the base and fluor [27]. The strong coupling sharply increases the speed and the light yield of the plastic scintillators.

Unfortunately, a fluor which fulfills other requirements is usually not completely adequate with respect to emission wavelength or attenuation length, so it is necessary to add yet another waveshifter (the "secondary" fluor), at fractional percent levels, and occasionally a third (not shown in Fig. 28.1).

External wavelength shifters: Light emitted from a plastic scintillator may be absorbed in a (nonscintillating) base doped with a wave-shifting fluor. Such wavelength shifters are widely used to aid light collection in complex geometries. The wavelength shifter must be insensitive to ionizing radiation and Cherenkov light. A typical wavelength shifter uses an acrylic base because of its good optical qualities, a single fluor to shift the light emerging from the plastic scintillator to the blue-green, and contains ultra-violet absorbing additives to deaden response to Cherenkov light.

28.2.2. ***Caveats and cautions*** : Plastic scintillators are reliable, robust, and convenient. However, they possess quirks to which the experimenter must be alert.

Aging and Handling: Plastic scintillators are subject to aging which diminishes the light yield. Exposure to solvent vapors, high temperatures, mechanical flexing, irradiation, or rough handling will aggravate the process. A particularly fragile region is the surface which can 'craze"—develop microcracks—which rapidly destroy the capability of plastic scintillators to transmit light by total internal reflection. Crazing is particularly likely where oils, solvents, or *fingerprints* have contacted the surface.

Attenuation length: The Stokes' shift is not the only factor determining attenuation length. Others are the concentration of fluors (the higher the concentration of a fluor, the greater will be its self-absorption); the optical clarity and uniformity of the bulk material; the quality of the surface; and absorption by additives, such as stabilizers, which may be present.

Afterglow: Plastic scintillators have a long-lived luminescence which does not follow a

simple exponential decay. Intensities at the 10^{-4} level of the initial fluorescence can persist for hundreds of ns [20,28].

Atmospheric quenching: Plastic scintillators will decrease their light yield with increasing partial pressure of oxygen. This can be a 10% effect in an artificial atmosphere [29]. It is not excluded that other gases may have similar quenching effects.

Magnetic field: The light yield of plastic scintillators may be changed by a magnetic field. The effect is very nonlinear and apparently not all types of plastic scintillators are so affected. Increases of $\approx 3\%$ at 0.45 T have been reported [30]. Data are sketchy and mechanisms are not understood.

Radiation damage: Irradiation of plastic scintillators creates color centers which absorb light more strongly in the UV and blue than at longer wavelengths. This poorly understood effect appears as a reduction both of light yield and attenuation length. Radiation damage depends not only on the integrated dose, but on the dose rate, atmosphere, and temperature, before, during and after irradiation, as well as the materials properties of the base such as glass transition temperature, polymer chain length, *etc.* Annealing also occurs, accelerated by the diffusion of atmospheric oxygen and elevated temperatures. The phenomena are complex, unpredictable, and not well understood [31]. Since color centers are less intrusive at longer wavelengths, the most reliable method of mitigating radiation damage is to shift emissions at every step to the longest practical wavelengths, *e.g.*, utilize fluors with large Stokes' shifts (aka the "Better red than dead" strategy).

28.2.3. *Scintillating and wavelength-shifting fibers* :

The clad optical fiber is an incarnation of scintillator and wavelength shifter (WLS) which is particularly useful [32]. Since the initial demonstration of the scintillating fiber (SCIFI) calorimeter [33], SCIFI techniques have become mainstream. SCIFI calorimeters are found, for example, in the $g-2$ experiment at Brookhaven [34] and at KLOE; SCIFI trackers are found at CHORUS and DØ ; WLS readout is used in both ATLAS and CMS hadron calorimeters [35].

SCIFI calorimeters are fast, dense, radiation hard, and can have leadglass-like resolution. SCIFI trackers can handle high rates and are radiation tolerant, but the low photon yield at the end of a long fiber (see below) forces the use of very sophisticated photodetectors such as VLPC's, such as are used in DØ . WLS scintillator readout of a calorimeter allows a very high level of hermeticity since the solid angle blocked by the fiber on its way to the photodetector is very small. The sensitive region of scintillating fibers can be controlled by splicing them onto clear (non-scintillating/non-WLS) fibers.

A typical configuration would be fibers with a core of polystyrene-based scintillator or WLS (index of refraction $n = 1.59$), surrounded by a cladding of PMMA ($n = 1.49$) a few microns thick, or, for added light capture, with another cladding of fluorinated PMMA with $n = 1.42$, for an overall diameter of 0.5 to 1 mm. The fiber is drawn from a boule and great care is taken during production to ensure that the intersurface between the core and the cladding has the highest possible uniformity and quality, so that the signal transmission via total internal reflection has a low loss. The fraction of generated light

which is transported down the optical pipe is denoted the capture fraction and is about 6% for the single-clad fiber and 10% for the double-clad fiber.

The number of photons from the fiber available at the photodetector is always smaller than desired, and increasing the light yield has proven difficult [36]. A minimum-ionizing particle traversing a high-quality 1 mm diameter fiber perpendicular to its axis will produce fewer than 2000 photons, of which about 200 are captured. Attenuation eliminates about 95% of these photons. DØ uses 0.775 mm diameter scintillating fibers in the tracker and obtains 9 photoelectrons with the VLPC reaching 85% quantum efficiency.

A scintillating or WLS fiber is often characterized by its "attenuation length," over which the signal is attenuated to $1/e$ of its original value. Many factors determine the attenuation length, including the importance of re-absorption of emitted photons by the polymer base or dissolved fluors, the level of crystallinity of the base polymer, and the quality of the total internal reflection boundary. Attenuation lengths of several meters are obtained by high quality fibers. However, it should be understood that the attenuation length is not necessarily a measure of fiber quality. Among other things, it is not constant with distance from the excitation source and it is wavelength dependent. So-called "cladding light" causes some of the distance dependence [37], but not all. The wavelength dependence is usually related to the higher re-absorption of shorter wavelength photons—once absorbed, re-emitted isotropically and lost with 90% probability—and to the lower absorption of longer wavelengths by polystyrene. Experimenters should be aware that measurements of attenuation length by a phototube with a bialkali photocathode, whose quantum efficiency drops below 10% at 480 nm, should not be naïvely compared to measurements utilizing a silicon photodiode, whose quantum efficiency is still rising at 600 nm.

28.3. Inorganic scintillators:

Revised September 2005 by C.L. Woody (BNL). and R.-Y. Zhu (California Inst. of Technology).

Inorganic crystals form a class of scintillating materials with much higher densities than organic plastic scintillators (typically $\sim$ 4–8 g/cm^3) with a variety of different properties for use as scintillation detectors. Due to their high density and high effective atomic number, they can be used in applications where high stopping power or a high conversion efficiency for electrons or photons is required. These include total absorption electromagnetic calorimeters (see Sec. 28.13.1), which consist of a totally active absorber (as opposed to a sampling calorimeter), as well as serving as gamma ray detectors over a wide range of energies. Many of these crystals also have very high light output, and can therefore provide excellent energy resolution down to very low energies ($\sim$ few hundred keV).

Some crystals are intrinsic scintillators in which the luminescence is produced by a part of the crystal lattice itself. However, other crystals require the addition of a dopant, typically fluorescent ions such as thallium (Tl) or cerium (Ce) which is responsible for producing the scintillation light. However, in both cases, the scintillation mechanism is

the same. Energy is deposited in the crystal by ionization, either directly by charged particles, or by the conversion of photons into electrons or positrons which subsequently produce ionization. This energy is transferred to the luminescent centers which then radiate scintillation photons. The efficiency η for the conversion of energy deposit in the crystal to scintillation light can be expressed by the relation [38]

$$\eta = \beta \cdot S \cdot Q \ . \tag{28.2}$$

where β is the efficiency of the energy conversion process, S is the efficiency of energy transfer to the luminescent center, and Q is the quantum efficiency of the luminescent center. The value of η ranges between 0.1 and ~ 1 depending on the crystal, and is the main factor in determining the intrinsic light output of the scintillator. In addition, the scintillation decay time is primarily determined by the energy transfer and emission process. The decay time of the scintillator is mainly dominated by the decay time of the luminescent center. For example, in the case of thallium doped sodium iodide (NaI(Tl)), the value of η is ~ 0.5, which results in a light output $\sim$ 40,000 photons per MeV of energy deposit. This high light output is largely due to the high quantum efficiency of the thallium ion (Q ~ 1), but the decay time is rather slow ($\tau \sim 250$ ns).

Table 28.4 lists the basic properties of some commonly used inorganic crystal scintillators. NaI(Tl) is one of the most common and widely used scintillators, with an emission that is well matched to a bialkali photomultiplier tube, but it is highly hygroscopic and difficult to work with, and has a rather low density. CsI(Tl) has high light yield, an emission that is well matched to solid state photodiodes, and is mechanically robust (high plasticity and resistance to cracking). However, it needs careful surface treatment and is slightly hygroscopic. Compared with CsI(Tl), pure CsI has identical mechanical properties, but faster emission at shorter wavelengths and light output approximately an order of magnitude lower. BaF_2 has a fast component with a sub-nanosecond decay time, and is the fastest known scintillator. However, it also has a slow component with a much longer decay time ($\sim$ 630 ns). Bismuth gemanate ($Bi_4Ge_3O_{12}$ or BGO) has a very high density, and consequently a short radiation length X_0 and Molière radius R_M. BGO's emission is well-matched to the spectral sensitivity of photodiodes, and it is easy to handle and not hygroscopic. Lead tungstate ($PbWO_4$ or PWO) has a very high density, with a very short X_0 and R_M, but its intrinsic light yield is rather low. Both cerium doped lutetium oxyorthosilicate (Lu_2SiO_5:Ce, or LSO:Ce) [39] and cerium doped gadolinium orthosilicate (Gd_2SiO_5:Ce, or GSO:Ce) [40] are dense crystal scintillators which have a high light yield and a fast decay time.

Beside the crystals listed in Table 28.4, a number of new crystals are being developed that may have potential applications in high energy or nuclear physics. Of particular interest is the family of yttrium and lutetium perovskites, which include YAP ($YAlO_3$:Ce) and LuAP ($LuAlO_3$:Ce) and their mixed compositions. These have been shown to be linear over a large energy range [41], and have the potential for providing extremely good intrinsic energy resolution. In addition, other fluoride crystals such as CeF_3 have been shown to provide excellent energy resolution in calorimeter applications.

Table 28.4 gives the light output of other crystals relative to NaI(Tl) as measured with a bialkalai photomultiplier tube. However, the useful signal produced by a scintillator is

usually quoted in terms of the number of photoelectrons per MeV produced by a given photodetector. The relationship between the number of photons/MeV produced and photoelectrons/MeV detected involves the factors for the light collection efficiency L and the quantum efficiency QE of the photodetector:

$$N_{\text{p.e.}}/\text{MeV} = L \cdot QE \cdot N_{\gamma}/\text{MeV} \tag{28.3}$$

L includes the transmission of scintillation light within the crystal (*i.e.*, the bulk attenuation length of the material), reflections and scattering from the surfaces, and the size and shape of the crystal. These factors can vary considerably depending on the sample, but can be in the range of ∼50–60%. However, the internal light transmission depends on the intrinsic properties of the material, as well as the number and type of impurities and defects that can produce internal absorption within the crystal, and can be highly affected by factors such as radiation damage, as discussed below.

The quantum efficiency depends on the type of photodetector used to detect the scintillation light, which is typically ∼15–20% for photomultiplier tubes and ∼70% for silicon photodiodes for visible wavelengths. The quantum efficiency of the detector is usually highly wavelength dependent, as shown in Fig. 28.2, and should be matched to the particular crystal of interest to give the highest quantum yield at the wavelength corresponding to the peak of the scintillation emission. The comparison of the light output given in Table 28.4 is for a standard photomultiplier tube with a bialkali photocathode. Results with different photodetectors can be significantly different. For example, the response of CsI(Tl) relative to NaI(Tl) with a silicon photodiode would be 140 rather than 45 due to its higher quantum efficiency at longer wavelengths. For scintillators which emit in the UV, a detector with a quartz window should be used.

One important issue related to the application of a crystal scintillator is its radiation hardness. Stability of its light output, or the ability to track and monitor the variation of its light output in a radiation environment, is required for high resolution and precision calibration [42]. All known crystal scintillators suffer from radiation damage. A common damage phenomenon is the appearance of radiation induced absorption caused by the formation of impurities or point defect related color centers. This radiation induced absorption reduces the light attenuation length in the crystal, and hence its light output. For crystals with high defect density, a severe reduction of light attenuation length may lead to a distortion of the light response uniformity, leading to a degradation of energy resolution. Additional radiation damage effects may include a reduced intrinsic scintillation light yield (damage to the luminescent centers) and an increased phosphorescence (afterglow). For crystals to be used in the construction a high precision calorimeter in a radiation environment, its scintillation mechanism must not be damaged and its light attenuation length in the expected radiation environment must be long enough so that its light response uniformity, and thus its energy resolution, does not change [43].

Table 28.4: Properties of several inorganic crystal scintillators. Most of the notation is defined in Sec. 6 of this *Review*.

Parameter: ρ	MP	X_0^*	R_M^*	dE/dx	λ_I^*	$\tau_{\rm decay}$	$\lambda_{\rm max}$	$n^\natural$	Relative output†	Hygro-scopic?	$d({\rm LY})/dT$
Units: g/cm^3	°C	cm	cm	MeV/cm	cm	ns	nm				%/°C‡
NaI(Tl) 3.67	651	2.59	4.13	4.8	42.9	230	410	1.85	100	yes	∼0
BGO 7.13	1050	1.12	2.23	9.0	22.8	300	480	2.15	9	no	−1.6
BaF_2 4.89	1280	2.03	3.10	6.6	30.7	630^s	300^s	1.50	21^s	no	−2^s
						0.9^f	220^f		2.7^f		∼0^f
CsI(Tl) 4.51	621	1.86	3.57	5.6	39.3	1300	560	1.79	45	slight	0.3
CsI(pure) 4.51	621	1.86	3.57	5.6	39.3	35^s	420^s	1.95	5.6^s	slight	−0.6
						6^f	310^f		2.3^f		
$PbWO_4$ 8.3	1123	0.89	2.00	10.2	20.7	50^s	560^s	2.20	0.1^s	no	−1.9
						10^f	420^f		0.6^f		
LSO(Ce) 7.40	2070	1.14	2.07	9.6	20.9	40	420	1.82	75	no	∼0
GSO(Ce) 6.71	1950	1.38	2.23	8.9	22.2	600^s	430	1.85	3^s	no	−0.1
						56^f			30^f		

* Numerical values calculated using formulae in this review.
$^\natural$ Refractive index at the wavelength of the emission maximum.
† Relative light yield measured with a bi-alkali cathode PMT.
‡ Variation of light yield with temperature evaluated at room temperature.
f = fast component, s = slow component

Most of the crystals listed in Table 28.4 have been used in high energy or nuclear physics experiments when the ultimate energy resolution for electrons and photons is desired. Examples are the Crystal Ball NaI(Tl) calorimeter at SPEAR, the L3 BGO calorimeter at LEP, the CLEO CsI(Tl) calorimeter at CESR, the KTeV CsI calorimeter at the Tevatron, and the BaBar and BELLE CsI(Tl) calorimeters at PEP-II and KEK. Because of its high density and low cost, PWO calorimeters are now being constructed by CMS and ALICE at LHC, by CLAS and PrimEx at CEBAF, and by BTeV at the Tevatron.

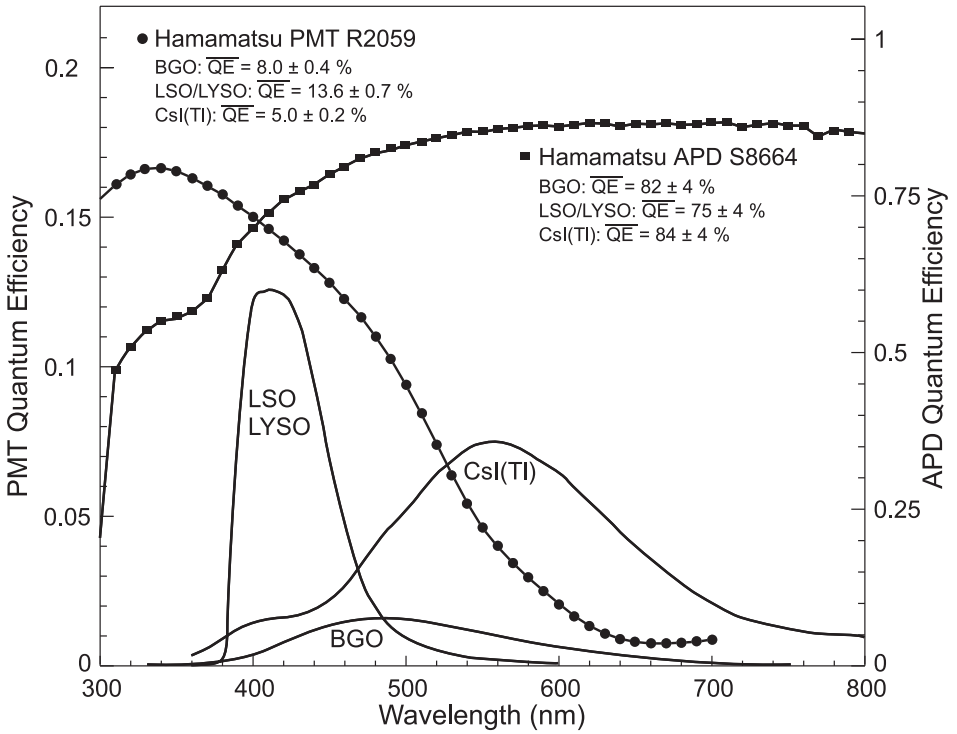


Figure 28.2: The quantum efficiencies of two photodetectors, a Hamamatsu R2059 PMT with bi-alkali cathode and a Hamamatsu S8664 avalanche photodiode (APD), are shown as a function of wavelength. Also shown in the figure are emission spectra of three crystal scintillators, BGO, LSO and CsI(Tl), and the numerical values of the emission weighted quantum efficiency. The area under each emission spectrum is proportional to crystal's light yield.

28.4. Cherenkov detectors

Written September 2003 by B.N. Ratcliff (SLAC).

Although devices using Cherenkov radiation are often thought of as particle identification (PID) detectors, in practice, they are widely used over a much broader range of applications; including (1) fast particle counters; (2) hadronic particle identification; and (3) tracking detectors performing complete event reconstruction. A few examples of specific applications from each category include; (1) the polarization detector of the SLD [44]; (2) the hadronic PID detectors at the B factory detectors (DIRC in BaBar [8] and the aerogel threshold Cherenkov in Belle [45]) ; and (3) large water Cherenkov counters such as Super-Kamiokande [46]. Cherenkov counters contain two main elements; (1) a radiator through which the charged particle passes, and (2) a photodetector. As Cherenkov radiation is a weak source of photons, light collection and detection must be as efficient as possible. The presence of the refractive index n and the path length of the particle in the radiator in the Cherenkov relations allows tuning these quantities for a particular experimental application.

Cherenkov detectors utilize one or more of the properties of Cherenkov radiation

discussed in the Passages of Particles through Matter section (Sec. 27 of this *Review*): the prompt emission of a light pulse; the existence of a velocity threshold for radiation; and the dependence of the Cherenkov cone half-angle θ_c and the number of emitted photons on the velocity of the particle.

The number of photoelectrons ($N_{\rm p.e.}$) detected in a given device is

$$N_{\rm p.e.} = L\frac{\alpha^2 z^2}{r_e\, m_e c^2}\int \epsilon(E)\, \sin^2\theta_c(E) dE\ , \tag{28.4}$$

where L is the path length in the radiator, ϵ(E) is the efficiency for collecting the Cherenkov light and transducing it in photoelectrons, and $\alpha^2/(r_e\, m_e c^2) = 370\ {\rm cm}^{-1}{\rm eV}^{-1}$.

The quantities ϵ and θ_c are functions of the photon energy E. However, since the typical energy dependent variation of the index of refraction is modest, a quantity called the *Cherenkov detector quality factor* N_0 can be defined as

$$N_0 = \frac{\alpha^2 z^2}{r_e\, m_e c^2}\ \int \epsilon\, dE\ , \tag{28.5}$$

so that

$$N_{\rm p.e.} \approx L N_0 \langle \sin^2\theta_c\rangle\ . \tag{28.6}$$

We take $z = 1$, the usual case in high-energy physics, in the following discussion.

This definition of the quality factor N_0 is not universal, nor, indeed, very useful for situations where the geometrical photon collection efficiency ($\epsilon_{\rm coll}$) varies substantially for different tracks. In this case, separate factors for photon collection and detection ($\epsilon_{\rm det}$), so that $\epsilon = \epsilon_{\rm coll}\epsilon_{\rm det}$, are sometimes included on the right hand side of the equation. A typical value of N_0 for a photomultiplier (PMT) detection system working in the visible and near UV, and collecting most of the Cherenkov light, is about 100 ${\rm cm}^{-1}$. Practical counters, utilizing a variety of different photodetectors, have values ranging between about 30 and 180 ${\rm cm}^{-1}$. Radiators can be chosen from a variety of transparent materials (Sec. 27 of this *Review* and Table 6.1). In addition to refractive index, the choice requires consideration of factors such as material density, radiation length, transmission bandwidth, absorption length, chromatic dispersion, optical workability (for solids), availability, and cost. Long radiator lengths are required to obtain sufficient numbers of photons when the momenta of the particle species to be separated are high. Recently, the gap in refractive index that has traditionally existed between gases and liquid or solid materials has been partially closed with transparent *silica aerogels* with indices that range between about 1.007 and 1.13.

Cherenkov counters may be classified as either *imaging* or *threshold* types, depending on whether they do or do not make use of Cherenkov angle (θ_c) information. Imaging counters may be used to track particles as well as identify them.

28.4.1. ***Threshold counters:*** : Threshold Cherenkov detectors [47], in their simplest form, make a yes/no decision based on whether the particle is above or below the Cherenkov threshold velocity $\beta_t = 1/n$. A straightforward enhancement of such detectors uses the number of observed photoelectrons (or a calibrated pulse height) to discriminate between species or to set probabilities for each particle species [48]. This strategy can increase the momentum range of particle separation by a modest amount (to a momentum some 20% above the threshold momentum of the heavier particle in a typical case).

Careful designs give $\langle \epsilon_{\rm coll} \rangle \gtrsim 90\%$. For a photomultiplier with a typical bialkali cathode, $\int \epsilon_{\rm det} dE \approx 0.27$, so that

$$N_{\rm p.e.}/L \approx 90 \text{ cm}^{-1} \langle \sin^2 \theta_c \rangle \qquad (i.e., \, N_0 = 90 \text{ cm}^{-1}) \; . \tag{28.7}$$

Suppose, for example, that n is chosen so that the threshold for species a is p_t; that is, at this momentum species a has velocity $\beta_a = 1/n$. A second, lighter, species b with the same momentum has velocity β_b, so $\cos\theta_c = \beta_a/\beta_b$, and

$$N_{\rm p.e.} \, L \approx 90 \text{ cm}^{-1} \, \frac{m_a^2 - m_b^2}{p_t^2 + m_a^2} \; . \tag{28.8}$$

For K/π separation at $p = p_t = 1(5)$ GeV/c, $N_{\rm p.e.}/L \approx 16(0.8)$ cm^{-1} for π's and (by design) 0 for K's.

For limited path lengths $N_{\rm p.e.}$ can be small, and a minimum number is required to trigger external electronics. The overall efficiency of the device is controlled by Poisson fluctuations, which can be especially critical for separation of species where one particle type is dominant. The effective number of photoelectrons is often less than the average number calculated above due to additional equivalent noise from the photodetector. It is common to design for at least 10 photoelectrons for the high velocity particle in order to obtain a robust counter. As rejection of the particle that is below threshold depends on *not* seeing a signal, electronic and other background noise can be important. Physics sources of light production for the below threshold particle, such as decay of the above threshold particle or the production of delta rays in the radiator, often limit the separation attainable, and need to be carefully considered. Well designed, modern multi-channel counters, such as the ACC at Belle [45], can attain good particle separation performance over a substantial momentum range for essentially the full solid angle of the spectrometer.

28.4.2. ***Imaging counters*** : The most powerful use of the information available from the Cherenkov process comes from measuring the ring-correlated angles of emission of the individual Cherenkov photons. Since low-energy photon detectors can measure only the position (and, perhaps, a precise detection time) of the individual Cherenkov photons (not the angles directly), the photons must be "imaged" onto a detector so that their angles can be derived [49]. In most cases the optics map the Cherenkov cone onto (a portion of) a distorted circle at the photodetector. Though this imaging process is directly analogous to the familiar imaging techniques used in telescopes and other optical instruments, there is a somewhat bewildering variety of methods used in a wide variety of counter types with different names. Some of the imaging methods used include (1)

focusing by a lens; (2) proximity focusing (i.e., focusing by limiting the emission region of the radiation); and (3) focusing through an aperture (a pinhole). In addition, the prompt Cherenkov emission coupled with the speed of modern photon detectors allows the use of time imaging, a method which is used much less frequently in conventional imaging technology. Finally, full tracking (and event reconstruction) can be performed in large water counters by combining the individual space position and time of each photon together with the constraint that Cherenkov photons are emitted from each track at a constant polar angle.

In a simple model of an imaging PID counter, the fractional error on the particle velocity (δ_β) is given by

$$\delta_\beta = \frac{\sigma_\beta}{\beta} = \tan\theta_c \sigma(\theta_c) \quad , \tag{28.9}$$

where

$$\sigma(\theta_c) = \frac{\langle\sigma(\theta_i)\rangle}{\sqrt{N_{\rm p.e.}}} \oplus C \ , \tag{28.10}$$

where $\langle\sigma(\theta_i)\rangle$ is the average single photoelectron resolution, as defined by the optics, detector resolution and the intrinsic chromaticity spread of the radiator index of refraction averaged over the photon detection bandwidth. C combines a number of other contributions to resolution including,(1) correlated terms such as tracking, alignment, and multiple scattering, (2) hit ambiguities, (3) background hits from random sources, and (4) hits coming from other tracks. In many practical cases, the resolution is limited by these effects.

For a $\beta \approx 1$ particle of momentum (p) well above threshold entering a radiator with index of refraction (n), the number of σ separation (N_σ) between particles of mass m_1 and m_2 is approximately

$$N_\sigma \approx \frac{|m_1^2 - m_2^2|}{2p^2\sigma(\theta_c)\sqrt{n^2-1}} \ . \tag{28.11}$$

In practical counters, the angular resolution term $\sigma(\theta_c)$ varies between about 0.1 and 5 mrad depending on the size, radiator, and photodetector type of the particular counter. The range of momenta over which a particular counter can separate particle species extends from the point at which the number of photons emitted becomes sufficient for the counter to operate efficiently as a threshold device (∼20% above the threshold for the lighter species) to the value in the imaging region given by the equation above. For example, for $\sigma(\theta_c) = 2$mrad, a fused silica radiator($n = 1.474$), or a flourocarbon gas radiator (C_5F_{12}, $n = 1.0017$), would separate π/K's from the threshold region starting around 0.15(3) GeV/c through the imaging region up to about 4.2(18) GeV/c at better than 3σ.

Many different imaging counters have been built during the last several decades [52]. Among the earliest examples of this class of counters are the very limited acceptance Differential Cherenkov detectors, designed for particle selection in high momentum beam lines. These devices use optical focusing and/or geometrical masking to select particles having velocities in a specified region. With careful design, a velocity resolution of $\sigma_\beta/\beta \approx 10^{-4}$–$10^{-5}$ can be obtained [47].

Practical multi-track Ring-Imaging Cherenkov detectors (generically called RICH counters) are a more recent development. They have been built in small-aperture and 4π geometries both as PID counters and as stand-alone detectors with complete tracking and event reconstruction as discussed more fully below. PID RICH counters are sometimes further classified by 'generations' that differ based on performance, design, and photodetection techniques.

A typical example of a first generation RICH used at the Z factory e^+e^- colliders [50,51] has both liquid (C_6F_{14}, $n = 1.276$) and gas (C_5F_{12}, $n = 1.0017$) radiators, the former being proximity imaged using the small radiator thickness while the latter use mirrors. The phototransducers are a TPC/wire-chamber combination having charge division or pads. They are made sensitive to photons by doping the TPC gas (usually, ethane/methane) with $\sim 0.05\%$ TMAE (tetrakis(dimethylamino)ethylene). Great attention to detail is required, (1) to avoid absorbing the UV photons to which TMAE is sensitive, (2) to avoid absorbing the single photoelectrons as they drift in the long TPC, and (3) to keep the chemically active TMAE vapor from interacting with materials in the system. In spite of their unforgiving operational characteristics, these counters attained good $e/\pi/K/p$ separation over wide momentum ranges during several years of operation. In particular, their π/K separation range extends over momenta from about 0.25 to 20 GeV/c.

Second and third generation counters [52] generally must operate at much higher particle rates than the first generation detectors, and utilize different photon detection bandwidths, with higher readout channel counts, and faster, more forgiving photon detection technology than the TMAE doped TPCs just described. Radiator choices have broadened to include materials such as lithium flouride, fused silica, and aerogel. Vacuum based photodetection systems (*e.g.*, photomultiplier tubes (PMT) or hybrid photodiodes (HPD)) have become increasingly common. They handle very high rates, can be used in either single or multi anode versions, and allow a wide choice of radiators. Other fast detection systems that use solid cesium iodide (CSI) photocathodes or triethylamine (TEA) doping in proportional chambers are useful with certain radiator types and geometries.

A DIRC (Detector of Internally Reflected Cherenkov light) is a third generation subtype of a RICH first used in the BaBar detector [8]. It "inverts" the usual principle for use of light from the radiator of a RICH by collecting and imaging the total internally reflected light, rather than the transmitted light. A DIRC utilizes the optical material of the radiator in two ways, simultaneously; first as a Cherenkov radiator, and second, as a light pipe for the Cherenkov light trapped in the radiator by total internal reflection. The DIRC makes use of the fact that the magnitudes of angles are preserved during reflection from a flat surface. This fact, coupled with the high reflection coefficients of the total internal reflection process (> 0.9995 for higly polished SiO_2), and the long attenuation length for photons in high purity fused silica, allows the photons of the ring image to be transported to a detector outside the path of the particle where they may be imaged. The BaBar DIRC uses 144 fused silica radiator bars ($1.7 \times 3.5 \times 490$ cm) with the light being focused onto 11 000 conventional PMT's located about 120 cm from the end of the bars by the "pinhole" of the bar end. DIRC performance can be understood

using the formula for (N_σ) discussed above. Typically, $N_{\rm p.e.}$ is rather large (between 15 and 60) and the Cherenkov polar angle is measured to about 2.5 mrad. The momentum range with good π/K separation extends up to about 4 GeV/c, matching the B decay momentum spectrum observed in BaBar.

28.5. Cherenkov tracking calorimeters

Written August 2003 by D. Casper (UC Irvine).

In addition to the specialized applications described in the previous section, Cherenkov radiation is also exploited in large, ring-imaging detectors with masses measured in kilotons or greater. Such devices are not subdetector components, but complete experiments with triggering, tracking, vertexing, particle identification and calorimetric capabilities, where the large mass of the transparent dielectric medium serves as an active target for neutrino interactions (or their secondary muons) and rare processes like nucleon decay.

For volumes of this scale, absorption and scattering of Cherenkov light are non-negligible, and a wavelength-dependent factor $e^{-d/L(\lambda)}$ (where d is the distance from emission to the sensor and $L(\lambda)$ is the attenuation length of the medium) must be included in the integral of Eq. (28.4) for the photoelectron yield. The choice of medium is therefore constrained by the refractive index and transparency in the region of photodetector sensitivity; highly-purified water is an inexpensive and effective choice; sea-water, mineral oil, polar ice, and D_2O are also used. Photo-multiplier tubes (PMTs) on either a volume or surface lattice measure the time of arrival and intensity of Cherenkov radiation. Hemispherical PMTs are favored for the widest angular acceptance, and sometimes mounted with reflectors or wavelength-shifting plates to increase the effective photosensitive area. Gains and calibration curves are measured with pulsed laser signals transmitted to each PMT individually via optical fiber or applied to the detector as a whole through one or more diffusing balls.

Volume instrumentation [53] is only cost-effective at low densities, with a spacing comparable to the attenuation (absorption and scattering) length of Cherenkov light in the medium (15–40 m for Antarctic ice and ∼45 m in the deep ocean). PMTs are deployed in vertical strings as modular units which include pressure housings, front-end electronics and calibration hardware. The effective photocathode coverage of such arrays is less than 1% but still adequate (using timing information and the Cherenkov angular constraint) to reconstruct the direction of TeV muons to 1° or better. The size of such "neutrino telescopes" is limited only by cost once the technical challenges of deployment, power, signal extraction and calibration in an inaccessible and inhospitable environment are addressed; arrays up to $(1\ \text{km})^3$ in size are under study or development.

Surface instrumentation [54] allows the target volume to be viewed with higher photocathode density by a number of PMTs which scales like $(\text{volume})^{2/3}$. To improve hermeticity and shielding, and to ensure that an outward-going particle's Cherenkov cone illuminates sufficient PMTs for reconstruction, a software-defined fiducial volume begins some distance (∼ 2 m) inside the photosensor surface. Events originating within the fiducial volume are classified as *fully-contained* if no particles exit the inner detector, or

partially-contained otherwise. An outer (veto) detector, optically separated from the inner volume and instrumented at reduced density, greatly assists in making this determination and also simplifies the selection of contained events. The maximum size of a pure surface array is limited by the attenuation length (~ 100 m has been achieved for large volumes using reverse-osmosis water purification), pressure tolerance of the PMTs (< 80 meters of water, without pressure housings) and structural integrity of the enclosing cavity, if underground. In practice, these limitations can be overcome by a segmented design involving multiple modules of the nominal maximum size; megaton-scale devices are under study.

Cherenkov detectors are excellent electromagnetic calorimeters, and the number of Cherenkov photons *produced* by an e/γ is nearly proportional to its kinetic energy. For massive particles, the number of photons produced is also related to the energy, but not linearly. For any type of particle, the *visible energy* $E_{\rm vis}$ is defined as the energy of an electron which would produce the same number of Cherenkov photons. The number of photoelectrons *collected* depends on a detector-specific scale factor, with event-by-event corrections for geometry and attenuation. For typical PMTs, in water $N_{\rm p.e.} \approx 15\,\xi\,E_{\rm vis}({\rm MeV})$, where ξ is the effective fractional photosensor coverage; for other materials, the photoelectron yield scales with the ratio of $\sin^2\theta_c$ over density. At solar neutrino energies, the visible energy resolution ($\sim 30\%/\sqrt{\xi\,E_{\rm vis}({\rm MeV})}$) is about 20% worse than photoelectron counting statistics would imply. For higher energies, multi-photoelectron hits are likely and the charge collected by each PMT (rather the number of PMTs firing) must be used; this degrades the energy resolution to approximately $2\%/\sqrt{\xi\,E_{\rm vis}({\rm GeV})}$. In addition, the absolute energy scale must be determined with sources of known energy. Using an electron LINAC and/or nuclear sources, 0.5–1.5% has been achieved at solar neutrino energies; for higher energies, cosmic-ray muons, Michel electrons and π^0 from neutrino interactions allow $\sim 3\%$ absolute energy calibration.

A trigger can be formed by the coincidence of PMTs within a window comparable to the detector's light crossing time; the coincidence level thus corresponds to a visible energy threshold. Physics analysis is usually not limited by the hardware trigger, but rather the ability to reconstruct events. The interaction vertex can be estimated using timing and refined by applying the Cherenkov angle constraint to identified ring edges. Multi-ring events are more strongly constrained, and their vertex resolution is 33–50% better than single rings. Vertex resolution depends on the photosensor density and detector size, with smaller detectors performing somewhat better than large ones (~ 25 cm is typical for existing devices). Angular resolution is limited by multiple scattering at solar neutrino energies (25–30°) and improves to a few degrees around $E_{vis} = 1$ GeV.

A non-showering $(\mu, \pi^{\pm}, p)$ track produces a sharp ring with small contributions from delta rays and other radiated secondaries, while the more diffuse pattern of a showering (e, γ) particle is actually the superposition of many individual rings from charged shower products. Using maximum likelihood techniques and the Cherenkov angle constraint, these two topologies can be distinguished with an efficiency which depends on the photosensor density and detector size [55]. This particle identification capability has been confirmed by using cosmic-rays and Michel electrons, as well as charged-

particle [56] and neutrino [57] beams. Large detectors perform somewhat better than smaller ones with identical photocathode coverage; a misidentification probability of $\sim 0.4\%/\xi$ in the sub-GeV range is consistent with the performance of several experiments for $4\% < \xi < 40\%$. Detection of a delayed coincidence from muon decay offers another, more indirect, means of particle identification; with suitable electronics, efficiency approaches 100% for μ^+ decays but is limited by nuclear absorption (22% probability in water) for μ^-.

Reconstruction of multiple Cherenkov rings presents a challenging pattern recognition problem, which must be attacked by some combination of heuristics, maximum likelihood fitting, Hough transforms and/or neural networks. The problem itself is somewhat ill-defined since, as noted, even a single showering primary produces many closely-overlapping rings. For $\pi^0 \to \gamma\gamma$ two-ring identification, performance falls off rapidly with increasing π^0 momentum, and selection criteria must be optimized with respect to the analysis-dependent cost-function for $e \leftrightarrow \pi^0$ mis-identification. Two representative cases for $\xi = 39\%$ will be illustrated. In an atmospheric neutrino experiment, where π^0 are relatively rare compared to $e^{\pm}$, one can isolate a $> 90\%$ pure 500 MeV/c π^0 sample with an efficiency of $\sim 40\%$. In a ν_e appearance experiment at $E_\nu \leq 1\,\mathrm{GeV}$, where $e^{\pm}$ are rare compared to π^0, a 99% pure 500 MeV/c electron sample can be identified with an efficiency of $\sim 70\%$. For constant ξ, a larger detector (with, perforce, a greater number of pixels to sample the light distribution) performs somewhat better at multi-ring separation than a smaller one. For a more detailed discussion of event reconstruction techniques, see Ref. 46.

Table 28.5: Properties of Cherenkov tracking calorimeters. LSND was a hybrid scintillation/Cherenkov detector; the estimated ratio of isotropic to Cherenkov photoelectrons was about 5:1. MiniBooNE's light yield also includes a small scintillation component.

Detector	Fiducial mass (kton)	PMTs (diameter, cm)	ξ	p.e./ MeV	Dates
IMB-1 [58]	3.3 H_2O	2048 (12.5)	1%	0.25	1982–85
IMB-3 [59]	3.3 H_2O	2048 (20 +plate)	4.5%	1.1	1987–90
KAM I [60,61]	0.88/0.78 H_2O	1000/948 (50)	20%	3.4	1983–85
KAM II [62]	1.04 H_2O	948 (50)	20%	3.4	1986–90
LSND [63]	0.084 oil+scint.	1220 (20)	25%	33	1993–98
SK-1 [64]	22.5 H_2O	11146 (50)	39%	6	1997–2001
SK-2	22.5 H_2O	5182 (50)	18%	3	2002–
K2K [65]	0.025 H_2O	680 (50)	39%	6	1999–
SNO [66]	1.0 D_2O	9456 (20+cone)	55%	9	1999–
MiniBooNE	0.445 oil	1280 (20)	10%	3–4	2002–

28.6. Transition radiation detectors (TRD's)

Revised September 2003 by D. Froidevaux (CERN).

It is clear from the discussion in the section on "Passages of Particles Through Matter" (Sec. 27 of this *Review*) that transition radiation (TR) only becomes useful for particle detectors when the signal can be observed as x rays emitted along the particle direction for Lorentz factors γ larger than 1000. In practice, TRD's are therefore used to provide electron/pion separation for 0.5 GeV/$c \lesssim p \lesssim$ 100 GeV/c. The charged-particle momenta have usually been measured elsewhere in the detector in the past [67].

Since soft x rays, in the useful energy range between 2 and 20 keV, are radiated with about 1% probability per boundary crossing, practical detectors use radiators with several hundred interfaces, *e.g.* foils or fibers of low-Z materials such as polypropylene (or, more rarely, lithium) in a gas. Absorption inside the radiator itself and in the inactive material of the x-ray detector is important and limits the usefulness of the softer x rays, but interference effects are even larger, and saturate the x-ray yield for electron energies above a few GeV [68,69].

A classical detector is composed of several similar modules, each consisting of a radiator and an x-ray detector, which is usually a wire chamber operated with a xenon-rich mixture, in order to efficiently absorb the x rays. The most prominent and recent examples of such detectors for large-scale experiments are the TRD detectors of NOMAD [70], ALICE [71], and PHENIX. Since transition-radiation photons are mostly emitted at very small angles with respect to the charged-particle direction, the x-ray detector most often detects the sum of the ionization loss (dE/dx) of the charged particle in the gas and energy deposition of the x rays. The discrimination between electrons and pions can be based on the charges measured in each detection module, on the number of energy clusters observed above an optimal threshold (usually in the 5 to 7 keV region), or on more sophisticated methods analyzing the pulse shape as a function of time. Once properly calibrated and optimized, most of these methods yield very similar results.

Development work over the past years for accelerator (ATLAS [72]) and space (AMS [73], PAMELA [74] applications has aimed at increasing the intrinsic quality of the TRD-performance by increasing the probability per detection module of observing a signal from TR-photons produced by electrons. This has been achieved experimentally by distributing small-diameter straw-tube detectors uniformly throughout the radiator material. This method has thereby also cured one of the major drawbacks of more classical TRD's, that is, their need to rely on another detector to measure the charged-particle trajectory. For example, in the ATLAS Transition Radiator Tracker [75] charged particles cross about 35 straw tubes embedded in the radiator material. Dedicated R&D work and detailed simulations have shown that the combination of charged-track measurement and particle identification in the same detector will provide a very powerful tool even at the highest LHC luminosity [76].

The major factor in the performance of any TRD is its overall length. This is illustrated in Fig. 28.3, which shows, for a variety of detectors, the measured (or predicted) pion efficiency at a fixed electron efficiency of 90% as a function of the overall detector length. The experimental data cover too wide a range of particle energies (from a few GeV to

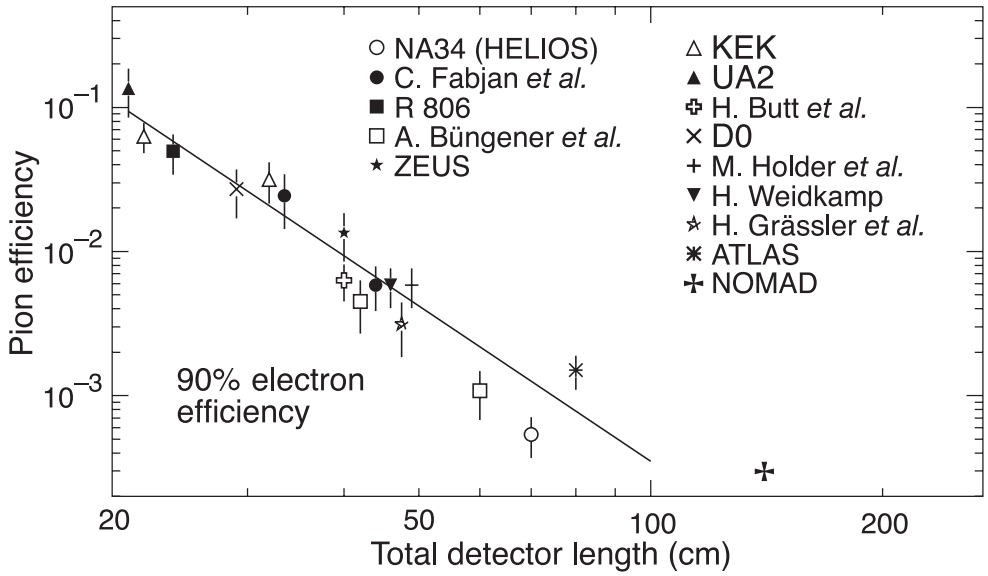


Figure 28.3: Pion efficiency measured (or predicted) for different TRDs as a function of the detector length for a fixed electron efficiency of 90%. The experimental data are directly taken or extrapolated from references [77–89,70](NA34 to NOMAD).

40 GeV) to allow for a quantitative fit to a universal curve. Fig. 28.3 shows that an order of magnitude in rejection power against pions is gained each time the detector length is increased by ~ 20 cm.

28.7. Wire chambers

Written October 1999 by A. Cattai and G. Rolandi (CERN).

A wire chamber relies on the detection of a large fraction of the charge created in a volume filled with an appropriate gas mixture. A charged particle traversing a gas layer of thickness Δ produces electron-ion pairs along its path (see Sec. 27.2). The yield $(1/\lambda)$ of ionization encounters for a minimum ionization particle (m.i.p.) (see Fig. 27.1) is given in Table 28.6.

The probability to have at least one ionization encounter is $1 - \exp(-\Delta/\lambda)$ and the thickness of the gas layer for 99% efficiency is $t_{99} = 4.6\lambda$. Depending on the gas, some 65–80% of the encounters result in the production of only one electron; the probability that a cluster has more than five electrons is smaller than 10%. However the tail of the distribution is very long and the yield of ionization electrons is 3–4 times that of the ionization encounters. The secondary ionization happens either in collisions of (primary) ionization electrons with atoms or through intermediate excited states. The process is non-linear and gas mixtures may have larger yields than each of their components. See also the discussion in Sec. 27.7.

Under the influence of electric and magnetic fields the ionization electrons drift inside

	Encounters/cm	t_{99}(mm)	Free electrons/cm
He	5	9.2	16
Ne	12	3.8	42
Ar	25	1.8	103
Xe	46	1.0	340
CH_4	27	1.7	62
CO_2	35	1.3	107
C_2H_6	43	1.1	113

Table 28.6: For various gases at STP: (a) yield of ionization encounters ($1/\lambda$) for m.i.p. [90], (b) t_{99}: thickness of the gas layer for 99% efficiency, and (c) the average number of free electrons produced by a m.i.p. (calculated using data from Ref. 91).

the gas with velocity $\boldsymbol{u}$ given by:

$$\boldsymbol{u} = \mu|\boldsymbol{E}|\frac{1}{1+\omega^2\tau^2}\left(\widehat{\boldsymbol{E}} + \omega\tau(\widehat{\boldsymbol{E}}\times\widehat{\boldsymbol{B}}) + \omega^2\tau^2(\widehat{\boldsymbol{E}}\cdot\widehat{\boldsymbol{B}})\widehat{\boldsymbol{B}}\right) \tag{28.12}$$

where $\widehat{\boldsymbol{E}}$ and $\widehat{\boldsymbol{B}}$ are unit vectors in the directions of the electric and magnetic fields respectively, μ is the electron mobility in the gas, ω is the cyclotron frequency eB/mc, and $\tau = \mu m/e$ is the mean time between collisions of the drifting electrons. The magnitude of the drift velocity depends on many parameters; typical values are in the range 1–8 cm/μs.

In a quite common geometry, the drift electric field is perpendicular to the magnetic field. In this case the electrons drift at an angle ψ with respect to the electric field direction such that $\tan\psi = \omega\tau$.

The ionization electrons are eventually collected by a thin (typically 10 μm radius) anode wire where a strong electric field—increasing as $1/r$—accelerates the electrons enough to produce secondary ionization and hence an avalanche. A quenching gas (organic molecules with large photo-absorption cross-section) absorbs the majority of the photons produced during the avalanche development, keeping the avalanche region localized. The gain achievable with a wire counter depends exponentially on the charge density on the wire, on the gas density ρ and—through it—on pressure and temperature: $dG/G \approx -Kd\rho/\rho$, where the coefficient K ranges between 5 and 8 in practical cases. Gains larger than 10^4 can be obtained in proportional mode.

The electrons produced in the avalanche are collected by the wire in a few nanoseconds. The positive ions move away from the wire and generate a signal that can be detected with an amplifier. Depending on whether the wire is treated as a current source or a voltage source, the signal is described respectively by:

$$I(t) = q\frac{d}{dt}F(t)\ ; \qquad \Delta V(t) = \frac{q}{C}F(t)\ , \tag{28.13}$$

where q is the positive charge in the avalanche, C is the capacitance between the anode wire and the cathodes and $F(t) = \ln(1+t/t_0)/\ln(1+t_{\rm max}/t_0)$. The constant t_0 is of the

order of one or few nanoseconds; the constant $t_{\rm max}$ (several microseconds) describes the time that it takes ions to reach the cathodes.

A sketch of the first multi-wire proportional chamber (MWPC) [92] is shown in Fig. 28.4. It consists of a plane of parallel sense wires with spacing s and length L inserted in a gap of thickness Δ. The potential distributions and fields in a proportional or drift chamber can usually be calculated with good accuracy from the exact formula for the potential around an array of parallel line charges q (coul/m) along z and located at $y = 0$, $x = 0$, $\pm s$, $\pm 2s$, ...,

$$V(x,y) = -\frac{q}{4\pi\,\epsilon_0} \ln\left\{4\left[\sin^2\left(\frac{\pi x}{s}\right) + \sinh^2\left(\frac{\pi y}{s}\right)\right]\right\} \ . \tag{28.14}$$

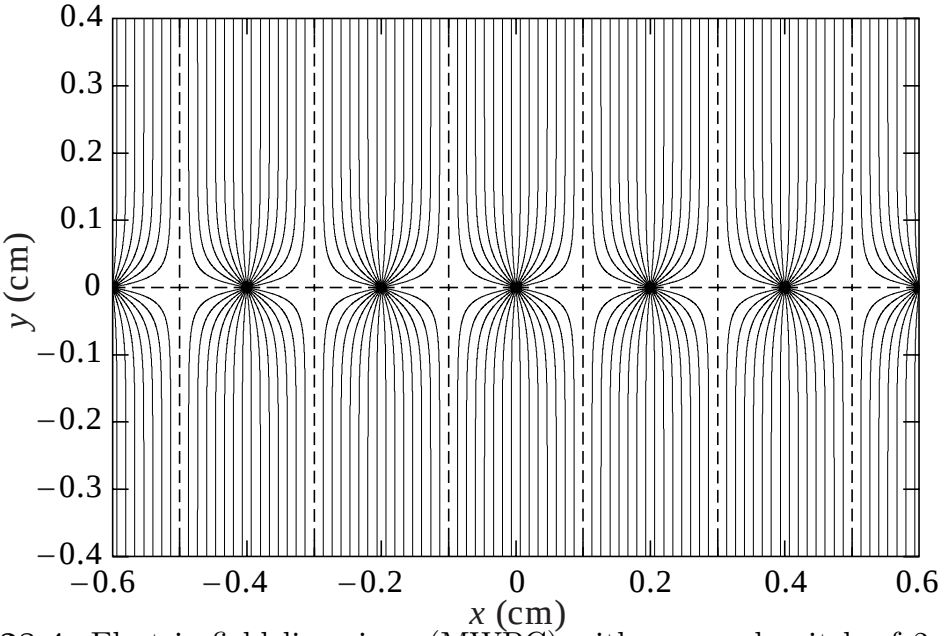


Figure 28.4: Electric field lines in a (MWPC) with an anode pitch of 2 mm as calculated with GARFIELD program [93].

With digital readout, the resolution in the direction perpendicular to the wire is $s/\sqrt{12}$, where s is typically 1–2 mm. Similar resolution can be achieved with a smaller channel density by measuring the difference in time between the arrival of electrons at the wire and the traversal of the particle, albeit with a longer response time. In the case of drift chambers, the spatial resolution is limited by the diffusion of ionization electrons during the drift and by the fluctuations of the ionization process. Depending on the gas mixture, the width of the diffusing cloud after 1 cm of drift is typically between 50 and 300 μm; small diffusion implies low drift velocity. With drift lengths up to 5 cm (1 μs), resolutions in the range 100–200 μm have been achieved in chambers with surface areas of several square meters [94]. The central detectors in many collider experiments are drift chambers with the wires parallel to the beam direction. Small volume chambers (0.1 m^3) have been used for vertex measurement achieving resolutions of 50 μm using high pressure

(2–4 bar) and low diffusion gas mixtures [95]. Large volume chambers (5–40 m^3) with several thousand wires of length of 1–2 meters are operated with resolution between 100 and 200 μm [96].

The spatial resolution cannot be improved by arbitrarily reducing the spacing of the wires. In addition to the practical difficulties of precisely stringing wires at a pitch below 1 mm, there is a fundamental limitation: the electrostatic force between the wires is balanced by the mechanical tension, which cannot exceed a critical value. This gives the following approximate stability condition:

$$\frac{s}{L} \geq 1.5 \times 10^{-3} V(\text{kV}) \sqrt{\frac{20\ \text{g}}{T}}\ , \tag{28.15}$$

where V is the voltage of the sense wire and T is the tension of the wire in grams-weight equivalent.

A review of the principle of particle detection with drift chambers can be found in [97]. A compilation of the mobilities, diffusion coefficients and drift deflection angles as a function of $\boldsymbol{E}$ and $\boldsymbol{B}$ for several gas mixtures used in proportional chambers can be found in [98]. A review of micro-strip gas chambers (MSGC) can be found in [99].

28.8. Micro-pattern Gas Detectors

Written October 2005 by M.T Ronan (LBNL).

New micro-pattern gas detectors (MPGD's) are replacing conventional wire chambers in many applications. In these devices electron gas amplification is obtained in very high fields generated by modest voltages (300–400 V) across 50–100 μm structures suitable for large-area applications. Typically gains of 10^3 to 10^4 are achieved with many gases under standard conditions. Two examples of MPGD's are gas electron multipliers (GEM's) [108] and Micromegas [109].

A GEM detector is typically fabricated by etching 50 μm holes at a 140 μm pitch in a thin (50 μm) metal-clad sheet of insulating material. A potential difference is maintained between opposite sides of the sheet, so that electron avalanches occur in the holes. Electrons more or less follow the field lines as shown in Fig. 28.5.

A Micromegas structure is made by separating a thin (5 μm) metallic mesh with 50 μm pitch holes from a conducting or resistive anode plane using small 50–100 μm-tall spacers. Electron and positive ion drift lines are shown in Fig. 28.6.

Fast electron and ion signals with risetimes of a few nsec and full widths of 20–100 nsec are obtained from direct charge collection with no loss due to ballistic deficit effects or the $1/t$ tails present in wire chambers. Positive ion feedback is determined by the ratio of drift and amplification fields, and by transfer fields in the case of multilayer GEM configurations. Typical drift or conversion fields of 100–1000 V/cm and amplification fields of 40–70 kV/cm leads to intrinsic feedback suppression of a few to several parts per mil.

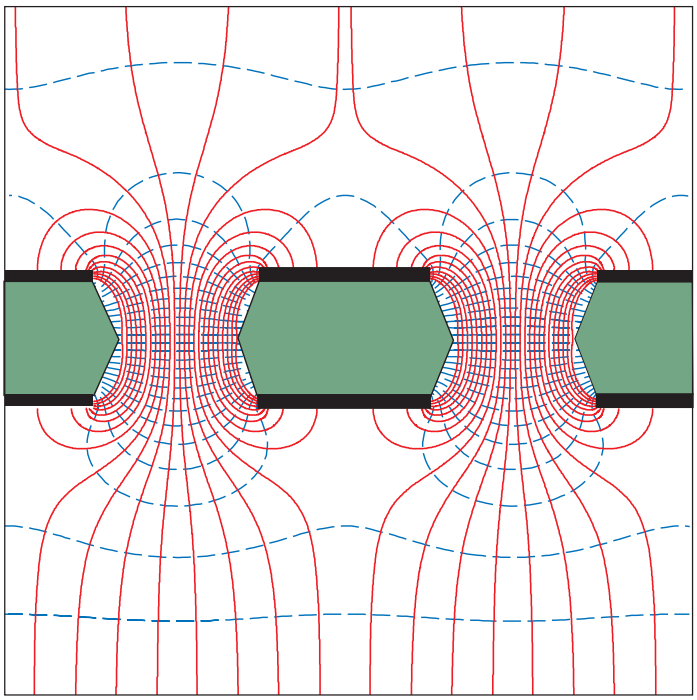


Figure 28.5: Detail of electric field lines (solid) and equipotentials (dashed) in the region of GEM holes. Electron transparencies are typically 100%. Most positive ions produced in the high field region within a hole drift back to the GEM's top side. Figure courtesy of F. Sauli.

28.9. Resistive-plate chambers

Written November 2005 by H.R. Band (U. Wisconsin) and J. Zhang (Beijing).

The resistive-plate chamber (RPC) was developed by Santonico and Cardarelli in the early 1980's [102] as a low-cost alternative to large scintillator planes.* Most commonly, an RPC is constructed from two parallel high-resistivity (10^9–10^{13} Ω-cm) glass or phenolic (Bakelite)/melamine laminate plates with a few-mm gap between them which is filled with atmospheric-pressure gas. The gas is chosen to absorb UV photons in order to limit transverse growth of discharges. The backs of the plates are coated with a lower-resistivity paint or ink ($\sim 10^5\,\Omega/\Box$), and a high potential (7–12 keV) is maintained between them. The passage of a charged particle initiates an electric discharge, whose size and duration are limited since the current reduces the local potential to below that needed to maintain the discharge. The sensitivity of the detector outside of this region is unaffected. The signal readout is via capacitive coupling to metallic strips on both sides of the detector which are separated from the high voltage coatings by thin insulating sheets. The x and y position of the discharge can be measured if the strips on opposite sides of the gap are orthogonal. When operated in streamer mode, the induced signals on the strips can be quite large (~300 mV), making sensitive electronics unnecessary. An example of an RPC structure is shown in Fig. 28.7.

* It was based on earlier work on a spark counter with one high-resistivity plate [103].

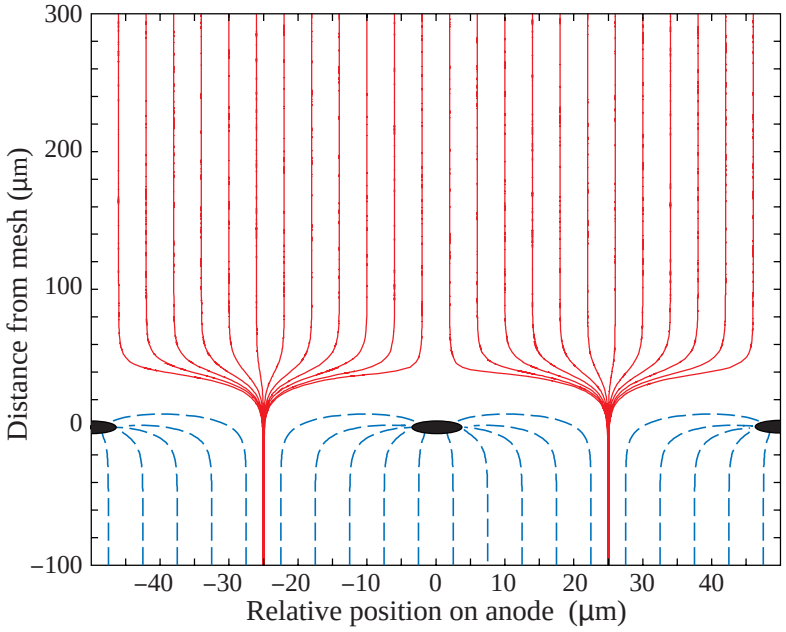


Figure 28.6: GARFIELD simulation of the electron (solid) and positive ion (dashed) drift lines in a Micromegas avalanche region. Drifting electrons are funneled through the mesh and spread across the anode surface by diffusion (not included). Most positive ions produced in the gap near the anode surface drift back to the Micromegas mesh.

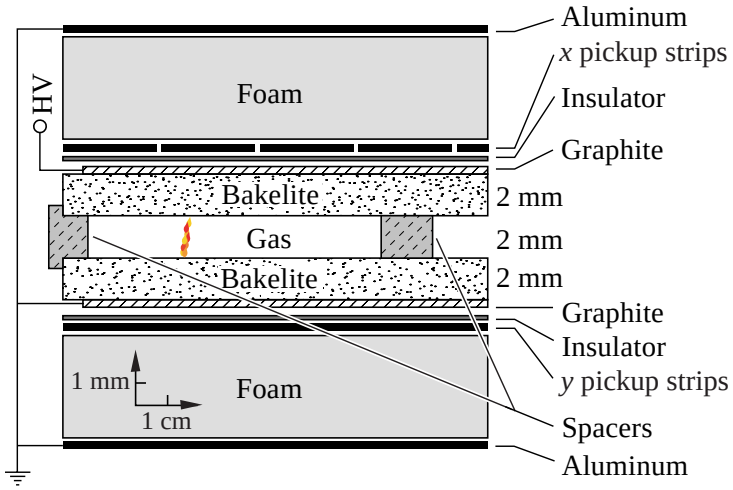


Figure 28.7: Schematic cross section of a typical RPC, in this case the single-gap streamer-mode BaBar RPC.

RPC's have inherent rate limitations since the time needed to re-establish the field after a discharge is proportional to the chamber capacitance and plate resistance. The average charge per streamer is 100–1000 pC. Typically, the efficiency of streamer-mode glass RPC's begins to fall above $\sim$0.4 Hz/cm^2. Because of Bakelite's lower bulk resistivity, Bakelite RPC's can be efficient at 10–100 Hz/cm^2. The need for higher rate capability led to the development of avalanche-mode RPC's, in which the gas and high voltage have

been tuned to limit the growth of the electric discharge, preventing streamer formation. Typical avalanche-mode RPC's have a signal charge of about 10 pC and can be efficient at 1 kHz/cm^2. The avalanche discharge produces a much smaller induced signal on the pickup strips ($\sim$1 mV) than streamers, and thus requires a more sophisticated and careful electronic design.

Many variations of the initial RPC design have been built for operation in either mode. Efficiencies of $\gtrsim$ 92% for single gaps can be improved by the use of two or more gas gaps with shared pickup strips. Non-flammable and more environmentally friendly gas mixtures have been developed. In streamer mode, various mixtures of argon with isobutane and tetrafluoroethane have been used. For avalanche mode operation, a gas mixture of tetrafluoroethane ($C_2H_2F_4$) with 2–5% isobutane and 0.4–10% sulphur hexafluoride (SF_6) is typical. An example of large-scale RPC use is provided by the muon system being built for the ATLAS detector, where three layers of pairs of RPC's are used to trigger the drift tube arrays between the pairs. The total area is about 10,000 m^2. These RPC's provide a spatial resolution of 1 cm and a time resolution of 1 ns at an efficiency $\geq$ 99%. Much better timing resolution (50 ps) has been obtained using thin multiple-gap RPC's. Fonte provides useful review [104] of RPC designs.

Operational experience with RPC's has been mixed. Several experiments (*e.g.*, L3 and HARP) have reported reliable performance. However, the severe problems experienced with the BaBar RPC's have raised concerns about the long-term reliability of Bakelite RPC's.

Glass RPC's have had fewer problems, as seen by the history of the BELLE chambers. A rapid growth in the noise rate and leakage current in some of the BELLE glass RPC's was observed during commissioning. It was found that water vapor in the input gas was reacting with fluorine (produced by the disassociation of the tetrafluoroethane in the streamers) to produce hydrofluoric acid. The acid etched the glass surfaces, leading to increased noise rates and lower efficiencies. The use of copper gas piping to insure the dryness of the input gas stopped the problem. The BELLE RPC's have now operated reliably for more than 5 years. Even so, high noise rates in the outermost endcap layers have significantly reduced efficiencies there.

Several different failure modes diagnosed in the first-generation BaBar Bakelite RPC's caused the average efficiency of the barrel RPC's to fall from $\gtrsim$90% to 35% in five years. The linseed oil which is used in Bakelite RPC's to coat the inner surface [105] had not been completely cured. Under warm conditions (32° C) and high voltage, oil collected on the spacers between the gaps or formed oil-drop bridges between the gaps. This led to large leakage currents (50–100 μA in some chambers) which persisted even when the temperature was regulated at 20° C. In addition, the graphite layer used to distribute the high voltage over the Bakelite became highly resistive (100 k$\Omega/\Box \rightarrow$ 10 M$\Omega/\Box$), resulting in lowered efficiency in some regions and the complete death of whole chambers.

The BaBar problems and the proposed use of Bakelite RPC's in the LHC detectors prompted detailed studies of RPC aging and have led to improved construction techniques and a better understanding of RPC operational limits. The graphite layer has been improved and should be stable with integrated currents of $\lesssim$ 600 mC/cm^2. Molded gas inlets and improved cleanliness during construction have reduced the noise rate of new

chambers. Unlike glass RPC's, Bakelite RPC's have been found to require humid input gases to prevent drying of the Bakelite (increasing the bulk resistivity) which would decrease the rate capability. Second-generation BaBar RPC's incorporating many of the above improvements have performed reliably for over two years [106].

With many of these problems solved, new-generation RPC's are now being or soon will be used in about a dozen cosmic-ray and HEP detectors. Their comparatively low cost, ease of construction, good time resolution, high efficiency, and moderate spatial resolution make them attractive in many situations, particularly those requiring fast timing and/or large-area coverage.

28.10. Time-projection chambers

Written November 1997 by M.T Ronan (LBNL); revised August 2005.

Detectors with long drift distances perpendicular to a readout pad plane provide three-dimensional information, with one being the time projection. A (typically strong) magnetic field parallel to the drift direction suppresses transverse diffusion ($\sigma = \sqrt{2Dt}$) by a factor

$$D(B)/D(0) = \frac{1}{1+\omega^2\tau^2} \ , \tag{28.16}$$

where D is the diffusion coefficient, $\omega = eB/mc$ is the cyclotron frequency, and τ is the mean time between collisions. Multiple measurements of energy deposit along the particle trajectory combined with the measurement of momentum in the magnetic field allows excellent particle identification [107], as can be seen in Fig. 28.8.

A typical gas-filled TPC consists of a long uniform drift region (1–2 m) generated by a central high-voltage membrane and precision concentric cylindrical field cages within a uniform, parallel magnetic field [97]. A multiwire proportional plane, Gas Electron Multiplier (GEM) [108] or Micromegas [109] micropattern device provides primary electron gas amplification. Details of construction and electron trajectories for a multiwire proportional anode plane are shown in Fig. 28.9. Signal shaping and processing using analog storage devices or FADC's allows excellent pattern recognition, track reconstruction, and particle identification within the same detector.

Typical values:

Gas: Ar + (10–20%) CH_4	Pressure(P) = 1–8.5 atm.
E/P = 100–200 V/cm/atm	B = 1–2 Tesla
$v_{\rm drift}$ = 5–7 cm/μs	$\omega\tau$ = 1–8
$\sigma_{x \text{ or } y}$ = 100–200 μm	σ_z = 0.2–1 mm
$\sigma_{E\,dep}$ = 2.5–5.5 %	

Truncated mean energy-deposit resolution depends on the number and size of samples, and gas pressure:

$$\sigma_{E\,dep} \propto N^{-0.43} \times (P\ell)^{-0.32} \ . \tag{28.17}$$

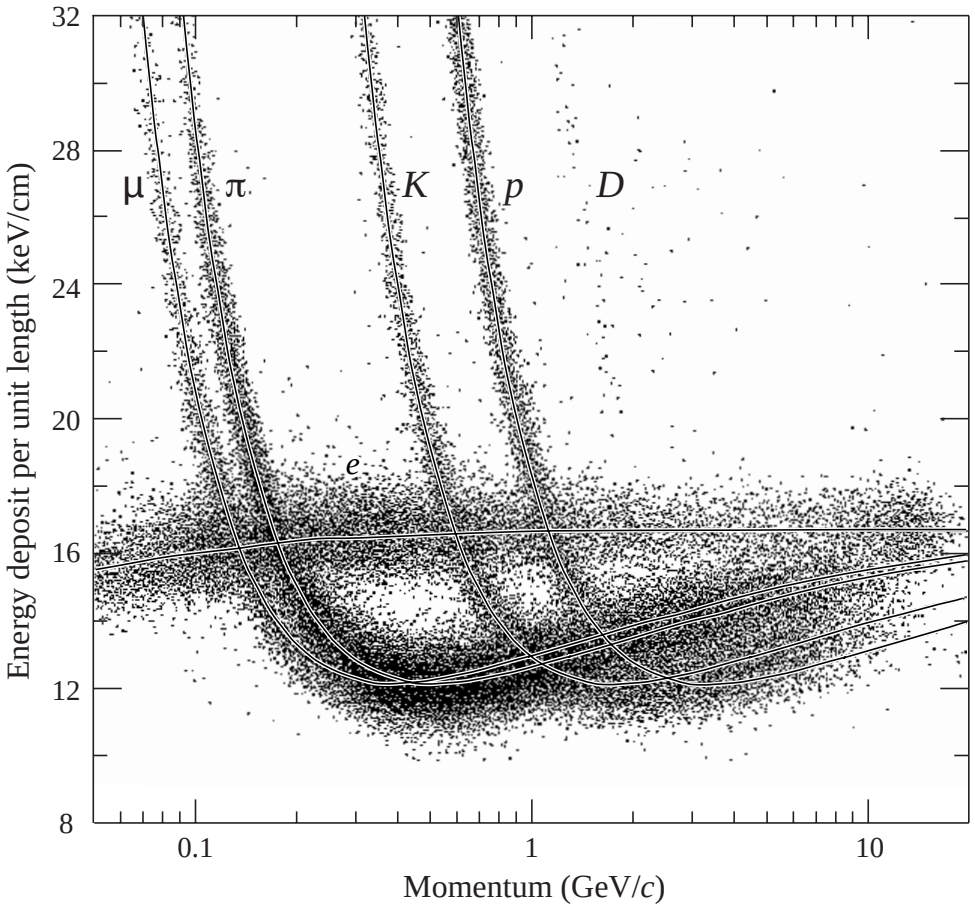


Figure 28.8: PEP4/9-TPC energy-deposit measurements (185 samples @8.5 atm Ar-CH_4 80–20%) in multihadron events. The electrons reach a Fermi plateau value of 1.4 times the most probably energy deposit at minimum ionization. Muons from pion decays are separated from pions at low momentum; π/K are separated over all momenta except in the cross-over region. (Low-momentum protons and deuterons originate from hadron-nucleus collisions in inner materials such as the beam pipe.)

Here N is the number of samples, ℓ is the sample size, and P is the pressure. Typical energy-deposit distributions are shown in Fig. 28.8. Good three-dimensional two-track resolutions of about 1–1.5 cm are routinely achieved.

$E \times B$ distortions arise from nonparallel E and B fields (see Eq. (28.12)), and from the curved drift of electrons to the anode wires in the amplification region. Position measurement errors include contributions from the anode-cathode geometry, the track crossing angle (α), $E \times B$ distortions, and from the drift diffusion of electrons

$$\sigma^2_{x \text{ or } y} = \sigma_0^2 + \sigma_D^2(1 + \tan^2\alpha)L/L_{\max} + \sigma_\alpha^2(\tan\alpha - \tan\psi)^2 \qquad (28.18)$$

where σ is the coordinate resolution, σ_0 includes the anode-cathode geometry contribution, ψ is the Lorentz angle, and L is the drift distance.

Space-charge distortions arise in high-rate environments, especially for low values of $\omega\tau$.

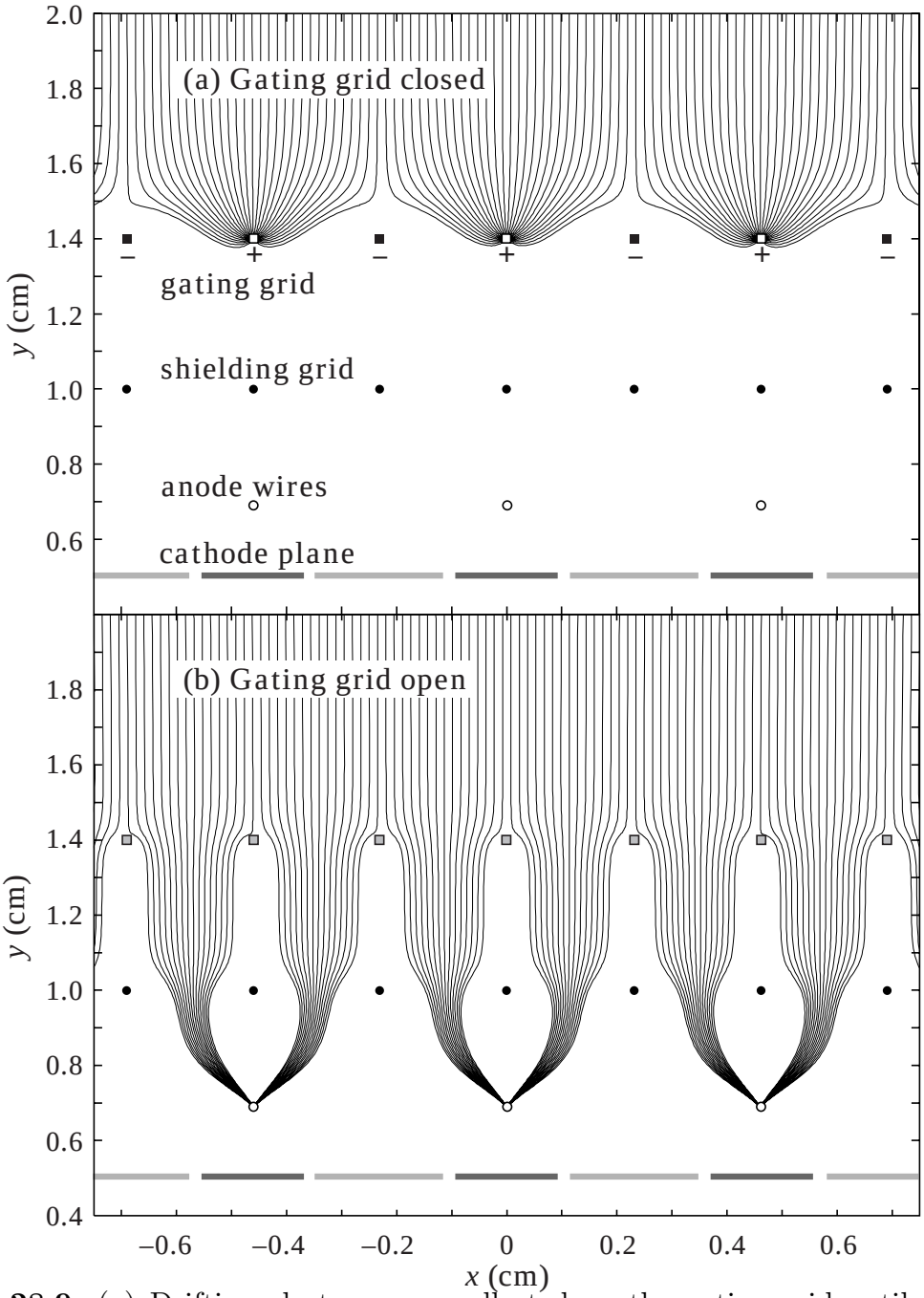


Figure 28.9: (a) Drifting electrons are collected on the gating grid until gated open by a triggering event. A shielding grid at ground potential is used to terminate the drift region. Electrons drifting through an open gating grid (b) pass through to the amplification region around the anode wires. Positive ions generated in the avalanche are detected on segmented cathode pads to provide precise measurements along the wire. The slow positive ions are blocked from entering the drift region by closing the gating grid after the electrons have drifted through.

However, they are mitigated by an effective gating grid (Fig. 28.9). Field uniformities of

$$\int (E_\perp/E)\, dz \lesssim 0.5\text{–}1 \text{ mm} , \tag{28.19}$$

over 10–40 m^3 volumes have been obtained. Laser tracks and calibration events allow mapping of any remnant drift non-uniformities.

28.11. Silicon semiconductor detectors

Updated September 2005 by H. Spieler (LBNL).

Semiconductor detectors are widely used in modern high-energy physics experiments. They are the key ingredient of high-resolution vertex and tracking detectors and are also used as photodetectors in scintillation calorimeters. The most commonly used material is silicon, but germanium, gallium-arsenide and diamond are also useful in some applications. Integrated circuit technology allows the formation of high-density micron-scale electrodes on large (10–15 cm diameter) wafers, providing excellent position resolution. Furthermore, the density of silicon and its small ionization energy result in adequate signals with active layers only 100–300 μm thick, so the signals are also fast (typically tens of ns). Semiconductor detectors depend crucially on low-noise electronics (see Sec. 28.12), so the detection sensitivity is determined by signal charge and capacitance. For a comprehensive discussion of semiconductor detectors and electronics see Ref. 110.

Silicon detectors are *p*-*n* junction diodes operated at reverse bias. This forms a sensitive region depleted of mobile charge and sets up an electric field that sweeps charge liberated by radiation to the electrodes. Detectors typically use an asymmetric structure, e.g. a highly doped p electrode and a lightly doped n region, so that the depletion region extends predominantly into the lightly doped volume. The thickness of the depleted region is

$$W = \sqrt{2\epsilon\,(V + V_{bi})/Ne} = \sqrt{2\rho\mu\epsilon(V + V_{bi})} , \tag{28.20}$$

where $V =$ external bias voltage
$V_{bi} =$ "built-in" voltage (≈ 0.5 V for resistivities typically used in detectors)
$N =$ doping concentration
$e =$ electronic charge
$\epsilon =$ dielectric constant $= 11.9\ \epsilon_0 \approx 1$ pF/cm
$\rho =$ resistivity (typically 1–10 kΩ cm)
$\mu =$ charge carrier mobility
$= 1350$ cm^2 V^{-1} s^{-1} for electrons
$= 450$ cm^2 V^{-1} s^{-1} for holes

or

$W = 0.5\ [\mu\text{m}/\sqrt{\Omega\text{-cm} \cdot \text{V}}] \times \sqrt{\rho(V + V_{bi})}$ for n-type material, and

$W = 0.3\ [\mu\text{m}/\sqrt{\Omega\text{-cm} \cdot \text{V}}] \times \sqrt{\rho(V + V_{bi})}$ for p-type material.

The conductive p and n regions together with the depleted volume form a capacitor with the capacitance per unit area

$$C = \epsilon/W \approx 1\,[\mathrm{pF/cm}]\,/W\ . \tag{28.21}$$

In strip and pixel detectors the capacitance is dominated by the fringing capacitance. For example, the strip-to-strip fringing capacitance is $\sim$ 1–1.5 pF cm^{-1} of strip length at a strip pitch of 25–50 μm.

Measurements on silicon photodiodes [111] show that for photon energies below 4 eV one electron-hole (e-h) pair is formed per incident photon. The mean energy E_i required to produce an e-h pair peaks at 4.4 eV for a photon energy around 6 eV. It assumes a constant value, 3.67 eV at room temperature, above $\sim$ 1.5 keV. It is larger than the bandgap energy because phonon excitation is required for momentum conservation. For minimum-ionizing particles, the most probable charge deposition in a 300 μm thick silicon detector is about 3.5 fC (22000 electrons). Since both electronic and lattice excitations are involved, the variance in the number of charge carriers $N = E/E_i$ produced by an absorbed energy E is reduced by the Fano factor F (about 0.1 in Si). Thus, $\sigma_N = \sqrt{FN}$ and the energy resolution $\sigma_E/E = \sqrt{FE_i/E}$. However, the measured signal fluctuations are usually dominated by electronic noise or energy loss fluctuations in the detector. Visible light can be detected with photon energies above the band gap. In optimized photodiodes quantum efficiencies $> 80\%$ for wavelengths between 400 nm and nearly 1 μm are achievable. UV-extended photodiodes have useful efficiency down to 200 nm.

Charge collection time decreases with increasing bias voltage, and can be reduced further by operating the detector with "overbias," *i.e.* a bias voltage exceeding the value required to fully deplete the device. The collection time is limited by velocity saturation at high fields (approaching 10^7 cm/s at $E > 10^4$ V/cm); at an average field of 10^4 V/cm the collection time is about 15 ps/μm for electrons and 30 ps/μm for holes. In typical fully-depleted detectors 300 μm thick, electrons are collected within about 10 ns, and holes within about 25 ns.

Position resolution is limited by transverse diffusion during charge collection (typically 5 μm for 300 μm thickness) and by knock-on electrons. Resolutions of 2–4 μm (rms) have been obtained in beam tests. In magnetic fields, the Lorentz drift deflects the electron and hole trajectories and the detector must be tilted to reduce spatial spreading (see "Hall effect" in semiconductor textbooks).

Radiation damage occurs through two basic mechanisms:

1. Bulk damage due to displacement of atoms from their lattice sites. This leads to increased leakage current, carrier trapping, and build-up of space charge that changes the required operating voltage. Displacement damage depends on the nonionizing energy loss and the energy imparted to the recoil atoms, which can initiate a chain of subsequent displacements, *i.e.*, damage clusters. Hence, it is critical to consider both particle type and energy.
2. Surface damage due to charge build-up in surface layers, which leads to increased surface leakage currents. In strip detectors the inter-strip isolation is affected. The

effects of charge build-up are strongly dependent on the device structure and on fabrication details. Since the damage is proportional to the absorbed energy (when ionization dominates), the dose can be specified in rad (or Gray) independent of particle type.

The increase in reverse bias current due to bulk damage is $\Delta I_r = \alpha\Phi$ per unit volume, where Φ is the particle fluence and α the damage coefficient ($\alpha \approx 3 \times 10^{-17}$ A/cm for minimum ionizing protons and pions after long-term annealing; $\alpha \approx 2 \times 10^{-17}$ A/cm for 1 MeV neutrons). The reverse bias current depends strongly on temperature

$$\frac{I_R(T_2)}{I_R(T_1)} = \left(\frac{T_2}{T_1}\right)^2 \exp\left[-\frac{E}{2k}\left(\frac{T_1 - T_2}{T_1 T_2}\right)\right] \tag{28.22}$$

where $E = 1.2$ eV, so rather modest cooling can reduce the current substantially ($\sim$ 6-fold current reduction in cooling from room temperature to 0°C).

The space-charge concentration in high-resistivity n-type Si changes as

$$N \approx N_0 e^{-\delta\Phi} - \beta\Phi \ , \tag{28.23}$$

where N_0 is the initial donor concentration, $\delta \approx 6 \times 10^{-14}$ cm^2 determines donor removal, and $\beta \approx 0.03$ cm^{-1} describes acceptor creation. The acceptor states trap electrons, building up a negative space charge, which in turn requires an increase in the applied voltage to sweep signal charge through the detector thickness. This has the same effect as a change in resistivity, *i.e.*, the required voltage drops initially with fluence, until the positive and negative space charge balance and very little voltage is required to collect all signal charge. At larger fluences the negative space charge dominates, and the required operating voltage increases ($V \propto N$). The safe limit on operating voltage ultimately limits the detector lifetime. Strip detectors specifically designed for high voltages have been operated at bias voltages >500V. Since the effect of radiation damage depends on the electronic activity of defects, various techniques have been applied to neutralize the damage sites. For example, additional doping with oxygen increases the allowable charged hadron fluence roughly three-fold [112]. The increase in leakage current with fluence, on the other hand, appears to be unaffected by resistivity and whether the material is n or p-type.

Strip and pixel detectors have remained functional at fluences beyond 10^{15} cm^{-2} for minimum ionizing protons. At this damage level, charge loss due to recombination and trapping also becomes significant and the high signal-to-noise ratio obtainable with low-capacitance pixel structures extends detector lifetime. The occupancy of the defect charge states is strongly temperature dependent; competing processes can increase or decrease the required operating voltage. It is critical to choose the operating temperature judiciously (−10 to 0°C in typical collider detectors) and limit warm-up periods during maintenance. For a more detailed summary see Ref. 113 and and the web-site of the ROSE collaboration at `http://RD48.web.cern.ch/rd48`.

Currently, the lifetime of detector systems is still limited by the detectors; in the electronics use of standard "deep submicron" CMOS fabrication processes with

appropriately designed circuitry has increased the radiation resistance to fluences $> 10^{15}$ cm^{-2} of minimum ionizing protons or pions. For a comprehensive discussion of radiation effects see Ref. 114.

28.12. Low-noise electronics

Revised August 2003 by H. Spieler (LBNL).

Many detectors rely critically on low-noise electronics, either to improve energy resolution or to allow a low detection threshold. A typical detector front-end is shown in Fig. 28.10.

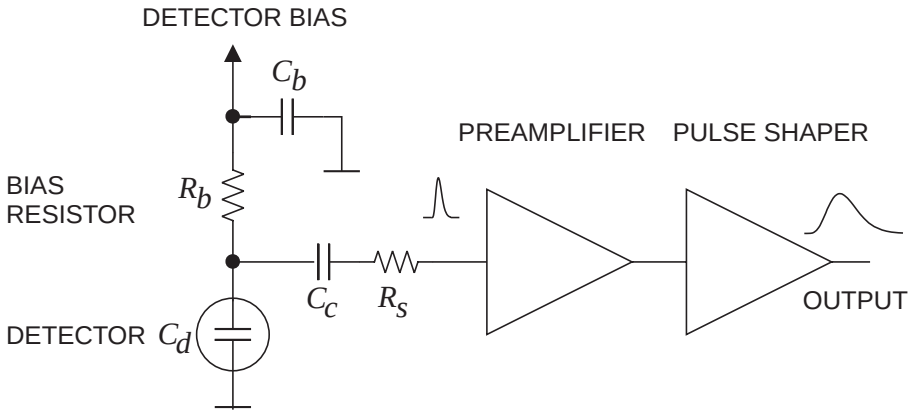


Figure 28.10: Typical detector front-end circuit.

The detector is represented by a capacitance C_d, a relevant model for most detectors. Bias voltage is applied through resistor R_b and the signal is coupled to the preamplifier through a blocking capacitor C_c. The series resistance R_s represents the sum of all resistances present in the input signal path, *e.g.* the electrode resistance, any input protection networks, and parasitic resistances in the input transistor. The preamplifier provides gain and feeds a pulse shaper, which tailors the overall frequency response to optimize signal-to-noise ratio while limiting the duration of the signal pulse to accommodate the signal pulse rate. Even if not explicitly stated, all amplifiers provide some form of pulse shaping due to their limited frequency response.

The equivalent circuit for the noise analysis (Fig. 28.11) includes both current and voltage noise sources. The leakage current of a semiconductor detector, for example, fluctuates due to electron emission statistics. This "shot noise" i_{nd} is represented by a current noise generator in parallel with the detector. Resistors exhibit noise due to thermal velocity fluctuations of the charge carriers. This noise source can be modeled either as a voltage or current generator. Generally, resistors shunting the input act as noise current sources and resistors in series with the input act as noise voltage sources (which is why some in the detector community refer to current and voltage noise as "parallel" and "series" noise). Since the bias resistor effectively shunts the input, as the capacitor C_b passes current fluctuations to ground, it acts as a current generator i_{nb} and its noise current has the same effect as the shot noise current from the detector.

Any other shunt resistances can be incorporated in the same way. Conversely, the series resistor R_s acts as a voltage generator. The electronic noise of the amplifier is described fully by a combination of voltage and current sources at its input, shown as e_{na} and i_{na}.

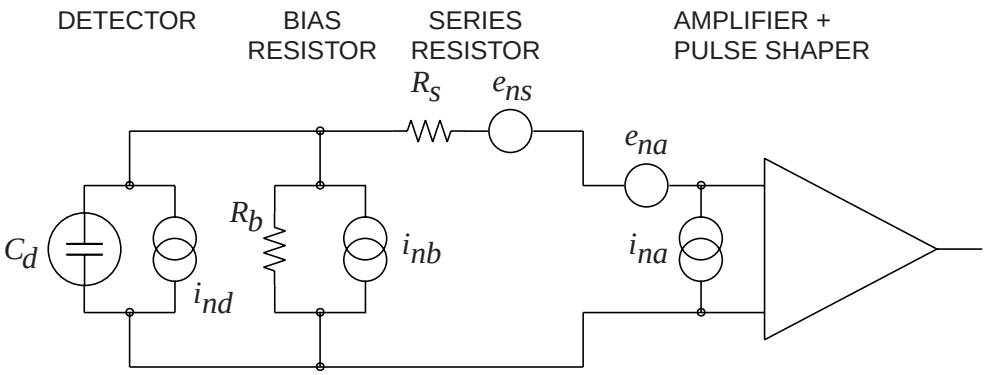


Figure 28.11: Equivalent circuit for noise analysis.

Shot noise and thermal noise have a "white" frequency distribution, *i.e.* the spectral power densities $dP_n/df \propto di_n^2/df \propto de_n^2/df$ are constant with the magnitudes

$$\begin{aligned} i_{nd}^2 &= 2eI_d \ , \\ i_{nb}^2 &= \frac{4kT}{R_b} \ , \\ e_{ns}^2 &= 4kTR_s \ , \end{aligned} \tag{28.24}$$

where e is the electronic charge, I_d the detector bias current, k the Boltzmann constant and T the temperature. Typical amplifier noise parameters e_{na} and i_{na} are of order $\mathrm{nV}/\sqrt{\mathrm{Hz}}$ and $\mathrm{pA}/\sqrt{\mathrm{Hz}}$. Trapping and detrapping processes in resistors, dielectrics and semiconductors can introduce additional fluctuations whose noise power frequently exhibits a $1/f$ spectrum. The spectral density of the $1/f$ noise voltage is

$$e_{nf}^2 = \frac{A_f}{f} \ , \tag{28.25}$$

where the noise coefficient A_f is device specific and of order 10^{-10}–$10^{-12}\,\mathrm{V}^2$.

A fraction of the noise current flows through the detector capacitance, resulting in a frequency-dependent noise voltage $i_n/(\omega C_d)$, which is added to the noise voltage in the input circuit. Since the individual noise contributions are random and uncorrelated, they add in quadrature. The total noise at the output of the pulse shaper is obtained by integrating over the full bandwidth of the system. Superimposed on repetitive detector signal pulses of constant magnitude, purely random noise produces a Gaussian signal distribution.

Since radiation detectors typically convert the deposited energy into charge, the system's noise level is conveniently expressed as an equivalent noise charge Q_n, which is

equal to the detector signal that yields a signal-to-noise ratio of one. The equivalent noise charge is commonly expressed in Coulombs, the corresponding number of electrons, or the equivalent deposited energy (eV). For a capacitive sensor

$$Q_n^2 = i_n^2 F_i T_S + e_n^2 F_v \frac{C^2}{T_S} + F_{vf} A_f C^2 \ , \tag{28.26}$$

where C is the sum of all capacitances shunting the input, F_i, F_v, and F_{vf} depend on the shape of the pulse determined by the shaper and T_s is a characteristic time, for example, the peaking time of a semi-gaussian pulse or the sampling interval in a correlated double sampler. The form factors F_i, F_v are easily calculated

$$F_i = \frac{1}{2T_S} \int_{-\infty}^{\infty} [W(t)]^2 \, dt \ , \qquad F_v = \frac{T_S}{2} \int_{-\infty}^{\infty} \left[\frac{dW(t)}{dt} \right]^2 dt \ , \tag{28.27}$$

where for time-invariant pulse-shaping $W(t)$ is simply the system's impulse response (the output signal seen on an oscilloscope) with the peak output signal normalized to unity. For more details see Refs. [115–116].

A pulse shaper formed by a single differentiator and integrator with equal time constants has $F_i = F_v = 0.9$ and $F_{vf} = 4$, independent of the shaping time constant. The overall noise bandwidth, however, depends on the time constant, *i.e.* the characteristic time T_s. The contribution from noise currents increases with shaping time, *i.e.*, pulse duration, whereas the voltage noise decreases with increasing shaping time. Noise with a $1/f$ spectrum depends only on the ratio of upper to lower cutoff frequencies (integrator to differentiator time constants), so for a given shaper topology the $1/f$ contribution to Q_n is independent of T_s. Furthermore, the contribution of noise voltage sources to Q_n increases with detector capacitance. Pulse shapers can be designed to reduce the effect of current noise, *e.g.*, mitigate radiation damage. Increasing pulse symmetry tends to decrease F_i and increase F_v (*e.g.*, to 0.45 and 1.0 for a shaper with one CR differentiator and four cascaded integrators). For the circuit shown in Fig. 28.11,

$$\begin{aligned} Q_n^2 = &\left(2eI_d + 4kT/R_b + i_{na}^2\right) F_i T_S \\ &+ \left(4kTR_s + e_{na}^2\right) F_v C_d^2 / T_S + F_{vf} A_f C_d^2 \ . \end{aligned} \tag{28.28}$$

As the characteristic time T_S is changed, the total noise goes through a minimum, where the current and voltage contributions are equal. Fig. 28.12 shows a typical example. At short shaping times the voltage noise dominates, whereas at long shaping times the current noise takes over. The noise minimum is flattened by the presence of $1/f$ noise. Increasing the detector capacitance will increase the voltage noise and shift the noise minimum to longer shaping times.

For quick estimates, one can use the following equation, which assumes an FET amplifier (negligible i_{na}) and a simple CR–RC shaper with time constants τ (equal to

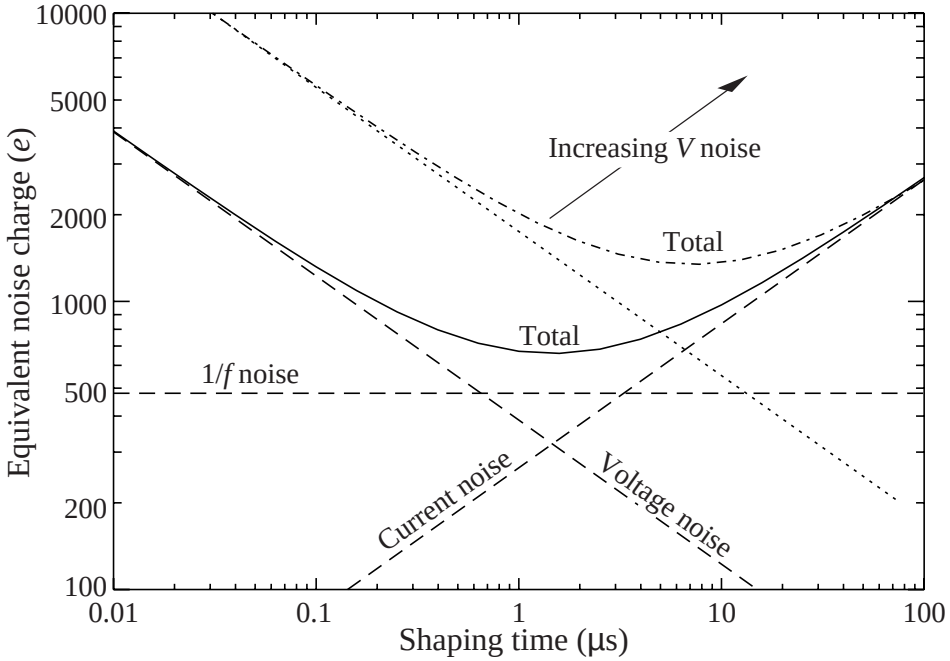


Figure 28.12: Equivalent noise charge *vs* shaping time. Changing the voltage or current noise contribution shifts the noise minimum. Increased voltage noise is shown as an example.

the peaking time):

$$(Q_n/e)^2 = 12\left[\frac{1}{\text{nA}\cdot\text{ns}}\right] I_d\tau + 6\times10^5\left[\frac{\text{k}\Omega}{\text{ns}}\right]\frac{\tau}{R_b} + 3.6\times10^4\left[\frac{\text{ns}}{(\text{pF})^2(\text{nV})^2/\text{Hz}}\right] e_n^2\frac{C^2}{\tau}\ . \tag{28.29}$$

Noise is improved by reducing the detector capacitance and leakage current, judiciously selecting all resistances in the input circuit, and choosing the optimum shaping time constant.

The noise parameters of the amplifier depend primarily on the input device. In field effect transistors, the noise current contribution is very small, so reducing the detector leakage current and increasing the bias resistance will allow long shaping times with correspondingly lower noise. In bipolar transistors, the base current sets a lower bound on the noise current, so these devices are best at short shaping times. In special cases where the noise of a transistor scales with geometry, *i.e.*, decreasing noise voltage with increasing input capacitance, the lowest noise is obtained when the input capacitance of the transistor is equal to the detector capacitance, albeit at the expense of power dissipation. Capacitive matching is useful with field-effect transistors, but not bipolar transistors. In bipolar transistors, the minimum obtainable noise is independent of shaping time, but

only at the optimum collector current I_C, which does depend on shaping time.

$$Q^2_{n,\min} = 4kT\frac{C}{\sqrt{\beta_{DC}}}\sqrt{F_i F_v} \;\text{ at }\; I_c = \frac{kT}{e}C\sqrt{\beta_{DC}}\sqrt{\frac{F_v}{F_i}}\frac{1}{T_S}\ , \tag{28.30}$$

where β_{DC} is the DC current gain. For a CR–RC shaper and $\beta_{DC} = 100$,

$$Q_{n,\min}/e \approx 250\sqrt{C/\text{pF}}\ . \tag{28.31}$$

Practical noise levels range from $\sim 1e$ for CCDs at long shaping times to $\sim 10^4$ e in high-capacitance liquid argon calorimeters. Silicon strip detectors typically operate at $\sim 10^3$ e electrons, whereas pixel detectors with fast readout provide noise of several hundred electrons.

In timing measurements, the slope-to-noise ratio must be optimized, rather than the signal-to-noise ratio alone, so the rise time t_r of the pulse is important. The "jitter" σ_t of the timing distribution is

$$\sigma_t = \frac{\sigma_n}{(dS/dt)_{S_T}} \approx \frac{t_r}{S/N}\ , \tag{28.32}$$

where σ_n is the rms noise and the derivative of the signal dS/dt is evaluated at the trigger level S_T. To increase dS/dt without incurring excessive noise, the amplifier bandwidth should match the rise-time of the detector signal. The 10 to 90% rise time of an amplifier with bandwidth f_U is $0.35/f_U$. For example, an oscilloscope with 350 MHz bandwidth has a 1 ns rise time. When amplifiers are cascaded, which is invariably necessary, the individual rise times add in quadrature.

$$t_r \approx \sqrt{t^2_{r1} + t^2_{r2} + ... + t^2_{rn}}$$

Increasing signal-to-noise ratio also improves time resolution, so minimizing the total capacitance at the input is also important. At high signal-to-noise ratios, the time jitter can be much smaller than the rise time. The timing distribution may shift with signal level ("walk"), but this can be corrected by various means, either in hardware or software [9].

For a more detailed introduction to detector signal processing and electronics see Ref. 110.

28.13. Calorimeters

28.13.1. *Electromagnetic calorimeters* :

Written August 2003 by R.-Y. Zhu (California Inst. of Technology).

The development of electromagnetic showers is discussed in the section on "Passage of Particles Through Matter" (Sec. 27 of this *Review*).

Formulae are given which approximately describe average showers, but since the physics of electromagnetic showers is well understood, detailed and reliable Monte Carlo simulation is possible. EGS4 [117] and GEANT [118] have emerged as the standards.

There are homogeneous and sampling electromagnetic calorimeters. In a homogeneous calorimeter the entire volume is sensitive, *i.e.*, contributes signal. Homogeneous electromagnetic calorimeters may be built with inorganic heavy (high-Z) scintillating crystals such as BGO, CsI, NaI, and PWO, non-scintillating Cherenkov radiators such as lead glass and lead fluoride, or ionizing noble liquids. Properties of commonly used inorganic crystal scintillators can be found in Table 28.4. A sampling calorimeter consists of an active medium which generates signal and a passive medium which functions as an absorber. The active medium may be a scintillator, an ionizing noble liquid, a gas chamber, or a semiconductor. The passive medium is usually a material of high density, such as lead, iron, copper, or depleted uranium.

The energy resolution σ_E/E of a calorimeter can be parametrized as $a/\sqrt{E} \oplus b \oplus c/E$, where $\oplus$ represents addition in quadrature and E is in GeV. The stochastic term a represents statistics-related fluctuations such as intrinsic shower fluctuations, photoelectron statistics, dead material at the front of the calorimeter, and sampling fluctuations. For a fixed number of radiation lengths, the stochastic term a for a sampling calorimeter is expected to be proportional to $\sqrt{t/f}$, where t is plate thickness and f is sampling fraction [119,120]. While a is at a few percent level for a homogeneous calorimeter, it is typically 10% for sampling calorimeters. The main contributions to the systematic, or constant, term b are detector non-uniformity and calibration uncertainty. In the case of the hadronic cascades discussed below, non-compensation also contributes to the constant term. One additional contribution to the constant term for calorimeters built for modern high-energy physics experiments, operated in a high-beam intensity environment, is radiation damage of the active medium. This can be minimized by developing radiation-hard active media [43] and by frequent *in situ* calibration and monitoring [42,120]. With effort, the constant term b can be reduced to below one percent. The term c is due to electronic noise summed over readout channels within a few Molière radii. The best energy resolution for electromagnetic shower measurement is obtained in total absorption homogeneous calorimeters, *e.g.* calorimeters built with heavy crystal scintillators. These are used when ultimate performance is pursued.

The position resolution depends on the effective Molière radius and the transverse granularity of the calorimeter. Like the energy resolution, it can be factored as $a/\sqrt{E} \oplus b$, where a is a few to 20 mm and b can be as small as a fraction of mm for a dense calorimeter with fine granularity. Electromagnetic calorimeters may also provide direction measurement for electrons and photons. This is important for photon-related physics when there are uncertainties in event origin, since photons do not leave information in

the particle tracking system. Typical photon angular resolution is about 45 $mrad/\sqrt{E}$, which can be provided by implementing longitudinal segmentation [121] for a sampling calorimeter or by adding a preshower detector [122] for a homogeneous calorimeter without longitudinal segmentation.

Novel technologies have been developed for electromagnetic calorimetry. New heavy crystal scintillators, such as PWO, LSO:Ce, and GSO:Ce (see Sec. 28.3), have attracted much attention for homogeneous calorimetry. In some cases, such as PWO, it has received broad applications in high-energy and nuclear physics experiments. The "spaghetti" structure has been developed for sampling calorimetry with scintillating fibers as the sensitive medium. The "accordion" structure has been developed for sampling calorimetry with ionizing noble liquid as the sensitive medium. Table 28.7 provides a brief description of typical electromagnetic calorimeters built recently for high-energy physics experiments. Also listed in this table are calorimeter depths in radiation lengths (X_0) and the achieved energy resolution. Whenever possible, the performance of calorimeters *in situ* is quoted, which is usually in good agreement with prototype test beam results as well as EGS or GEANT simulations, provided that all systematic effects are properly included. Detailed references on detector design and performance can be found in Appendix C of reference [120] and Proceedings of the International Conference series on Calorimetry in Particle Physics.

28.13.2. ***Hadronic calorimeters*** : [120,123] The length scale appropriate for hadronic cascades is the nuclear interaction length, given very roughly by

$$\lambda_I \approx 35 \text{ g cm}^{-2} A^{1/3} \ . \tag{28.33}$$

Longitudinal energy deposition profiles are characterized by a sharp peak near the first interaction point (from the fairly local deposition of EM energy resulting from π^0's produced in the first interaction), followed by a more gradual development with a maximum at

$$x/\lambda_I \equiv t_{\max} \approx 0.2 \ln(E/1 \text{ GeV}) + 0.7 \tag{28.34}$$

as measured from the front of the detector.

The depth required for containment of a fixed fraction of the energy also increases logarithmically with incident particle energy. The thickness of iron required for 95% (99%) containment of cascades induced by single hadrons is shown in Fig. 28.13 [124]. Two of the sets of data are from large neutrino experiments, while the third is from a commonly-used parameterization. Depths as measured in nuclear interaction lengths presumably scale to other materials. From the same data it can be concluded that the requirement that 95% of the energy in 95% of the showers be contained requires 40 to 50 cm (2.4 to 3.0 λ_I) more material material than for an average 95% containment. The transverse dimensions of hadronic showers also scale as λ_I, although most of the energy is contained in a narrow core.

The energy deposit in a hadronic cascade consists of a prompt EM component due to π^0 production and a somewhat slower component mainly due to low-energy hadronic activity. An induction argument verified by Monte-Carlo simulations has shown that the

Table 28.7: Resolution of typical electromagnetic calorimeters. E is in GeV.

Technology (Experiment)	Depth	Energy resolution	Date
NaI(Tl) (Crystal Ball)	$20X_0$	$2.7\%/\mathrm{E}^{1/4}$	1983
$Bi_4Ge_3O_{12}$ (BGO) (L3)	$22X_0$	$2\%/\sqrt{E} \oplus 0.7\%$	1993
CsI (KTeV)	$27X_0$	$2\%/\sqrt{E} \oplus 0.45\%$	1996
CsI(Tl) (BaBar)	16–$18X_0$	$2.3\%/E^{1/4} \oplus 1.4\%$	1999
CsI(Tl) (BELLE)	$16X_0$	1.7% for $E_\gamma > 3.5$ GeV	1998
$PbWO_4$ (PWO) (CMS)	$25X_0$	$3\%/\sqrt{E} \oplus 0.5\% \oplus 0.2/E$	1997
Lead glass (OPAL)	$20.5X_0$	$5\%/\sqrt{E}$	1990
Liquid Kr (NA48)	$27X_0$	$3.2\%/\sqrt{E} \oplus 0.42\% \oplus 0.09/E$	1998
Scintillator/depleted U (ZEUS)	20–$30X_0$	$18\%/\sqrt{E}$	1988
Scintillator/Pb (CDF)	$18X_0$	$13.5\%/\sqrt{E}$	1988
Scintillator fiber/Pb spaghetti (KLOE)	$15X_0$	$5.7\%/\sqrt{E} \oplus 0.6\%$	1995
Liquid Ar/Pb (NA31)	$27X_0$	$7.5\%/\sqrt{E} \oplus 0.5\% \oplus 0.1/E$	1988
Liquid Ar/Pb (SLD)	$21X_0$	$8\%/\sqrt{E}$	1993
Liquid Ar/Pb (H1)	20–$30X_0$	$12\%/\sqrt{E} \oplus 1\%$	1998
Liquid Ar/depl. U (DØ)	$20.5X_0$	$16\%/\sqrt{E} \oplus 0.3\% \oplus 0.3/E$	1993
Liquid Ar/Pb accordion (ATLAS)	$25X_0$	$10\%/\sqrt{E} \oplus 0.4\% \oplus 0.3/E$	1996

fraction of hadronic energy in a cascade is $(E/E_0)^{m-1}$, where $0.80 \lesssim m \lesssim 0.85$ [125]. E_0 is about 1 GeV for incident pions, and the power-law description is approximately valid for incident energy E greater than a few tens of GeV. In general, the electromagnetic and hadronic energy depositions are converted to electrical signals with different efficiencies. The ratio of the conversion efficiencies is usually called the intrinsic e/h ratio. It follows in the power-law approximation the ratio of the responses for incident pions and incident electrons is given by "π/e"$= 1 - (1 - h/e)(E/E_0)^{m-1}$. With or without the power-law approximation the response for pions is not a linear function of energy for $e/h \neq 1$. (But in any case, as the energy increases a larger and larger fraction of the energy is transferred to π^0's, and "π/e"$\rightarrow 1$.) If $e/h = 1.0$ the calorimeter is said to be *compensating*. If e/h differs from unity by more than 5% or 10%, detector performance is compromised because of fluctuations in the π^0 content of the cascades. This results in (a) a skewed

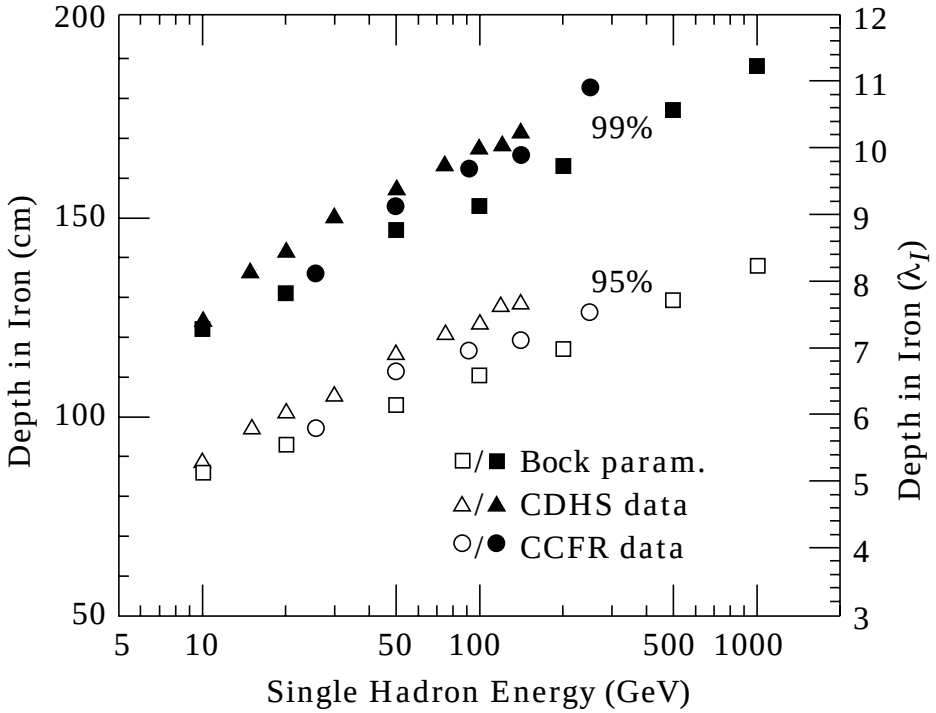


Figure 28.13: Required calorimeter thickness for 95% and 99% hadronic cascade containment in iron, on the basis of data from two large neutrino detectors and the parameterization of Bock *et al.* [124].

signal distribution and (b) an almost-constant contribution to detector resolution which is proportional to the degree of noncompensation $|1 - h/e|$. The coefficient relating the size of the constant term to $|1 - h/e|$ is 14% according to FLUKA simulations [125], and 21% according to Wigmans' calculations [126]. (Wigmans now prefers a different approach to the "constant term" [120].)

The formula for "π/e" given above is valid for a large uniform calorimeter. Real calorimeters usually have an EM front structure which is different, and so modifications must be made in modeling the response.

In most cases e/h is greater than unity, particularly if little hydrogen is present or if the gate time is short. This is because much of the low-energy hadronic energy is "hidden" in nuclear binding energy release, low-energy spallation products, *etc.* Partial correction for these losses occurs in a sampling calorimeter with high-Z absorbers, because a disproportionate fraction of electromagnetic energy is deposited in the inactive region. For this reason, a fully sensitive detector such as scintillator or glass cannot be made compensating.

The average electromagnetic energy fraction in a high-energy cascade is smaller for incident protons than for pions; $E_0 \approx 2.6$ GeV rather than ≈ 1 GeV. As a result "π/e">"p/e" (if $e/h > 1$) in a noncompensating calorimeter [125]. This difference has now been measured [127].

Circa 1990 compensation was thought to be very important in hadronic calorimeter

design. Motivated very much by the work of Wigmans [126], several calorimeters were built with $e/h \approx 1 \pm 0.02$. These include

- ZEUS [128] 2.6 cm thick scintillator sheets sandwiched between 3.3 mm depleted uranium plates; a resolution of $0.35/\sqrt{E}$ was obtained;
- ZEUS prototype study [129], with 10 mm lead plates and 2.5 mm scintillator sheets; $0.44/\sqrt{E}$;
- D0 [130], where the sandwich cell consists of a 4–6 mm thick depleted uranium plate, 2.3 mm LAr, a G-10 signal board, and another 2.3 mm LAr gap; $45\%/\sqrt{E}$.

Approximately Gaussian signal distributions were observed.

More recently, compensation has not been considered as important, and, in addition, the new generation of calorimeters for LHC experiments operate in a different energy regime and can tolerate poorer resolution in return for simpler, deeper structures. For example, the ATLAS endcaps consist of iron plates with scintillating fiber readout [131]. The fraction of the structure consisting of low-Z active region (scintillator in this case) is much smaller than would be necessary to achieve compensation. Test beam results with these modules show a resolution of $\approx 46\%/\sqrt{E}$, and $e/h = 1.5$–1.6.

28.13.3. ***Free electron drift velocities in liquid ionization sensors*** : Velocities as a function of electric field strength are given in Refs. 132–135 and are plotted in Fig. 28.14. Recent precise measurements of the free electron drift velocity in LAr have been published by W. Walkowiak [136]. These measurements were motivated by the design of the ATLAS electromagnetic calorimeter and inconsistencies in the previous literature. Velocities are systematically higher than those shown in Fig. 28.14.

28.14. Superconducting magnets for collider detectors

Revised October 2001 by R.D. Kephart (FNAL); revised September 2005 by A. Yamamoto (KEK)

28.14.1. ***Solenoid Magnets*** :

In all cases SI unit are assumed, so that the magnetic field, B, is in Tesla, the stored energy, E, is in joules, the dimensions are in meters, and $\mu_0 = 4\pi \times 10^{-7}$.

The magnetic field (B) in an ideal solenoid with a flux return iron yoke, in which the magnetic field is < 2 T, is given by

$$B = \mu_0 \, n \, I \tag{28.35}$$

where n is the number of turns/meter and I is the current. In an air-core solenoid, the central field is given by

$$B(0,0) = \mu_0 \, n \, I \frac{L}{\sqrt{L^2 + 4R^2}} \ , \tag{28.36}$$

where L is the coil length and R is the coil radius.

In most cases, momentum analysis is made by measuring the circular trajectory of the passing particles according to $p = mv\gamma = q\,rB$, where p is the momentum, m the mass, q

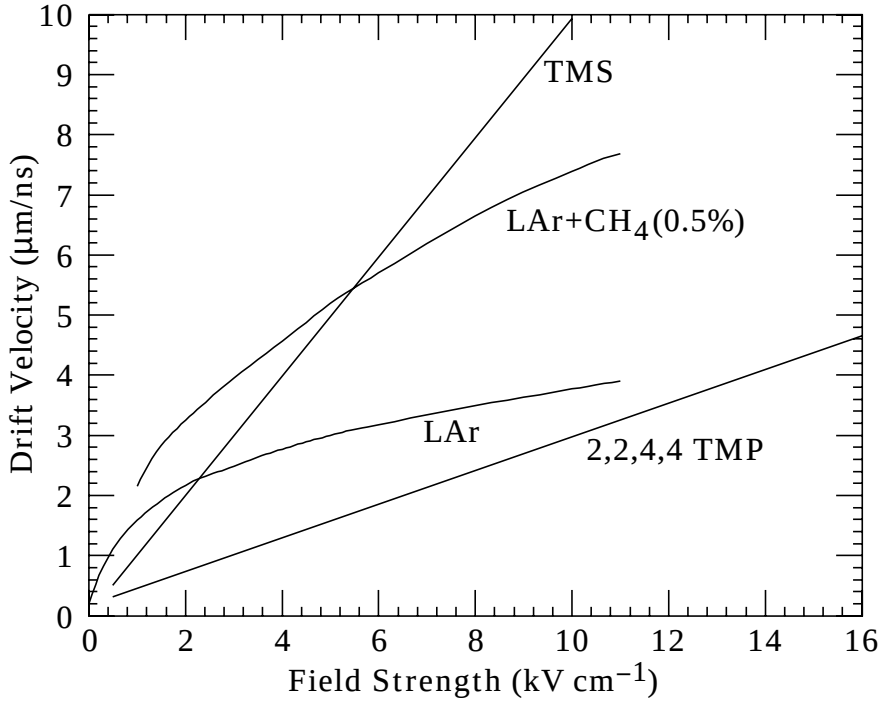


Figure 28.14: Electron drift velocity as a function of field strength for commonly used liquids.

the charge, r the bending radius. The sagitta, s, of the trajectory is given by

$$s = q\,B\,\ell^2 l/8p\ , \tag{28.37}$$

where ℓ is the path length in the magnetic field. In a practical momentum measurement in colliding beam detectors, it is more effective to increase the magnetic volume than the field strength, since

$$dp/p \propto p/B\,\ell^2\ , \tag{28.38}$$

where ℓ corresponds to the solenoid coil radius R.

The energy stored in the magnetic field of any magnet is calculated by integrating B^2 over all space:

$$E = \frac{1}{2\mu_0}\int B^2 dV \tag{28.39}$$

If the coil thin, (which is the case if it is to superconducting coil), then

$$E \approx (B^2/2\mu_0)\pi R^2 L\ . \tag{28.40}$$

For a detector in which the calorimetry is outside the aperture of the solenoid, the coil must be thin in terms of radiation and absorption lengths. This usually means that the coil is superconducting and that the vacuum vessel encasing it is of minimum real thickness and fabricated of a material with long radiation length. There are two major contributors to the thickness of a thin solenoid:

1) The conductor consisting of the current-carrying superconducting material (usually NbTi/Cu) and the quench protecting stabilizer (usually aluminum) are wound on the inside of a structural support cylinder (usually aluminum also). The coil thickness scales as B^2R, so the thickness in radiation lengths (X_0) is

$$t_{\rm coil}/X_0 = (R/\sigma_h X_0)(B^2/2\mu_0) \ , \tag{28.41}$$

where $t_{\rm coil}$ is the physical thickness of the coil, X_0 the average radiation length of the coil/stabilizer material, and σ_h is the hoop stress in the coil [137]. $B^2/2\mu_0$ is the magnetic pressure. In large detector solenoids, the aluminum stabilizer and support cylinders dominate the thickness; the superconductor (NbTI/Cu) contributes a smaller fraction. The coil package including the cryostat typically contributes about 2/3 of the total thickness in radiation lengths.

2) Another contribution to the material comes from the outer cylindrical shell of the vacuum vessel. Since this shell is susceptible to buckling collapse, its thickness is determined by the diameter, length and the modulus of the material of which it is fabricated. The outer vacuum shell represents about 1/3 of the total thickness in radiation length.

28.14.2. *Properties of collider detector magnets* :

The physical dimensions, central field stored energy and thickness in radiation lengths normal to the beam line of the superconducting solenoids associated with the major collider are given in Table 28.8 [139]. Fig. 28.15 shows thickness in radiation lengths as a function of B^2R in various collider detector solenoids.

The ratio of stored energy to cold mass (E/M) is a useful performance measure. It can also be expressed as the ratio of the stress, σ_h, to twice the equivalent density, ρ, in the coil [137]:

$$\frac{E}{M} = \frac{\int (B^2/2\mu_0) dV}{\rho\, V_{\rm coil}} \approx \frac{\sigma_h}{2\rho} \tag{28.42}$$

The E/M ratio in the coil is approximately equivalent to H,* the enthalpy of the coil, and it determines the average coil temperature rise after energy absorption in a quench:

$$E/M = H(T_2) - H(T_1) \approx H(T_2) \tag{28.43}$$

where T_2 is the average coil temperature after the full energy absorption in a quench, and T_1 is the initial temperature. E/M ratios of 5, 10, and 20 kJ/kg correspond to ∼65, ∼80, and ∼100 K, respectively. The E/M ratios of various detector magnets are shown in Fig. 28.16 as a function of total stored energy. One would like the cold mass to be as small as possible to minimize the thickness, but temperature rise during a quench must also be minimized. An E/M ratio as large as 12 kJ/kg is designed into the CMS solenoid, with the possibility that about half of the stored energy can go to an external

* The enthalpy, or heat content, is called H in the thermodynamics literature. It is not to be confused with the magnetic field intensity B/μ.

Table 28.8: Progress of superconducting magnets for particle physics detectors.

Experiment	Laboratory	B [T]	Radius [m]	Length [m]	Energy [MJ]	X/X_0	E/M [kJ/kg]
TOPAZ*	KEK	1.2	1.45	5.4	20	0.70	4.3
CDF	Tsukuba/Fermi	1.5	1.5	5.07	30	0.84	5.4
VENUS*	KEK	0.75	1.75	5.64	12	0.52	2.8
AMY*	KEK	3	1.29	3	40	‡	
CLEO-II	Cornell	1.5	1.55	3.8	25	2.5	3.7
ALEPH*	Saclay/CERN	1.5	2.75	7.0	130	2.0	5.5
DELPHI*	RAL/CERN	1.2	2.8	7.4	109	1.7	4.2
ZEUS	INFN/DESY	1.8	1.5	2.85	11	0.9	5.5
H1	RAL/DESY	1.2	2.8	5.75	120	1.8	4.8
BaBar	INFN/SLAC	1.5	1.5	3.46	27	‡	3.6
D0	Fermi	2.0	0.6	2.73	5.6	0.9	3.7
BELLE	KEK	1.5	1.8	4	42	‡	5.3
BES-III†	IHEP	1.0	1.475	3.5	9.5	‡	2.6
ATLAS-CS†	ATLAS/CERN	2.0	1.25	5.3	38	0.66	7.0
ATLAS-BT†	ATLAS/CERN	1	4.7–9.75	26	1080	(Toroid)	
ATLAS-ET†	ATLAS/CERN	1	0.825–5.35	5	2×250	(Toroid)	
CMS†	CMS/CERN	4	6	12.5	2600	‡	12

* No longer in service

† Detector under construction

‡ EM calorimeter is inside solenoid, so small X/X_0 is not a goal

dump resistor. Thus the coil temperature can be kept below 80 K if the energy extraction system work well. The limit is set by the maximum temperature that the coil design can tolerate during a quench. This maximum local temperature should be <130 K (50 K + 80 K), so that thermal expansion effects in the coil are manageable.

28.14.3. *Toroidal magnets* :

Toroidal coils uniquely provide a closed magnetic field without the necessity of an iron flux-return yoke. Because no field exists at the collision point and along the beam line, there is, in principle, no effect on the beam. On the other hand, the field profile generally has $1/r$ dependence. The particle momentum may be determined by measurements of the deflection angle combined with the sagitta. The deflection (bending) power BL is

$$BL \approx \int_{R_i}^{R_0} \frac{B_i R_i \, dR}{R \sin\theta} = \frac{B_i \, R_i}{\sin\theta} \ln(R_0/R_i) \ , \tag{28.44}$$

where R_i is the inner coil radius, R_0 is the outer coil radius, and θ is the angle between the particle trajectory and the beam line axis . The momentum resolution given by the

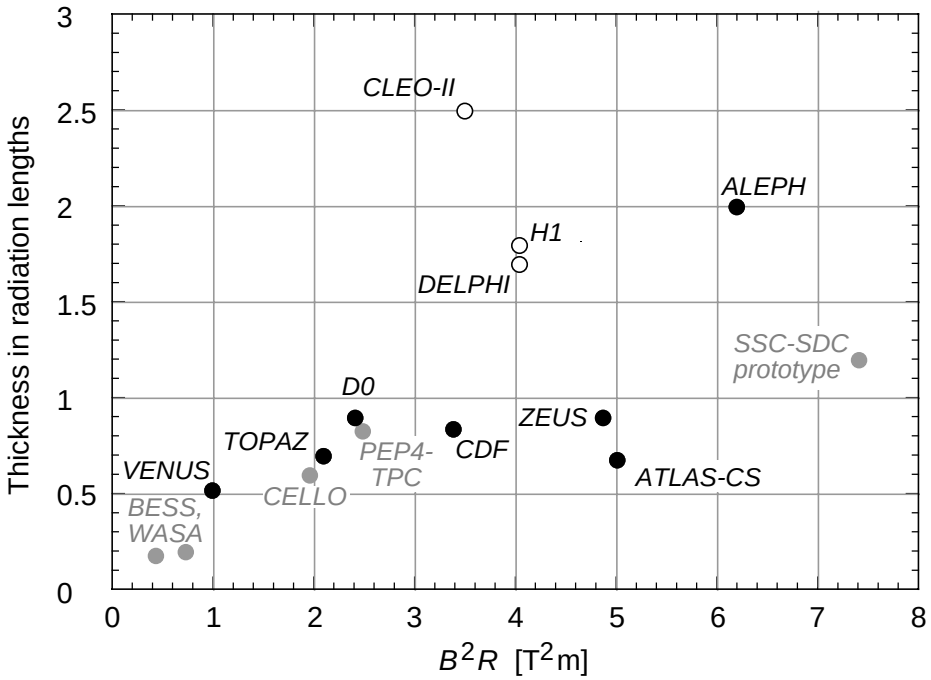


Figure 28.15: Magnet wall thickness in radiation length as a function of B^2R for various detector solenoids. Gray entries are for magnets not listed in Table 28.8. Open circles are for magnets not designed to be "thin." The SSC-SDC prototype provided important R&D for LHC magnets.

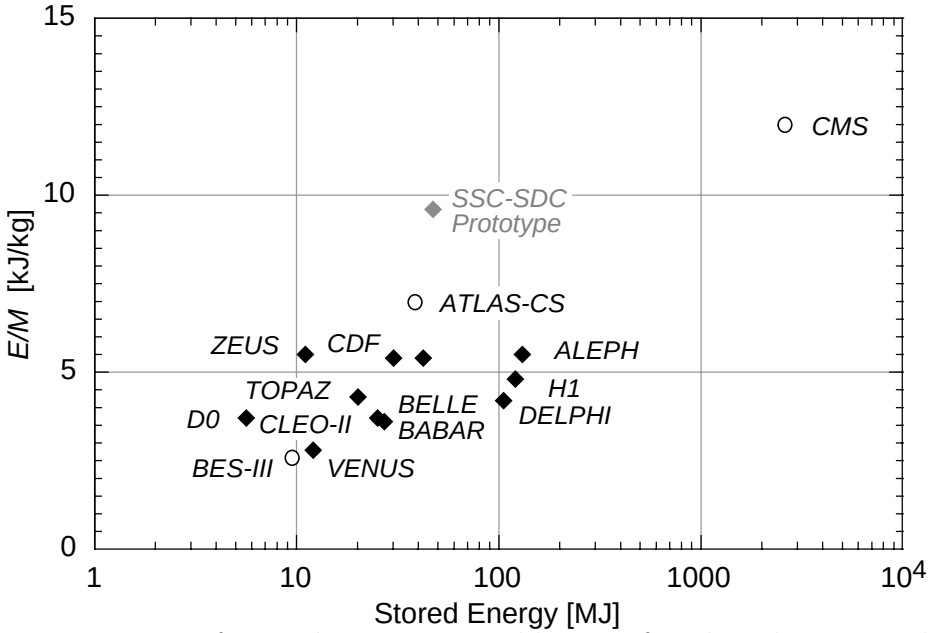


Figure 28.16: Ratio of stored energy to cold mass for thin detector solenoids. Open circles indicate magnets under construction.

deflection may be expressed as

$$\frac{\Delta p}{p} \propto \frac{p}{BL} \approx \frac{p \sin\theta}{B_i R_i \ln(R_0/R_i)} \ . \tag{28.45}$$

The momentum resolution is better in the forward/backward (smaller θ) direction. The geometry has been found to be optimal when $R_0/R_i \approx 3$–4. In practical designs, the coil is divided into 6–12 lumped coils in order to have reasonable acceptance and accessibility. This causes the coil design to be much more complex. The mechanical structure needs to sustain the decentering force between adjacent coils, and the peak field in the coil is 3–5 times higher than the useful magnetic field for the momentum analysis [138].

28.15. Measurement of particle momenta in a uniform magnetic field [140,141]

The trajectory of a particle with momentum p (in GeV/c) and charge ze in a constant magnetic field $\vec{B}$ is a helix, with radius of curvature R and pitch angle λ. The radius of curvature and momentum component perpendicular to $\vec{B}$ are related by

$$p \cos\lambda = 0.3\, z\, B\, R \ , \tag{28.46}$$

where B is in tesla and R is in meters.

The distribution of measurements of the curvature $k \equiv 1/R$ is approximately Gaussian. The curvature error for a large number of uniformly spaced measurements on the trajectory of a charged particle in a uniform magnetic field can be approximated by

$$(\delta k)^2 = (\delta k_{\rm res})^2 + (\delta k_{\rm ms})^2 \ , \tag{28.47}$$

where $\delta k =$ curvature error
$\delta k_{\rm res} =$ curvature error due to finite measurement resolution
$\delta k_{\rm ms} =$ curvature error due to multiple scattering.

If many (≥ 10) uniformly spaced position measurements are made along a trajectory in a uniform medium,

$$\delta k_{\rm res} = \frac{\epsilon}{L'^2} \sqrt{\frac{720}{N+4}} \ , \tag{28.48}$$

where $N =$ number of points measured along track
$L' =$ the projected length of the track onto the bending plane
$\epsilon =$ measurement error for each point, perpendicular to the trajectory.

If a vertex constraint is applied at the origin of the track, the coefficient under the radical becomes 320.

For arbitrary spacing of coordinates s_i measured along the projected trajectory and with variable measurement errors ϵ_i the curvature error $\delta k_{\rm res}$ is calculated from:

$$(\delta k_{\rm res})^2 = \frac{4}{w} \ \frac{V_{ss}}{V_{ss} V_{s^2 s^2} - (V_{ss^2})^2} \ , \tag{28.49}$$

where V are covariances defined as $V_{s^m s^n} = \langle s^m s^n \rangle - \langle s^m \rangle \langle s^n \rangle$ with $\langle s^m \rangle = w^{-1} \sum ({s_i}^m / {\epsilon_i}^2)$ and $w = \sum {\epsilon_i}^{-2}$.

The contribution due to multiple Coulomb scattering is approximately

$$\delta k_{\rm ms} \approx \frac{(0.016)({\rm GeV}/c)z}{Lp\beta\cos^2\lambda}\sqrt{\frac{L}{X_0}}\ , \tag{28.50}$$

where $p =$ momentum (GeV/c)
$z =$ charge of incident particle in units of e
$L =$ the total track length
$X_0 =$ radiation length of the scattering medium (in units of length; the X_0 defined elsewhere must be multiplied by density)
$\beta =$ the kinematic variable v/c.

More accurate approximations for multiple scattering may be found in the section on Passage of Particles Through Matter (Sec. 27 of this *Review*). The contribution to the curvature error is given approximately by $\delta k_{\rm ms} \approx 8 s^{\rm rms}_{\rm plane}/L^2$, where $s^{\rm rms}_{\rm plane}$ is defined there.

References:

1. *Experimental Techniques in High Energy Physics*, T. Ferbel (ed.) (Addison-Wesley, Menlo Park, CA, 1987).
2. Claus Grupen, *Particle Detectors,* Cambridge Monographs on Particle Physics, Nuclear Physics and Cosmology, # 5, Cambridge University Press (1996).
3. K. Kleinknecht, *Detectors for Particle Radiation*, Cambridge University Press (1998).
4. G.F. Knoll, *Radiation Detection and Measurement*, 3nd edition, John Wiley & Sons, New York (1999).
5. Dan Green, *The Physics of Particle Detectors,* Cambridge Monographs on Particle Physics, Nuclear Physics and Cosmology, # 12, Cambridge University Press (2000).
6. [Icarus Collaboration], ICARUS-TM/2001-09; LGNS-EXP 13/89 add 2-01.
7. E. Albert *et al.*, Nucl. Instrum. Methods **A409**, 70 (1998).
8. B. Aubert *et al.*, [BaBar Collaboration], Nucl. Instrum. Methods **A479**, 1 (2002).
9. H. Spieler, IEEE Trans. Nucl. Sci. **NS-29**, 1142 (1982).
10. K. Arisaka, Nucl. Instrum. Methods **A442**, 80 (2000).
11. Hamamatsu K.K. Electron Tube Division, *Photomultiplier Tubes: Basics and Applications*, 2nd edition (2002);
    Can be found under "Photomultiplier Tube Handbook" at
    `http://sales.hamamatsu.com/en/products/electron-tube-division/detectors/photomultiplier-tubes-pmts.php`.
12. R. Arnold *et al.*, Nucl. Instrum. Methods **A314**, 465 (1992).
13. P. Mangeot *et al.*, Nucl. Instrum. Methods **A216**, 79 (1983);
    R. Apsimon *et al.*, IEEE. Trans. Nucl. Sci. **33**, 122 (1986);
    R. Arnold *et al.*, Nucl. Instrum. Methods **A270**, 255, 289 (1988);
    D. Aston *et al.*, Nucl. Instrum. Methods **A283**, 582 (1989).

14. J. Janesick *Scientific charge-coupled devices*, SPIE Press, Bellingham, WA (2001).
15. R. Haitz *et al.*, J. Appl. Phys. **36**, 3123 (1965);
R. McIntyre, IEEE Trans. Electron Devices **13**, 164 (1966);
H. Dautet *et al.*, Applied Optics, **32**, (21), 3894 (1993);
Perkin-Elmer Optoelectronics, *Avalanche Photodiodes: A User's Guide*.
16. P. Buzhan *et al.*, Nucl. Instrum. Methods **A504**, 48 (2003);
Z. Sadygov *et al.*, Nucl. Instrum. Methods **A504**, 301 (2003);
V. Golovin and V. Saveliev, Nucl. Instrum. Methods **A518**, 560 (2004).
17. M. Landstrass *et al.*, Diam. & Rel. Matter, **2**, 1033 (1993);
R. McKeag and R. Jackman, Diam. & Rel. Matter, **7**, 513 (1998);
R. Brascia *et al.*, Phys. Stat. Sol., **199**, 113 (2003).
18. M. Petrov, M. Stapelbroek, and W. Kleinhans, Appl. Phys. Lett. **51**, 406 (1987);
M. Atac, M. Petrov, IEEE Trans. Nucl. Sci. **36** 163 (1989);
M. Atac *et al.*, Nucl. Instrum. Methods **A314**, 56 (1994).
19. A. Braem *et al.*, Nucl. Instrum. Methods **A518**, 574 (2004).
20. J.B. Birks, *The Theory and Practice of Scintillation Counting*, (Pergamon, London, 1964).
21. D. Clark, Nucl. Instrum. Methods **117**, 295 (1974).
22. J.B. Birks, Proc. Phys. Soc. **A64**, 874 (1951).
23. B. Bengston and M. Moszynski, Nucl. Instrum. Methods **117**, 227 (1974);
J. Bialkowski *et al.*, Nucl. Instrum. Methods **117**, 221 (1974).
24. *Proceedings of the Symposium on Detector Research and Development for the Superconducting Supercollider*, eds. T. Dombeck, V. Kelly, and G.P. Yost (World Scientific, Singapore, 1991).
25. I.B. Berlman, *Handbook of Fluorescence Spectra of Aromatic Molecules*, 2nd edition (Academic Press, New York, 1971).
26. C. Zorn, in *Instrumentation in High Energy Physics,* ed. F. Sauli, (1992, World Scientific, Singapore) pp. 218–279.
27. T. Foerster, Ann. Phys. **2**, 55 (1948).
28. J.M. Fluornoy, Conference on Radiation-Tolerant Plastic Scintillators and Detectors, K.F. Johnson and R.L. Clough editors, Rad. Phys. and Chem., **41** 389 (1993).
29. D. Horstman and U. Holm, *ibid,* 395.
30. D. Blomker *et al.*, Nucl. Instrum. Methods **A311**, 505 (1992);
J. Mainusch *et al.*, Nucl. Instrum. Methods **A312**, 451 (1992).
31. Conference on Radiation-Tolerant Plastic Scintillators and Detectors, K.F. Johnson and R.L. Clough editors, Rad. Phys. and Chem., **41** (1993).
32. S.R. Borenstein and R.C. Strand, IEEE Trans. Nuc. Sci. **NS-31(1)**, 396 (1984).
33. P. Sonderegger, Nucl. Instrum. Methods **A257**, 523 (1987).
34. S.A. Sedykh *et al.*, Nucl. Instrum. Methods **A455**, 346 (2000).
35. SCIFI 97: Conference on Scintillating Fiber Detectors, 1997 University of Notre Dame, Indiana, eds. A. Bross, R. Ruchti, and M. Wayne.
36. K.F. Johnson, Nucl. Instrum. Methods **A344**, 432 (1994).
37. C.M. Hawkes *et al.*, Nucl. Instrum. Methods **A292**, 329 (1990).

38. A. Lempicki *et al.*, Nucl. Instrum. Methods **A333**, 304 (1993); G. Blasse, *Proceedings of the Crystal 2000 International Workshop on Heavy Scintillators for Scientific and Industrial Applications*, Chamonix, France, Sept. (1992), Edition Frontieres.
39. C. Melcher and J. Schweitzer, Nucl. Instrum. Methods **A314**, 212 (1992).
40. K. Takagi and T. Fakazawa, Appl. Phys. Lett. **42**, 43 (1983).
41. C. Kuntner *et al.*, Nucl. Instrum. Methods **A493**, 131 (2002).
42. G. Gratta, H. Newman, and R.Y. Zhu, Ann. Rev. Nucl. and Part. Sci. **44**, 453 (1994).
43. R.Y. Zhu, Nucl. Instrum. Methods **A413**, 297 (1998).
44. M. Woods *et al.*, SPIN96 (QCD161:S921:1996) 843.
45. A. Abashian *et al.*, Nucl. Instrum. Methods **A479**, 117 (2002).
46. M. Shiozawa, [Super-Kamiokande Collaboration], Nucl. Instrum. Methods **A433**, 240 (1999).
47. J. Litt and R. Meunier, Ann. Rev. Nucl. Sci. **23**, 1 (1973).
48. D. Bartlett *et al.*, Nucl. Instrum. Methods **A260**, 55 (1987).
49. B. Ratcliff, Nucl. Instrum. Methods **A502**, 211 (2003).
50. M. Cavalli-Sforza *et al.*, "Construction and Testing of the SLC Cherenkov Ring Imaging Detector," IEEE **37**, N3:1132 (1990).
51. E.G. Anassontzis *et al.*, "Recent Results from the DELPHI Barrel Ring Imaging Cherenkov Counter," IEEE **38**, N2:417 (1991).
52. See the RICH Workshop series: Nucl. Instrum. Methods **A343**, 1 (1993); Nucl. Instrum. Methods **A371**, 1 (1996); Nucl. Instrum. Methods **A433**, 1 (1999); Nucl. Instrum. Methods **A502**, 1 (2003).
53. H. Blood *et al.*, FERMILAB-PUB-76-051-EXP.
54. L. Sulak, HUEP-252 Presented at the Workshop on Proton Stability, Madison, Wisc. (1978).
55. K.S. Hirata *et al.*, Phys. Lett. **B205**, 416 (1988).
56. S. Kasuga *et al.*, Phys. Lett. **B374**, 238 (1996).
57. M.H. Ahn *et al.*, Phys. Rev. Lett. **90**, 041801 (2003).
58. R.M. Bionta *et al.*, Phys. Rev. Lett. **51**, 27 (1983); [Erratum-ibid. **51**, 522 (1983)].
59. R. Becker-Szendy *et al.*, Nucl. Instrum. Methods **A324**, 363 (1993).
60. H. Ikeda *et al.*, UTLICEPP-82-04.
61. K. Arisaka *et al.*, J. Phys. Soc. Jap. **54**, 3213 (1985).
62. K.S. Hirata *et al.*, Phys. Rev. **D38**, 448 (1988).
63. C. Athanassopoulos *et al.*, Nucl. Instrum. Methods **A388**, 149 (1997).
64. Y. Fukuda *et al.*, Nucl. Instrum. Methods **A501**, 418 (2003).
65. S.H. Ahn *et al.*, Phys. Lett. **B511**, 178 (2001).
66. J. Boger *et al.*, Nucl. Instrum. Methods **A449**, 172 (2000).
67. B. Dolgoshein Nucl. Instrum. Methods **A326**, 434 (1993).
68. X. Artru *et al.*, Phys. Rev. **D12**, 1289 (1975).
69. G.M. Garibian *et al.*, Nucl. Instrum. Methods **125**, 133 (1975).
70. G. Bassompierre *et al.*, Nucl. Instrum. Methods **411**, 63 (1998).

71. ALICE Collaboration, "Technical Design Report of the Transition Radiation Tracker," CERN/LHCC/ 2001-021 (2001).
72. RD6 Collaboration, CERN/DRDC 90-38 (1990); CERN/DRDC 91-47 (1991); CERN/DRDC 93-46 (1993).
73. T. Kirn et al., *Proceedings of TRDs for the 3rd millenium,* Workshop on advanced transition radiation detectors for accelerator and space applications, eds N. Giglietto and P. Spinelli, Frascati Physics Series, Vol. XXV, 161 (2002).
74. P. Spinelli et al., Proceedings of "TRDs for the 3rd millenium", Workshop on advanced transition radiation detectors for accelerator and space applications,"eds N. Giglietto and P. Spinelli, Frascati Physics Series, Vol. XXV, 177 (2002).
75. ATLAS Collaboration, ATLAS Inner Detector Technical Design Report, v 2, ATLAS TDR 5, CERN/LHCC/97-16 (30 April 1997).
76. ATLAS Collaboration, Detector and Physics Performance Technical Design Report, CERN/LHCC/99-14, 71 (1999).
77. B. Dolgoshein, Nucl. Instrum. Methods **252**, 137 (1986).
78. C.W. Fabjan *et al.*, Nucl. Instrum. Methods **185**, 119 (1981).
79. J. Cobb *et al.*, Nucl. Instrum. Methods **140**, 413 (1977).
80. A. Büngener *et al.*, Nucl. Instrum. Methods **214**, 261 (1983).
81. R.D. Appuhn *et al.*, Nucl. Instrum. Methods **263**, 309 (1988).
82. Y. Watase *et al.*, Nucl. Instrum. Methods **248**, 379 (1986).
83. R. Ansari *et al.*, Nucl. Instrum. Methods **263**, 51 (1988).
84. H.J. Butt *et al.*, Nucl. Instrum. Methods **252**, 483 (1986).
85. J.F. Detoeuf *et al.*, Nucl. Instrum. Methods **265**, 157 (1988).
86. M. Holder *et al.*, Nucl. Instrum. Methods **263**, 319 (1988).
87. H. Weidkamp, DiplomArbeit, Rhein-Westf. Tech. Hochschule Aachen (1984).
88. H. Grässler *et al.*, Proc. Vienna Wire Chamber Conference (1989).
89. T. Akesson *et al.*, Nucl. InstrumMethods A412, 200 (1998).
90. F.F. Rieke and W. Prepejchal, Phys. Rev. **A6**, 1507 (1972).
91. L.G. Christophorou, "Atomic and molecular radiation physics" (Wiley, London 1991).
92. G. Charpak *et al.*, Nucl. Instrum. Methods **62**, 262 (1968).
93. R. Veenhof, GARFIELD program: simulation of gaseous detectors, version 6.32, CERN Program Library Pool W999 (W5050).
94. As representative examples see: B. Adeva *et al.*, Nucl. Instrum. Methods **A287**, 35 (1990).
95. As representative example see: A. Alexander *et al.*, Nucl. Instrum. Methods **A276**, 42 (1989).
96. As representative examples see: F. Bedeschi *et al.*, Nucl. Instrum. Methods **A268**, 50 (1988);
    OpalCollaboration: Nucl. Instrum. Methods **A305**, 275 (1991).
97. W. Blum and G. Rolandi, *Particle Detection with Drift Chambers*, Springer-Verlag (1994).
98. A. Peisert and F. Sauli, CERN-84-08 (Jul 1984).
99. R. Bellazzini and A. M. Spezziga, Rivista del Nuovo Cimento **17**, 1 (1994).

100. F. Sauli, Nucl. Instrum. Methods **A386**, 531 (1997).
101. Y. Giomataris, Ph. Rebougeard, J.P. Robert and G. Charpak, Nucl. Instrum. Methods **A376**, 29 (1996).
102. R. Santonico and R. Cardarelli, Nucl. Instrum. Methods **A187**, 377 (1981).
103. V. V. Parkhomchuck, Yu. N. Pestov, & N. V. Petrovykh, Nucl. Instrum. Methods **93**, 269 (1971).
104. P. Fonte, IEEE. Trans. Nucl. Sci. **49**, 881 (2002).
105. M. Abbrescia *et al.*, Nucl. Instrum. Methods **A394**, 13 (1997).
106. F. Anulli *et al.*, Nucl. Instrum. Methods **A552**, 276 (2005).
107. D.R. Nygren and J.N. Marx, "The Time Projection Chamber," Phys. Today **31**, 46 (1978).
108. F. Sauli, Nucl. Instrum. Methods **386**, 531 (1997).
109. Y. Giomataris, Ph. Rebougeard, J.P. Robert and G. Charpak, Nucl. Instrum. Methods **376**, 29 (1996).
110. H. Spieler, *Semiconductor Detector Systems*, Oxford Univ. Press, Oxford (2005) ISBN 0-19-852784-5.
111. F. Scholze *et al.*, Nucl. Instrum. Methods **A439**, 208 (2000).
112. G. Lindström *et al.*, Nucl. Instrum. Methods **A465**, 60 (2001).
113. G. Lindström *et al.*, Nucl. Instrum. Methods **A426**, 1 (1999).
114. A. Holmes-Siedle and L. Adams, *Handbook of Radiation Effects*, 2nd ed., Oxford 2002, ISBN 0-19-850733-X, QC474.H59 2001.
115. V. Radeka, IEEE Trans. Nucl. Sci. **NS-15/3**, 455 (1968);
V. Radeka, IEEE Trans. Nucl. Sci. **NS-21**, 51 (1974).
116. F.S. Goulding, Nucl. Instrum. Methods **100**, 493 (1972);
F.S. Goulding and D.A. Landis, IEEE Trans. Nucl. Sci. **NS-29**, 1125 (1982).
117. W.R. Nelson, H. Hirayama, and D.W.O. Rogers, "The EGS4 Code System," SLAC-265, Stanford Linear Accelerator Center (Dec. 1985).
118. R. Brun *et al.*, *GEANT3*, `CERN DD/EE/84-1` (1987).
119. D. Hitlin *et al.*, Nucl. Instrum. Methods **137**, 225 (1976). See also W. J. Willis and V. Radeka, Nucl. Instrum. Methods **120**, 221 (1974), for a more detailed discussion.
120. R. Wigmans, *Calorimetry: Energy Measurement in Particle Physics*, Clarendon, Oxford (2000).
121. ATLAS Collaboration, *The ATLAS Liquid Argon Calorimeter Technical Design Report*, CERN/LHCC 96-41 (1996).
122. CMS Collaboration, *The CMS Electromagnetic Calorimeter Technical Design Report*, CERN/LHCC 97-33 (1997).
123. C. Leroy and P.-G. Rancoita, Rep. Prog. Phys. **63**, 505 (2000).
124. D. Bintinger, in *Proceedings of the Workshop on Calorimetry for the Supercollider*, Tuscaloosa, AL, March 13–17, 1989, edited by R. Donaldson and M.G.D. Gilchriese (World Scientific, Teaneck, NJ, 1989), p. 91;
R.K. Bock, T. Hansl-Kozanecka, and T.P. Shah, Nucl. Instrum. Methods **186**, 533 (1981).
125. T.A. Gabriel, D.E. Groom, P.K. Job, N.V. Mokhov, and G.R. Stevenson, Nucl. Instrum. Methods **A338**, 336 (1994).

126. R. Wigmans, Nucl. Instrum. Methods **A259**, 389 (1987);
R. Wigmans, Nucl. Instrum. Methods **A265**, 273 (1988).
127. N. Akchurian *et al.*, Nucl. Instrum. Methods **A408**, 380 (1998).
128. U. Behrens *et al.*, Nucl. Instrum. Methods **A289**, 115 (1990);
A. Bernstein *et al.*, Nucl. Instrum. Methods **A336**, 23 (1993).
129. E. Bernardi *et al.*, Nucl. Instrum. Methods **A262**, 229 (1987).
130. S. Abachi *et al.*, Nucl. Instrum. Methods **A324**, 53 (1993).
131. F. Ariztizabal *et al.*, Nucl. Instrum. Methods **A349**, 384 (1994).
132. E. Shibamura *et al.*, Nucl. Instrum. Methods **131**, 249 (1975).
133. T.G. Ryan and G.R. Freeman, J. Chem. Phys. **68**, 5144 (1978).
134. W.F. Schmidt, "Electron Migration in Liquids and Gases," HMI B156 (1974).
135. A.O. Allen, "Drift Mobilities and Conduction Band Energies of Excess Electrons in Dielectric Liquids," NSRDS-NBS-58 (1976).
136. W. Walkowiak, Nucl. Instrum. Methods **A449**, 288 (2000).
137. A. Yamamoto, Nucl. Instr. Meth. **A453**, 445 (2000).
138. T. Taylor, Phys. Scr. **23**, 459 (1980).
139. A. Yamamoto, Nucl. Instr. Meth. **A494**, 255 (2003).
140. R.L. Gluckstern, Nucl. Instrum. Methods **24**, 381 (1963).
141. V. Karimäki, Nucl. Instrum. Methods **A410**, 284 (1998).

29. RADIOACTIVITY & RADIATION PROTECTION

Revised Sept. 1996 by R.S. Donahue (LBNL) and A. Fassó (SLAC);
revised Sept. 2005 by J.C. Liu (SLAC) and S. Roesler (CERN).

29.1. Definitions

The International Commission on Radiation Units and Measurements (ICRU) recommends the use of SI units. Therefore we list SI units first, followed by cgs (or other common) units in parentheses, where they differ.

• **Unit of activity** = becquerel (curie):
1 Bq = 1 disintegration s^{-1} [$= 1/(3.7 \times 10^{10})$ Ci]

• **Unit of absorbed dose in any material** = gray (rad):
1 Gy = 1 joule kg^{-1} ($= 10^4$ erg $g^{-1} = 100$ rad)
$= 6.24 \times 10^{12}$ MeV kg^{-1} deposited energy

• **Unit of exposure**, A measure of photon fluence at a certain point in space integrated over time, in terms of ion charge of either sign produced by secondary electrons in a small volume of air about the point:
= 1 C kg^{-1} of air (roentgen; 1 R = 1 esu cm^{-3} in air = 2.58×10^{-4} C kg^{-1})
Implicit in the definition is the assumption that the small test volume is embedded in a sufficiently large uniformly irradiated volume that the number of secondary electrons entering the volume equals the number leaving (so-called charged particle equilibrium). This unit is somewhat historical, but appears on many measuring instruments.

• **Unit of equivalent dose** (for biological damage) = sievert [= 100 rem (**r**oentgen **e**quivalent for **m**an)]: Equivalent dose H_T (Sv) in an organ T is equal to the absorbed dose in the organ (Gy) times the radiation weighting factor w_R (formerly the quality factor Q, but w_R is defined for the radiation incident on the body). It expresses long-term risks (primarily cancer and leukemia) from low-level chronic exposure. It depends upon the type of radiation and other factors, as follows [1]:

Table 29.1: Radiation weighting factors.

Radiation	w_R
X- and γ-rays, all energies	1
Electrons and muons, all energies	1
Neutrons $<$ 10 keV	5
10–100 keV	10
$>$ 100 keV to 2 MeV	20
2–20 MeV	10
$>$ 20 MeV	5
Protons (other than recoils) $>$ 2 MeV	5
Alphas, fission fragments, & heavy nuclei	20

The sum of the equivalent doses, weighted by the tissue weighting factors w_T of several organs and tissues of the body that are considered to be most sensitive [1], is called "effective dose" E:

$$E = \sum_T w_T \times H_T \tag{29.1}$$

29.2. Radiation levels [2]

• **Natural annual background**, all sources: Most world areas, whole-body equivalent dose rate ≈ (0.4–4) mSv (40–400 mrem). Can range up to 50 mSv (5 rem) in certain areas. U.S. average ≈ 3.6 mSv, including ≈ 2 mSv (≈ 200 mrem) from inhaled natural radioactivity, mostly radon and radon daughters (Average is for a typical house and varies by more than an order of magnitude. It can be more than two orders of magnitude higher in poorly ventilated mines. 0.1–0.2 mSv in open areas).

• **Cosmic ray background** in counters (sea level, mostly muons):
~ 1 min^{-1} cm^{-2} sr^{-1}. For more accurate estimates and details, see the Cosmic Rays section (Sec. 24 of this *Review*).

• **Fluence** (per cm^2) to deposit one Gy, assuming uniform irradiation:

≈ (**charged particles**) $6.24{\times}10^9/(dE/dx)$, where dE/dx (MeV g^{-1} cm^2), the energy loss per unit length, may be obtained from Figs. 27.3 and 27.4 in Sec. 27 of this *Review*, and `pdg.lbl.gov/AtomicNuclearProperties`.

≈ 3.5×10^9 cm^{-2} minimum-ionizing singly-charged particles in carbon.

≈ (**photons**) $6.24{\times}10^9/[Ef/\ell]$, for photons of energy E (MeV), attenuation length ℓ (g cm^{-2}), and fraction $f \lesssim 1$ expressing the fraction of the photon's energy deposited in a small volume of thickness $\ll \ell$ but large enough to contain the secondary electrons.

≈ 2×10^{11} photons cm^{-2} for 1 MeV photons on carbon ($f \approx 1/2$).

• **Recommended limits to exposure of radiation workers (whole-body dose):**[*]

EU/Switzerland: 20 mSv yr^{-1}

U.S.: 50 mSv yr^{-1} (5 rem yr^{-1})[†]

• **Lethal dose**: The whole-body dose from penetrating ionizing radiation resulting in 50% mortality in 30 days (assuming no medical treatment) is 2.5–4.5 Gy (250–450 rad), as measured internally on body longitudinal center line. Surface dose varies due to variable body attenuation and may be a strong function of energy.

• **Cancer induction by low LET radiation**: The cancer induction probability is about 5% per Sv on average for the entire population. [1]

29.3. Prompt neutrons at accelerators

Neutrons dominate the particle environment outside thick shielding (> 1 m) for high energy (> a few hundred MeV) electron and hadron accelerators.

29.3.1. ***Electron beams*** : At electron accelerators neutrons are generated via photonuclear reactions from bremsstrahlung photons. Neutron yields from semi-infinite targets per unit electron beam power are plotted in Fig. 29.1 as a function of electron beam energy [3]. In the photon energy range 10–30 MeV neutron production results from the giant photonuclear resonance mechanism. Neutrons are produced roughly isotropically (within a factor of 2) and with a Maxwellian energy distribution described as:

$$\frac{dN}{dE_n} = \frac{E_n}{T^2} e^{-E_n/T} \ , \tag{29.2}$$

where T is the nuclear temperature characteristic of the target nucleus, generally in the range of $T = 0.5$–1.0 MeV. For higher energy photons the quasi-deuteron and photopion production mechanisms become important.

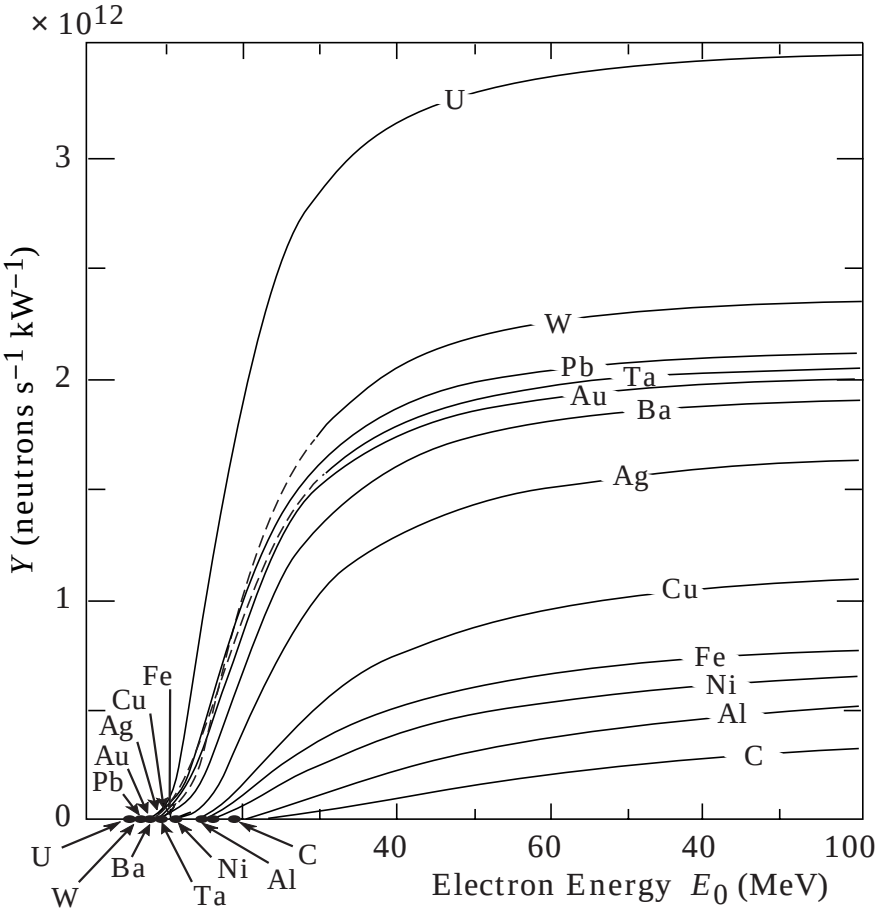


Figure 29.1: Neutron yields from semi-infinite targets, per kW of electron beam power, as a function of electron beam energy, disregarding target self-shielding.

29.3.2. ***Proton beams*** : At proton accelerators neutron yields emitted per incident proton by different target materials are roughly independent [4] of proton energy between 20 MeV and 1 GeV and are given by the ratio C:Al:Cu-Fe:Sn:Ta-Pb $= 0.3 : 0.6 : 1.0 : 1.5 : 1.7$. Above 1 GeV the neutron yield [5] is proportional to E^m, where $0.80 \leq m \leq 0.85$.

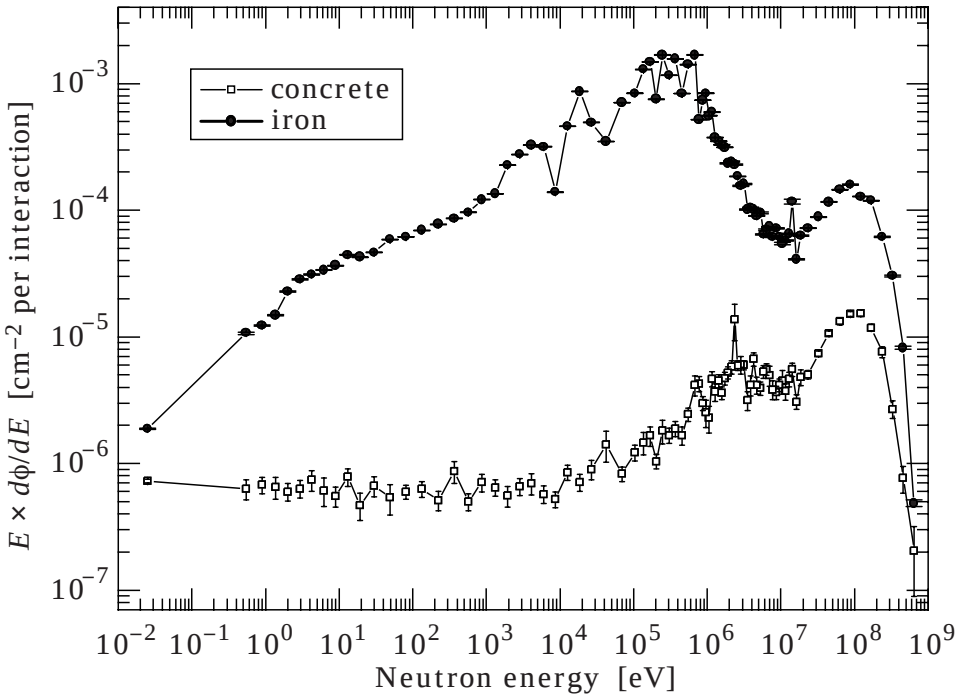


Figure 29.2: Calculated neutron spectrum from 205 GeV/c hadrons (2/3 protons and 1/3 π^+) on a thick copper target [6]. Spectra are evaluated at 90° to beam and through 80 cm of normal density concrete or 40 cm of iron.

A typical neutron spectrum [6] outside a proton accelerator concrete shield is shown in Fig. 29.2. The shape of these spectra are generally characterized as having a thermal-energy peak which is very dependent on geometry and the presence of hydrogenous material, a low-energy evaporation peak around $1 - 2$ MeV, and a high-energy spallation shoulder at around $70 - 80$ MeV.

The neutron-attenuation length, is shown in Fig. 29.3 for concrete and mono-energetic broad-beam conditions.

Letaw's [7] formula for the energy dependence of the inelastic proton cross-section for $E < 2$ GeV is:

$$\sigma(E) = \sigma_{\text{asympt}} \left[1 - 0.62e^{-E/200} \sin(10.9E^{-0.28})\right] \ , \tag{29.3}$$

and for $E > 2$ GeV:

$$\sigma_{\text{asympt}} = 45A^{0.7} \left[1 + 0.016 \sin(5.3 - 2.63 \ln\ A)\right] \ , \tag{29.4}$$

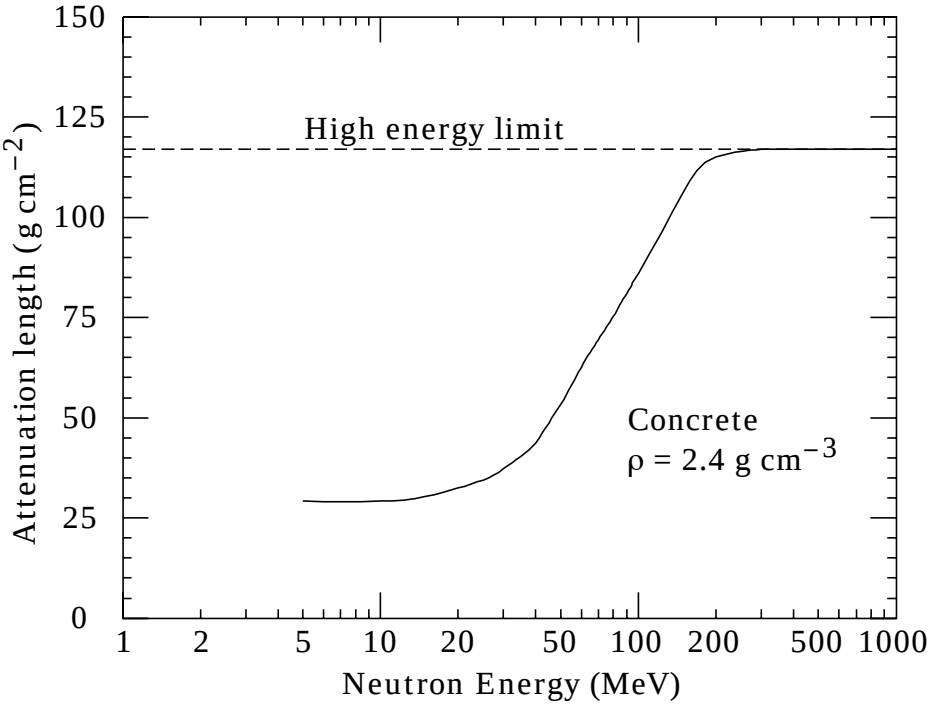


Figure 29.3: The variation of the attenuation length for mono-energetic neutrons in concrete as a function of neutron energy [4].

where σ is in mb, E is the proton energy in MeV and A is the mass number.

29.4. Dose conversion factors

Conversion coefficients from fluence to effective dose are given for anterior-posterior irradiation and various particles in Fig. 29.4 [8]. These factors can be used for converting particle fluence to dose for personnel protection purposes. For example, the effective dose from an anterior-posterior irradiation in a field of 1-MeV neutrons with a fluence of 1 neutron / cm^2 is about 290 pSv.

29.5. Accelerator-induced activity

The dose rate at 1 m due to spallation-induced activity by high energy hadrons in a 1 g medium atomic weight target can be estimated [9] from the following expression:

$$D = D_0 \, \Phi \, \ln[(T+t)/t] \, , \tag{29.5}$$

where T is the irradiation time, t is the decay time since irradiation, Φ is the flux of irradiating hadrons (hadrons cm^{-2} s^{-1}) and D_0 has a value of 5.2×10^{-17} [(Sv hr^{-1})/(hadron cm^{-2} s^{-1})]. This relation is essentially independent of hadron energy above 200 MeV.

Dose due to accelerator-produced induced activity can also be estimated with the use of "ω factors" [4]. These factors give the dose rate per unit star density (inelastic

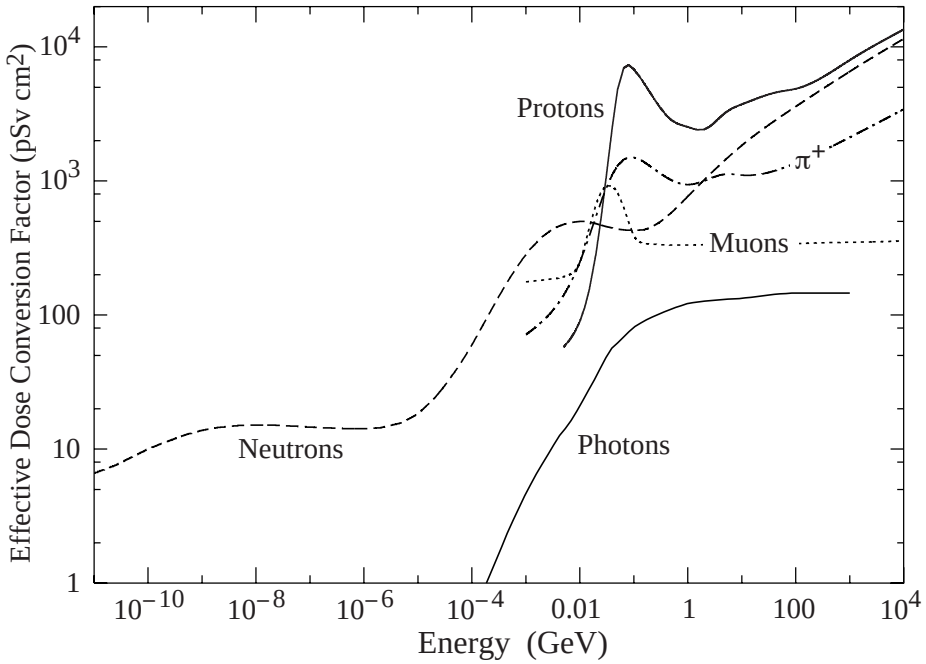


Figure 29.4: Fluence to effective dose conversion factors for anterior-posterior irradiation and various particles [8].

reaction for $E > 50$ MeV) after a 30-day irradiation and 1-day decay. The ω factor for steel or iron is $\simeq 3 \times 10^{-12}$ (Sv cm^3/star). This does not include possible contributions from thermal-neutron activation. Induced activity in concrete can vary widely depending on concrete composition, particularly with the concentration of trace quantities such as sodium. Additional information can be found in Barbier [10].

29.6. Photon sources

The dose rate in air from a gamma point source of C Curies emitting one photon of energy $0.07 < E < 4$ MeV per disintegration at a distance of 30 cm is about $6\,C\,E$ (rem/hr), or $60\,C\,E$ (mSv/hr), $\pm 20\%$. In general, the dependence of the dose rate from a point source on the distance r follows a $1/r^2$ behaviour

$$D(r) = \frac{D(r_0)}{(r/r_0)^2}. \tag{29.6}$$

The dose rate in air from a semi-infinite uniform photon source of specific activity C (μCi/g) and gamma energy E (MeV) is about $1.07CE$ (rem/hr), or $10.7CE$ (mSv/hr).

Footnotes:

* The ICRP recommendation [1] is 20 mSv yr^{-1} averaged over 5 years, with the dose in any one year $\leq$ 50 mSv.

† Many laboratories in the U.S. and elsewhere set lower limits.

References:

1. ICRP Publication 60, *1990 Recommendation of the International Commission on Radiological Protection* Pergamon Press (1991).
2. See E. Pochin, *Nuclear Radiation: Risks and Benefits*, Clarendon Press, Oxford, 1983.
3. W.P. Swanson, *Radiological Safety Aspects of the operation of Electron Linear Accelerators,* IAEA Technical Reports Series No. 188 (1979).
4. R.H. Thomas and G.R. Stevenson, *Radiological Safety Aspects of the Operation of Proton Accelerators,* IAEA Technical Report Series No. 283 (1988).
5. T.A. Gabriel *et al.*, "Energy Dependence of Hadronic Activity," Nucl. Instrum. Methods **A338**, 336 (1994).
6. C. Birattari *et al.*, "Measurements and characterization of high energy neutron fields," Nucl. Instrum. Methods **A338**, 534 (1994).
7. J.R. Letaw, R. Silberberg, and C.H. Tsao, "Proton-nucleus Total Inelastic Cross Sections: An Empirical Formula for E > 10 MeV," *Astrophysical Journal Supplement Series,* **51**, 271 (1983);
   For improvements to this formula see Shen Qing-bang, "Systematics of intermediate energy proton nonelastic and neutron total cross section," International Nuclear Data Committee INDC(CPR)-020 (July 1991).
8. M. Pelliccioni, "Overview of fluence-to-effective dose and fluence-to-ambient dose equivalent conversion coefficients for high energy radiation calculated using the FLUKA code," Radiation Protection Dosimetry 88, 279 (2000).
9. A.H. Sullivan *A Guide To Radiation and Radioactivity Levels Near High Energy Particle Accelerators,* Nuclear Technology Publishing, Ashford, Kent, England (1992).
10. M. Barbier, *Induced Activity,* North-Holland, Amsterdam (1969).

30. COMMONLY USED RADIOACTIVE SOURCES

Table 30.1. Revised November 1993 by E. Browne (LBNL).

			Particle		Photon	
Nuclide	Half-life	Type of decay	Energy (MeV)	Emission prob.	Energy (MeV)	Emission prob.
$^{22}_{11}$Na	2.603 y	β^+, EC	0.545	90%	0.511	Annih.
					1.275	100%
$^{54}_{25}$Mn	0.855 y	EC			0.835	100%
					Cr K x rays	26%
$^{55}_{26}$Fe	2.73 y	EC			Mn K x rays:	
					0.00590	24.4%
					0.00649	2.86%
$^{57}_{27}$Co	0.744 y	EC			0.014	9%
					0.122	86%
					0.136	11%
					Fe K x rays	58%
$^{60}_{27}$Co	5.271 y	β^-	0.316	100%	1.173	100%
					1.333	100%
$^{68}_{32}$Ge	0.742 y	EC			Ga K x rays	44%
$\rightarrow {}^{68}_{31}$Ga		β^+, EC	1.899	90%	0.511	Annih.
					1.077	3%
$^{90}_{38}$Sr	28.5 y	β^-	0.546	100%		
$\rightarrow {}^{90}_{39}$Y		β^-	2.283	100%		
$^{106}_{44}$Ru	1.020 y	β^-	0.039	100%		
$\rightarrow {}^{106}_{45}$Rh		β^-	3.541	79%	0.512	21%
					0.622	10%
$^{109}_{48}$Cd	1.267 y	EC	0.063 e^-	41%	0.088	3.6%
			0.084 e^-	45%	Ag K x rays	100%
			0.087 e^-	9%		
$^{113}_{50}$Sn	0.315 y	EC	0.364 e^-	29%	0.392	65%
			0.388 e^-	6%	In K x rays	97%
$^{137}_{55}$Cs	30.2 y	β^-	0.514 e^-	94%	0.662	85%
			1.176 e^-	6%		
$^{133}_{56}$Ba	10.54 y	EC	0.045 e^-	50%	0.081	34%
			0.075 e^-	6%	0.356	62%
					Cs K x rays	121%
$^{207}_{83}$Bi	31.8 y	EC	0.481 e^-	2%	0.569	98%
			0.975 e^-	7%	1.063	75%
			1.047 e^-	2%	1.770	7%
					Pb K x rays	78%
$^{228}_{90}$Th	1.912 y	6α:	5.341 to 8.785		0.239	44%
		3β^-:	0.334 to 2.246		0.583	31%
					2.614	36%
($\rightarrow {}^{224}_{88}$Ra	$\rightarrow {}^{220}_{86}$Rn	$\rightarrow {}^{216}_{84}$Po	$\rightarrow {}^{212}_{82}$Pb	$\rightarrow {}^{212}_{83}$Bi	$\rightarrow {}^{212}_{84}$Po)	
$^{241}_{95}$Am	432.7 y	α	5.443	13%	0.060	36%
			5.486	85%	Np L x rays	38%
$^{241}_{95}$Am/Be	432.2 y	6×10^{-5} neutrons (4–8 MeV) and				
		$4 \times 10^{-5}\gamma$'s (4.43 MeV) per Am decay				
$^{244}_{96}$Cm	18.11 y	α	5.763	24%	Pu L x rays	$\sim 9\%$
			5.805	76%		
$^{252}_{98}$Cf	2.645 y	α (97%)	6.076	15%		
			6.118	82%		
		Fission (3.1%)				
		$\approx 20\ \gamma$'s/fission; 80% < 1 MeV				
		≈ 4 neutrons/fission; $\langle E_n \rangle = 2.14$ MeV				

"Emission probability" is the probability per decay of a given emission; because of cascades these may total more than 100%. Only principal emissions are listed. EC means electron capture, and e^- means monoenergetic internal conversion (Auger) electron. The intensity of 0.511 MeV e^+e^- annihilation photons depends upon the number of stopped positrons. Endpoint $\beta^\pm$ energies are listed. In some cases when energies are closely spaced, the γ-ray values are approximate weighted averages. Radiation from short-lived daughter isotopes is included where relevant.

Half-lives, energies, and intensities are from E. Browne and R.B. Firestone, *Table of Radioactive Isotopes* (John Wiley & Sons, New York, 1986), recent *Nuclear Data Sheets*, and *X-ray and Gamma-ray Standards for Detector Calibration*, IAEA-TECDOC-619 (1991).

Neutron data are from *Neutron Sources for Basic Physics and Applications* (Pergamon Press, 1983).

31. PROBABILITY

Revised May 1996 by D.E. Groom (LBNL) and F. James (CERN), September 1999 by R. Cousins (UCLA), October 2001 and October 2003 by G. Cowan (RHUL).

31.1. General [1–8]

An abstract definition of probability can be given by considering a set S, called the sample space, and possible subsets $A, B, \ldots$, the interpretation of which is left open. The probability P is a real-valued function defined by the following axioms due to Kolmogorov [9]:

1. For every subset A in S, $P(A) \geq 0$.
2. For disjoint subsets (*i.e.*, $A \cap B = \emptyset$), $P(A \cup B) = P(A) + P(B)$.
3. $P(S) = 1$.

In addition one defines the conditional probability $P(A|B)$ (read P of A given B) as

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \ . \tag{31.1}$$

From this definition and using the fact that $A \cap B$ and $B \cap A$ are the same, one obtains *Bayes' theorem*,

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \ . \tag{31.2}$$

From the three axioms of probability and the definition of conditional probability, one obtains the *law of total probability*,

$$P(B) = \sum_i P(B|A_i)P(A_i) \ , \tag{31.3}$$

for any subset B and for disjoint A_i with $\cup_i A_i = S$. This can be combined with Bayes' theorem Eq. (31.2) to give

$$P(A|B) = \frac{P(B|A)P(A)}{\sum_i (B|A_i)P(A_i)} \ , \tag{31.4}$$

where the subset A could, for example, be one of the A_i.

The most commonly used interpretation of the subsets of the sample space are outcomes of a repeatable experiment. The probability $P(A)$ is assigned a value equal to the limiting frequency of occurrence of A. This interpretation forms the basis of *frequentist statistics*.

The subsets of the sample space can also be interpreted as *hypotheses*, *i.e.*, statements that are either true or false, such as 'The mass of the W boson lies between 80.3 and 80.5 GeV'. In the frequency interpretation, such statements are either always or never true, *i.e.*, the corresponding probabilities would be 0 or 1. Using *subjective probability*, however, $P(A)$ is interpreted as the degree of belief that the hypothesis A is true.

Subjective probability is used in *Bayesian* (as opposed to frequentist) statistics. Bayes' theorem can be written

$$P(\text{theory}|\text{data}) \propto P(\text{data}|\text{theory})P(\text{theory}) \ , \tag{31.5}$$

where 'theory' represents some hypothesis and 'data' is the outcome of the experiment. Here $P(\text{theory})$ is the *prior* probability for the theory, which reflects the experimenter's degree of belief before carrying out the measurement, and $P(\text{data}|\text{theory})$ is the probability to have gotten the data actually obtained, given the theory, which is also called the *likelihood*.

Bayesian statistics provides no fundamental rule for obtaining the prior probability; this is necessarily subjective and may depend on previous measurements, theoretical prejudices, etc. Once this has been specified, however, Eq. (31.5) tells how the probability for the theory must be modified in the light of the new data to give the *posterior* probability, $P(\text{theory}|\text{data})$. As Eq. (31.5) is stated as a proportionality, the probability must be normalized by summing (or integrating) over all possible hypotheses.

31.2. Random variables

A *random variable* is a numerical characteristic assigned to an element of the sample space. In the frequency interpretation of probability, it corresponds to an outcome of a repeatable experiment. Let x be a possible outcome of an observation. If x can take on any value from a continuous range, we write $f(x;\theta)dx$ as the probability that the measurement's outcome lies between x and $x + dx$. The function $f(x;\theta)$ is called the *probability density function* (p.d.f.), which may depend on one or more parameters θ. If x can take on only discrete values (*e.g.*, the non-negative integers), then $f(x;\theta)$ is itself a probability.

The p.d.f. is always normalized to unit area (unit sum, if discrete). Both x and θ may have multiple components and are then often written as vectors. If θ is unknown, we may wish to estimate its value from a given set of measurements of x; this is a central topic of *statistics* (see Sec. 32).

The *cumulative distribution function* $F(a)$ is the probability that $x \leq a$:

$$F(a) = \int_{-\infty}^{a} f(x)\, dx \ . \tag{31.6}$$

Here and below, if x is discrete-valued, the integral is replaced by a sum. The endpoint a is expressly included in the integral or sum. Then $0 \leq F(x) \leq 1$, $F(x)$ is nondecreasing, and $P(a < x \leq b) = F(b) - F(a)$. If x is discrete, $F(x)$ is flat except at allowed values of x, where it has discontinuous jumps equal to $f(x)$.

Any function of random variables is itself a random variable, with (in general) a different p.d.f. The *expectation value* of any function $u(x)$ is

$$E[u(x)] = \int_{-\infty}^{\infty} u(x)\, f(x)\, dx \ , \tag{31.7}$$

assuming the integral is finite. For $u(x)$ and $v(x)$ any two functions of x, $E[u+v] = E[u]+E[v]$. For c and k constants, $E[cu+k] = cE[u]+k$.

The n^{th} moment of a random variable is

$$\alpha_n \equiv E[x^n] = \int_{-\infty}^{\infty} x^n f(x)\,dx\ , \tag{31.8a}$$

and the n^{th} central moment of x (or moment about the mean, α_1) is

$$m_n \equiv E[(x-\alpha_1)^n] = \int_{-\infty}^{\infty} (x-\alpha_1)^n f(x)\,dx\ . \tag{31.8b}$$

The most commonly used moments are the mean μ and variance σ^2:

$$\mu \equiv \alpha_1\ , \tag{31.9a}$$
$$\sigma^2 \equiv V[x] \equiv m_2 = \alpha_2 - \mu^2\ . \tag{31.9b}$$

The mean is the location of the "center of mass" of the p.d.f., and the variance is a measure of the square of its width. Note that $V[cx+k] = c^2V[x]$. It is often convenient to use the *standard deviation* of x, σ, defined as the square root of the variance.

Any odd moment about the mean is a measure of the skewness of the p.d.f. The simplest of these is the dimensionless coefficient of skewness $\gamma_1 = m_3/\sigma^3$.

The fourth central moment m_4 provides a convenient measure of the tails of a distribution. For the Gaussian distribution (see Sec. 31.4) one has $m_4 = 3\sigma^4$. The *kurtosis* is defined as $\gamma_2 = m_4/\sigma^4 - 3$, *i.e.*, it is zero for a Gaussian, positive for a *leptokurtic* distribution with longer tails, and negative for a *platykurtic* distribution with tails that die off more quickly than those of a Gaussian.

Besides the mean, another useful indicator of the "middle" of the probability distribution is the *median*, $x_{\rm med}$, defined by $F(x_{\rm med}) = 1/2$, *i.e.*, half the probability lies above and half lies below $x_{\rm med}$. (More rigorously, $x_{\rm med}$ is a median if $P(x \geq x_{\rm med}) \geq 1/2$ and $P(x \leq x_{\rm med}) \geq 1/2$. If only one value exists it is called 'the median'.)

Let x and y be two random variables with a *joint* p.d.f. $f(x,y)$. The *marginal* p.d.f. of x (the distribution of x with y unobserved) is

$$f_1(x) = \int_{-\infty}^{\infty} f(x,y)\,dy\ , \tag{31.10}$$

and similarly for the marginal p.d.f. $f_2(y)$. The *conditional* p.d.f. of y given fixed x (with $f_1(x) \neq 0$) is defined by $f_3(y|x) = f(x,y)/f_1(x)$ and similarly $f_4(x|y) = f(x,y)/f_2(y)$. From these we immediately obtain Bayes' theorem (see Eqs. (31.2) and (31.4)),

$$f_4(x|y) = \frac{f_3(y|x)f_1(x)}{f_2(y)} = \frac{f_3(y|x)f_1(x)}{\int f_3(y|x')f_1(x')\,dx'}\ . \tag{31.11}$$

The mean of x is

$$\mu_x = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\, f(x,y)\, dx\, dy = \int_{-\infty}^{\infty} x\, f_1(x)\, dx\ , \tag{31.12}$$

and similarly for y. The *covariance* of x and y is

$$\mathrm{cov}[x,y] = E[(x-\mu_x)(y-\mu_y)] = E[xy] - \mu_x\mu_y\ . \tag{31.13}$$

A dimensionless measure of the covariance of x and y is given by the *correlation coefficient*,

$$\rho_{xy} = \mathrm{cov}[x,y]/\sigma_x\sigma_y\ , \tag{31.14}$$

where σ_x and σ_y are the standard deviations of x and y. It can be shown that $-1 \le \rho_{xy} \le 1$.

Two random variables x and y are *independent* if and only if

$$f(x,y) = f_1(x) f_2(y)\ . \tag{31.15}$$

If x and y are independent then $\rho_{xy} = 0$; the converse is not necessarily true. If x and y are independent, $E[u(x)v(y)] = E[u(x)]E[v(y)]$, and $V[x+y] = V[x] + V[y]$; otherwise, $V[x+y] = V[x] + V[y] + 2\mathrm{cov}[x,y]$ and $E[uv]$ does not necessarily factorize.

Consider a set of n continuous random variables $\boldsymbol{x} = (x_1, \ldots, x_n)$ with joint p.d.f. $f(\boldsymbol{x})$ and a set of n new variables $\boldsymbol{y} = (y_1, \ldots, y_n)$, related to $\boldsymbol{x}$ by means of a function $\boldsymbol{y}(\boldsymbol{x})$ that is one-to-one, *i.e.*, the inverse $\boldsymbol{x}(\boldsymbol{y})$ exists. The joint p.d.f. for $\boldsymbol{y}$ is given by

$$g(\boldsymbol{y}) = f(\boldsymbol{x}(\boldsymbol{y}))|J|\ \ , \tag{31.16}$$

where $|J|$ is the absolute value of the determinant of the square matrix $J_{ij} = \partial x_i/\partial y_j$ (the Jacobian determinant). If the transformation from $\boldsymbol{x}$ to $\boldsymbol{y}$ is not one-to-one, the $\boldsymbol{x}$-space must be broken in to regions where the function $\boldsymbol{y}(\boldsymbol{x})$ can be inverted and the contributions to $g(\boldsymbol{y})$ from each region summed.

Given a set of functions $\boldsymbol{y} = (y_1, \ldots, y_m)$ with $m < n$, one can construct $n - m$ additional independent functions, apply the procedure above, then integrate the resulting $g(\boldsymbol{y})$ over the unwanted y_i to find the marginal distribution of those of interest.

To change variables for discrete random variables simply substitute; no Jacobian is necessary because now f is a probability rather than a probability density. If f depends on a set of parameters $\boldsymbol{\theta}$, a change to a different parameter set $\boldsymbol{\eta}(\boldsymbol{\theta})$ is made by simple substitution; no Jacobian is used.

31.3. Characteristic functions

The characteristic function $\phi(u)$ associated with the p.d.f. $f(x)$ is essentially its Fourier transform, or the expectation value of e^{iux}:

$$\phi(u) = E\left[e^{iux}\right] = \int_{-\infty}^{\infty} e^{iux} f(x)\, dx \ . \tag{31.17}$$

Once $\phi(u)$ is specified, the p.d.f. $f(x)$ is uniquely determined and vice versa; knowing one is equivalent to the other. Characteristic functions are useful in deriving a number of important results about moments and sums of random variables.

It follows from Eqs. (31.8a) and (31.17) that the n^{th} moment of a random variable x that follows $f(x)$ is given by

$$i^{-n} \left.\frac{d^n \phi}{du^n}\right|_{u=0} = \int_{-\infty}^{\infty} x^n f(x)\, dx = \alpha_n \ . \tag{31.18}$$

Thus it is often easy to calculate all the moments of a distribution defined by $\phi(u)$, even when $f(x)$ cannot be written down explicitly.

If the p.d.f.s $f_1(x)$ and $f_2(y)$ for independent random variables x and y have characteristic functions $\phi_1(u)$ and $\phi_2(u)$, then the characteristic function of the weighted sum $ax + by$ is $\phi_1(au)\phi_2(bu)$. The addition rules for several important distributions (*e.g.*, that the sum of two Gaussian distributed variables also follows a Gaussian distribution) easily follow from this observation.

Let the (partial) characteristic function corresponding to the conditional p.d.f. $f_2(x|z)$ be $\phi_2(u|z)$, and the p.d.f. of z be $f_1(z)$. The characteristic function after integration over the conditional value is

$$\phi(u) = \int \phi_2(u|z) f_1(z)\, dz \ . \tag{31.19}$$

Suppose we can write ϕ_2 in the form

$$\phi_2(u|z) = A(u) e^{ig(u)z} \ . \tag{31.20}$$

Then

$$\phi(u) = A(u)\phi_1(g(u)) \ . \tag{31.21}$$

The semi-invariants κ_n are defined by

$$\phi(u) = \exp\left[\sum_{n=1}^{\infty} \frac{\kappa_n}{n!}(iu)^n\right] = \exp\left(i\kappa_1 u - \tfrac{1}{2}\kappa_2 u^2 + \ldots\right) \ . \tag{31.22}$$

The values κ_n are related to the moments α_n and m_n. The first few relations are

$$\begin{aligned}\kappa_1 &= \alpha_1 \ (= \mu, \text{ the mean}) \\ \kappa_2 &= m_2 \ = \alpha_2 - \alpha_1^2 \ (= \sigma^2, \ \text{ the variance}) \\ \kappa_3 &= m_3 = \alpha_3 - 3\alpha_1\alpha_2 + 2\alpha_1^2 \ . \end{aligned} \tag{31.23}$$

31.4. Some probability distributions

Table 31.1 gives a number of common probability density functions and corresponding characteristic functions, means, and variances. Further information may be found in Refs. [1– 8] and [10]; Ref. [10] has particularly detailed tables. Monte Carlo techniques for generating each of them may be found in our Sec. 33.4. We comment below on all except the trivial uniform distribution.

31.4.1. *Binomial distribution* :

A random process with exactly two possible outcomes which occur with fixed probabilities is called a *Bernoulli* process. If the probability of obtaining a certain outcome (a "success") in each trail is p, then the probability of obtaining exactly r successes ($r = 0, 1, 2, \ldots, N$) in N independent trials, without regard to the order of the successes and failures, is given by the binomial distribution $f(r; N, p)$ in Table 31.1. If r and s are binomially distributed with parameters (N_r, p) and (N_s, p), then $t = r + s$ follows a binomial distribution with parameters $(N_r + N_s, p)$.

31.4.2. *Poisson distribution* :

The Poisson distribution $f(n; \nu)$ gives the probability of finding exactly n events in a given interval of x (*e.g.*, space and time) when the events occur independently of one another and of x at an average rate of ν per the given interval. The variance σ^2 equals ν. It is the limiting case $p \to 0$, $N \to \infty$, $Np = \nu$ of the binomial distribution. The Poisson distribution approaches the Gaussian distribution for large ν.

31.4.3. *Normal or Gaussian distribution* :

The normal (or Gaussian) probability density function $f(x; \mu, \sigma^2)$ given in Table 31.1 has mean $E[x] = \mu$ and variance $V[x] = \sigma^2$. Comparison of the characteristic function $\phi(u)$ given in Table 31.1 with Eq. (31.22) shows that all semi-invariants κ_n beyond κ_2 vanish; this is a unique property of the Gaussian distribution. Some other properties are:

$P(x \text{ in range } \mu \pm \sigma) = 0.6827$,

$P(x \text{ in range } \mu \pm 0.6745\sigma) = 0.5$,

$E[|x - \mu|] = \sqrt{2/\pi}\sigma = 0.7979\sigma$,

half-width at half maximum $= \sqrt{2 \ln 2}\sigma = 1.177\sigma$.

For a Gaussian with $\mu = 0$ and $\sigma^2 = 1$ (the *standard* Gaussian), the cumulative distribution, Eq. (31.6), is related to the error function $\mathrm{erf}(y)$ by

$$F(x; 0, 1) = \tfrac{1}{2}\left[1 + \mathrm{erf}(x/\sqrt{2})\right] \ . \tag{31.24}$$

The error function and standard Gaussian are tabulated in many references (*e.g.*, Ref. [10]) and are available in libraries of computer routines such as CERNLIB. For a mean μ and variance σ^2, replace x by $(x-\mu)/\sigma$. The probability of x in a given range can be calculated with Eq. (32.43).

For x and y independent and normally distributed, $z = ax + by$ follows $f(z; a\mu_x + b\mu_y, a^2\sigma_x^2 + b^2\sigma_y^2)$; that is, the weighted means and variances add.

The Gaussian derives its importance in large part from the *central limit theorem*: If independent random variables $x_1, \ldots, x_n$ are distributed according to *any* p.d.f.s with finite means and variances, then the sum $y = \sum_{i=1}^{n} x_i$ will have a p.d.f. that approaches a Gaussian for large n. The mean and variance are given by the sums of corresponding terms from the individual x_i. Therefore the sum of a large number of fluctuations x_i will be distributed as a Gaussian, even if the x_i themselves are not.

(Note that the *product* of a large number of random variables is not Gaussian, but its logarithm is. The p.d.f. of the product is *log-normal*. See Ref. [8] for details.)

For a set of n Gaussian random variables $\boldsymbol{x}$ with means $\boldsymbol{\mu}$ and corresponding Fourier variables $\boldsymbol{u}$, the characteristic function for a one-dimensional Gaussian is generalized to

$$\phi(\boldsymbol{u}; \boldsymbol{\mu}, V) = \exp\left[i\boldsymbol{\mu}\cdot\boldsymbol{u} - \tfrac{1}{2}\boldsymbol{u}^T V \boldsymbol{u}\right] \ . \tag{31.25}$$

From Eq. (31.18), the covariance of x_i and x_j is

$$E\left[(x_i - \mu_i)(x_j - \mu_j)\right] = V_{ij} \ . \tag{31.26}$$

If the components of $\boldsymbol{x}$ are independent, then $V_{ij} = \delta_{ij}\sigma_i^2$, and Eq. (31.25) is the product of the c.f.s of n Gaussians.

The covariance matrix V can be related to the correlation matrix defined by Eq. (31.14) (a sort of normalized covariance matrix) as $\rho_{ij} = V_{ij}/\sigma_i\sigma_j$. Note that by construction $\rho_{ii} = 1$, since $V_{ii} = \sigma_i^2$.

The characteristic function may be inverted to find the corresponding p.d.f.,

$$f(\boldsymbol{x}; \boldsymbol{\mu}, V) = \frac{1}{(2\pi)^{n/2}\sqrt{|V|}} \exp\left[-\tfrac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T V^{-1} (\boldsymbol{x}-\boldsymbol{\mu})\right] \tag{31.27}$$

where the determinant $|V|$ must be greater than 0. For diagonal V (independent variables), $f(\boldsymbol{x}; \boldsymbol{\mu}, V)$ is the product of the p.d.f.s of n Gaussian distributions.

For $n = 2$, $f(\boldsymbol{x}; \boldsymbol{\mu}, V)$ is

$$f(x_1, x_2;\ \mu_1, \mu_2, \sigma_1, \sigma_2, \rho) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}$$
$$\times\ \exp\left\{\frac{-1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} - \frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]\right\}. \tag{31.28}$$

The marginal distribution of any x_i is a Gaussian with mean μ_i and variance V_{ii}. V is $n \times n$, symmetric, and positive definite. Therefore for any vector $\boldsymbol{X}$, the quadratic form $\boldsymbol{X}^T V^{-1} \boldsymbol{X} = C$, where C is any positive number, traces an n-dimensional ellipsoid as $\boldsymbol{X}$ varies. If $X_i = x_i - \mu_i$, then C is a random variable obeying the χ^2 distribution with n degrees of freedom, discussed in the following section. The probability that $\boldsymbol{X}$ corresponding to a set of Gaussian random variables x_i lies outside the ellipsoid characterized by a given value of C $(= \chi^2)$ is given by $1 - F_{\chi^2}(C; n)$, where F_{χ^2} is the cumulative χ^2 distribution. This may be read from Fig. 32.1. For example, the "s-standard-deviation ellipsoid" occurs at $C = s^2$. For the two-variable case ($n = 2$), the point $\boldsymbol{X}$ lies outside the one-standard-deviation ellipsoid with 61% probability. The use of these ellipsoids as indicators of probable error is described in Sec. 32.3.2.3; the validity of those indicators assumes that $\boldsymbol{\mu}$ and V are correct.

31.4.4. χ^2 *distribution* :

If $x_1, \ldots, x_n$ are independent Gaussian random variables, the sum $z = \sum_{i=1}^{n} (x_i - \mu_i)^2/\sigma_i^2$ follows the χ^2 p.d.f. with n *degrees of freedom*, which we denote by $\chi^2(n)$. Under a linear transformation to n correlated Gaussian variables x_i', the value of z is invariant; then $z = \boldsymbol{X'}^T V^{-1} \boldsymbol{X'}$ as in the previous section. For a set of z_i, each of which follows $\chi^2(n_i)$, $\sum z_i$ follows $\chi^2(\sum n_i)$. For large n, the χ^2 p.d.f. approaches a Gaussian with mean $\mu = n$ and variance $\sigma^2 = 2n$.

The χ^2 p.d.f. is often used in evaluating the level of compatibility between observed data and a hypothesis for the p.d.f. that the data might follow. This is discussed further in Sec. 32.2.2 on tests of goodness-of-fit.

31.4.5. *Student's t distribution* :

Suppose that x and $x_1, \ldots, x_n$ are independent and Gaussian distributed with mean 0 and variance 1. We then define

$$z = \sum_{i=1}^{n} x_i^2 \quad \text{and} \quad t = \frac{x}{\sqrt{z/n}}\,. \tag{31.29}$$

The variable z thus follows a $\chi^2(n)$ distribution. Then t is distributed according to Student's t distribution with n degrees of freedom, $f(t; n)$, given in Table 31.1.

The Student's t distribution resembles a Gaussian with wide tails. As $n \to \infty$, the distribution approaches a Gaussian. If $n = 1$, it is a *Cauchy* or *Breit–Wigner* distribution. The mean is finite only for $n > 1$ and the variance is finite only for $n > 2$, so the central limit theorem is not applicable to sums of random variables following the t distribution for $n = 1$ or 2.

As an example, consider the *sample mean* $\overline{x} = \sum x_i/n$ and the *sample variance* $s^2 = \sum(x_i - \overline{x})^2/(n-1)$ for normally distributed x_i with unknown mean μ and variance σ^2. The sample mean has a Gaussian distribution with a variance σ^2/n, so the variable $(\overline{x} - \mu)/\sqrt{\sigma^2/n}$ is normal with mean 0 and variance 1. Similarly, $(n-1)s^2/\sigma^2$ is independent of this and follows $\chi^2(n-1)$. The ratio

$$t = \frac{(\overline{x} - \mu)/\sqrt{\sigma^2/n}}{\sqrt{(n-1)s^2/\sigma^2(n-1)}} = \frac{\overline{x} - \mu}{\sqrt{s^2/n}} \tag{31.30}$$

is distributed as $f(t; n-1)$. The unknown variance σ^2 cancels, and t can be used to test the probability that the true mean is some particular value μ.

In Table 31.1, n in $f(t; n)$ is not required to be an integer. A Student's t distribution with non-integral $n > 0$ is useful in certain applications.

31.4.6. *Gamma distribution* :

For a process that generates events as a function of x (*e.g.*, space or time) according to a Poisson distribution, the distance in x from an arbitrary starting point (which may be some particular event) to the k^{th} event follows a *gamma* distribution, $f(x; \lambda, k)$. The Poisson parameter μ is λ per unit x. The special case $k = 1$ (*i.e.*, $f(x; \lambda, 1) = \lambda e^{-\lambda x}$) is called the *exponential* distribution. A sum of k' exponential random variables x_i is distributed as $f(\sum x_i; \lambda, k')$.

The parameter k is not required to be an integer. For $\lambda = 1/2$ and $k = n/2$, the gamma distribution reduces to the $\chi^2(n)$ distribution.

Table 31.1. Some common probability density functions, with corresponding characteristic functions and means and variances. In the Table, $\Gamma(k)$ is the gamma function, equal to $(k-1)!$ when k is an integer.

Distribution	Probability density function f (variable; parameters)	Characteristic function $\phi(u)$	Mean	Variance σ^2
Uniform	$f(x;a,b) = \begin{cases} 1/(b-a) & a \le x \le b \\ 0 & \text{otherwise} \end{cases}$	$\dfrac{e^{ibu} - e^{iau}}{(b-a)iu}$	$\dfrac{a+b}{2}$	$\dfrac{(b-a)^2}{12}$
Binomial	$f(r;N,p) = \dfrac{N!}{r!(N-r)!}\ p^r q^{N-r}$ $r = 0, 1, 2, \ldots, N\ ;\quad 0 \le p \le 1\ ;\quad q = 1-p$	$(q + pe^{iu})^N$	Np	Npq
Poisson	$f(n;\nu) = \dfrac{\nu^n e^{-\nu}}{n!}\ ;\quad n = 0, 1, 2, \ldots\ ;\quad \nu > 0$	$\exp[\nu(e^{iu} - 1)]$	ν	ν
Normal (Gaussian)	$f(x;\mu,\sigma^2) = \dfrac{1}{\sigma\sqrt{2\pi}}\ \exp(-(x-\mu)^2/2\sigma^2)$ $-\infty < x < \infty\ ;\quad -\infty < \mu < \infty\ ;\quad \sigma > 0$	$\exp(i\mu u - \frac{1}{2}\sigma^2 u^2)$	μ	σ^2
Multivariate Gaussian	$f(\boldsymbol{x};\boldsymbol{\mu},V) = \dfrac{1}{(2\pi)^{n/2}\sqrt{\lvert V\rvert}}$ $\times \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T V^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right]$ $-\infty < x_j < \infty;\quad -\infty < \mu_j < \infty;\quad \det V > 0$	$\exp\left[i\boldsymbol{\mu}\cdot\boldsymbol{u} - \frac{1}{2}\boldsymbol{u}^T V \boldsymbol{u}\right]$	$\boldsymbol{\mu}$	V_{jk}
χ^2	$f(z;n) = \dfrac{z^{n/2-1}e^{-z/2}}{2^{n/2}\Gamma(n/2)}\ ;\quad z \ge 0$	$(1-2iu)^{-n/2}$	n	$2n$
Student's t	$f(t;n) = \dfrac{1}{\sqrt{n\pi}}\ \dfrac{\Gamma[(n+1)/2]}{\Gamma(n/2)}\left(1 + \dfrac{t^2}{n}\right)^{-(n+1)/2}$ $-\infty < t < \infty\ ;\quad n$ not required to be integer	—	0 for $n \ge 2$	$n/(n-2)$ for $n \ge 3$
Gamma	$f(x;\lambda,k) = \dfrac{x^{k-1}\lambda^k e^{-\lambda x}}{\Gamma(k)}\ ;\quad 0 < x < \infty\ ;$ k not required to be integer	$(1 - iu/\lambda)^{-k}$	k/λ	k/λ^2

References:

1. H. Cramér, *Mathematical Methods of Statistics,* (Princeton Univ. Press, New Jersey, 1958).
2. A. Stuart and A.K. Ord, *Kendall's Advanced Theory of Statistics*, Vol. 1 *Distribution Theory* 5th Ed., (Oxford Univ. Press, New York, 1987), and earlier editions by Kendall and Stuart.
3. W.T. Eadie, D. Drijard, F.E. James, M. Roos, and B. Sadoulet, *Statistical Methods in Experimental Physics* (North Holland, Amsterdam, and London, 1971).
4. L. Lyons, *Statistics for Nuclear and Particle Physicists* (Cambridge University Press, New York, 1986).
5. B.R. Roe, *Probability and Statistics in Experimental Physics,* 2nd Ed., (Springer, New York, 2001).
6. R.J. Barlow, *Statistics: A Guide to the Use of Statistical Methods in the Physical Sciences* (John Wiley, New York, 1989).
7. S. Brandt, *Data Analysis*, 3rd Ed., (Springer, New York, 1999).
8. G. Cowan, *Statistical Data Analysis* (Oxford University Press, Oxford, 1998).
9. A.N. Kolmogorov, *Grundbegriffe der Wahrscheinlichkeitsrechnung* (Springer, Berlin 1933); *Foundations of the Theory of Probability*, 2nd Ed., (Chelsea, New York 1956).
10. M. Abramowitz and I. Stegun, eds., *Handbook of Mathematical Functions* (Dover, New York, 1972).

32. STATISTICS

Revised April 1998 by F. James (CERN); February 2000 by R. Cousins (UCLA); October 2001, October 2003, and August 2005 by G. Cowan (RHUL).

This chapter gives an overview of statistical methods used in High Energy Physics. In statistics we are interested in using a given sample of data to make inferences about a probabilistic model, *e.g.*, to assess the model's validity or to determine the values of its parameters. There are two main approaches to statistical inference, which we may call frequentist and Bayesian. In frequentist statistics, probability is interpreted as the frequency of the outcome of a repeatable experiment. The most important tools in this framework are parameter estimation, covered in Section 32.1, and statistical tests, discussed in Section 32.2. Frequentist confidence intervals, which are constructed so as to cover the true value of a parameter with a specified probability, are treated in Section 32.3.2. Note that in frequentist statistics one does not define a probability for a hypothesis or for a parameter.

Frequentist statistics provides the usual tools for reporting objectively the outcome of an experiment without needing to incorporate prior beliefs concerning the parameter being measured or the theory being tested. As such they are used for reporting essentially all measurements and their statistical uncertainties in High Energy Physics.

In Bayesian statistics, the interpretation of probability is more general and includes *degree of belief*. One can then speak of a probability density function (p.d.f.) for a parameter, which expresses one's state of knowledge about where its true value lies. Bayesian methods allow for a natural way to input additional information such as physical boundaries and subjective information; in fact they *require* as input the *prior* p.d.f. for the parameters, *i.e.*, the degree of belief about the parameters' values before carrying out the measurement. Using Bayes' theorem Eq. (31.4), the prior degree of belief is updated by the data from the experiment. Bayesian methods for interval estimation are discussed in Sections 32.3.1 and 32.3.2.5

Bayesian techniques are often used to treat systematic uncertainties, where the author's subjective beliefs about, say, the accuracy of the measuring device may enter. Bayesian statistics also provides a useful framework for discussing the validity of different theoretical interpretations of the data. This aspect of a measurement, however, will usually be treated separately from the reporting of the result.

For many inference problems, the frequentist and Bayesian approaches give the same numerical answers, even though they are based on fundamentally different interpretations of probability. For small data samples, however, and for measurements of a parameter near a physical boundary, the different approaches may yield different results, so we are forced to make a choice. For a discussion of Bayesian vs. non-Bayesian methods, see References written by a statistician[1], by a physicist[2], or the more detailed comparison in Ref. [3].

Following common usage in physics, the word "error" is often used in this chapter to mean "uncertainty". More specifically it can indicate the size of an interval as in "the standard error" or "error propagation", where the term refers to the standard deviation of an estimator.

32.1. Parameter estimation

Here we review *point estimation* of parameters. An *estimator* $\widehat{\theta}$ (written with a hat) is a function of the data whose value, the *estimate*, is intended as a meaningful guess for the value of the parameter θ.

There is no fundamental rule dictating how an estimator must be constructed. One tries therefore to choose that estimator which has the best properties. The most important of these are (a) *consistency*, (b) *bias*, (c) *efficiency*, and (d) *robustness*.

(a) An estimator is said to be *consistent* if the estimate $\widehat{\theta}$ converges to the true value θ as the amount of data increases. This property is so important that it is possessed by all commonly used estimators.

(b) The *bias*, $b = E[\widehat{\theta}] - \theta$, is the difference between the expectation value of the estimator and the true value of the parameter. The expectation value is taken over a hypothetical set of similar experiments in which $\widehat{\theta}$ is constructed in the same way. When $b = 0$ the estimator is said to be unbiased. The bias depends on the chosen metric, *i.e.*, if $\widehat{\theta}$ is an unbiased estimator of θ, then $\widehat{\theta}^2$ is not in general an unbiased estimator for θ^2. If we have an estimate $\widehat{b}$ for the bias we can subtract it from $\widehat{\theta}$ to obtain a new $\widehat{\theta}' = \widehat{\theta} - \widehat{b}$. The estimate $\widehat{b}$ may, however, be subject to statistical or systematic uncertainties that are larger than the bias itself, so that the new estimator may not be better than the original.

(c) *Efficiency* is the inverse of the ratio of the variance $V[\widehat{\theta}]$ to its minimum possible value. Under rather general conditions, the minimum variance is given by the Rao-Cramér-Frechet bound,

$$\sigma^2_{\rm min} = \left(1 + \frac{\partial b}{\partial \theta}\right)^2 / I(\theta) \,, \tag{32.1}$$

where

$$I(\theta) = E\left[\left(\frac{\partial}{\partial \theta} \sum_i \ln f(x_i;\theta)\right)^2\right] \tag{32.2}$$

is the *Fisher information*. The sum is over all data, assumed independent and distributed according to the p.d.f. $f(x;\theta)$, b is the bias, if any, and the allowed range of x must not depend on θ.

The *mean-squared error*,

$$\text{MSE} = E[(\widehat{\theta} - \theta)^2] = V[\widehat{\theta}] + b^2 \,, \tag{32.3}$$

is a convenient quantity which combines the uncertainties in an estimate due to bias and variance.

(d) *Robustness* is the property of being insensitive to departures from assumptions in the p.d.f. owing to factors such as noise.

For some common estimators the properties above are known exactly. More generally, it is possible to evaluate them by Monte Carlo simulation. Note that they will often depend on the unknown θ.

32.1.1. ***Estimators for mean, variance and median*** :

Suppose we have a set of N independent measurements x_i assumed to be unbiased measurements of the same unknown quantity μ with a common, but unknown, variance σ^2. Then

$$\widehat{\mu} = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{32.4}$$

$$\widehat{\sigma^2} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \widehat{\mu})^2 \tag{32.5}$$

are unbiased estimators of μ and σ^2. The variance of $\widehat{\mu}$ is σ^2/N and the variance of $\widehat{\sigma^2}$ is

$$V\left[\widehat{\sigma^2}\right] = \frac{1}{N}\left(m_4 - \frac{N-3}{N-1}\sigma^4\right) , \tag{32.6}$$

where m_4 is the 4th central moment of x. For Gaussian distributed x_i this becomes $2\sigma^4/(N-1)$ for any $N \geq 2$, and for large N the standard deviation of $\widehat{\sigma}$ (the "error of the error") is $\sigma/\sqrt{2N}$. Again if the x_i are Gaussian, $\widehat{\mu}$ is an efficient estimator for μ and the estimators $\widehat{\mu}$ and $\widehat{\sigma^2}$ are uncorrelated. Otherwise the arithmetic mean (32.4) is not necessarily the most efficient estimator; this is discussed in more detail in [4] Sec. 8.7

If σ^2 is known, it does not improve the estimate $\widehat{\mu}$, as can be seen from Eq. (32.4); however, if μ is known, substitute it for $\widehat{\mu}$ in Eq. (32.5) and replace $N-1$ by N to obtain a somewhat better estimator of σ^2. If the x_i have different, known variances σ_i^2, then the weighted average

$$\widehat{\mu} = \frac{1}{w}\sum_{i=1}^{N} w_i x_i \tag{32.7}$$

is an unbiased estimator for μ with a smaller variance than an unweighted average; here $w_i = 1/\sigma_i^2$ and $w = \sum_i w_i$. The standard deviation of $\widehat{\mu}$ is $1/\sqrt{w}$.

As an estimator for the median $x_{\rm med}$ one can use the value $\widehat{x}_{\rm med}$ such that half the x_i are below and half above (the sample median). If the sample median lies between

two observed values, it is set by convention halfway between them. If the p.d.f. of x has the form $f(x-\mu)$ and μ is both mean and median, then for large N the variance of the sample median approaches $1/[4Nf^2(0)]$, provided $f(0) > 0$. Although estimating the median can often be more difficult computationally than the mean, the resulting estimator is generally more robust, as it is insensitive to the exact shape of the tails of a distribution.

32.1.2. *The method of maximum likelihood* :

"From a theoretical point of view, the most important general method of estimation so far known is the *method of maximum likelihood*" [5]. We suppose that a set of N independently measured quantities x_i came from a p.d.f. $f(x;\boldsymbol{\theta})$, where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)$ is set of n parameters whose values are unknown. The method of maximum likelihood takes the estimators $\widehat{\boldsymbol{\theta}}$ to be those values of $\boldsymbol{\theta}$ that maximize the *likelihood function*,

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{N} f(x_i;\boldsymbol{\theta}) \; . \tag{32.8}$$

The likelihood function is the joint p.d.f. for the data, evaluated with the data obtained in the experiment and regarded as a function of the parameters. Note that the likelihood function is *not* a p.d.f. for the parameters $\boldsymbol{\theta}$; in frequentist statistics this is not defined. In Bayesian statistics one can obtain from the likelihood the posterior p.d.f. for $\boldsymbol{\theta}$, but this requires multiplying by a prior p.d.f. (see Sec. 32.3.1).

It is usually easier to work with $\ln L$, and since both are maximized for the same parameter values $\boldsymbol{\theta}$, the maximum likelihood (ML) estimators can be found by solving the *likelihood equations*,

$$\frac{\partial \ln L}{\partial \theta_i} = 0 \; , \qquad i = 1, \ldots, n \; . \tag{32.9}$$

Maximum likelihood estimators are important because they are approximately unbiased and efficient for large data samples, under quite general conditions, and the method has a wide range of applicability.

In evaluating the likelihood function, it is important that any normalization factors in the p.d.f. that involve $\boldsymbol{\theta}$ be included. However, we will only be interested in the maximum of L and in ratios of L at different values of the parameters; hence any multiplicative factors that do not involve the parameters that we want to estimate may be dropped, including factors that depend on the data but not on $\boldsymbol{\theta}$.

Under a one-to-one change of parameters from $\boldsymbol{\theta}$ to $\boldsymbol{\eta}$, the ML estimators $\widehat{\boldsymbol{\theta}}$ transform to $\boldsymbol{\eta}(\widehat{\boldsymbol{\theta}})$. That is, the ML solution is invariant under change of parameter. However, other properties of ML estimators, in particular the bias, are not invariant under change of parameter.

The inverse V^{-1} of the covariance matrix $V_{ij} = \mathrm{cov}[\widehat{\theta}_i, \widehat{\theta}_j]$ for a set of ML estimators can be estimated by using

$$(\widehat{V}^{-1})_{ij} = -\left.\frac{\partial^2 \ln L}{\partial\theta_i \partial\theta_j}\right|_{\widehat{\theta}} \ . \tag{32.10}$$

For finite samples, however, Eq. (32.10) can result in an underestimate of the variances. In the large sample limit (or in a linear model with Gaussian errors), L has a Gaussian form and $\ln L$ is (hyper)parabolic. In this case it can be seen that a numerically equivalent way of determining s-standard-deviation errors is from the contour given by the $\boldsymbol{\theta'}$ such that

$$\ln L(\boldsymbol{\theta'}) = \ln L_{\max} - s^2/2 \ , \tag{32.11}$$

where $\ln L_{\max}$ is the value of $\ln L$ at the solution point (compare with Eq. (32.46)). The extreme limits of this contour on the θ_i axis give an approximate s-standard-deviation confidence interval for θ_i (see Section 32.3.2.3).

In the case where the size n of the data sample $x_1, \ldots, x_n$ is small, the unbinned maximum likelihood method, i.e., use of equation (32.8), is preferred since binning can only result in a loss of information and hence larger statistical errors for the parameter estimates. The sample size n can be regarded as fixed or the user can choose to treat it as a Poisson-distributed variable; this latter option is sometimes called "extended maximum likelihood" (see, *e.g.*, [6, 7, 8]). If the sample is large it can be convenient to bin the values in a histogram, so that one obtains a vector of data $\boldsymbol{n} = (n_1, \ldots, n_N)$ with expectation values $\boldsymbol{\nu} = E[\boldsymbol{n}]$ and probabilities $f(\boldsymbol{n}; \boldsymbol{\nu})$. Then one may maximize the likelihood function based on the contents of the bins (so i labels bins). This is equivalent to maximizing the likelihood ratio $\lambda(\boldsymbol{\theta}) = f(\boldsymbol{n}; \boldsymbol{\nu}(\boldsymbol{\theta}))/f(\boldsymbol{n}; \boldsymbol{n})$, or to minimizing the quantity [9]

$$-2 \ln \lambda(\boldsymbol{\theta}) = 2 \sum_{i=1}^{N} \left[\nu_i(\boldsymbol{\theta}) - n_i + n_i \ln \frac{n_i}{\nu_i(\boldsymbol{\theta})} \right] \ , \tag{32.12}$$

where in bins where $n_i = 0$, the last term in (32.12) is zero. In the limit of zero bin width, maximizing (32.12) is equivalent to maximizing the unbinned likelihood function (32.8).

A benefit of binning is that it allows for a goodness-of-fit test (see Sec. 32.2.2). The minimum of $-2 \ln \lambda$ as defined by Eq. (32.12) follows a χ^2 distribution in the large sample limit. If there are N bins and m fitted parameters, then the number of degrees of freedom for the χ^2 distribution is $N - m - 1$ if the data are treated as multinomially distributed and $N - m$ if the n_i are Poisson variables with $\nu_{\rm tot} = \sum_i \nu_i$ fixed. If the n_i are Poisson distributed and $\nu_{\rm tot}$ is also fitted, then by minimizing Eq. (32.12) one obtains that the area under the fitted function is equal to the sum of the histogram contents, *i.e.*, $\sum_i \nu_i = \sum_i n_i$. This is not the case for parameter estimation methods based on a least-squares procedure with traditional weights (see, *e.g.*, Ref. [8]).

32.1.3. *The method of least squares* :

The *method of least squares* (LS) coincides with the method of maximum likelihood in the following special case. Consider a set of N independent measurements y_i at known points x_i. The measurement y_i is assumed to be Gaussian distributed with mean $F(x_i;\boldsymbol{\theta})$ and known variance σ_i^2. The goal is to construct estimators for the unknown parameters $\boldsymbol{\theta}$. The likelihood function contains the sum of squares

$$\chi^2(\boldsymbol{\theta}) = -2\ln L(\boldsymbol{\theta}) + \text{ constant } = \sum_{i=1}^{N} \frac{(y_i - F(x_i;\boldsymbol{\theta}))^2}{\sigma_i^2} \ . \tag{32.13}$$

The set of parameters $\boldsymbol{\theta}$ which maximize L is the same as those which minimize χ^2.

The minimum of Equation (32.13) defines the least-squares estimators $\widehat{\boldsymbol{\theta}}$ for the more general case where the y_i are not Gaussian distributed as long as they are independent. If they are not independent but rather have a covariance matrix $V_{ij} = \mathrm{cov}[y_i, y_j]$, then the LS estimators are determined by the minimum of

$$\chi^2(\boldsymbol{\theta}) = (\boldsymbol{y} - \boldsymbol{F}(\boldsymbol{\theta}))^T V^{-1} (\boldsymbol{y} - \boldsymbol{F}(\boldsymbol{\theta})) \ , \tag{32.14}$$

where $\boldsymbol{y} = (y_1, \ldots, y_N)$ is the vector of measurements, $\boldsymbol{F}(\boldsymbol{\theta})$ is the corresponding vector of predicted values (understood as a column vector in (32.14)), and the superscript T denotes transposed (*i.e.*, row) vector.

In many practical cases one further restricts the problem to the situation where $F(x_i;\boldsymbol{\theta})$ is a linear function of the parameters, *i.e.*,

$$F(x_i;\boldsymbol{\theta}) = \sum_{j=1}^{m} \theta_j h_j(x_i) \ . \tag{32.15}$$

Here the $h_j(x)$ are m linearly independent functions, *e.g.*, $1, x, x^2, \ldots, x^{m-1}$, or Legendre polynomials. We require $m < N$ and at least m of the x_i must be distinct.

Minimizing χ^2 in this case with m parameters reduces to solving a system of m linear equations. Defining $H_{ij} = h_j(x_i)$ and minimizing χ^2 by setting its derivatives with respect to the θ_i equal to zero gives the LS estimators,

$$\widehat{\boldsymbol{\theta}} = (H^T V^{-1} H)^{-1} H^T V^{-1} \boldsymbol{y} \equiv D\boldsymbol{y} \ . \tag{32.16}$$

The covariance matrix for the estimators $U_{ij} = \mathrm{cov}[\widehat{\theta}_i, \widehat{\theta}_j]$ is given by

$$U = DVD^T = (H^T V^{-1} H)^{-1} \ , \tag{32.17}$$

or equivalently, its inverse U^{-1} can be found from

$$(U^{-1})_{ij} = \frac{1}{2} \frac{\partial^2 \chi^2}{\partial\theta_i \partial\theta_j}\bigg|_{\theta=\widehat{\theta}} = \sum_{k,l=1}^{N} h_i(x_k)(V^{-1})_{kl} h_j(x_l) \ . \tag{32.18}$$

The LS estimators can also be found from the expression

$$\widehat{\boldsymbol{\theta}} = U\boldsymbol{g} \ , \tag{32.19}$$

where the vector $\boldsymbol{g}$ is defined by

$$g_i = \sum_{j,k=1}^{N} y_j h_i(x_k)(V^{-1})_{jk} \ . \tag{32.20}$$

For the case of uncorrelated y_i, for example, one can use (32.19) with

$$(U^{-1})_{ij} = \sum_{k=1}^{N} \frac{h_i(x_k)h_j(x_k)}{\sigma_k^2} \ , \tag{32.21}$$

$$g_i = \sum_{k=1}^{N} \frac{y_k h_i(x_k)}{\sigma_k^2} \ . \tag{32.22}$$

Expanding $\chi^2(\boldsymbol{\theta})$ about $\widehat{\boldsymbol{\theta}}$, one finds that the contour in parameter space defined by

$$\chi^2(\boldsymbol{\theta}) = \chi^2(\widehat{\boldsymbol{\theta}}) + 1 = \chi^2_{\rm min} + 1 \tag{32.23}$$

has tangent planes located at plus or minus one standard deviation $\sigma_{\widehat{\theta}}$ from the LS estimates $\widehat{\boldsymbol{\theta}}$.

In constructing the quantity $\chi^2(\boldsymbol{\theta})$, one requires the variances or, in the case of correlated measurements, the covariance matrix. Often these quantities are not known *a priori* and must be estimated from the data; an important example is where the measured value y_i represents a counted number of events in the bin of a histogram. If, for example, y_i represents a Poisson variable, for which the variance is equal to the mean, then one can either estimate the variance from the predicted value, $F(x_i;\boldsymbol{\theta})$, or from the observed number itself, y_i. In the first option, the variances become functions of the fitted parameters, which may lead to calculational difficulties. The second option can be undefined if y_i is zero, and in both cases for small y_i the variance will be poorly estimated. In either case one should constrain the normalization of the fitted curve to the correct value, *e.g.*, one should determine the area under the fitted curve directly from the number of entries in the histogram (see [8] Section 7.4). A further alternative is to use the method of maximum likelihood; for binned data this can be done by minimizing Eq. (32.12)

As the minimum value of the χ^2 represents the level of agreement between the measurements and the fitted function, it can be used for assessing the goodness-of-fit; this is discussed further in Section 32.2.2.

32.1.4. *Propagation of errors* :

Consider a set of n quantities $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)$ and a set of m functions $\boldsymbol{\eta}(\boldsymbol{\theta}) = (\eta_1(\boldsymbol{\theta}), \ldots, \eta_m(\boldsymbol{\theta}))$. Suppose we have estimates $\widehat{\boldsymbol{\theta}} = (\widehat{\theta}_1, \ldots, \widehat{\theta}_n)$, using, say, maximum likelihood or least squares, and we also know or have estimated the covariance matrix $V_{ij} = \mathrm{cov}[\widehat{\theta}_i, \widehat{\theta}_j]$. The goal of *error propagation* is to determine the covariance matrix for the functions, $U_{ij} = \mathrm{cov}[\widehat{\eta}_i, \widehat{\eta}_j]$, where $\widehat{\boldsymbol{\eta}} = \boldsymbol{\eta}(\widehat{\boldsymbol{\theta}})$. In particular, the diagonal elements $U_{ii} = V[\widehat{\eta}_i]$ give the variances. The new covariance matrix can be found by expanding the functions $\boldsymbol{\eta}(\boldsymbol{\theta})$ about the estimates $\widehat{\boldsymbol{\theta}}$ to first order in a Taylor series. Using this one finds

$$U_{ij} \approx \sum_{k,l} \left.\frac{\partial \eta_i}{\partial \theta_k} \frac{\partial \eta_j}{\partial \theta_l}\right|_{\widehat{\theta}} V_{kl} \; . \tag{32.24}$$

This can be written in matrix notation as $U \approx AVA^T$ where the matrix of derivatives A is

$$A_{ij} = \left.\frac{\partial \eta_i}{\partial \theta_j}\right|_{\widehat{\theta}} \tag{32.25}$$

and A^T is its transpose. The approximation is exact if $\boldsymbol{\eta}(\boldsymbol{\theta})$ is linear (it holds, for example, in equation (32.17)). If this is not the case the approximation can break down if, for example, $\boldsymbol{\eta}(\boldsymbol{\theta})$ is significantly nonlinear close to $\widehat{\boldsymbol{\theta}}$ in a region of a size comparable to the standard deviations of $\widehat{\boldsymbol{\theta}}$.

32.2. Statistical tests

In addition to estimating parameters, one often wants to assess the validity of certain statements concerning the data's underlying distribution. *Hypothesis tests* provide a rule for accepting or rejecting hypotheses depending on the outcome of a measurement. In *goodness-of-fit tests* one gives the probability to obtain a level of incompatibility with a certain hypothesis that is greater than or equal to the level observed with the actual data.

32.2.1. *Hypothesis tests* :

Consider an experiment whose outcome is characterized by a vector of data $\boldsymbol{x}$. A *hypothesis* is a statement about the distribution of $\boldsymbol{x}$. It could, for example, define completely the p.d.f. for the data (a simple hypothesis) or it could specify only the functional form of the p.d.f., with the values of one or more parameters left open (a composite hypothesis).

A *statistical test* is a rule that states for which values of $\boldsymbol{x}$ a given hypothesis (often called the null hypothesis, H_0) should be rejected. This is done by defining a region of $\boldsymbol{x}$-space called the critical region; if the outcome if the experiment lands in this region, H_0 is rejected. Equivalently one can say that the hypothesis is accepted if $\boldsymbol{x}$ is observed in the acceptance region, *i.e.*, the complement of the critical region. Here 'accepted' is understood to mean simply that the test did not reject H_0.

Rejecting H_0 if it is true is called an error of the first kind. The probability for this to occur is called the *significance level* of the test, α, which is often chosen to be equal to some pre-specified value. It can also happen that H_0 is false and the true hypothesis is given by some alternative, H_1. If H_0 is accepted in such a case, this is called an error of the second kind. The probability for this to occur, β, depends on the alternative hypothesis, say, H_1, and $1-\beta$ is called the *power* of the test to reject H_1.

In High Energy Physics the components of $\boldsymbol{x}$ might represent the measured properties of candidate events, and the acceptance region is defined by the cuts that one imposes in order to select events of a certain desired type. That is, H_0 could represent the signal hypothesis, and various alternatives, H_1, H_2, etc., could represent background processes.

Often rather than using the full data sample $\boldsymbol{x}$ it is convenient to define a *test statistic*, t, which can be a single number or in any case a vector with fewer components than $\boldsymbol{x}$. Each hypothesis for the distribution of $\boldsymbol{x}$ will determine a distribution for t, and the acceptance region in $\boldsymbol{x}$-space will correspond to a specific range of values of t. In constructing t one attempts to reduce the volume of data without losing the ability to discriminate between different hypotheses.

In particle physics terminology, the probability to accept the signal hypothesis, H_0, is the selection efficiency, *i.e.*, one minus the significance level. The efficiencies for the various background processes are given by one minus the power. Often one tries to construct a test to minimize the background efficiency for a given signal efficiency. The *Neyman–Pearson lemma* states that this is done by defining the acceptance region such that, for $\boldsymbol{x}$ in that region, the ratio of p.d.f.s for the hypotheses H_0 and H_1,

$$\lambda(\boldsymbol{x}) = \frac{f(\boldsymbol{x}|H_0)}{f(\boldsymbol{x}|H_1)} \,, \tag{32.26}$$

is greater than a given constant, the value of which is chosen to give the desired signal efficiency. This is equivalent to the statement that (32.26) represents the test statistic with which one may obtain the highest purity sample for a given signal efficiency. It can be difficult in practice, however, to determine $\lambda(\boldsymbol{x})$, since this requires knowledge of the joint p.d.f.s $f(\boldsymbol{x}|H_0)$ and $f(\boldsymbol{x}|H_1)$. Instead, test statistics based on *neural networks* or *Fisher discriminants* are often used (see [10]).

32.2.2. *Goodness-of-fit tests* :

Often one wants to quantify the level of agreement between the data and a hypothesis without explicit reference to alternative hypotheses. This can be done by defining a *goodness-of-fit statistic*, t, which is a function of the data whose value reflects in some way the level of agreement between the data and the hypothesis. The user must decide what values of the statistic correspond to better or worse levels of agreement with the hypothesis in question; for many goodness-of-fit statistics there is an obvious choice.

The hypothesis in question, say, H_0, will determine the p.d.f. $g(t|H_0)$ for the statistic. The goodness-of-fit is quantified by giving the p-value, defined as the probability to find t in the region of equal or lesser compatibility with H_0 than the level of compatibility

observed with the actual data. For example, if t is defined such that large values correspond to poor agreement with the hypothesis, then the p-value would be

$$p = \int_{t_{\rm obs}}^{\infty} g(t|H_0)\, dt \ , \tag{32.27}$$

where $t_{\rm obs}$ is the value of the statistic obtained in the actual experiment. The p-value should not be confused with the significance level of a test or the confidence level of a confidence interval (Section 32.3), both of which are pre-specified constants.

The p-value is a function of the data and is therefore itself a random variable. If the hypothesis used to compute the p-value is true, then for continuous data, p will be uniformly distributed between zero and one. Note that the p-value is not the probability for the hypothesis; in frequentist statistics this is not defined. Rather, the p-value is the probability, under the assumption of a hypothesis H_0, of obtaining data at least as incompatible with H_0 as the data actually observed.

When estimating parameters using the method of least squares, one obtains the minimum value of the quantity χ^2 (32.13), which can be used as a goodness-of-fit statistic. It may also happen that no parameters are estimated from the data, but that one simply wants to compare a histogram, *e.g.*, a vector of Poisson distributed numbers $\boldsymbol{n} = (n_1, \ldots, n_N)$, with a hypothesis for their expectation values $\nu_i = E[n_i]$. As the distribution is Poisson with variances $\sigma_i^2 = \nu_i$, the χ^2 (32.13) becomes *Pearson's* χ^2 *statistic*,

$$\chi^2 = \sum_{i=1}^{N} \frac{(n_i - \nu_i)^2}{\nu_i} \ . \tag{32.28}$$

If the hypothesis $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_N)$ is correct and if the measured values n_i in (32.28) are sufficiently large (in practice, this will be a good approximation if all $n_i > 5$), then the χ^2 statistic will follow the χ^2 p.d.f. with the number of degrees of freedom equal to the number of measurements N minus the number of fitted parameters. The same holds for the minimized χ^2 from Eq. (32.13) if the y_i are Gaussian.

Alternatively one may fit parameters and evaluate goodness-of-fit by minimizing $-2\ln\lambda$ from Eq. (32.12). One finds that the distribution of this statistic approaches the asymptotic limit faster than does Pearson's χ^2 and thus computing the p-value with the χ^2 p.d.f. will in general be better justified (see [9] and references therein).

Assuming the goodness-of-fit statistic follows a χ^2 p.d.f., the p-value for the hypothesis is then

$$p = \int_{\chi^2}^{\infty} f(z; n_{\rm d})\, dz \ , \tag{32.29}$$

where $f(z; n_{\rm d})$ is the χ^2 p.d.f. and $n_{\rm d}$ is the appropriate number of degrees of freedom. Values can be obtained from Fig. 32.1 or from the CERNLIB routine `PROB`. If the

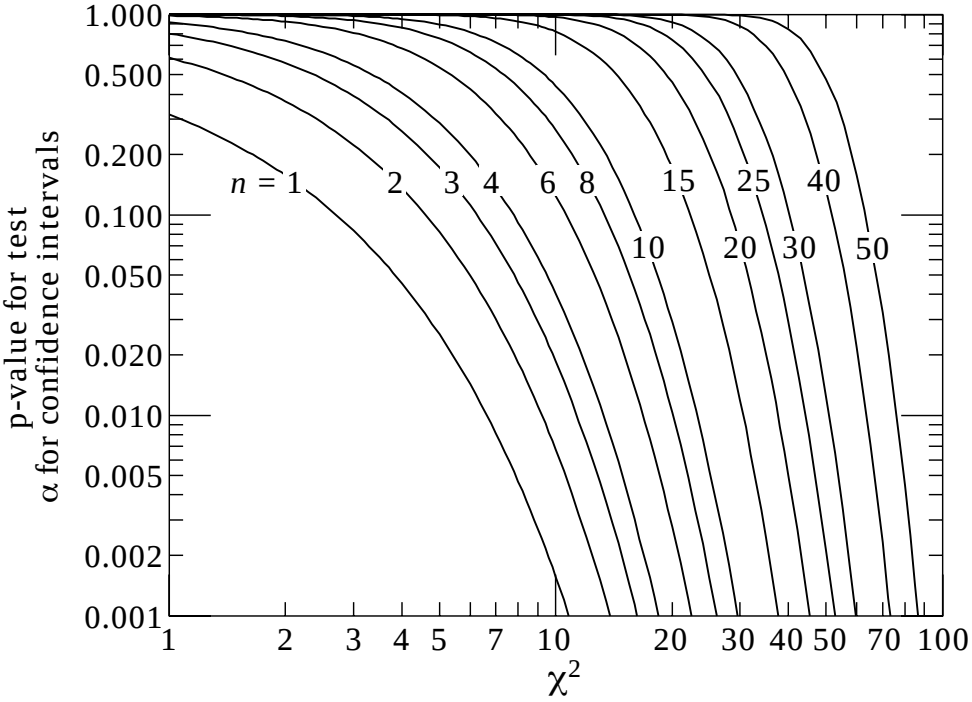


Figure 32.1: One minus the χ^2 cumulative distribution, $1 - F(\chi^2; n)$, for n degrees of freedom. This gives the p-value for the χ^2 goodness-of-fit test as well as one minus the coverage probability for confidence regions (see Sec. 32.3.2.3).

conditions for using the χ^2 p.d.f. do not hold, the statistic can still be defined as before, but its p.d.f. must be determined by other means in order to obtain the p-value, *e.g.*, using a Monte Carlo calculation.

If one finds a χ^2 value much greater than n_d and a correspondingly small p-value, one may be tempted to expect a high degree of uncertainty for any fitted parameters. Although this may be true for systematic errors in the parameters, it is not in general the case for statistical uncertainties. If, for example, the error bars (or covariance matrix) used in constructing the χ^2 are underestimated, then this will lead to underestimated statistical errors for the fitted parameters. But in such a case an estimate $\hat{\theta}$ can differ from the true value θ by an amount much greater than its estimated statistical error. The standard deviations of estimators that one finds from, say, equation (32.11) reflect how widely the estimates would be distributed if one were to repeat the measurement many times, assuming that the measurement errors used in the χ^2 are also correct. They do not include the systematic error which may result from an incorrect hypothesis or incorrectly estimated measurement errors in the χ^2.

Since the mean of the χ^2 distribution is equal to n_d, one expects in a "reasonable" experiment to obtain $\chi^2 \approx n_\mathrm{d}$. Hence the quantity χ^2/n_d is sometimes reported. Since the p.d.f. of χ^2/n_d depends on n_d, however, one must report n_d as well in order to make a meaningful statement. The p-values obtained for different values of χ^2/n_d are shown in Fig. 32.2.

32.3. Confidence intervals and limits

When the goal of an experiment is to determine a parameter θ, the result is usually expressed by quoting, in addition to the point estimate, some sort of interval which reflects the statistical precision of the measurement. In the simplest case this can be given by the parameter's estimated value $\widehat{\theta}$ plus or minus an estimate of the standard deviation of $\widehat{\theta}$, $\sigma_{\widehat{\theta}}$. If, however, the p.d.f. of the estimator is not Gaussian or if there are physical boundaries on the possible values of the parameter, then one usually quotes instead an interval according to one of the procedures described below.

In reporting an interval or limit, the experimenter may wish to

- communicate as objectively as possible the result of the experiment;
- provide an interval that is constructed to cover the true value of the parameter with a specified probability;
- provide the information needed by the consumer of the result to draw conclusions about the parameter or to make a particular decision;
- draw conclusions about the parameter that incorporate stated prior beliefs.

With a sufficiently large data sample, the point estimate and standard deviation (or for the multiparameter case, the parameter estimates and covariance matrix) satisfy essentially all of these goals. For finite data samples, no single method for quoting an interval will achieve all of them. In particular, drawing conclusions about the parameter in the framework of Bayesian statistics necessarily requires subjective input.

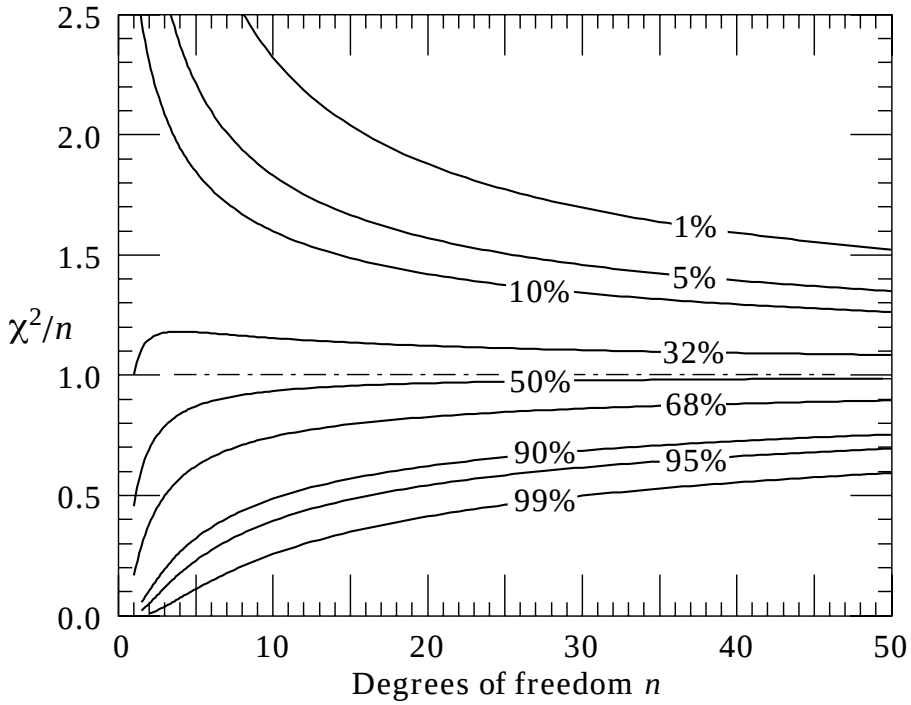


Figure 32.2: The 'reduced' χ^2, equal to χ^2/n, for n degrees of freedom. The curves show as a function of n the χ^2/n that corresponds to a given p-value.

In addition to the goals listed above, the choice of method may be influenced by practical considerations such as ease of producing an interval from the results of several measurements. Of course the experimenter is not restricted to quoting a single interval or limit; one may choose, for example, first to communicate the result with a confidence interval having certain frequentist properties, and then in addition to draw conclusions about a parameter using Bayesian statistics. It is recommended, however, that there be a clear separation between these two aspects of reporting a result. In the remainder of this section we assess the extent to which various types of intervals achieve the goals stated here.

32.3.1. *The Bayesian approach* :

Suppose the outcome of the experiment is characterized by a vector of data $\boldsymbol{x}$, whose probability distribution depends on an unknown parameter (or parameters) $\boldsymbol{\theta}$ that we wish to determine. In Bayesian statistics, all knowledge about $\boldsymbol{\theta}$ is summarized by the posterior p.d.f. $p(\boldsymbol{\theta}|\boldsymbol{x})$, which gives the degree of belief for $\boldsymbol{\theta}$ to take on values in a certain region given the data $\boldsymbol{x}$. It is obtained by using Bayes' theorem,

$$p(\boldsymbol{\theta}|\boldsymbol{x}) = \frac{L(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int L(\boldsymbol{x}|\boldsymbol{\theta}')\pi(\boldsymbol{\theta}')\,d\boldsymbol{\theta}'} \,, \tag{32.30}$$

where $L(\boldsymbol{x}|\boldsymbol{\theta})$ is the likelihood function, *i.e.*, the joint p.d.f. for the data given a certain value of $\boldsymbol{\theta}$, evaluated with the data actually obtained in the experiment, and $\pi(\boldsymbol{\theta})$ is the prior p.d.f. for $\boldsymbol{\theta}$. Note that the denominator in (32.30) serves simply to normalize the posterior p.d.f. to unity.

Bayesian statistics supplies no fundamental rule for determining $\pi(\boldsymbol{\theta})$; this reflects the experimenter's subjective degree of belief about $\boldsymbol{\theta}$ before the measurement was carried out. By itself, therefore, the posterior p.d.f. is not a good way to report objectively the result of an observation, since it contains both the result (through the likelihood function) and the experimenter's prior beliefs. Without the likelihood function, someone with different prior beliefs would be unable to substitute these to determine his or her own posterior p.d.f. This is an important reason, therefore, to publish wherever possible the likelihood function or an appropriate summary of it. Often this can be achieved by reporting the ML estimate and one or several low order derivatives of L evaluated at the estimate.

In the single parameter case, for example, an interval (called a Bayesian or credible interval) $[\theta_{\rm lo}, \theta_{\rm up}]$ can be determined which contains a given fraction $1-\alpha$ of the probability, *i.e.*,

$$1-\alpha = \int_{\theta_{\rm lo}}^{\theta_{\rm up}} p(\theta|\boldsymbol{x})\,d\theta \,. \tag{32.31}$$

Sometimes an upper or lower limit is desired, *i.e.*, $\theta_{\rm lo}$ can be set to zero or $\theta_{\rm up}$ to infinity. In other cases one might choose $\theta_{\rm lo}$ and $\theta_{\rm up}$ such that $p(\theta|\boldsymbol{x})$ is higher everywhere inside the interval than outside; these are called *highest posterior density* (HPD) intervals. Note that HPD intervals are not invariant under a nonlinear transformation of the parameter.

The main difficulty with Bayesian intervals is in quantifying the prior beliefs. Sometimes one attempts to construct $\pi(\boldsymbol{\theta})$ to represent complete ignorance about the parameters by setting it equal to a constant. A problem here is that if the prior p.d.f. is flat in $\boldsymbol{\theta}$, then it is not flat for a nonlinear function of $\boldsymbol{\theta}$, and so a different parametrization of the problem would lead in general to a different posterior p.d.f. In fact, one rarely chooses a flat prior as a true expression of degree of belief about a parameter; rather, it is used as a recipe to construct an interval, which in the end will have certain frequentist properties.

If a parameter is constrained to be non-negative, then the prior p.d.f. can simply be set to zero for negative values. An important example is the case of a Poisson variable n which counts signal events with unknown mean s as well as background with mean b, assumed known. For the signal mean s one often uses the prior

$$\pi(s) = \begin{cases} 0 & s < 0 \\ 1 & s \geq 0 \end{cases} \quad . \tag{32.32}$$

As mentioned above, this is regarded as providing an interval whose frequentist properties can be studied, rather than as representing a degree of belief. In the absence of a clear discovery, (*e.g.*, if $n = 0$ or if in any case n is compatible with the expected background), one usually wishes to place an upper limit on s. Using the likelihood function for Poisson distributed n,

$$L(n|s) = \frac{(s+b)^n}{n!} e^{-(s+b)} \ , \tag{32.33}$$

along with the prior (32.32) in (32.30) gives the posterior density for s. An upper limit $s_{\rm up}$ at confidence level $1-\alpha$ can be obtained by requiring

$$1 - \alpha = \int_{-\infty}^{s_{\rm up}} p(s|n) ds = \frac{\int_{-\infty}^{s_{\rm up}} L(n|s)\, \pi(s)\, ds}{\int_{-\infty}^{\infty} L(n|s)\, \pi(s)\, ds} \ , \tag{32.34}$$

where the lower limit of integration is effectively zero because of the cut-off in $\pi(s)$. By relating the integrals in Eq. (32.34) to incomplete gamma functions, the equation reduces to

$$\alpha = e^{-s_{\rm up}} \frac{\sum_{m=0}^{n} (s_{\rm up} + b)^m / m!}{\sum_{m=0}^{n} b^m / m!} \ . \tag{32.35}$$

This must be solved numerically for the limit $s_{\rm up}$. For the special case of $b = 0$, the sums can be related to the *quantile* $F^{-1}_{\chi^2}$ of the χ^2 distribution (inverse of the cumulative distribution) to give

$$s_{\rm up} = \tfrac{1}{2} F^{-1}_{\chi^2}(1 - \alpha; n_{\rm d}) \ , \tag{32.36}$$

where the number of degrees of freedom is $n_{\rm d} = 2(n+1)$. The quantile of the χ^2 distribution can be obtained using the CERNLIB routine `CHISIN`. It so happens that for the case of $b = 0$, the upper limits from Eq. (32.36) coincide numerically with the values of the frequentist upper limits discussed in Section 32.3.2.4. Values for $1 - \alpha = 0.9$ and 0.95 are given by the values $\nu_{\rm up}$ in Table 32.3. The frequentist properties of confidence intervals for the Poisson mean obtained in this way are discussed in Refs. [2] and [11].

Bayesian statistics provides a framework for incorporating systematic uncertainties into a result. Suppose, for example, that a model depends not only on parameters of interest $\boldsymbol{\theta}$ but on *nuisance parameters* $\boldsymbol{\nu}$, whose values are known with some limited accuracy. For a single nuisance parameter ν, for example, one might have a p.d.f. centered about its nominal value with a certain standard deviation σ_ν. Often a Gaussian p.d.f. provides a reasonable model for one's degree of belief about a nuisance parameter; in other cases more complicated shapes may be appropriate. The likelihood function, prior and posterior p.d.f.s then all depend on both $\boldsymbol{\theta}$ and $\boldsymbol{\nu}$ and are related by Bayes' theorem as usual. One can obtain the posterior p.d.f. for $\boldsymbol{\theta}$ alone by integrating over the nuisance parameters, *i.e.*,

$$p(\boldsymbol{\theta}|\boldsymbol{x}) = \int p(\boldsymbol{\theta}, \boldsymbol{\nu}|\boldsymbol{x})\, d\boldsymbol{\nu} \; . \tag{32.37}$$

If the prior joint p.d.f. for $\boldsymbol{\theta}$ and $\boldsymbol{\nu}$ factorizes, then integrating the posterior p.d.f. over $\boldsymbol{\nu}$ is equivalent to replacing the likelihood function by (see Ref. [12]),

$$L'(\boldsymbol{x}|\boldsymbol{\theta}) = \int L(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{\nu})\pi(\boldsymbol{\nu})\, d\boldsymbol{\nu} \; . \tag{32.38}$$

The function $L'(\boldsymbol{x}|\boldsymbol{\theta})$ can also be used together with frequentist methods that employ the likelihood function such as ML estimation of parameters. The results then have a mixed frequentist/Bayesian character, where the systematic uncertainty due to limited knowledge of the nuisance parameters is built in. Although this may make it more difficult to disentangle statistical from systematic effects, such a hybrid approach may satisfy the objective of reporting the result in a convenient way.

Even if the subjective Bayesian approach is not used explicitly, Bayes' theorem represents the way that people evaluate the impact of a new result on their beliefs. One of the criteria in choosing a method for reporting a measurement, therefore, should be the ease and convenience with which the consumer of the result can carry out this exercise.

32.3.2. *Frequentist confidence intervals* :

The unqualified phrase "confidence intervals" refers to frequentist intervals obtained with a procedure due to Neyman [13], described below. These are intervals (or in the multiparameter case, regions) constructed so as to include the true value of the parameter with a probability greater than or equal to a specified level, called the *coverage probability*. In this section we discuss several techniques for producing intervals that have, at least approximately, this property.

32.3.2.1. *The Neyman construction for confidence intervals:*

Consider a p.d.f. $f(x;\theta)$ where x represents the outcome of the experiment and θ is the unknown parameter for which we want to construct a confidence interval. The variable x could (and often does) represent an estimator for θ. Using $f(x;\theta)$ we can find for a pre-specified probability $1-\alpha$ and for every value of θ a set of values $x_1(\theta,\alpha)$ and $x_2(\theta,\alpha)$ such that

$$P(x_1 < x < x_2;\theta) = 1-\alpha = \int_{x_1}^{x_2} f(x;\theta)\,dx\,. \tag{32.39}$$

This is illustrated in Fig. 32.3: a horizontal line segment $[x_1(\theta,\alpha), x_2(\theta,\alpha)]$ is drawn for representative values of θ. The union of such intervals for all values of θ, designated in the figure as $D(\alpha)$, is known as the *confidence belt*. Typically the curves $x_1(\theta,\alpha)$ and $x_2(\theta,\alpha)$ are monotonic functions of θ, which we assume for this discussion.

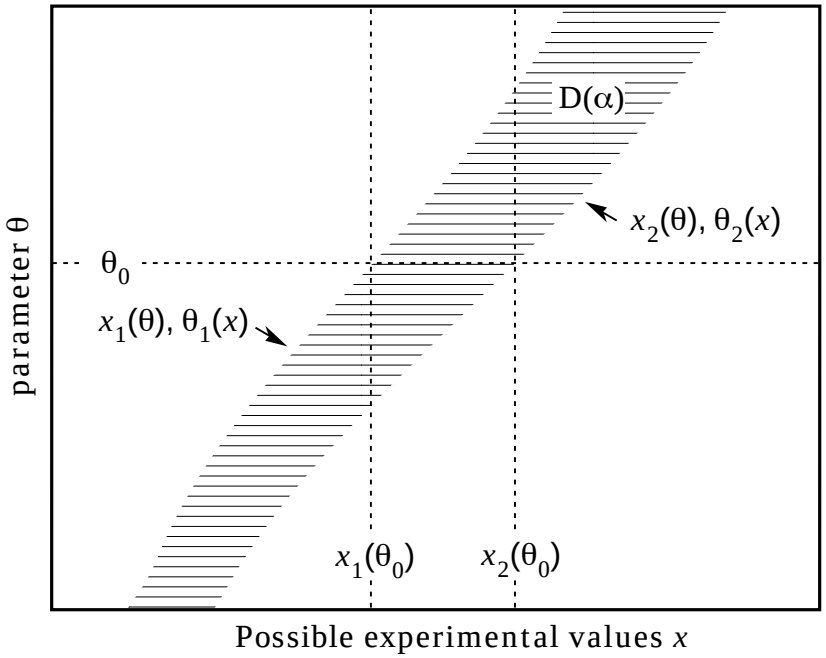


Figure 32.3: Construction of the confidence belt (see text).

Upon performing an experiment to measure x and obtaining a value x_0, one draws a vertical line through x_0. The confidence interval for θ is the set of all values of θ for which the corresponding line segment $[x_1(\theta,\alpha), x_2(\theta,\alpha)]$ is intercepted by this vertical line. Such confidence intervals are said to have a *confidence level* (CL) equal to $1-\alpha$.

Now suppose that the true value of θ is θ_0, indicated in the figure. We see from the figure that θ_0 lies between $\theta_1(x)$ and $\theta_2(x)$ if and only if x lies between $x_1(\theta_0)$ and $x_2(\theta_0)$. The two events thus have the same probability, and since this is true for any value θ_0, we can drop the subscript 0 and obtain

$$1 - \alpha = P(x_1(\theta) < x < x_2(\theta)) = P(\theta_2(x) < \theta < \theta_1(x)) \; . \tag{32.40}$$

In this probability statement $\theta_1(x)$ and $\theta_2(x)$, *i.e.*, the endpoints of the interval, are the random variables and θ is an unknown constant. If the experiment were to be repeated a large number of times, the interval $[\theta_1, \theta_2]$ would vary, covering the fixed value θ in a fraction $1 - \alpha$ of the experiments.

The condition of coverage Eq. (32.39) does not determine x_1 and x_2 uniquely and additional criteria are needed. The most common criterion is to choose *central intervals* such that the probabilities excluded below x_1 and above x_2 are each $\alpha/2$. In other cases one may want to report only an upper or lower limit, in which case the probability excluded below x_1 or above x_2 can be set to zero. Another principle based on *likelihood ratio ordering* for determining which values of x should be included in the confidence belt is discussed in Sec. 32.3.2.2

When the observed random variable x is continuous, the coverage probability obtained with the Neyman construction is $1 - \alpha$, regardless of the true value of the parameter. If x is discrete, however, it is not possible to find segments $[x_1(\theta, \alpha), x_2(\theta, \alpha)]$ that satisfy (32.39) exactly for all values of θ. By convention one constructs the confidence belt requiring the probability $P(x_1 < x < x_2)$ to be *greater than or equal to* $1 - \alpha$. This gives confidence intervals that include the true parameter with a probability greater than or equal to $1 - \alpha$.

32.3.2.2. *Relationship between intervals and tests:*

An equivalent method of constructing confidence intervals is to consider a test (see Sec. 32.2) of the hypothesis that the parameter's true value is θ. One then excludes all values of θ where the hypothesis would be rejected at a significance level less than α. The remaining values constitute the confidence interval at confidence level $1 - \alpha$.

In this procedure one is still free to choose the test to be used; this corresponds to the freedom in the Neyman construction as to which values of the data are included in the confidence belt. One possibility is use a test statistic based on the *likelihood ratio*,

$$\lambda = \frac{f(x; \theta)}{f(x; \widehat{\theta}\,)} \; , \tag{32.41}$$

where $\widehat{\theta}$ is the value of the parameter which, out of all allowed values, maximizes $f(x; \theta)$. This results in the intervals described in [14] by Feldman and Cousins. The same intervals can be obtained from the Neyman construction described in the previous section by including in the confidence belt those values of x which give the greatest values of λ.

Another technique that can be formulated in the language of statistical tests has been used to set limits on the Higgs mass from measurements at LEP [15,16]. For each value of the Higgs mass, a statistic called $\mathrm{CL_s}$ is determined from the ratio

$$\mathrm{CL_s} = \frac{p\text{-value of signal plus background hypothesis}}{1 - p\text{-value of hypothesis of background only}} \; . \tag{32.42}$$

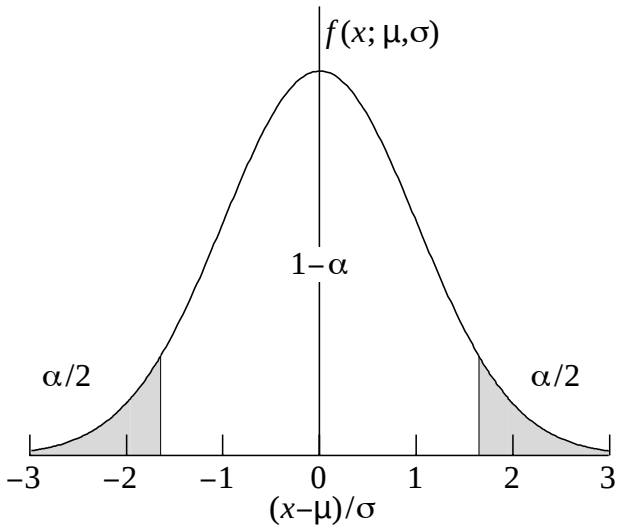


Figure 32.4: Illustration of a symmetric 90% confidence interval (unshaded) for a measurement of a single quantity with Gaussian errors. Integrated probabilities, defined by α, are as shown.

The p-values in (32.42) are themselves based on a goodness-of-fit statistic which depends in general on the signal being tested, *i.e.*, on the hypothesized Higgs mass. Smaller $\mathrm{CL_s}$ corresponds to a lesser level of agreement with the signal hypothesis.

In the usual procedure for constructing confidence intervals, one would exclude the signal hypothesis if the probability to obtain a value of $\mathrm{CL_s}$ less than the one actually observed is less than α. The LEP Higgs group has in fact followed a more conservative approach and excludes the signal at a confidence level $1-\alpha$ if $\mathrm{CL_s}$ itself (not the probability to obtain a lower $\mathrm{CL_s}$ value) is less than α. This results in a coverage probability that is in general greater than $1-\alpha$. The interpretation of such intervals is discussed in [15,16].

32.3.2.3. *Gaussian distributed measurements:*

An important example of constructing a confidence interval is when the data consists of a single random variable x that follows a Gaussian distribution; this is often the case when x represents an estimator for a parameter and one has a sufficiently large data sample. If there is more than one parameter being estimated, the multivariate Gaussian is used. For the univariate case with known σ,

$$1-\alpha = \frac{1}{\sqrt{2\pi}\sigma}\int_{\mu-\delta}^{\mu+\delta} e^{-(x-\mu)^2/2\sigma^2}\,dx = \mathrm{erf}\left(\frac{\delta}{\sqrt{2}\,\sigma}\right) \tag{32.43}$$

is the probability that the measured value x will fall within $\pm\delta$ of the true value μ. From the symmetry of the Gaussian with respect to x and μ, this is also the probability for the interval $x \pm \delta$ to include μ. Fig. 32.4 shows a $\delta = 1.64\sigma$ confidence interval unshaded. The choice $\delta = \sigma$ gives an interval called the *standard error* which has $1-\alpha = 68.27\%$ if σ is known. Values of α for other frequently used choices of δ are given in Table 32.1.

Table 32.1: Area of the tails α outside $\pm\delta$ from the mean of a Gaussian distribution.

α	δ	α	δ
0.3173	1σ	0.2	1.28σ
$4.55\ \times 10^{-2}$	2σ	0.1	1.64σ
$2.7\ \times 10^{-3}$	3σ	0.05	1.96σ
6.3×10^{-5}	4σ	0.01	2.58σ
5.7×10^{-7}	5σ	0.001	3.29σ
2.0×10^{-9}	6σ	10^{-4}	3.89σ

We can set a one-sided (upper or lower) limit by excluding above $x + \delta$ (or below $x - \delta$). The values of α for such limits are half the values in Table 32.1.

In addition to Eq. (32.43), α and δ are also related by the cumulative distribution function for the χ^2 distribution,

$$\alpha = 1 - F(\chi^2; n)\ , \tag{32.44}$$

for $\chi^2 = (\delta/\sigma)^2$ and $n = 1$ degree of freedom. This can be obtained from Fig. 32.1 on the $n = 1$ curve or by using the CERNLIB routine `PROB`.

For multivariate measurements of, say, n parameter estimates $\widehat{\boldsymbol{\theta}} = (\widehat{\theta}_1, \ldots, \widehat{\theta}_n)$, one requires the full covariance matrix $V_{ij} = \mathrm{cov}[\widehat{\theta}_i, \widehat{\theta}_j]$, which can be estimated as described in Sections 32.1.2 and 32.1.3. Under fairly general conditions with the methods of maximum-likelihood or least-squares in the large sample limit, the estimators will be distributed according to a multivariate Gaussian centered about the true (unknown) values $\boldsymbol{\theta}$, and furthermore the likelihood function itself takes on a Gaussian shape.

The standard error ellipse for the pair $(\widehat{\theta}_i, \widehat{\theta}_j)$ is shown in Fig. 32.5, corresponding to a contour $\chi^2 = \chi^2_{\mathrm{min}} + 1$ or $\ln L = \ln L_{\mathrm{max}} - 1/2$. The ellipse is centered about the estimated values $\widehat{\boldsymbol{\theta}}$, and the tangents to the ellipse give the standard deviations of the estimators, σ_i and σ_j. The angle of the major axis of the ellipse is given by

$$\tan 2\phi = \frac{2\rho_{ij}\sigma_i\sigma_j}{\sigma_i^2 - \sigma_j^2}\ , \tag{32.45}$$

where $\rho_{ij} = \mathrm{cov}[\widehat{\theta}_i, \widehat{\theta}_j]/\sigma_i\sigma_j$ is the correlation coefficient.

The correlation coefficient can be visualized as the fraction of the distance σ_i from the ellipse's horizontal centerline at which the ellipse becomes tangent to vertical, *i.e.* at the distance $\rho_{ij}\sigma_i$ below the centerline as shown. As ρ_{ij} goes to $+1$ or -1, the ellipse thins to a diagonal line.

It could happen that one of the parameters, say, θ_j, is known from previous measurements to a precision much better than σ_j so that the current measurement contributes almost nothing to the knowledge of θ_j. However, the current measurement of

of θ_i and its dependence on θ_j may still be important. In this case, instead of quoting both parameter estimates and their correlation, one sometimes reports the value of θ_i which minimizes χ^2 at a fixed value of θ_j, such as the PDG best value. This θ_i value lies along the dotted line between the points where the ellipse becomes tangent to vertical, and has statistical error $\sigma_{\rm inner}$ as shown on the figure, where $\sigma_{\rm inner} = (1-\rho_{ij}^2)^{1/2}\sigma_i$. Instead of the correlation ρ_{ij}, one reports the dependency $d\widehat{\theta}_i/d\theta_j$ which is the slope of the dotted line. This slope is related to the correlation coefficient by $d\widehat{\theta}_i/d\theta_j = \rho_{ij} \times \frac{\sigma_i}{\sigma_j}$.

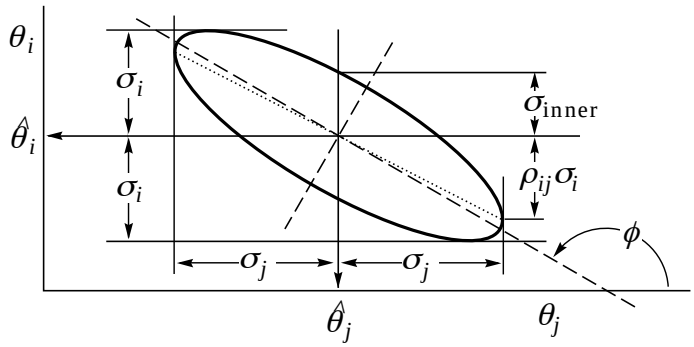


Figure 32.5: Standard error ellipse for the estimators $\widehat{\theta}_i$ and $\widehat{\theta}_j$. In this case the correlation is negative.

As in the single-variable case, because of the symmetry of the Gaussian function between $\boldsymbol{\theta}$ and $\widehat{\boldsymbol{\theta}}$, one finds that contours of constant $\ln L$ or χ^2 cover the true values with a certain, fixed probability. That is, the confidence region is determined by

$$\ln L(\boldsymbol{\theta}) \geq \ln L_{\rm max} - \Delta \ln L \ , \tag{32.46}$$

or where a χ^2 has been defined for use with the method of least squares,

$$\chi^2(\boldsymbol{\theta}) \leq \chi^2_{\rm min} + \Delta\chi^2 \ . \tag{32.47}$$

Values of $\Delta\chi^2$ or $2\Delta \ln L$ are given in Table 32.2 for several values of the coverage probability and number of fitted parameters.

For finite data samples, the probability for the regions determined by Equations (32.46) or (32.47) to cover the true value of $\boldsymbol{\theta}$ will depend on $\boldsymbol{\theta}$, so these are not exact confidence regions according to our previous definition. Nevertheless, they can still have a coverage probability only weakly dependent on the true parameter and approximately as given in Table 32.2. In any case the coverage probability of the intervals or regions obtained according to this procedure can in principle be determined as a function of the true parameter(s), for example, using a Monte Carlo calculation.

One of the practical advantages of intervals that can be constructed from the log-likelihood function or χ^2 is that it is relatively simple to produce the interval for the combination of several experiments. If N independent measurements result in

Table 32.2: $\Delta\chi^2$ or $2\Delta \ln L$ corresponding to a coverage probability $1-\alpha$ in the large data sample limit, for joint estimation of m parameters.

$(1-\alpha)$ (%)	$m=1$	$m=2$	$m=3$
68.27	1.00	2.30	3.53
90.	2.71	4.61	6.25
95.	3.84	5.99	7.82
95.45	4.00	6.18	8.03
99.	6.63	9.21	11.34
99.73	9.00	11.83	14.16

log-likelihood functions $\ln L_i(\boldsymbol{\theta})$, then the combined log-likelihood function is simply the sum,

$$\ln L(\boldsymbol{\theta}) = \sum_{i=1}^{N} \ln L_i(\boldsymbol{\theta}) \; . \tag{32.48}$$

This can then be used to determine an approximate confidence interval or region with Equation (32.46), just as with a single experiment.

32.3.2.4. *Poisson or binomial data:*

Another important class of measurements consists of counting a certain number of events n. In this section we will assume these are all events of the desired type, *i.e.*, there is no background. If n represents the number of events produced in a reaction with cross section σ, say, in a fixed integrated luminosity $\mathcal{L}$, then it follows a Poisson distribution with mean $\nu = \sigma\mathcal{L}$. If, on the other hand, one has selected a larger sample of N events and found n of them to have a particular property, then n follows a binomial distribution where the parameter p gives the probability for the event to possess the property in question. This is appropriate, *e.g.*, for estimates of branching ratios or selection efficiencies based on a given total number of events.

For the case of Poisson distributed n, the upper and lower limits on the mean value ν can be found from the Neyman procedure to be

$$\nu_{\rm lo} = \tfrac{1}{2} F^{-1}_{\chi^2}(\alpha_{\rm lo}; 2n) \; , \tag{32.49a}$$

$$\nu_{\rm up} = \tfrac{1}{2} F^{-1}_{\chi^2}(1-\alpha_{\rm up}; 2(n+1)) \; , \tag{32.49b}$$

where the upper and lower limits are at confidence levels of $1-\alpha_{\rm lo}$ and $1-\alpha_{\rm up}$, respectively, and $F^{-1}_{\chi^2}$ is the *quantile* of the χ^2 distribution (inverse of the cumulative distribution). The quantiles $F^{-1}_{\chi^2}$ can be obtained from standard tables or from the

CERNLIB routine `CHISIN`. For central confidence intervals at confidence level $1-\alpha$, set $\alpha_{\rm lo} = \alpha_{\rm up} = \alpha/2$.

It happens that the upper limit from (32.49a) coincides numerically with the Bayesian upper limit for a Poisson parameter using a uniform prior p.d.f. for ν. Values for confidence levels of 90% and 95% are shown in Table 32.3.

Table 32.3: Lower and upper (one-sided) limits for the mean ν of a Poisson variable given n observed events in the absence of background, for confidence levels of 90% and 95%.

	$1-\alpha$ =90%		$1-\alpha$ =95%	
n	$\nu_{\rm lo}$	$\nu_{\rm up}$	$\nu_{\rm lo}$	$\nu_{\rm up}$
0	–	2.30	–	3.00
1	0.105	3.89	0.051	4.74
2	0.532	5.32	0.355	6.30
3	1.10	6.68	0.818	7.75
4	1.74	7.99	1.37	9.15
5	2.43	9.27	1.97	10.51
6	3.15	10.53	2.61	11.84
7	3.89	11.77	3.29	13.15
8	4.66	12.99	3.98	14.43
9	5.43	14.21	4.70	15.71
10	6.22	15.41	5.43	16.96

For the case of binomially distributed n successes out of N trials with probability of success p, the upper and lower limits on p are found to be

$$p_{\rm lo} = \frac{nF_F^{-1}[\alpha_{\rm lo}; 2n, 2(N-n+1)]}{N-n+1 \,+\, nF_F^{-1}[\alpha_{\rm lo}; 2n, 2(N-n+1)]} \,, \tag{32.50a}$$

$$p_{\rm up} = \frac{(n+1)F_F^{-1}[1-\alpha_{\rm up}; 2(n+1), 2(N-n)]}{(N-n) \,+\, (n+1)F_F^{-1}[1-\alpha_{\rm up}; 2(n+1), 2(N-n)]} \,. \tag{32.50b}$$

Here F_F^{-1} is the quantile of the F distribution (also called the Fisher–Snedecor distribution; see Ref. [4]).

32.3.2.5. *Difficulties with intervals near a boundary:*

A number of issues arise in the construction and interpretation of confidence intervals when the parameter can only take on values in a restricted range. An important example is where the mean of a Gaussian variable is constrained on physical grounds to be non-negative. This arises, for example, when the square of the neutrino mass is estimated from $\widehat{m}^2 = \widehat{E}^2 - \widehat{p}^2$, where $\widehat{E}$ and $\widehat{p}$ are independent, Gaussian distributed estimates of the energy and momentum. Although the true m^2 is constrained to be positive, random errors in $\widehat{E}$ and $\widehat{p}$ can easily lead to negative values for the estimate $\widehat{m}^2$.

If one uses the prescription given above for Gaussian distributed measurements, which says to construct the interval by taking the estimate plus or minus one standard deviation, then this can give intervals that are partially or entirely in the unphysical region. In fact, by following strictly the Neyman construction for the central confidence interval, one finds that the interval is truncated below zero; nevertheless an extremely small or even a zero-length interval can result.

An additional important example is where the experiment consists of counting a certain number of events, n, which is assumed to be Poisson distributed. Suppose the expectation value $E[n] = \nu$ is equal to $s + b$, where s and b are the means for signal and background processes, and assume further that b is a known constant. Then $\widehat{s} = n - b$ is an unbiased estimator for s. Depending on true magnitudes of s and b, the estimate $\widehat{s}$ can easily fall in the negative region. Similar to the Gaussian case with the positive mean, the central confidence interval or even the upper limit for s may be of zero length.

The confidence interval is in fact designed not to cover the parameter with a probability of at most α, and if a zero-length interval results, then this is evidently one of those experiments. So although the construction is behaving as it should, a null interval is an unsatisfying result to report and several solutions to this type of problem are possible.

An additional difficulty arises when a parameter estimate is not significantly far away from the boundary, in which case it is natural to report a one-sided confidence interval (often an upper limit). It is straightforward to force the Neyman prescription to produce only an upper limit by setting $x_2 = \infty$ in Eq. 32.39. Then x_1 is uniquely determined and the upper limit can be obtained. If, however, the data come out such that the parameter estimate is not so close to the boundary, one might wish to report a central (*i.e.*, two-sided) confidence interval. As pointed out by Feldman and Cousins [14], however, if the decision to report an upper limit or two-sided interval is made by looking at the data ("flip-flopping"), then the resulting intervals will not in general cover the parameter with the probability $1 - \alpha$.

With the confidence intervals suggested in [14], the prescription determines whether the interval is one- or two-sided in a way which preserves the coverage probability. Intervals with this property are said to be *unified*. Furthermore, the Feldman–Cousins prescription is such that null intervals do not occur. For a given choice of $1 - \alpha$, if the parameter estimate is sufficiently close to the boundary, then the method gives a one-sided limit. In the case of a Poisson variable in the presence of background, for example, this would occur if the number of observed events is compatible with the expected background. For parameter estimates increasingly far away from the boundary, i.e., for increasing signal

significance, the interval makes a smooth transition from one- to two-sided, and far away from the boundary one obtains a central interval.

The intervals according to this method for the mean of Poisson variable in the absence of background are given in Table 32.4. (Note that α in [14] is defined following Neyman [13] as the coverage probability; this is opposite the modern convention used here in which the coverage probability is $1-\alpha$.) The values of $1-\alpha$ given here refer to the coverage of the true parameter by the whole interval $[\nu_1, \nu_2]$. In Table 32.3 for the one-sided upper and lower limits, however, $1-\alpha$ refers to the probability to have individually $\nu_{\rm up} \geq \nu$ or $\nu_{\rm lo} \leq \nu$.

Table 32.4: Unified confidence intervals $[\nu_1, \nu_2]$ for a the mean of a Poisson variable given n observed events in the absence of background, for confidence levels of 90% and 95%.

	$1-\alpha$ =90%		$1-\alpha$ =95%	
n	ν_1	ν_2	ν_1	ν_2
0	0.00	2.44	0.00	3.09
1	0.11	4.36	0.05	5.14
2	0.53	5.91	0.36	6.72
3	1.10	7.42	0.82	8.25
4	1.47	8.60	1.37	9.76
5	1.84	9.99	1.84	11.26
6	2.21	11.47	2.21	12.75
7	3.56	12.53	2.58	13.81
8	3.96	13.99	2.94	15.29
9	4.36	15.30	4.36	16.77
10	5.50	16.50	4.75	17.82

A potential difficulty with unified intervals arises if, for example, one constructs such an interval for a Poisson parameter s of some yet to be discovered signal process with, say, $1-\alpha = 0.9$. If the true signal parameter is zero, or in any case much less than the expected background, one will usually obtain a one-sided upper limit on s. In a certain fraction of the experiments, however, a two-sided interval for s will result. Since, however, one typically chooses $1-\alpha$ to be only 0.9 or 0.95 when searching for a new effect, the value $s = 0$ may be excluded from the interval before the existence of the effect is well established. It must then be communicated carefully that in excluding $s = 0$ from the interval, one is not necessarily claiming to have discovered the effect.

The intervals constructed according to the unified procedure in [14] for a Poisson variable n consisting of signal and background have the property that for $n = 0$

observed events, the upper limit decreases for increasing expected background. This is counter-intuitive, since it is known that if $n = 0$ for the experiment in question, then no background was observed, and therefore one may argue that the expected background should not be relevant. The extent to which one should regard this feature as a drawback is a subject of some controversy (see, *e.g.*, Ref. [18]).

Another possibility is to construct a Bayesian interval as described in Section 32.3.1. The presence of the boundary can be incorporated simply by setting the prior density to zero in the unphysical region. Priors based on invariance principles (rather than subjective degree of belief) for the Poisson mean are rarely used in high energy physics; they diverge for the case of zero events observed, and they give upper limits which undercover when evaluated by the frequentist definition of coverage [2]. Rather, priors uniform in the Poisson mean have been used, although as previously mentioned, this is generally not done to reflect the experimenter's degree of belief but rather as a procedure for obtaining an interval with certain frequentist properties. The resulting upper limits have a coverage probability that depends on the true value of the Poisson parameter and is everywhere greater than the stated probability content. Lower limits and two-sided intervals for the Poisson mean based on flat priors undercover, however, for some values of the parameter, although to an extent that in practical cases may not be too severe [2, 11]. Intervals constructed in this way have the advantage of being easy to derive; if several independent measurements are to be combined then one simply multiplies the likelihood functions (cf. Eq. (32.48)).

An additional alternative is presented by the intervals found from the likelihood function or χ^2 using the prescription of Equations (32.46) or (32.47). As in the case of the Bayesian intervals, the coverage probability is not, in general, independent of the true parameter. Furthermore, these intervals can for some parameter values undercover. The coverage probability can of course be determined with some extra effort and reported with the result.

Also as in the Bayesian case, intervals derived from the value of the likelihood function from a combination of independent experiments can be determined simply by multiplying the likelihood functions. These intervals are also invariant under transformation of the parameter; this is not true for Bayesian intervals with a conventional flat prior, because a uniform distribution in, say, θ will not be uniform if transformed to θ^2. Use of the likelihood function to determine approximate confidence intervals is discussed further in [17].

In any case it is important always to report sufficient information so that the result can be combined with other measurements. Often this means giving an unbiased estimator and its standard deviation, even if the estimated value is in the unphysical region.

Regardless of the type of interval reported, the consumer of that result will almost certainly use it to derive some impression about the value of the parameter. This will inevitably be done, either explicitly or intuitively, with Bayes' theorem,

$$p(\theta|\text{result}) \propto L(\text{result}|\theta)\pi(\theta) \ , \tag{32.51}$$

where the reader supplies his or her own prior beliefs $\pi(\theta)$ about the parameter, and the

'result' is whatever sort of interval or other information the author has reported. For all of the intervals discussed, therefore, it is not sufficient to know the result; one must also know the probability to have obtained this result as a function of the parameter, *i.e.*, the likelihood. Contours of constant likelihood, for example, provide this information, and so an interval obtained from $\ln L = \ln L_{\max} - \Delta \ln L$ already takes one step in this direction.

It can also be useful with a frequentist interval to calculate its subjective probability content using the posterior p.d.f. based on one or several reasonable guesses for the prior p.d.f. If it turns out to be significantly less than the stated confidence level, this warns that it would be particularly misleading to draw conclusions about the parameter's value without further information from the likelihood.

References:

1. B. Efron, Am. Stat. **40**, 11 (1986).
2. R.D. Cousins, Am. J. Phys. **63**, 398 (1995).
3. A. Stuart, A.K. Ord, and Arnold, *Kendall's Advanced Theory of Statistics*, Vol. 2 *Classical Inference and Relationship* 6th Ed., (Oxford Univ. Press, 1998), and earlier editions by Kendall and Stuart. The likelihood-ratio ordering principle is described at the beginning of Ch. 23. Chapter 26 compares different schools of statistical inference.
4. W.T. Eadie, D. Drijard, F.E. James, M. Roos, and B. Sadoulet, *Statistical Methods in Experimental Physics* (North Holland, Amsterdam and London, 1971).
5. H. Cramér, *Mathematical Methods of Statistics,* Princeton Univ. Press, New Jersey (1958).
6. L. Lyons, *Statistics for Nuclear and Particle Physicists* (Cambridge University Press, New York, 1986).
7. R. Barlow, Nucl. Inst. Meth. A **297**, 496 (1990).
8. G. Cowan, *Statistical Data Analysis* (Oxford University Press, Oxford, 1998).
9. For a review, see S. Baker and R. Cousins, Nucl. Instrum. Methods **221**, 437 (1984).
10. For information on neural networks and related topics, see *e.g.* C.M. Bishop, *Neural Networks for Pattern Recognition*, Clarendon Press, Oxford (1995); C. Peterson and T. Rögnvaldsson, An Introduction to Artificial Neural Networks, in *Proceedings of the 1991 CERN School of Computing*, C. Verkerk (ed.), CERN 92-02 (1992).
11. Byron P. Roe and Michael B. Woodroofe, Phys. Rev. **D63**, 13009 (2001).
12. Paul H. Garthwaite, Ian T. Jolliffe and Byron Jones, *Statistical Inference* (Prentice Hall, 1995).
13. J. Neyman, Phil. Trans. Royal Soc. London, Series A, **236**, 333 (1937), reprinted in *A Selection of Early Statistical Papers on J. Neyman* (University of California Press, Berkeley, 1967).
14. G.J. Feldman and R.D. Cousins, Phys. Rev. **D57**, 3873 (1998). This paper does not specify what to do if the ordering principle gives equal rank to some values of x. Eq. 23.6 of Ref. 3 gives the rule: all such points are included in the acceptance region (the domain $D(\alpha)$). Some authors have assumed the contrary, and shown that one can then obtain null intervals.
15. T. Junk, Nucl. Inst. Meth. A **434**, 435 (1999).

16. A.L. Read, *Modified frequentist analysis of search results (the CL_s method)*, in F. James, L. Lyons and Y. Perrin (eds.), *Workshop on Confidence Limits*, CERN Yellow Report 2000-005, available through `weblib.cern.ch`.
17. F. Porter, Nucl. Inst. Meth. A **368**, 793 (1996).
18. Workshop on Confidence Limits, CERN, 17-18 Jan. 2000, `www.cern.ch/CERN/Divisions/EP/Events/CLW/`. The proceedings, F. James, L. Lyons, and Y. Perrin (eds.), CERN Yellow Report 2000-005, are available through `weblib.cern.ch`. See also the later Fermilab workshop linked to the CERN web page.

33. MONTE CARLO TECHNIQUES

Revised July 1995 by S. Youssef (SCRI, Florida State University).Updated February 2000 by R. Cousins (UCLA) in consultation with F. James (CERN); October 2003 by G. Cowan (RHUL) and R. Miquel (LBNL), and September 2005 by G. Cowan (RHUL).

Monte Carlo techniques are often the only practical way to evaluate difficult integrals or to sample random variables governed by complicated probability density functions. Here we describe an assortment of methods for sampling some commonly occurring probability density functions.

33.1. Sampling the uniform distribution

Most Monte Carlo sampling or integration techniques assume a "random number generator" which generates uniform statistically independent values on the half open interval $[0, 1)$. There is a long history of problems with various generators on a finite digital computer, but recently, the RANLUX generator [1] has emerged with a solid theoretical basis in chaos theory. Based on the method of Lüscher, it allows the user to select different quality levels, trading off quality with speed.

Other generators are also available which pass extensive batteries of tests for statistical independence and which have periods which are so long that, for practical purposes, values from these generators can be considered to be uniform and statistically independent. In particular, the lagged-Fibonacci based generator introduced by Marsaglia, Zaman, and Tsang [2] is efficient, has a period of approximately 10^{43}, produces identical sequences on a wide variety of computers and, passes the extensive "DIEHARD" battery of tests [3]. Many commonly available congruential generators fail these tests and often have sequences (typically with periods less than 2^{32}) which can be easily exhausted on modern computers and should therefore be avoided [4].

33.2. Inverse transform method

If the desired probability density function is $f(x)$ on the range $-\infty < x < \infty$, its cumulative distribution function (expressing the probability that $x \leq a$) is given by Eq. (31.6). If a is chosen with probability density $f(a)$, then the integrated probability up to point a, $F(a)$, is itself a random variable which will occur with uniform probability density on $[0, 1]$. If x can take on any value, and ignoring the endpoints, we can then find a unique x chosen from the p.d.f. $f(s)$ for a given u if we set

$$u = F(x) \ , \tag{33.1}$$

provided we can find an inverse of F, defined by

$$x = F^{-1}(u) \ . \tag{33.2}$$

This method is shown in Fig. 33.1a. It is most convenient when one can calculate by hand the inverse function of the indefinite integral of f. This is the case for some common functions $f(x)$ such as $\exp(x)$, $(1 - x)^n$, and $1/(1 + x^2)$ (Cauchy or Breit-Wigner), although it does not necessarily produce the fastest generator. CERNLIB contains routines to implement this method numerically, working from functions or histograms.

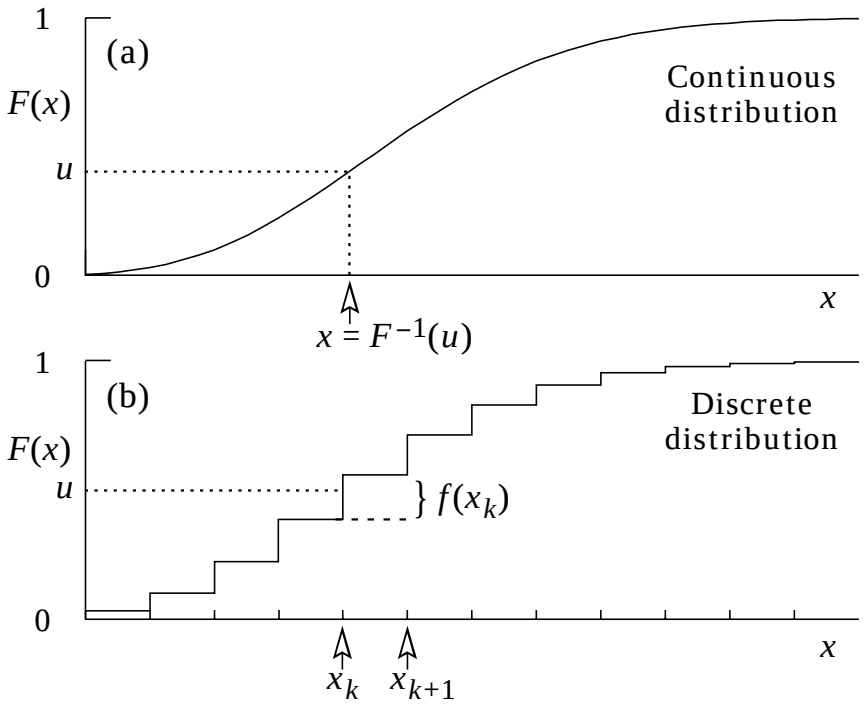


Figure 33.1: Use of a random number u chosen from a uniform distribution (0,1) to find a random number x from a distribution with cumulative distribution function $F(x)$.

For a discrete distribution, $F(x)$ will have a discontinuous jump of size $f(x_k)$ at each allowed $x_k, k = 1, 2, \cdots$. Choose u from a uniform distribution on (0,1) as before. Find x_k such that

$$F(x_{k-1}) < u \leq F(x_k) \equiv \text{Prob}\ (x \leq x_k) = \sum_{i=1}^{k} f(x_i)\ ; \tag{33.3}$$

then x_k is the value we seek (note: $F(x_0) \equiv 0$). This algorithm is illustrated in Fig. 33.1b.

33.3. Acceptance-rejection method (Von Neumann)

Very commonly an analytic form for $F(x)$ is unknown or too complex to work with, so that obtaining an inverse as in Eq. (33.2) is impractical. We suppose that for any given value of x the probability density function $f(x)$ can be computed and further that enough is known about $f(x)$ that we can enclose it entirely inside a shape which is C times an easily generated distribution $h(x)$ as illustrated in Fig. 33.2.

Frequently $h(x)$ is uniform or is a normalized sum of uniform distributions. Note that both $f(x)$ and $h(x)$ must be normalized to unit area and therefore the proportionality constant $C > 1$. To generate $f(x)$, first generate a candidate x according to $h(x)$. Calculate $f(x)$ and the height of the envelope $C\,h(x)$; generate u and test if $uC\,h(x) \leq f(x)$. If so, accept x; if not reject x and try again. If we regard x and

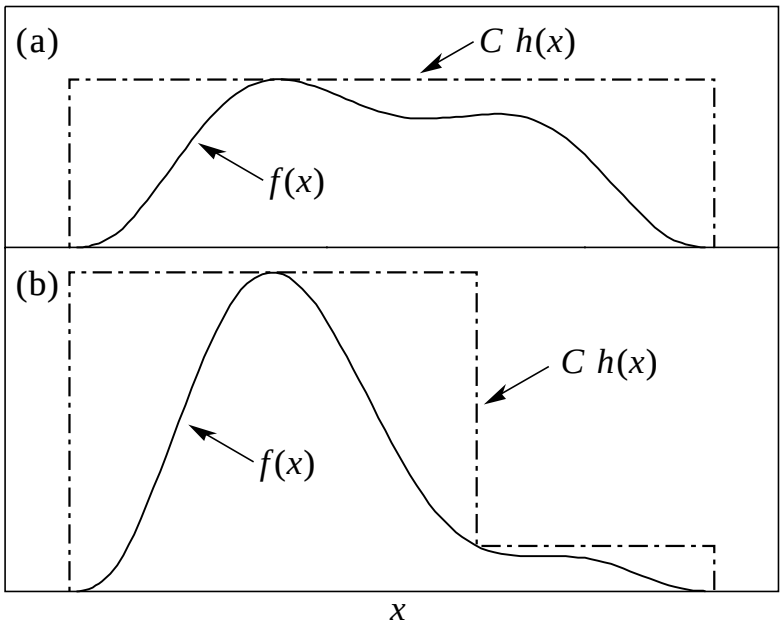


Figure 33.2: Illustration of the acceptance-rejection method. Random points are chosen inside the upper bounding figure, and rejected if the ordinate exceeds $f(x)$. Lower figure illustrates importance sampling.

$uC\,h(x)$ as the abscissa and ordinate of a point in a two-dimensional plot, these points will populate the entire area $C\,h(x)$ in a smooth manner; then we accept those which fall under $f(x)$. The efficiency is the ratio of areas, which must equal $1/C$; therefore we must keep C as close as possible to 1.0. Therefore we try to choose $C\,h(x)$ to be as close to $f(x)$ as convenience dictates, as in the lower part of Fig. 33.2. This practice is called importance sampling, because we generate more trial values of x in the region where $f(x)$ is most important.

33.4. Algorithms

Algorithms for generating random numbers belonging to many different distributions are given by Press [5], Ahrens and Dieter [6], Rubinstein [7], Everett and Cashwell [8], Devroye [9], and Walck [10]. For many distributions alternative algorithms exist, varying in complexity, speed, and accuracy. For time-critical applications, these algorithms may be coded in-line to remove the significant overhead often encountered in making function calls. Variables named "u" are assumed to be independent and uniform on [0,1). (Hence, u must be verified to be non-zero where relevant.)

In the examples given below, we use the notation for the variables and parameters given in Table 31.1.

33.4.1. ***Exponential decay*** :

This is a common application of the inverse transform method and uses the fact that if u is uniformly distributed in $[0,1]$ then $(1-u)$ is as well. Consider an exponential p.d.f. $f(t) = (1/\tau)\exp(-t/\tau)$ that is truncated so as to lie between two values, a and b, and renormalized to unit area. To generate decay times t according to this p.d.f., first let $\alpha = \exp(-a/\tau)$ and $\beta = \exp(-b/\tau)$; then generate u and let

$$t = -\tau \ln(\beta + u(\alpha - \beta)). \tag{33.4}$$

For $(a,b) = (0,\infty)$, we have simply $t = -\tau \ln u$. (See also Sec. 33.4.6.)

33.4.2. ***Isotropic direction in 3D*** :

Isotropy means the density is proportional to solid angle, the differential element of which is $d\Omega = d(\cos\theta)d\phi$. Hence $\cos\theta$ is uniform $(2u_1 - 1)$ and ϕ is uniform $(2\pi u_2)$. For alternative generation of $\sin\phi$ and $\cos\phi$, see the next subsection.

33.4.3. ***Sine and cosine of random angle in 2D*** :

Generate u_1 and u_2. Then $v_1 = 2u_1 - 1$ is uniform on (−1,1), and $v_2 = u_2$ is uniform on (0,1). Calculate $r^2 = v_1^2 + v_2^2$. If $r^2 > 1$, start over. Otherwise, the sine (S) and cosine (C) of a random angle (*i.e.*, uniformly distributed between zero and 2π) are given by

$$S = 2v_1v_2/r^2 \quad \text{and} \quad C = (v_1^2 - v_2^2)/r^2 \ . \tag{33.5}$$

33.4.4. ***Gaussian distribution*** :

If u_1 and u_2 are uniform on (0,1), then

$$z_1 = \sin 2\pi u_1 \sqrt{-2\ln u_2} \quad \text{and} \quad z_2 = \cos 2\pi u_1 \sqrt{-2\ln u_2} \tag{33.6}$$

are independent and Gaussian distributed with mean 0 and $\sigma = 1$.

There are many faster variants of this basic algorithm. For example, construct $v_1 = 2u_1 - 1$ and $v_2 = 2u_2 - 1$, which are uniform on (−1,1). Calculate $r^2 = v_1^2 + v_2^2$, and if $r^2 > 1$ start over. If $r^2 < 1$, it is uniform on (0,1). Then

$$z_1 = v_1\sqrt{\frac{-2\ln r^2}{r^2}} \quad \text{and} \quad z_2 = v_2\sqrt{\frac{-2\ln r^2}{r^2}} \tag{33.7}$$

are independent numbers chosen from a normal distribution with mean 0 and variance 1. $z_i' = \mu + \sigma z_i$ distributes with mean μ and variance σ^2. A recent implementation of the fast algorithm of Leva Ref. 11 is in CERNLIB.

For a multivariate Gaussian with an $n \times n$ covariance matrix V, one can start by generating n independent Gaussian variables, $\{\eta_j\}$, with mean 0 and variance 1 as above. Then the new set $\{x_i\}$ is obtained as $x_i = \mu_i + \sum_j L_{ij}\eta_j$, where μ_i is the mean of x_i, and L_{ij} are the components of L, the unique lower triangular matrix that fulfils $V = LL^T$.

The matrix L can be easily computed by the following recursive relation (Cholesky's method):

$$L_{jj} = \left(V_{jj} - \sum_{k=1}^{j-1} L_{jk}^2 \right)^{1/2} , \tag{33.8a}$$

$$L_{ij} = \frac{V_{ij} - \sum_{k=1}^{j-1} L_{ik} L_{jk}}{L_{jj}} , \; j = 1, ..., n \, ; \; i = j+1, ..., n, \tag{33.8b}$$

where $V_{ij} = \rho_{ij} \sigma_i \sigma_j$ are the components of V. For $n = 2$ one has

$$L = \begin{pmatrix} \sigma_1 & 0 \\ \rho\sigma_2 & \sqrt{1-\rho^2}\,\sigma_2 \end{pmatrix} , \tag{33.9}$$

and therefore the correlated Gaussian variables are generated as $x_1 = \mu_1 + \sigma_1 \eta_1$, $x_2 = \mu_2 + \rho\sigma_2\eta_1 + \sqrt{1-\rho^2}\,\sigma_2\eta_2$.

33.4.5. $\chi^2(n)$ distribution :

For n even, generate $n/2$ uniform numbers u_i; then

$$y = -2 \ln \left(\prod_{i=1}^{n/2} u_i \right) \quad \text{is} \quad \chi^2(n) \; . \tag{33.10}$$

For n odd, generate $(n-1)/2$ uniform numbers u_i and one Gaussian z as in Sec. 33.4.4; then

$$y = -2 \ln \left(\prod_{i=1}^{(n-1)/2} u_i \right) + z^2 \quad \text{is} \quad \chi^2(n) \; . \tag{33.11}$$

For $n \gtrsim 30$ the much faster Gaussian approximation for the χ^2 may be preferable: generate z as in Sec. 33.4.4. If $z \geq -\sqrt{2n-1}$ use $y = \left[z + \sqrt{2n-1} \right]^2 /2$; otherwise reject.

33.4.6. *Gamma distribution* :

All of the following algorithms are given for $\lambda = 1$. For $\lambda \neq 1$, divide the resulting random number x by λ.

- If $k = 1$ (the *exponential* distribution), accept $x = -\ln u$. (See also Sec. 33.4.1.)

- If $0 < k < 1$, initialize with $v_1 = (e+k)/e$ (with $e = 2.71828...$ being the natural log base). Generate u_1, u_2. Define $v_2 = v_1 u_1$.

Case 1: $v_2 \leq 1$. Define $x = v_2^{1/k}$. If $u_2 \leq e^{-x}$, accept x and stop, else restart by generating new u_1, u_2.
Case 2: $v_2 > 1$. Define $x = -\ln([v_1 - v_2]/k)$. If $u_2 \leq x^{k-1}$, accept x and stop, else restart by generating new u_1, u_2. Note that, for $k < 1$, the probability density has a pole at $x = 0$, so that return values of zero due to underflow must be accepted or otherwise dealt with.

- Otherwise, if $k > 1$, initialize with $c = 3k - 0.75$. Generate u_1 and compute $v_1 = u_1(1 - u_1)$ and $v_2 = (u_1 - 0.5)\sqrt{c/v_1}$. If $x = k + v_2 - 1 \leq 0$, go back and generate new u_1; otherwise generate u_2 and compute $v_3 = 64v_1^3u_2^2$. If $v_3 \leq 1 - 2v_2^2/x$ **or** if $\ln v_3 \leq 2\{[k-1]\ln[x/(k-1)] - v_2\}$, accept x and stop; otherwise go back and generate new u_1.

33.4.7. *Binomial distribution* :

Begin with $k = 0$ and generate u uniform in $[0, 1)$. Compute $P_k = (1 - p)^n$ and store P_k into B. If $u \leq B$ accept $r_k = k$ and stop. Otherwise, increment k by one; compute the next P_k as $P_k \cdot (p/(1 - p)) \cdot (n - k)/(k + 1)$; add this to B. Again if $u \leq B$ accept $r_k = k$ and stop, otherwise iterate until a value is accepted. If $p > 1/2$ it will be more efficient to generate r from $f(r; n, q)$, *i.e.*, with p and q interchanged, and then set $r_k = n - r$.

33.4.8. *Poisson distribution* :

Iterate until a successful choice is made: Begin with $k = 1$ and set $A = 1$ to start. Generate u. Replace A with uA; if now $A < \exp(-\mu)$, where μ is the Poisson parameter, accept $n_k = k - 1$ and stop. Otherwise increment k by 1, generate a new u and repeat, always starting with the value of A left from the previous try. For large $\mu(\gtrsim 10)$ it may be satisfactory (and much faster) to approximate the Poisson distribution by a Gaussian distribution (see our Probability chapter, Sec. 31.4.3). Generate z from a Gaussian with zero mean and unit standard deviation; then use $x = \max(0, [\mu + z\sqrt{\mu} + 0.5])$ where [ ] signifies the greatest integer $\leq$ the expression. [12]

33.4.9. *Student's t* distribution :

For $n > 0$ degrees of freedom (n not necessarily integer), generate x from a Gaussian with mean 0 and $\sigma^2 = 1$ according to the method of 33.4.4. Next generate y, an independent gamma random variate, according to 33.4.6 with $\lambda = 1/2$ and $k = n/2$. Then $z = x/(\sqrt{y}/n)$ is distributed as a t with n degrees of freedom.

For the special case $n = 1$, the Breit-Wigner distribution, generate u_1 and u_2; set $v_1 = 2u_1 - 1$ and $v_2 = 2u_2 - 1$. If $v_1^2 + v_2^2 \leq 1$ accept $z = v_1/v_2$ as a Breit-Wigner distribution with unit area, center at 0.0, and FWHM 2.0. Otherwise start over. For center M_0 and FWHM Γ, use $W = z\Gamma/2 + M_0$.

References:

1. F. James, Comp. Phys. Comm. **79** 111 (1994), based on M. Lüscher, Comp. Phys. Comm. **79** 100 (1994). This generator is available as the CERNLIB routine V115, RANLUX.

2. G. Marsaglia, A. Zaman, and W.W. Tsang, *Towards a Universal Random Number Generator*, Supercomputer Computations Research Institute, Florida State University technical report FSU-SCRI-87-50 (1987). This generator is available as the CERNLIB routine V113, RANMAR, by F. Carminati and F. James.
3. Much of DIEHARD is described in: G. Marsaglia, *A Current View of Random Number Generators*, keynote address, *Computer Science and Statistics: 16th Symposium on the Interface*, Elsevier (1985).
4. Newer generators with periods even longer than the lagged-Fibonacci based generator are described in G. Marsaglia and A. Zaman, *Some Portable Very-Long-Period Random Number Generators*, Compt. Phys. **8**, 117 (1994). The Numerical Recipes generator **ran2** [W.H. Press and S.A. Teukolsky, *Portable Random Number Generators*, Compt. Phys. **6**, 521 (1992)] is also known to pass the DIEHARD tests.
5. W.H. Press *et al.*, *Numerical Recipes* (Cambridge University Press, New York, 1986).
6. J.H. Ahrens and U. Dieter, Computing **12**, 223 (1974).
7. R.Y. Rubinstein, *Simulation and the Monte Carlo Method* (John Wiley and Sons, Inc., New York, 1981).
8. C.J. Everett and E.D. Cashwell, *A Third Monte Carlo Sampler*, Los Alamos report LA-9721-MS (1983).
9. L. Devroye, *Non-Uniform Random Variate Generation* (Springer-Verlag, New York, 1986).
10. Ch. Walck, *Random Number Generation*, University of Stockholm Physics Department Report 1987-10-20 (Vers. 3.0).
11. J.L. Leva, ACM Trans. Math. Softw. **18** 449 (1992). This generator has been implemented by F. James in the CERNLIB routine V120, RNORML.
12. This generator has been implemented by D. Drijard and K. Kölbig in the CERNLIB routine V136, RNPSSN.

35. CLEBSCH-GORDAN COEFFICIENTS, SPHERICAL HARMONICS, AND *d* FUNCTIONS

Note: A square-root sign is to be understood over *every* coefficient, *e.g.*, for $-8/15$ read $-\sqrt{8/15}$.

Notation:

J	J	...
M	M	...
m_1	m_2	Coefficients
m_1	m_2	
.	.	
.	.	
.	.	

1/2×1/2

	1		
+1/2 +1/2	1	1	0
		0	0

+1/2 −1/2	1/2	1/2	1
−1/2 +1/2	1/2	−1/2	−1
	−1/2	−1/2	1

$$Y_1^0 = \sqrt{\frac{3}{4\pi}}\cos\theta$$

$$Y_1^1 = -\sqrt{\frac{3}{8\pi}}\sin\theta\, e^{i\phi}$$

$$Y_2^0 = \sqrt{\frac{5}{4\pi}}\left(\frac{3}{2}\cos^2\theta - \frac{1}{2}\right)$$

$$Y_2^1 = -\sqrt{\frac{15}{8\pi}}\sin\theta\cos\theta\, e^{i\phi}$$

$$Y_2^2 = \frac{1}{4}\sqrt{\frac{15}{2\pi}}\sin^2\theta\, e^{2i\phi}$$

2×1/2

	5/2		
+2 +1/2	1	5/2	3/2
		+3/2	+3/2

+2 −1/2	1/5	4/5	5/2	3/2
+1 +1/2	4/5	−1/5	+1/2	+1/2

+1 −1/2	2/5	3/5	5/2	3/2
0 +1/2	3/5	−2/5	−1/2	−1/2

0 −1/2	3/5	2/5	5/2	3/2
−1 +1/2	2/5	−3/5	−3/2	−3/2

−1 −1/2	4/5	1/5	5/2
−2 +1/2	1/5	−4/5	−5/2
	−2	−1/2	1

1×1/2

	3/2		
+1 +1/2	1	3/2	1/2
		+1/2	+1/2

+1 −1/2	1/3	2/3	3/2	1/2
0 +1/2	2/3	−1/3	−1/2	−1/2

0 −1/2	2/3	1/3	3/2
−1 +1/2	1/3	−2/3	−3/2
	−1	−1/2	1

3/2×1/2

	2		
+3/2 +1/2	1	2	1
		+1	+1

+3/2 −1/2	1/4	3/4	2	1
+1/2 +1/2	3/4	−1/4	0	0

+1/2 −1/2	1/2	1/2	2	1
−1/2 +1/2	1/2	−1/2	−1	−1

−1/2 −1/2	3/4	1/4	2
−3/2 +1/2	1/4	−3/4	−2
	−3/2	−1/2	1

2×1

	3		
+2 +1	1	3	2
		+2	+2

+2 0	1/3	2/3	3	2	1
+1 +1	2/3	−1/3	+1	+1	+1

+2 −1	1/15	1/3	3/5	
+1 0	8/15	1/6	−3/10	
0 +1	2/5	−1/2	1/10	

3	2	1	
0	0	0	
+1 −1	1/5	1/2	3/10
0 0	3/5	0	−2/5
−1 +1	1/5	−1/2	3/10

3/2×1

	5/2		
+3/2 +1	1	5/2	3/2
		+3/2	+3/2

+3/2 0	2/5	3/5	5/2	3/2	1/2
+1/2 +1	3/5	−2/5	+1/2	+1/2	+1/2

+3/2 −1	1/10	2/5	1/2	
+1/2 0	3/5	1/15	−1/3	
−1/2 +1	3/10	−8/15	1/6	

5/2	3/2	1/2	
−1/2	−1/2	−1/2	
+1/2 −1	3/10	8/15	1/6
−1/2 0	3/5	−1/15	−1/3
−3/2 +1	1/10	−2/5	1/2

1×1

	2		
+1 +1	1	2	1
		+1	+1

+1 0	1/2	1/2	2	1	0
0 +1	1/2	−1/2	0	0	0

+1 −1	1/6	1/2	1/3	
0 0	2/3	0	−1/3	
−1 +1	1/6	−1/2	1/3	

2	1		
0	0		
+1 −1	1/5	1/2	3/10
0 0	3/5	0	−2/5
−1 +1	1/5	−1/2	3/10

3	2	1	
−1	−1	−1	
0 −1	2/5	1/2	1/10
−1 0	8/15	−1/6	−3/10
−2 +1	1/15	−1/3	3/5

3	2		
−2	−2		
−1 −1	2/3	1/3	3
−2 0	1/3	−2/3	−3
	−2	−1	1

5/2	3/2	1/2	
−3/2	−3/2	−3/2	
−1/2 −1	3/5	2/5	5/2
−3/2 0	2/5	−3/5	−5/2
	−3/2	−1	1

$$Y_\ell^{-m} = (-1)^m Y_\ell^{m*}$$

$$d^\ell_{m,0} = \sqrt{\frac{4\pi}{2\ell+1}}\, Y_\ell^m\, e^{-im\phi}$$

$$\langle j_1 j_2 m_1 m_2 | j_1 j_2 J M\rangle = (-1)^{J-j_1-j_2}\langle j_2 j_1 m_2 m_1 | j_2 j_1 J M\rangle$$

$$d^j_{m',m} = (-1)^{m-m'} d^j_{m,m'} = d^j_{-m,-m'}$$

3/2×3/2

	3		
+3/2 +3/2	1	3	2
		+2	+2

+3/2 +1/2	1/2	1/2	3	2	1
+1/2 +3/2	1/2	−1/2	+1	+1	+1

+3/2 −1/2	1/5	1/2	3/10	
+1/2 +1/2	3/5	0	−2/5	
−1/2 +3/2	1/5	−1/2	3/10	

3	2	1	0	
0	0	0	0	
+3/2 −3/2	1/20	1/4	9/20	1/4
+1/2 −1/2	9/20	1/4	−1/20	−1/4
−1/2 +1/2	9/20	−1/4	−1/20	1/4
−3/2 +3/2	1/20	−1/4	9/20	−1/4

3	2	1		
−1	−1	−1		
+1/2 −3/2	1/5	1/2	3/10	
−1/2 −1/2	3/5	0	−2/5	
−3/2 +1/2	1/5	−1/2	3/10	

3	2			
−2	−2			
−1/2	−3/2	1/2	1/2	3
−3/2	−1/2	1/2	−1/2	−3
	−3/2	−3/2		1

2×3/2

	7/2		
+2 +3/2	1	7/2	5/2
		+5/2	+5/2

+2 +1/2	3/7	4/7	7/2	5/2	3/2
+1 +3/2	4/7	−3/7	+3/2	+3/2	+3/2

+2 −1/2	1/7	16/35	2/5	
+1 +1/2	4/7	1/35	−2/5	
0 +3/2	2/7	−18/35	1/5	

7/2	5/2	3/2	1/2	
+1/2	+1/2	+1/2	+1/2	
+2 −3/2	1/35	6/35	2/5	2/5
+1 −1/2	12/35	5/14	0	−3/10
0 +1/2	18/35	−3/35	−1/5	1/5
−1 +3/2	4/35	−27/70	2/5	−1/10

7/2	5/2	3/2	1/2	
−1/2	−1/2	−1/2	−1/2	
+1 −3/2	4/35	27/70	2/5	1/10
0 −1/2	18/35	3/35	−1/5	−1/5
−1 +1/2	12/35	−5/14	0	3/10
−2 +3/2	1/35	−6/35	2/5	−2/5

7/2	5/2	3/2		
−3/2	−3/2	−3/2		
0 −3/2	2/7	18/35	1/5	
−1 −1/2	4/7	−1/35	−2/5	
−2 +1/2	1/7	−16/35	2/5	

7/2	5/2		
−5/2	−5/2		
−1 −3/2	4/7	3/7	7/2
−2 −1/2	3/7	−4/7	−7/2
	−2	−3/2	1

2×2

	4		
+2 +2	1	4	3
		+3	+3

+2 +1	1/2	1/2	4	3	2
+1 +2	1/2	−1/2	+2	+2	+2

+2 0	3/14	1/2	2/7	
+1 +1	4/7	0	−3/7	
0 +2	3/14	−1/2	2/7	

4	3	2	1	
+1	+1	+1	+1	
+2 −1	1/14	3/10	3/7	1/5
+1 0	3/7	1/5	−1/14	−3/10
0 +1	3/7	−1/5	−1/14	3/10
−1 +2	1/14	−3/10	3/7	−1/5

4	3	2	1	0
0	0	0	0	0
+2 −2	1/70	1/10	2/7	2/5	1/5
+1 −1	8/35	2/5	1/14	−1/10	−1/5
0 0	18/35	0	−2/7	0	1/5
−1 +1	8/35	−2/5	1/14	1/10	−1/5
−2 +2	1/70	−1/10	2/7	−2/5	1/5

4	3	2	1	
−1	−1	−1	−1	
+1 −2	1/14	3/10	3/7	1/5
0 −1	3/7	1/5	−1/14	−3/10
−1 0	3/7	−1/5	−1/14	3/10
−2 +1	1/14	−3/10	3/7	−1/5

4	3	2		
−2	−2	−2		
0 −2	3/14	1/2	2/7	
−1 −1	4/7	0	−3/7	
−2 0	3/14	−1/2	2/7	

4	3			
−3	−3			
−1 −2	1/2	1/2	4	
−2 −1	1/2	−1/2	−4	
	−2	−2	1	

$$d^{3/2}_{3/2,3/2} = \frac{1+\cos\theta}{2}\cos\frac{\theta}{2}$$

$$d^{3/2}_{3/2,1/2} = -\sqrt{3}\frac{1+\cos\theta}{2}\sin\frac{\theta}{2} \qquad d^2_{2,2} = \left(\frac{1+\cos\theta}{2}\right)^2$$

$$d^{3/2}_{3/2,-1/2} = \sqrt{3}\frac{1-\cos\theta}{2}\cos\frac{\theta}{2} \qquad d^2_{2,1} = -\frac{1+\cos\theta}{2}\sin\theta$$

$$d^{3/2}_{3/2,-3/2} = -\frac{1-\cos\theta}{2}\sin\frac{\theta}{2} \qquad d^2_{2,0} = \frac{\sqrt{6}}{4}\sin^2\theta \qquad d^2_{1,1} = \frac{1+\cos\theta}{2}(2\cos\theta - 1)$$

$$d^{3/2}_{1/2,1/2} = \frac{3\cos\theta - 1}{2}\cos\frac{\theta}{2} \qquad d^2_{2,-1} = -\frac{1-\cos\theta}{2}\sin\theta \qquad d^2_{1,0} = -\sqrt{\frac{3}{2}}\sin\theta\cos\theta$$

$$d^{3/2}_{1/2,-1/2} = -\frac{3\cos\theta + 1}{2}\sin\frac{\theta}{2} \qquad d^2_{2,-2} = \left(\frac{1-\cos\theta}{2}\right)^2 \qquad d^2_{1,-1} = \frac{1-\cos\theta}{2}(2\cos\theta + 1) \qquad d^2_{0,0} = \left(\frac{3}{2}\cos^2\theta - \frac{1}{2}\right)$$

$$d^1_{0,0} = \cos\theta \qquad d^{1/2}_{1/2,1/2} = \cos\frac{\theta}{2} \qquad d^1_{1,1} = \frac{1+\cos\theta}{2}$$

$$d^{1/2}_{1/2,-1/2} = -\sin\frac{\theta}{2} \qquad d^1_{1,0} = -\frac{\sin\theta}{\sqrt{2}}$$

$$d^1_{1,-1} = \frac{1-\cos\theta}{2}$$

Figure 35.1: The sign convention is that of Wigner (*Group Theory*, Academic Press, New York, 1959), also used by Condon and Shortley (*The Theory of Atomic Spectra*, Cambridge Univ. Press, New York, 1953), Rose (*Elementary Theory of Angular Momentum*, Wiley, New York, 1957), and Cohen (*Tables of the Clebsch-Gordan Coefficients*, North American Rockwell Science Center, Thousand Oaks, Calif., 1974). The coefficients here have been calculated using computer programs written independently by Cohen and at LBNL.

36. SU(3) ISOSCALAR FACTORS AND REPRESENTATION MATRICES

Written by R.L. Kelly (LBNL).

The most commonly used SU(3) isoscalar factors, corresponding to the singlet, octet, and decuplet content of $8 \otimes 8$ and $10 \otimes 8$, are shown at the right. The notation uses particle names to identify the coefficients, so that the pattern of relative couplings may be seen at a glance. We illustrate the use of the coefficients below. See J.J de Swart, Rev. Mod. Phys. **35**, 916 (1963) for detailed explanations and phase conventions.

A $\sqrt{\ }$ is to be understood over every integer in the matrices; the exponent 1/2 on each matrix is a reminder of this. For example, the $\Xi \to \Omega K$ element of the $10 \to 10 \otimes 8$ matrix is $-\sqrt{6}/\sqrt{24} = -1/2$.

Intramultiplet relative decay strengths may be read directly from the matrices. For example, in decuplet → octet + octet decays, the ratio of $\Omega^* \to \Xi\overline{K}$ and $\Delta \to N\pi$ partial widths is, from the $10 \to 8 \times 8$ matrix,

$$\frac{\Gamma\left(\Omega^* \to \Xi\overline{K}\right)}{\Gamma\left(\Delta \to N\pi\right)} = \frac{12}{6} \times \text{ (phase space factors)} \ . \tag{36.1}$$

Including isospin Clebsch-Gordan coefficients, we obtain, *e.g.*,

$$\frac{\Gamma(\Omega^{*-} \to \Xi^0 K^-)}{\Gamma(\Delta^+ \to p\,\pi^0)} = \frac{1/2}{2/3} \times \frac{12}{6} \times p.s.f. = \frac{3}{2} \times p.s.f. \tag{36.2}$$

Partial widths for $8 \to 8 \otimes 8$ involve a linear superposition of 8_1 (symmetric) and 8_2 (antisymmetric) couplings. For example,

$$\Gamma(\Xi^* \to \Xi\pi) \sim \left(-\sqrt{\frac{9}{20}}\, g_1 + \sqrt{\frac{3}{12}}\, g_2\right)^2 \ . \tag{36.3}$$

The relations between g_1 and g_2 (with de Swart's normalization) and the standard D and F couplings that appear in the interaction Lagrangian,

$$\mathscr{L} = -\sqrt{2}\, D\, Tr\left(\{\overline{B}, B\}M\right) + \sqrt{2}\, F\, Tr\left([\overline{B}, B]\, M\right) \ , \tag{36.4}$$

where $[\overline{B}, B] \equiv \overline{B}B - B\overline{B}$ and $\{\overline{B}, B\} \equiv \overline{B}B + B\overline{B}$, are

$$D = \frac{\sqrt{30}}{40}\, g_1 \ , \qquad F = \frac{\sqrt{6}}{24}\, g_2 \ . \tag{36.5}$$

Thus, for example,

$$\Gamma(\Xi^* \to \Xi\pi) \sim (F - D)^2 \sim (1 - 2\alpha)^2 \ , \tag{36.6}$$

where $\alpha \equiv F/(D + F)$. (This definition of α is de Swart's. The alternative $D/(D + F)$, due to Gell-Mann, is also used.)

The generators of SU(3) transformations, λ_a ($a = 1, 8$), are 3×3 matrices that obey the following commutation and anticommutation relationships:

$$[\lambda_a,\ \lambda_b] \equiv \lambda_a\lambda_b - \lambda_b\lambda_a = 2if_{abc}\lambda_c \tag{36.7}$$

$$\{\lambda_a,\ \lambda_b\} \equiv \lambda_a\lambda_b + \lambda_b\lambda_a = \frac{4}{3}\delta_{ab}I + 2d_{abc}\lambda_c \ , \tag{36.8}$$

where I is the 3×3 identity matrix, and δ_{ab} is the Kronecker delta symbol. The f_{abc} are odd under the permutation of any pair of indices, while the d_{abc} are even. The nonzero values are

$\mathbf{1} \to \mathbf{8} \otimes \mathbf{8}$

$$\begin{pmatrix} \Lambda \end{pmatrix} \quad \to \begin{pmatrix} N\overline{K} & \Sigma\pi & \Lambda\eta & \Xi K \end{pmatrix} \qquad = \frac{1}{\sqrt{8}} \begin{pmatrix} 2 & 3 & -1 & -2 \end{pmatrix}^{1/2}$$

$\mathbf{8_1} \to \mathbf{8} \otimes \mathbf{8}$

$$\begin{pmatrix} N \\ \Sigma \\ \Lambda \\ \Xi \end{pmatrix} \to \begin{pmatrix} N\pi & N\eta & \Sigma K & \Lambda K \\ N\overline{K} & \Sigma\pi & \Lambda\pi & \Sigma\eta & \Xi K \\ N\overline{K} & \Sigma\pi & \Lambda\eta & \Xi K \\ \Sigma\overline{K} & \Lambda\overline{K} & \Xi\pi & \Xi\eta \end{pmatrix} = \frac{1}{\sqrt{20}} \begin{pmatrix} 9 & -1 & -9 & -1 \\ -6 & 0 & 4 & 4 & -6 \\ 2 & -12 & -4 & -2 \\ 9 & -1 & -9 & -1 \end{pmatrix}^{1/2}$$

$\mathbf{8_2} \to \mathbf{8} \otimes \mathbf{8}$

$$\begin{pmatrix} N \\ \Sigma \\ \Lambda \\ \Xi \end{pmatrix} \to \begin{pmatrix} N\pi & N\eta & \Sigma K & \Lambda K \\ N\overline{K} & \Sigma\pi & \Lambda\pi & \Sigma\eta & \Xi K \\ N\overline{K} & \Sigma\pi & \Lambda\eta & \Xi K \\ \Sigma\overline{K} & \Lambda\overline{K} & \Xi\pi & \Xi\eta \end{pmatrix} = \frac{1}{\sqrt{12}} \begin{pmatrix} 3 & 3 & 3 & -3 \\ 2 & 8 & 0 & 0 & -2 \\ 6 & 0 & 0 & 6 \\ 3 & 3 & 3 & -3 \end{pmatrix}^{1/2}$$

$\mathbf{10} \to \mathbf{8} \otimes \mathbf{8}$

$$\begin{pmatrix} \Delta \\ \Sigma \\ \Xi \\ \Omega \end{pmatrix} \to \begin{pmatrix} N\pi & \Sigma K \\ N\overline{K} & \Sigma\pi & \Lambda\pi & \Sigma\eta & \Xi K \\ \Sigma\overline{K} & \Lambda\overline{K} & \Xi\pi & \Xi\eta \\ \Xi\overline{K} \end{pmatrix} = \frac{1}{\sqrt{12}} \begin{pmatrix} -6 & 6 \\ -2 & 2 & -3 & 3 & 2 \\ 3 & -3 & 3 & 3 \\ 12 \end{pmatrix}^{1/2}$$

$\mathbf{8} \to \mathbf{10} \otimes \mathbf{8}$

$$\begin{pmatrix} N \\ \Sigma \\ \Lambda \\ \Xi \end{pmatrix} \to \begin{pmatrix} \Delta\pi & \Sigma K \\ \Delta\overline{K} & \Sigma\pi & \Sigma\eta & \Xi K \\ \Sigma\pi & \Xi K \\ \Sigma\overline{K} & \Xi\pi & \Xi\eta & \Omega K \end{pmatrix} \qquad = \frac{1}{\sqrt{15}} \begin{pmatrix} -12 & 3 \\ 8 & -2 & -3 & 2 \\ -9 & 6 \\ 3 & -3 & -3 & 6 \end{pmatrix}^{1/2}$$

$\mathbf{10} \to \mathbf{10} \otimes \mathbf{8}$

$$\begin{pmatrix} \Delta \\ \Sigma \\ \Xi \\ \Omega \end{pmatrix} \to \begin{pmatrix} \Delta\pi & \Delta\eta & \Sigma K \\ \Delta\overline{K} & \Sigma\pi & \Sigma\eta & \Xi K \\ \Sigma\overline{K} & \Xi\pi & \Xi\eta & \Omega K \\ \Xi\overline{K} & \Omega\eta \end{pmatrix} \qquad = \frac{1}{\sqrt{24}} \begin{pmatrix} 15 & 3 & -6 \\ 8 & 8 & 0 & -8 \\ 12 & 3 & -3 & -6 \\ 12 & -12 \end{pmatrix}^{1/2}$$

abc	f_{abc}
123	1
147	1/2
156	−1/2
246	1/2
257	1/2
345	1/2
367	−1/2
458	$\sqrt{3}/2$
678	$\sqrt{3}/2$

abc	d_{abc}
118	$1/\sqrt{3}$
146	1/2
157	1/2
228	$1/\sqrt{3}$
247	−1/2
256	1/2
338	$1/\sqrt{3}$
344	1/2

abc	d_{abc}
355	1/2
366	−1/2
377	−1/2
448	$-1/(2\sqrt{3})$
558	$-1/(2\sqrt{3})$
668	$-1/(2\sqrt{3})$
778	$-1/(2\sqrt{3})$
888	$-1/\sqrt{3}$

The λ_a's are

$$\lambda_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \lambda_2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \lambda_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$\lambda_4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \quad \lambda_5 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix} \quad \lambda_6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

$$\lambda_7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix} \quad \lambda_8 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}$$

Equation (36.7) defines the Lie algebra of SU(3). A general d-dimensional representation is given by a set of $d \times d$ matrices satisfying Eq. (36.7) with the f_{abc} given above. Equation (36.8) is specific to the defining 3-dimensional representation.

37. SU(n) MULTIPLETS AND YOUNG DIAGRAMS

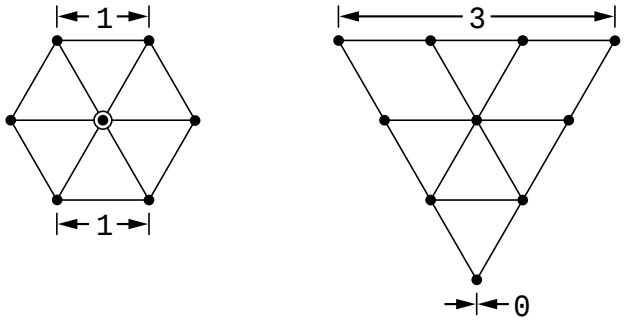
Written by C.G. Wohl (LBNL).

This note tells (1) how SU(n) particle multiplets are identified or labeled, (2) how to find the number of particles in a multiplet from its label, (3) how to draw the Young diagram for a multiplet, and (4) how to use Young diagrams to determine the overall multiplet structure of a composite system, such as a 3-quark or a meson-baryon system.

In much of the literature, the word "representation" is used where we use "multiplet," and "tableau" is used where we use "diagram."

37.1. Multiplet labels

An SU(n) multiplet is uniquely identified by a string of ($n-1$) nonnegative integers: $(\alpha, \beta, \gamma, \ldots)$. Any such set of integers specifies a multiplet. For an SU(2) multiplet such as an isospin multiplet, the single integer α is the number of *steps* from one end of the multiplet to the other (*i.e.*, it is one fewer than the number of particles in the multiplet). In SU(3), the two integers α and β are the numbers of steps across the top and bottom levels of the multiplet diagram. Thus the labels for the SU(3) octet and decuplet



are (1,1) and (3,0). For larger n, the interpretation of the integers in terms of the geometry of the multiplets, which exist in an ($n-1$)-dimensional space, is not so readily apparent.

The label for the SU(n) singlet is $(0, 0, \ldots, 0)$. In a flavor SU(n), the n quarks together form a $(1, 0, \ldots, 0)$ multiplet, and the n antiquarks belong to a $(0, \ldots, 0, 1)$ multiplet. These two multiplets are *conjugate* to one another, which means their labels are related by $(\alpha, \beta, \ldots) \leftrightarrow (\ldots, \beta, \alpha)$.

37.2. Number of particles

The number of particles in a multiplet, $N = N(\alpha, \beta, \ldots)$, is given as follows (note the pattern of the equations).

In SU(2), $N = N(\alpha)$ is

$$N = \frac{(\alpha+1)}{1} \ . \tag{37.1}$$

In SU(3), $N = N(\alpha, \beta)$ is

$$N = \frac{(\alpha+1)}{1} \cdot \frac{(\beta+1)}{1} \cdot \frac{(\alpha+\beta+2)}{2} \ . \tag{37.2}$$

In SU(4), $N = N(\alpha, \beta, \gamma)$ is

$$N = \frac{(\alpha+1)}{1} \cdot \frac{(\beta+1)}{1} \cdot \frac{(\gamma+1)}{1} \cdot \frac{(\alpha+\beta+2)}{2} \cdot \frac{(\beta+\gamma+2)}{2} \cdot \frac{(\alpha+\beta+\gamma+3)}{3} \ . \tag{37.3}$$

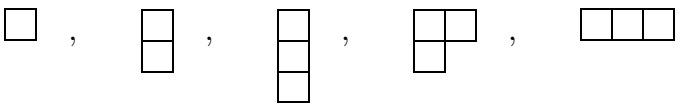
Note that in Eq. (37.3) there is no factor with $(\alpha+\gamma+2)$: only a *consecutive* sequence of the label integers appears in any factor. One more example should make the pattern clear for any SU(n). In SU(5), $N = N(\alpha, \beta, \gamma, \delta)$ is

$$\begin{aligned} N = {} & \frac{(\alpha+1)}{1} \cdot \frac{(\beta+1)}{1} \cdot \frac{(\gamma+1)}{1} \cdot \frac{(\delta+1)}{1} \cdot \frac{(\alpha+\beta+2)}{2} \cdot \frac{(\beta+\gamma+2)}{2} \\ & \times \frac{(\gamma+\delta+2)}{2} \cdot \frac{(\alpha+\beta+\gamma+3)}{3} \cdot \frac{(\beta+\gamma+\delta+3)}{3} \cdot \frac{(\alpha+\beta+\gamma+\delta+4)}{4} \ . \end{aligned} \tag{37.4}$$

From the symmetry of these equations, it is clear that multiplets that are conjugate to one another have the same number of particles, but so can other multiplets. For example, the SU(4) multiplets (3,0,0) and (1,1,0) each have 20 particles. Try the equations and see.

37.3. Young diagrams

A Young diagram consists of an array of boxes (or some other symbol) arranged in one or more *left-justified* rows, with each row being *at least as long* as the row beneath. The correspondence between a diagram and a multiplet label is: The top row juts out α boxes to the right past the end of the second row, the second row juts out β boxes to the right past the end of the third row, *etc.* A diagram in SU(n) has at most n rows. There can be any number of "completed" columns of n boxes buttressing the left of a diagram; these don't affect the label. Thus in SU(3) the diagrams



represent the multiplets (1,0), (0,1), (0,0), (1,1), and (3,0). In any SU(n), the quark multiplet is represented by a single box, the antiquark multiplet by a column of $(n-1)$ boxes, and a singlet by a completed column of n boxes.

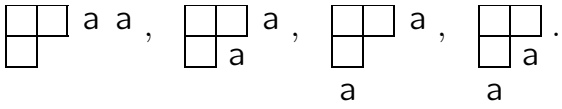
37.4. Coupling multiplets together

The following recipe tells how to find the multiplets that occur in coupling two multiplets together. To couple together more than two multiplets, first couple two, then couple a third with each of the multiplets obtained from the first two, *etc.*

First a definition: A sequence of the letters $a, b, c, \ldots$ is *admissible* if at any point in the sequence at least as many a's have occurred as b's, at least as many b's have occurred as c's, *etc.* Thus *abcd* and *aabcb* are admissible sequences and *abb* and *acb* are not. Now the recipe:

(a) Draw the Young diagrams for the two multiplets, but in one of the diagrams replace the boxes in the first row with a's, the boxes in the second row with b's, *etc.* Thus, to couple two SU(3) octets (such as the π-meson octet and the baryon octet), we start with  and  . The *unlettered* diagram forms the *upper left-hand corner* of all the enlarged diagrams constructed below.

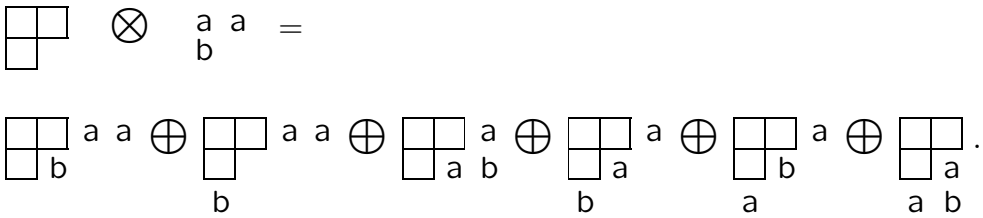
(b) Add the a's from the lettered diagram to the right-hand ends of the rows of the unlettered diagram to form all possible legitimate Young diagrams that have no more than one a per column. In general, there will be several distinct diagrams, and all the a's appear in each diagram. At this stage, for the coupling of the two SU(3) octets, we have:



(c) Use the b's to further enlarge the diagrams already obtained, subject to the same rules. Then throw away any diagram in which the full sequence of letters formed by reading *right to left* in the first row, then the second row, *etc.*, is not admissible.

(d) Proceed as in (c) with the c's (if any), *etc.*

The final result of the coupling of the two SU(3) octets is:



Here only the diagrams with admissible sequences of a's and b's and with fewer than four rows (since $n = 3$) have been kept. In terms of multiplet labels, the above may be written

$$(1,1) \otimes (1,1) = (2,2) \oplus (3,0) \oplus (0,3) \oplus (1,1) \oplus (1,1) \oplus (0,0) \ .$$

In terms of numbers of particles, it may be written

$$\mathbf{8} \otimes \mathbf{8} = \mathbf{27} \oplus \mathbf{10} \oplus \overline{\mathbf{10}} \oplus \mathbf{8} \oplus \mathbf{8} \oplus \mathbf{1} \ .$$

The product of the numbers on the left here is equal to the sum on the right, a useful check. (See also Sec. 14 on the Quark Model.)

38. KINEMATICS

Revised January 2000 by J.D. Jackson (LBNL).

Throughout this section units are used in which $\hbar = c = 1$. The following conversions are useful: $\hbar c = 197.3$ MeV fm, $(\hbar c)^2 = 0.3894$ $(\mathrm{GeV})^2$ mb.

38.1. Lorentz transformations

The energy E and 3-momentum $\boldsymbol{p}$ of a particle of mass m form a 4-vector $p = (E, \boldsymbol{p})$ whose square $p^2 \equiv E^2 - |\boldsymbol{p}|^2 = m^2$. The velocity of the particle is $\boldsymbol{\beta} = \boldsymbol{p}/E$. The energy and momentum $(E^*, \boldsymbol{p}^*)$ viewed from a frame moving with velocity $\boldsymbol{\beta}_f$ are given by

$$\begin{pmatrix} E^* \\ p^*_\parallel \end{pmatrix} = \begin{pmatrix} \gamma_f & -\gamma_f \beta_f \\ -\gamma_f \beta_f & \gamma_f \end{pmatrix} \begin{pmatrix} E \\ p_\parallel \end{pmatrix} , \quad p^*_T = p_T , \tag{38.1}$$

where $\gamma_f = (1 - \beta_f^2)^{-1/2}$ and p_T ($p_\parallel$) are the components of $\boldsymbol{p}$ perpendicular (parallel) to $\boldsymbol{\beta}_f$. Other 4-vectors, such as the space-time coordinates of events, of course transform in the same way. The scalar product of two 4-momenta $p_1 \cdot p_2 = E_1 E_2 - \boldsymbol{p}_1 \cdot \boldsymbol{p}_2$ is invariant (frame independent).

38.2. Center-of-mass energy and momentum

In the collision of two particles of masses m_1 and m_2 the total center-of-mass energy can be expressed in the Lorentz-invariant form

$$\begin{aligned} E_{\mathrm{cm}} &= \left[(E_1 + E_2)^2 - (\boldsymbol{p}_1 + \boldsymbol{p}_2)^2 \right]^{1/2} , \\ &= \left[m_1^2 + m_2^2 + 2E_1 E_2 (1 - \beta_1 \beta_2 \cos\theta) \right]^{1/2} , \end{aligned} \tag{38.2}$$

where θ is the angle between the particles. In the frame where one particle (of mass m_2) is at rest (lab frame),

$$E_{\mathrm{cm}} = (m_1^2 + m_2^2 + 2E_{1\,\mathrm{lab}}\, m_2)^{1/2} . \tag{38.3}$$

The velocity of the center-of-mass in the lab frame is

$$\boldsymbol{\beta}_{\mathrm{cm}} = \boldsymbol{p}_{\mathrm{lab}}/(E_{1\,\mathrm{lab}} + m_2) , \tag{38.4}$$

where $\boldsymbol{p}_{\mathrm{lab}} \equiv \boldsymbol{p}_{1\,\mathrm{lab}}$ and

$$\gamma_{\mathrm{cm}} = (E_{1\,\mathrm{lab}} + m_2)/E_{\mathrm{cm}} . \tag{38.5}$$

The c.m. momenta of particles 1 and 2 are of magnitude

$$p_{\mathrm{cm}} = p_{\mathrm{lab}} \frac{m_2}{E_{\mathrm{cm}}} . \tag{38.6}$$

For example, if a 0.80 GeV/c kaon beam is incident on a proton target, the center of mass energy is 1.699 GeV and the center of mass momentum of either particle is 0.442 GeV/c. It is also useful to note that

$$E_{\mathrm{cm}}\, dE_{\mathrm{cm}} = m_2\, dE_{1\,\mathrm{lab}} = m_2\, \beta_{1\,\mathrm{lab}}\, dp_{\mathrm{lab}} . \tag{38.7}$$

38.3. Lorentz-invariant amplitudes

The matrix elements for a scattering or decay process are written in terms of an invariant amplitude $-i\mathscr{M}$. As an example, the S-matrix for $2 \to 2$ scattering is related to $\mathscr{M}$ by

$$\langle p_1' p_2' \,|S|\, p_1 p_2 \rangle = I - i(2\pi)^4\, \delta^4(p_1 + p_2 - p_1' - p_2') \times \frac{\mathscr{M}(p_1,\, p_2;\, p_1',\, p_2')}{(2E_1)^{1/2}\,(2E_2)^{1/2}\,(2E_1')^{1/2}\,(2E_2')^{1/2}}\ . \tag{38.8}$$

The state normalization is such that

$$\langle p'|p\rangle = (2\pi)^3 \delta^3(\boldsymbol{p} - \boldsymbol{p'})\ . \tag{38.9}$$

38.4. Particle decays

The partial decay rate of a particle of mass M into n bodies in its rest frame is given in terms of the Lorentz-invariant matrix element $\mathscr{M}$ by

$$d\Gamma = \frac{(2\pi)^4}{2M}\, |\mathscr{M}|^2\, d\Phi_n\,(P;\, p_1,\, \ldots,\, p_n), \tag{38.10}$$

where $d\Phi_n$ is an element of n-body phase space given by

$$d\Phi_n(P;\, p_1,\, \ldots,\, p_n) = \delta^4\,(P - \sum_{i=1}^{n} p_i)\ \prod_{i=1}^{n}\ \frac{d^3 p_i}{(2\pi)^3 2E_i}\ . \tag{38.11}$$

This phase space can be generated recursively, viz.

$$d\Phi_n(P;\, p_1,\, \ldots,\, p_n) = d\Phi_j(q;\, p_1,\, \ldots,\, p_j) \times\ d\Phi_{n-j+1}\,(P;\, q,\, p_{i+1},\, \ldots,\, p_n)(2\pi)^3 dq^2\ , \tag{38.12}$$

where $q^2 = (\sum_{i=1}^{j} E_i)^2 - \left|\sum_{i=1}^{j} \boldsymbol{p}_i\right|^2$. This form is particularly useful in the case where a particle decays into another particle that subsequently decays.

38.4.1. ***Survival probability*** : If a particle of mass M has mean proper lifetime τ $(= 1/\Gamma)$ and has momentum $(E, \boldsymbol{p})$, then the probability that it lives for a time t_0 or greater before decaying is given by

$$P(t_0) = e^{-t_0\,\Gamma/\gamma} = e^{-Mt_0\,\Gamma/E}\ , \tag{38.13}$$

and the probability that it travels a distance x_0 or greater is

$$P(x_0) = e^{-Mx_0\,\Gamma/|\boldsymbol{p}|}\ . \tag{38.14}$$

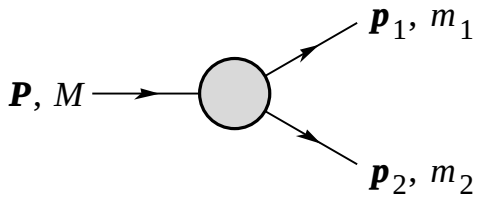


Figure 38.1: Definitions of variables for two-body decays.

38.4.2. ***Two-body decays*** :

In the rest frame of a particle of mass M, decaying into 2 particles labeled 1 and 2,

$$E_1 = \frac{M^2 - m_2^2 + m_1^2}{2M} \ , \tag{38.15}$$

$$|\boldsymbol{p}_1| = |\boldsymbol{p}_2|$$

$$= \frac{\left[\left(M^2 - (m_1 + m_2)^2\right)\left(M^2 - (m_1 - m_2)^2\right)\right]^{1/2}}{2M} \ , \tag{38.16}$$

and

$$d\Gamma = \frac{1}{32\pi^2}\,|\mathscr{M}|^2\,\frac{|\boldsymbol{p}_1|}{M^2}\,d\Omega \ , \tag{38.17}$$

where $d\Omega = d\phi_1 d(\cos\theta_1)$ is the solid angle of particle 1.

38.4.3. ***Three-body decays*** :

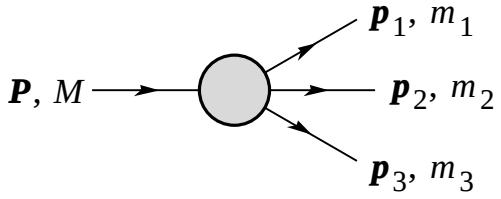


Figure 38.2: Definitions of variables for three-body decays.

Defining $p_{ij} = p_i + p_j$ and $m_{ij}^2 = p_{ij}^2$, then $m_{12}^2 + m_{23}^2 + m_{13}^2 = M^2 + m_1^2 + m_2^2 + m_3^2$ and $m_{12}^2 = (P - p_3)^2 = M^2 + m_3^2 - 2ME_3$, where E_3 is the energy of particle 3 in the rest frame of M. In that frame, the momenta of the three decay particles lie in a plane. The relative orientation of these three momenta is fixed if their energies are known. The momenta can therefore be specified in space by giving three Euler angles (α, β, γ) that specify the orientation of the final system relative to the initial particle [1]. Then

$$d\Gamma = \frac{1}{(2\pi)^5}\frac{1}{16M}\,|\mathscr{M}|^2\,dE_1\,dE_2\,d\alpha\,d(\cos\beta)\,d\gamma \ . \tag{38.18}$$

Alternatively

$$d\Gamma = \frac{1}{(2\pi)^5}\frac{1}{16M^2}\,|\mathscr{M}|^2\,|\boldsymbol{p}_1^*|\,|\boldsymbol{p}_3|\,dm_{12}\,d\Omega_1^*\,d\Omega_3\ , \tag{38.19}$$

where $(|\boldsymbol{p}_1^*|,\ \Omega_1^*)$ is the momentum of particle 1 in the rest frame of 1 and 2, and Ω_3 is the angle of particle 3 in the rest frame of the decaying particle. $|\boldsymbol{p}_1^*|$ and $|\boldsymbol{p}_3|$ are given by

$$|\boldsymbol{p}_1^*| = \frac{\left[\left(m_{12}^2-(m_1+m_2)^2\right)\left(m_{12}^2-(m_1-m_2)^2\right)\right]^{1/2}}{2m_{12}}\ , \tag{38.20a}$$

and

$$|\boldsymbol{p}_3| = \frac{\left[\left(M^2-(m_{12}+m_3)^2\right)\left(M^2-(m_{12}-m_3)^2\right)\right]^{1/2}}{2M}\ . \tag{38.20b}$$

[Compare with Eq. (38.16).]

If the decaying particle is a scalar or we average over its spin states, then integration over the angles in Eq. (38.18) gives

$$\begin{aligned} d\Gamma &= \frac{1}{(2\pi)^3}\frac{1}{8M}\,\overline{|\mathscr{M}|^2}\,dE_1\,dE_2 \\ &= \frac{1}{(2\pi)^3}\frac{1}{32M^3}\,\overline{|\mathscr{M}|^2}\,dm_{12}^2\,dm_{23}^2\ . \end{aligned} \tag{38.21}$$

This is the standard form for the Dalitz plot.

38.4.3.1. *Dalitz plot:* For a given value of m_{12}^2, the range of m_{23}^2 is determined by its values when $\boldsymbol{p}_2$ is parallel or antiparallel to $\boldsymbol{p}_3$:

$$(m_{23}^2)_{\max} = (E_2^*+E_3^*)^2 - \left(\sqrt{E_2^{*2}-m_2^2}-\sqrt{E_3^{*2}-m_3^2}\right)^2\ , \tag{38.22a}$$

$$(m_{23}^2)_{\min} = (E_2^*+E_3^*)^2 - \left(\sqrt{E_2^{*2}-m_2^2}+\sqrt{E_3^{*2}-m_3^2}\right)^2\ . \tag{38.22b}$$

Here $E_2^* = (m_{12}^2-m_1^2+m_2^2)/2m_{12}$ and $E_3^* = (M^2-m_{12}^2-m_3^2)/2m_{12}$ are the energies of particles 2 and 3 in the m_{12} rest frame. The scatter plot in m_{12}^2 and m_{23}^2 is called a Dalitz plot. If $\overline{|\mathscr{M}|^2}$ is constant, the allowed region of the plot will be uniformly populated with events [see Eq. (38.21)]. A nonuniformity in the plot gives immediate information on $|\mathscr{M}|^2$. For example, in the case of $D \to K\pi\pi$, bands appear when $m_{(K\pi)} = m_{K^*(892)}$, reflecting the appearance of the decay chain $D \to K^*(892)\pi \to K\pi\pi$.

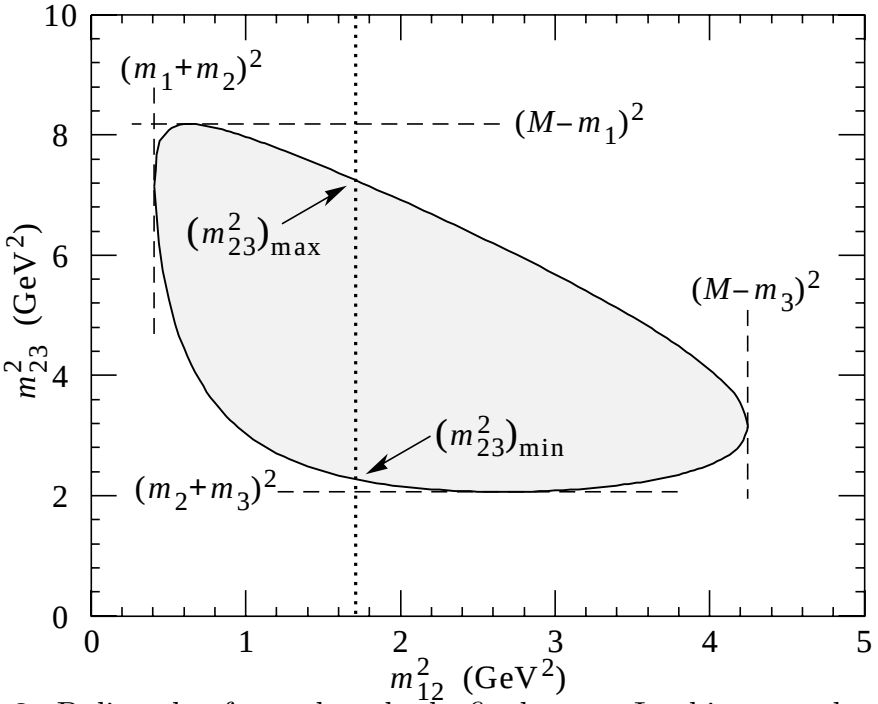


Figure 38.3: Dalitz plot for a three-body final state. In this example, the state is $\pi^+\overline{K}^0 p$ at 3 GeV. Four-momentum conservation restricts events to the shaded region.

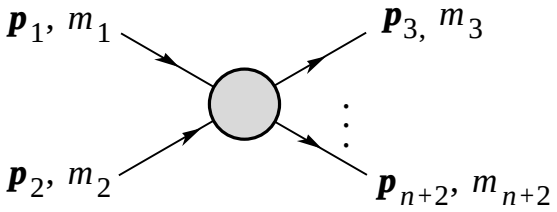


Figure 38.4: Definitions of variables for production of an n-body final state.

38.4.4. ***Kinematic limits*** : In a three-body decay the maximum of $|\boldsymbol{p}_3|$, [given by Eq. (38.20)], is achieved when $m_{12} = m_1 + m_2$, *i.e.*, particles 1 and 2 have the same vector velocity in the rest frame of the decaying particle. If, in addition, $m_3 > m_1, m_2$, then $|\boldsymbol{p}_3|_{\rm max} > |\boldsymbol{p}_1|_{\rm max}$, $|\boldsymbol{p}_2|_{\rm max}$.

38.4.5. ***Multibody decays*** : The above results may be generalized to final states containing any number of particles by combining some of the particles into "effective particles" and treating the final states as 2 or 3 "effective particle" states. Thus, if $p_{ijk\ldots} = p_i + p_j + p_k + \ldots$, then

$$m_{ijk\ldots} = \sqrt{{p^2}_{ijk\ldots}}\,, \tag{38.23}$$

and $m_{ijk\ldots}$ may be used in place of *e.g.*, m_{12} in the relations in Sec. 38.4.3 or 38.4.3.1 above.

38.5. Cross sections

The differential cross section is given by

$$d\sigma = \frac{(2\pi)^4 |\mathscr{M}|^2}{4\sqrt{(p_1 \cdot p_2)^2 - m_1^2 m_2^2}}$$

$$\times \; d\Phi_n(p_1 + p_2;\, p_3, \ldots, p_{n+2}) \; . \tag{38.24}$$

[See Eq. (38.11).] In the rest frame of m_2(lab),

$$\sqrt{(p_1 \cdot p_2)^2 - m_1^2 m_2^2} = m_2 p_{1\,\mathrm{lab}} \; ; \tag{38.25a}$$

while in the center-of-mass frame

$$\sqrt{(p_1 \cdot p_2)^2 - m_1^2 m_2^2} = p_{1\mathrm{cm}} \sqrt{s} \; . \tag{38.25b}$$

38.5.1. *Two-body reactions* :

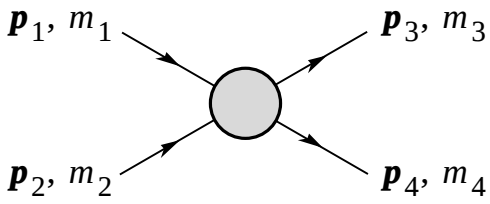


Figure 38.5: Definitions of variables for a two-body final state.

Two particles of momenta p_1 and p_2 and masses m_1 and m_2 scatter to particles of momenta p_3 and p_4 and masses m_3 and m_4; the Lorentz-invariant Mandelstam variables are defined by

$$\begin{aligned} s &= (p_1 + p_2)^2 = (p_3 + p_4)^2 \\ &= m_1^2 + 2E_1E_2 - 2\boldsymbol{p}_1 \cdot \boldsymbol{p}_2 + m_2^2 \; , \end{aligned} \tag{38.26}$$

$$\begin{aligned} t &= (p_1 - p_3)^2 = (p_2 - p_4)^2 \\ &= m_1^2 - 2E_1E_3 + 2\boldsymbol{p}_1 \cdot \boldsymbol{p}_3 + m_3^2 \; , \end{aligned} \tag{38.27}$$

$$\begin{aligned} u &= (p_1 - p_4)^2 = (p_2 - p_3)^2 \\ &= m_1^2 - 2E_1E_4 + 2\boldsymbol{p}_1 \cdot \boldsymbol{p}_4 + m_4^2 \; , \end{aligned} \tag{38.28}$$

and they satisfy

$$s + t + u = m_1^2 + m_2^2 + m_3^2 + m_4^2 \; . \tag{38.29}$$

The two-body cross section may be written as

$$\frac{d\sigma}{dt} = \frac{1}{64\pi s}\,\frac{1}{|\boldsymbol{p}_{1\mathrm{cm}}|^2}\,|\mathscr{M}|^2\ . \tag{38.30}$$

In the center-of-mass frame

$$\begin{aligned} t &= (E_{1\mathrm{cm}} - E_{3\mathrm{cm}})^2 - (p_{1\mathrm{cm}} - p_{3\mathrm{cm}})^2 \qquad - 4p_{1\mathrm{cm}}\,p_{3\mathrm{cm}}\,\sin^2(\theta_{\mathrm{cm}}/2) \\ &= t_0 - 4p_{1\mathrm{cm}}\,p_{3\mathrm{cm}}\,\sin^2(\theta_{\mathrm{cm}}/2)\ , \end{aligned} \tag{38.31}$$

where θ_{cm} is the angle between particle 1 and 3. The limiting values t_0 ($\theta_{\mathrm{cm}} = 0$) and t_1 ($\theta_{\mathrm{cm}} = \pi$) for $2 \to 2$ scattering are

$$t_0(t_1) = \left[\frac{m_1^2 - m_3^2 - m_2^2 + m_4^2}{2\sqrt{s}}\right]^2 - (p_{1\,\mathrm{cm}} \mp p_{3\,\mathrm{cm}})^2\ . \tag{38.32}$$

In the literature the notation t_{min} (t_{max}) for t_0 (t_1) is sometimes used, which should be discouraged since $t_0 > t_1$. The center-of-mass energies and momenta of the incoming particles are

$$E_{1\mathrm{cm}} = \frac{s + m_1^2 - m_2^2}{2\sqrt{s}}\ , \qquad E_{2\mathrm{cm}} = \frac{s + m_2^2 - m_1^2}{2\sqrt{s}}\ , \tag{38.33}$$

For $E_{3\mathrm{cm}}$ and $E_{4\mathrm{cm}}$, change m_1 to m_3 and m_2 to m_4. Then

$$p_{i\,\mathrm{cm}} = \sqrt{E_{i\,\mathrm{cm}}^2 - m_i^2}\ \text{and}\ p_{1\mathrm{cm}} = \frac{p_{1\,\mathrm{lab}}\,m_2}{\sqrt{s}}\ . \tag{38.34}$$

Here the subscript lab refers to the frame where particle 2 is at rest. [For other relations see Eqs. (38.2)–(38.4).]

38.5.2. ***Inclusive reactions*** : Choose some direction (usually the beam direction) for the z-axis; then the energy and momentum of a particle can be written as

$$E = m_{_T}\cosh y\ ,\ p_x\ ,\ p_y\ ,\ p_z = m_{_T}\sinh y\ , \tag{38.35}$$

where $m_{_T}$ is the transverse mass

$$m_{_T}^2 = m^2 + p_x^2 + p_y^2\ , \tag{38.36}$$

and the rapidity y is defined by

$$y = \frac{1}{2}\ln\left(\frac{E + p_z}{E - p_z}\right)$$

$$= \ln\left(\frac{E + p_z}{m_{_T}}\right) = \tanh^{-1}\left(\frac{p_z}{E}\right)\ . \tag{38.37}$$

Under a boost in the z-direction to a frame with velocity β, $y \to y - \tanh^{-1}\beta$. Hence the shape of the rapidity distribution dN/dy is invariant. The invariant cross section may also be rewritten

$$E\frac{d^3\sigma}{d^3p} = \frac{d^3\sigma}{d\phi\, dy\, p_T dp_T} \Longrightarrow \frac{d^2\sigma}{\pi\, dy\, d(p_T^2)} \ . \tag{38.38}$$

The second form is obtained using the identity $dy/dp_z = 1/E$, and the third form represents the average over ϕ.

Feynman's x variable is given by

$$x = \frac{p_z}{p_{z\,\mathrm{max}}} \approx \frac{E + p_z}{(E + p_z)_{\mathrm{max}}} \quad (p_T \ll |p_z|) \ . \tag{38.39}$$

In the c.m. frame,

$$x \approx \frac{2p_{z\,\mathrm{cm}}}{\sqrt{s}} = \frac{2m_T \sinh y_{\mathrm{cm}}}{\sqrt{s}} \tag{38.40}$$

and

$$= (y_{\mathrm{cm}})_{\mathrm{max}} = \ln(\sqrt{s}/m) \ . \tag{38.41}$$

For $p \gg m$, the rapidity [Eq. (38.37)] may be expanded to obtain

$$y = \frac{1}{2}\ln\frac{\cos^2(\theta/2) + m^2/4p^2 + \ldots}{\sin^2(\theta/2) + m^2/4p^2 + \ldots}$$

$$\approx -\ln\tan(\theta/2) \equiv \eta \tag{38.42}$$

where $\cos\theta = p_z/p$. The pseudorapidity η defined by the second line is approximately equal to the rapidity y for $p \gg m$ and $\theta \gg 1/\gamma$, and in any case can be measured when the mass and momentum of the particle is unknown. From the definition one can obtain the identities

$$\sinh\eta = \cot\theta \ , \quad \cosh\eta = 1/\sin\theta \ , \quad \tanh\eta = \cos\theta \ . \tag{38.43}$$

38.5.3. ***Partial waves*** : The amplitude in the center of mass for elastic scattering of spinless particles may be expanded in Legendre polynomials

$$f(k,\theta) = \frac{1}{k}\sum_\ell (2\ell+1)a_\ell P_\ell(\cos\theta) \ , \tag{38.44}$$

where k is the c.m. momentum, θ is the c.m. scattering angle, $a_\ell = (\eta_\ell e^{2i\delta_\ell} - 1)/2i$, $0 \le \eta_\ell \le 1$, and δ_ℓ is the phase shift of the ℓ^{th} partial wave. For purely elastic scattering, $\eta_\ell = 1$. The differential cross section is

$$\frac{d\sigma}{d\Omega} = |f(k,\theta)|^2 \ . \tag{38.45}$$

The optical theorem states that

$$\sigma_{\rm tot} = \frac{4\pi}{k} {\rm Im}\, f(k,0) \ , \tag{38.46}$$

and the cross section in the ℓ^{th} partial wave is therefore bounded:

$$\sigma_\ell = \frac{4\pi}{k^2}(2\ell+1)|a_\ell|^2 \leq \frac{4\pi(2\ell+1)}{k^2} \ . \tag{38.47}$$

The evolution with energy of a partial-wave amplitude a_ℓ can be displayed as a trajectory in an Argand plot, as shown in Fig. 38.6.

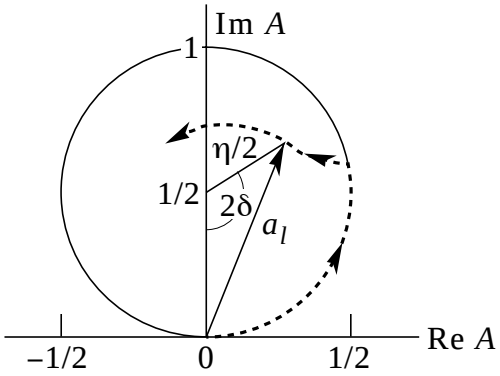


Figure 38.6: Argand plot showing a partial-wave amplitude a_ℓ as a function of energy. The amplitude leaves the unitary circle where inelasticity sets in ($\eta_\ell < 1$).

The usual Lorentz-invariant matrix element $\mathscr{M}$ (see Sec. 38.3 above) for the elastic process is related to $f(k,\theta)$ by

$$\mathscr{M} = -8\pi\sqrt{s}\ f(k,\theta) \ , \tag{38.48}$$

so

$$\sigma_{\rm tot} = -\frac{1}{2p_{\rm lab}\, m_2}\ {\rm Im}\, \mathscr{M}(t=0) \ , \tag{38.49}$$

where s and t are the center-of-mass energy squared and momentum transfer squared, respectively (see Sec. 38.4.1).

38.5.3.1. *Resonances:* The Breit-Wigner (nonrelativistic) form for an elastic amplitude a_ℓ with a resonance at c.m. energy E_R, elastic width $\Gamma_{\rm el}$, and total width $\Gamma_{\rm tot}$ is

$$a_\ell = \frac{\Gamma_{\rm el}/2}{E_R - E - i\Gamma_{\rm tot}/2} \ , \tag{38.50}$$

where E is the c.m. energy. As shown in Fig. 38.7, in the absence of background the elastic amplitude traces a counterclockwise circle with center $ix_{\rm el}/2$ and radius $x_{\rm el}/2$, where the elasticity $x_{\rm el} = \Gamma_{\rm el}/\Gamma_{\rm tot}$. The amplitude has a pole at $E = E_R - i\Gamma_{\rm tot}/2$.

The spin-averaged Breit-Wigner cross section for a spin-J resonance produced in the collision of particles of spin S_1 and S_2 is

$$\sigma_{BW}(E) = \frac{(2J+1)}{(2S_1+1)(2S_2+1)} \ \frac{\pi}{k^2} \ \frac{B_{\rm in}B_{\rm out}\Gamma_{\rm tot}^2}{(E-E_R)^2 + \Gamma_{\rm tot}^2/4} \ , \tag{38.51}$$

where k is the c.m. momentum, E is the c.m. energy, and $B_{\rm in}$ and $B_{\rm out}$ are the branching fractions of the resonance into the entrance and exit channels. The $2S+1$ factors are the multiplicities of the incident spin states, and are replaced by 2 for photons. This expression is valid only for an isolated state. If the width is not small, $\Gamma_{\rm tot}$ cannot be treated as a constant independent of E. There are many other forms for σ_{BW}, all of which are equivalent to the one given here in the narrow-width case. Some of these forms may be more appropriate if the resonance is broad.

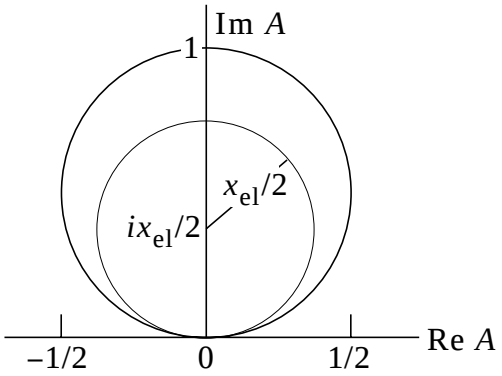


Figure 38.7: Argand plot for a resonance.

The relativistic Breit-Wigner form corresponding to Eq. (38.50) is:

$$a_\ell = \frac{-m\Gamma_{\rm el}}{s - m^2 + im\Gamma_{\rm tot}} \ . \tag{38.52}$$

A better form incorporates the known kinematic dependences, replacing $m\Gamma_{\rm tot}$ by $\sqrt{s}\,\Gamma_{\rm tot}(s)$, where $\Gamma_{\rm tot}(s)$ is the width the resonance particle would have if its mass

were $\sqrt{s}$, and correspondingly $m\Gamma_{\rm el}$ by $\sqrt{s}\,\Gamma_{\rm el}(s)$ where $\Gamma_{\rm el}(s)$ is the partial width in the incident channel for a mass $\sqrt{s}$:

$$a_\ell = \frac{-\sqrt{s}\,\Gamma_{\rm el}(s)}{s - m^2 + i\sqrt{s}\,\Gamma_{\rm tot}(s)} \ . \tag{38.53}$$

For the Z boson, all the decays are to particles whose masses are small enough to be ignored, so on dimensional grounds $\Gamma_{\rm tot}(s) = \sqrt{s}\,\Gamma_0/m_Z$, where Γ_0 defines the width of the Z, and $\Gamma_{\rm el}(s)/\Gamma_{\rm tot}(s)$ is constant. A full treatment of the line shape requires consideration of dynamics, not just kinematics. For the Z this is done by calculating the radiative corrections in the Standard Model.

References:

1. See, for example, J.J. Sakurai, *Modern Quantum Mechnaics*, Addison-Wesley (1985), p. 172, or D.M. Brink and G.R. Satchler, *Angular Momentum*, 2nd ed., Oxford University Press (1968), p. 20.

39. CROSS-SECTION FORMULAE FOR SPECIFIC PROCESSES

Revised September 2005 by R.N. Cahn (LBNL).

Setting aside leptoproduction (for which, see Sec. 16), the cross sections of primary interest are those with light incident particles, e^+e^-, $\gamma\gamma$, $q\overline{q}$, gq , gg, *etc.*, where g and q represent gluons and light quarks. The produced particles include both light particles and heavy ones - t, W, Z, and the Higgs boson H. We provide the production cross sections calculated within the Standard Model for several such processes.

39.1. Resonance Formation

Resonant cross sections are generally described by the Breit-Wigner formula (Sec. 16 of this *Review*).

$$\sigma(E) = \frac{2J+1}{(2S_1+1)(2S_2+1)}\frac{4\pi}{k^2}\left[\frac{\Gamma^2/4}{(E-E_0)^2+\Gamma^2/4}\right]B_{in}B_{out}, \tag{39.1}$$

where E is the c.m. energy, J is the spin of the resonance, and the number of polarization states of the two incident particles are $2S_1+1$ and $2S_2+2$. The c.m. momentum in the initial state is k, E_0 is the c.m. energy at the resonance, and Γ is the full width at half maximum height of the resonance. The branching fraction for the resonance into the initial-state channel is B_{in} and into the final-state channel is B_{out}. For a narrow resonance, the factor in square brackets may be replaced by $\pi\Gamma\delta(E-E_0)/2$.

39.2. Production of light particles

The production of point-like, spin-1/2 fermions in e^+e^- annihilation through a virtual photon, $e^+e^- \to \gamma^* \to f\overline{f}$, at c.m. energy squared s is given by

$$\frac{d\sigma}{d\Omega} = N_c\frac{\alpha^2}{4s}\beta\left[1+\cos^2\theta+(1-\beta^2)\sin^2\theta\right]Q_f^2 \ , \tag{39.2}$$

where β is v/c for the produced fermions in the c.m., θ is the c.m. scattering angle, and Q_f is the charge of the fermion. The factor N_c is 1 for charged leptons and 3 for quarks. In the ultrarelativistic limit, $\beta \to 1$,

$$\sigma = N_cQ_f^2\frac{4\pi\alpha^2}{3s} = N_cQ_f^2\frac{86.8\ nb}{s(\mathrm{GeV}^2)^2} \ . \tag{39.3}$$

The cross section for the annihilation of a $q\overline{q}$ pair into a distinct pair $q'\overline{q}'$ through a gluon is completely analogous up to color factors, with the replacement $\alpha \to \alpha_s$. Treating all quarks as massless, averaging over the colors of the initial quarks and defining $t = -s\sin^2(\theta/2)$, $u = -s\cos^2(\theta/2)$, one finds [1]

$$\frac{d\sigma}{d\Omega}(q\overline{q} \to q'\overline{q}') = \frac{\alpha^2}{9s}\frac{t^2+u^2}{s^2} \ . \tag{39.4}$$

Crossing symmetry gives

$$\frac{d\sigma}{d\Omega}(q\overline{q} \to q'\overline{q}') = \frac{\alpha^2}{9s}\frac{s^2+u^2}{t^2} \ . \tag{39.5}$$

If the quarks q and q' are identical, we have

$$\frac{d\sigma}{d\Omega}(q\overline{q} \to q\overline{q}) = \frac{\alpha_s^2}{9s}\left[\frac{t^2+u^2}{s^2} + \frac{s^2+u^2}{t^2} - \frac{2u^2}{3st}\right] \ , \tag{39.6}$$

and by crossing

$$\frac{d\sigma}{d\Omega}(qq \to qq) = \frac{\alpha_s^2}{9s}\left[\frac{t^2+s^2}{u^2} + \frac{s^2+u^2}{t^2} - \frac{2s^2}{3ut}\right] . \tag{39.7}$$

Annihilation of e^+e^- into $\gamma\gamma$ has the cross section

$$\frac{d\sigma}{d\Omega}(e^+e^- \to \gamma\gamma) = \frac{\alpha^2}{2s}\frac{u^2+t^2}{tu} \ . \tag{39.8}$$

The related QCD process also has a triple-gluon coupling. The cross section is

$$\frac{d\sigma}{d\Omega}(q\overline{q} \to qq) = \frac{8\alpha_s^2}{27s}(t^2+u^2)\left(\frac{1}{tu} - \frac{9}{4s^2}\right) \ . \tag{39.9}$$

The crossed reactions are

$$\frac{d\sigma}{d\Omega}(qg \to qg) = \frac{\alpha_s^2}{9s}(s^2+u^2)(-\frac{1}{su} + \frac{9}{4t^2}) \tag{39.10}$$

and

$$\frac{d\sigma}{d\Omega}(gg \to q\overline{q}) = \frac{\alpha_s^2}{24s}(t^2+u^2)(\frac{1}{tu} - \frac{9}{4s^2}) \ . \tag{39.11}$$

Finally,

$$\frac{d\sigma}{d\Omega}(gg \to gg) = \frac{9\alpha_s^2}{8s}(3 - \frac{ut}{s^2} - \frac{su}{t^2} - \frac{st}{u^2}) \ . \tag{39.12}$$

39.3. Production of Weak Gauge Bosons

39.3.1. *W and Z resonant production* :

Resonant production of a single W or Z is governed by the partial widths

$$\Gamma(W \to \ell_i \overline{\nu}_i) = \frac{\sqrt{2} G_F m_W^3}{12\pi} \tag{39.13}$$

$$\Gamma(W \to q_i \overline{q}_j) = 3\frac{\sqrt{2} G_F |V_{ij}|^2 m_W^3}{12\pi} \tag{39.14}$$

$$\Gamma(Z \to f\overline{f}) = N_c \frac{\sqrt{2} G_F m_Z^3}{6\pi} \left[(T_3 - Q_f \sin^2\theta_W)^2 + (Q_f \sin\theta_W)^2\right] \tag{39.15}$$

The weak mixing angle is θ_W. The CKM matrix elements are indicated by V_{ij} and N_c is 3 for $q\overline{q}$ final states and 1 for leptonic final states.

The full differential cross section for $f_i \overline{f}_j \to (W, Z) \to f_{i'} \overline{f}_{j'}$ is given by

$$\begin{aligned}\frac{d\sigma}{d\Omega} = {} & \frac{N_c^f}{N_c^i} \cdot \frac{1}{256\pi^2 s} \cdot \frac{s^2}{(s - M^2)^2 + s\Gamma^2} \\ & \times \Big[(L^2 + R^2)(L'^2 + R'^2)(1 + \cos^2\theta) \\ & + (L^2 - R^2)(L'^2 - R'^2) 2\, cos\theta\Big]\end{aligned} \tag{39.16}$$

where M is the mass of the W or Z. The couplings for the W are $L = (8 G_F m_W^2/\sqrt{2})^{1/2} V_{ij}/\sqrt{2}; R = 0$ where V_{ij} is the corresponding CKM matrix element, with an analogous expression for L' and R'. For Z, the couplings are $L = (8 G_F m_Z^2/\sqrt{2})^{1/2}(T_3 - \sin^2\theta_W Q); R = -(8 G_F m_Z^2/\sqrt{2})^{1/2} \sin^2\theta_W Q$, where T_3 is the weak isospin of the initial left-handed fermion and Q is the initial fermion's electric charge. The expressions for L' and R' are analogous. The color factors $N_c^{i,f}$ are 3 for initial or final quarks and 1 for initial or final leptons.

39.3.2. *Production of pairs of weak gauge bosons* :

The cross section for $f\overline{f} \to W^+W^-$ is given in term of the couplings of the left-handed and right-handed fermion f, $\ell = 2(T_3 - Q x_W)$, $r = -2Q x_W$, where T_3 is the third component of weak isospin for the left-handed f, Q is its electric charge (in units of the proton charge), and $x_W = \sin^2\theta_W$:

$$\frac{d\sigma}{dt} = \frac{2\pi\alpha^2}{N_c s^2}\left\{\left[\left(Q + \frac{\ell + r}{4x_W}\frac{s}{s-m_Z^2}\right)^2 + \left(\frac{\ell + r}{4x_W}\frac{s}{s-m_Z^2}\right)^2\right]A(s,t,u) + \frac{1}{2x_W}\left(Q + \frac{\ell}{2x_W}\frac{s}{s-m_Z^2}\right)\left(\Theta(-Q)I(s,t,u) - \Theta(Q)I(s,u,t)\right) + \frac{1}{8x_W^2}\left(\Theta(-Q)E(s,t,u) + \Theta(Q)E(s,u,t)\right)\right\}, \tag{39.17}$$

where $\Theta(x)$ is 1 for $x > 0$ and 0 for $x < 0$, and where

$$\begin{aligned} A(s,t,u) &= \left(\frac{tu}{m_W^4} - 1\right)\left(\frac{1}{4} - \frac{m_W^2}{s} + 3\frac{m_W^4}{s^2}\right) + \frac{s}{m_W^2} - 4, \\ I(s,t,u) &= \left(\frac{tu}{m_W^4} - 1\right)\left(\frac{1}{4} - \frac{m_W^2}{2s} - \frac{m_W^4}{st}\right) + \frac{s}{m_W^2} - 2 + 2\frac{m_W^2}{t}, \\ E(s,t,u) &= \left(\frac{tu}{m_W^4} - 1\right)\left(\frac{1}{4} + \frac{m_W^2}{t}\right) + \frac{s}{m_W^2}, \end{aligned} \tag{39.18}$$

and s, t, u are the usual Mandelstam variables with $s = (p_f + p_{\overline{f}})^2, t = (p_f - p_{W^-})^2, u = (p_f - p_{W^+})^2$. The factor N_c is 3 for quarks and 1 for leptons.

The analogous cross-section for $q_i\overline{q}_j \to W^\pm Z^0$ is

$$\begin{aligned} \frac{d\sigma}{dt} = \frac{\pi\alpha^2|V_{ij}|^2}{6s^2x_W^2}\Bigg\{ & \left(\frac{1}{s-m_W^2}\right)^2\left[\left(\frac{9-8x_W}{4}\right)\left(ut - m_W^2m_Z^2\right) + (8x_W - 6)\,s\left(m_W^2 + m_Z^2\right)\right] \\ & + \left[\frac{ut - m_W^2m_Z^2 - s(m_W^2 + m_Z^2)}{s - m_W^2}\right]\left[\frac{\ell_j}{t} - \frac{\ell_i}{u}\right] \\ & + \frac{ut - m_W^2m_Z^2}{4(1-x_W)}\left[\frac{\ell_j^2}{t^2} + \frac{\ell_i^2}{u^2}\right] + \frac{s(m_W^2 + m_Z^2)}{2(1-x_W)}\frac{\ell_i\ell_j}{tu}\Bigg\}, \end{aligned} \tag{39.19}$$

where ℓ_i and ℓ_j are the couplings of the left-handed q_i and q_j as defined above. The CKM matrix element between q_i and q_j is V_{ij}.

The cross section for $q_i\overline{q}_i \to Z^0Z^0$ is

$$\frac{d\sigma}{dt} = \frac{\pi\alpha^2}{96}\frac{\ell_i^4 + r_i^4}{x_W^2(1-x_W^2)s^2}\left[\frac{t}{u} + \frac{u}{t} + \frac{4m_Z^2 s}{tu} - m_Z^4\left(\frac{1}{t^2} + \frac{1}{u^2}\right)\right]. \tag{39.20}$$

39.4. Production of Higgs Bosons

39.4.1. *Resonant Production* :

The Higgs boson of the Standard Model can be produced resonantly in the collisions of quarks, leptons, W or Z bosons, gluons, or photons. The production cross section is thus controlled by the partial width of the Higgs boson into the entrance channel and its total width. The branching fractions for the Standard Model Higgs boson are shown in Fig. 1 of the "Searches for Higgs bosons" review in the Particle Listings section, as a function of the Higgs boson mass. The partial widths are given by the relations

$$\Gamma(H \to f\overline{f}) = \frac{G_F m_f^2 m_H N_c}{4\pi\sqrt{2}}\left(1 - 4m_f^2/m_H^2\right)^{3/2}, \tag{39.21}$$

$$\Gamma(H \to W^+W^-) = \frac{G_F m_H^3 \beta_W}{32\pi\sqrt{2}}\left(4 - 4a_W + 3a_W^2\right), \tag{39.22}$$

$$\Gamma(H \to ZZ) = \frac{G_F m_H^3 \beta_Z}{64\pi\sqrt{2}}\left(4 - 4a_Z + 3a_Z^2\right), \tag{39.23}$$

where N_c is 3 for quarks and 1 for leptons and where $a_W = 1 - \beta_W^2 = 4m_W^2/m_H^2$ and $a_Z = 1 - \beta_Z^2 = 4m_Z^2/m_H^2$. The decay to two gluons proceeds through quark loops, with the t quark dominating [2]. Explicitly,

$$\Gamma(H \to gg) = \frac{\alpha_s^2 G_F m_H^3}{36\pi^3\sqrt{2}}\left|\sum_q I(m_q^2/m_H^2)\right|^2, \tag{39.24}$$

where $I(z)$ is complex for $z < 1/4$. For $z < 2\times10^{-3}$, $|I(z)|$ is small so the light quarks contribute negligibly. For $m_H < 2m_t$, $z > 1/4$ and

$$I(z) = 3\left[2z + 2z(1-4z)\left(\sin^{-1}\frac{1}{2\sqrt{z}}\right)^2\right], \tag{39.25}$$

which has the limit $I(z) \to 1$ as $z \to \infty$.

39.4.2. *Higgs Boson Production in W^* and Z^* decay* :

The Standard Model Higgs boson can be produced in the decay of a virtual W or Z ("Higgstrahlung") [3,4]: In particular, if k is the c.m. momentum of the Higgs boson,

$$\sigma(q_i\overline{q}_j \to WH) = \frac{\pi\alpha^2|V_{ij}|^2}{36\sin^4\theta_W}\,\frac{2k}{\sqrt{s}}\,\frac{k^2+3m_W^3}{(s-m_W^2)^2} \tag{39.26}$$

$$\sigma(f\overline{f} \to ZH) = \frac{2\pi\alpha^2(\ell_f^2+r_f^2)}{48N_c\sin^4\theta_W\cos^4\theta_W}\,\frac{2k}{\sqrt{s}}\,\frac{k^2+3m_Z^3}{(s-m_Z^2)^2}, \tag{39.27}$$

where ℓ and r are defined as above.

39.4.3. *W and Z Fusion* :

Just as high-energy electrons can be regarded as sources of virtual photon beams, at very high energies they are sources of virtual W and Z beams. For Higgs boson production, it is the longitudinal components of the Ws and Zs that are important [5]. The distribution of longitudinal Ws carrying a fraction y of the electron's energy is [6]

$$f(y) = \frac{g^2}{16\pi^2}\frac{1-y}{y}, \tag{39.28}$$

where $g = e/\sin\theta_W$. In the limit $s \gg m_H \gg m_W$, the partial decay rate is $\Gamma(H \to W_LW_L) = (g^2/16\pi^2)^3(m_H^3/8\pi)$ and in the equivalent W approximation [7]

$$\sigma(e^+e^- \to \overline{\nu}_e\nu_e H) = \frac{1}{16m_W^2}\left(\frac{\alpha}{\sin^2\theta_W}\right)^3 \left[\left(1+\frac{m_H^2}{s}\right)\log\frac{s}{m_H^2} - 2 + 2\frac{m_H^2}{s}\right]. \tag{39.29}$$

There are significant corrections to this relation when m_H is not large compared to m_W [8]. For $m_H = 150$ GeV, the estimate is too high by 51% for $\sqrt{s} = 1000$ GeV, 32% too high at $\sqrt{s} = 2000$ GeV, and 22% too high at $\sqrt{s} = 4000$ GeV. Fusion of ZZ to make a Higgs boson can be treated similarly. Identical formulae apply for Higgs production in the collisions of quarks whose charges permit the emission of a W^+ and a W^-, except that QCD corrections and CKM matrix elements are required. Even in the absence of QCD corrections, the fine-structure constant ought to be evaluated at the scale of the collision, say m_W. All quarks contribute to the ZZ fusion process.

39.5. Inclusive hadronic reactions

One-particle inclusive cross sections $Ed^3\sigma/d^3p$ for the production of a particle of momentum p are conveniently expressed in terms of rapidity y (see above) and the momentum p_T transverse to the beam direction (in the c.m.):

$$E\frac{d^3\sigma}{d^3p} = \frac{d^3\sigma}{d\phi\, dy\, p_T dp_T^2}\ . \tag{39.30}$$

In appropriate circumstances, the cross section may be decomposed as a partonic cross section multiplied by the probabilities of finding partons of the prescribed momenta:

$$\sigma_{\rm hadronic} = \sum_{ij}\int dx_1\ dx_2\ f_i(x_1)\ f_j(x_2)\ d\widehat{\sigma}_{\rm partonic}\ , \tag{39.31}$$

The probability that a parton of type i carries a fraction of the incident particle's that lies between x_1 and $x_1 + dx_1$ is $f_i(x_1)dx_1$ and similarly for partons in the other incident particle. The partonic collision is specified by its c.m. energy squared $\hat{s} = x_1x_2s$ and the momentum transfer squared $\hat{t}$. The final hadronic state is more conveniently specified by the rapidities y_1, y_2 of the two jets resulting from the collision and the transverse momentum p_T. The connection between the differentials is

$$dx_1dx_2d\hat{t} = dy_1dy_2\frac{\hat{s}}{s}dp_T^2, \tag{39.32}$$

so that

$$\frac{d^3\sigma}{dy_1dy_2dp_T^2} = \frac{\hat{s}}{s}\left[f_i(x_1)f_j(x_2)\frac{d\hat{\sigma}}{d\hat{t}}(\hat{s},\hat{t},\hat{u}) + f_i(x_2)f_j(x_1)\frac{d\hat{\sigma}}{d\hat{t}}(\hat{s},\hat{u},\hat{t})\right], \tag{39.33}$$

where we have taken into account the possibility that the incident parton types might arise from either incident particle. The second term should be dropped if the types are identical: $i = j$.

39.6. Two-photon processes

In the Weizsäcker-Williams picture, a high-energy electron beam is accompanied by a spectrum of virtual photons of energies ω and invariant-mass squared $q^2 = -Q^2$, for which the photon number density is

$$dn = \frac{\alpha}{\pi}\left[1 - \frac{\omega}{E} + \frac{\omega^2}{E^2} - \frac{m_e^2\ \omega^2}{Q^2E^2}\right]\frac{d\omega}{\omega}\frac{dQ^2}{Q^2}, \tag{39.34}$$

where E is the energy of the electron beam. The cross section for $e^+e^- \to e^+e^-X$ is then [9]

$$d\sigma_{e^+e^-\to e^+e^-X}(s) = dn_1 dn_2 d\sigma_{\gamma\gamma\to X}(W^2), \tag{39.35}$$

where $W^2 = m_X^2$. Integrating from the lower limit $Q^2 = m_e^2 \dfrac{\omega_i^2}{E_i(E_i-\omega_i)}$ to a maximum Q^2 gives

$$\sigma_{e^+e^-\to e^+e^-X}(s) = \frac{\alpha^2}{\pi^2}\int_{z_{th}}^1 \frac{dz}{z} \left[\left(\ln\frac{Q^2_{max}}{zm_e^2} - 1\right)^2 f(z) + \frac{1}{3}(\ln z)^3\right]\sigma_{\gamma\gamma\to X}(zs), \tag{39.36}$$

where

$$f(z) = \left(1+\tfrac{1}{2}z\right)^2 \ln(1/z) - \tfrac{1}{2}(1-z)(3+z). \tag{39.37}$$

The appropriate value of Q^2_{max} depends on the properties of the produced system X. For production of hadronic systems, $Q^2_{max} \approx m_\rho^2$, while for lepton-pair production, $Q^2 \approx W^2$. For production of a resonance with spin $J \neq 1$, we have

$$\sigma_{e^+e^-\to e^+e^-R}(s) = (2J+1)\frac{8\alpha^2\Gamma_{R\to\gamma\gamma}}{m_R^3} \times \left[f(m_R^2/s)\left(\ln\frac{m_V^2 s}{m_e^2 m_R^2} - 1\right)^2 - \frac{1}{3}\left(\ln\frac{s}{M_R^2}\right)^3\right] \tag{39.38}$$

where m_V is the mass that enters into the form factor for the $\gamma\gamma \to R$ transition, typically m_ρ.

References:

1. G.F. Owens *et al.*, Phys. Rev. **D18**, 1501 (1978). Note that cross section given in previous editions of RPP for $gg \to q\overline{q}$ lacked a factor of π.
2. F. Wilczek, Phys. Rev. Lett. **39**, 1304 (1977).
3. B.L. Ioffe and V.Khoze, Leningrad Report 274, 1976; *Sov. J. Nucl. Part.* **9**,50 (1978).
4. J. Ellis *et al.*, Nucl. Phys. **B106**, 292 (1976).
5. R.N. Cahn and S. Dawson, Phys. Lett. **B136**, 196 (1984), erratum, Phys. Lett. **B138**, 464 (1984).
6. S. Dawson, Nucl. Phys. **B249**, 42 (1985).
7. M.S. Chanowitz and M.K. Gaillard, Phys. Lett. **B142**, 85 (1984).
8. R.N. Cahn, Nucl. Phys. **B255**, 341 (1985).
9. For an exhaustive treatment, see V.M. Budnev *et al.*, Phys. Reports **15C**, 181(1975).

40. PLOTS OF CROSS SECTIONS AND RELATED QUANTITIES

Jet Production in pp and $\overline{p}p$ Interactions

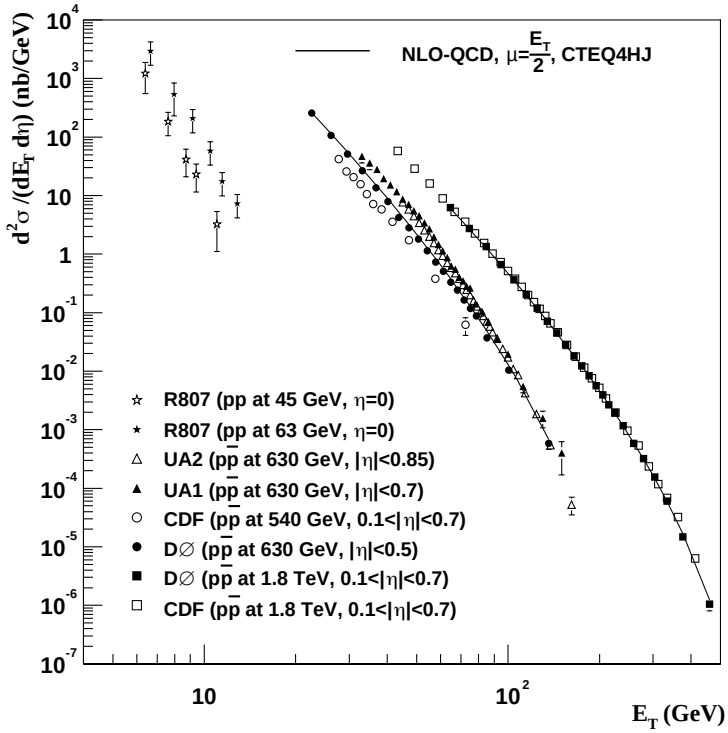


Figure 40.1: Transverse energy dependence of the inclusive differential jet cross sections in the central pseudorapidity region. The error bars are either statistical (DØ), statistical and p_T dependent (UA2), statistical and energy dependent from unsmearing (UA1), uncorrelated (CDF), or total (R806) uncertainties. Comparison of the different experimental results is not straight forward, since the different experiments used different jet reconstruction algorithms. For instance, DØ and CDF used a fixed cone algorithm with a size $\mathcal{R}$=0.7 for all their measurements, compared to a cone size of 1.3 for UA2. **DØ**: Phys. Rev. **D64**, 032003 (2001); **CDF**: Phys. Rev. **D64**, 032001 (2001); **UA1**: Phys. Lett. **B172**, 461 (1986); **UA2**: Phys. Lett. **B257**, 232 (1991); **R807**: Phys. Lett. **B123**, 133 (1983). Next-to-Leading order QCD predictions, using CTEQ4HJ pdfs and $\mu_{R,F} = E_T/2$, are shown for $p\overline{p}$ at 630 GeV and 1.8 TeV. (Courtesy of V.D. Elvira, Fermilab, 2001)

Direct γ Production in $\overline{p}p$ Interactions

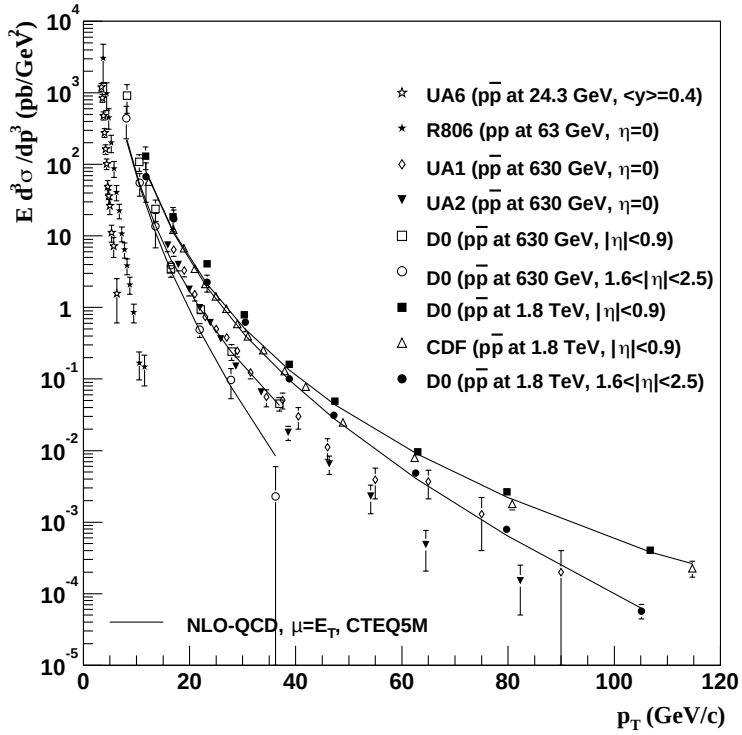


Figure 40.2: Transverse energy dependence of isolated photon cross sections. The error bars are either statistical (CDF), uncorrelated (DØ), or total (UA1, UA2, R806) uncertainties. **DØ**: Phys. Rev. Lett. **87**, 251805 (2001); **CDF**: Phys. Rev. **D73**, 2662 (1994); **UA6**: Phys. Lett. **B206**, 163 (1988); **UA1**: Phys. Lett. **B209**, 385 (1988); **UA2**: Phys. Lett. **B288**, 386 (1992); **R806**: Z. Phys. **C13**, 277 (1982). Next-to-Leading order QCD predictions are shown for $p\overline{p}$ at 630 GeV and 1.8 TeV. (Courtesy of V.D. Elvira, Fermilab, 2001)

Differential Cross Section for W and Z Boson Production

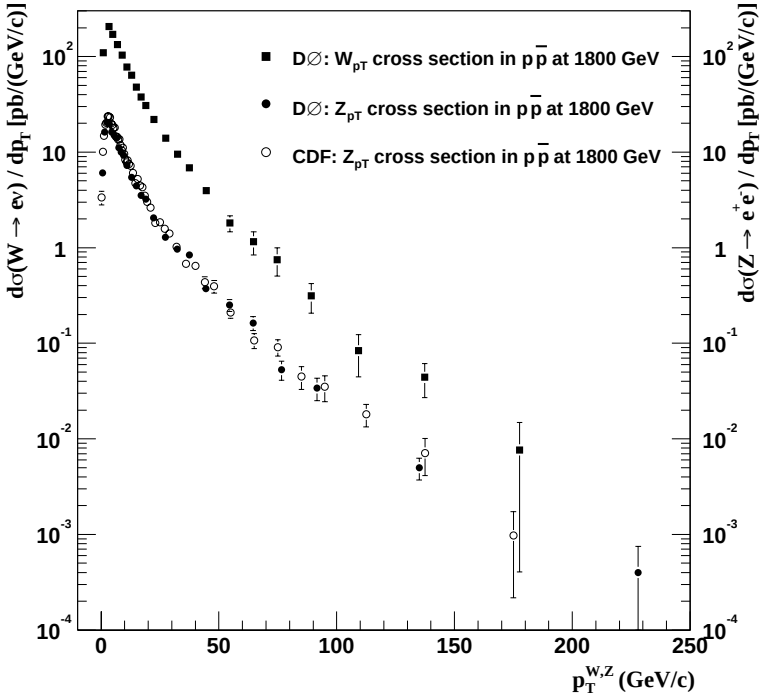


Figure 40.3: Differential cross section for W and Z boson production. The error bars are total errors, excluding the DØ (CDF) 4.4% (3.9%) luminosity uncertainty. **DØ**: Phys. Lett. **B513**, 292 (2001), Phys. Rev. Lett. **84**, 2792 (2000). **CDF**: Phys. Rev. Lett. **84**, 845 (2000). (Courtesy of V.D. Elvira, Fermilab, 2001)

Pseudorapidity Distributions in $\overline{p}p$ Interactions

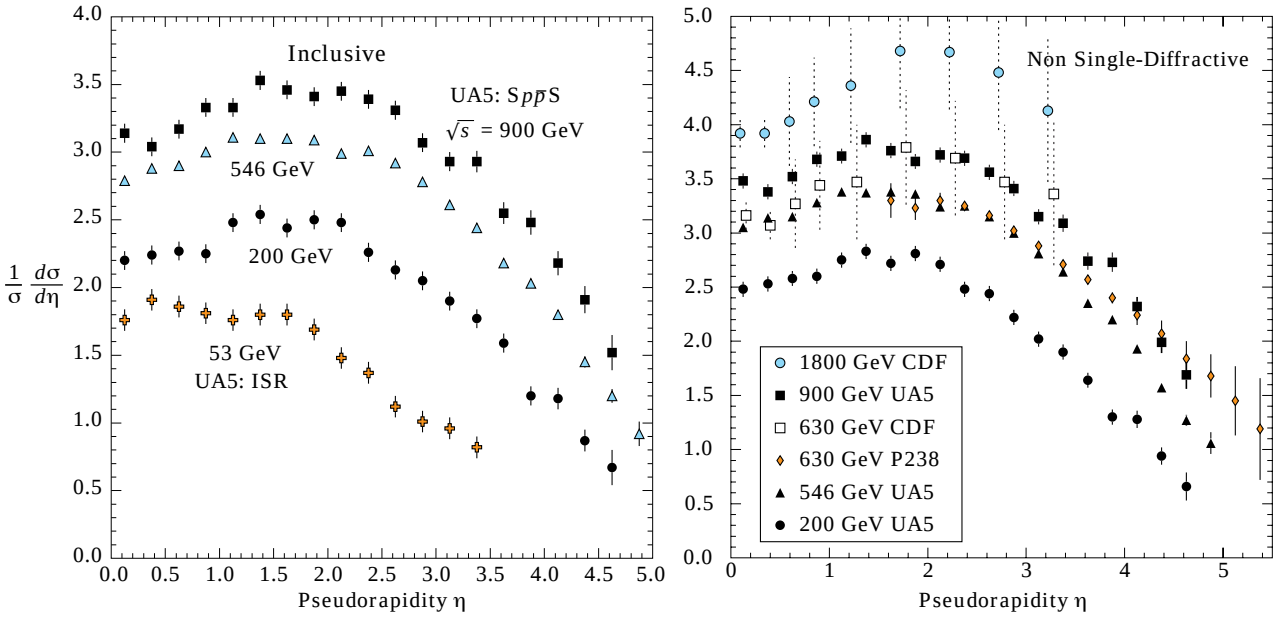


Figure 40.4: Charged particle pseudorapidity distributions in $p\overline{p}$ collisions for 53 GeV $\leq \sqrt{s} \leq$ 1800 GeV. UA5 data from the S$p\overline{p}$S are taken from G.J. Alner *et al.*, Z. Phys. **C33**, 1 (1986), and from the ISR from K. Alpgøard *et al.*, Phys. Lett. **112B**, 193 (1982). The UA5 data are shown for both the full inelastic cross-section and with singly diffractive events excluded. Additional non single-diffractive measurements are available from CDF at the Tevatron, F. Abe *et al.*, Phys. Rev. **D41**, 2330 (1990) and Experiment P238 at the S$p\overline{p}$S, R. Harr *et al.*, Phys. Lett. **B401**, 176 (1997). (Courtesy of D.R. Ward, Cambridge Univ., 1999.)

Average Hadron Multiplicities in Hadronic e^+e^- Annihilation Events

Table 40.1: Average hadron multiplicities per hadronic e^+e^- annihilation event at $\sqrt{s} \approx 10$, 29–35, 91, and 130–200 GeV. The rates given include decay products from resonances with $c\tau < 10$ cm, and include the corresponding anti-particle state. Correlations of the systemaic uncertainties were considered for the calculation of the averages. (Updated July 2005 by O. Biebel, LMU, Munich.)

Particle	$\sqrt{s} \approx 10$ GeV	$\sqrt{s} = 29$–35 GeV	$\sqrt{s} = 91$ GeV	$\sqrt{s} = 130$–200 GeV
Pseudoscalar mesons:				
π^+	6.6 ± 0.2	10.3 ± 0.4	17.02 ± 0.19	21.24 ± 0.39
π^0	3.2 ± 0.3	5.83 ± 0.28	9.42 ± 0.32	
K^+	0.90 ± 0.04	1.48 ± 0.09	2.228 ± 0.059	2.82 ± 0.19
K^0	0.91 ± 0.05	1.48 ± 0.07	2.049 ± 0.026	2.10 ± 0.12
η	0.20 ± 0.04	0.61 ± 0.07	1.049 ± 0.080	
$\eta'(958)$	0.03 ± 0.01	0.26 ± 0.10	0.152 ± 0.020	
D^+	$0.194 \pm 0.019^{(k)}$	0.17 ± 0.03	0.175 ± 0.016	
D^0	$0.446 \pm 0.032^{(k)}$	0.45 ± 0.07	0.454 ± 0.030	
D_s^+	$0.063 \pm 0.014^{(k)}$	$0.45 \pm 0.20^{(a)}$	0.131 ± 0.021	
B^+, B_d^0	—	—	$0.165 \pm 0.026^{(b)}$	
B_u^+	—	—	$0.178 \pm 0.006^{(b)}$	
B_s^0	—	—	$0.057 \pm 0.013^{(b)}$	
Scalar mesons:				
$f_0(980)$	0.024 ± 0.006	$0.05 \pm 0.02^{(c)}$	0.146 ± 0.012	
$a_0(980)^\pm$	—	—	$0.27 \pm 0.11^{(d)}$	
Vector mesons:				
$\rho(770)^0$	0.35 ± 0.04	0.81 ± 0.08	1.231 ± 0.098	
$\rho(770)^\pm$	—	—	$2.40 \pm 0.43^{(d)}$	
$\omega(782)$	0.30 ± 0.08	—	1.016 ± 0.065	
$K^*(892)^+$	0.27 ± 0.03	0.64 ± 0.05	0.715 ± 0.059	
$K^*(892)^0$	0.29 ± 0.03	0.56 ± 0.06	0.738 ± 0.024	
$\phi(1020)$	0.044 ± 0.003	0.085 ± 0.011	0.0963 ± 0.0032	
$D^*(2010)^+$	$0.177 \pm 0.022^{(k)}$	0.43 ± 0.07	$0.1937 \pm 0.0057^{(j)}$	
$D^*(2007)^0$	$0.168 \pm 0.019^{(k)}$	0.27 ± 0.11	—	
$D_s^*(2112)^+$	$0.048 \pm 0.014^{(k)}$	—	$0.101 \pm 0.048^{(g)}$	
$B^{*\ (e)}$	—	—	0.288 ± 0.026	
$J/\psi(1S)$	$0.00050 \pm 0.00005^{(k)}$	—	$0.0052 \pm 0.0004^{(f)}$	
$\psi(2S)$	—	—	$0.0023 \pm 0.0004^{(f)}$	
$\Upsilon(1S)$	—	—	$0.00014 \pm 0.00007^{(f)}$	
Pseudovector mesons:				
$f_1(1285)$	—	—	0.165 ± 0.051	
$f_1(1420)$	—	—	0.056 ± 0.012	
$\chi_{c1}(3510)$	—	—	$0.0041 \pm 0.0011^{(f)}$	
Tensor mesons:				
$f_2(1270)$	0.09 ± 0.02	0.14 ± 0.04	0.166 ± 0.020	
$f_2'(1525)$	—	—	0.012 ± 0.006	
$K_2^*(1430)^+$	—	0.09 ± 0.03	—	
$K_2^*(1430)^0$	—	0.12 ± 0.06	0.084 ± 0.022	
$B^{**\ (i)}$	—	—	0.118 ± 0.024	
$D_{s1}^\pm$	—	—	$0.0052 \pm 0.0011^{(\ell)}$	
$D_{s2}^{*\pm}$	—	—	$0.0083 \pm 0.0031^{(\ell)}$	
Baryons:				
p	0.253 ± 0.016	0.640 ± 0.050	1.050 ± 0.032	1.41 ± 0.18
Λ	0.080 ± 0.007	0.205 ± 0.010	0.3915 ± 0.0065	0.39 ± 0.03
Σ^0	0.023 ± 0.008	—	0.076 ± 0.011	
Σ^-	—	—	0.081 ± 0.010	
Σ^+	—	—	0.107 ± 0.011	
$\Sigma^\pm$	—	—	0.174 ± 0.009	
Ξ^-	0.0059 ± 0.0007	0.0176 ± 0.0027	0.0258 ± 0.0010	
$\Delta(1232)^{++}$	0.040 ± 0.010	—	0.085 ± 0.014	
$\Sigma(1385)^-$	0.006 ± 0.002	0.017 ± 0.004	0.0240 ± 0.0017	
$\Sigma(1385)^+$	0.005 ± 0.001	0.017 ± 0.004	0.0239 ± 0.0015	
$\Sigma(1385)^\pm$	0.0106 ± 0.0020	0.033 ± 0.008	0.0462 ± 0.0028	
$\Xi(1530)^0$	0.0015 ± 0.0006	—	0.0055 ± 0.0005	
Ω^-	0.0007 ± 0.0004	0.014 ± 0.007	0.0016 ± 0.0003	
Λ_c^+	$0.074 \pm 0.031^{(i)}$	0.110 ± 0.050	0.078 ± 0.017	
Λ_b^0	—	—	0.031 ± 0.016	
$\Sigma_c^{++}, \Sigma_c^0$	0.014 ± 0.007	—	—	
$\Lambda(1520)$	0.008 ± 0.002	—	0.0222 ± 0.0027	

Notes for Table 40.1:

(*a*) B($D_s \to \eta\pi,\ \eta'\pi$) was used (RPP1994).

(*b*) The Standard Model B($Z \to b\overline{b}$) = 0.217 was used.

(*c*) $x_p = p/p_{\rm beam} > 0.1$ only.

(*d*) Both charge states.

(*e*) Any charge state (*i.e.*, B_d^*, B_u^*, or B_s^*).

(*f*) B($Z \to$ hadrons) = 0.699 was used (RPP1994).

(*g*) B($D_s^* \to D_S^+\gamma$), B($D_s^+ \to \phi\pi^+$), B($\phi \to K^+K^-$) have been used (RPP1998).

(*h*) Any charge state (*i.e.*, B_d^{**}, B_u^{**}, or B_s^{**}).

(*i*) The value was derived from the cross section of $\Lambda_c^+ \to p\pi K$ using (k) and assuming the branching fraction to be $(5.0 \pm 1.3)\%$ (RPP2004).

(*j*) B($D^*(2010)^+ \to D^0\pi^+$) × B($D^0 \to K^-\pi^+$) has been used (RPP2000).

(*k*) $\sigma_{\rm had} = 3.33 \pm 0.05 \pm 0.21$ nb (**CLEO**: Phys. Rev. **D29**, 1254 (1984)) has been used in converting the measured cross sections to average hadron multiplicities.

(ℓ) Assumes B($D_{s1}^+ \to D^{*+}K^0 + D^{*0}K+$) = 100% and B($D_{s2}^+ \to D^0K^+$) = 45%.

References for Table 40.1:

RPP1992: Phys. Rev. **D45** (1992) and references therein.

RPP1994: Phys. Rev. **D50**, 1173 (1994) and references therein.

RPP1996: Phys. Rev. **D54**, 1 (1996) and references therein.

RPP1998: Eur. Phys. J. **C3**, 1 (1998) and references therein.

RPP2000: Eur. Phys. J. **C15**, 1 (2000) and references therein.

RPP2002: Phys. Rev. **D66**, 010001 (2002) and references therein.

RPP2004: Phys. Lett. **B592**, 1 (2004) and references therein.

R. Marshall, Rep. Prog. Phys. **52**,1329(1989).

A. De Angelis, J. Phys. **G19**, 1233 (1993) and references therein.

ALEPH: D. Buskulic *et al.*: Phys. Lett. **B295**, 396 (1992); Z. Phys. **C64**, 361 (1994); **C69**, 15 (1996); **C69**, 379 (1996); **C73**, 409 (1997); and
R. Barate *et al.*: Z. Phys. **C74**, 451 (1997); Phys. Reports **294**, 1 (1998); Eur. Phys. J. **C5**, 205 (1998); **C16**, 597 (2000); **C16**, 613 (2000); and
A. Heister *et al.*: Phys. Lett. **B526**, 34 (2002); **B528**, 19 (2002).

ARGUS: H. Albrecht *et al.*: Phys. Lett. **230B**, 169 (1989); Z. Phys. **C44**, 547 (1989); **C46**, 15 (1990); **C54**, 1 (1992); **C58**, 199 (1993); **C61**, 1 (1994); Phys. Rep. **276**, 223 (1996).

BaBar: B. Aubert *et al.*: Phys. Rev. Lett. **87**, 162002 (2001); Phys. Rev. **D65**, 091104 (2002).

Belle: K. Abe *et al.*: Phys. Rev. Lett. **88**, 052001 (2002); and
R. Seuster *et al.*: hep-ex/0506068.

CELLO: H.J. Behrend *et al.*: Z. Phys. **C46**, 397 (1990); **C47**, 1 (1990).

CLEO: D. Bortoletto *et al.*, Phys. Rev. **D37**, 1719 (1988); erratum ibid **D39**, 1471 (1989); and
M. Artuso *et al.*, Phys. Rev. **D70**, 112001 (2004).

Crystal Ball: Ch. Bieler *et al.*, Z. Phys. **C49**, 225 (1991).

DELPHI: P. Abreu *et al.*: Z. Phys. **C57**, 181 (1993); **C59**, 533 (1993); **C61**, 40 7(1994); **C65**, 587 (1995); **C67**, 543 (1995); **C68**, 353 (1995); **C73**, 61 (1996); Nucl. Phys. **B444**, 3 (1995); Phys. Lett. **B341**, 109 (1994); **B345**, 598 (1995); **B361**, 207 (1995); **B372**, 172 (1996); **B379**, 309 (1996); **B416**, 233 (1998); **B449**, 364 (1999); **B475**, 429 (2000); Eur. Phys. J. **C6**, 19 (1999); **C5**, 585 (1998); **C18**, 203 (2000); and
J. Abdallah *et al.*, Phys. Lett. **B569**, 129 (2003); Phys. Lett. **B576**, 29 (2003).

HRS: S. Abachi *et al.*, Phys. Rev. Lett. **57**, 1990 (1986); and
M. Derrick *et al.*, Phys. Rev. **D35**, 2639 (1987).

L3: M. Acciarri *et al.*: Phys. Lett. **B328**, 223 (1994); **B345**, 589 (1995); **B371**, 126 (1996); **B371**, 137 (1996); **B393**, 465 (1997); **B404**, 390 (1997); **B407**, 351 (1997); **B407**, 389 (1997), erratum ibid. **B427**, 409 (1998); **B453**, 94 (1999); **B479**, 79 (2000).

MARK II: H. Schellman *et al.*, Phys. Rev. **D31**, 3013 (1985); and
G. Wormser *et al.*, Phys. Rev. Lett. **61**, 1057 (1988).

JADE: W. Bartel *et al.*, Z. Phys. **C20**, 187 (1983); andD.D. Pietzl *et al.*, Z. Phys. **C46**, 1 (1990).

OPAL: R. Akers *et al.*: Z. Phys. **C63**, 181 (1994); **C66**, 555 (1995); **C67**, 389 (1995); **C68**, 1 (1995); and
G. Alexander *et al.*: Phys. Lett. **B358**, 162 (1995); Z. Phys. **C70**, 197 (1996); **C72**, 1 (1996); **C72**, 191 (1996); **C73**, 569 (1997); **C73**, 587 (1997); Phys. Lett. **B370**, 185 (1996); and
K. Ackerstaff *et al.*: Z. Phys. **C75**, 192 (1997); Phys. Lett. **B412**, 210 (1997); Eur. Phys. J. **C1**, 439 (1998); **C4**, 19 (1998); **C5**, 1 (1998); **C5**, 411 (1998); and
G. Abbiendi *et al.*: Eur. Phys. J. **C16**, 185 (2000); **C16**, 185 (2000).

PLUTO: Ch. Berger *et al.*, Phys. Lett. **104B**, 79 (1981).

SLD: K. Abe, Phys. Rev. **D59**, 052001 (1999); Phys. Rev. **D69**, 072003 (2004).

TASSO: H. Aihara *et al.*, Z. Phys. **C27**, 27 (1985).

TPC: H. Aihara *et al.*, Phys. Rev. Lett. **53**, 2378 (1984).

Average e^+e^-, pp, and $\overline{p}p$ Multiplicity

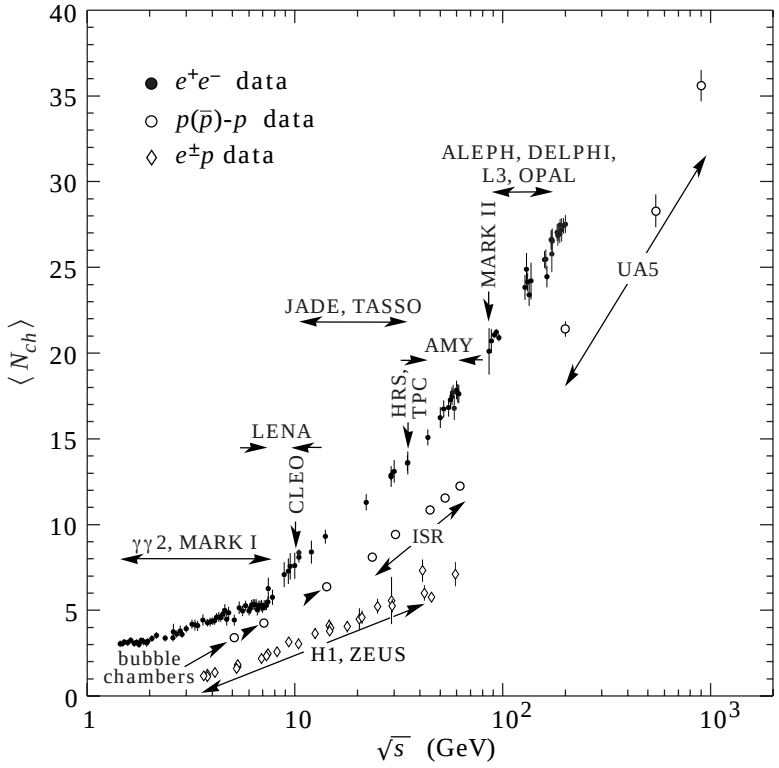


Figure 40.5: Average multiplicity as a function of $\sqrt{s}$ for e^+e^- and $p\overline{p}$ annihilations, and pp and ep collisions. The indicated errors are statistical and systematic errors added in quadrature, except when no systematic errors are given. Files of the data shown in this figure are given in `http://home.cern.ch/b/biebel/www/RPP06/`.

—

e^+e^-: Most e^+e^- measurements include contributions from K^0_S and Λ decays. The $\gamma\gamma 2$ and MARK I measurements contain a systematic 5% error. Points at identical energies have been spread horizontally for clarity:

ALEPH: D. Buskulic *et al.*, Z. Phys. **C69**, 15 (1995); and Z. Phys. **C73**, 409 (1997);
A. Heister *et al.*, Eur. Phys. J. **C35**, 457 (2004).

ARGUS: H. Albrecht *et al.*, Z. Phys. **C54**, 13 (1992).

DELPHI: P. Abreu *et al.*, Eur. Phys. J. **C6**, 19 (1999); Phys. Lett. **B372**, 172 (1996); Phys. Lett. **B416**, 233 (1998); and Eur. Phys. J. **C18**, 203 (2000).

L3: M. Acciarri *et al.*, Phys. Lett. **B371**, 137 (1996); Phys. Lett. **B404**, 390 (1997); and Phys. Lett. **B444**, 569 (1998);
P. Achard *et al.*, Phys. Reports **339**, 71 (2004).

OPAL: G. Abbiendi *et al.*, Eur. Phys. J. **C16**, 185 (2000); and Eur. Phys. J. **C37**, 25 (2004);
K. Ackerstaff *et al.*, Z. Phys. **C75**, 193 (1997);
P.D. Acton *et al.*, Z. Phys. **C53**, 539 (1992) and references therein;
R. Akers *et al.*, Z. Phys. **C68**, 203 (1995).

TOPAZ: K. Nakabayashi *et al.*, Phys. Lett. **B413**, 447 (1997).

VENUS: K. Okabe *et al.*, Phys. Lett. **B423**, 407 (1998).

—

$e^{\pm}p$: Multiplicities have been measured in the current fragmentation region of the Breit frame:

H1: C. Adloff *et al.*, Nucl. Phys. **B504**, 3 (1997).

ZEUS: J. Breitweg *et al.*, Eur. Phys. J. **C11**, 251 (1999);
S. Chekanov *et al.*, Phys. Lett. **B510**, 36 (2001).

—

$p(\overline{p})$: The errors of the $p(\overline{p})$ measurements are the quadratically added statistical and systematic errors, except for the bubble chamber measurements for which only statistical errors are given in the references. The values measured by UA5 exclude single diffractive dissociation:

bubble chamber: J. Benecke *et al.*, Nucl. Phys. **B76**, 29 (1976); W.M. Morse *et al.*, Phys. Rev. **D15**, 66 (1977).

ISR: A. Breakstone *et al.*, Phys. Rev. **D30**, 528 (1984).

UA5: G.J. Alner *et al.*, Phys. Lett. **167B**, 476 (1986);
R.E. Ansorge *et al.*, Z. Phys. **C43**, 357 (1989).

(Courtesy of O. Biebel, LMU, München, 2005.)

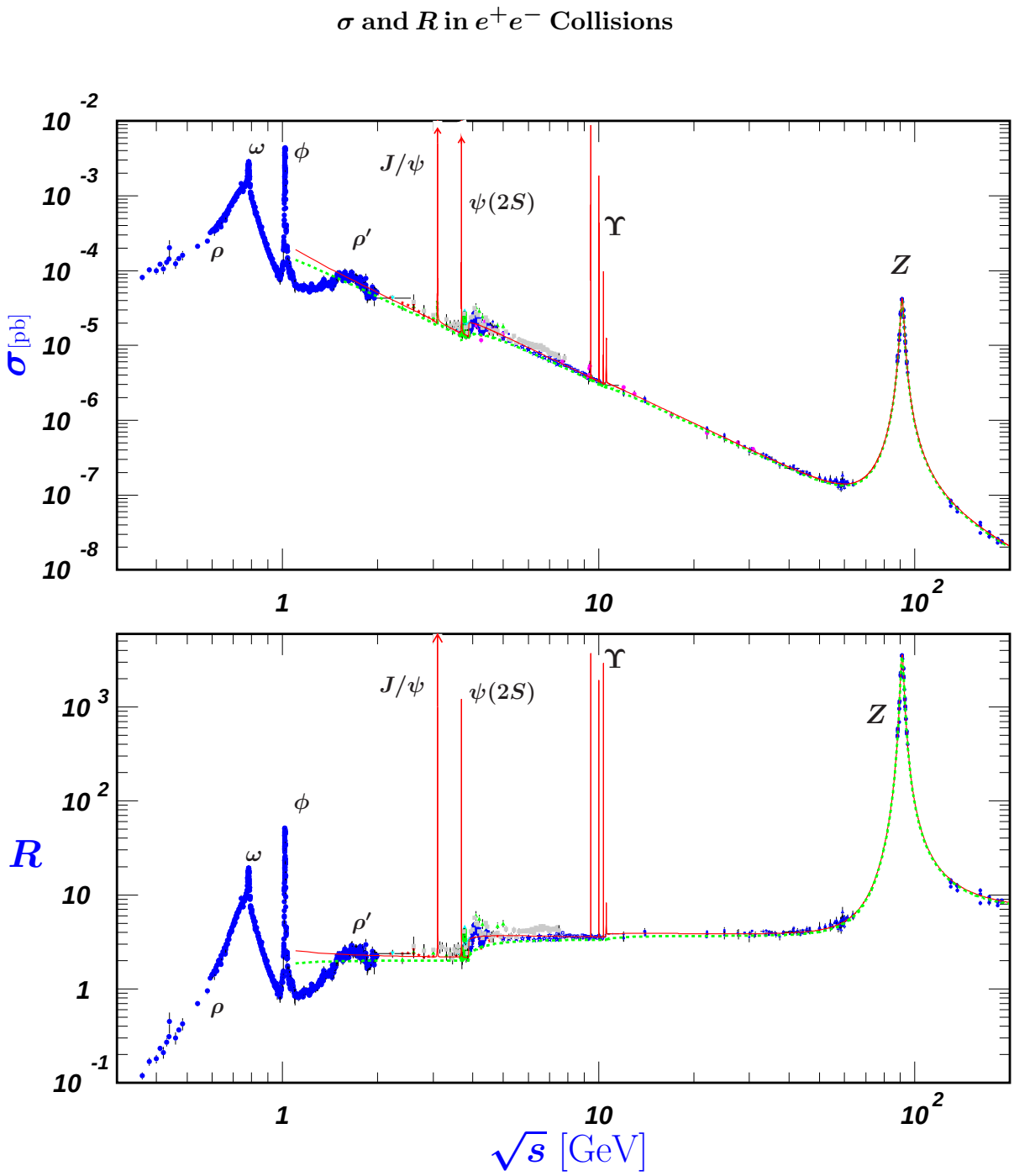


Figure 40.6: World data on the total cross section of $e^+e^- \to hadrons$ and the ratio $R(s) = \sigma(e^+e^- \to hadrons, s)/\sigma(e^+e^- \to \mu^+\mu^-, s)$. $\sigma(e^+e^- \to hadrons, s)$ is the experimental cross section corrected for initial state radiation and electron-positron vertex loops, $\sigma(e^+e^- \to \mu^+\mu^-, s) = 4\pi\alpha^2(s)/3s$. Data errors are total below 2 GeV and statistical above 2 GeV. The curves are an educative guide: the broken one (green) is a naive quark-parton model prediction and the solid one (red) is 3-loop pQCD prediction (see "Quantum Chromodynamics" section of this *Review*, Eq. (9.12) or, for more details, K. G. Chetyrkin *et al.*, Nucl. Phys. **B586**, 56 (2000) (Erratum *ibid.* **B634**, 413 (2002)). Breit-Wigner parameterizations of J/ψ, $\psi(2S)$, and $\Upsilon(nS), n = 1, 2, 3, 4$ are also shown. The full list of references to the original data and the details of the R ratio extraction from them can be found in `[arXiv:hep-ph/0312114]`. Corresponding computer-readable data files are available at `http://pdg.ihep.su/xsect/contents.html`. (Courtesy of the COMPAS(Protvino) and HEPDATA(Durham) Groups, August 2005. Corrections by P. Janot (CERN) and M. Schmitt (Northwestern U.)) See full-color version on color pages at end of book.

R in Light-Flavor, Charm, and Beauty Threshold Regions

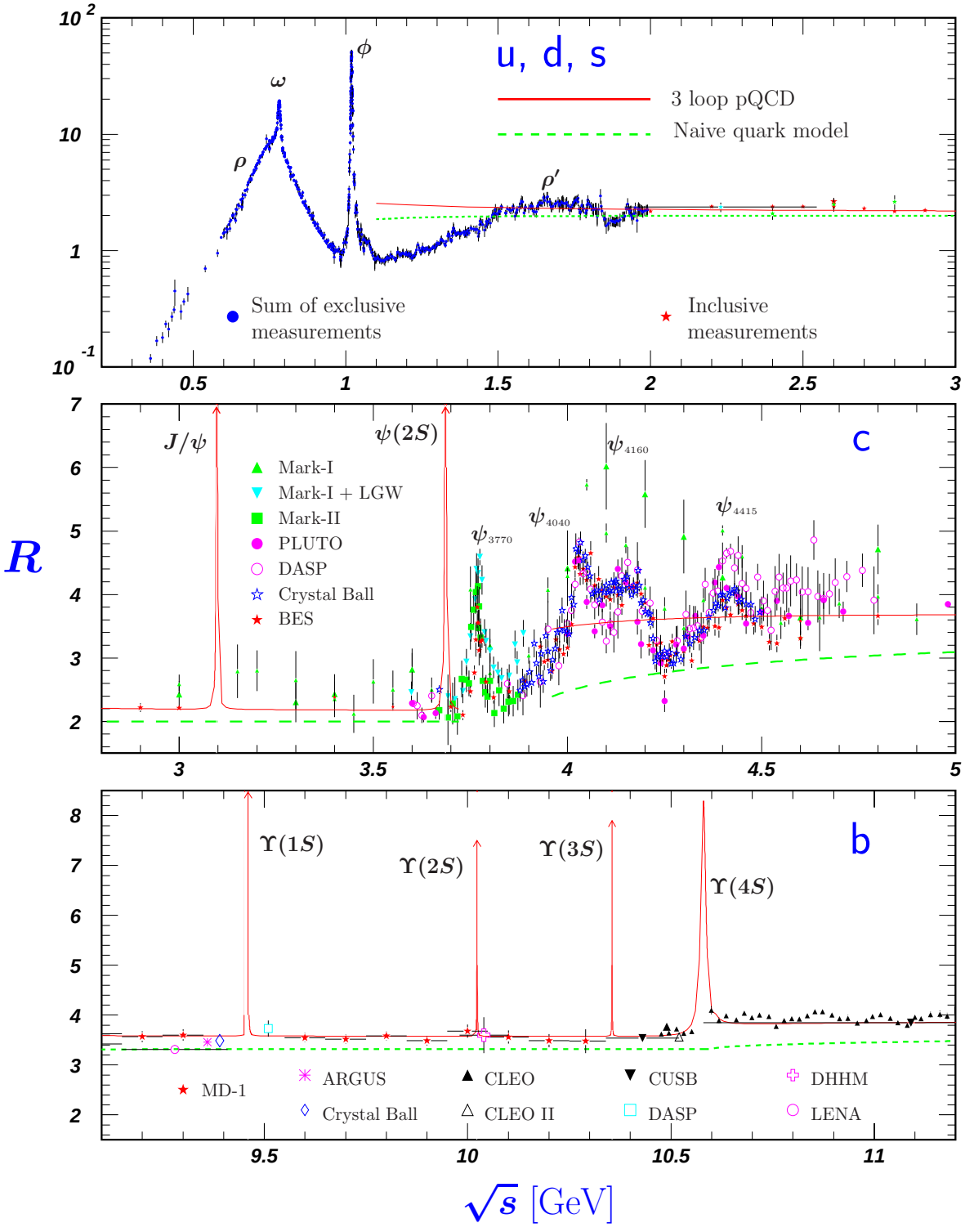


Figure 40.7: R in the light-flavor, charm, and beauty threshold regions. Data errors are total below 2 GeV and statistical above 2 GeV. The curves are the same as in Fig. 40.6. **Note:** CLEO data above $\Upsilon(4S)$ were not fully corrected for radiative effects, and we retain them on the plot only for illustrative purposes with a normalization factor of 0.8. The full list of references to the original data and the details of the R ratio extraction from them can be found in [arXiv:hep-ph/0312114]. The computer-readable data are available at http://pdg.ihep.su/xsect/contents.html (Courtesy of the COMPAS(Protvino) and HEPDATA(Durham) Groups, August 2005.) See full-color version on color pages at end of book.

Annihilation Cross Section Near M_Z

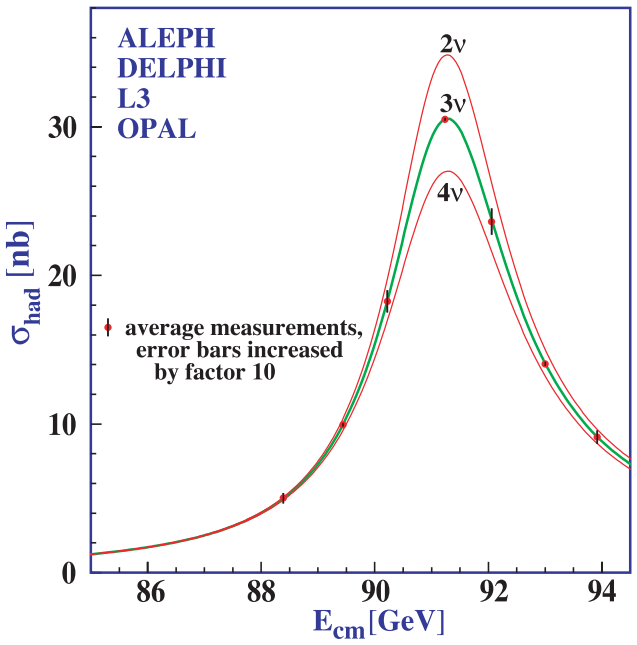


Figure 40.8: Combined data from the ALEPH, DELPHI, L3, and OPAL Collaborations for the cross section in e^+e^- annihilation into hadronic final states as a function of the center-of-mass energy near the Z pole. The curves show the predictions of the Standard Model with two, three, and four species of light neutrinos. The asymmetry of the curve is produced by initial-state radiation. Note that the error bars have been increased by a factor ten for display purposes. References:

ALEPH: R. Barate *et al.*, Eur. Phys. J. **C14**, 1 (2000).
DELPHI: P. Abreu *et al.*, Eur. Phys. J. **C16**, 371 (2000).
L3: M. Acciarri *et al.*, Eur. Phys. J. **C16**, 1 (2000).
OPAL: G. Abbiendi *et al.*, Eur. Phys. J. **C19**, 587 (2001).
Combination: The Four LEP Collaborations (ALEPH, DELPHI, L3, OPAL) and the Lineshape Sub-group of the LEP Electroweak Working Group, `hep-ph/0101027`.

(Courtesy of M. Grünewald and the LEP Electroweak Working Group, 2003)

Muon Neutrino and Anti-Neutrino Charged-Current Total Cross Section

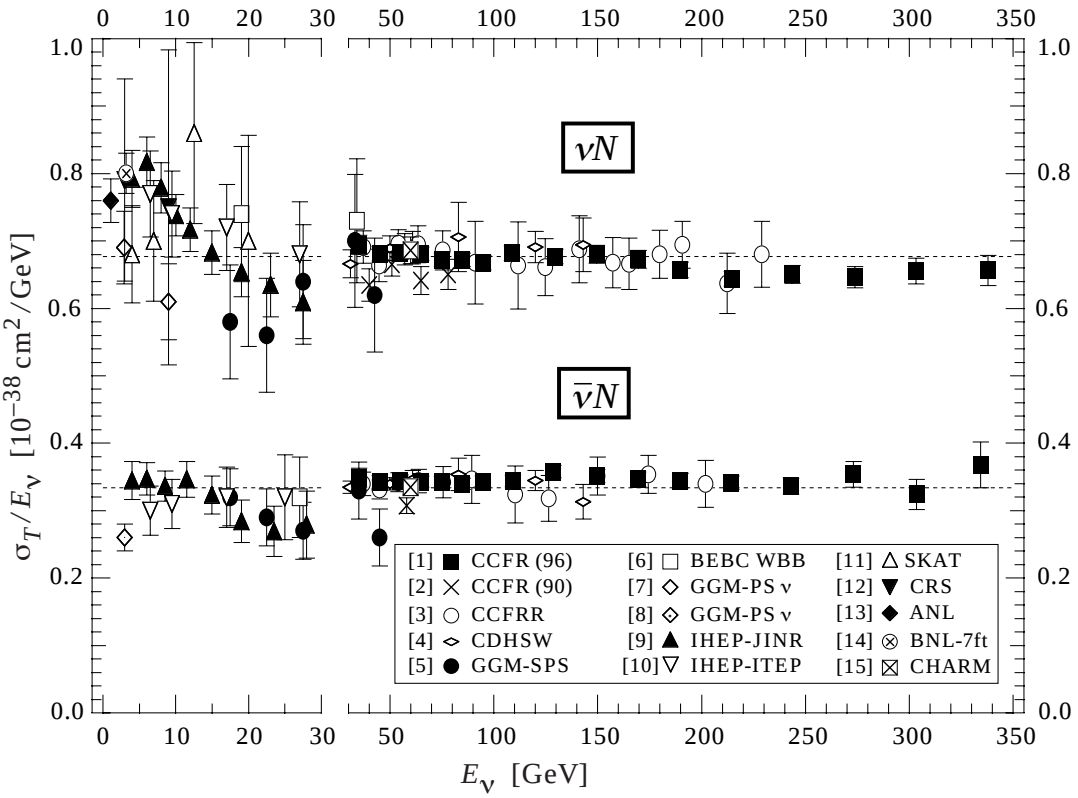


Figure 40.9: σ_T/E_ν, for the muon neutrino and anti-neutrino charged-current total cross section as a function of neutrino energy. The error bars include both statistical and systematic errors. The straight lines are the averaged values over all energies as measured by the experiments in Refs. [1–4]: $= 0.677 \pm 0.014$ $(0.334 \pm 0.008) \times 10^{-38}$ cm^2/GeV. Note the change in the energy scale at 30 GeV. (Courtesy W. Seligman and M.H. Shaevitz, Columbia University, 2001.)

[1] W. Seligman, Ph.D. Thesis, Nevis Report 292 (1996);
[2] P.S. Auchincloss *et al.*, Z. Phys. **C48**, 411 (1990);
[3] D.B. MacFarlane *et al.*, Z. Phys. **C26**, 1 (1984);
[4] P. Berge *et al.*, Z. Phys. **C35**, 443 (1987);
[5] J. Morfin *et al.*, Phys. Lett. **104B**, 235 (1981);
[6] D.C. Colley *et al.*, Z. Phys. **C2**, 187 (1979);
[7] S. Campolillo *et al.*, Phys. Lett. **84B**, 281 (1979);
[8] O. Erriquez *et al.*, Phys. Lett. **80B**, 309 (1979);
[9] V.B. Anikeev *et al.*, Z. Phys. **C70**, 39 (1996);
[10] A.S. Vovenko *et al.*, Sov. J. Nucl. Phys. **30**, 527 (1979);
[11] D.S. Baranov *et al.*, Phys. Lett. **81B**, 255 (1979);
[12] C. Baltay *et al.*, Phys. Rev. Lett. **44**, 916 (1980);
[13] S.J. Barish *et al.*, Phys. Rev. **D19**, 2521 (1979);
[14] N.J. Baker *et al.*, Phys. Rev. **D25**, 617 (1982);
[15] J.V. Allaby *et al.*, Z. Phys. **C38**, 403 (1988).

Table 40.2: Total hadronic cross section. Analytic S-matrix and Regge theory suggest a variety of parameterizations of total cross sections at high energies with different areas of applicability and fits quality.

A ranking procedure, based on measures of different aspects of the quality of the fits to the current evaluated experimental database, allows one to single out the following parameterization of highest rank[1]

$$\sigma^{ab} = Z^{ab} + B\log^2(s/s_0) + Y_1^{ab}(s_1/s)^{\eta_1} - Y_2^{ab}(s_1/s)^{\eta_2}\ , \qquad \sigma^{\overline{a}b} = Z^{ab} + B\log^2(s/s_0) + Y_1^{ab}(s_1/s)^{\eta_1} + Y_2^{ab}(s_1/s)^{\eta_2}$$

where Z^{ab}, B, Y_i^{ab} are in mb, s, s_1, and s_0 are in GeV^2. The scales s_0, s_1, the rate of universal rise of the cross sections B, and exponents η_1 and η_2 are independent of the colliding particles. The scale s_1 is fixed at 1 GeV^2. Terms $Z^{ab} + B\log^2(s/s_0)$ represent the pomerons. The exponents η_1 and η_2 represent lower-lying C-even and C-odd exchanges, respectively. Requiring $\eta_1 = \eta_2$ results in somewhat poorer fits. In addition to total cross sections σ, the measured ratios of the real-to-imaginary parts of the forward scattering amplitudes $\rho = \mathrm{Re}(T)/\mathrm{Im}(T)$ were included in the fits by using s to u crossing symmetry. Global fits were made to the 2005-updated data for $\overline{p}(p)p$, $\Sigma^- p$, $\pi^\pm p$, $K^\pm p$, γp, and $\gamma\gamma$ collisions.

Exact factorization hypothesis in the form $(Z^{\gamma p}, B^{\gamma p}) = \delta \cdot (Z^{pp}, B)$, $(Z^{\gamma\gamma}, B^{\gamma\gamma}) = \delta^2 \cdot (Z^{pp}, B)$ was used to extend the universal rise of the total hadronic cross sections to the $\gamma p \to$ *hadrons* and $\gamma\gamma \to$ *hadrons* collisions. This resulted in reducing the number of adjusted parameters from 21 used for the 2002 edition to 19, and in the higher quality rank of the parameterization. The asymptotic parameters thus obtained were then fixed and used as inputs to a fit to a larger data sample that included cross sections on deuterons (d) and neutrons (n). All fits included data above $\sqrt{s_{\min}} = 5$ GeV.

Fits to $\overline{p}(p)\,p$, $\Sigma^- p$, $\pi^\pm p$, $K^\pm p$, γp, $\gamma\gamma$			Beam/	Fits to groups				χ^2/dof
Z	Y_1	Y_2	Target	Z	Y_1	Y_2	B	by groups
35.45(48)	42.53(1.35)	33.34(1.04)	$\overline{p}(p)/p$	35.45(48)	42.53(23)	33.34(33)	0.308(10)	
			$\overline{p}(p)n$	35.80(16)	40.15(1.59)	30.00(96)	0.308(10)	1.029
35.20(1.46)	−199(102)	−264(126)	Σ^-/p	35.20(1.41)	−199(86)	−264(112)	0.308(10)	0.565
20.86(40)	19.24(1.22)	6.03(19)	$\pi^\pm/p$	20.86(3)	19.24(18)	6.03(9)	0.308(10)	0.955
17.91(36)	7.1(1.5)	13.45(40)	$K^\pm/p$	17.91(3)	7.14(25)	13.45(13)	0.308(10)	
			$K^\pm/n$	17.87(6)	5.17(50)	7.23(28)	0.308(10)	0.669
0.0317(6)			γ/p		0.0320(40)		0.308(10)	
−0.61(62)E−3			γ/γ		−0.58(61)E−3		0.308(10)	0.766
$\chi^2/dof = 0.971$,	$B = 0.308(10)$ mb,		$\overline{p}(p)/d$	64.35(38)	130(3)	85.5(1.3)	0.537(31)	1.432
$\eta_1 = 0.458(17)$,	$\eta_2 = 0.545(7)$		$\pi^\pm/d$	38.62(21)	59.62(1.53)	1.60(41)	0.461(14)	0.735
$\delta = 0.00308(2)$,	$\sqrt{s_0} = 5.38(50)$ GeV		$K^\pm/d$	33.41(20)	23.66(1.45)	28.70(37)	0.449(14)	0.814

The fitted functions are shown in the following figures, along with one-standard-deviation error bands. When the reduced χ^2 is greater than one, a scale factor has been included to evaluate the parameter values, and to draw the error bands. Where appropriate, statistical and systematic errors were combined quadratically in constructing weights for all fits. On the plots, only statistical error bars are shown. Vertical arrows indicate lower limits on the $p_{\rm lab}$ or $E_{\rm cm}$ range used in the fits.

One can find the details of the global fits and ranking procedure, in the paper [1]. Database is practically the same as for the 2004 edition (it was slightly changed in the low energy regions not used in the fits).

Recently, the statement in [1] that the models with $\log^2(s/s_0)$ asymptotic terms work much better than the models with $\log(s/s_0)$ or $(s/s_0)^\epsilon$ terms was confirmed in [2] and [3], based on matching traditional asymptotic parameterizations with low energy data in different ways. Both these references, however, questioned the statement in [1] on the universality of the coefficient of the $\log^2(s/s_0)$ term for all processes with nucleon and gamma targets. The two references give different predictions at superhigh energies: $\sigma^{as}_{\pi N} > \sigma^{as}_{NN}$ [2] and $\sigma^{as}_{\pi N} \sim 2/3\sigma^{as}_{NN}$ [3]. A broader universality of σ^{as}_{tot} has been recently advocated in [4] for hadron-nucleus collisions. It should be noted that asymptotic rate universality in hadron-deuteron collisions has not been established at available energies (see table).

Computer-readable data files extracted from the PPDS (`http://wwwppds.ihep.su:8001/ppds.html`) are also available at `pdg.lbl.gov`. (Courtesy of the COMPAS group, IHEP, Protvino, August 2005.)

On-line "Predictor" to calculate σ and ρ for any energy from five high rank models is also available at `http://nuclth02.phys.ulg.ac.be/compete/predictor.html`.

References:

1. J.R. Cudell *et al.* (COMPETE Collab.), Phys. Rev. **D65**, 074024 (2002).
2. K. Igi and M. Ishida, Phys. Rev. **D66**, 034023 (2002), Phys. Lett. **B622**, 286 (2005)
3. M. M. Block and F. Halzen, Phys. Rev. **D70**, 091901 (2004), Phys. Rev. **D72**, 036006 (2005)
4. L. Frankfurt, M. Strikman, and M. Zhalov, Phys. Lett. **B616**, 59 (2005)

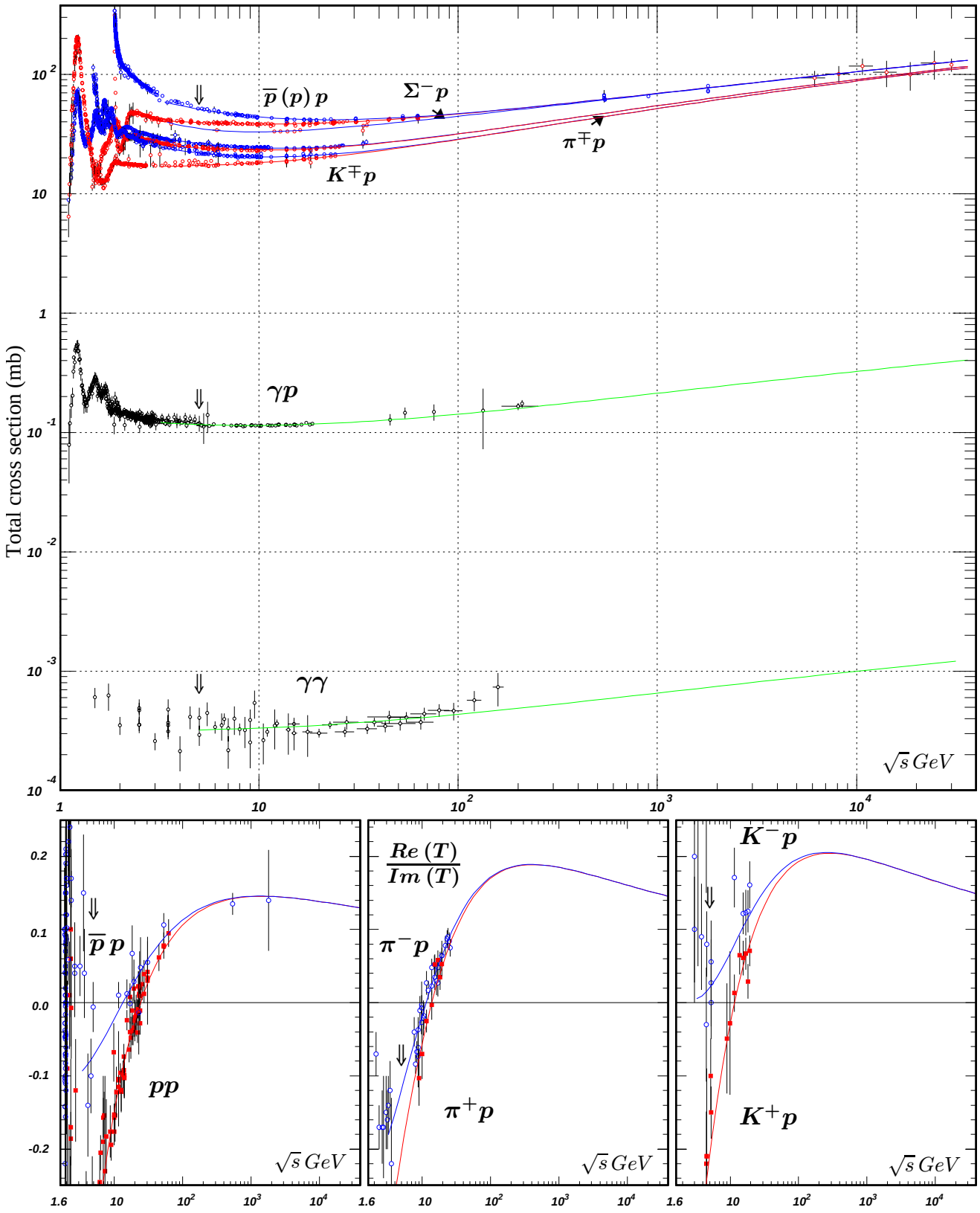


Figure 40.10: Summary of hadronic, γp, and $\gamma\gamma$ total cross sections, and ratio of the real to imaginary parts of the forward hadronic amplitudes. Corresponding computer-readable data files may be found at `http://pdg.lbl.gov/xsect/contents.html`. (Courtesy of the COMPAS group, IHEP, Protvino, August 2005.) See full-color version on color pages at end of book.

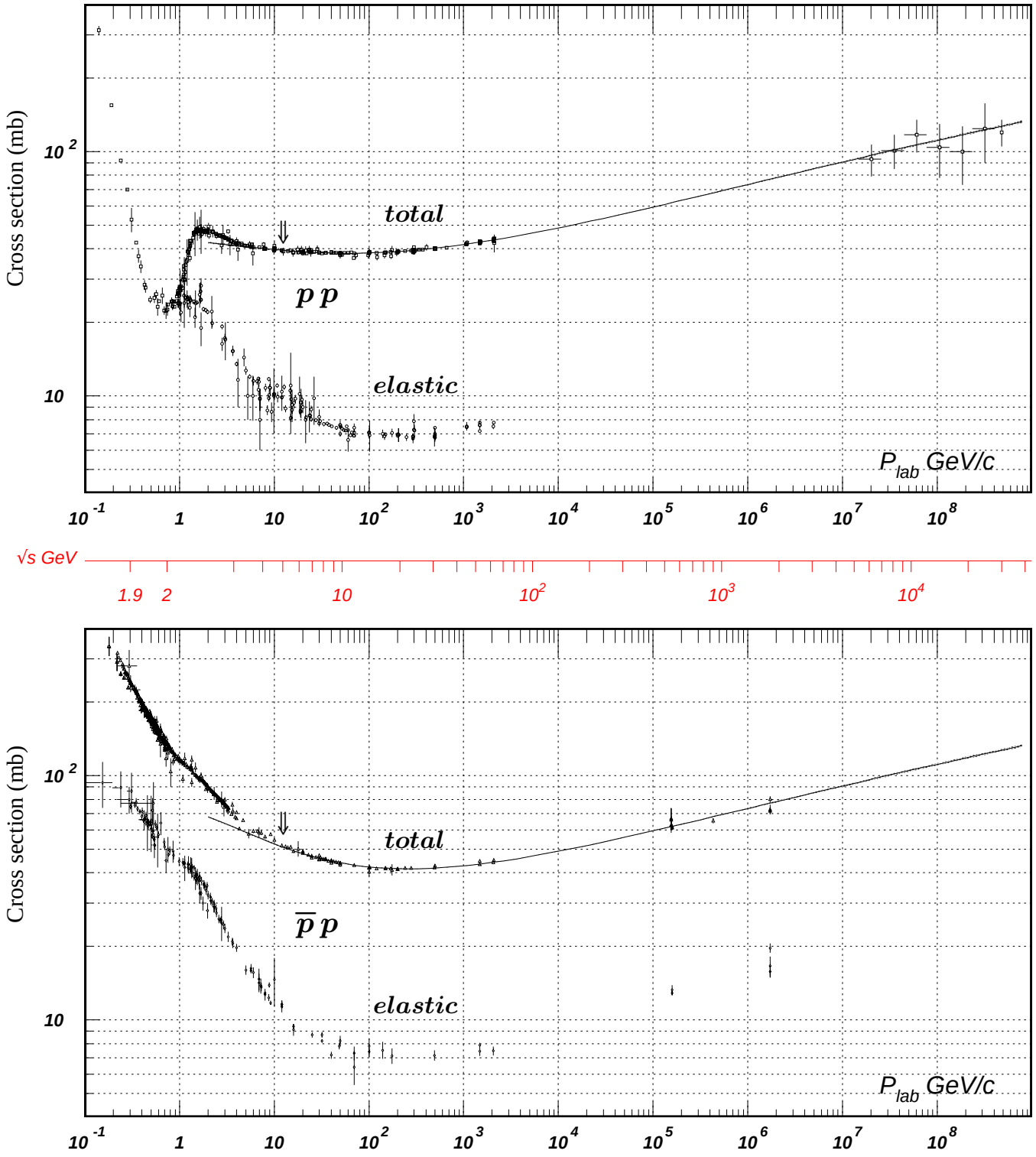


Figure 40.11: Total and elastic cross sections for pp and $\overline{p}p$ collisions as a function of laboratory beam momentum and total center-of-mass energy. Corresponding computer-readable data files may be found at `http://pdg.lbl.gov/xsect/contents.html`. (Courtesy of the COMPAS group, IHEP, Protvino, August 2005.)

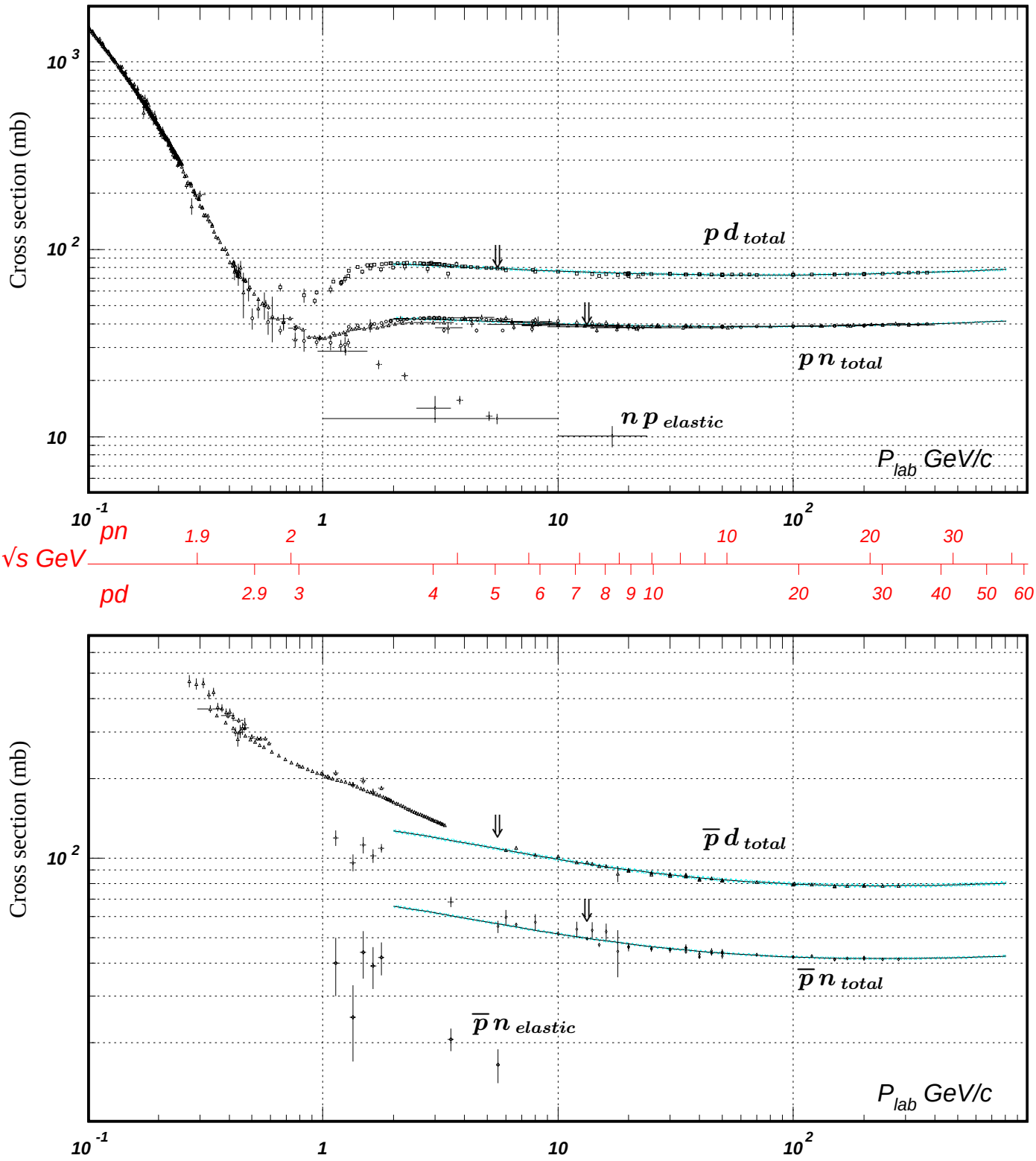


Figure 40.12: Total and elastic cross sections for pd (total only), np, $\overline{p}d$ (total only), and $\overline{p}n$ collisions as a function of laboratory beam momentum and total center-of-mass energy. Corresponding computer-readable data files may be found at `http://pdg.lbl.gov/xsect/contents.html`. (Courtesy of the COMPAS Group, IHEP, Protvino, August 2005.)

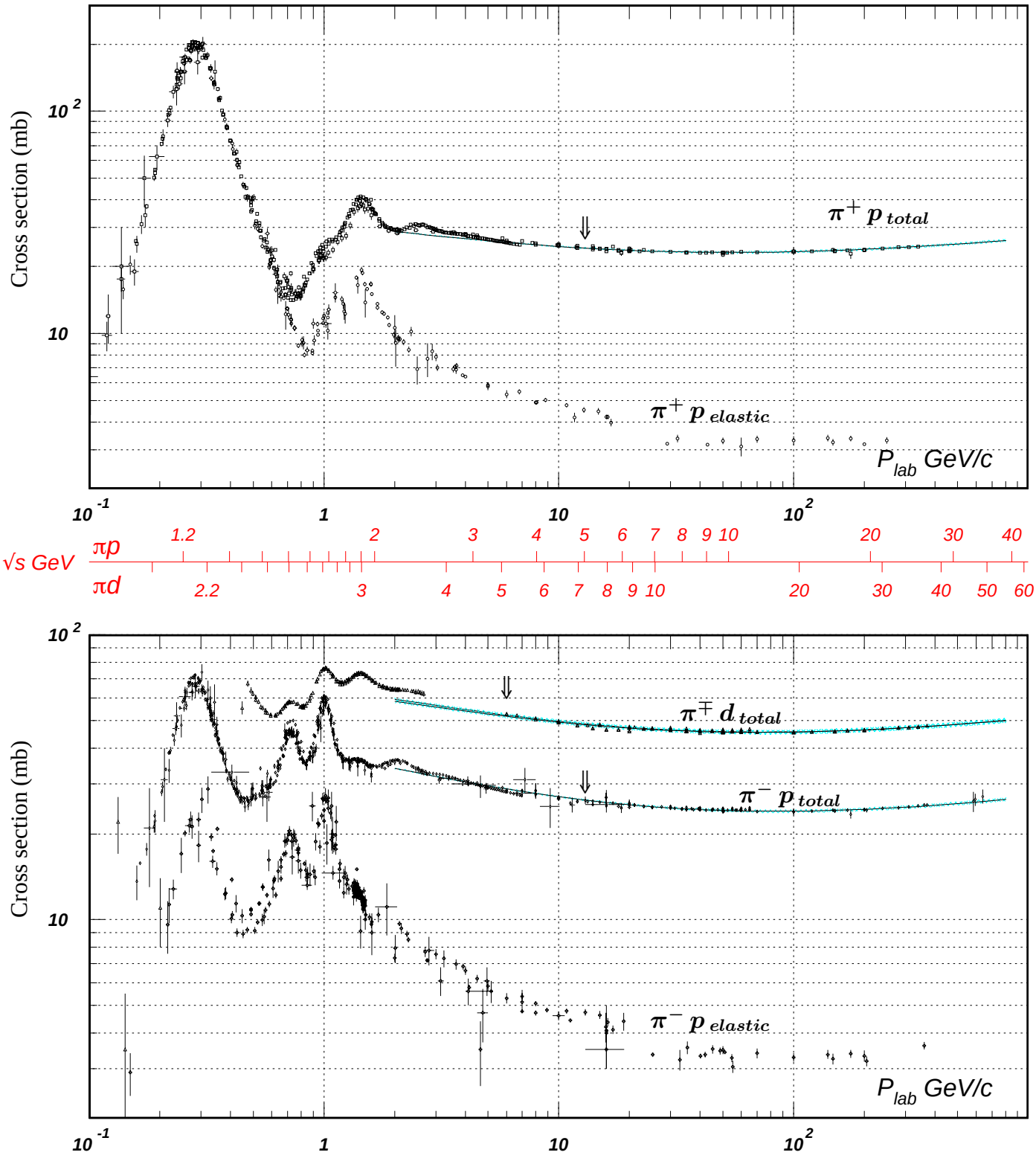


Figure 40.13: Total and elastic cross sections for $\pi^{\pm}p$ and $\pi^{\pm}d$ (total only) collisions as a function of laboratory beam momentum and total center-of-mass energy. Corresponding computer-readable data files may be found at `http://pdg.lbl.gov/xsect/contents.html`. (Courtesy of the COMPAS Group, IHEP, Protvino, August 2005.)

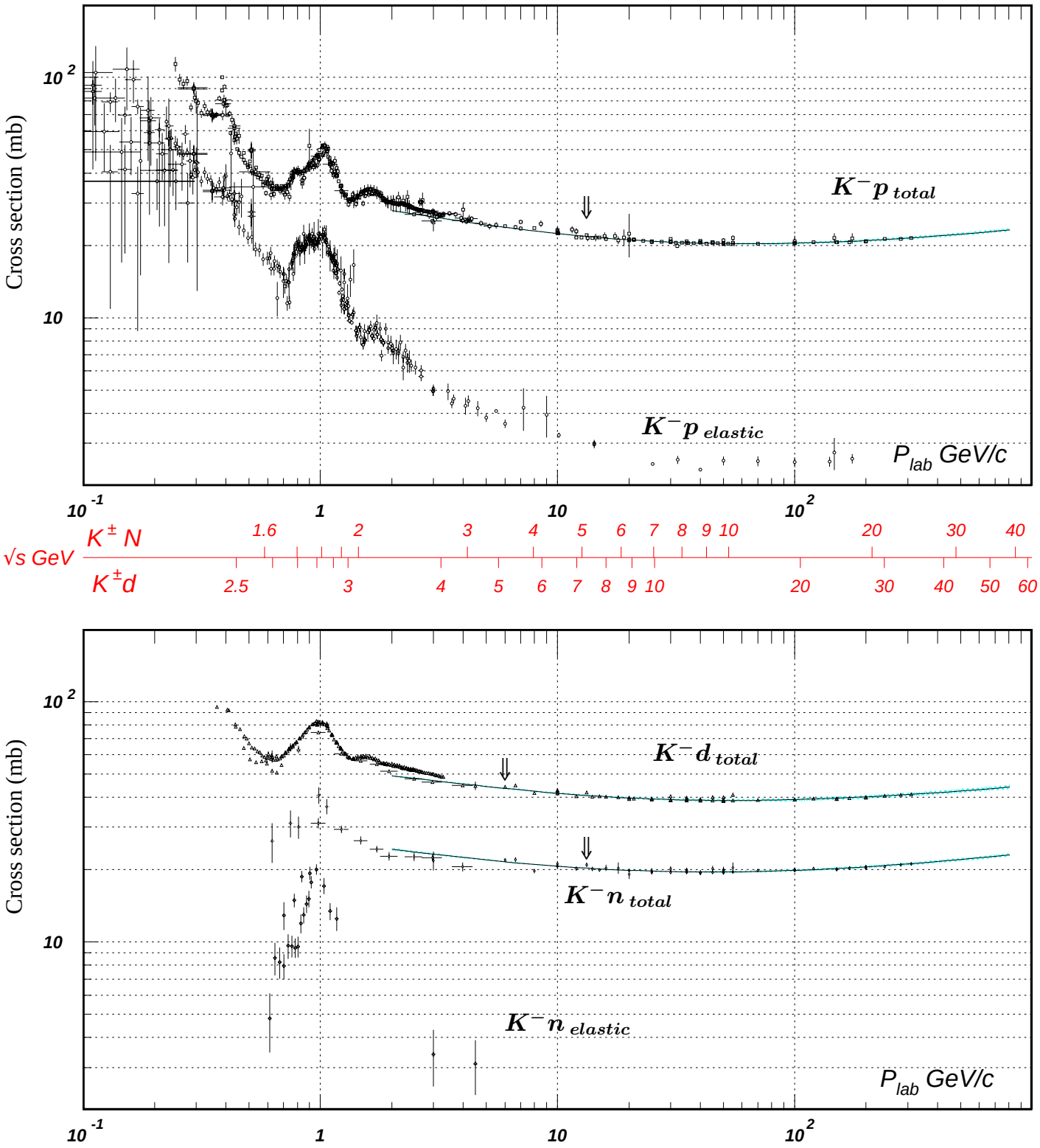


Figure 40.14: Total and elastic cross sections for $K^- p$ and $K^- d$ (total only), and $K^- n$ collisions as a function of laboratory beam momentum and total center-of-mass energy. Corresponding computer-readable data files may be found at `http://pdg.lbl.gov/xsect/contents.html`. (Courtesy of the COMPAS Group, IHEP, Protvino, August 2005.)

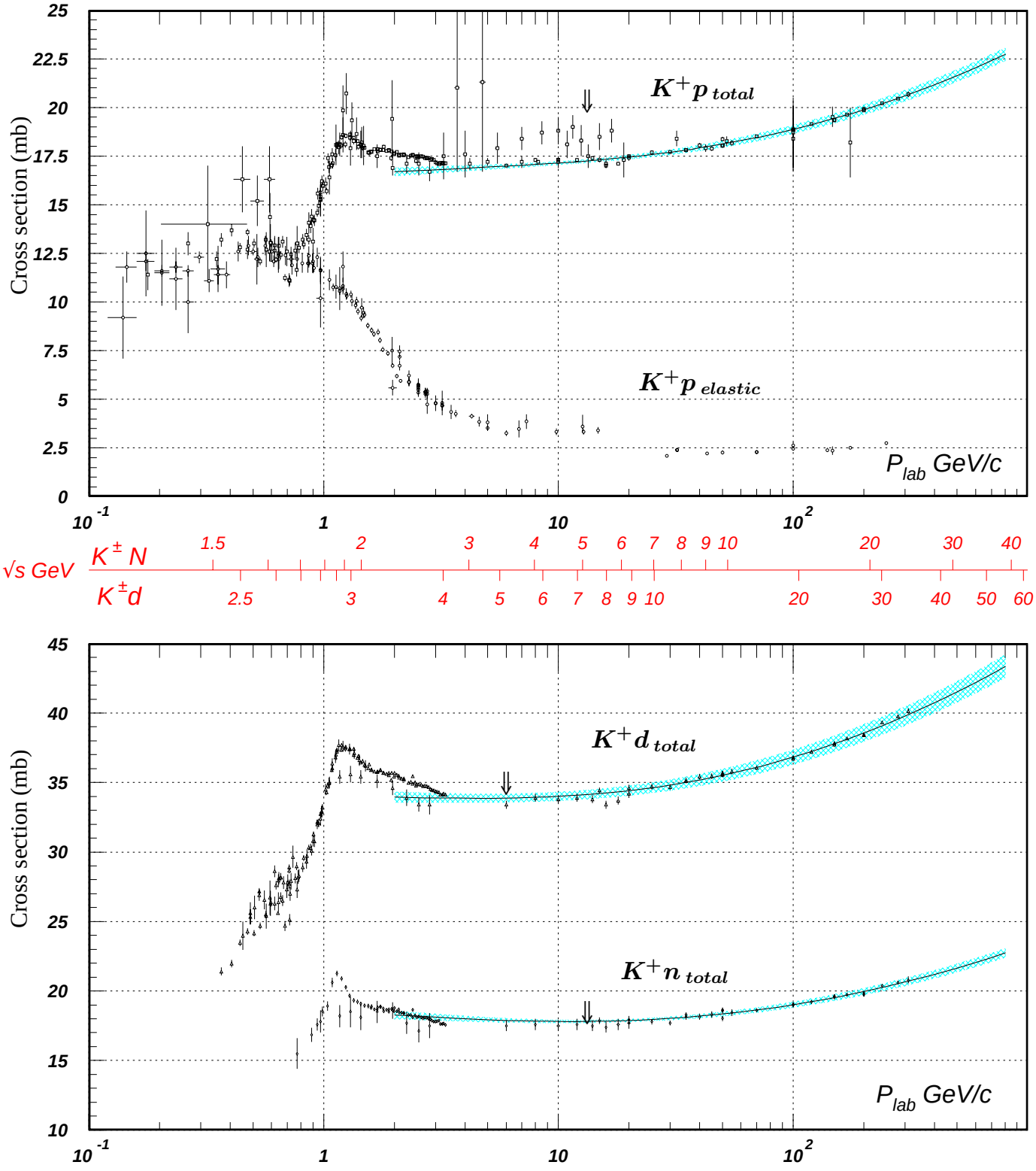


Figure 40.15: Total and elastic cross sections for K^+p and total cross sections for K^+d and K^+n collisions as a function of laboratory beam momentum and total center-of-mass energy. Corresponding computer-readable data files may be found at `http://pdg.lbl.gov/xsect/contents.html`. (Courtesy of the COMPAS Group, IHEP, Protvino, August 2005.)

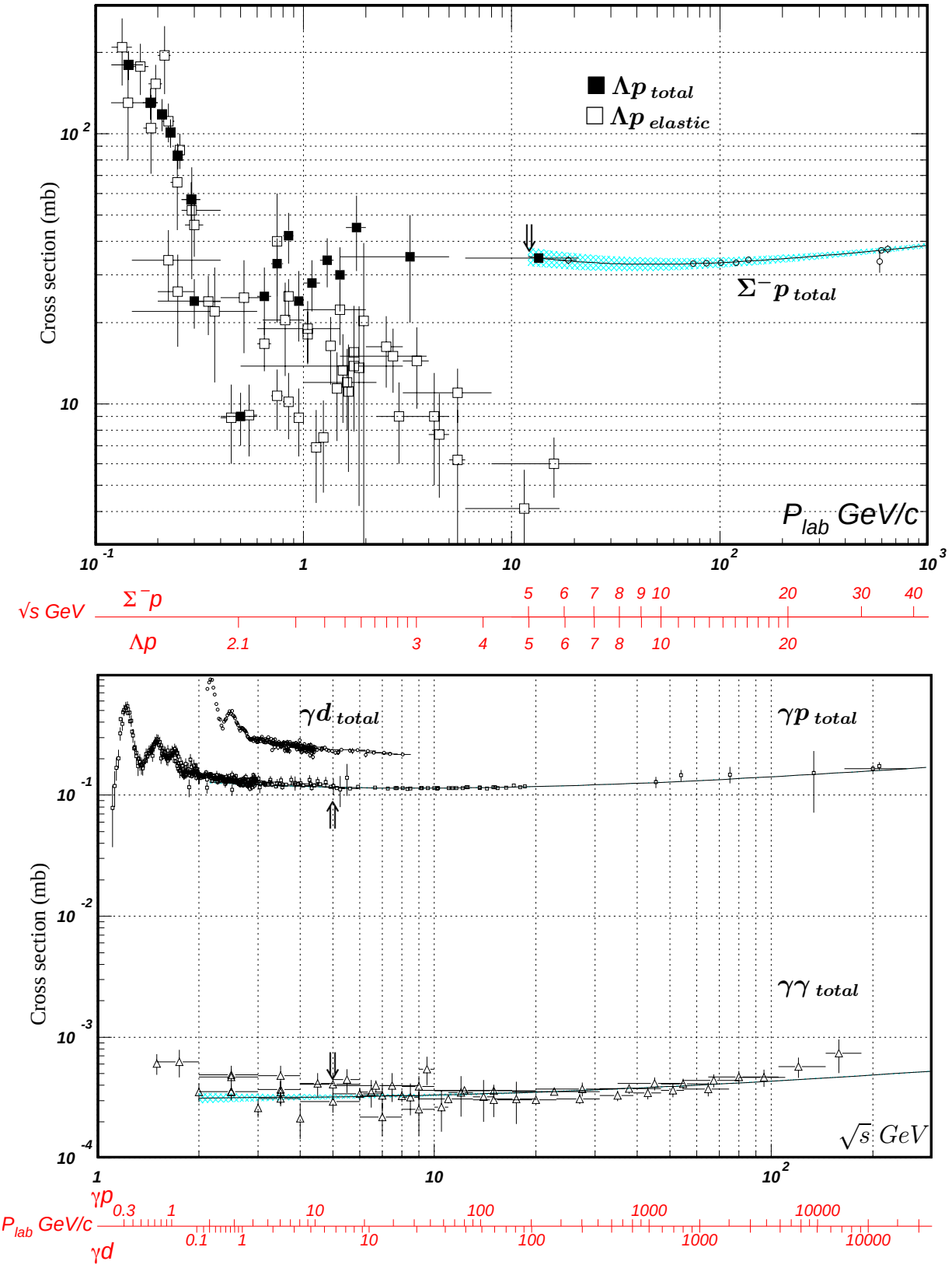


Figure 40.16: Total and elastic cross sections for Λp, total cross section for $\Sigma^- p$, and total hadronic cross sections for γd, γp, and $\gamma\gamma$ collisions as a function of laboratory beam momentum and the total center-of-mass energy. Corresponding computer-readable data files may be found at `http://pdg.lbl.gov/xsect/contents.html`. (Courtesy of the COMPAS group, IHEP, Protvino, August 2005.)